A BIDIRECTIONAL LMS ALGORITHM FOR ESTIMATION OF FAST
TIME-VARYING CHANNELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YAVUZ YAPICI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

MAY 2011

Approval of the thesis:

A BIDIRECTIONAL LMS ALGORITHM FOR ESTIMATION OF FAST TIME-VARYING CHANNELS

submitted by **YAVUZ YAPICI** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering**

Assoc. Prof. Dr. A. Özgür Yılmaz
Supervisor, **Electrical and Electronics Engineering**

Examining Committee Members:

Prof. Dr. Yalçın Tanık
Electrical and Electronics Engineering, METU

Assoc. Prof. Dr. A. Özgür Yılmaz
Electrical and Electronics Engineering, METU

Assoc. Prof. Dr. Emre Aktaş
Electrical and Electronics Engineering, Hacettepe University

Prof. Dr. T. Engin Tuncer
Electrical and Electronics Engineering, METU

Assoc. Prof. Dr. Çağatay Candan
Electrical and Electronics Engineering, METU

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: YAVUZ YAPICI

Signature   :

ABSTRACT

A BIDIRECTIONAL LMS ALGORITHM FOR ESTIMATION OF FAST
TIME-VARYING CHANNELS

Yapıcı, Yavuz

Ph.D., Department of Electrical and Electronics Engineering

Supervisor   : Assoc. Prof. Dr. A. Özgür Yılmaz

May 2011, 186 pages

Effort to estimate unknown time-varying channels as a part of high-speed mobile communication systems is of interest especially for next-generation wireless systems. The high computational complexity of the optimal Wiener estimator usually makes its use impractical in fast time-varying channels. As a powerful candidate, the adaptive least mean squares (LMS) algorithm offers a computationally efficient solution with its simple first-order weight-vector update equation. However, the performance of the LMS algorithm deteriorates in time-varying channels as a result of the eigenvalue disparity, i.e., spread, of the input correlation matrix in such channels.

In this work, we incorporate the LMS algorithm into the well-known bidirectional processing idea to produce an extension called the *bidirectional LMS*. This algorithm is shown to be robust to the adverse effects of time-varying channels such as large eigenvalue spread. The associated tracking performance is observed to be very close to that of the optimal Wiener filter in many cases and the bidirectional LMS algorithm is therefore referred to as near-optimal. The computational complexity is observed to increase by the bidirectional employment of the LMS algorithm, but nevertheless is significantly lower than that of the optimal Wiener filter.

The tracking behavior of the bidirectional LMS algorithm is also analyzed and eventually a steady-state step-size dependent mean square error (MSE) expression is derived for single-antenna flat-fading channels with various correlation properties. The aforementioned analysis is then generalized to include single-antenna frequency-selective channels where the so-called independence assumption is no more applicable due to the channel memory at hand, and then to multi-antenna flat-fading channels. The optimal selection of the step-size values is also presented using the results of the MSE analysis. The numerical evaluations show a very good match between the theoretical and the experimental results under various scenarios. The tracking analysis of the bidirectional LMS algorithm is believed to be novel in the sense that although there are several works in the literature on the bidirectional estimation, none of them provides a theoretical analysis on the underlying estimators.

An iterative channel estimation scheme is also presented as a more realistic application for each of the estimation algorithms and the channel models under consideration. As a result, the bidirectional LMS algorithm is observed to be very successful for this real-life application with its increased but still practical level of complexity, the near-optimal tracking performance and robustness to the imperfect initialization.

Keywords: Least mean squares (LMS), channel estimation, low-complexity, mean square error (MSE), optimal filter.

ÖZ

ZAMANLA HIZLI DEĞİŞEN KANALLARIN KESTİRİMİ İÇİN İKİ YÖNLÜ LMS ALGORİTMASI

Yapıcı, Yavuz

Doktora, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Yöneticisi   : Doç. Dr. A. Özgür Yılmaz

Mayıs 2011, 186 sayfa

Yüksek hızlı mobil iletişim sistemlerinin bir bileşeni olan zamanla değişen haberleşme kanallarının kestirimi, özellikle yeni nesil kablosuz sistemlerin geliştirilmesi bakımından ilgi çekmektedir. En iyi Wiener kestiricisinin yüksek hesaplama karmaşıklığı, bu süzgecin zamanla hızlı değişen kanallarda kullanımını elverişsiz hale getirmektedir. Güçlü bir alternatif olan uyarlamalı en küçük ortalama kareler (LMS) algoritması ise, sahip olduğu basit birinci-derece ağırlık-vektörü güncelleme denklemi ile hesaplama karmaşıklığı bakımından oldukça etkin bir çözüm sunmaktadır. Fakat, zamanla değişen kanallarda girdi korelasyon matrisine ait özdeğerlerin birbirinden farklılığı sebebiyle LMS algoritmasının bu tür kanallardaki başarımı kötüleşmektedir.

Bu çalışmada, LMS algoritmasını iyi bilinen iki yönlü işleme fikriyle biraraya getirerek iki yönlü LMS algoritmasını ürettik. Bu algoritmanın, zamanla değişen kanallarda bulunan yüksek özdeğer farklılığı gibi olumsuz etkilere karşı gürbüz olduğu gösterilmiştir. Sonuçtaki takip performansı, en iyi Wiener süzgecininkine oldukça yakın kalmakta ve bu sebeple de en iyinin yakınında olarak ifade edilmektedir. Hesaplama karmaşıklığının, LMS algoritmasının iki yönlü kullanımı ile arttığı gözlense de bu sonuç en iyi Wiener süzgecininkine göre oldukça

düşük ve klasik LMS algoritması ile de benzer düzeydedir.

İki yönlü LMS algoritmasının takip davranışı ayrıca analiz edilmiş ve sonuçta tek-antenli düz-sönümlemeli kanallarda çeşitli korelasyon özellikleri için basamak-büyüklüğüne bağlı yatışkın-durum ortalama kare hatası (MSE) ifadesi elde edilmiştir. Bahsedilen analiz daha sonra, kanal hafızasından dolayı bağımsızlık varsıyımının uygulanabilir olmadığı tek-antenli frekans-seçici kanallar ve çok-antenli düz-sönümlemeli kanalları da içerecek şekilde genelleş-tirilmiştir. Basamak-büyüklüğünün en iyi seçimi de, elde edilen MSE analiz sonuçları kul-lanılarak verilmiştir. Sayısal hesaplamalar, teorik ve deneysel sonuçların çok çeşitli senar-yolar altında iyi bir uyum sergilediğini göstermiştir. Literatürde iki yönlü kestirme konusunda çeşitli çalışmalar bulunsa da, bu çalışmalarda ele alınan kestiriciler için herhangi bir teorik analiz yapılmamıştır. Bu yönüyle, iki yönlü LMS algoritması için sunulan takip analizinin yenilikçi olduğu düşünülmektedir.

Ayrıca, yinelemeli bir kanal kestirim düzeni, ele alınan bütün kanal modelleri ve kestirim algoritmaları için sunulmuştur. Sonuç olarak, iki yönlü LMS algoritmasının ele alınan bu gerçek-hayat uygulaması için gerek artmış fakat yine de pratik karmaşıklık seviyesi, en iyinin yakınındaki takip performansı ve hatalı ilklendirmeye karşı gürbüzlüğü ile oldukça başarılı olduğu gözlemlenmiştir.

Anahtar Kelimeler: En küçük ortalama kareler (LMS), kanal kestirimi, düşük karmaşıklık,ortalama kare hatası (MSE), en iyi süzgeç.

To my dear family;
My parents Emine and Durmuş,
My beloved Songül and my honey daughter İclal Gökçe...

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

The ultimate goal in many communication systems is to design an optimal receiver to minimize the probability of detection error. Most of these receivers require the knowledge of the communication channel under consideration in order to perform coherent demodulation and detection, and for some other tasks including frame synchronization and decoding. Towards this end, inserting some amount of a priori known pilot symbols into the transmitted data sequence to be processed by various channel estimation algorithms at the receiver appears to be a practical solution. However, as emerging communication technologies pushed by the end-user demands involve fast time-varying channels, the conventional approaches result in an excessive computational complexity in obtaining an accurate estimate of the unknown channel. With this motivation, the efficient channel estimation algorithms in the mean square error (MSE) sense are therefore the subject of this work assuming a packet-based transmission scheme in which neither the transmitter nor the receiver has the exact knowledge of the communication channel in use.

1.1 Motivation

Although the well-known adaptive least mean squares (LMS) algorithm suggests a practical way of estimating unknown channels in any communication system, the associated performance over time-varying channels are known to be far behind that of the optimal Wiener filter especially as the speed of time-variations increases. The main reason behind this degradation is the large eigenvalue spread of the input correlation matrix for fast time-varying channels [1]. This motivates us to explore for a suitable adaptive algorithm as an extension of the conventional LMS algorithm which will be robust to adverse effects of fast fading chan-

nels such as the large eigenvalue disparity and hopefully achieve a significantly improved performance yet at a still practical level of complexity as compared to the original algorithm as well as to the optimal Wiener filter.

In the literature, there are several works on the forward-backward signal processing techniques applied to communication problems with a promise of improved overall performance. In [2], a forward-backward LMS (FBLMS) adaptive line enhancer is proposed for stationary systems which makes use of the forward and the backward prediction errors jointly to update the weight-vector which eventually achieves a lower misadjustment. This algorithm is further elaborated in [3] which establishes the same performance with a less computational burden. In [4], a different approach is preferred in which the adaptations are performed in the forward and the backward directions independently along each of the paths present in the trellis using a per-survivor processing (PSP) based approach [5, 6]. These estimates are then combined using some optimal binding strategies for which the final performance improvement is significant, but unfortunately with an excessively large overall processing complexity .

In this work, we benefit from the aforementioned works from the literature, and offer to operate the conventional LMS algorithm in both the forward and the backward directions independently along a transmitted block assuming a packet-based transmission scenario. The resulting forward and backward estimates are then combined in a suboptimal way to produce the final channel coefficient estimate in a computationally efficient manner. We call this algorithm the *bidirectional LMS* and abbreviate as BiLMS whereas the conventional LMS algorithm is referred to as the *unidirectional LMS* throughout this work with the abbreviation UniLMS. Through computer simulations, the bidirectional LMS algorithm is shown to achieve a superior performance over the unidirectional LMS algorithm in fast time-varying channels at an increased but still practical level of computational complexity. The resulting performance is even very close to that of the optimal Wiener filter in most cases and is therefore referred to as near-optimal. Note that the optimality of the Wiener filter under consideration is in the linear minimum mean-squared error (MMSE) sense so that MMSE and Wiener terms are used interchangeably in this work to specify the desired filter.

Having introduced the bidirectional LMS algorithm, it is of significant interest to analyze the resulting performance from both practical and theoretical perspectives. We therefore investigate the theoretical foundations of the steady-state tracking characteristics of the bidirectional

LMS algorithm and finally come up with a mean square error (MSE) expression which is verified over a number of channel models including flat-fading, frequency-selective fading and multi-input multi-output (MIMO) fast time-varying channels with various Doppler spectrums. We have also realized that although there are some relevant work present in the literature as summarized above, none of them concerns with the theoretical analysis of the adaptive algorithms under consideration. Therefore, our work not only considers a practical extension of the LMS algorithm with near-optimal performance but also presents a novel theoretical analysis for the bidirectional LMS algorithm in time-varying channels with a satisfactorily high accuracy.

1.2 Thesis Overview

We will present a detailed literature survey in Chapter 2 on the time-varying channel estimation problem. It considers the basics of the transmission schemes for the purpose of channel estimation including pilot-based and (semi-)blind schemes together with superimposed training and the PSAM approach. Adaptive algorithms, and particularly the LMS algorithm, are introduced with the resulting complexity benefits and performance issues. And lastly, the concept of iterative channel estimation is introduced with the considerations on some recent papers.

Chapter 3 considers the bidirectional LMS algorithm in time-varying flat-fading channels. The tracking behavior of the bidirectional LMS algorithm is analyzed and a novel steady-state MSE expression is obtained which includes the effect of the adaptation step-size and is valid for a variety of correlation characteristics. The optimal selection for the forward and the backward step-size values is also investigated together with the effect of joint Doppler and SNR estimation and imperfect initialization. The theoretical results are observed to match the experimental ones very well both of which exhibit a significant improvement in the tracking performance. The analysis is then revisited with an assumption of a special correlation characteristics specified by an auto-regressive (AR) process and a more compact steady-state MSE expression is obtained by eliminating the necessity for numerical integration methods. Finally, a coded communication system is considered in which coherent detection is performed using the estimate of the unknown time-varying channel refined through iterations. The bidirectional LMS algorithm is shown to offer a practical solution for this scheme, as well, with

its computationally efficient structure and near-optimal tracking performance.

The findings for the time-varying flat-fading channels are then generalized to frequency-selective channels in Chapter 4. The estimation of frequency-selective channels is usually challenging since the parameter to be estimated at each time instant is a vector composed of the associated channel taps. The bidirectional LMS algorithm which updates the estimate of the channel vector at each time epoch proposes a very good tracking performance at still low level of complexity, as for the flat-fading case. The tracking behavior of the bidirectional LMS algorithm is also investigated together with a discussion about inapplicability of the so-called independence assumption to frequency-selective channels. The performance of the algorithm is verified also in a coded system over a frequency-selective channel with various power delay profiles in which the unknown channel is estimated in an iterative fashion.

Finally, Chapter 5 deals with a multi-antenna system in which subchannels between each of the transmitter and the receiver antenna pair experience time-varying frequency-flat fading. The optimal Wiener filter has a high complexity in such channels which appreciates the improved performance of the bidirectional LMS algorithm at a practical level of complexity. A steady-state tracking analysis is again provided which results in a good match with the experimental data in many cases, and some tolerable differences for the rest. Finally, we summarize the contributions of the thesis and mark some topics as future work in Chapter 6.

CHAPTER 2

THE TIME-VARYING CHANNEL ESTIMATION

User mobility in current and emerging communication technologies has changed the focus to time-varying communication channels. Although time-invariant channels are well-understood from channel estimation point of view, the associated processing complexity makes the time-varying counterparts really challenging. This work therefore considers low-complexity adaptive channel estimation algorithms with improved estimation and tracking performances.

This chapter is devoted to a literature survey in detail on the basics of the problem of time-varying channel estimation. Training-based and blind channel estimation methods, adaptive algorithms and particularly the conventional LMS algorithm and finally iterative channel estimation approach are some of the subjects covered in this chapter.

2.1 Transmitter and Receiver Structures for Time-Varying Channels

One of the well-known transmission schemes for the purpose of channel estimation is called the *pilot-* or *training-based* transmission in which a set of a priori known pilot symbols are transmitted along with the information-bearing data sequence. The pattern in which pilots are used in transmission heavily depends on the channel characteristics under consideration. In some communication systems of time-varying nature, the variation is relatively small as compared to the transmission intervals so that the channel is referred to as quasi-static or piecewise time-invariant. For such channels, the pilot-based transmission is employed in a *train-before-transmit* scheme [7] where all the available pilots are located at the beginning of the transmitted block. Although this scheme is suitable for quasi-static channels such as the one assumed in global system of mobile communications (GSM) protocol [8], which results

in an overhead of 22.4% on the overall system throughput [9], it is not efficient for fast time-varying channels.

The simplest alternative solution for time-varying channels is to repeat the training blocks frequent enough [10]. However, this strategy is known to be usually insufficient to track the dynamics of the underlying channel and is severely bandwidth consuming. Instead of periodic transmission of the short training sequences, distribution of pilot symbols along the transmitted block is a much more efficient way to jointly estimate and track time-varying channels. In the literature, the periodic insertion of single pilot symbols into the transmitting sequence is called *pilot symbol assisted modulation* (PSAM) [11, 12, 13]. A theoretical performance analysis for PSAM transmission is provided in [14] and [15] over flat-fading and frequency-selective fading channels, respectively.

Another well-known pilot-based transmission scheme is the *superimposed training* in which an a priori known pilot sequence with a relatively low power is superimposed on the information-bearing data sequence. The superimposed channel estimation, which is also called the *spread spectrum pilot technique* in [16], is first offered by [17] and further elaborated in some subsequent papers including [16, 18]. In [19], the superimposed approach is theoretically analyzed in terms of overall estimation performance, power allocation and frame synchronization. A blind estimation algorithm independent of the channel characteristics and modulation format is proposed in [20] where the superimposed pilot sequence is referred to as "hidden". In order to lower the correlation between the pilot and the data sequences, a different method based on selective usage of the superimposed sequence from a larger set is introduced in [21]. Recently, the superimposed training based channel estimation technique is further investigated using exponential basis models [22], for a multi-antenna channel estimation and symbol detection problem [23], for an iterative joint channel estimation and data detection problem with analytical performance results [24], and finally with an important comparison between the pilot-based training for next generation terrestrial digital video broadcasting (DVB-T) technique using orthogonal frequency division multiplexing (OFDM) [25] to circumvent the frequency selectivity of the underlying communication channel.

The optimal choice of pilot symbols to be transmitted is of great importance in order to improve the overall performance of the pilot-based approaches. The optimal pilot design for PSAM-based transmission over time-invariant flat-fading, frequency-selective fading and

6

MIMO channels are considered in [26] by obtaining a Cramer-Rao Bound (CRB) which is independent of the channel estimator in use. This investigation is further extended to include time-varying channels in [27]. In [28], a periodic pattern for the superimposed pilot sequence is proposed by deriving the corresponding CRB. An interesting alternative technique is presented in [29] where a more general superimposed pilot sequence is shown to be possible through the utilization of underlying cyclostationary nature. A rich survey on pilot-based transmission schemes is presented in [30] with more details for interested readers.

Blind algorithms constitute another class of strategy to estimate the unknown communication channel without any need for pilot symbols. In the blind channel estimation, which is also called *train-while-transmit* scheme in [7], the channel statistics and therefore the corresponding estimate are obtained directly from observations. By this way, the throughput overhead due to the pilot transmission present in the training-based approach is saved for some other useful mechanisms such as channel coding [31]. This discussion implies that the bandwidth available for communication is not wasted to pilot transmission in the blind channel estimation. In the literature, blind channel estimation algorithms are reported to achieve as low as a 1 dB mean SNR loss for a GSM data transmission over both fast and slowly fading mobile channels [31]. Beside this performance loss, blind channel estimation algorithms are also known to suffer from a high computational complexity such that the aforementioned result reported in [31] is achieved by using the fourth order statistics of the underlying communication channel. In order to have reliable higher order statistics experimentally, blind algorithms have to wait for sufficient number of observations which tends to loose their practical value. The slow convergence rate and possibility of convergence to the wrong solution are also argued to be some of the other major drawbacks of blind algorithms [32].

For interested readers, an initial investigation into the general channel equalization problem in a blind fashion without resorting to a training sequence is considered in two seminal papers [33, 34] followed by the work [35] in which a blind correction of any channel impairments is studied. Another survey of interest on blind processing is presented in [36] and the concept is reconsidered from signal processing view referred to as blind signal separation and estimation in [37]. In [38], a wideband communication channel is considered together with OFDM signaling where the unknown channel is estimated blindly by making use of the cyclic-prefix (CP) in subspace-based methods along with training-based semi-blind algorithms. In a recent paper [39], performance of the semiblind channel estimation algorithm is analyzed

in a code-division multiple access (CDMA) system and the superiority of the subspace-based methods is reported for some certain cases with large channel orders.

The channel estimation problem for MIMO systems is another research area of great importance in the sense that the unknown coefficient set is in a matrix-form which results in some computational difficulties[40]. The initial studies in MIMO channel estimation problem commonly assume quasi-static communication channel [41]. As the coherence time becomes smaller as a result of increasing end-user demands, the variation over a transmitted block could not be ignored any more and the channel model in hand turns out to be time-varying. For such continuous fading channels, Kalman-based approaches [42] are employed by making use of the statistics of the underlying channel. In [43, 44], a time-varying flat-fading MIMO system is considered with maximum likelihood (ML) and interpolation based channel estimators in which short training sequences are transmitted periodically. It is argued in these work that although tracking significantly improves the quality of the channel estimate, the resulting complexity makes this choice really impractical as is also stated in [45].

There is a rich literature in the area of MIMO channel estimation problem for both flat- and frequency-selective fading channels examples of which are summarized in the following. A superimposed type channel estimation approach is presented in [46] for a quasi-static Rayleigh fading MIMO channel where low-power pilots are transmitted continuously along with the data symbols. In [47], a spatially correlated MIMO channel is considered from optimal transmitted signal design point of view with the assumption of perfect knowledge of the second order channel statistics at the receiver. A theoretical study is carried on by [48] in which a lower bound on the the error correlation matrix of training-based channel estimators for quasi-static frequency-selective MIMO channels. The parameter estimation problem including the unknown channel coefficients for a flat-fading MIMO system in the presence of frequency offset is considered in [49] in a generalized fashion such that each of the transmitter antennas experiences independent frequency offset levels. The expectation-maximization (EM) algorithm is employed in [50] in order to estimate the unknown frequency-selective fading MIMO channel where the order of the underlying multipath channel is estimated using the conditional model order estimator (CME). In a recent work [51], LS (Least Squares) and LMMSE (Linear Minimum Mean Square Error) methods are considered for MIMO with some conclusion about the robustness of space-time coding and receiver diversity to the channel estimation error.

From an information theoretic point of view, the achievable rate for a coded modulation flat-fading MIMO system with PSAM based channel estimation is investigated in [52] under a perfect interleaving scenario. The effect of channel estimation error is further investigated in [53] for multiple antenna fading channels with a conclusion of mutual information bounds and proper power allocation schemes. The optimal training sequence for the ML channel estimator for a quasi-static flat-fading MIMO channel is first shown in [41] to be orthogonal across the transmitting antennas and the associated design issues are investigated in detail by [54]. In [55], optimal pilot signal design is addressed for frequency-selective block-fading MIMO-OFDM channels by maximizing a lower bound on the average capacity and the same problem is revisited in [56] with a result of a more general class of optimal pilot symbols by using some basic features of the Discrete Fourier Transform (DFT).

2.2 Adaptive Estimation of Time-Varying Channels

2.2.1 An Introduction to Adaptive Algorithms

Adaptive algorithms have a wide variety of application areas due to their self-learning characterics, computational efficiency and convergence to the optimal non-adaptive solutions. In digital communications, the LMS algorithm is one of the well-known adaptive algorithm that is commonly used in numerous applications including equalization and channel estimation. In order to provide a solid background on the subject, we first talk about the basis of adaptive algorithms by following [57] and then continue with the associated stationary and nonstationary characteristics.

For any adaptive algorithm, an associated error function is defined which is sometimes called the (error) performance surface. For quadratic error functions, the corresponding error performance surface is a bowl-shaped hyper paraboloid where the order of this geometry is determined by the number of weights of the algorithm. The horizontal cross-section of this quadratic performance surface is an ellipse. For such surfaces, there is a single minimum point at the bottom which specifies the minimum mean-square error (MMSE) and the associated projection onto the weight-vector plane is the optimal weight-vector [57].

Most adaptive algorithms try to find the optimal weight-vector solution by seeking the minimum point of the performance surface using gradient-based search algorithms. The gradient

of any point on the performance surface specifies a vector with the direction pointing the greatest rate of increase of the surface and the magnitude which is the greatest rate of change at that point. In real-life problems of interest, such surfaces are not known exactly due to unknown parameters (weights etc.) except for some points on the surface which could be measured or estimated in some ways. The exhaustive search over the performance surface to find the optimal point is therefore not possible for a variety of applications. As such, the adaptive algorithms are of practical interest in that they seek the minimum point of the performance surface by proceeding in small steps towards the direction of the negative of the gradient or a matrix scale of it [57] .

Of gradient-based search algorithms, the Newton's method and the steepest descent are the two well-known examples. Newton's method adapts its weights in the direction of the minimum point of the performance surface, but unfortunately has some practical limitations such as the necessity of the inverse of the input correlation matrix. On the contrary, the steepest descent method is much more practical in that it does not require the knowledge of the input correlation matrix and just progresses in the direction of the negative of the gradient at each step which is not necessarily towards exactly to the minimum point of the performance surface. Although the Newton's method is much faster than the steepest descent algorithm, the slow adaptation of the latter behaves as a low pass filter on the overall estimation noise which diminishes the final misadjustment as compared to the former [57].

2.2.2   The Adaptive LMS Algorithm

Although being more practical in its use than the Newton's method, the steepest descent algorithm has an important restriction in that it still requires the estimate of the gradient to be computed in an off-line fashion for many cases. The LMS algorithm appears first in [58] to overcome this necessity by employing an instantaneous estimate of the gradient in the steepest-descent algorithm rather than a long-term average, and it has been recognized as one of the most practical adaptive algorithms since then [59].

The conventional LMS algorithm in estimating a flat-fading communication channel is given as [57]

$$\hat{f}_{k+1} = \hat{f}_k + 2\mu\, e_k\, a_k \tag{2.1}$$

10

where $\hat{f}_k$ is the estimate of the fading coefficient, a_k is the input symbol, μ is the step-size value of the algorithm and e_k is the estimation error given for this particular case as

$$e_k = y_k - \hat{f}_k \, a_k \qquad (2.2)$$

where y_k is the associated observation symbol. Note that, although the factor '2' in (2.1) could be safely integrated into the step-size value, we prefer to write it separately in order to provide an easy comparison with some previous works in the literature.

Before proceeding further, we would also like to specify the Wiener filter as a counterpart for (2.1) which is given as

$$\hat{f}_k \;\; = \;\; \mathbf{w}_k^T \, \mathbf{y}_k \qquad (2.3)$$

where $\mathbf{y}_k$ is a $K \times 1$ vector including the observations to be used and the $K \times 1$ Wiener filter is given as

$$\mathbf{w}_k = \mathbf{R}_{yy}^{-1} \, \mathbf{P}_{fy} \, . \qquad (2.4)$$

In (2.4), $\mathbf{R}_{yy} \;=\; E\{\mathbf{y}_k \, \mathbf{y}_k^H\}$ is the $K \times K$ autocorrelation matrix and $\mathbf{P}_{fy} \;=\; E\{\mathbf{y}_k \, f_k^*\}$ is the $K \times 1$ cross-correlation vector. Note that, $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ stand for the transpose, complex conjugate and Hermitian operations, respectively, in above equations.

As in the other adaptive algorithms, the function of the LMS algorithm in stationary environments is to *estimate* the unknown quantity whereas this task turns out to first *estimate* and then *track* the variations in the unknown quantity in nonstationary environments. Within the scope of this work, the stationary and nonstationary environments are considered as time-invariant and time-varying channels, respectively. Therefore, the LMS algorithm has to both estimate and track the unknown time-varying channel under consideration. Because a packet-based transmission scheme with short blocklengths is assumed throughout this work, we provide proper initial conditions to the LMS algorithm with varying accuracy to skip the acquisition, i.e., transition, phase at the beginning of the transmission block and directly switch to the tracking phase. Hence, transition behavior of the LMS based algorithms is considered to be beyond the scope of this work while a steady-state tracking performance of these algorithms are investigated in detail here.

Adaptive filters differ from the well-known Wiener filter in that they do not need to know the statistical behavior of the underlying processes which is a basic requirement for the Wiener

filter as in (2.4) [60]. In addition, Wiener estimate of each channel tap requires first a matrix inverse and a matrix multiplication as in (2.4) to compute the optimal filter coefficients and then a vector multiplication as in (2.3) to find the desired estimate unlike the significantly simpler 1-tap LMS update equation. Therefore, although the Wiener filter is optimal in the MSE sense, the adaptive filters including the LMS algorithm are much more practical.

It is shown in [57] that the LMS algorithm will converge in mean to the Wiener solution for stationary inputs except for a misadjustment term. In time-invariant environments with stationary inputs, the rate at which the adaptive filter converges to the Wiener solution is closely related to the step-size value or adaptation constant in more general terms. For large step-size values, the convergence is fast but unfortunately results in a large misadjustment whereas the convergence is slow for small step-size values but hopefully with a small final misadjustment. The eigenvalue spread of the input data covariance matrix appears to be important in the sense that the performance of the conventional LMS algorithm deteriorates over time-varying channels where the eigenvalue spread is large. In [61], the optimal weight vector in time-varying channels is also shown to change with time.

A well-known measure for the estimation and/or tracking quality of the LMS algorithm is the resulting MSE value together with the convergence speed which is of importance for stationary environments. The speed of convergence and the MSE analysis of the LMS algorithm are given first in [61] and then in [62] for both the stationary and nonstationary environments. This work employs the so-called "independence assumption" while deriving the steady-state MSE expression in which the weight-vector is assumed to be statistically independent of the current input data vector. The MSE analysis of the LMS algorithm for nonstationary environments is also presented in [61] through the transfer function method by considering a 1-st order auto-regressive (AR) process with a rational power spectrum. In [63], the analysis is generalized to include any kind of unknown nonstationary processes with a known power spectrum.

In communication applications, the MSE analysis of the LMS algorithm in frequency-selective channels has some differences from those presented in [61] and [63]. In frequency-selective channels, the weight-vector is surely correlated with the input data vector due to the channel memory at hand. The independence assumption therefore does not apply and the previous MSE analysis has to be revisited. The work given in [64] pioneers the studies of the MSE

analysis of the LMS algorithm over vector channels. The scenario in which the elements of the input data vector are also correlated due to some other reasons such as channel coding is considered in [65] and a proper MSE analysis is offered for the LMS algorithm which requires to solve a set of equations in an iterative fashion. There are some other useful works in the literature on the MSE analysis of the LMS algorithm without invoking the independence assumption such as [66, 67, 68] together with analysis of the transient behavior given in [69, 70] and [71] which may be useful for interested readers.

It is important to realize that the performance of adaptive filters is subject to change as the underlying channel models also change. An interesting example to this issue is that although the recursive least squares (RLS) algorithm is known to exhibit a faster convergence rate than the LMS algorithm for time-invariant channels, the LMS algorithm is reported to achieve a superior tracking performance than the RLS algorithm in time-varying channels [72]. The reason behind this somewhat surprising result is about the foundations of these algorithms. Although both of them are of adaptive nature, the LMS originates from a statistical model in the sense that it adapts itself according to the gradient search based steepest descent algorithm whereas the RLS algorithm has a deterministic nature such that it derives a filter to minimize the sum of squares of a given prediction error. For this reason, the RLS algorithm is sometimes classified as "model dependent" which is the main cause for performance degradation observed in time-varying environments. In the literature, there are various attempts to modify the conventional RLS algorithm to make it optimum for time-varying systems, e.g., the extended RLS (ERLS) [73] and optimal exponential factor selection in the context of adaptive antenna arrays [74]. From a different perspective, it is argued in [72] and [73] that the convergence is a transient phenomenon whereas the tracking is a steady-state behavior. Therefore, the adaptive algorithms with an appealing convergence behavior in nonstationary environments, such as the RLS algorithm, do not necessarily exhibit a satisfactory tracking performance in nonstationary environments. This is one of the reasons why the LMS algorithm becomes the subject of this work.

Finally, an interesting application of the adaptive LMS algorithm is in the area of equalization in the presence of a temporally colored narrowband interferer. In usual equalization problems with or without a white narrowband interference, the performance of the LMS algorithm is considered to be bounded by that of the optimal Wiener filter. However, this way of analysis is shown in [75] to be incorrect when the interferer has a correlation along the time axis which

reveals the nonlinear nature of the LMS algorithm. Furthermore, the MSE performance of the LMS algorithm is unexpectedly shown to be better than that of the Wiener filter [75] for such situations which is referred to in the literature as a non-Wiener effect of the LMS algorithm [76]. Indeed, the basic problem with the conventional way of MSE analysis for the LMS algorithm in such situations lies in the so-called independence assumption which cannot be invoked since the LMS weight-vector and the input data, i.e., observations, to be processed become strongly correlated as the bandwidth of the interferer gets smaller than that of the signal and additive noise. In order to analyze the MSE performance of the LMS algorithm in the presence of a temporally correlated interferer, a transfer filter approach is presented as a combination of steady-state and time-varying filters in [75], and the subject is further investigated in [77] by deriving a proper bound. A similar study is carried on in [78] to perfectly analyze the nonlinear characteristics of the multi-step LMS algorithm from the adaptive prediction point of view as a different application area, and with a practical two-stage structure proposal for the decision-feedback equalization (DFE) problem in [79].

2.3 Iterative Channel Estimation for Time-Varying Channels

Iterative processing has one of its major application areas in estimation of unknown communication channels. The algorithm called the *iterative channel estimation* is first proposed by [80] and further elaborated in [81, 82] and [83] with a promise of a significant improvement on the quality of the channel estimate.

In training-based transmission, the number of pilots in use has a major effect on the estimation quality which is limited in many applications due to the short blocklength or insufficient training length to avoid excess pilot overhead on the transmission throughput. In iterative channel estimation, decisions on the data symbols are also employed in the receiver after an initial coarse channel estimation using the pilot symbols only. As a result, the overall quality of the channel estimate improves over iterations. The decisions on the data symbols are produced using various mechanisms such as a soft-decoder, and are fed back to the estimator unit in either hard or soft fashion.

In [84], the well-known PSAM transmission technique is employed over a slow flat-fading channel together with Wiener and moving average (MA) channel estimators. A time-invariant

frequency-selective channel is considered in [85] together with an iterative channel estimation scheme. An interesting investigation is given in [86] in which it is argued that the use of soft statistics cannot always result in a better channel quality over the time-invariant channels. Iterative channel estimation is revisited for time-invariant and time-varying frequency-selective channels in [87] and [88], respectively, and the LMS algorithm with soft feedback is decided to perform the closest to the optimal performance as compared to the various scenarios including the RLS algorithm and hard feedback. As a different application, a new ISI and multiple-access interference (MAI) equalization scheme is proposed in [89] together with a properly adjusted iterative channel estimation method.

There are various work in the literature on the low-complexity algorithms for the iterative channel estimation. Among these attempts, complexity is still a problem in [90] and [91] due to the PSP-originated ideas. The Kalman filter is employed in both [92] and [93], however the resulting complexity is declared there to be several times greater than that of the LMS algorithm. In addition, a bidirectional strategy is followed also in [94] over a trellis structure again together with a high computational complexity.

In a recent paper [95], the Gaussian message passing (GMP) algorithm [96] is considered for channel estimation in a bidirectional manner. The work in [95] prefers to express the underlying channel using a 1-st order AR process having a rational power spectrum in order to be able to express the iterative process using a Forney style graph [97] in accordance with the GMP algorithm without any theoretical analysis. If the correlation property is different than what is assumed, e.g., such as the one in the Jakes' model [98], then a serious mismatch naturally appears as an important weakness of the proposed algorithm. Although the order of complexity is discussed to be linear with the number of channel taps, the number of complex multiplications and additions required to estimate even a single channel tap is much greater than that of the bidirectional LMS algorithm.

2.4   Conclusion

In training-based schemes, estimation of time-varying channels requires more processing power and number of pilot symbols as compared to the time-invariant scenarios. As the underlying channel turns out to be frequency-selective and/or multiple antennas are used at

15

both terminals, the overall complexity and necessity to more pilot symbols increase even more since the channel to be estimated becomes a vector or a matrix, respectively. The efficient use of the available pilots is therefore of interest and the transmission technique known in the literature as PSAM becomes a powerful choice in joint estimation and tracking of unknown channels. This is the reason why we prefer the PSAM technique in real-life scenarios of the subsequent chapters.

The optimal Wiener filter in estimation of the unknown time-varying channel has an excessive computational complexity. The adaptive LMS algorithm therefore appears to be an efficient solution in time-varying channels with its simple first-order weight-vector update equation. However, the eigenvalue spread of the input correlation matrix in time-varying channels result in a performance degradation for the LMS algorithm. This work, therefore, presents a much more robust extension of the LMS algorithm called the bidirectional LMS together with a detailed analysis. Finally, the iterative channel estimation technique offers a significant improvement in the channel estimate by incorporating the soft estimates of the data symbols as well as the pilot symbols. It is shown in the subsequent chapters that the overall complexity of the iterative channel estimation due to recurring iterations decreases significantly by employing the low-complexity bidirectional LMS algorithm with almost no significant change in the final estimation quality under many scenarios.

CHAPTER 3

THE BIDIRECTIONAL LMS ALGORITHM FOR FLAT-FADING CHANNELS

In order to deal with the high processing complexity present in the estimation of time-varying communication channels, the bidirectional LMS algorithm is considered in [99] and [100] for time-varying single-antenna and multi-antenna systems, respectively, with a promise of near-optimal tracking performance over fast time-varying channels at a practical level of complexity. The tracking performance of the bidirectional LMS algorithm is also shown to be very close to that of the optimal Wiener filter and is remarkably better than that of the conventional unidirectional LMS algorithm.

This chapter considers time-varying flat-fading channels as a simple yet sufficient framework on which the bidirectional LMS algorithm is introduced and the associated tracking performance is analyzed. A steady-state MSE analysis for the bidirectional LMS algorithm is performed over Rayleigh fading channels with both rational and nonrational power spectrums. An analysis on the step-size optimization for the bidirectional LMS algorithm is also provided to enable the use of the best step-size value prior to transmission for a given channel. The iterative channel estimation idea is also considered as an application close to the real-life scenarios for which the bidirectional LMS algorithm is employed to estimate coded flat-fading time-varying channels.

3.1   System Model for Flat-Fading Channels

We consider a time-varying flat-fading communication channel which is represented by a set of complex fading coefficients $\{f_k\}_{k=1}^{L}$ where L is the observation length. The fading coeffi-

cients are assumed to have unity variance, i.e., $E\{|f_k|^2\} = 1$, and to be unknown at either the transmitter and the receiver. A sequence of independent data symbols $\{a_k\}_{k=1}^L$ with symbol energy $E_s = E\{|a_k|^2\}$ is formed using a finite alphabet $\mathcal{A}$, and is transmitted through the channel. The corresponding output symbols $\{y_k\}_{k=1}^L$ are observed according to the following equivalent discrete-time complex baseband channel model given as

$$y_k = f_k a_k + n_k \tag{3.1}$$

where n_k is a sample from a circularly symmetric white complex Gaussian process with zero-mean and variance N_0. We assume perfect synchronization in the sense that the timing of the transmitted block and each of the transmitted symbols are known perfectly. Furthermore, any frequency offset due to the imperfect knowledge of the carrier frequency is considered to be beyond the scope of this work, as (3.1) implies.

The received signal-to-noise ratio (SNR), denoted as γ_r, is given as

$$\gamma_r = \frac{E\{|f_k a_k|^2\}}{E\{|n_k|^2\}} = \frac{E\{|f_k|^2\} E\{|a_k|^2\}}{E\{|n_k|^2\}} = \frac{E_s}{N_0} \tag{3.2}$$

and the information symbol SNR is therefore $\gamma = \gamma_r/R$ where R is the overall transmission rate of the communication system including the effect of channel codes, pilot symbols and other relevant techniques in use.

3.2 Estimation Algorithms for Flat-Fading Channels

In this section, we will give basic definitions of the conventional MMSE and the unidirectional LMS channel estimation algorithms for the flat-fading channel model given in Section 3.1. As an extension, the bidirectional LMS algorithm is also introduced and corresponding complexities are compared.

3.2.1 The MMSE Channel Estimation

In MMSE channel estimation, the ultimate goal is to design an FIR filter which produces an estimate of the unknown fading coefficient f_k by filtering a set of noisy observations such that the resulting mean square error is minimized [101]. This problem is illustrated in Fig.3.1 where $W_k(z)$ is the desired FIR filter associated with the k-th fading coefficient, $\hat{f}_k$ is the

associated estimate of f_k, and e_k is the estimation error defined to be $e_k = f_k - \hat{f}_k$ for this particular case.
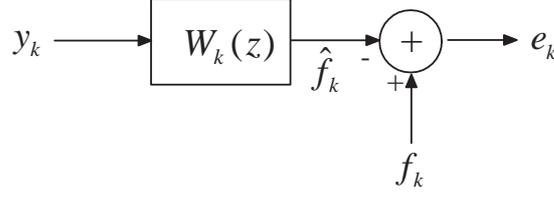
Figure 3.1: The general MMSE channel estimation problem for flat-fading channels.

We assume that the desired MMSE filter $W_k(z)$ for the k-th fading coefficient is a K-tap linear transversal filter which is given as

$$W_k(z) = \sum_{n=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,n} \, z^{-n} \tag{3.3}$$

where $\lfloor \cdot \rfloor$ describes the floor operation which returns the greatest integer smaller than or equal to its argument. The estimate of the k-th fading coefficient is then accordingly computed as

$$\hat{f}_k = \sum_{n=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,n} \, y_{k-n} \tag{3.4}$$

$$= \mathbf{w}_k^T \, \mathbf{y}_k \tag{3.5}$$

where $\mathbf{w}_k = [w_{k,-\lfloor K/2 \rfloor} \; \ldots \; w_{k,\lfloor K/2 \rfloor}]^T$ and $\mathbf{y}_k = [y_{k+\lfloor K/2 \rfloor} \; \ldots \; y_{k-\lfloor K/2 \rfloor}]^T$ are the tap-weight and the observation vectors, respectively.

The optimal tap-weight vector that minimizes the MSE associated with this problem could be computed using the well-known Wiener-Hopf's equations which are derived in Appendix A.1 and given as

$$\sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,l} \, r_{yy} \, (n - l) = r_{fy} \, (n), \tag{3.6}$$

for $n = -\lfloor K/2 \rfloor, \ldots, \lfloor K/2 \rfloor$. In (3.6), $r_{yy} \, (n - l) = E\left\{ y_{k-l} \, y_{k-n}^* \right\}$ is the the autocorrelation of observations and $r_{fy} \, (n) = E\left\{ f_k \, y_{k-n}^* \right\}$ is the cross-correlation between fading coefficients and observations. The Wiener-Hopf's equations for this problem could also be expressed in vector form as follows

$$E\{\mathbf{y}_k \, \mathbf{y}_k^H\} \, \mathbf{w}_k = E\{\mathbf{y}_k \, f_k^*\}, \tag{3.7}$$

19

or equivalently

$$\mathbf{w}_k = \mathbf{R}_{yy}^{-1} \mathbf{P}_{fy}, \tag{3.8}$$

where $\mathbf{R}_{yy} = E\{\mathbf{y}_k \mathbf{y}_k^H\}$ is the $K \times K$ autocorrelation matrix and $\mathbf{P}_{fy} = E\{\mathbf{y}_k f_k^*\}$ is the $K \times 1$ cross-correlation vector.

In order to evaluate the complexity of the MMSE channel estimation, one should consider both the pre-filtering and the filtering stages. In the pre-filtering stage, optimal coefficients of the desired MMSE filter should be computed by using (3.8) which requires a matrix inversion of complexity $O(K^3)$ and a matrix multiplication of complexity $O(K^2)$. In the filtering stage, the desired estimate is obtained according to (3.4) which results in $(K - 1)$ complex additions and K complex multiplications.

In a packet-based transmission, it is sometimes possible to reduce the overall computational complexity associated with the pre-filtering stage under certain circumstances. As an example, if the alphabet $\mathcal{A}$ has equal-energy symbols, e.g., M-ary PSK, and the symbols corresponding to the taps of the estimation filter are known perfectly, then it is sufficient to compute a single global MMSE filter for each transmitted packet. Otherwise, $\mathbf{R}_{yy}$ and $\mathbf{P}_{fy}$ generally depend on the transmitted symbols and should therefore be recomputed for each time epoch k. In Section 3.8, we make use of this simplification for the initial channel estimation and show that the MMSE filter has to be recomputed for each time epoch k in the subsequent iterations since the data symbols corresponding to filter taps are not known perfectly.

3.2.2   The Unidirectional and The Bidirectional LMS Algorithms

The conventional unidirectional LMS algorithm is known to be one of the simplest adaptive algorithm. For the flat-fading channel model introduced in Section 3.1, the unidirectional LMS algorithm is given as [57]

$$\hat{f}_{k+1} = \hat{f}_k + 2\mu e_k a_k \tag{3.9}$$

where μ is the step-size value of the algorithm and e_k is the estimation error given for this particular case as

$$e_k = y_k - \hat{f}_k a_k. \tag{3.10}$$

We observe that the unidirectional LMS algorithm requires 2 complex additions and 3 complex multiplications in order to estimate a single fading coefficient and that there is no pre-filtering stage as for MMSE channel estimation.

Despite its simplicity, the conventional unidirectional LMS algorithm suffers from any time variation present in the unknown channel. We therefore explore an extension of this adaptive algorithm with improved tracking performance over fast time-varying channels at a practical level of complexity. Towards this end, we consider the bidirectional LMS algorithm which is basically an application of the two-way processing technique in which the conventional LMS algorithm is employed in both the forward and the backward directions along a data block.

In order to formulate the bidirectional LMS algorithm, we first define $\hat{f}_k^f$ and $\hat{f}_k^b$ to be the estimates of the fading coefficients in the forward and the backward directions, respectively. The bidirectional LMS algorithm is then given as

$$\hat{f}_{k+1}^f = \hat{f}_k^f + 2\mu^f e_k^f a_k \tag{3.11}$$

$$\hat{f}_{k-1}^b = \hat{f}_k^b + 2\mu^b e_k^b a_k \tag{3.12}$$

where μ^f and μ^b are the forward and the backward step-size values of the algorithm, e_k^f and e_k^b are the forward and the backward estimation errors defined as

$$e_k^f = y_k - \hat{f}_k^f a_k \tag{3.13}$$

$$e_k^b = y_k - \hat{f}_k^b a_k. \tag{3.14}$$

Throughout this work, the forward and the backward step-size values are assumed to be the same, i.e., $\mu^f = \mu^b = \mu$, unless otherwise stated.

In order to obtain the final estimates of the fading coefficients from the forward and the backward estimates, the arithmetic average operation is preferred among various choices as a simple yet efficient combining strategy which is described as

$$\hat{f}_k = \frac{\hat{f}_k^f + \hat{f}_k^b}{2}, \tag{3.15}$$

where $\hat{f}_k$ represents the final estimates of the fading coefficients.

The computational complexity of the bidirectional LMS algorithm is observed to be as low as the conventional LMS algorithm and much smaller than that of the optimal Wiener filter with

only 5 complex additions and 9 complex multiplications per iteration. The bidirectional LMS algorithm inherently do not need any pre-filtering stage as is the case for the conventional LMS algorithm which is an important advantage of these adaptive algorithms over the MMSE filter.

As a final point, we would like to mention that the definitions of both LMS algorithms including the term a_k explicitly are appropriate for system identification type problems in which the ultimate concern is to estimate the channel only, and the transmitted symbols $\{a_k\}_{k=1}^{L}$ are therefore assumed to be known a priori at the receiver. This type of operation is sometimes called the training or genie-aided mode, and is very important in understanding and qualifying the tracking algorithm under consideration.

However, the basic motivation in real-life communication systems is generally to estimate the transmitted symbols carrying the desired information. In such a case, channel estimation and tracking is crucial to perform coherent detection and decoding, and we modify the representation of both the unidirectional and the bidirectional LMS algorithms accordingly by just replacing a_k with the corresponding estimate $\hat{a}_k$. This estimate could be obtained either through symbol-by-symbol detection in an uncoded scenario, or from some previous stages as is the case for the iterative channel estimation to be considered in Section 3.8.

3.3 Tracking Performance of Bidirectional LMS over Flat-Fading Channels

In this section, we evaluate the tracking performance of the bidirectional LMS algorithm over flat-fading channels. Therefore, the problem under consideration is of channel identification type so that the overall system is operating in the training mode, i.e., all the transmitted symbols $\{a_k\}_{k=1}^{L}$ are known a priori. As a result, we concentrate only on obtaining a measure which illustrates how well the bidirectional LMS algorithm tracks the unknown time-varying channel rather than evaluating the overall detection performance through bit error rate (BER) or block error rate (BLER) results which will be investigated in the subsequent sections.

In order to characterize the tracking behavior of the bidirectional LMS algorithm, one of the best choice is to derive an MSE expression for the algorithm in a fast time-varying communication channel. Since we deal with how well the bidirectional LMS algorithm follows the time-varying channel rather than being interested in how quickly it converges, the analysis is

performed at the steady-state and the transition analysis is therefore beyond the scope of this work. In the following analysis, the effect of the step-size value of the adaptations is taken into account, as well, as one of the major performance parameters. This theoretical derivation is also very general in the sense that it does not assume any particular fading model and is valid for any type of fading choice with a known power spectrum.

Throughout this section, we highly benefit from the results presented in [61, 63] associated with the conventional unidirectional LMS algorithm. Beside developing an analysis for the bidirectional LMS algorithm depending on these previous results, we also clarify some derivation steps which are not obvious in those original papers. We begin with defining the error performance surface, or equivalently the MSE, for this particular case given as

$$
\begin{align}
J_{MSE,k} &= E\left\{|e_k|^2\right\} \tag{3.16} \\
&= E\left\{\left|y_k - \hat{f}_k a_k\right|^2\right\} \tag{3.17} \\
&= E\left\{\left|\left(f_k - \hat{f}_k\right) a_k + n_k\right|^2\right\} \tag{3.18} \\
&= \underbrace{E\left\{|n_k|^2\right\}}_{J_{min}} + \underbrace{E\left\{|a_k|^2\right\}}_{E_s}\left|f_k - \hat{f}_k\right|^2 \tag{3.19}
\end{align}
$$

where e_k is the overall tracking error and J_{min} is the minimum achievable MSE due to the presence of additive noise and is equal to N_0. The second term in the last line is sometimes called the excess MSE, and is denoted as

$$
J_{ex,k} = E_s \left|f_k - \hat{f}_k\right|^2 \tag{3.20}
$$

which is due to the noisy gradient estimation and the time variation [61]. The steady-state MSE could be obtained accordingly by taking expected value of $J_{MSE,k}$ as follows

$$
\begin{align}
J_{MSE} &= E\{J_{MSE,k}\} \tag{3.21} \\
&= J_{min} + E_s \underbrace{E\left\{\left|f_k - \hat{f}_k\right|^2\right\}}_{J_{MSIE}} \tag{3.22}
\end{align}
$$

where J_{MSIE} is defined as the mean square identification error (MSIE) [63] which is the basic expression to be derived to characterize J_{MSE}. Note that there is a relation between the average excess MSE J_{ex} and J_{MSIE} given as $J_{MSIE} = J_{ex}/E_s$ where $J_{ex} \triangleq E\{J_{ex,k}\}$.

In this work, the expectations in the context of MSE computation is assumed to be over a special ensemble presented in [61] in which the same time-varying channel is used for trans-

mission of an independent set of information symbols. According to this ensemble model, the observation symbols associated with each ensemble member is obtained by adding independent Gaussian noise samples with the same variance which is $E\{|n_k|^2\} = J_{min} = N_0$. The observations belonging to each ensemble member together with the associated input symbols are then sent to the tracking algorithm in order to obtain an estimate. One of the consequences of this construction is that the expected value of the unknown time-varying fading coefficient over this ensemble at a time instant k is equal to itself, i.e., $E\{f_k\} = f_k$, since the unknown fading coefficient is unchanged over the ensemble.

We now proceed to further elaborate the J_{MSIE} expression by making use of the aforementioned ensemble construction and the corresponding results as follows

$$
\begin{aligned}
J_{MSIE} &= E\left\{ \left| f_k - \hat{f}_k \right|^2 \right\} & (3.23) \\
&= E\left\{ \left| \left(\hat{f}_k - E\{\hat{f}_k\} \right) + \left(E\{\hat{f}_k\} - f_k \right) \right|^2 \right\} & (3.24) \\
&= E\left\{ \left| \hat{f}_k - E\{\hat{f}_k\} \right|^2 \right\} + E\left\{ \left| E\{\hat{f}_k\} - f_k \right|^2 \right\} \\
&\quad + 2\,\mathrm{Re}\left\{ E\left\{ \left(\hat{f}_k - E\{\hat{f}_k\} \right) \left(E\{\hat{f}_k\} - f_k \right)^* \right\} \right\} & (3.25)
\end{aligned}
$$

where the time reverse of (3.23) is used in (3.24). The last term in (3.25) could be further simplified as in [61] by making use of the fact that the time-varying channel is unchanged over the ensemble so that $E\{f_k\} = f_k$ for any time epoch k. We accordingly obtain

$$
\begin{aligned}
E\left\{ \left(\hat{f}_k - E\{\hat{f}_k\} \right) \left(E\{\hat{f}_k\} - f_k \right)^* \right\} &= E\left\{ \hat{f}_k \left(E\{\hat{f}_k\} \right)^* \right\} - E\left\{ \hat{f}_k f_k^* \right\} - E\left\{ E\{\hat{f}_k\} \left(E\{\hat{f}_k\} \right)^* \right\} \\
&\quad + E\left\{ E\{\hat{f}_k\} f_k^* \right\} \\
&= \left| E\{\hat{f}_k\} \right|^2 - E\{\hat{f}_k\} f_k^* - \left| E\{\hat{f}_k\} \right|^2 + E\{\hat{f}_k\} f_k^* \\
&= 0. & (3.26)
\end{aligned}
$$

Therefore, J_{MSIE} of the bidirectional LMS in time-varying channels is expressed as a sum of two terms which are called the self-noise (J_{self}) and the lag (J_{lag}) components [61], and is given as

$$
\begin{aligned}
J_{MSIE} &= E\left\{ \left| f_k - \hat{f}_k \right|^2 \right\} & (3.27) \\
&= \underbrace{E\left\{ \left| \hat{f}_k - E\{\hat{f}_k\} \right|^2 \right\}}_{J_{self}} + \underbrace{E\left\{ \left| E\{\hat{f}_k\} - f_k \right|^2 \right\}}_{J_{lag}} & (3.28) \\
&= J_{self} + J_{lag}. & (3.29)
\end{aligned}
$$

With the light of (3.28)-(3.29) and the discussion in [61], any deviation of the estimate of the unknown fading coefficient, $\hat{f}_k$, from the ensemble mean of the estimate, $E\{\hat{f}_k\}$, contributes to the self-noise part while differences between the ensemble mean of the estimate and the unknown fading coefficient itself amplify the lag part. According to this result, one could conclude that J_{self} arises from the noisy gradient estimation of the error performance surface under consideration whereas J_{lag} is just due to the time variation. Because J_{MSIE} is perfectly expressed as a sum of the self-noise and the lag components, we will separately derive the steady-state expressions for J_{self} and J_{lag} in this section in order to come up with a final expression for the steady-state MSE.

3.3.1 Derivation of the Self-Noise Component (J_{self})

As explained before, the source of the self-noise is the noisy gradient estimation of the error performance surface given in (3.16) for the system identification problem under consideration. We therefore ignore any time variation while deriving J_{self} and focus only on the effect of the gradient estimation error by extending the findings of [57, 61] on the analysis of the conventional unidirectional LMS algorithm.

In order to evaluate J_{self}, we first model the gradient estimates in the forward and the backward directions as follows

$$\hat{\nabla}_k^f = \nabla_k^f + \epsilon_k^f = 2E_s(\hat{f}_k^f - f_k) + \epsilon_k^f \tag{3.30}$$

$$\hat{\nabla}_k^b = \nabla_k^b + \epsilon_k^b = 2E_s(\hat{f}_k^b - f_k) + \epsilon_k^b, \tag{3.31}$$

respectively, where $\nabla_k^f = 2E_s(\hat{f}_k^f - f_k)$ and $\nabla_k^b = 2E_s(\hat{f}_k^b - f_k)$ are the true gradients, and ϵ_k^f and ϵ_k^b are the associated estimation errors in the forward and the backward directions, respectively [57]. In order to incorporate the effect of gradient estimation error into the adaptations, we prefer to express the conventional LMS adaptations in the forward and the backward directions as follows

$$\hat{f}_{k+1}^f = \hat{f}_k^f - \mu \hat{\nabla}_k^f \tag{3.32}$$

$$\hat{f}_{k-1}^b = \hat{f}_k^b - \mu \hat{\nabla}_k^b. \tag{3.33}$$

Note that, the gradient estimates for the conventional LMS algorithm are $\hat{\nabla}_k^f = -2\,e_k^f a_k$ and $\hat{\nabla}_k^b = -2\,e_k^b a_k$ where the adaptation errors e_k^f and e_k^f are given in (3.13)-(3.14) [102]. Instead

25

of this representation, we prefer to use the gradient estimates given in (3.30)-(3.31), and the adaptations given by (3.32)-(3.33) accordingly become

$$\hat{f}_{k+1}^f = \hat{f}_k^f - 2\mu E_s (\hat{f}_k^f - f_k) - \mu \epsilon_k^f \tag{3.34}$$

$$\hat{f}_{k-1}^b = \hat{f}_k^b - 2\mu E_s (\hat{f}_k^b - f_k) - \mu \epsilon_k^b. \tag{3.35}$$

In order to characterize the error due to the noisy gradient estimation, we define the forward and the backward tap-weight tracking errors as $v_k^f = \hat{f}_k^f - f_k$ and $v_k^b = \hat{f}_k^b - f_k$, respectively. The algorithms given in (3.34)-(3.35) could be expressed in terms of tap-weight tracking errors by subtracting f_{k+1} from both side of the adaptations as follows

$$\hat{f}_{k+1}^f - f_{k+1} = \hat{f}_k^f - f_k - 2\mu E_s (\hat{f}_k^f - f_k) - \mu \epsilon_k^f \tag{3.36}$$

$$\hat{f}_{k-1}^b - f_{k+1} = \hat{f}_k^b - f_k - 2\mu E_s (\hat{f}_k^b - f_k) - \mu \epsilon_k^b. \tag{3.37}$$

where we make use of the time invariance assumption, i.e., $f_{k+1} = f_k$, at the right side of (3.36)-(3.37) which is previously made for the self-noise derivation. As a result, the adaptations in terms of the tap-weight tracking error become

$$\hat{v}_{k+1}^f = (1 - 2\mu E_s)\hat{v}_k^f - \mu \epsilon_k^f \tag{3.38}$$

$$\hat{v}_{k-1}^b = (1 - 2\mu E_s)\hat{v}_k^b - \mu \epsilon_k^b. \tag{3.39}$$

In order to express the self-noise in terms of the overall tap-weight tracking error, which is defined as $v_k = \hat{f}_k - f_k$, we further elaborate the expression given in (3.28). In order to get rid of the inner expectation $E\{\hat{f}_k\}$ present as a part of the self-noise expression, we take expectations of (3.34)-(3.35) as follows

$$\underbrace{E\{\hat{f}_{k+1}^f\}}_{=E\{\hat{f}_k^f\}} = E\{\hat{f}_k^f\} - 2\mu E_s \left(E\{\hat{f}_k^f\} - \underbrace{E\{f_k\}}_{=f_k}\right) - \mu \underbrace{E\{\epsilon_k^f\}}_{=0} \tag{3.40}$$

$$\underbrace{E\{\hat{f}_{k-1}^b\}}_{=E\{\hat{f}_k^b\}} = E\{\hat{f}_k^b\} - 2\mu E_s \left(E\{\hat{f}_k^b\} - \underbrace{E\{f_k\}}_{=f_k}\right) - \mu \underbrace{E\{\epsilon_k^b\}}_{=0}, \tag{3.41}$$

where $E\{\hat{f}_{k+1}^f\} = E\{\hat{f}_k^f\}$ and $E\{\hat{f}_{k-1}^b\} = E\{\hat{f}_k^b\}$ follows from the time invariance assumption, $E\{f_k\} = f_k$ is a previous result due to the fact that f_k is common across the ensemble, and ϵ_k^f ϵ_k^b are zero-mean Gaussian random variables by definition [57]. After some straightforward steps in (3.40)-(3.41), we have $E\{\hat{f}_k^f\} = f_k$ and $E\{\hat{f}_k^b\} = f_k$, and $E\{\hat{f}_k\}$ is therefore found to be

$$E\{\hat{f}_k\} = \frac{E\{\hat{f}_k^f\} + E\{\hat{f}_k^b\}}{2} = \frac{f_k + f_k}{2} = f_k \tag{3.42}$$

with a help of (3.15). As a result of these findings, the self-noise expression given in (3.28) becomes

$$J_{self} = E\left\{\left|\hat{f}_k - E\left\{\hat{f}_k\right\}\right|^2\right\} = E\left\{\left|\hat{f}_k - f_k\right|^2\right\} = E\left\{|v_k|^2\right\} \tag{3.43}$$

where the overall tap-weight tracking error v_k is given as

$$v_k = \hat{f}_k - f_k \tag{3.44}$$

$$= \frac{\hat{f}_k^f + \hat{f}_k^b}{2} - f_k = \frac{\left(\hat{f}_k^f - f_k\right) + \left(\hat{f}_k^b - f_k\right)}{2} = \frac{v_k^f + v_k^b}{2}. \tag{3.45}$$

Using (3.45), the self-noise defined in (3.43) could then be evaluated as

$$J_{self} = E\left\{|v_k|^2\right\} \tag{3.46}$$

$$= E\left\{\left(\frac{v_k^f + v_k^b}{2}\right)\left(\frac{v_k^f + v_k^b}{2}\right)^*\right\} \tag{3.47}$$

$$= \frac{E\left\{|v_k^f|^2\right\}}{4} + \frac{E\left\{|v_k^b|^2\right\}}{4} + \frac{\text{Re}\left(E\left\{v_k^f(v_k^b)^*\right\}\right)}{2} \tag{3.48}$$

where $E\{|v_k^f|^2\}$ and $E\{|v_k^b|^2\}$ are known through the steady-state analysis of the conventional LMS algorithm, and is given in [57] as $\mu/(1-\mu)\, J_{min}$. In order to simplify the expression found in (3.48), the expectation in the last term could be further elaborated using (3.38)-(3.39) as follows

$$E\{v_k^f(v_k^b)^*\} = E\left\{\left[(1 - 2\mu E_s)v_{k-1}^f - \mu\epsilon_{k-1}^f\right]\left[(1 - 2\mu E_s)(v_{k+1}^b)^* - \mu(\epsilon_{k+1}^b)^*\right]\right\} \tag{3.49}$$

$$= (1 - 2\mu E_s)^2 E\left\{v_{k-1}^f(v_{k+1}^b)^*\right\} - \mu(1 - 2\mu E_s)E\left\{v_{k-1}^f(\epsilon_{k+1}^b)^*\right\}$$

$$- \mu(1 - 2\mu E_s)E\left\{\epsilon_{k-1}^f(v_{k+1}^b)^*\right\} + \mu^2 E\left\{\epsilon_{k-1}^f(\epsilon_{k+1}^b)^*\right\} \tag{3.50}$$

$$= (1 - 2\mu E_s)^2 E\left\{v_{k-1}^f(v_{k+1}^b)^*\right\} \tag{3.51}$$

where (3.51) makes use of the assumptions that ϵ_k^f and ϵ_l^b are assumed to be zero-mean random variables, and they are mutually independent of each other and of v_k^f and v_k^b, which directly follows from [57]. Through iterations, (3.51) becomes

$$E\{v_k^f(v_k^b)^*\} = (1 - 2\mu E_s)^L E\{v_0^f(v_L^b)^*\}$$

which could safely be ignored since $(1 - 2\mu E_s)^L \ll 1$ due to the fact that $|1 - 2\mu E_s| < 1$ is known to be the stability condition of the conventional LMS algorithm. Therefore, the

self-noise term becomes

$$J_{self} \quad = \quad E\{|v_k|^2\} = \frac{E\left\{\left|v_k^f\right|^2\right\}}{4} + \frac{E\left\{\left|v_k^b\right|^2\right\}}{4} \tag{3.52}$$

$$= \quad \frac{\mu/(1-\mu) J_{min}}{4} + \frac{\mu/(1-\mu) J_{min}}{4} = \frac{\mu}{2(1-\mu)} J_{min}, \tag{3.53}$$

which is observed to depend purely on the step-size μ and the minimum achievable MSE which is equal to the noise variance for this particular case. As a result, the self-noise for the bidirectional LMS algorithm is half that of the conventional LMS algorithm. Since no time variation is assumed while deriving self-noise part, MSE of the bidirectional algorithm in a quasi-static channel is also expected to be half that of the conventional LMS algorithm.

3.3.2 Derivation of the Lag Component (J_{lag})

We have analyzed the contribution of the noisy gradient estimation into the steady-state MSE by deriving the self-noise expression in the previous section. We therefore assume a perfect gradient estimation while analyzing the lag component, and concentrate only on the effect of the time variation. By making use of this assumption and the results of [61, 63], the forward and the backward adaptations given in (3.32)-(3.33) become

$$\hat{f}_{k+1}^f \quad = \quad \hat{f}_k^f - \mu \nabla_k^f = (1 - 2\mu E_s) \hat{f}_k^f + 2\mu E_s f_k \tag{3.54}$$

$$\hat{f}_{k-1}^b \quad = \quad \hat{f}_k^b - \mu \nabla_k^b = (1 - 2\mu E_s) \hat{f}_k^b + 2\mu E_s f_k. \tag{3.55}$$

In order to cope with the time variation, we prefer to express the adaptations in the frequency domain, as in [61]. For this purpose, z-transform of (3.54) and (3.55) are first computed, and the results are then rearranged as follows

$$\hat{f}^f(z) \quad = \quad Z\left\{\hat{f}_k^f\right\} = \frac{1-\beta}{z-\beta} f(z) \tag{3.56}$$

$$\hat{f}^b(z) \quad = \quad Z\left\{\hat{f}_k^b\right\} = \frac{1-\beta}{z^{-1}-\beta} f(z) \tag{3.57}$$

where $\beta = 1 - 2\mu E_s$ is the geometric ratio of the adjustments [61], $Z\{\cdot\}$ denotes the z-transform [103] defined as

$$Z\{x_k\} = \sum_{k=-\infty}^{\infty} x_k z^{-k}, \tag{3.58}$$

28

and $f(z) = Z\{f_k\}$. The z-transform of $\hat{f}_k$ is then found with a help of (3.15) as follows

$$\hat{f}(z) = \frac{\hat{f}^f(z) + \hat{f}^b(z)}{2} = \frac{1}{2}\left(\frac{1-\beta}{z-\beta} + \frac{1-\beta}{z^{-1}-\beta}\right)f(z). \tag{3.59}$$

The z-transform of the tracking error $v(z)$ becomes

$$v(z) = \hat{f}(z) - f(z) = H(z)\, f(z) \tag{3.60}$$

where $H(z)$ is the transfer function for the bidirectional LMS algorithm which is independent of the channel characteristics to be estimated, and is given as

$$H(z) = \frac{1}{2}\left(\frac{1-\beta}{z-\beta} + \frac{1-\beta}{z^{-1}-\beta}\right) - 1 \tag{3.61}$$

$$= -\frac{1+\beta}{2\beta} + \frac{1-\beta}{2\beta}\left(\frac{1}{1-\beta z^{-1}} - \frac{1}{1-\frac{1}{\beta}z^{-1}}\right). \tag{3.62}$$

In (3.62), it is observed that the poles of the transfer function $H(z)$ are $z_{p1} = \beta$ and $z_{p2} = 1/\beta$ both of which are real. Remembering that $\beta = |1 - 2\mu E_s| < 1$ is the necessary condition for the convergence of the LMS algorithm in the mean, the first pole $z_{p1} = \beta$ lies inside the unit circle while the second pole $z_{p2} = 1/\beta$ is outside of it.

The transfer function could also be expressed in the frequency domain as follows

$$H(e^{jw}) = \frac{1}{2}\left(\frac{1-\beta}{e^{jw}-\beta} + \frac{1-\beta}{e^{-jw}-\beta}\right) - 1 \tag{3.63}$$

$$= -\frac{(1+\beta)(1-\cos w)}{1+\beta^2 - 2\beta\cos w}. \tag{3.64}$$

From (3.64), one could observe that $H(e^{jw})$ is a real-valued function. Note also that, the term in the denominator of (3.64) is nonnegative since

$$1 + \beta^2 - 2\beta\cos w \geq 1 + \beta^2 - 2\beta$$
$$= (1-\beta)^2$$
$$\geq 0,$$

and the numerator of (3.64) is also nonnegative since $|\beta| < 1$ is the mean-convergence condition as explained before. As a result, $H(e^{jw})$ is always a negative function of w. We emphasize that these findings about $H(e^{jw})$ is valid for the previous assumption that is the forward and the backward step-size values are equal. In Section 3.4.2, we also investigate the effect of

a more general setting in which the forward and the backward step-size values are chosen independently which result in a complex-valued transfer function in case of unequal forward and backward values.
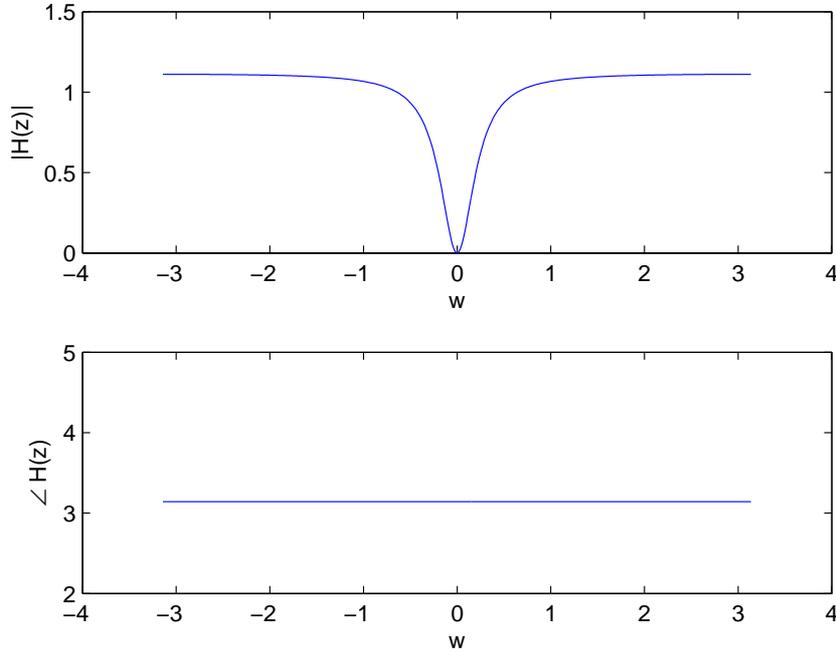
Figure 3.2: Transfer function characteristics for $\mu = 0.1$.

The magnitude and the phase of $H(e^{jw})$ are given in Fig. 3.3 for $\mu = 0.1$, $E_s = 1$ and $w \in [-\pi, \pi]$, as an example. We observe that the phase of the transfer function is $\angle H(e^{jw}) = -\pi$ for any choice of w which implies a real and negative value, as stated before. The magnitude of $H(e^{jw})$ is almost constant for $w > \pi/2$ for this particular choice, and has the minimum at $w = 0$.

We now turn to the derivation of the lag component. Remember that the ensemble over which the expectations are taken is defined at the beginning of this section to be a set of systems, or equivalently adaptive processes, which have the same time-varying channel to be estimated and are fed by a set of independent input signals. In addition, since the gradient of the error performance surface is assumed to be estimated perfectly for this particular case, the channel estimates associated with different ensemble members are the same due to the adaptations given in (3.54)-(3.55). In order to provide a better understanding of this point, we also note

30

that the adaptations (3.54)-(3.55) do not depend on the input signal a_k which is the major source of difference across the ensemble, and this dependency is considered in the self-noise derivation by (3.32)-(3.33) where the forward and the backward gradient estimates are given to be a function of a_k. As a result, we conclude that $E\{\hat{f}_k\} = \hat{f}_k$ since $\hat{f}_k$ is the same across the ensemble for this particular case.

With a help of the previous discussion which concludes that $E\{\hat{f}_k\} = \hat{f}_k$, the lag component given in (3.26) becomes

$$J_{lag} = E\left\{\left|E\left\{\hat{f}_k\right\} - f_k\right|^2\right\} = E\left\{\left|\hat{f}_k - f_k\right|^2\right\}, \tag{3.65}$$

and J_{lag} is therefore interpreted as the mean-square energy of the tracking error. In order to evaluate this energy, the input/output relation given in (3.60) is also represented by a block diagram as follows

$$f_k \longrightarrow \boxed{H(z)} \longrightarrow \hat{f}_k - f_k$$
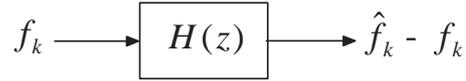
Figure 3.3: Transfer function representation of the bidirectional LMS algorithm.

which implies that the mean-square energy in the tracking error could be evaluated in the frequency domain as [63]

$$J_{lag} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S(w)\, dw \tag{3.66}$$

where $S(w)$ is the power spectrum of the fading process under consideration.

As a result, the final expression for the steady state MSIE is given as

$$\begin{aligned} J_{MSIE} &= J_{self} + J_{lag} \\ &= \frac{\mu}{2(1-\mu)} J_{min} + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S(w)\, dw, \end{aligned} \tag{3.67}$$

and the steady-state MSE is therefore found to be

$$\begin{aligned} J_{MSE} &= J_{min} + E_s J_{MSIE} \\ &= \left(1 + \frac{\mu E_s}{2(1-\mu)}\right) J_{min} + \frac{E_s}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S(w)\, dw. \end{aligned} \tag{3.68}$$

31

3.3.3 Numerical Results

In this section, our goal is to verify the steady-state MSE analysis of the bidirectional LMS algorithm given in the previous section by comparing it with the experimental data obtained through Monte Carlo simulations, and to evaluate the overall tracking performance of the bidirectional LMS algorithm algorithm through a number of comparisons for the resulting MSE with that of the conventional unidirectional LMS algorithm and the optimal MMSE filter under various circumstances.

Throughout simulations, an independent set of $L = 100$ information symbols are chosen from a BPSK alphabet such that $\mathcal{A} = \{-1, +1\}$ and $E_s = 1$. Unless otherwise stated, we assume Rayleigh fading with the well-known Jakes' power spectrum [98] for all simulations with the temporal autocorrelation given as

$$r(\tau) = J_0\left(2\pi f_d T_s \tau\right) \tag{3.69}$$

where $J_0(.)$ is the zeroth order Bessel function of the first kind, $f_d T_s$ is the maximum normalized Doppler frequency, and τ is the time difference in the discrete domain. We also consider the Gaussian and the double-Gaussian power spectrums to discuss the generality of performance superiority of the bidirectional LMS algorithm and the correctness of the associated theoretical derivations. Note that, the choice of both the Gaussian and the double-Gaussian power spectrums makes also practical sense since they are known in the literature to well suit to the HF channels [104].

The speed of the Rayleigh fading channel for Jakes' spectrum is determined by the maximum normalized Doppler frequency represented by $f_d T_s$. The variations in the magnitude of the channel along a block of 200 symbols as a result of a set Doppler values $f_d T_s = \{0.005, 0.01, 0.02\}$ are depicted in Fig. 3.4. We observe that even the slowest channel realization is changing significantly during the transmission of 100 symbols. In the subsequent simulations, we prefer to use much faster channel realizations by choosing $f_d T_s = \{0.01, 0.02\}$ unless otherwise stated.

In order to gain more insight on the Doppler spread, consider a mobile vehicle making use of a GSM service at 900 MHz. If the vehicle is moving at a high speed of 120 km/h, then the

maximum Doppler frequency is

$$f_d = f_c \frac{v}{c} = 900 \times 10^6 \, \text{Hz} \, \frac{120 \, \text{km/h}}{3 \times 10^8 \, \text{m/s}} = 100 \, \text{Hz} \tag{3.70}$$

which is therefore referred to as a very fast channel in the literature.
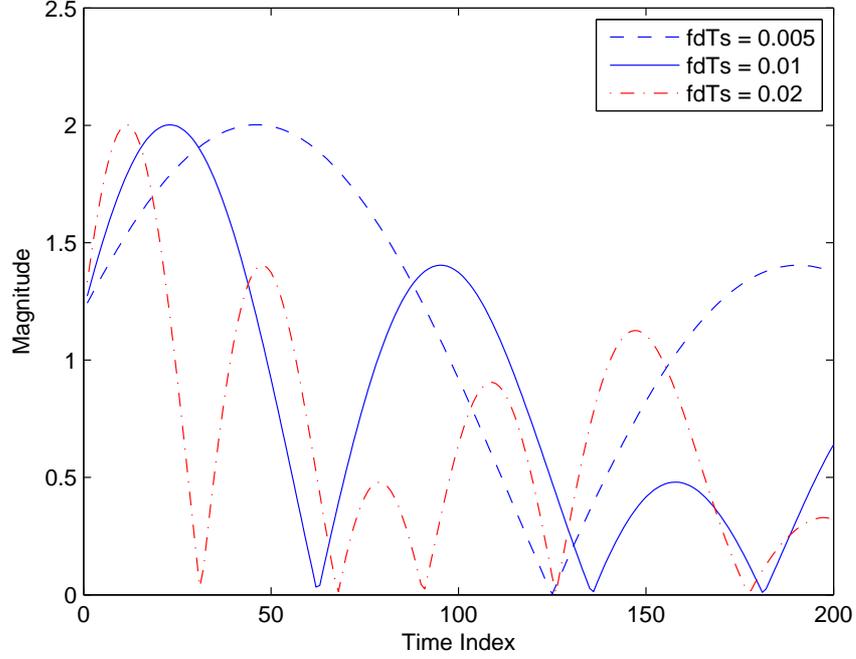
Figure 3.4: Magnitude of a set of Rayleigh fading realizations corresponding to $f_d T_s = \{0.005, 0.01, 0.02\}$.

In order to achieve the aforementioned goals of this section, we depict the experimental and theoretical MSIE, i.e., J_{MSIE}, in Fig. 3.8 associated with both the unidirectional and the bidirectional LMS algorithms for Jakes' spectrum with respect to varying step-size values at $\gamma = 10$ dB. The theoretical MSIE for the unidirectional LMS is given in [63] as

$$J_{MSIE} \;=\; \frac{\mu}{1-\mu} J_{min} + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) \right|^2 S(w) \, dw, \tag{3.71}$$

except the term $1 - \mu$ in the denominator of the self-noise part, i.e., J_{self}, and the associated transfer function is defined therein to be

$$H(e^{jw}) = \frac{2(1 - \cos w)}{(1 - 2\beta \cos w + \beta^2)}. \tag{3.72}$$

33
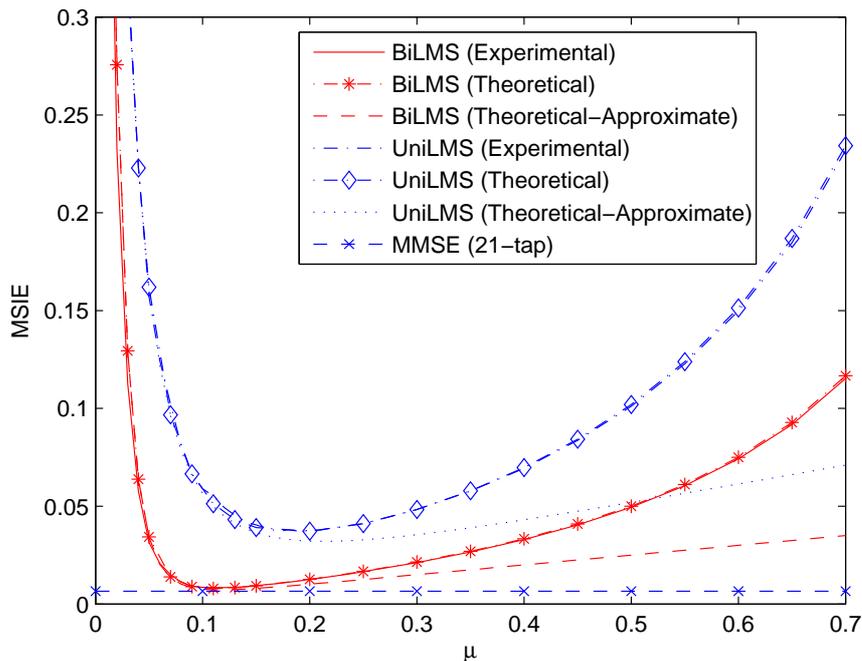
Figure 3.5: Theoretical and experimental MSIE for UniLMS and BiLMS for varying step-size, μ, at SNR=10 dB and $f_d T_s = 0.01$. The experimental MSIE for a 21-tap MMSE and theoretical approximate MSIE ignoring $(1 - \mu)$ term are also provided.

The MSIE value corresponding to the optimal MMSE filter with 21 taps is also provided in Fig. 3.5 as a bound. We make a number of observations from these results. First, there is almost a perfect match between the theoretical results obtained using the steady-state MSIE expression given in (3.67) and the experimental ones for the bidirectional LMS algorithm for any choice of the step-size value. This result together with the subsequent simulations at different Doppler and SNR values verifies the derivation of the steady-state MSE for the bidirectional LMS algorithm performed in the previous section. Second, the minimum MSIE of the bidirectional LMS algorithm, which is achieved when the optimal step-size value is used, is very close to that of the optimal MMSE filter with 21 taps and is significantly smaller than that of the conventional unidirectional LMS algorithm. As a result, the bidirectional LMS algorithm is said to have a near-optimal tracking performance. Furthermore, this appealing performance is obtained with a practical level of complexity which is argued previously to be comparable to that of the unidirectional LMS algorithm and significantly smaller than that of the optimal MMSE filter. Finally, we also observe significant deviations in theoretical

MSIE values when it is computed by ignoring the term $(1 - \mu)$ in the denominator of J_{self} as in [61, 63] for both the conventional unidirectional and bidirectional LMS algorithms.
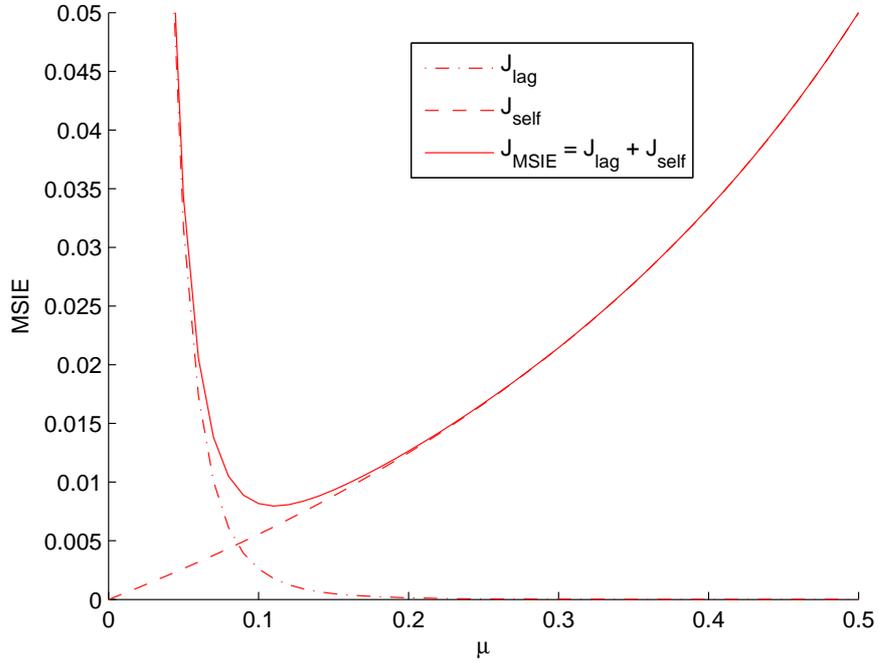
Figure 3.6: J_{self} and J_{lag} terms for BiLMS with varying step-size at SNR=10 dB and $f_d T_s = 0.01$.

In derivation of MSE, it is assumed that the resulting MSIE, i.e., J_{MSIE}, is a sum of self-noise, i.e., J_{self}, and lag, i.e., J_{lag}, parts. In Fig. 3.6, we depict these two parts individually and also together with the resulting J_{MSIE} for the same settings to provide further understanding. We observe that J_{lag} is a decreasing function of the step-size and is dominant for the small step-size values whereas J_{self} is an increasing function of the step-size and is dominant for large step-size values. The optimal step-size value, i.e., μ_{opt}, is therefore observed to be a compromise between J_{self} and J_{lag}.

In Fig. 3.7, we plot the experimental and theoretical MSIE performances of the unidirectional LMS and the bidirectional LMS algorithms for varying SNR where both algorithms employ the optimal step-size values. The experimental MSIE of the 31-tap MMSE filter is also provided. We observe that the theoretical MSIE values exactly follow the experimental MSIE values of the bidirectional LMS algorithm, and that the MSIE performance of the bidirec-

tional LMS algorithm is very close to that of the optimal MMSE filter and is much better than that of the unidirectional LMS algorithm.
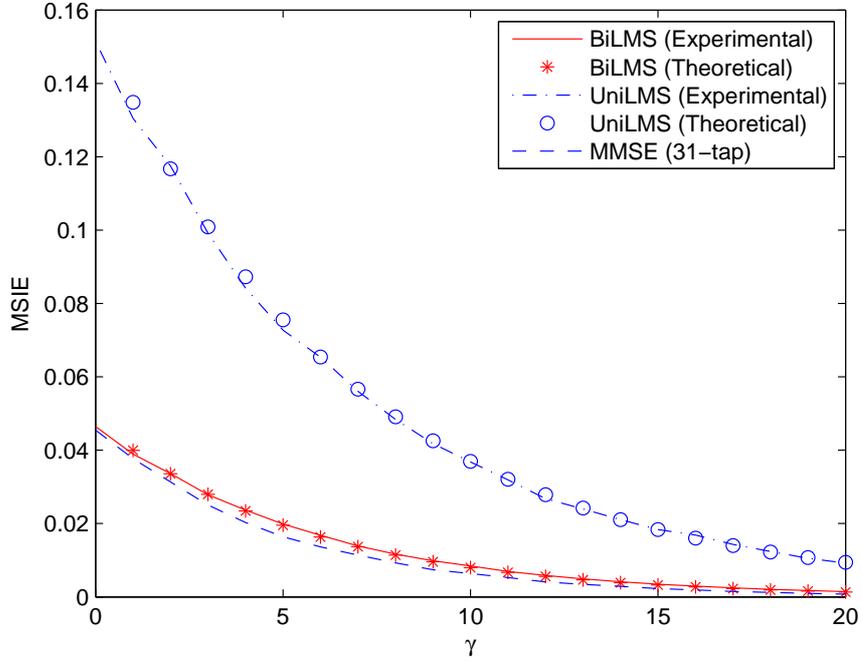
Figure 3.7: Theoretical and experimental MSIE for UniLMS and BiLMS with the optimal step-size, μ_{opt}, with varying SNR and $f_d T_s = 0.01$. The experimental MSIE for a 31-tap MMSE filter is also provided.

Although the details of MMSE filter design is considered in Section 3.2.1, we briefly present the effect of the associated filter length. For this purpose, the experimental MSIE values of the optimal MMSE filter with various number of taps are depicted in Fig. 3.8. It is observed that performance improvement of the optimal MMSE approach as a result of increase in the number of filter taps diminishes as the number of filter taps gets bigger. A careful consideration of Fig. 3.7 and Fig. 3.8 shows that the bidirectional LMS algorithm results in an MSIE which is smaller than that of, let us say, an 11-tap MMSE filter for a definite SNR region. We therefore use the MMSE filter with a sufficient number of taps in the simulations to provide a fair performance comparison.
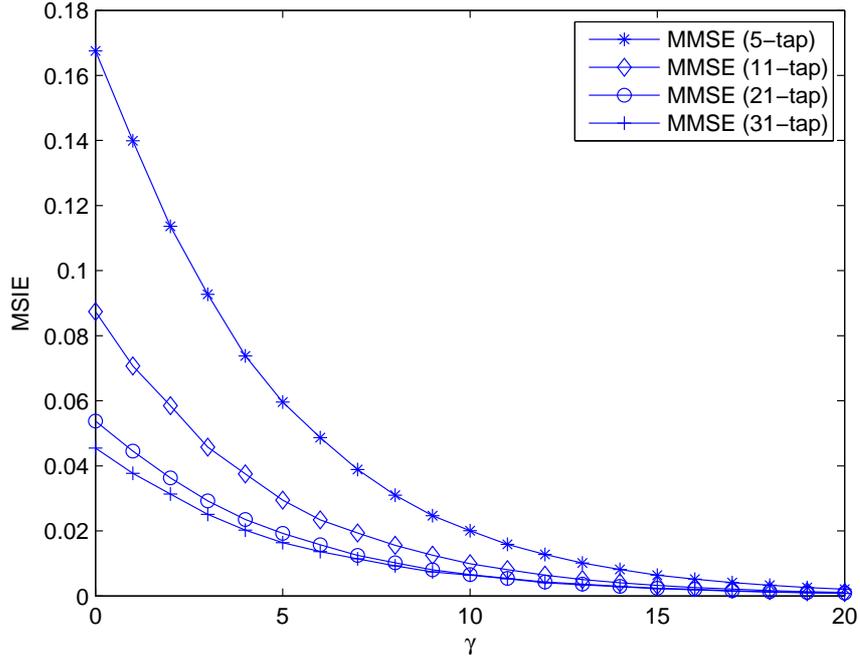
Figure 3.8: Experimental MSIE for MMSE filters with various taps for $f_d T_s = 0.01$.

Before proceeding further, we would like to investigate the robustness of the MMSE filter and the bidirectional LMS algorithm when the transmitted symbols are known imperfectly. This is similar to the scenario in the iterative channel estimation to be considered in Section 3.8 where the imperfect estimates of the transmitted symbols are used in channel estimation after the initial acquisition. In order to gain insight for this future issue, we run the estimation algorithms under considerations with the imperfect values of the transmitted symbols which are obtained such that a target BER will be achieved in case of hard decision. Let

$$\hat{a}_k = a_k + \eta_k \qquad (3.73)$$

be the imperfect estimate of the transmitted symbol a_k to be used in channel estimation where η_k is a sample from a zero-mean white Gaussian noise with the variance σ_η^2. If we perform hard decision on $\hat{a}_k$, the corresponding BER for BPSK signalling is given as [102]

$$\text{BER} = \frac{1}{2} \text{erfc} \left(\frac{1}{\sqrt{2} \, \sigma_\eta} \right) \qquad (3.74)$$

37

where erfc(·) is the complementary error function defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt \,. \tag{3.75}$$

In order to achieve a target BER, we therefore choose σ_η^2 to be

$$\sigma_\eta^2 = \frac{1}{2 \, \text{erfcinv}^2(2 \, \text{BER})} \tag{3.76}$$

where erfcinv(·) is the inverse of erfc(·).
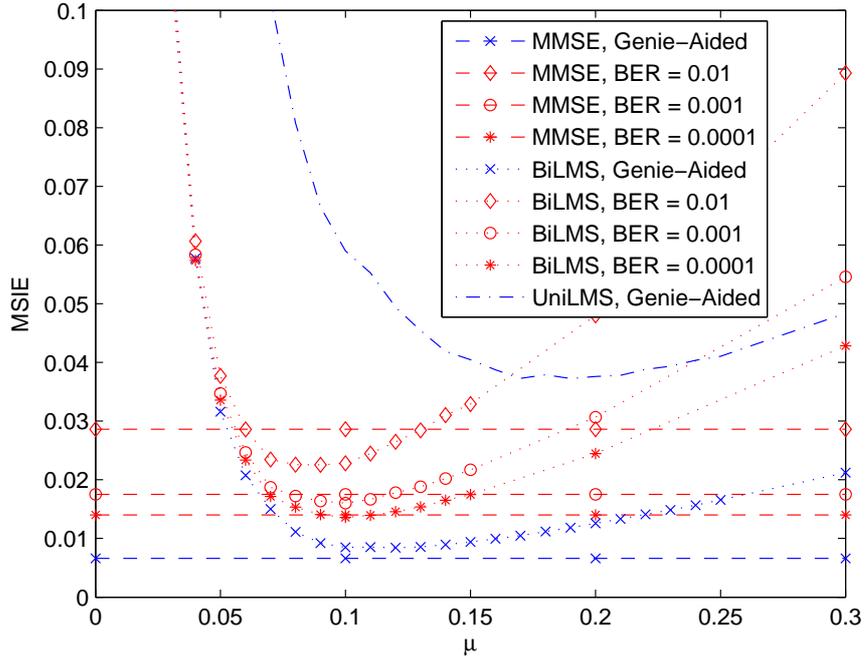
Figure 3.9: Experimental MSIE for UniLMS, BiLMS and 21-tap MMSE at SNR=10 dB and $f_d T_s = 0.01$.

In Fig. 3.9, we present the MSIE results of the 21-tap MMSE filter and the bidirectional LMS algorithm for the above scenario with various BER choices at $\gamma = 10$ for $f_d T_s = 0.01$. We observe that although the MMSE filter performs slightly better than the bidirectional LMS algorithm in a genie-aided mode, it performs worse as the imperfect estimates in use become more noisy. This result is of practical importance in understanding and evaluating the results of the iterative channel estimation technique considered in this and the subsequent chapters.

We now investigate the effect of a much faster Rayleigh fading channel with $f_d T_s = 0.02$ on

the tracking performance of the bidirectional LMS algorithm together with the accuracy of the associated steady-state MSE expression. In Fig. 3.10, the theoretical and the experimental MSIE values for the unidirectional and the bidirectional LMS algorithms are depicted for the varying step-size at SNR= 15 dB. The theoretical MSIE of the bidirectional LMS algorithm is observed to match exactly the experimental values minimum of which is very close to that of the optimal MMSE filter with 21-tap and is significantly better than that of the unidirectional LMS algorithm. The approximate MSIE of both the unidirectional and the bidirectional LMS algorithm in which the term $(1 - \mu)$ in the denominator of J_{self} is ignored is again provided with a result of an obvious deviation from the actual experimental data.
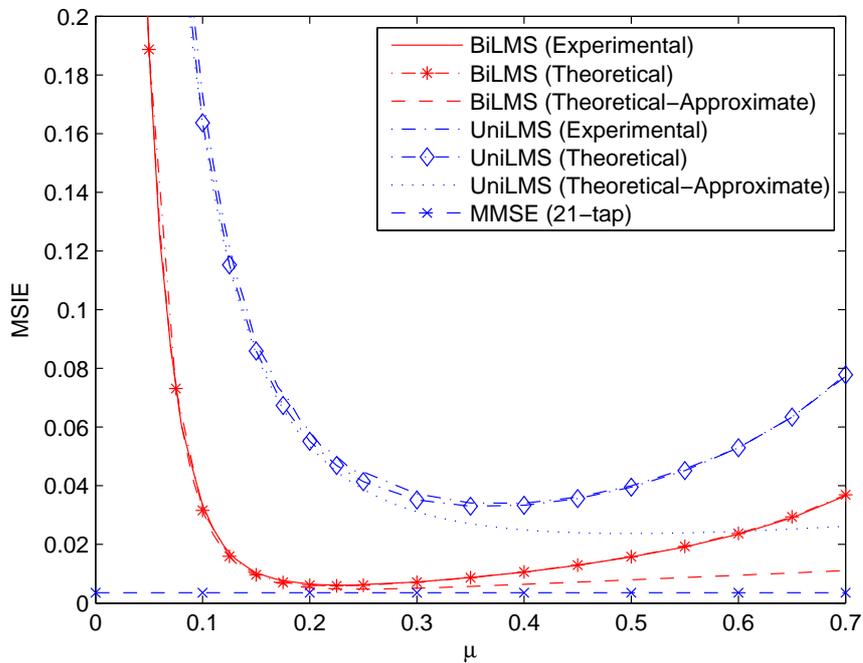
Figure 3.10: Theoretical and experimental MSIE for UniLMS and BiLMS for varying step-size, μ, at SNR=15 dB and $f_d T_s = 0.02$. The experimental MSIE for a 21-tap MMSE and theoretical MSIE ignoring $(1 - \mu)$ term are also provided.

In Fig. 3.11, the theoretical and the experimental MSIE values for the bidirectional and the unidirectional LMS algorithms are presented with optimal step-size and varying SNR for $f_d T_s = 0.02$. The experimental MSIE of an optimal MMSE filter with 21-tap is also added as a benchmark. For this significantly fast Rayleigh fading channel, the theoretical MSIE values of the bidirectional LMS algorithm are observed to exactly follow the associated experimental

ones. The resulting MSIE of the bidirectional LMS algorithm is also very close to that of the MMSE filter under consideration and is substantially better than that of the unidirectional LMS algorithm. For comparison purposes, we also provide the MSIE results for optimal MMSE filters with various number of taps in Fig. 3.12 for this fast Rayleigh fading channel. In this case, because the channel is changing very fast, the correlation between the consecutive discrete time indices decreases as compared to relatively slow scenarios. As a result, the MSIE performance of both the 21-tap and 31-tap optimal MMSE filters are almost the same.

Finally, we present the experimental MSIE results for varying $f_d T_s$ at $\gamma = 10$ dB in Fig. 3.13. We observe that, although the bidirectional LMS algorithm keeps its superiority over the unidirectional LMS algorithm for any choice of $f_d T_s$, the associated performance gets worser than that of the MMSE filter as $f_d T_s$ becomes larger.
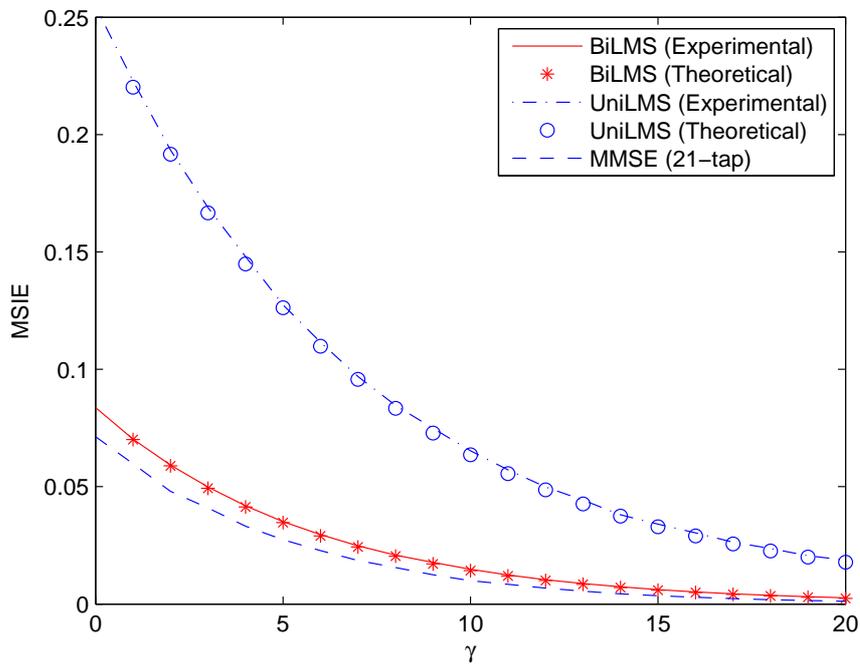
Figure 3.11: Theoretical and experimental MSIE for UniLMS and BiLMS with the optimal step-size, μ_{opt}, with varying SNR and $f_d T_s = 0.02$. The experimental MSIE for a 21-tap MMSE filter is also provided.
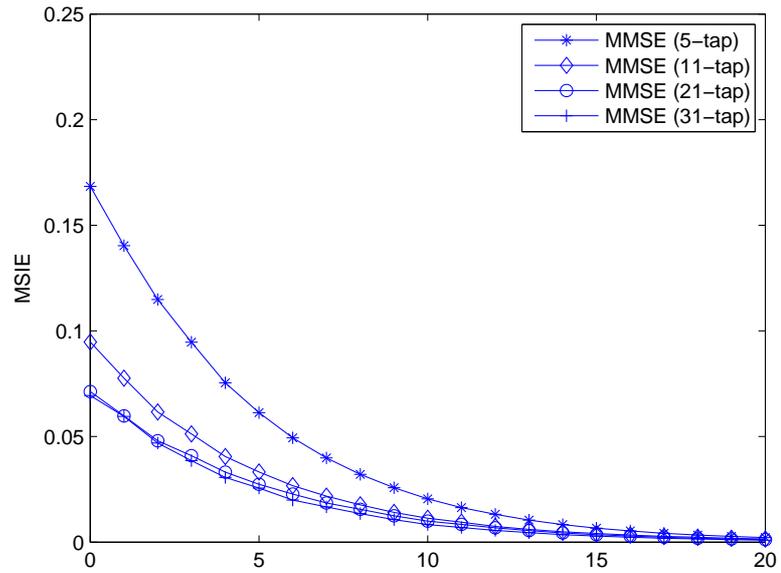
Figure 3.12: Experimental MSIE for MMSE filters with various taps for $f_dT_s = 0.02$.
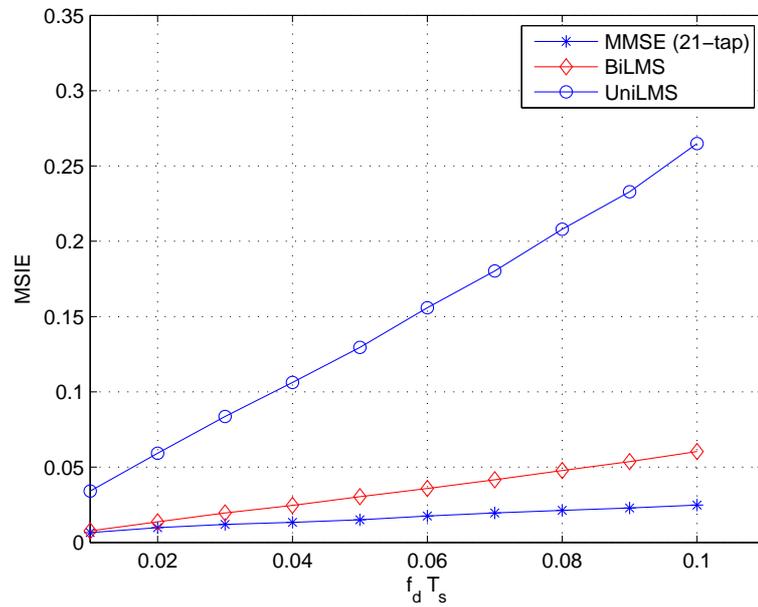
Figure 3.13: Experimental MSIE for varying f_dT_s at $\gamma = 10$ dB.

We now consider the Gaussian and the double-Gaussian spectrums which are discussed in the

literature to be well suited to the HF communications. In [104], the double-Gaussian spectrum is represented by two magnetoionic components with the following tap-gain correlation function

$$r(\triangle t) = C_a \exp\left\{-2\pi^2 \sigma_a^2 (\triangle t)^2 + j 2\pi f_a \triangle t\right\} + C_b \exp\left\{-2\pi^2 \sigma_b^2 (\triangle t)^2 + j 2\pi f_b \triangle t\right\} \qquad (3.77)$$

and the spectrum

$$S(f) = \frac{C_a}{\sqrt{2\pi}\,\sigma_a} \exp\left\{-\frac{(f - f_a)^2}{2\sigma_a^2}\right\} + \frac{C_b}{\sqrt{2\pi}\,\sigma_b} \exp\left\{-\frac{(f - f_b)^2}{2\sigma_b^2}\right\} \qquad (3.78)$$

where C_a and C_b are the power ratios, f_a and f_b are the frequency shifts and σ_a and σ_b are the frequency spreads of the two magnetoionic components specified by the subscripts a and b. We conclude from (3.78) that the double-Gaussian spectrum is in fact sum of two Gaussian spectrums so that we use (3.77) and (3.78) with proper parameters while dealing with Gaussian spectrums, as well.

Throughout simulations, we choose $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz to characterize the double-Gaussian spectrum given in Fig.3.14 and $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz for the Gaussian spectrum in Fig.3.15. A set of sample fading realizations for each spectrum is also provided in Fig.3.16 and Fig.3.17
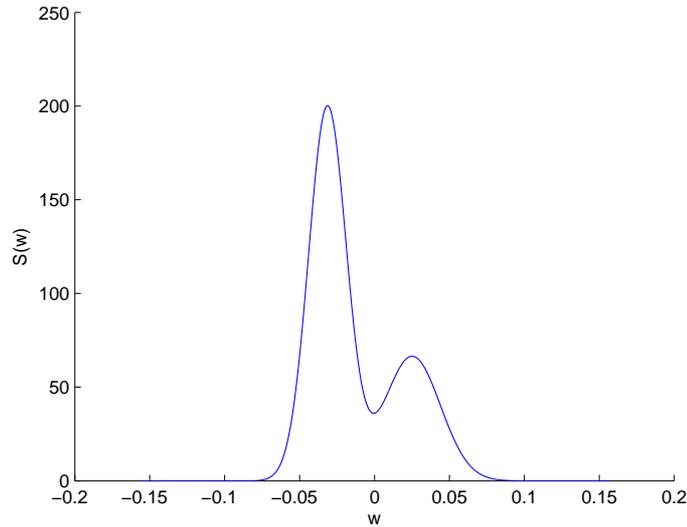
Figure 3.14: Double-Gaussian spectrum with $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz.
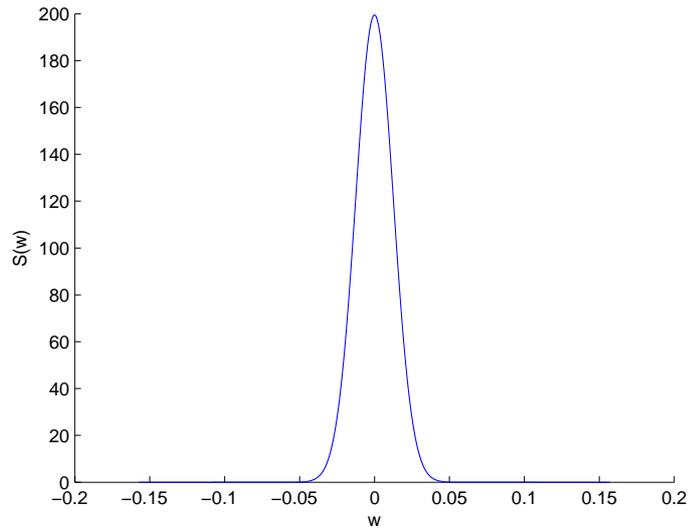
Figure 3.15: Gaussian spectrum with $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz.
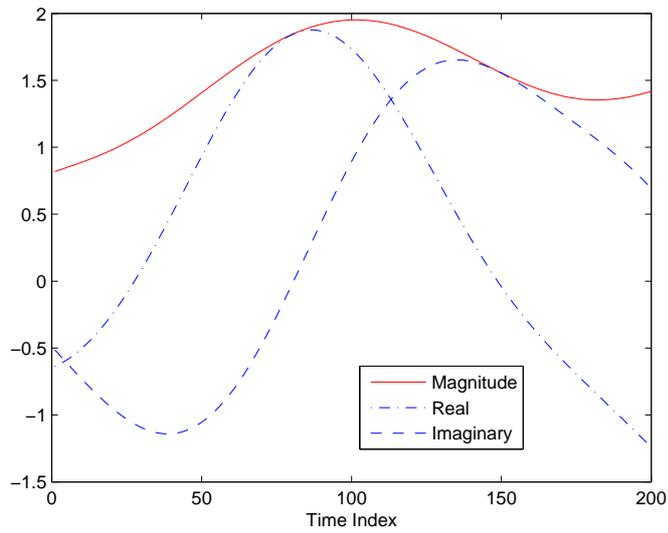
Figure 3.16: A fading realization for the double-Gaussian spectrum with $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz.
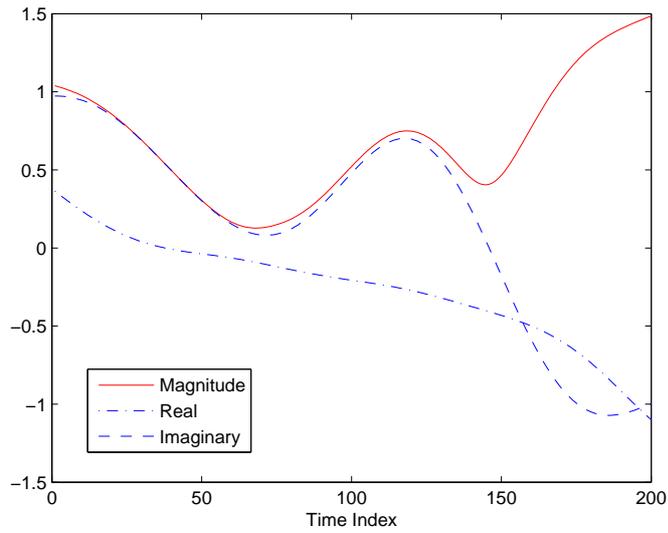
Figure 3.17: A realization for the Gaussian spectrum with $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz.

Keeping the other system parameters the same, we also plot MSIE values as before for the double-Gaussian and the Gaussian spectrums specified above at $\gamma = 5$ dB and $\gamma = 15$ dB in Fig.3.18-3.21 with varying step-size values. These results are observed to be very similar to the previous cases with Jakes' spectrum.
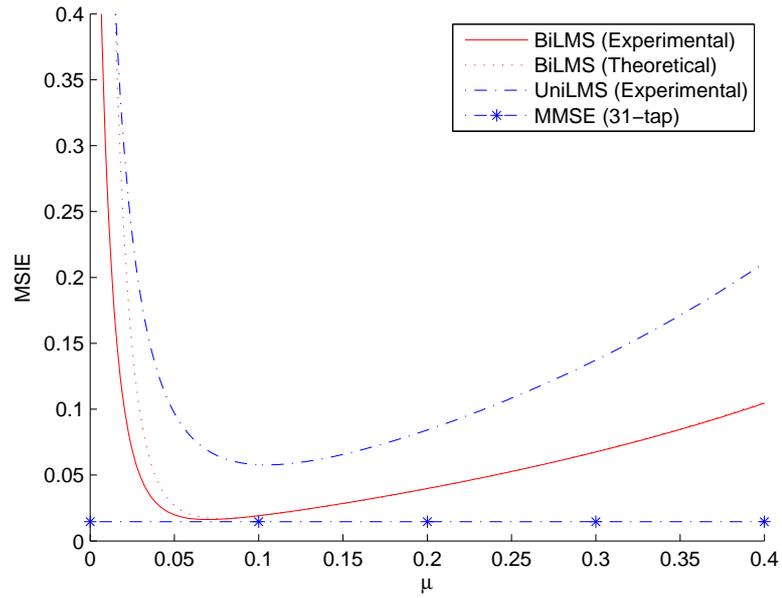
Figure 3.18: MSIE for UniLMS, BiLMS and 31-tap MMSE for the double-Gaussian spectrum with $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz at $\gamma = 5$ dB.
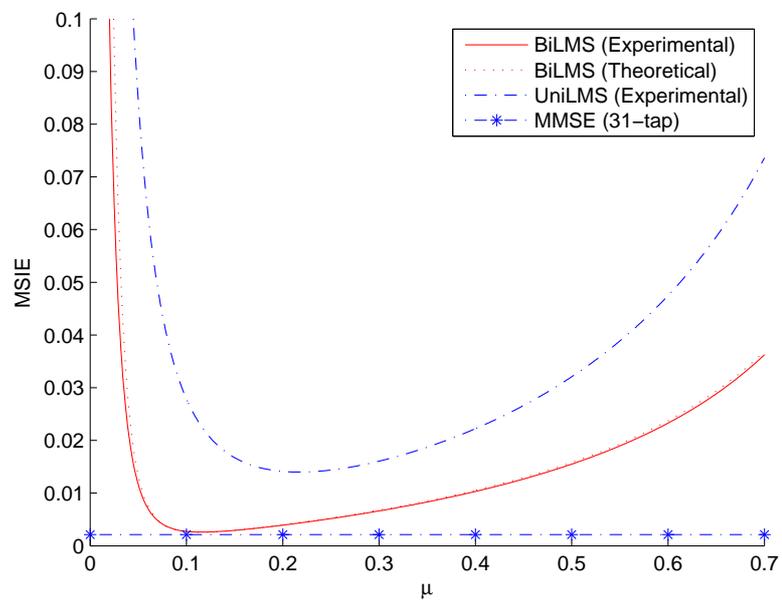
Figure 3.19: MSIE for UniLMS, BiLMS and 31-tap MMSE for the double-Gaussian spectrum with $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz at $\gamma = 15$ dB.
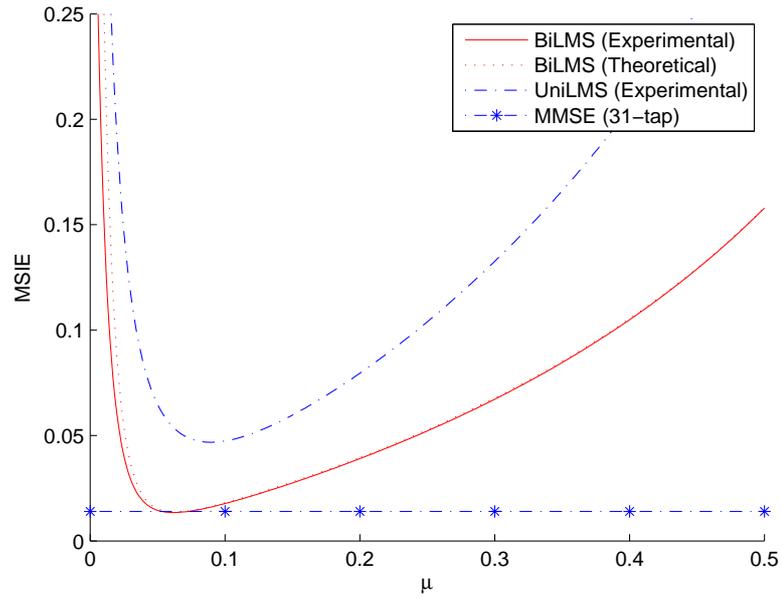
Figure 3.20: MSIE for UniLMS, BiLMS and 31-tap MMSE for the Gaussian spectrum with $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz at $\gamma = 5$ dB.
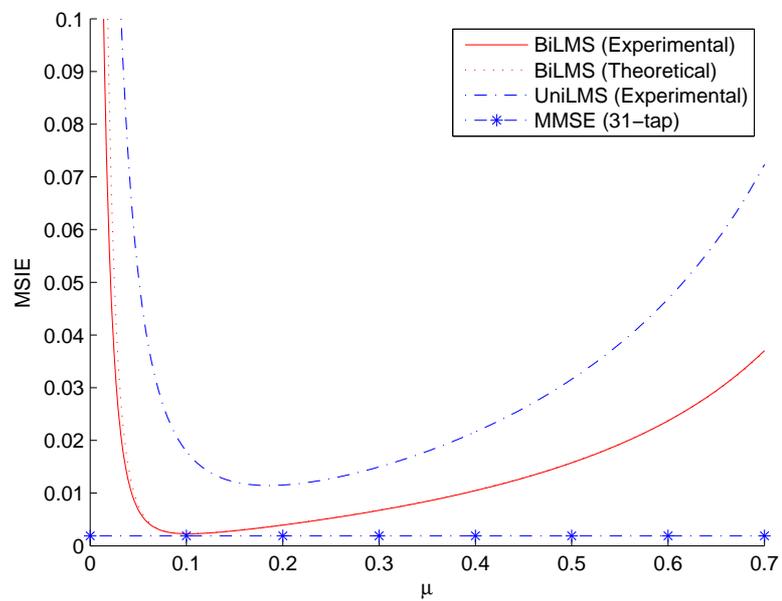
Figure 3.21: MSIE for UniLMS, BiLMS and 31-tap MMSE for the Gaussian spectrum with $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz at $\gamma = 15$ dB.

3.4 Step-Size Optimization in Flat-Fading Channels

3.4.1 Equal Forward and Backward Step-Size Values

As is the case for the conventional LMS algorithm, we observe from (3.68) that the tracking performance of the bidirectional LMS algorithm highly depends on the step-size value of the adaptations. Therefore, the optimal choice for the step-size value is of interest to characterize the minimum achievable MSE for the bidirectional LMS algorithm at the steady-state. We remind that the step-size value is chosen to be the same for the forward and the backward adaptations while deriving (3.68), and leave the effect of the independent step-size choice for the forward and the backward adaptations to the next section.

In order to derive the optimal step-size expression, μ_{opt}, theoretically, we first express (3.68) in terms of only β, and then take derivative with respect to β as follows

$$
\begin{aligned}
\frac{\partial J_{MSE}}{\partial \beta} &= \frac{\partial}{\partial \beta} \left\{ \left(1 + \frac{(1-\beta)E_s}{2(2E_s - 1 + \beta)} \right) J_{min} + \frac{E_s}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) \right|^2 S(w)\, dw \right\} \\
&= -\frac{E_s^2}{(2E_s - 1 + \beta)^2} J_{min} + \frac{E_s}{\pi} \int_{-\pi}^{\pi} H(e^{jw}) \frac{\partial H(e^{jw})}{\partial \beta} S(w)\, dw
\end{aligned}
\tag{3.79}
$$

where $\partial H(e^{jw})/\partial \beta$ is given as

$$
\begin{aligned}
\frac{\partial H(e^{jw})}{\partial \beta} &= \frac{1}{2\beta^2} - \frac{1}{2\beta^2} \left(\frac{1}{1 - \beta e^{-jw}} - \frac{1}{1 - \frac{1}{\beta}e^{-jw}} \right) \\
&\quad + \frac{(1-\beta)e^{-jw}}{2\beta} \left(\frac{1}{(1 - \beta e^{-jw})^2} + \frac{1/\beta^2}{\left(1 - \frac{1}{\beta}e^{-jw} \right)^2} \right),
\end{aligned}
\tag{3.80}
$$

or, equivalently, as follows

$$
\frac{\partial H(e^{jw})}{\partial \beta} = -\frac{(1 - \cos w)(1 - \beta^2 - 2\beta + 2\cos w)}{(1 + \beta^2 - 2\beta \cos w)^2}.
\tag{3.81}
$$

The optimal geometric ratio β_{opt} could then be evaluated numerically using (3.79) and (3.81) as follows

$$
\left. \frac{\partial J_{MSE}}{\partial \beta} \right|_{\beta = \beta_{opt}} = 0
\tag{3.82}
$$

and the optimal step-size μ_{opt} could be found as $\mu_{opt} = (1 - \beta_{opt})/2E_s$.

47

We now investigate the effectiveness of the theoretical optimal step-size values, i.e., μ_{opt}'s, computed according to (3.79) and (3.81). We present the resulting theoretical values for μ_{opt} together with the associated experimental ones for $f_d T_s = 0.01$ and $f_d T_s = 0.02$ in Table 3.1 and Table 3.2, respectively. Note that the experimental optimal step-size values are determined using 0.01 steps. We observe from Table 3.1 and Table 3.2 that the theoretical μ_{opt} values are very close to the experimental results for a variety of SNR and Doppler choices. This result is believed to have a significant practical importance since it eliminates the necessity of excessive experiments to find μ_{opt} for a genie-aided scenario.

Table 3.1: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for a Rayleigh Fading Channel with $f_d T_s = 0.01$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.070	0.070	0.090	0.100	0.100	0.120	0.130	0.140	0.140
Theoretical	0.068	0.075	0.083	0.091	0.101	0.110	0.121	0.132	0.144

Table 3.2: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for a Rayleigh Fading Channel with $f_d T_s = 0.02$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.110	0.120	0.140	0.160	0.170	0.180	0.200	0.220	0.230
Theoretical	0.111	0.124	0.137	0.150	0.165	0.180	0.196	0.213	0.231

3.4.2   Independent Forward and Backward Step-Size Values

In order to investigate the step-size optimization in a more comprehensive fashion, we now examine a different scenario in which we do not impose the forward and the backward adaptations to use strictly the same step-size value. As a result, the forward and the backward step-size values, i.e., μ^f and μ^b, could be chosen independently in order to explore for a better tracking performance. Indeed, one could argue intuitively that the forward and the backward step-size values should be chosen equally by symmetry for the best performance. Nevertheless, the following investigation is of value since it provides further understanding of the effect of the step-size and the characteristics of the transfer function.

In Section 3.2.2, the forward and the backward adaptive algorithms are introduced in (3.11) and (3.12) with independent step-size values μ^f and μ^b, respectively. We now define the

associated geometric ratios as $\beta^f = 1 - 2\mu^f E_s$ and $\beta^b = 1 - 2\mu^b E_s$. In order to examine the effect of independent choice of μ^f and μ^b on the overall tracking performance, we will first derive a corresponding MSE expression, and then try to determine the optimal values μ^f_{opt} and μ^b_{opt} analytically. To this end, we use the previous definition of J_{MSE} given in (3.22), and derive the self-noise and the lag components separately to determine J_{MSIE} according to (3.29).

We use the result of (3.52) to obtain a proper J_{self} expression associated with the independent forward and the backward step-size values as follows

$$J_{self} = \frac{E\{|v^f_k|^2\}}{4} + \frac{E\{|v^b_k|^2\}}{4} \tag{3.83}$$

$$= \frac{1}{4}\left(\frac{\mu^f}{1-\mu^f} + \frac{\mu^b}{1-\mu^b}\right) J_{min} \tag{3.84}$$

$$= \frac{1}{4}\left(\frac{1-\beta^f}{2E_s - 1 + \beta^f} + \frac{1-\beta^b}{2E_s - 1 + \beta^b}\right) J_{min} \tag{3.85}$$

where $E\{|v^f_k|^2\}$ and $E\{|v^b_k|^2\}$ is known through the steady-state MSE analysis of the the conventional LMS algorithm [57], as explained before.

In order to determine the corresponding lag component, we follow the steps of Section 3.3.2. For this purpose, we first replace μ in (3.54) and (3.55) with μ^f and μ^b, respectively, and then take the z-transforms as before to obtain

$$\hat{f}^f(z) = Z\{\hat{f}^f_k\} = \frac{1-\beta^f}{z - \beta^f} f(z) \tag{3.86}$$

$$\hat{f}^b(z) = Z\{\hat{f}^b_k\} = \frac{1-\beta^b}{z^{-1} - \beta^b} f(z). \tag{3.87}$$

The z-transform of the tracking error now becomes

$$\hat{f}(z) - f(z) = \frac{\hat{f}^f(z) + \hat{f}^b(z)}{2} - f(z) \tag{3.88}$$

$$= \frac{1}{2}\left(\frac{1-\beta^f}{z - \beta^f} + \frac{1-\beta^b}{z^{-1} - \beta^b}\right) f(z) - f(z), \tag{3.89}$$

and the transfer function $H(z)$ is found to be

$$H(z) = \frac{1}{2}\left(\frac{1-\beta^f}{z - \beta^f} + \frac{1-\beta^b}{z^{-1} - \beta^b}\right) - 1 \tag{3.90}$$

$$= -\frac{1+\beta^f}{2\beta^f} + \frac{1}{2}\left(\frac{1-\beta^f}{\beta^f}\frac{1}{1-\beta^f z^{-1}} - \frac{1-\beta^b}{\beta^b}\frac{1}{1-\frac{1}{\beta^b}z^{-1}}\right). \tag{3.91}$$

We observe from (3.91) that the poles of the transfer function $H(z)$ are $z_{p1} = \beta^f$ and $z_{p2} = 1/\beta^b$ where the first pole $z_{p1} = \beta^f$ associated with the forward adaptation lies inside the unit circle whereas the second pole $z_{p2} = 1/\beta^b$ associated with the backward adaptation lies outside of it due to the mean-convergence property of the LMS algorithm. As opposed to the equal forward and the backward step-size case, this time the transfer function is complex-valued as long as $\mu^f \neq \mu^b$. The frequency domain representation of the transfer function, which is given as

$$H(e^{jw}) = \frac{1}{2}\left(\frac{1-\beta^f}{e^{jw}-\beta^f} + \frac{1-\beta^b}{e^{-jw}-\beta^b}\right) - 1, \tag{3.92}$$

is also examined in terms of its magnitude and phase in Fig. 3.22 for $\mu_f = 0.1$ and $\mu_b = 0.3$, as an example. We observe that the phase response $\angle H(e^{jw})$ is now a nonlinear function of the frequency which implies that $H(e^{jw})$ is now complex-valued as stated before. The magnitude response is now observed to be highly nonlinear for this particular case.
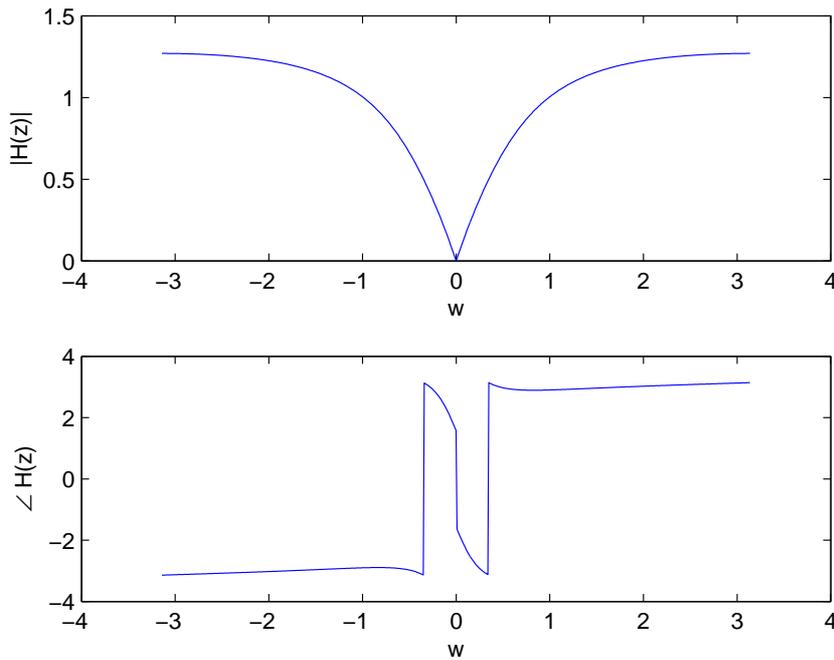
Figure 3.22: Transfer function characteristics for $\mu^f = 0.1$ and $\mu^b = 0.3$

The lag component is then evaluated using (3.66) and (3.92) as follows

$$J_{lag} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S_f(w)\,dw. \tag{3.93}$$

As a result, the overall MSIE associated with the different forward and the backward step-size case is given as

$$J_{MSIE} = J_{self} + J_{lag} \tag{3.94}$$

$$= \frac{1}{4} \left[\frac{1 - \beta^f}{2E_s - 1 + \beta^f} + \frac{1 - \beta^b}{2E_s - 1 + \beta^b} \right] J_{min}$$

$$+ \frac{E_s}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) \right|^2 S_f(w) \, dw. \tag{3.95}$$

and the corresponding MSE is

$$J_{MSE} = J_{min} + E_s \left(J_{self} + J_{lag} \right) \tag{3.96}$$

$$= \left(1 + \frac{E_s}{4} \left[\frac{1 - \beta^f}{2E_s - 1 + \beta^f} + \frac{1 - \beta^b}{2E_s - 1 + \beta^b} \right] \right) J_{min}$$

$$+ \frac{E_s}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) \right|^2 S_f(w) \, dw. \tag{3.97}$$

Since we are dealing with the expressions for the optimal forward and the backward step-size values, we take derivative of (3.97) with respect to β^f and β^b to jointly find the roots of the resulting equations as follows

$$\left. \frac{\partial J_{MSE}}{\partial \beta^f} \right|_{\beta^f = \beta^f_{opt}, \, \beta^b = \beta^b_{opt}} = 0 \tag{3.98}$$

$$\left. \frac{\partial J_{MSE}}{\partial \beta^b} \right|_{\beta^f = \beta^f_{opt}, \, \beta^b = \beta^b_{opt}} = 0. \tag{3.99}$$

Let us consider the derivative of J_{MSE} given in (3.97) with respect to β^f as follows

$$\frac{\partial J_{MSE}}{\partial \beta^f} = -\frac{E_s^2}{2(2E_s - 1 + \beta^f)^2} J_{min} + \frac{E_s}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \left| H(e^{jw}) \right|^2}{\partial \beta^f} S(w) \, dw \tag{3.100}$$

where $\partial \left| H(e^{jw}) \right|^2 / \partial \beta^f$ is given as

$$\frac{\partial \left| H(e^{jw}) \right|^2}{\partial \beta^f} = \frac{\partial H(e^{jw})}{\partial \beta^f} H(e^{jw})^* + H(e^{jw}) \frac{\partial H(e^{jw})^*}{\partial \beta^f} \tag{3.101}$$

$$= 2 \operatorname{Re} \left\{ \frac{\partial H(e^{jw})}{\partial \beta^f} H(e^{jw})^* \right\} \tag{3.102}$$

$$= 2 \operatorname{Re} \left\{ \frac{1 - e^{jw}}{2 \left(e^{jw} - \beta^f \right)^2} \left[\frac{1}{2} \left(\frac{1 - \beta^f}{e^{-jw} - \beta^f} + \frac{1 - \beta^b}{e^{jw} - \beta^b} \right) - 1 \right] \right\}. \tag{3.103}$$

Similarly, the derivative of (3.97) with respect to β^b is given as

$$\frac{\partial J_{MSE}}{\partial \beta^b} = -\frac{E_s^2}{2(2E_s - 1 + \beta^b)^2} J_{min} + \frac{E_s}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \left| H(e^{jw}) \right|^2}{\partial \beta^b} S(w) \, dw \tag{3.104}$$

51

where $\partial \left|H(e^{jw})\right|^2 /\partial \beta^b$ is

$$\frac{\partial \left|H(e^{jw})\right|^2}{\partial \beta^b} = 2\,\mathrm{Re}\left\{\frac{\partial H(e^{jw})^*}{\partial \beta^b} H(e^{jw})\right\} \tag{3.105}$$

$$= 2\,\mathrm{Re}\left\{\frac{1-e^{jw}}{2\left(e^{jw}-\beta^b\right)^2}\left[\frac{1}{2}\left(\frac{1-\beta^f}{e^{jw}-\beta^f}+\frac{1-\beta^b}{e^{-jw}-\beta^b}\right)-1\right]\right\}. \tag{3.106}$$

Indeed, β_{opt}^f and β_{opt}^b could be found using (3.100) and (3.104) together with (3.98) and (3.99). Instead, we made an important observation here to simplify the subsequent equations. If we evaluate $\frac{\partial J_{MSE}}{\partial \beta^f}$ for $(\beta^f,\beta^b)=(\beta_{opt}^b,\beta_{opt}^f)$, we have exactly the same expression $\frac{\partial J_{MSE}}{\partial \beta^b}$ evaluated for $(\beta^f,\beta^b)=(\beta_{opt}^f,\beta_{opt}^b)$ which is equal to zero by (3.99). It implies that, if $(\beta^f,\beta^b)=(\beta_{opt}^f,\beta_{opt}^b)$ is a root of $\frac{\partial J_{MSE}}{\partial \beta^f}$, then $(\beta^f,\beta^b)=(\beta_{opt}^b,\beta_{opt}^f)$ is also a root. In addition, since the quadratic error performance surfaces have a global minimum, then we obtain that $\beta_{opt}^f=\beta_{opt}^b$. Therefore, we do not proceed any more to derive β_{opt}^f and β_{opt}^b since the problem under consideration now turns out to be the one considered in Section 3.4.1.

As a result of this discussion, we conclude that one could not have further tracking performance improvement for the bidirectional LMS algorithm by just employing different step-size values for the forward and the backward adaptations.

We also verify this theoretical results through some illustrative examples. We assume a Rayleigh fading channel with Jakes' spectrum for which the normalized maximum Doppler frequency is $f_D T_s = 0.01$ with $E_s = 1$. We plot J_{MSIE} for SNR values $\gamma = 0$ dB and $\gamma = 10$ dB in Fig. 3.23 and Fig. 3.24, respectively, by making use of J_{MSIE} expression given in (3.95). We observe that the forward and the backward step-size values are equal to achieve the minimum MSIE, i.e., $J_{MSIE,min}$, and that the optimal step-size values are $\mu_{opt}^f = \mu_{opt}^b = 0.07$ for $\gamma = 0$ where the minimum MSIE is $J_{MSIE,min} = 0.047$, and $\mu_{opt}^f = \mu_{opt}^b = 0.11$ for $\gamma = 0$ where the minimum MSIE is $J_{MSIE,min} = 0.008$.
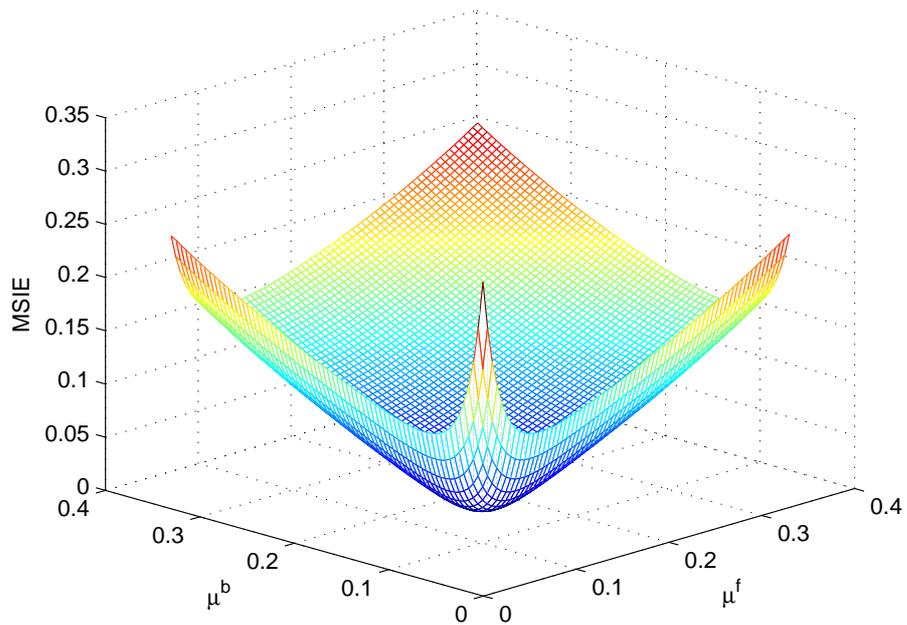
Figure 3.23: J_{MSIE} for the independent μ^f and μ^b over Rayleigh fading with $f_D T_s = 0.01$. The optimal step-size values are $\mu^f_{opt} = \mu^b_{opt} = 0.07$ with $J_{MSIE,min} = 0.047$.
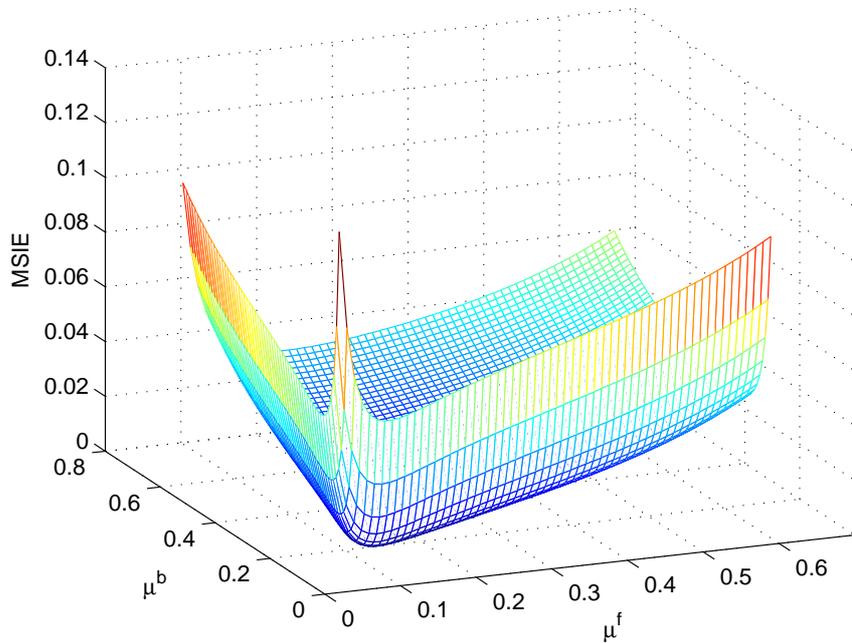
Figure 3.24: J_{MSIE} for the independent μ^f and μ^b over Rayleigh fading with $f_D T_s = 0.01$ and M-ary PSK. The optimal step-size values are $\mu^f_{opt} = \mu^b_{opt} = 0.11$ with $J_{MSIE,min} = 0.008$.

3.5 Effect of Imperfect Doppler and SNR Information

Although some a priori information on the communication channel is available in most practical systems, the statistical parameters of the underlying time-varying channel should be derived by means of the observations especially for coherent receivers. The Doppler spread of the time-varying channel and the SNR are two such examples that should be estimated carefully. In this section, we will consider the proper estimation methods for these parameters and the effects of their imperfect estimates on the estimation algorithms under consideration.

Before going further, we want to distinguish the necessity for the Doppler spread and the SNR of the estimation algorithms. The MMSE algorithm requires both of these parameters in order to design a proper filter whereas the LMS based algorithms do not have such a need to adaptively estimate the unknown fading coefficients. However, in order to achieve a satisfactory estimation performance, the LMS based algorithms require the choice of a good step-size value as is discussed in the previous sections. Although such a choice can be done by trial and error in the training mode, we have shown that the best step-size can also be determined through an analytical expression which is a function of the Doppler spread and the SNR. We will therefore use the estimates of these parameters in order to determine the best step-size value to be employed in LMS adaptations.

For both Doppler and SNR estimation, we prefer data-aided approach such that the unknown parameters are estimated by continuous transmission of a pilot sequence of length L_T which is consist of independent and identically distributed pilot symbols prior to the data block. The equivalent observation model for (3.1) during this training period is given in vector form as follows

$$\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{n} \tag{3.107}$$

where $\mathbf{y} = [y_0 \, y_1 \, \ldots \, y_{L_T-1}]$ is the observation vector, $\mathbf{A} = \text{diag}\{a_0, \, a_1, \, \ldots \, a_{L_T-1}\}$ is the diagonal pilot matrix, $\mathbf{f} = [f_0 \, f_1 \, \ldots \, f_{L_T-1}]$ is the fading vector and $\mathbf{n} = [n_0 \, n_1 \, \ldots \, n_{L_T-1}]$ is the noise vector.

54

3.5.1 Doppler Spread Estimation

There is a rich literature on the estimation of the mobile speed which in turn determines the Doppler spread. In [105], the Doppler estimation methods present in the literature are surveyed and grouped into two basic approaches which are the level crossing rate (LCR) based and covariance-based methods. The common problem of such methods is the necessity for long training or observation periods in order to have statistically efficient estimates which makes them hard to be employed in communication scenarios with short block lengths such as the one considered in this work.

In [106], an optimal maximum-likelihood (ML) approach together with a suboptimal least-squares (LS) method is proposed for Doppler spread estimation for flat-fading time-varying systems and employed in vector channels in a recent paper [107]. In [106], it is assumed that the Doppler spread does not change during Q transmitted block which is suitable for our study due to the short blocklengths. Since the proposed ML estimation requires the information of SNR which is not available a priori, we prefer to use LS estimation which is shown in [106] to achieve almost the same performance with the ML approach. This LS cost function is given as

$$F(f_d) = \frac{1}{Q} \sum_{q=1}^{Q} \sum_{l=1}^{L_T-1} \left| \frac{\hat{K}_q(l)}{\hat{K}_q(0)} - \frac{r(l; f_d)}{r(0; f_d)} \right|^2 \qquad (3.108)$$

where $\hat{K}_q(l)$ is an estimate of the autocorrelation function at the q-th block defined as

$$\hat{K}_q(l) = \frac{1}{L_T - l} \sum_{k=0}^{L_T-l} \tilde{f}_k^q (\tilde{f}_{k+l}^q)^* \qquad (3.109)$$

where $\tilde{f}_k^q$ is an estimate of the channel coefficient f_k at the q-th transmitted block [106]. In (3.108), $r(.; f_d)$ is the true autocorrelation value, e.g., the one in (3.69) for Jakes' spectrum. Note that, an estimate of the unknown fading vector during the training period for any transmitted block could be estimated by least-squares (LS) method without any need to the Doppler spread and the SNR as follows

$$\hat{\mathbf{f}} = \mathbf{A}^H \left(\mathbf{A}\mathbf{A}^H \right)^{-1} \mathbf{y}. \qquad (3.110)$$

As a result, the final Doppler estimate is given as

$$\hat{f}_d = \underset{f_d}{\operatorname{argmin}} \ F(f_d) . \tag{3.111}$$

Before the corresponding simulation results presented in Section 3.5.3, we discuss the robustness of the estimation algorithms under consideration to the imperfect Doppler estimate for $L = 200$ and BPSK modulation with $T_s = 0.1$ ms over a flat Rayleigh fading channel with Jakes' spectrum and $f_d = 100$ Hz. The degradation in MSIE due to imperfect estimate of the Doppler spread $\hat{f}_d$ is depicted through a mismatched ratio, $\hat{f}_d / f_d$, in Fig. 3.25 and Fig. 3.26 for $\gamma = 0$ dB and $\gamma = 10$ dB, respectively.
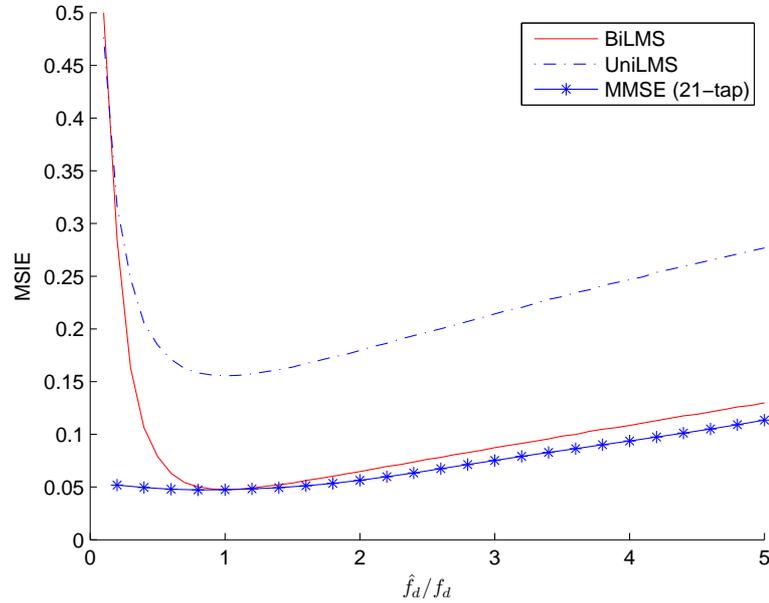
Figure 3.25: MSIE for UniLMS, BiLMS and 21-tap MMSE for imperfect Doppler spread estimate, $\hat{f}_d$, at $\gamma = 0$ dB, $f_d = 100$ Hz, $T_s = 0.1$ ms.

A number of observations could be made on the results presented in Fig. 3.25-3.26. First, degradation in MSIE is almost the same for both the bidirectional LMS algorithm and the MMSE filter at low SNR when $\hat{f}_d > f_d$ whereas the MMSE filter is observed to be more robust than the bidirectional LMS algorithm at high SNR when $\hat{f}_d > 2fd$. For any SNR value, the optimal MSIE performance does not change much when $f_d < \hat{f}_d < 2f_d$. Second, both LMS algorithms experience a serious MSIE degradation when $\hat{f}_d < 0.7f_d$. Note that underestimating the Doppler spread results in a step-size which is smaller than the optimal

one and therefore maps to the lag part of the MSIE. We know from the previous sections that the MSIE curve shows a sharp increase for the lag part as the step-size value becomes smaller than the optimal one. As a result, we could conclude that if the Doppler estimation algorithm results in an estimate such that $0.7 f_d < \hat{f}_d < 2 f_d$, we expect to observe almost no change in the minimum achievable MSIE.
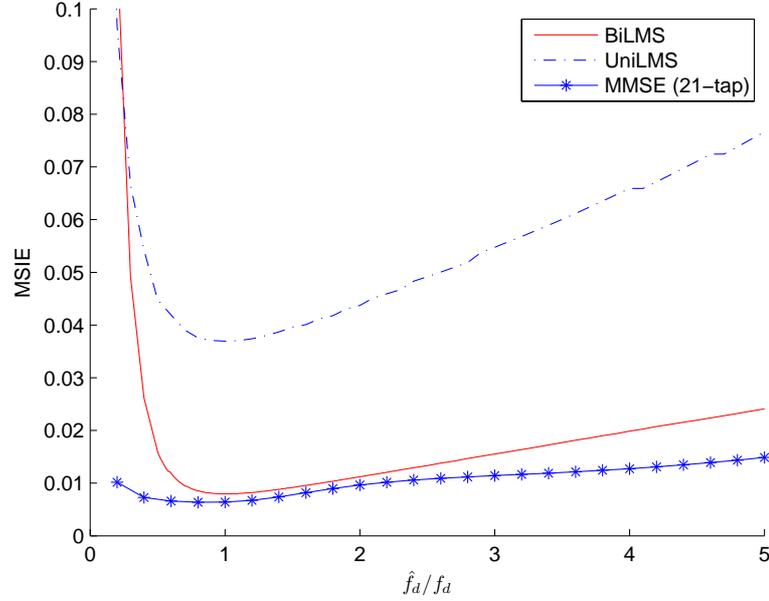
Figure 3.26: MSIE for UniLMS, BiLMS and 21-tap MMSE for imperfect Doppler spread estimate, $\hat{f}_d$, at $\gamma = 10$ dB, $f_d = 100$ Hz, $T_s = 0.1$ ms.

3.5.2 SNR Estimation

In this section, we consider ML based SNR estimation for optimal operation as is discussed in [108]. Towards this end, the probability density function of the observation vector is found to be

$$p(\mathbf{y}; \gamma | \mathbf{A}) = \frac{1}{\pi^{L_T} |\mathbf{R}_y|} \exp\left\{-\mathbf{y}^H \mathbf{R}_y^{-1} \mathbf{y}\right\} \tag{3.112}$$

and therefore the ML estimate of the SNR is given as

$$
\begin{aligned}
\hat{\gamma} &= \underset{\gamma}{\operatorname{argmax}} \ \ln p(\mathbf{y}; \gamma | \mathbf{A}) \\
&= \underset{\gamma}{\operatorname{argmax}} \left\{ -\ln |\mathbf{R}_y| - \mathbf{y}^H \mathbf{R}_y^{-1} \mathbf{y} \right\}
\end{aligned} \tag{3.113}
$$

where the observation autocorrelation matrix $\mathbf{R}_y$ is

$$\mathbf{R}_y = E\{\mathbf{y}\,\mathbf{y}^H\} = \mathbf{A}\mathbf{R}_f\mathbf{A}^H + \frac{E_s}{\gamma}\,\mathbf{I} \tag{3.114}$$

which is obviously a function of γ. Note that, we employ the Doppler estimate $\hat{f}_d$ obtained in the previous section in (3.114) in order to compute $\mathbf{R}_f$. Note also that, one could further smooth $\hat{\gamma}$ by averaging the results associated with the independent training blocks.

3.5.3 Joint Estimation of Doppler Spread and SNR

In this section, we discuss the effect of joint estimation of the Doppler spread and the SNR by means of the methods presented in the previous sections on the MSIE statistics. For this purpose, we consider flat Rayleigh fading channel with Jakes' spectrum and $f_d = 100$ Hz. A training sequence consisting of length L_T prior to a data sequence of length $L = 200$ is assumed both of which employs randomly chosen BPSK symbols with durations $T_s = 0.1$ ms. Based on the previous results, the length of the MMSE filter is chosen to be 21 to obtain the best performance.
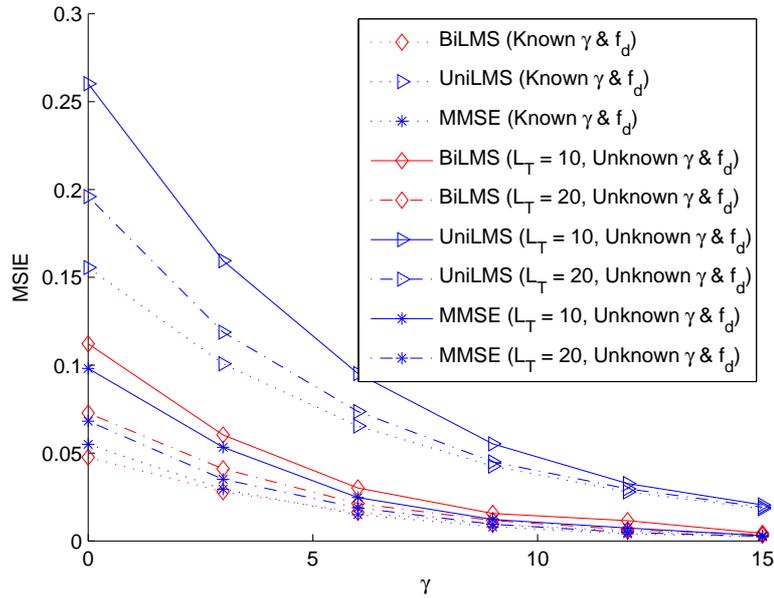
Figure 3.27: MSIE for UniLMS, BiLMS and 21-tap MMSE for known and estimated Doppler and SNR over flat Rayleigh fading channel with Jakes' spectrum and $f_d = 100$ Hz, $Q = 4$.

Fig. 3.27 presents the degradation of MSIE due to the noisy estimate of the Doppler spread

58

and the SNR for all the channel estimation algorithms under consideration after 4 fames, i.e., $Q = 4$. For the bidirectional LMS algorithm and the MMSE filter, MSIE statistics associated with a training sequence of length $L_T = 20$ is observed to be sufficiently close to the known case for any choice of the SNR. A relatively short training sequence of length $L_T = 10$ is also reported to exhibit very good performance at moderate and high SNR values. We conclude that the degradations in MSIE associated with the bidirectional LMS algorithm and the MMSE filter are very close to each other.
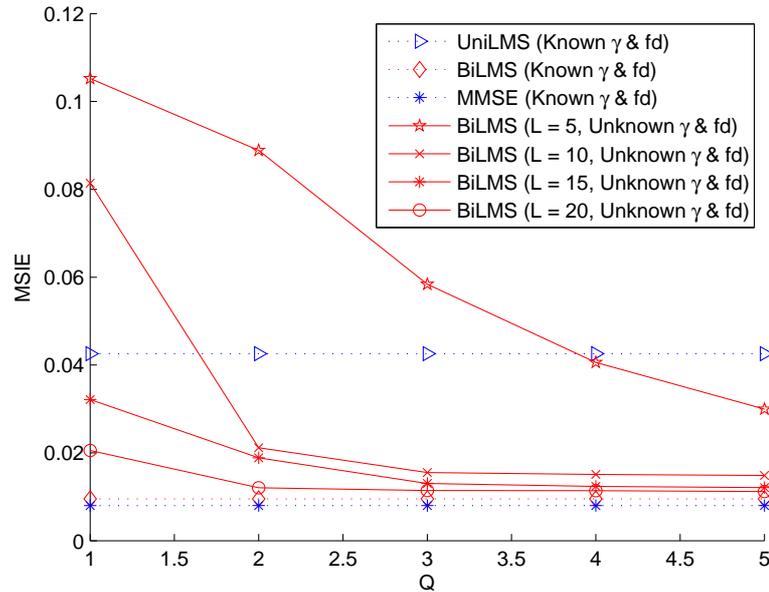
Figure 3.28: MSIE for UniLMS, BiLMS and 21-tap MMSE for known and estimated Doppler and SNR over flat Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ Hz at $\gamma = 9$ dB.

In Fig. 3.28 and Fig. 3.29, we present the effect of using multiple frames in estimating the Doppler spread and the SNR for $\gamma = 9$ dB and $\gamma = 15$ dB, respectively, for the bidirectional LMS algorithm. We observe that employing multiple frames significantly improves the MSIE performance. For example, a single training sequence of length $L_T = 10$ results in a MSIE statistics for the bidirectional LMS algorithm which is two times worse than that for the unidirectional LMS algorithm whereas it achieves a very close performance to the known case after employing $Q = 3$ frames at $\gamma = 9$ dB.
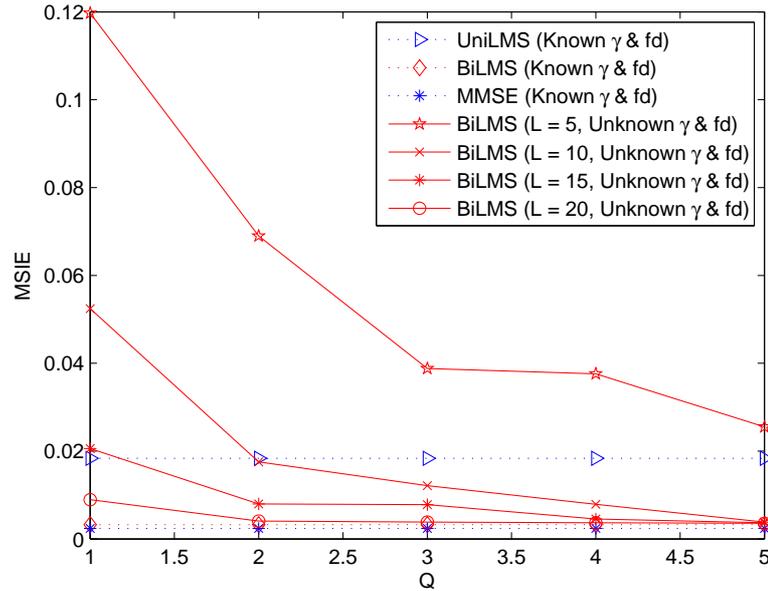
Figure 3.29: MSIE for UniLMS, BiLMS and 21-tap MMSE for known and estimated Doppler and SNR over flat Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ Hz at $\gamma = 15$ dB.

As a result, we conclude that by employing multiple independent training blocks in estimating the unknown Doppler spread and the SNR with relatively short lengths, the bidirectional LMS algorithm together with the MMSE filter stays very close to the known performance. And that, any MSIE degradation due to the imperfect estimates of these parameters are very close to each other for the aforementioned channel estimation algorithms.

3.6   Effect of Imperfect Initialization

In this section, we consider the effect of imperfect initialization on the overall performance of the bidirectional LMS algorithm. Although the scope of this work is related to the steady-state performance of the aforementioned algorithm in time-varying systems, we believed that it is of practical interest to explore the transient behavior of the algorithm when the initial value of the channel is known imperfectly or there is no such information at all. To this end, we first present the results for some simple initialization methods and then consider the associated Cramer-Rao Bound (CRB). In Section 3.8, we also provide results for the imperfect

60

initialization scenario in which unknown channel coefficient is initialized with the estimates from the previous estimation iteration.

3.6.1 Practical Initialization Methods

We employ the same data-aided framework presented in Section 3.5 where a number of L_T independent and identically distributed pilot symbols are transmitted prior to the data block. With a help of these L_T symbols, our aim is to estimate the most recent channel coefficient in the training block which will be used as an initial value for the channel coefficient at the beginning of the data block. For continuous transmission, the bidirectional LMS algorithm need only one training block which is sent prior to the data block since the backward initialization could be performed by using the training block associated with the next data block.

For the flat-fading channels under consideration, we consider both the zero and the ML initialization methods. In the former, the unknown fading coefficient at the beginning of the data block is initialized with 0 which results in the following MSIE

$$J_{Zero} = E\left\{ \left| \hat{f}_{L_T} - f_{L_T} \right|^2 \right\} = E\left\{ \left| f_{L_T} \right|^2 \right\} = 1. \tag{3.115}$$

The ML initialization is defined as

$$\hat{f}_{L_T} = a_{L_T}^* \, y_{L_T} \tag{3.116}$$

which minimizes the following probability density function

$$p\left(y_k \mid f_k, a_k\right) = \frac{1}{\pi N_0} \exp\left\{ -\frac{|y_k - f_k \, a_k|^2}{N_0} \right\} \tag{3.117}$$

according to the channel model given in (3.1). Note that the ML initialization makes use of a single pilot symbol which results in the following MSIE

$$J_{ML} = E\left\{ \left| \hat{f}_{L_T} - f_{L_T} \right|^2 \right\} = \sigma^2 , \tag{3.118}$$

as is intuitively expected.

The Monte Carlo results for the bidirectional LMS algorithm with the aforementioned initialization methods for various data block lengths and SNR are presented in Fig.3.30-3.32. We assume a flat Rayleigh fading channel with Jakes' spectrum for which $f_d T_s = 0.01$. We

observe that even when the channel is initialized with zero, a data block of length $L = 200$ is very close to the perfect initialization scenario. In addition, a short data block of length $L = 100$ with ML initialization results in an MSIE statistics which is almost the same with that of the perfect case.
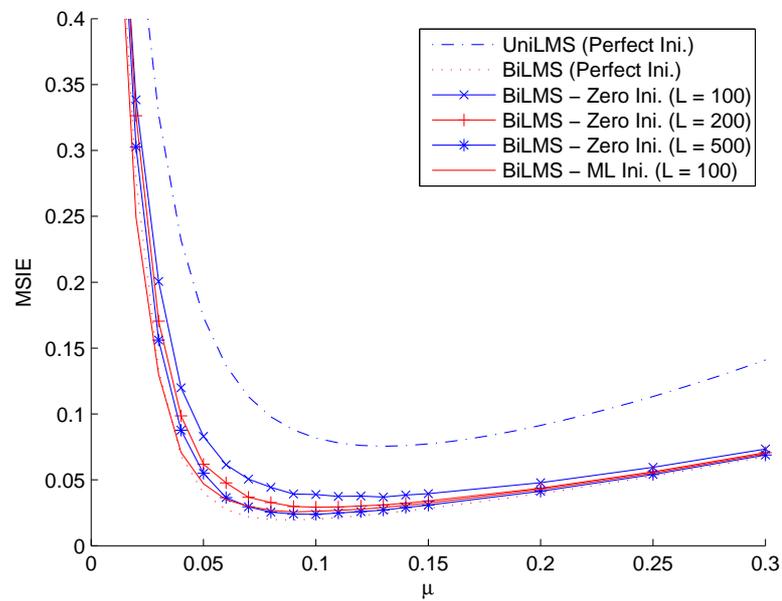
Figure 3.30: MSIE for BiLMS with zero and ML initializations together with perfectly initialized UniLMS over flat Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ at $\gamma = 5$ dB.
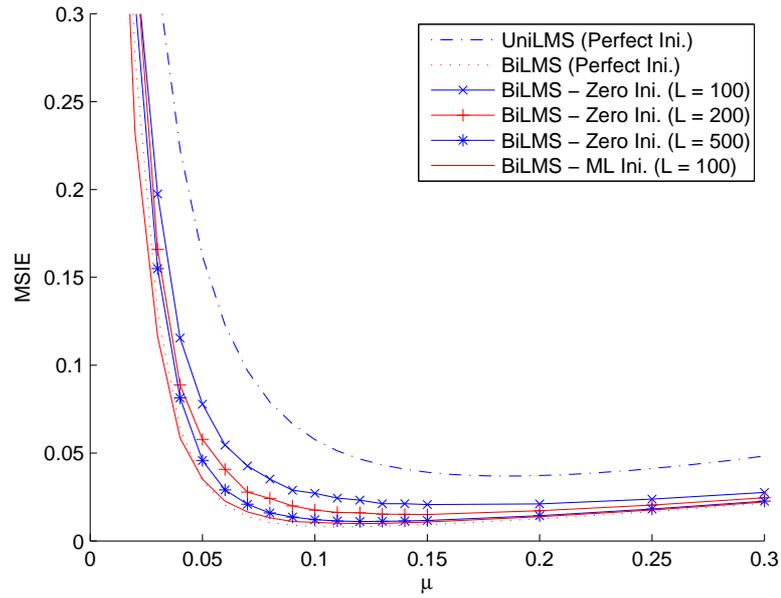
Figure 3.31: MSIE for BiLMS with zero and ML initializations together with perfectly initialized UniLMS over flat Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ at $\gamma = 10$ dB.
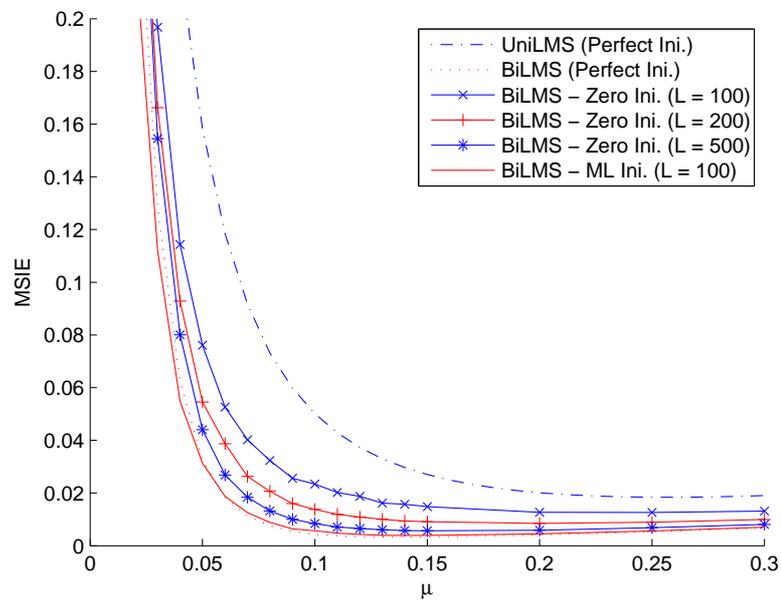
Figure 3.32: MSIE for BiLMS with zero and ML initializations together with perfectly initialized UniLMS over flat Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ at $\gamma = 15$ dB.

3.6.2 Cramer-Rao Bound for Imperfect Initialization

In this section, we derive the CRB for estimation of an unknown flat Rayleigh fading channel in order to obtain a bound on the initialization error. We assume that a number of L_T independent and identically distributed BPSK symbols are transmitted and the associated observations are received according to the system model given in (3.1). We also assume real entries for ease of presentation where the generalization to the complex case is straightforward.

In [109], the CRB is given as

$$E\left\{\left|\hat{f}_i - f_i\right|^2\right\} \geq \left(\mathbf{J}^{-1}\right)_{ii} \tag{3.119}$$

where $\mathbf{J}$ is the Fisher's Information Matrix (FIM) defined as

$$J_{ij} = E_{\mathbf{y},\mathbf{f}}\left\{\frac{\partial \ln p(\mathbf{y},\mathbf{f})}{\partial f_i}\frac{\partial \ln p(\mathbf{y},\mathbf{f})}{\partial f_j}\right\} \tag{3.120}$$

$$= -E_{\mathbf{y},\mathbf{f}}\left\{\frac{\partial^2 \ln p(\mathbf{y},\mathbf{f})}{\partial f_i \partial f_j}\right\} \tag{3.121}$$

for which the expectations are over both $\mathbf{y}$ and $\mathbf{f}$. We may further elaborate (3.121) as follows

$$J_{ij} = \underbrace{-E_{\mathbf{f}}\left\{E_{\mathbf{y}|\mathbf{f}}\left\{\frac{\partial^2 \ln p(\mathbf{y}|\mathbf{f})}{\partial f_i \partial f_j}\right\}\right\}}_{J_{ij}^1} \underbrace{-E_{\mathbf{f}}\left\{\frac{\partial^2 \ln p(\mathbf{f})}{\partial f_i \partial f_j}\right\}}_{J_{ij}^2} \tag{3.122}$$

$$= J_{ij}^1 + J_{ij}^2. \tag{3.123}$$

In order to compute J_{ij}^1, we consider the following probability density function

$$p(y_k | f_k, a_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_k - f_k a_k)^2}{2\sigma^2}\right\} \tag{3.124}$$

where σ^2 is the variance of the additive white Gaussian noise presence in the channel. In order to get rid of a_k dependency, (3.124) should be averaged over the BPSK alphabet which yields

$$p(y_k | f_k) = \sum_{i=\{-1,+1\}} P(a_k = i)\, p(y_k | f_k, a_k = i) \tag{3.125}$$

$$= \frac{1}{2\sqrt{2\pi\sigma^2}}\left(\exp\left\{-\frac{(y_k - f_k)^2}{2\sigma^2}\right\} + \exp\left\{-\frac{(y_k + f_k)^2}{2\sigma^2}\right\}\right), \tag{3.126}$$

and the following probability density function is obtained

$$p\left(\mathbf{y}\,|\,\mathbf{f}\right) = \prod_{k=1}^{L_T} p\left(y_k\,|\,f_k\right) \tag{3.127}$$

$$= (8\pi\sigma^2)^{-\frac{L_T}{2}} \prod_{k=1}^{L_T} \left(\exp\left\{-\frac{(y_k - f_k)^2}{2\,\sigma^2}\right\} + \exp\left\{-\frac{(y_k + f_k)^2}{2\,\sigma^2}\right\}\right). \tag{3.128}$$

Taking logarithm of (3.128) gives

$$\ln p\left(\mathbf{y}\,|\,\mathbf{f}\right) = -\frac{L_T}{2}\,\ln(8\pi\sigma^2) + \sum_{k=1}^{L_T} \ln\left(\exp\left\{-\frac{(y_k - f_k)^2}{2\,\sigma^2}\right\} + \exp\left\{-\frac{(y_k + f_k)^2}{2\,\sigma^2}\right\}\right) \tag{3.129}$$

with which one could compute the first order derivative as follows

$$\frac{\partial \ln p\left(\mathbf{y}\,|\,\mathbf{f}\right)}{\partial f_i} = \frac{\partial}{\partial f_i}\,\ln\left(\exp\left\{-\frac{(y_i - f_i)^2}{2\,\sigma^2}\right\} + \exp\left\{-\frac{(y_i + f_i)^2}{2\,\sigma^2}\right\}\right) \tag{3.130}$$

$$= \frac{y_i}{\sigma^2}\,\tanh\left(\frac{y_i}{\sigma^2}\,f_i\right) - \frac{f_i}{\sigma^2}, \tag{3.131}$$

and the second order derivative is readily obtained as

$$\frac{\partial^2 \ln p\left(\mathbf{y}\,|\,\mathbf{f}\right)}{\partial f_i\,\partial f_j} = \left(\frac{y_i}{\sigma^2}\,\operatorname{sech}^2\left(\frac{y_i}{\sigma^2}\,f_i\right) - \frac{1}{\sigma^2}\right)\delta_{ij}. \tag{3.132}$$

Finally, we end up with

$$J_{ij}^1 = -E_{\mathbf{f}}\left\{E_{\mathbf{y}|\mathbf{f}}\left\{\frac{y_i}{\sigma^2}\,\operatorname{sech}^2\left(\frac{y_i}{\sigma^2}\,f_i\right)\right\}\right\}\delta_{ij} + \frac{1}{\sigma^2}\,\delta_{ij} \tag{3.133}$$

$$= \left(-\frac{\sigma}{\sqrt{2\pi}}\iint_{-\infty}^{\infty} \frac{u^2 \exp\left\{-\frac{\sigma^2}{2}u^2\right\}}{\cosh(u\,v)}\,\exp\left\{-\frac{1+\sigma^2}{2\sigma^2}v^2\right\}\,du\,dv + \frac{1}{\sigma^2}\right)\delta_{ij} \tag{3.134}$$

by a change of variable which is $u = y_i/\sigma^2$ and $v = f_i$. Note that, no closed form for the expression (3.134) could be obtained, but it could fortunately be computed through numerical methods.

Now, consider the joint probability density function of the unknown channel coefficients given as

$$p\left(\mathbf{f}\right) = \frac{1}{(2\pi)^{L/2}\left|\mathbf{R}_f\right|^{1/2}}\,\exp\left\{-\frac{1}{2}\mathbf{f}^T\mathbf{R}_f\,\mathbf{f}\right\}. \tag{3.135}$$

The associated first order derivative is computed to be

$$\frac{\partial \ln p\left(\mathbf{f}\right)}{\partial f_i} = -\frac{1}{2}\sum_{m=1}^{L_T} f_m\left[\left(\mathbf{R}_f^{-1}\right)_{im} + \left(\mathbf{R}_f^{-1}\right)_{mi}\right] \tag{3.136}$$

$$= -\sum_{m=1}^{L_T} f_m\left[\left(\mathbf{R}_f^{-1}\right)_{im}\right], \tag{3.137}$$

65

and the second order differentiation therefore becomes

$$\frac{\partial^2 \ln p(\mathbf{f})}{\partial f_i \, \partial f_j} = -\left(\mathbf{R}_f^{-1}\right)_{ij} .$$

(3.138)

where (3.137) employs the fact that $\mathbf{R}_f$ is a symmetric matrix. We therefore obtained J_{ij}^2 as

$$J_{ij}^2 = -E_{\mathbf{f}}\left\{\left(-\mathbf{R}_f^{-1}\right)_{ij}\right\} = \left(\mathbf{R}_f^{-1}\right)_{ij} .$$

(3.139)

As a result, the final expression for FIM is given as

$$J_{ij} = \left(-\frac{\sigma}{\sqrt{2\pi}} \iint_{-\infty}^{\infty} \frac{u^2 \exp\left\{-\frac{\sigma^2}{2}u^2\right\}}{\cosh(u\,v)} \exp\left\{-\frac{1+\sigma^2}{2\sigma^2}v^2\right\} du\,dv + \frac{1}{\sigma^2}\right) \delta_{ij} + \left(\mathbf{R}_f^{-1}\right)_{ij} .$$

(3.140)

By using (3.119) and (3.140), we may compute the bound on the initialization error. In Table 3.3, we present the MSIE for ML initialization and the associated CRB over a flat Rayleigh fading channel with Jakes' spectrum where $f_d T_s = 0.01$ and $L_T = 1$. Note that, zero initialization yields an MSIE which is 1 all the time, as argued in the previous section. Note also that, it is possible to yield lower MSIE values and accordingly lower CRB for $L_T > 1$, but we do not explore for such scenarios since the $L_T = 1$ is shown in the previous section to achieve satisfactory performance.

Table 3.3: CRB and MSIE for ML Initialization over Flat Rayleigh Fading with $f_d T_s = 0.01$ and $L_T = 1$.

SNR	5 dB	10 dB	15 dB
CRB	0.2403	0.0909	0.0307
ML	0.3162	0.1000	0.0316

3.7 Tracking Performance of the Bidirectional LMS over AR Channels

In Section 3.3, the tracking performance of the bidirectional LMS algorithm is analyzed by deriving an MSE expression at the steady-state where the effect of the step-size choice is considered, as well. The MSE expression given by (3.68) is valid for any type of fading channels with a known spectrum. However, this generality comes with a cost such that the resulting MSE expression is not compact enough and requires a frequency domain energy computation which involves numerical integration. Although this previous result could not

be further elaborated for the fading channels with a nonrational spectrum such as the Jakes' model [98], there is a possibility to obtain a much simplified MSE expression for the fading channels with rational spectrums.

As is known, most linear time-varying communication channels could be expressed using some simplified stochastic models which have a rational power spectrum. Among various choices, auto-regressive (AR) models are commonly used to specify the correlation between the fading coefficients [110]. In such channels, the complex Gaussian fading coefficients are generated using an innovation process characterized by an AR model. This section considers the aforementioned model referred to as AR channel throughout the thesis with a purpose of simplifying the steady-state MSE expression obtained in the previous section. To this end, we carry the frequency domain computation of the lag component of the MSE into the time domain by making use of the rationality of the power spectrum for the AR channel [61]. In the end, we come up with a more compact closed form expression for the steady-state MSE which depends on the adaptation step-size, as before. Finally, an analytical expression for the optimal step-size of the adaptation is also derived as a function of SNR and AR correlation metric.

3.7.1   AR Channel Model

In Section 3.1, a general system model is introduced for flat-fading channels without any constraint on the temporal correlation characteristics. In this section, we consider a 1-st order AR process to specify the correlations between fading coefficients keeping the other model parameters the same. As a brief summary, we aim to estimate a realization of a time-varying process which is given by a sequence of complex coefficients $\{f_k\}_{k=1}^{L}$ where L is the observation length, as before. A sequence of known complex input symbols $\{a_k\}_{k=1}^{L}$ are chosen from a finite discrete alphabet $\mathcal{A}$ with symbol energies $E_s = E\{|a_k|^2\}$. The corresponding complex output symbols $\{y_k\}_{k=1}^{L}$ are observed in the presence of additive noise with the following discrete-time channel model given as

$$y_k = f_k a_k + n_k, \tag{3.141}$$

where n_k is a sample from a circularly symmetric white complex Gaussian process with zero-mean and variance N_0. The unknown fading coefficients $\{f_k\}_{k=1}^{L}$ are assumed to be generated

using an 1st order AR innovation process, without any loss of generalization, given as

$$f_k = \alpha f_{k-1} + w_k \tag{3.142}$$

where the correlation metric α is called the AR constant which is usually chosen as $\alpha = 1 - \epsilon$ to obtain a stable realization with $\epsilon > 0$ is a small positive constant. The noise term w_k in (3.142) is a sample from a circularly symmetric white complex Gaussian process with zero-mean and variance σ_w^2, and is independent of the transmitted symbols and the fading coefficients. The temporal autocorrelation associated with the AR channel under consideration is given as

$$
\begin{align}
r(n) &= E\left\{ f_{k+n} f_k^* \right\} \tag{3.143} \\
&= \alpha^n E\left\{ |f_k|^2 \right\} \tag{3.144}
\end{align}
$$

Note that, mean-square energy of the fading coefficient is found using (3.142) as follows

$$
\begin{align}
E\left\{ |f_k|^2 \right\} &= E\left\{ |\alpha f_{k-1} + w_k|^2 \right\} \tag{3.145} \\
&= \alpha^2 E\left\{ |f_{k-1}|^2 \right\} + E\left\{ |w_k|^2 \right\} \tag{3.146}
\end{align}
$$

where the assumption of statistically independence of the noise and the fading coefficient is employed in (3.144) and (3.146). Since the AR process under consideration is stationary, we have $E\{|f_{k-1}|^2\} = E\{|f_k|^2\}$, and therefore

$$E\left\{ |f_k|^2 \right\} = \frac{\sigma_w^2}{1 - \alpha^2}. \tag{3.147}$$

As a result, the temporal autocorrelation given in (3.144) becomes

$$r(n) = \alpha^n \frac{\sigma_w^2}{1 - \alpha^2}. \tag{3.148}$$

The SNR for this scenario is then given as

$$\gamma = \frac{E\{|f_k a_k|^2\}}{E\{|n_k|^2\}} = \frac{E\{|f_k|^2\} E\{|a_k|^2\}}{E\{|n_k|^2\}} = \frac{\sigma_w^2}{1 - \alpha^2} \frac{E_s}{N_0}. \tag{3.149}$$

Note that one could choose the variance of the input noise as $\sigma_w^2 = 1 - \alpha^2$ to have the customary SNR expression that is $\gamma = E_s/N_0$.

3.7.2 Steady-State MSE Analysis for AR Channels

As stated before, our purpose in this section is to further simplify the steady-state MSE expression obtained in (3.68) for AR channels and to derive a compact closed form expression for the steady-state MSE by making use of the approach of [61]. Using the results given in (3.22) and (3.29), the steady-state MSE is given as

$$J_{MSE} = J_{min} + E_s (J_{self} + J_{lag}) \tag{3.150}$$

where J_{self} and J_{lag} are previously found as

$$J_{self} = \frac{\mu}{2(1-\mu)} J_{min} \tag{3.151}$$

$$J_{lag} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) \right|^2 S(w) \, dw \tag{3.152}$$

where the transfer function $H(e^{jw})$ is given by (3.62) in the z-domain, and $S(w)$ is the power spectrum of the unknown fading channel. We observe from (3.150)-(3.152) that the self-noise part of the MSE is in a simplified form, which solely depends on the step-size of the adaptations and the minimum achievable MSE, whereas the lag component is not as compact as the former one and involves numerical integration which results in an unavoidable complexity in frequency domain energy computation. We therefore focus on the derivation of the lag component and will derive it in the time domain to simplify the resulting expression as opposed to the the frequency domain evaluation as in (3.152).

In order to better understand and utilize the AR innovation process, we take z-transform of (3.142) as follows [61]

$$\underbrace{Z\{f_k\}}_{f(z)} = \alpha \underbrace{Z\{f_{k-1}\}}_{z^{-1} f(z)} + \underbrace{Z\{w_k\}}_{w(z)} \tag{3.153}$$

by which we obtain a relation between the z-transform of the fading coefficient f_k and the input noise w_k, i.e., $f(z)$ and $w(z)$, as follows

$$f(z) = \frac{1}{1 - \alpha e^{-jw}} w(z) \tag{3.154}$$

$$= H_{AR}(z) w(z) \tag{3.155}$$

where $H_{AR(z)}$ is the transfer function associated with the first-order AR process which is given as

$$H_{AR}(e^{jw}) = \frac{1}{1 - \alpha e^{-jw}}. \tag{3.156}$$

69

Because our aim is to compute the lag part of MSE which is discussed previously to be the average energy in the tap-weight tracking error, $\hat{f}_k - f_k$, under the assumption of perfect gradient estimation, we incorporate the result of (3.60) into (3.155) as follows

$$\hat{f}(z) - f(z) = H(z) f(z) \tag{3.157}$$

$$= \underbrace{H(z) H_{AR}(z)}_{H_T(z)} w(z). \tag{3.158}$$

where $H_T(z)$ is defined to be the overall transfer function of the AR channel.
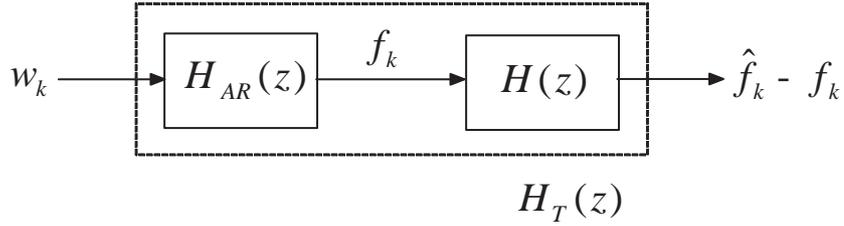
Figure 3.33: The overall transfer function $H_T(z)$ to obtain the tracking error out of input noise over AR channels.

In Fig. 3.33, the filtering operation given in (3.158) is expressed in a block diagram representation. By making use of the relation given in Fig. 3.33, the average energy in the tap-weight tracking error, which is the lag part to be computed, could be evaluated as follows

$$J_{lag} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) H_{AR}(e^{jw}) \right|^2 S_w(w) \, dw \tag{3.159}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| H_T(e^{jw}) \right|^2 \sigma_w^2 \, dw \tag{3.160}$$

where $S_w(w)$ is the power spectral density of the white Gaussian noise process w_k which is given as $S_w(w) = \sigma_w^2$. Note that the power spectral density $S(w)$ of the fading coefficients is a rational function of the frequency as follows

$$S(w) = |H_{AR}(e^{jw})|^2 S_w(w) \tag{3.161}$$

$$= \frac{\sigma_w^2}{\left| 1 - \alpha e^{-jw} \right|^2}. \tag{3.162}$$

In order to avoid from the numerical integration present in (3.160) which is the ultimate purpose of this section, we make use of the well-known Parseval's theorem [102] and translate the infinite integration in the frequency domain into the infinite summation in the discrete time

domain as follows

$$J_{lag} = \frac{\sigma_w^2}{2\pi} \int_{-\pi}^{\pi} \left| H_T(e^{jw}) \right|^2 dw = \sigma_w^2 \sum_{n=-\infty}^{\infty} |h_T[n]|^2. \qquad (3.163)$$

In order to proceed with the purpose of computing the energy present in (3.163), we should evaluate the unit step response $h_T[n]$. To this end, we first compute the partial fraction expansion of the overall transfer function $H_T(z)$ as follows

$$
\begin{aligned}
H_T(z) &= H(z)\,H_{AR}(z) = \left(-\frac{1+\beta}{2\beta} + \frac{1-\beta}{2\beta}\left(\frac{1}{1-\beta z^{-1}} - \frac{1}{1-\frac{1}{\beta}z^{-1}} \right) \right) \left(\frac{1}{1-\alpha z^{-1}} \right) \\
&= -\left(\frac{1+\beta}{2\beta} \right) \frac{1}{1-\alpha z^{-1}} + \left(\frac{1-\beta}{2\beta} \right) \left[\frac{1}{1-\alpha z^{-1}} \frac{1}{1-\beta z^{-1}} - \frac{1}{1-\alpha z^{-1}} \frac{1}{1-\frac{1}{\beta}z^{-1}} \right] \\
&= A(\alpha,\beta)\frac{1}{1-\alpha z^{-1}} - B(\alpha,\beta)\frac{1}{1-\beta z^{-1}} - C(\alpha,\beta)\frac{1}{1-\frac{1}{\beta}z^{-1}} \qquad (3.164)
\end{aligned}
$$

where the coefficients are given as a function of α and β as follows

$$
\begin{aligned}
A(a,\beta) &= \frac{(1+\beta)(1-\alpha)^2}{2(\alpha-\beta)(1-\alpha\beta)} & (3.165) \\
B(a,\beta) &= \frac{(1-\beta)}{2(\alpha-\beta)} & (3.166) \\
C(a,\beta) &= \frac{(1-\beta)}{2\beta(1-\alpha\beta)}. & (3.167)
\end{aligned}
$$

In order to obtain a stable discrete-time sequence $h_T[n]$, we should specify a proper region of convergence (ROC) such that the unit circle is included [103]. The locations of the poles of $H_T(z)$ which are $z_{p1} = \alpha$, $z_{p2} = \beta$ and $z_{p3} = 1/\beta$ are determined as follows

- Due to the assumption made in Section 3.7.1 that ϵ is a small positive constant, $z_{p1} = \alpha$ lies inside the unit circle at a position very close to 1 with the relation $|\alpha| = |1 - \epsilon| < 1$.

- Since $|\beta| = |1 - 2\mu E_s| < 1$ is the mean-convergence condition of the conventional LMS algorithm, $z_{p2} = \beta$ lies inside the unit circle and $z_{p3} = 1/\beta$, therefore, lies outside the unit circle.

As a result, the desired ROC including the unit circle is a ring given as

$$\text{ROC: } \max(\alpha,\beta) < |z| < 1/\beta,$$

71
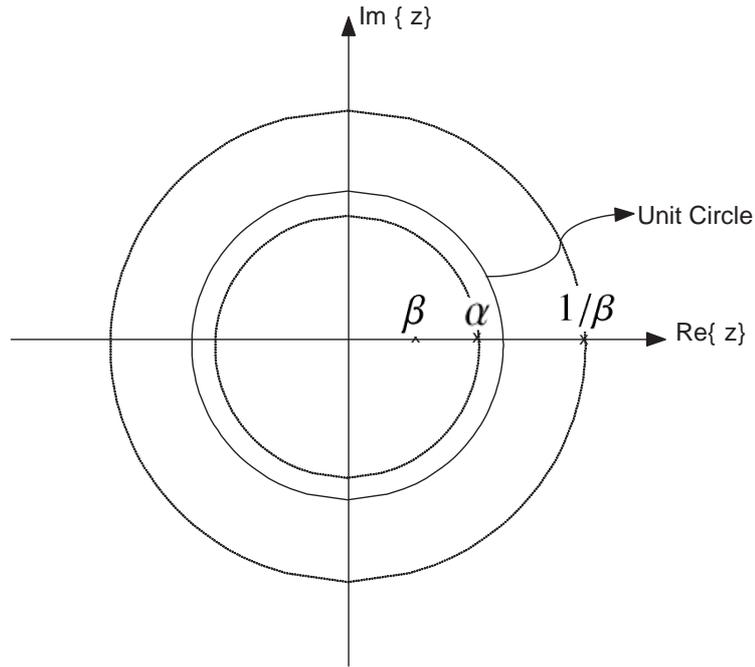
Figure 3.34: ROC of the overall transfer function $H_T(z)$.

and is depicted in Fig. 3.34 assuming $\max(\alpha,\beta) = \alpha$ and $\beta > 0$ without any loss of generality.

Using the final ROC given as $\max(\alpha,\beta) < |z| < 1/\beta$, the unit step response $h_T[n]$ is computed to be

$$h_T[n] = A(\alpha,\beta)\,\alpha^n u[n] \; - \; B(\alpha,\beta)\beta^n u[n] \; + \; C(\alpha,\beta)\left(\frac{1}{\beta}\right)^n u[-n-1], \tag{3.168}$$

which may be expressed equivalently as

$$h_T[n] = \begin{cases} A(\alpha,\beta)\,\alpha^n \; - \; B(\alpha,\beta)\beta^n, & n \geq 0; \\ C(\alpha,\beta)\left(\frac{1}{\beta}\right)^n, & n < 0. \end{cases} \tag{3.169}$$

The average energy of $h_T[n]$ is then computed in time domain as follows

$$\sum_{n=-\infty}^{\infty} |h_T[n]|^2 \quad = \quad \sum_{n=0}^{\infty} \left| A(\alpha,\beta)\,\alpha^n - B(\alpha,\beta)\beta^n \right|^2 + \sum_{n=-\infty}^{-1} \left| C(\alpha,\beta)\beta^{-n} \right|^2 \tag{3.170}$$

$$= \quad \sum_{n=0}^{\infty} \left(A^2(\alpha,\beta)\,\alpha^{2n} + B^2(\alpha,\beta)\beta^{2n} - 2A(\alpha,\beta)\,B(\alpha,\beta)\,(\alpha\beta)^n \right)$$

$$+ \sum_{n=-\infty}^{-1} C^2(\alpha,\beta)\beta^{-2n} \tag{3.171}$$

$$= \quad \frac{A(\alpha,\beta)^2}{1-\alpha^2} + \frac{B(\alpha,\beta)^2 + C(\alpha,\beta)^2\beta^2}{1-\beta^2} - \frac{2A(\alpha,\beta)\,B(\alpha,\beta)}{1-\alpha\beta}. \tag{3.172}$$

Consequently, J_{lag} is found with the help of (3.163) and (3.172) as follows

$$J_{lag} \quad = \quad \sigma_w^2 \Big(\frac{A(\alpha,\beta)^2}{1-\alpha^2} + \frac{B(\alpha,\beta)^2 + C(\alpha,\beta)^2\beta^2}{1-\beta^2} - \frac{2A(\alpha,\beta)\,B(\alpha,\beta)}{1-\alpha\beta} \Big). \tag{3.173}$$

As a result, the overall MSIE expression becomes

$$J_{MSIE} = J_{self} + J_{lag}$$

$$= \frac{\mu}{2(1-\mu)} J_{min} + \sigma_w^2 \left(\frac{A(\alpha,\beta)^2}{1-\alpha^2} + \frac{B(\alpha,\beta)^2 + C(\alpha,\beta)^2\beta^2}{1-\beta^2} - \frac{2A(\alpha,\beta)\,B(\alpha,\beta)}{1-\alpha\beta} \right),$$

$$\tag{3.174}$$

and the associated MSE expression is

$$J_{MSE} = \left(1 + \frac{\mu E_s}{2(1-\mu)} \right) J_{min} + \sigma_w^2 E_s \left(\frac{A(\alpha,\beta)^2}{1-\alpha^2} + \frac{B(\alpha,\beta)^2 + C(\alpha,\beta)^2\beta^2}{1-\beta^2} - \frac{2A(\alpha,\beta)\,B(\alpha,\beta)}{1-\alpha\beta} \right).$$

$$\tag{3.175}$$

When we compare the final MSE expressions in (3.68) and (3.175), we observe that (3.175) associated with the AR channel appears to be in a more compact form in the sense that it does not require numerical integration as opposed to (3.68).

We could also derive the optimal choice for the step-size, i.e., μ_{opt}, using (3.175). Since the derivative of (3.175) with respect to the geometric ratio β is extensively complicated, we explore for ways to further simplify the final MSE expression in (3.175). To this end, we first observe that the term $A(\alpha,\beta)$ has a relatively small value as compared to $B(\alpha,\beta)$ and $C(\alpha,\beta)$ due to the $(1-\alpha^2)$ factor in the numerator since the AR constant α was previously chosen to be close to 1. Therefore, J_{MSIE} could be approximated by the following expression as a

function of β as follows

$$J_{MSIE} \approx \frac{\mu}{2(1-\mu)}J_{min} + \sigma_w^2 \frac{B(\alpha,\beta)^2 + C(\alpha,\beta)^2 \beta^2}{1-\beta^2}$$

$$= \frac{1-\beta}{2(2E_s - 1 + \beta)}J_{min} + \frac{\sigma_w^2}{4}\frac{1-\beta}{1+\beta}\left(\frac{1}{(\alpha-\beta)^2} + \frac{1}{(1-\alpha\beta)^2}\right) \quad (3.176)$$

with high accuracy especially when $\alpha \to 1$ which corresponds to a moderate time variation. This assumption is justified in the next section through numerical examples. The corresponding J_{MSE} expression becomes

$$J_{MSE} \approx \left(1 + \frac{(1-\beta)E_s}{2(2E_s - 1 + \beta)}\right)J_{min} + \frac{\sigma_w^2 E_s}{4}\frac{1-\beta}{1+\beta}\left(\frac{1}{(\alpha-\beta)^2} + \frac{1}{(1-\alpha\beta)^2}\right)$$

with the following derivative with respect to β as follows

$$\frac{\partial J_{MSE}}{\partial \beta} \approx -\frac{E_s^2}{(2E_s - 1 + \beta)^2}J_{min} + \frac{\sigma_w^2 E_s}{2}\left[-\frac{1}{(1+\beta)^2}\left(\frac{1}{(\alpha-\beta)^2} + \frac{1}{(1-\alpha\beta)^2}\right)\right.$$

$$\left. + \frac{1-\beta}{1+\beta}\left(\frac{1}{(\alpha-\beta)^3} + \frac{\alpha}{(1-\alpha\beta)^3}\right)\right]. \quad (3.177)$$

The optimal values β_{opt} and μ_{opt} could then be evaluated numerically as follows

$$\left.\frac{\partial J_{MSE}}{\partial \beta}\right|_{\beta=\beta_{opt}} = 0 \quad (3.178)$$

$$\mu_{opt} = \frac{1-\beta_{opt}}{2E_s}. \quad (3.179)$$

3.7.3   Numerical Results

In this section, we verify the theoretical MSE derivation for the bidirectional LMS algorithm operating at the steady-state over a communication channel characterized by a 1-st order AR process. To this end, we perform extensive Monte Carlo simulations and compare the associated MSIE results with the theoretical ones computed using (3.174). We assume BPSK modulation so that the transmitted symbols, i.e., $\{a_k\}_{k=1}^L$, are chosen from the binary alphabet $\mathcal{A} = \{-1, +1\}$ in an equally likely fashion where $L = 1000$, and the symbol energy is therefore given as $E_s = 1$.
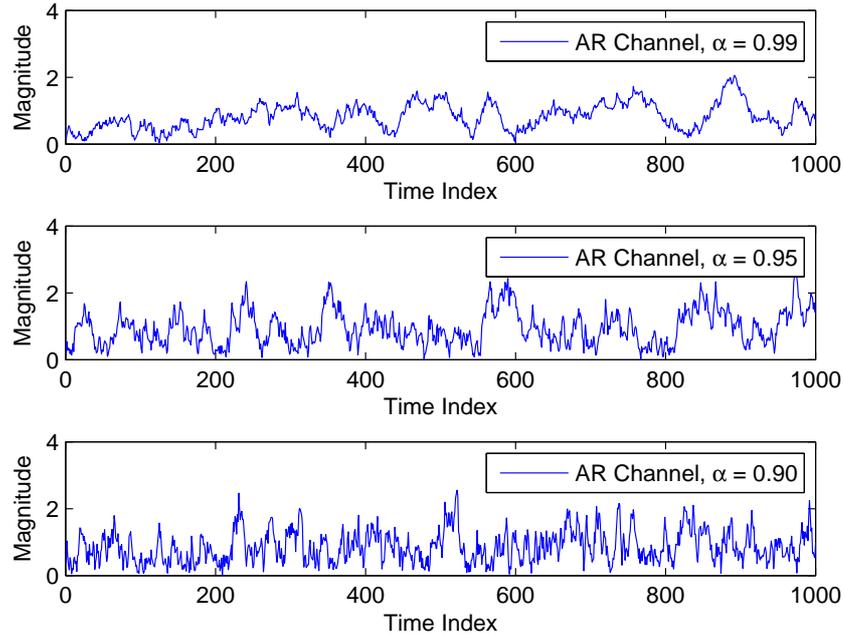
Figure 3.35: AR channel realizations generated by a white complex Gaussian noise with the variance $\sigma_w^2 = 1 - \alpha^2$ where $\alpha = \{0.99, 0.95, 0.90\}$.

Fading channel realizations are generated according to (3.142) with $\sigma_w^2 = 1 - \alpha^2$. As a result, we have unit energy fading coefficients, i.e., $E\{|f_k|^2\} = 1$, and the resulting SNR therefore becomes $\gamma = 1/\sigma^2$. We consider three different channel speeds which are characterized by a set of AR constants $\alpha = \{0.99, 0.95, 0.90\}$. In Fig. 3.35, a single realization of the channel with one of the AR constants under consideration is depicted. We observe that even a small increase in the AR constant α, the speed of the corresponding channel increases significantly. In the sequel, we therefore consider the cases with $\alpha = \{0.99, 0.95, 0.90\}$ for the purpose of presenting the accuracy of the steady-state MSE and optimal step-size expressions, and no other comparisons are provided since such cases have no practical meaning in the scope of this work.
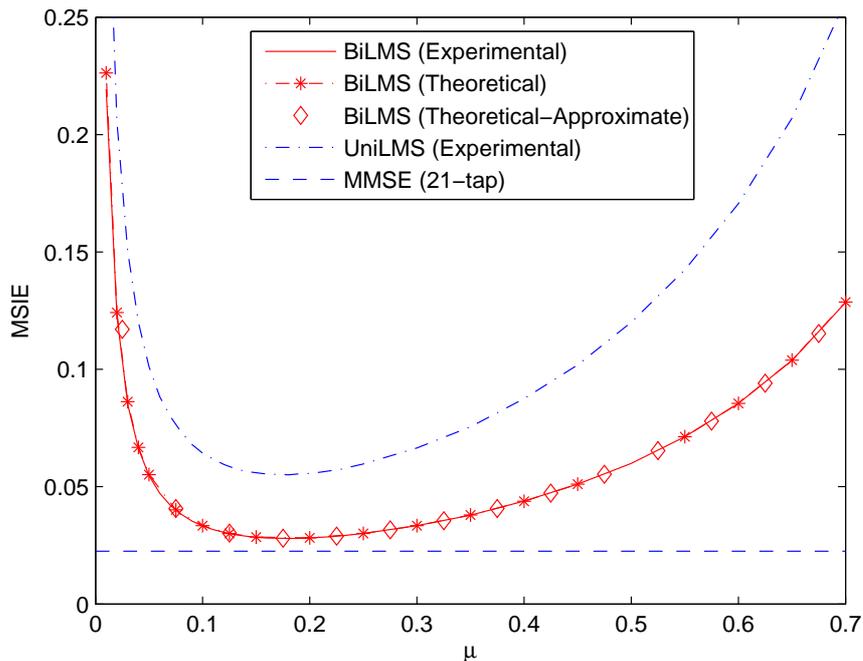
Figure 3.36: Theoretical and experimental MSIE for BiLMS together with the experimental MSIE for UniLMS and a 21-tap MMSE filter with varying step-size and $\alpha = 0.99$ at $\gamma = 10$ dB SNR. The approximate theoretical MSIE for BiLMS given in (3.176) is also provided.

In Fig. 3.36, the theoretical and experimental MSIE for the bidirectional LMS algorithm are depicted together with the approximate MSIE given by (3.176) for varying step-size and $\alpha = 0.99$ at $\gamma = 10$ dB SNR. The experimental MSIE for the unidirectional LMS algorithm and the optimal 21-tap MMSE filter are also presented for comparison purposes. We first observe that the theoretical results obtained numerically using MSIE expression given in (3.174) perfectly match the experimental results for any choice of the adaptation step-size. Therefore, the associated MSE derivation ended up with the expression given by (3.175) is verified. Second, the theoretical result for the approximation to (3.174) which is given by (3.176) is also observed to follow the original exact MSIE results perfectly showing the accuracy of the approximation to be used in the optimal step-size computations. Third, the superiority of the bidirectional LMS algorithm in tracking of the unknown AR channel is verified by a much better performance than the conventional unidirectional LMS and by a similar performance with the optimal Wiener filter.

In Fig. 3.37, the exact/approximate theoretical and the experimental MSIE results for the

76

bidirectional LMS algorithm together with the experimental MSIE for the unidirectional LMS algorithm are presented for optimal step-size and varying SNR. The experimental MSIE for the optimal MMSE filter with 21-tap is again added. We observe from Fig. 3.37 that the conclusions made for Fig. 3.36 are also valid for the varying SNR case. We also depict the MSIE performance of the optimal Wiener filter with various number of taps in Fig. 3.38.
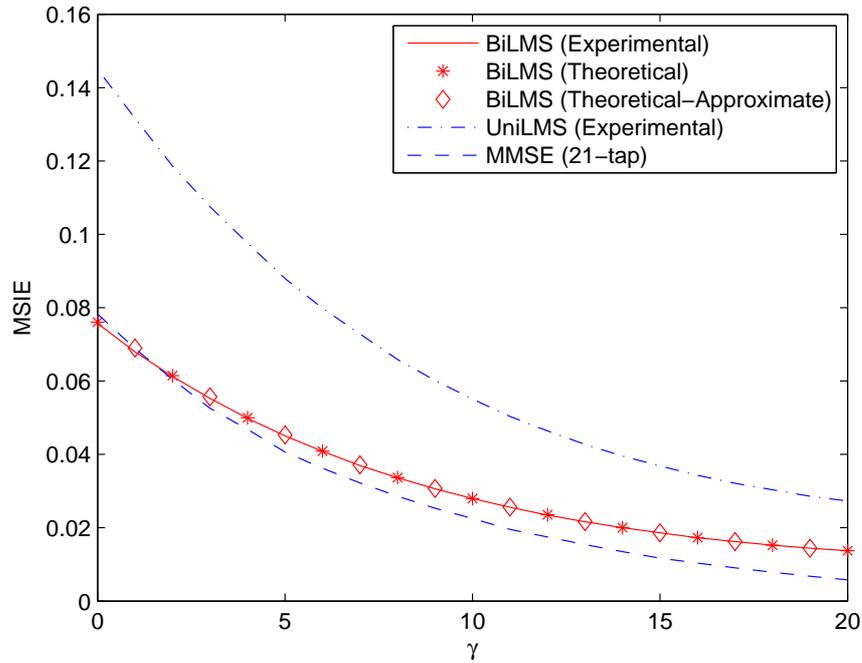
Figure 3.37: Theoretical and experimental MSIE for BiLMS together with the experimental MSIE for UniLMS and a 21-tap MMSE filter with optimal step-size (μ_{opt}) and $\alpha = 0.99$ for varying SNR. The approximate theoretical MSIE (3.176) for BiLMS with (μ_{opt}) is also provided.

In Fig. 3.39 and Fig. 3.40, the MSE derivation for the bidirectional LMS algorithm is verified over much faster AR channels with $\alpha = 0.95$ and $\alpha = 0.9$, respectively. We observe from Fig. 3.39 and Fig. 3.40 that the resulting MSIE, and equivalently MSE, expression for the bidirectional LMS algorithm has a very good match to the experimental results even for very fast AR channels. We also observe that the approximate theoretical MSIE result has some deviation from the original statistics under these highly time varying environments especially for small step-size values, and is very close to the original one at the optimal step-size values, i.e., at the minimum of the performance surface. Therefore, this approximation is again of value in computing of the optimal step-size.

77
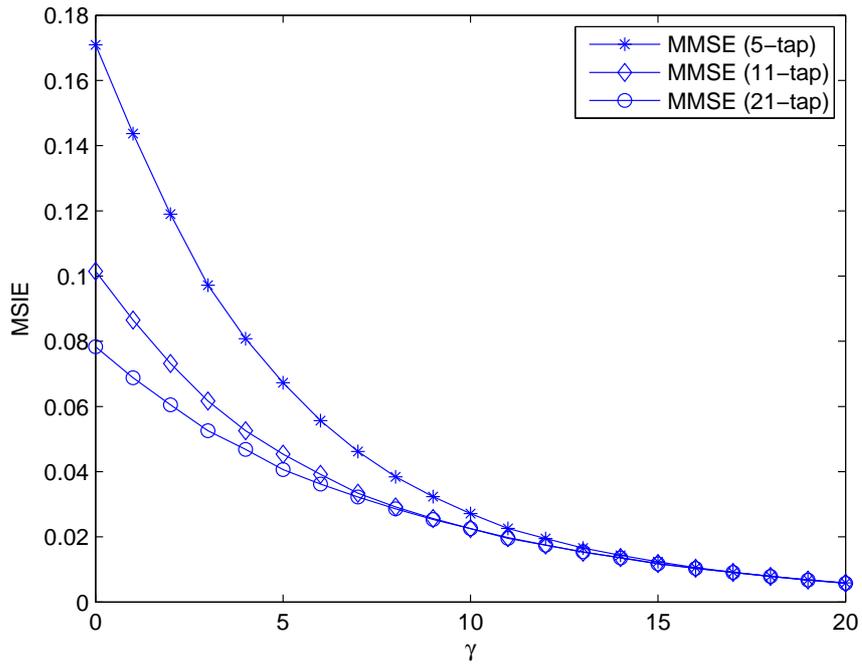
Figure 3.38: Experimental MSIE for the MMSE filter with various number of taps for $\alpha = 0.99$.
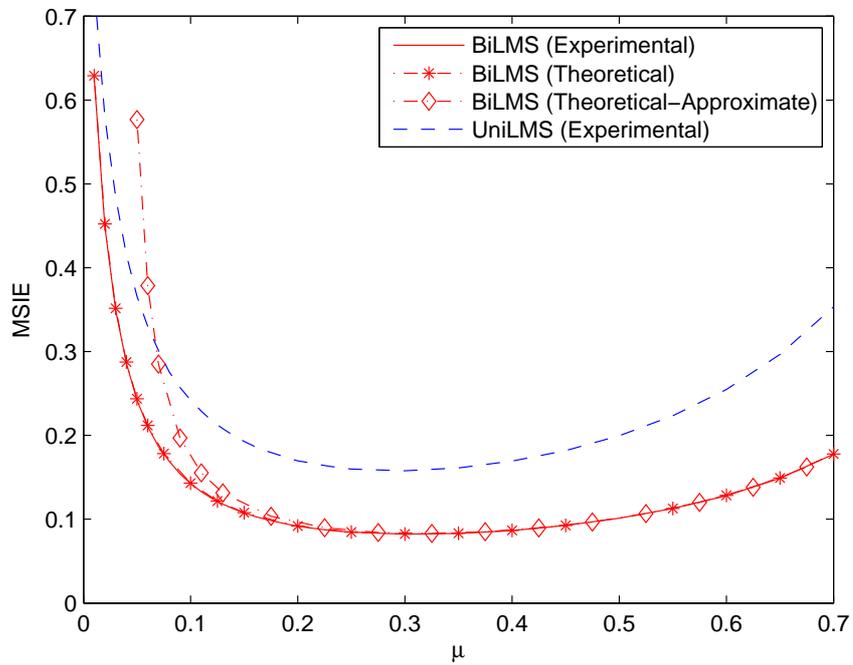
Figure 3.39: Theoretical and experimental MSIE for BiLMS together with the experimental MSIE for UniLMS with varying step-size and $\alpha = 0.95$ at $\gamma = 10$ dB SNR.
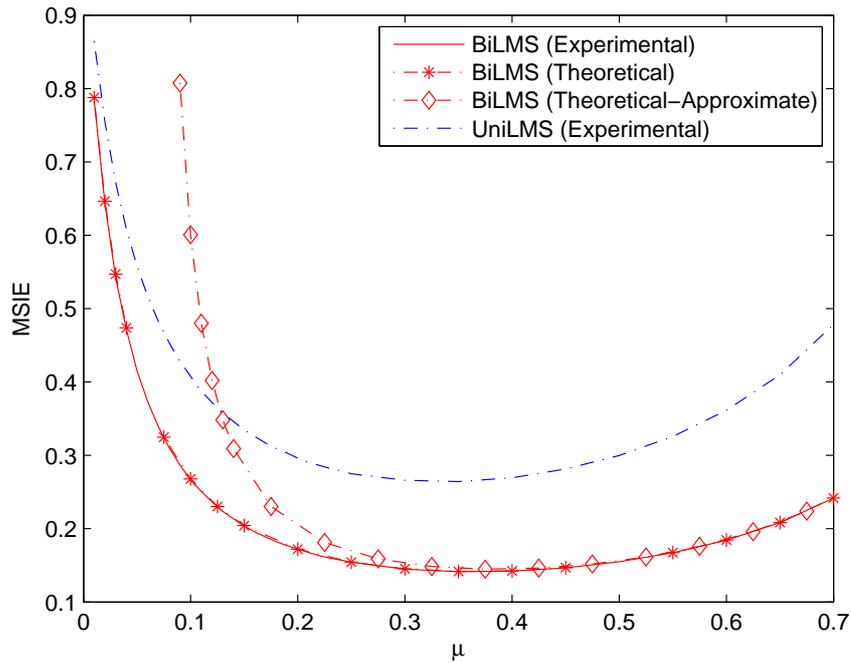
Figure 3.40: Theoretical and experimental MSIE for BiLMS together with the experimental MSIE for UniLMS with varying step-size and $\alpha = 0.90$ at $\gamma = 10$ dB SNR.

Because the choice of the adaptation step-size of the bidirectional LMS algorithm has a major effect on the overall tracking performance, we investigate the the effectiveness of the theoretical optimal steps-size values, i.e., μ_{opt}'s, computed according to (3.177) and (3.178). Table 3.4-3.6 present the theoretical and experimental results for μ_{opt} under various SNR levels and for a set of AR constants given as $\alpha = \{0.99, 0.95, 0.90\}$, respectively. We observe that the theoretical steps-size values are close to the experimental results for many cases. We again note that, since the MSE is almost constant around the neighborhood of μ_{opt} as seen from Fig. 3.39 and 3.40, the moderate differences between the theoretical and experimental values have no impact on the overall tracking performance.

Table 3.4: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for an AR Channel with $\alpha = 0.99$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.060	0.080	0.100	0.130	0.150	0.190	0.220	0.250	0.300
Theoretical	0.050	0.061	0.075	0.092	0.113	0.138	0.167	0.201	0.240

Table 3.5: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for an AR Channel with $\alpha = 0.95$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.130	0.170	0.190	0.230	0.250	0.300	0.350	0.300	0.400
Theoretical	0.119	0.140	0.165	0.195	0.230	0.270	0.312	0.356	0.397

Table 3.6: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for an AR Channel with $\alpha = 0.90$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.170	0.210	0.250	0.300	0.350	0.350	0.400	0.400	0.450
Theoretical	0.180	0.206	0.236	0.271	0.310	0.352	0.392	0.427	0.454

3.8 Iterative Channel Estimation for Flat-Fading Channels

Iterative channel estimation is a well-known technique in which estimates of the transmitted symbols are employed together with the a priori known pilot symbols in order to improve the quality of estimation in an iterative fashion. One of the basic drawbacks of this approach is the increasing computational complexity through the recurring iterations. The bidirectional LMS algorithm with its computationally efficient adaptations is therefore considered to be a good choice to be employed with iterative channel estimation idea.

This section deals with a more realistic communication system with a channel code and iteratively employed pilot-aided channel estimation. The basic goal of this section is to compare the estimation and tracking performances of the estimation algorithms under consideration in a communication system very close to real-life applications.

3.8.1 Transmitter and Receiver Models

In this section, we make use of the same equivalent discrete-time complex baseband channel model given in (3.1) except that the transmitted symbols $\{a_k\}_{k=1}^{L}$ are not chosen independently from the finite alphabet $\mathcal{A}$ any more. In order to adopt the PSAM transmission and the iterative channel estimation techniques, we employ the transmitter and the receiver models given in Fig. 3.41 and Fig. 3.42, respectively.
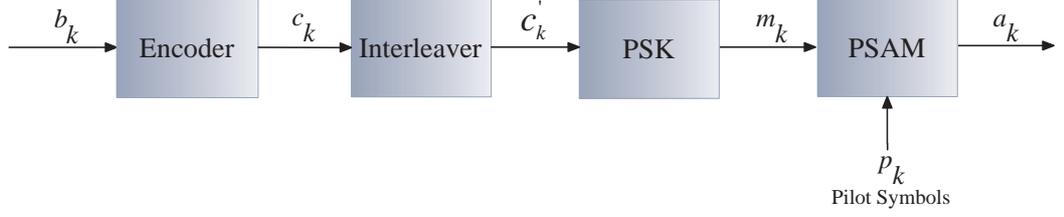
Figure 3.41: Transmitter model including channel encoder, interleaver, PSK modulator and PSAM block.

At the transmitter shown in Fig. 3.41, a set of binary information symbols $\{b_k\}_{k=1}^{L_d}$ of length L_d, which are chosen from the binary set $\{0, 1\}$ in an independent and identical fashion, are first encoded by a channel code of rate R_c. The set of coded symbols $\{c_k\}_{k=1}^{L_c}$ at the output of the encoder are then interleaved to combat with the burst errors, and the resulting set of symbols $\{c'_k\}_{k=1}^{L_c}$ are modulated using a PSK modulator with a finite modulation alphabet $\mathcal{A}$ where L_c is the block length of the coded symbols with $L_c = \frac{L_d}{R_c}$. The modulated symbols $\{m_k\}_{k=1}^{L_m}$, where $L_m = \frac{L_d}{R_c \log_2 |\mathcal{A}|}$ is the associated block length and $|\mathcal{A}|$ is the cardinality of the alphabet $\mathcal{A}$, are then multiplexed with a set of pilot symbols $\{p_k\}_{k=1}^{L_p}$ which are known a priori at the receiver and are chosen from the same alphabet $\mathcal{A}$ where L_p is the number of pilot symbols in use. The multiplexing operation is performed such that the modulated symbols are first split into the groups of $M_p - 1$ symbols and a single pilot symbol is inserted periodically into the center of each of these groups. This transmission scheme is known as PSAM where M_p is referred to as the pilot symbol spacing and is assumed to be odd as discussed in [14].

According to the aforementioned transmission scheme, the necessary number of pilot symbols are given as

$$L_p = \left\lfloor \frac{L_m - \frac{M_p-1}{2}}{M_p - 1} \right\rfloor + 1 = \left\lfloor \frac{\frac{L_d}{R_c \log_2 |\mathcal{A}|} - \frac{M_p-1}{2}}{M_p - 1} \right\rfloor + 1. \tag{3.180}$$

As a result, the data sequence $\{a_k\}_{k=1}^{L}$ which is produced after multiplexing is expressed as follows

$$a_k = \begin{cases} m_{k - \left\lfloor \frac{k-(M_p+1)/2}{M_p} \right\rfloor}, & k \neq (i - 1) M_p + \frac{M_p+1}{2}, \ i = 1, 2, \ldots, L_p; \\ p_{\frac{k-(M_p+1)/2}{M_p}}, & k = (i - 1) M_p + \frac{M_p+1}{2}, \ i = 1, 2, \ldots, L_p. \end{cases} \tag{3.181}$$

where the associated frame length becomes

$$L = \frac{L_d}{R_c \log_2 |\mathcal{A}|} + L_p = \frac{L_d}{R_c \log_2 |\mathcal{A}|} + \left\lfloor \frac{\frac{L_d}{R_c \log_2 |\mathcal{A}|} - \frac{M_p-1}{2}}{M_p - 1} \right\rfloor + 1 \tag{3.182}$$

81

by (3.180). We also define the set of indices associated with the pilot symbols as follows

$$P_p = \left\{ k_p(i) \right\}_{i=1}^{L_p} = \left\{ \frac{M_p + 1}{2}, \frac{3M_p + 1}{2}, \ldots, \frac{(2L_p - 1)M_p + 1}{2} \right\} \qquad (3.183)$$

where $k_p(\cdot)$ stands for any of the indices of the pilot symbols in a transmitted block and is given as

$$k_p(i) = (i - 1)\, M_p + \frac{M_p + 1}{2}. \qquad (3.184)$$

The data sequence $\{a_k\}_{k=1}^{L}$ is then transmitted over the flat-fading channel defined in (3.1). Note that since the optimal design of the patterns or the values for the pilot symbols is not considered in this work, we choose p_k to be equal to one of the elements of the modulation alphabet $\mathcal{A}$ without any loss of generality.
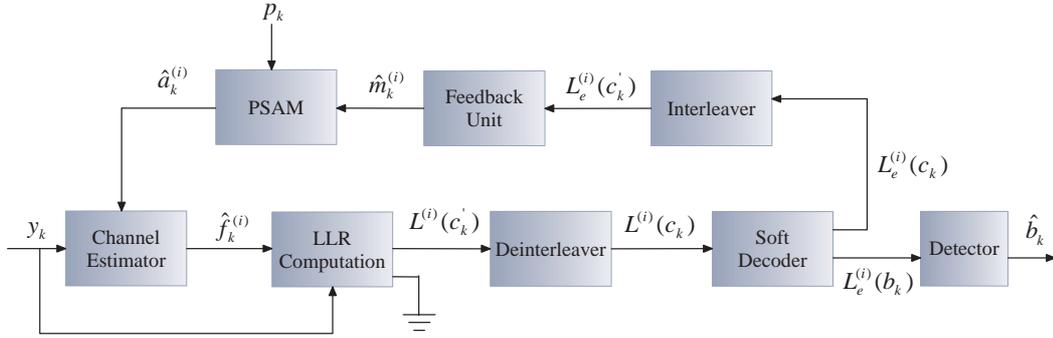
Figure 3.42: Receiver model for a time-varying flat-fading channel with iterative channel estimation.

At the receiver side, an iterative channel estimation technique is employed with a soft-input soft-output decoder as shown in Fig. 3.42. Therefore, any parameter or the function devoted to the i-th channel estimation iteration takes the superscript (i), and is updated at each estimation iteration. In this scheme, the observations are first fed to the channel estimator unit which employs one of the estimation algorithms explored later in this section. The initial estimate of the channel is obtained by making use of the observations and the a priori known pilot symbols only. In the subsequent estimation iterations, the quality of the channel estimate is improved by making use of the soft information on the coded symbols provided by the decoder.

Once an estimate of the unknown channel is obtained, the log-likelihood values (LLRs) of the

transmitted symbols are computed at the i-th iteration as follows

$$L^{(i)}(a_k) = \log \frac{P\left(y_k \,\middle|\, \hat{f}_k^{(i)}, a_k \in \mathcal{A} \setminus \{a_{ref}\}\right)}{P\left(y_k \,\middle|\, \hat{f}_k^{(i)}, a_{ref}\right)}, \tag{3.185}$$

where a_{ref} is a reference symbol from the modulation alphabet $\mathcal{A}$ and $\hat{f}_k^{(i)}$ is the estimate of the k-th fading coefficient at the i-th iteration. The probabilities are computed in (3.185) with the assumption that $\hat{f}_k^{(i)}$ is the actual fading coefficient and its incorrectness is therefore not accounted for. Since the decoder needs the LLRs for the binary set $\{c_k\}_{k=1}^L$, the LLRs associated with the pilot symbols are removed from the set $\left\{L^{(i)}(a_k)\right\}_{k=1}^L$, and the residual LLRs are then deinterleaved. The remaining symbol level LLRs, i.e., $L^{(i)}(m_k)$'s, are then converted into the desired bit level LLRs, i.e., $L^{(i)}(c_k)$'s, details of which are given in [111, 112, 113].

There are a number of observations to be made about this formulation. First, a number of $|\mathcal{A}| - 1$ symbol level LLRs are to be computed according to (3.185). Second, the computation in (3.185) is not optimal, but is hopefully suboptimal. One reason for this claim is that it is assumed that $L^{(i)}(a_k)$'s for different k's are uncorrelated although they are not since these values are function of the estimates of the fading coefficients $\hat{f}_k^{(i)}$ which are obviously correlated. As a final note, if the higher order modulation alphabets are employed, i.e., $|\mathcal{A}| > 2$, LLR computation and feedback units given in Fig. 3.41 are capable of symbol-to-bit level LLR conversion and vice versa.

Assuming BPSK modulation, LLR computation given in (3.185) becomes

$$L^{(i)}(a_k) = \log \frac{P(y_k \,|\, \hat{f}_k, a_k = +1)}{P(y_k \,|\, \hat{f}_k, a_{ref} = -1)}. \tag{3.186}$$

If the additive noise present in the channel is Gaussian, then we obtain

$$L^{(i)}(a_k) = \log\left(\frac{1}{\pi N_0} \exp\left\{-\frac{|y_k - \hat{f}_k|^2}{N_0}\right\}\right) - \log\left(\frac{1}{\pi N_0} \exp\left\{-\frac{|y_k + \hat{f}_k|^2}{N_0}\right\}\right) \tag{3.187}$$

$$= \frac{|y_k + \hat{f}_k|^2}{N_0} - \frac{|y_k - \hat{f}_k|^2}{N_0} \tag{3.188}$$

$$= \frac{4\,\mathrm{Re}\left\{y_k^* \hat{f}_k\right\}}{N_0}. \tag{3.189}$$

where we have used the complex Gaussian distribution with zero mean and N_0 variance given as

$$f_x(x) = \frac{1}{\pi N_0} \exp\left\{-\frac{|x|^2}{N_0}\right\}. \tag{3.190}$$

The soft decoder in the receiver employs the maximum a posteriori (MAP) algorithm in the log domain, which is referred to as the log-MAP algorithm and described in detail in [114]. The values given in (3.189) are processed by the soft decoder after removal of those associated with the pilot symbols and the extrinsic LLRs of both the coded, i.e., $L_e^{(i)}(c_k)$'s, and the uncoded symbols, i.e., $L_e^{(i)}(b_k)$'s, are computed accordingly [114]. At each channel estimation iteration except the last one, $L_e^{(i)}(c_k)$'s are sent to the feedback unit to refine the channel estimate of the previous iteration. After the last iteration, $L_e^{(N_I)}(b_k)$'s are sent to the detector to obtain the estimates of the information bits as follows

$$
\hat{b}_k = \begin{cases} 1, & L_e^{(N_I)}(b_k) \geq 0; \\ 0, & L_e^{(N_I)}(b_k) < 0. \end{cases} \tag{3.191}
$$

where N_I is the number of channel estimation iterations.

The feedback unit in Fig. 3.41 involves a set of operations which are to convert the extrinsic LLRs of c_k's to those of m_k's for $|\mathcal{A}| > 2$ in the same way explained in [112], and to evaluate the estimates of the modulated symbols either in the soft or the hard manner. The soft estimate of a modulated symbol is indeed the ensemble average of the corresponding LLRs and given in [82] for BPSK as

$$
\hat{m}_k^{(i)} = E\left\{ L_e^{(i)}(m_k) \right\} = \tanh\left(\frac{L_e^{(i)}(m_k)}{2} \right), \tag{3.192}
$$

When we choose Gray-coded QPSK with the constellation diagram in Fig.3.43, the soft estimates are given as [115]

$$
\hat{m}_k^{(i)} = -\frac{1+j}{2} \tanh\left(\frac{L_e^{(i)}(c_{2k})}{2} \right) - \frac{1-j}{2} \tanh\left(\frac{L_e^{(i)}(c_{2k+1})}{2} \right), \tag{3.193}
$$

and 8-PSK modulation given in Fig.3.43 results in

$$
\hat{m}_k^{(i)} = A \, \tanh\left(\frac{L_e^{(i)}(c_{2k})}{2} \right) + B \, \tanh\left(\frac{L_e^{(i)}(c_{2k+1})}{2} \right)
$$
$$
+ \left[C \, \tanh\left(\frac{L_e^{(i)}(c_{2k})}{2} \right) + D \, \tanh\left(\frac{L_e^{(i)}(c_{2k+1})}{2} \right) \right] \tanh\left(\frac{L_e^{(i)}(c_{2k+2})}{2} \right) \tag{3.194}
$$

where $A = \frac{1-(\sqrt{2}+1)j}{4}$, $B = \frac{1+\sqrt{2}+j}{4}$, $C = \frac{1-\sqrt{2}+j}{4}$, $D = \frac{1+(\sqrt{2}-1)j}{4}$.
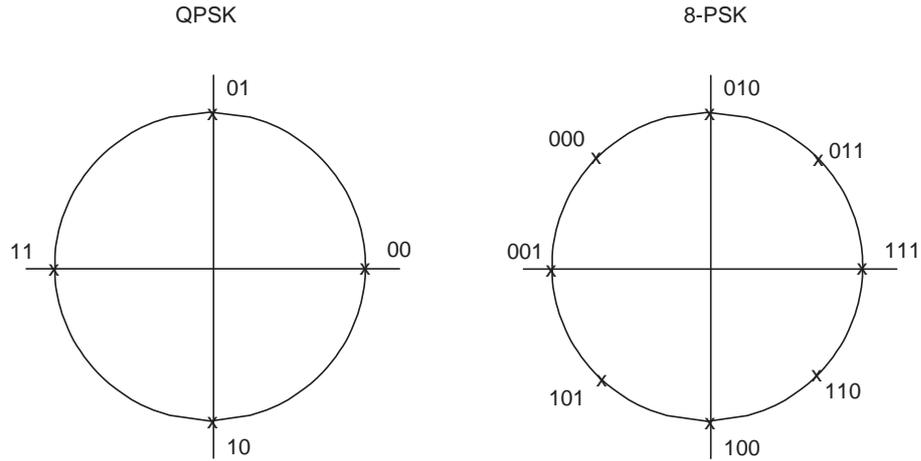
Figure 3.43: QPSK and 8-PSK constellation diagrams.

Finally, the hard decision feedback is performed as

$$\hat{m}_k^{(i)} = \begin{cases} 1, & L_e^{(i)}(m_k) \geq 0; \\ -1, & L_e^{(i)}(m_k) < 0. \end{cases} \qquad (3.195)$$

These estimates are then multiplexed with the known pilot symbols as in the transmitter to produce the overall estimates of the transmitted symbols $\hat{a}_k^{(i)}$ at the estimation iteration i.

After the initial channel estimation, the soft or the hard estimates of the modulated symbols $\hat{m}_k^{(i)}$ are incorporated into the channel estimation algorithm together with the pilot symbols as if they were the actual transmitted symbols. A new channel estimate is then computed using not only the pilot symbols but also the estimated values of modulated symbols. By this technique, quality of the channel estimate is expected to improve through iterations.

3.8.2 Channel Estimation Algorithms for Flat-Fading Channels

In this section, we revisit the channel estimation algorithms introduced in Section 3.2 to perform some modifications explicitly which are necessary for the communication scenario under consideration. We assume that the initial channel estimation is performed by the MMSE filter using the pilot symbols only since both the LMS algorithms need the knowledge of the transmitted symbols which is not available initially for the data symbols. Once an initial estimate is obtained, we could use any of the LMS algorithms as well as the Wiener filter since estimates

85

of the data symbols are available as explained in the previous section.

3.8.2.1   MMSE Channel Estimation

In iterative channel estimation, the MMSE filters associated with the initial and the subsequent iterations are different since the former is using only the pilots which are apart from each other by a number of data symbols while the latter is employing the data symbols as well so that the symbols in use are next to each other. We therefore present the MMSE channel estimation and derivation of the associated filters for the initial and the subsequent estimation iterations separately.

An initial estimate of the unknown channel is obtained by using only the pilot symbols as follows

$$\hat{f}_k^{(1)} = \sum_{n=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,n}\, a_{n_p(k)-nM_p}\, y_{n_p(k)-nM_p}. \tag{3.196}$$

where $n_p(k)$ is the index of the pilot symbol closest to a_k, i.e., $a_{n_p(k)}$ is a pilot symbol, and is given as

$$n_p(k) = \begin{cases} \left\lfloor \dfrac{k-\frac{M_p+1}{2}}{M_p} \right\rfloor M_p + \dfrac{M_p+1}{2}\,, & k < \frac{1}{2} + \left(\left\lfloor \dfrac{k-\frac{M_p+1}{2}}{M_p} \right\rfloor + 1 \right) M_p \;\; ; \\[2em] \left(\left\lfloor \dfrac{k-\frac{M_p+1}{2}}{M_p} \right\rfloor + 1 \right) M_p + \dfrac{M_p+1}{2}\,, & k > \frac{1}{2} + \left(\left\lfloor \dfrac{k-\frac{M_p+1}{2}}{M_p} \right\rfloor + 1 \right) M_p \;\; . \end{cases} \tag{3.197}$$

with a special case at the right edge of the observation block due to the PSAM transmission given as

$$n_p(k) = \frac{(2L_p-1)M_p+1}{2} \qquad \text{if} \quad \left\lfloor \frac{k-\frac{M_p+1}{2}}{M_p} \right\rfloor + 1 > L_p\,. \tag{3.198}$$

Note that, the filtering operation given in (3.196) includes the pilot symbol term explicitly unlike the conventional formulation given in (3.4). The reason behind this choice is to obtain a single global MMSE filter for each transmitted packet detail of which is given in Appendix A.2.1. The associated Wiener-Hopf's equations is also derived in Appendix A.2.1 with the final form given as

$$\sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,l} \left\{ r_f\big((n-l)M_p\big) + N_0\,\delta_{nl} \right\} = r_f\left(k - n_p(k) + nM_p\right) \tag{3.199}$$

for $n = -\lfloor K/2 \rfloor, \ldots, \lfloor K/2 \rfloor$ where δ_{nl} is the delta function defined to be

$$\delta_{nl} = \begin{cases} 1, & n = l; \\ 0, & n \neq l. \end{cases} \tag{3.200}$$

We observe that (3.199) is independent of the transmitted symbols and this result implies a significant decrease in the overall computational complexity of the MMSE channel estimation, as is argued in Section 3.2.1.

In the subsequent channel estimation iterations, estimates of the data symbols are incorporated into the channel estimation as well as the pilot symbols. The desired estimates are computed as follows

$$\hat{f}_k^{(i)} = \sum_{n=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,n}^{(i)} y_{k-n} \tag{3.201}$$

for $i > 1$. The optimal filter coefficients at the i-th iteration are computed by using the Wiener-Hopf's equations given as follows

$$\sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,l}^{(i)} \left\{ \hat{a}_{k-l}^{(i)} r_f (n - l) (\hat{a}_{k-n}^{(i)})^* + N_0 \delta_{nl} \right\} = r_f (n) (\hat{a}_{k-n}^{(i)})^*, \tag{3.202}$$

for $n = -\lfloor K/2 \rfloor, \ldots, \lfloor K/2 \rfloor$. The details of above formulation for subsequent estimation iterations with some practical considerations is given in Appendix A.2.2.

In the subsequent iterations, it is not possible to design the MMSE filter without any dependency to the transmitted symbols since $\left(\hat{a}_k^{(i)}\right)^* a_k \neq 1$ for soft decision feedback. Indeed, this point is believed to be overlooked in [84] so that the resulting procedure followed there is not optimal in the MMSE sense.

3.8.2.2 Unidirectional LMS Channel Estimation

Following the result of Section 3.2.2, the conventional unidirectional LMS algorithm adopted to the iterative channel estimation scheme under consideration is given for the i-th iteration as follows

$$\hat{f}_{k+1}^{(i)} = \hat{f}_k^{(i)} + 2\mu\, e_k^{(i)}\, \hat{a}_k^{(i)} \tag{3.203}$$

where $i > 1$ so that we employ the algorithm after the initial channel estimation iteration, μ is the step-size value of the adaptation and e_k is the error term given as

$$e_k^{(i)} = y_k - \hat{f}_k^{(i)} \hat{a}_k^{(i)}. \tag{3.204}$$

3.8.2.3 Bidirectional LMS Channel Estimation

As described in Section 3.2.2, the forward and the backward adaptations of the bidirectional LMS algorithm adopted to the iterative channel estimation scheme under consideration are given for the i-th iteration as follows

$$\hat{f}_{k+1}^{f,(i)} = \hat{f}_k^{f,(i)} + 2\mu\, e_k^{f,(i)}\, \hat{a}_k^{(i)} \tag{3.205}$$

$$\hat{f}_{k-1}^{b,(i)} = \hat{f}_k^{b,(i)} + 2\mu\, e_k^{b,(i)}\, \hat{a}_k^{(i)} \tag{3.206}$$

where $i > 1$ as before, μ is the common step-size value and $e_k^{f,(i)}$ and $e_k^{b,(i)}$ are the associated error terms given as

$$e_k^{f,(i)} = y_k - \hat{f}_k^{f,(i)}\, \hat{a}_k^{(i)} \tag{3.207}$$

$$e_k^{b,(i)} = y_k - \hat{f}_k^{b,(i)}\, \hat{a}_k^{(i)}. \tag{3.208}$$

The final fading coefficient estimate $\hat{f}(i)_k$ is again given to be

$$\hat{f}_k^{(i)} = \frac{\hat{f}_k^{f,(i)} + \hat{f}_k^{b,(i)}}{2}. \tag{3.209}$$

3.8.3 Numerical Results

In this section, we evaluate the performance of the channel estimation algorithms considered in Section 3.8.2 with the transmitter and receiver structures given in Section 3.8.1. To this end, we perform Monte Carlo simulations which assumes a Rayleigh fading channel together with the Jakes' spectrum [98], without any loss of generality. The associated temporal autocorrelation is given by (3.69) where the normalized maximum Doppler frequency is chosen to be $f_d T_s = 0.01$ and $f_d T_s = 0.02$ throughout the simulations, if otherwise stated. We perform sufficient number of Monte Carlo runs to have appropriate results in terms of statistical significance. We assume BPSK alphabet throughout the simulations except for some examples using QPSK and 8-PSK symbols with some special settings.

At the transmitter, a set of $L_d = 98$ bits are chosen from the set $\{0, 1\}$ in an independent and identical fashion. A convolutional encoder with generator $(1, 5/7)_8$ and rate $R_c = 1/2$ is then used which employs 2 bits for termination. By this way, a number of $L_c = 200$ coded symbols are produced at the output of the encoder. The coded symbols are passed through a random interleaver and then modulated using a BPSK modulator with the alphabet $\mathcal{A} = \{-1, +1\}$. The resulting set of modulated symbols of length $L_m = 200$ are passed through the PSAM block and the final set of symbols $\{a_k\}_{k=1}^{L}$ are then sent to the time-varying flat-fading channel under consideration.

For the PSAM operation, the pilot symbol spacing is chosen to be $M_p = 11$ and $M_p = 21$ throughout the simulations which results in a number of $L_p = 20$ and $L_p = 10$ pilots, respectively, for each of the transmitted block of length $L = 220$ and $L = 210$, respectively. These settings result in a pilot overhead percentage of 0.0909 and 0.0476 for $M_p = 11$ and $M_p = 21$, respectively. The overall transmission rate of the system associated with these pilot settings, which is defined to be $R = L_d/L$, becomes 0.4667 and 0.4455, respectively.

At the receiver, the unknown channel is estimated iteratively such that an initial estimate is obtained using the MMSE estimator making use of the pilot symbols only. This estimate is then refined over iterations by employing the soft decisions of the coded symbols with any of the estimation algorithms under consideration. Note that hard decision feedback could also be preferred in simulations which is known in the literature to achieve a degraded error performance under some circumstances. Throughout simulations, we set the number of channel estimation iterations to 3 after which no significant improvement is observed. The remaining details of the receiver operations are explained in Section 3.8.1.

Note that when we are re-computing taps of the MMSE filter through the iterations of the channel estimation, we use the available soft decisions on the transmitted symbols instead of their true values. Therefore, this mismatch destroys the optimality of the MMSE filter and may cause an error performance degradation since MMSE filter is known to be not robust to any parameter mismatch (see [116] and references therein).

Before presenting the simulation results, we want to make a final note from a practical point of view. While analyzing the steady-state MSE behavior of the bidirectional LMS algorithm, we naturally assume perfect initialization such that the fading coefficients at the beginning and end of each transmitted block are known without any error. In this section, because we

are dealing with a more practical scenario, we propose an alternative solution in which the forward and the backward adaptations are initialized by using not the perfect values of the fading coefficients which are unknown a priori, but the associated estimates from the previous estimation iterations. In the following, we provide the simulation results corresponding to this imperfect initialization case, as well.
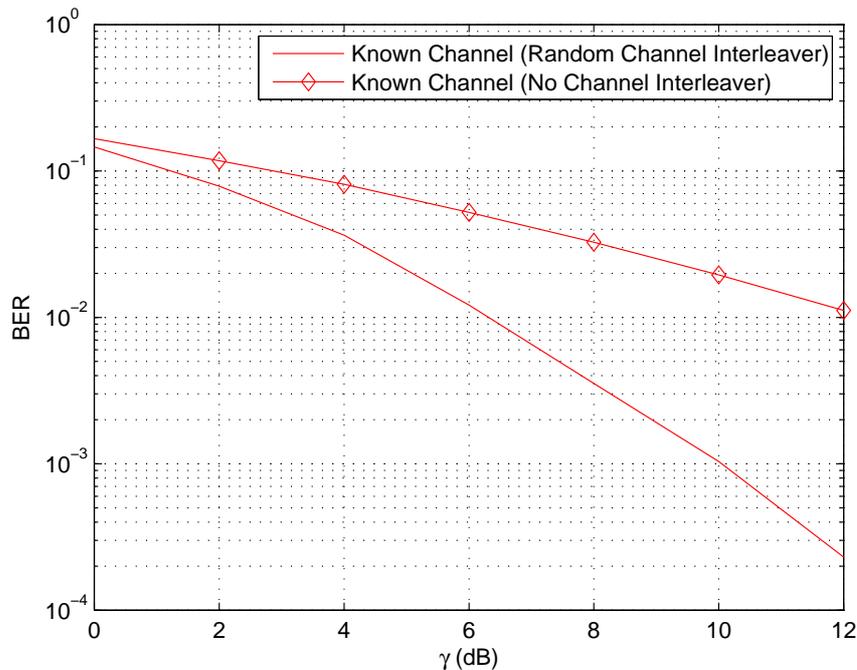
Figure 3.44: BER for random channel interleaver and no channel interleaver cases where the channel is known a priori for both cases and $f_d T_s = 0.01$.

We begin with presenting simulation results for the known channel case with a comparison on the use of a channel interleaver. Fig. 3.44 depicts the BER results for a random channel interleaver and no channel interleaver cases for the normalized maximum Doppler frequency value of $f_d T_s = 0.01$. As is expected, interleaving the set of modulated symbols prior to PSAM block achieves a significant performance improvement. Therefore, we employ channel interleaver for the transmission models in the rest of this thesis unless otherwise stated.

In the subsequent figures, we present performances of the bidirectional LMS and the uni-directional LMS algorithms together with the MMSE filter after 3 estimation iterations for various choices of the pilot symbol spacing (M_p), the number of MMSE filter taps (K) and the normalized maximum Doppler frequency ($f_d T_s$). The results for the MMSE filter using

pilot symbols only and for the bidirectional LMS algorithm with imperfect initialization are also provided together with the known channel bound. In addition, both types of the LMS algorithm employs the optimal step-size values, i.e., μ_{opt}, in a trial and error basis to achieve the best possible performance.
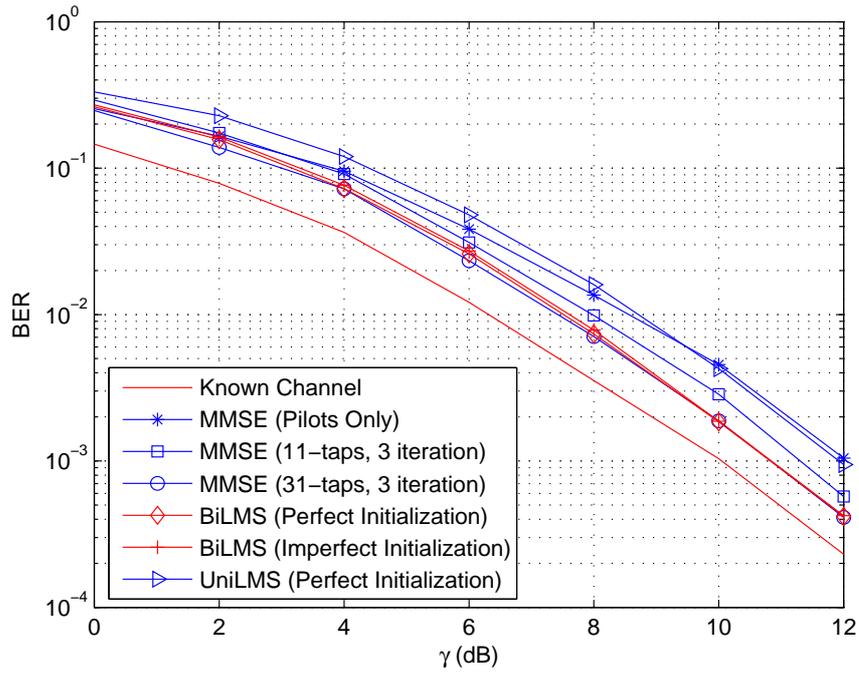
Figure 3.45: BER for BiLMS, UniLMS and MMSE with $M_p = 11$ and $f_d T_s = 0.01$. BER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.

Fig. 3.45 and 3.46 depict the BER and BLER results, respectively, for $M_p = 11$ and $f_d T_s = 0.01$. As explained before, the MMSE filter employed in the initial estimation iteration uses all available pilot symbols, so the associated number of taps is L_p. In the subsequent iterations, performances of a 11-tap and 31-tap MMSE filters are compared and approximately a 0.5 dB SNR improvement is observed at BER= 10^{-3} when the latter is employed instead of the former.
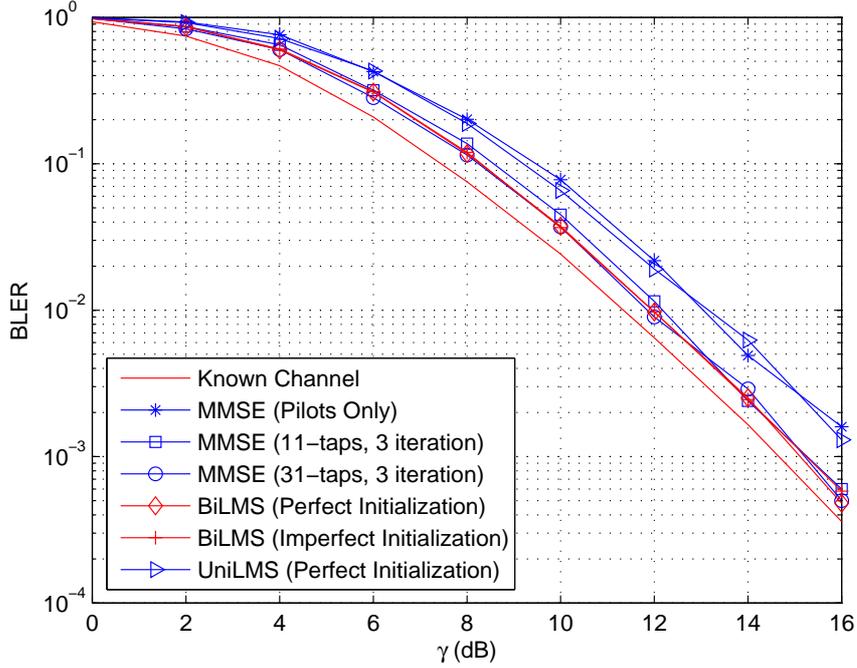
Figure 3.46: BLER for BiLMS, UniLMS and MMSE with $M_p = 11$ and $f_d T_s = 0.01$. BLER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.

We make a number of observations from Fig. 3.45- 3.46 some of which are as follows. The 1-tap bidirectional LMS algorithm has almost the same BER and BLER performance with the 31-tap MMSE filter which verifies its low-complexity and near-optimal tracking performance. We observe that in order for MMSE filter to achieve the same error performance with the bidirectional LMS algorithm, even an 11-tap filter is not sufficient. It is also observed that by employing the soft estimates of the coded symbols with any of the MMSE filter or the bidirectional LMS algorithm, 1 dB SNR improvement is achieved over 3 iterations as compared to pilot only case. The final BER result for the bidirectional LMS algorithm is off the known channel bound by only 1 dB whereas this gap diminishes to 0.5 dB approximately for the BLER statistics. Interestingly, the conventional unidirectional LMS algorithm with soft estimates provide no improvement over the pilot only case, even deteriorates at low and moderate SNR values. As a final remark, the initialization of the bidirectional LMS algorithm with the imperfect estimates result in no performance degradation so that this algorithm could be argued to be robust to initialization imperfections.

The effect of pilot symbol spacing is partially explored for the system under consideration

by increasing its value to $M_p = 21$ which means less number of pilot symbols to be used. Fig. 3.47- 3.48 demonstrate the associated results keeping the other system parameters the same. In this case, the performances of the bidirectional LMS algorithm after 3 iterations is almost the same with that of the 31-tap MMSE filter, as before, with an unchanged gap from the known channel bound. Since the number of pilot symbols decrease, both BER and BLER performances associated with the initial channel estimate deteriorate and are far from the known channel bound by 3 dB where this gap was 2 dB for $M_p = 11$ case. As before, the imperfect initialization of the bidirectional LMS algorithm does not cause any performance degradation and 11-tap MMSE filter could not achieve the error performance of the bidirectional LMS algorithm. As a final note, the unidirectional LMS algorithm provides some improvement with the use of the soft estimates of the coded symbols, unfortunately, which remains somewhat marginal as compared with the other algorithms.
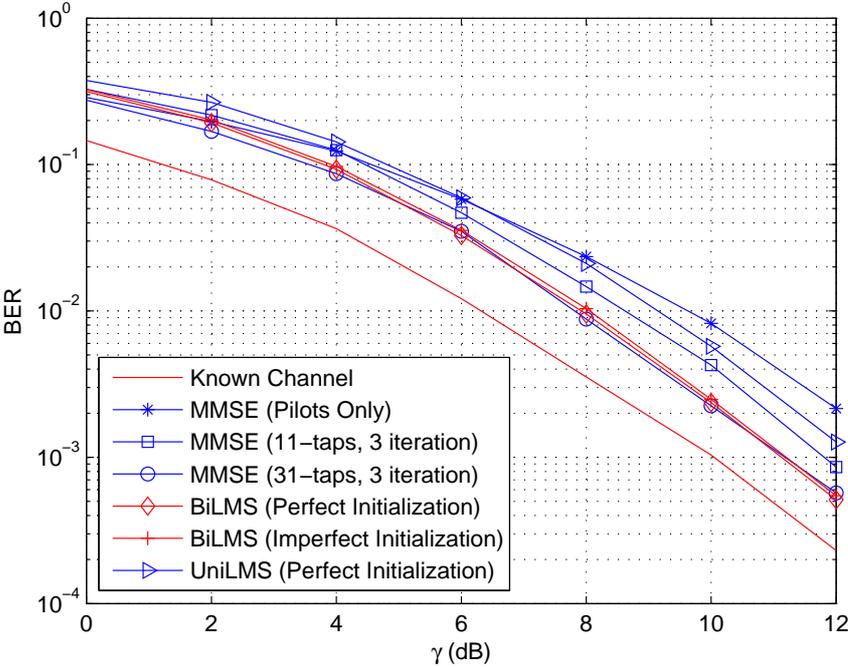
Figure 3.47: BER for BiLMS, UniLMS and MMSE with $M_p = 21$ and $f_d T_s = 0.01$. BER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.
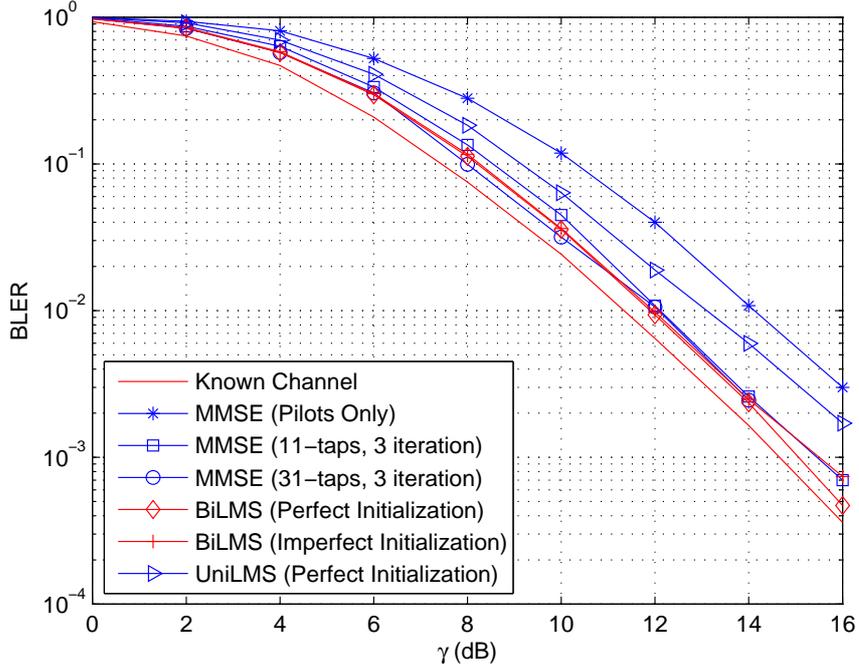
Figure 3.48: BLER for BiLMS, UniLMS and MMSE with $M_p = 21$ and $f_d T_s = 0.01$. BLER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.

We now explore the effect of the Doppler frequency or equivalently the speed of the channel variation by simply choosing a faster channel in which the maximum normalized Doppler frequency is $f_d T_s = 0.02$. Fig. 3.49- 3.50 depicts the associated BER and BLER performances, respectively, for $M_p = 11$, and Fig. 3.51- 3.52 demonstrates the same statistics for $M_p = 21$. The difficulty in estimating and tracking such a fast time-varying channel could be observed through the increase in SNR gap between the known channel bound and the associated error performances of the estimation algorithms. As an example, BER of the bidirectional LMS algorithm after 3 iterations is observed to be off the known channel bound by 2 dB in Fig. 3.51 whereas this gap was only 1 dB for $f_d T_s = 0.01$ as is shown in Fig. 3.47. Nevertheless, the bidirectional LMS algorithm achieves almost the same error performance with the MMSE filter with a sufficient number of taps even under this challenging environment. The robustness of the bidirectional LMS algorithm to initialization imperfections are again verified for these particular choices. As before, the unidirectional LMS algorithm could not provide any significant error performance improvement through iterations.

94
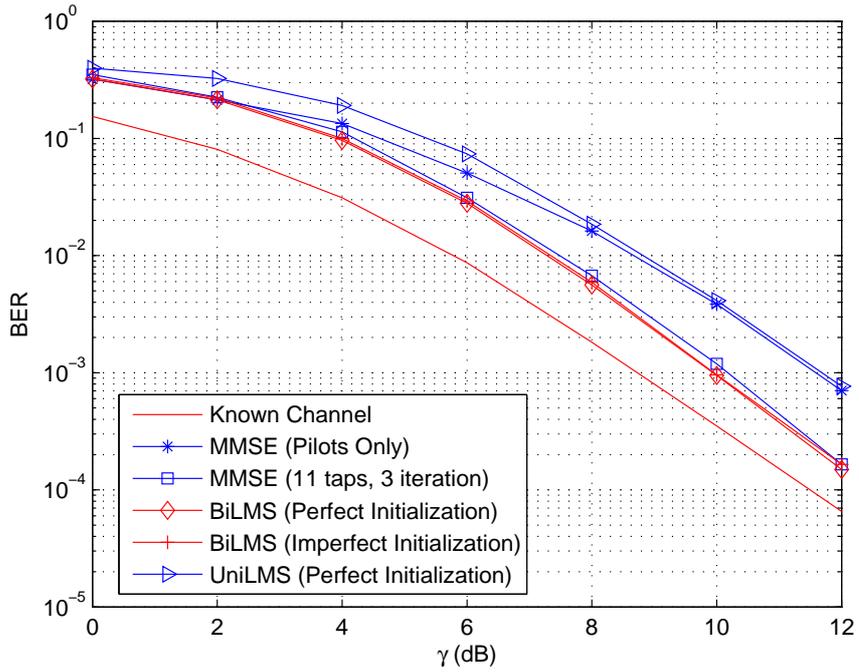
Figure 3.49: BER for BiLMS, UniLMS and MMSE with $M_p = 11$ and $f_d T_s = 0.02$. BER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.
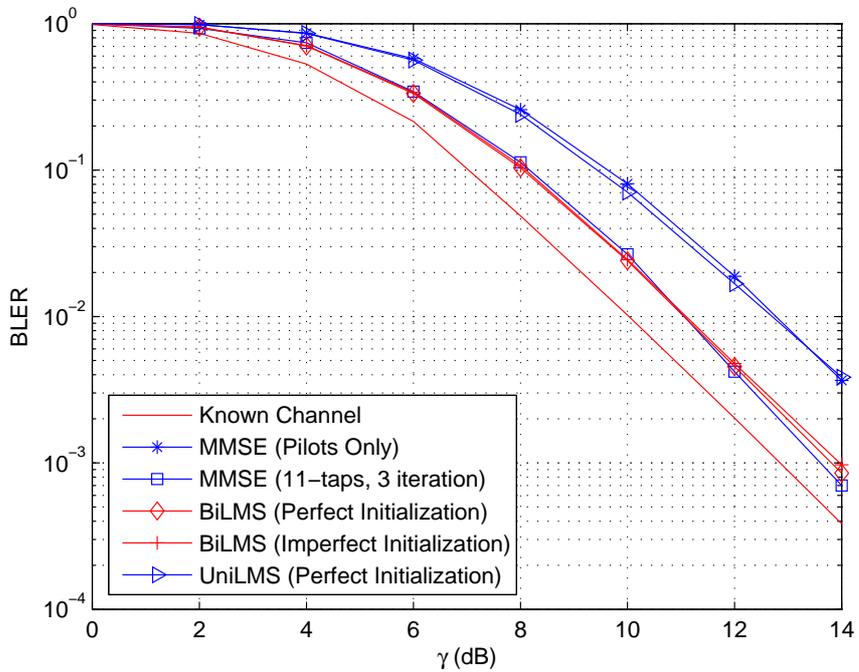
Figure 3.50: BLER for BiLMS, UniLMS and MMSE with $M_p = 11$ and $f_d T_s = 0.02$. BLER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.
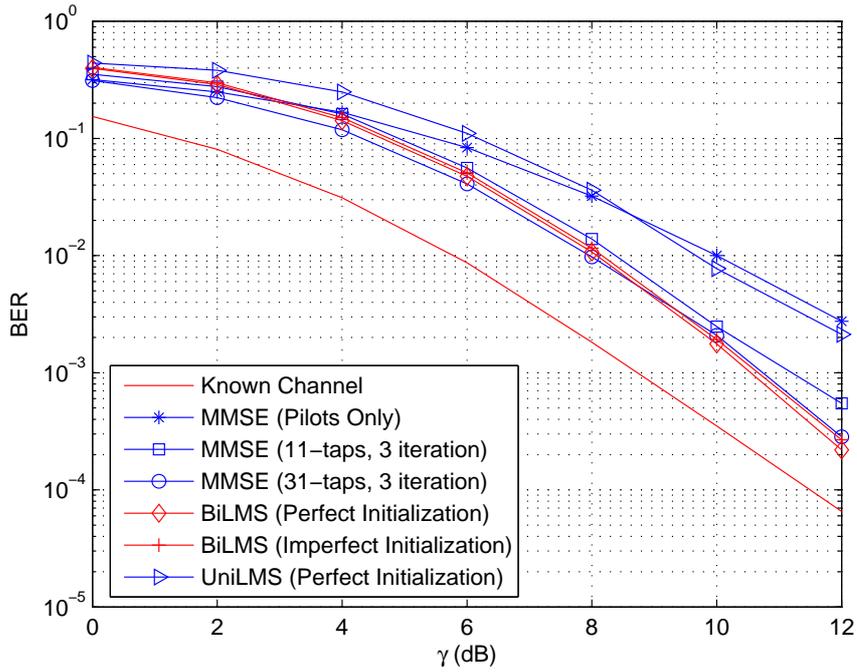
Figure 3.51: BER for BiLMS, UniLMS and MMSE with $M_p = 21$ and $f_d T_s = 0.02$. BER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.
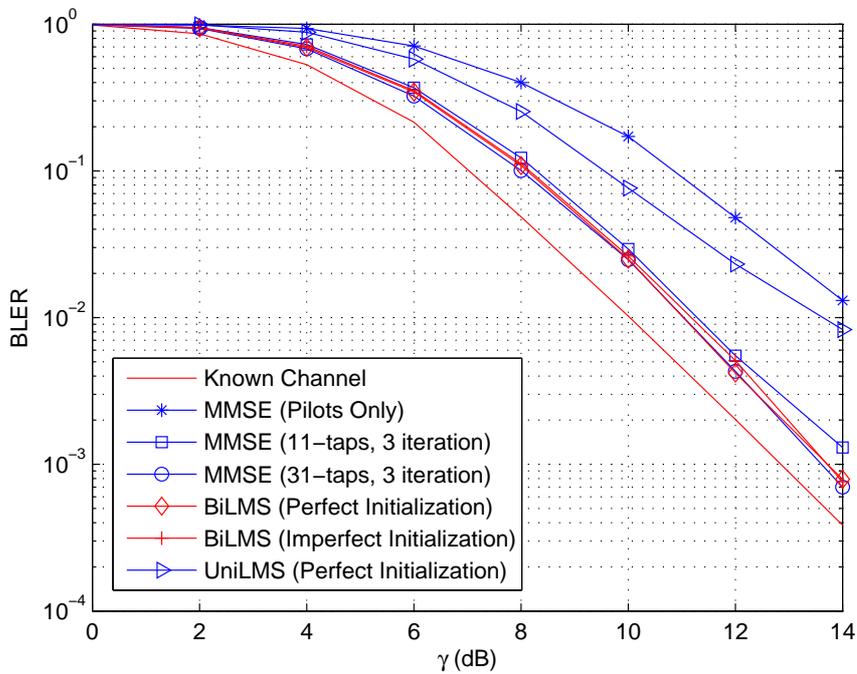
Figure 3.52: BLER for BiLMS, UniLMS and MMSE with $M_p = 21$ and $f_d T_s = 0.02$. BLER for MMSE using pilots only and for imperfectly initialized BiLMS are also provided.

In Fig. 3.53-3.54, we also present BER results for QPSK and 8-PSK modulations with the constellations given in Fig. 3.43. The iterative estimator works for 3 times which again exploits a satisfactory convergence. We observe from both cases that the unidirectional LMS algorithm deteriorates the quality of the channel estimate obtained by using the pilots only and therefore achieves an error performance which is the worst of all unlike the case for BPSK modulation with the same channel speed. We also observe that both the bidirectional LMS algorithm and the MMSE filter with the soft estimates of the coded symbols provide almost a 1 dB SNR gain over the pilots only case and are off the known channel bound by again 1 dB.

Finally, we present the BER results for BPSK over an extremely fast channel with $f_d T_s = 0.1$ in Fig. 3.55. Under such a scenario, the pilot symbol spacing value should be as small as $M_p = 5$ in order to achieve reasonable performance for which the performance of the MMSE filter with 11-tap is better than that for the bidirectional LMS algorithm.
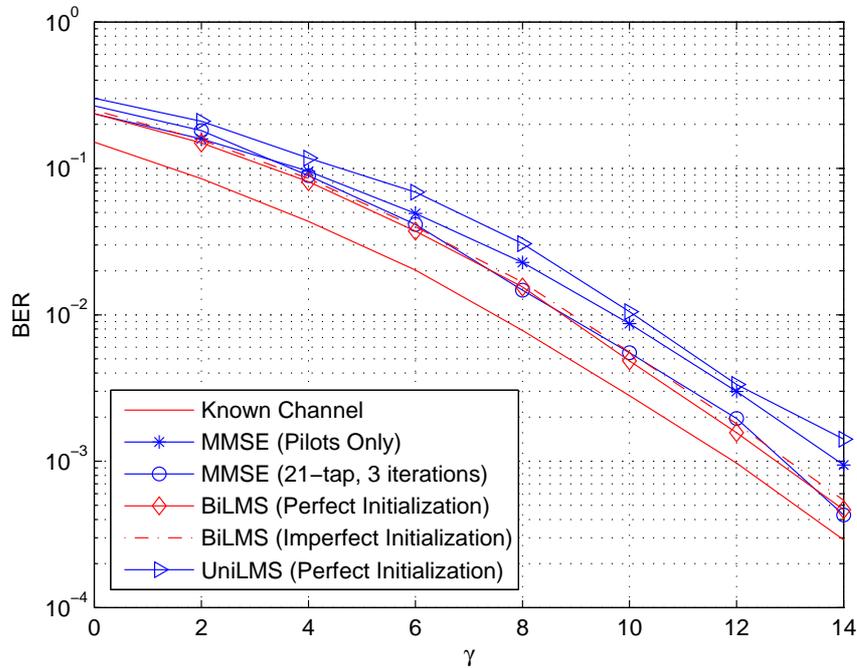
Figure 3.53: BER for BiLMS, UniLMS and MMSE with QPSK modulation for $M_p = 11$, $L_d = 98$ and $f_d T_s = 0.01$.
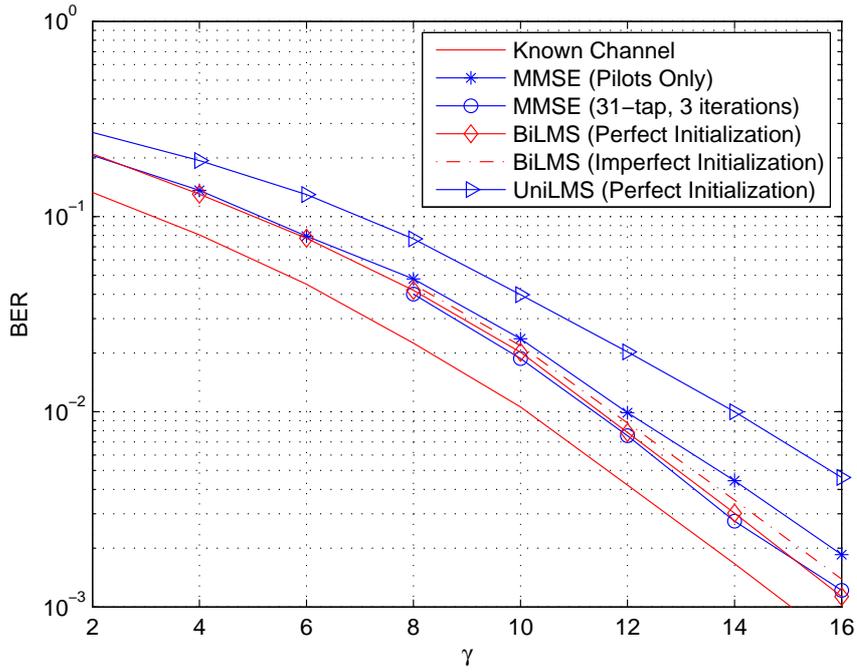
Figure 3.54: BER for BiLMS, UniLMS and MMSE with 8-PSK modulation for $M_p = 11$, $L_d = 97$ and $f_d T_s = 0.01$.
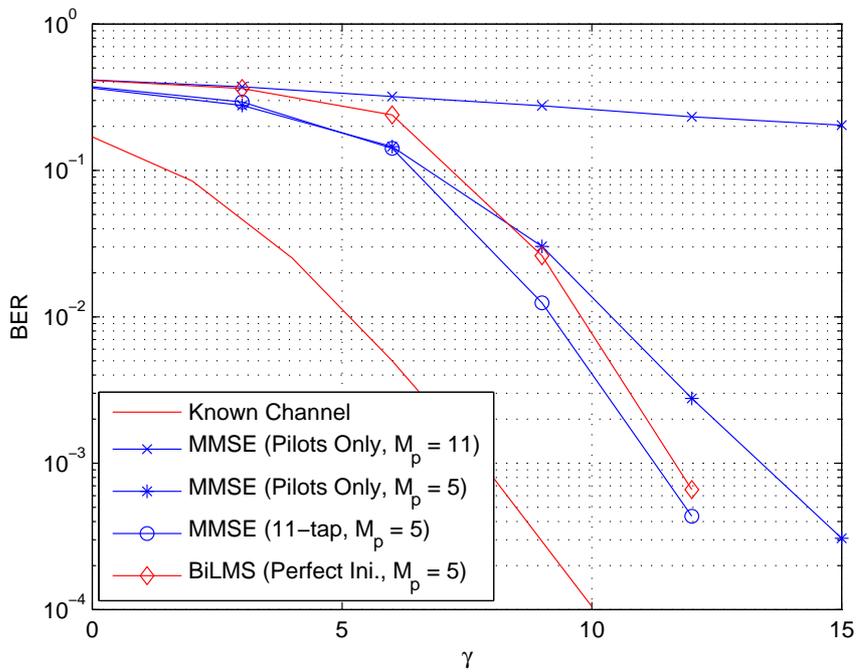
Figure 3.55: BER for BiLMS and 11-tap MMSE with BPSK modulation for $L_d = 98$ and $f_d T_s = 0.1$.

3.9 Conclusion

This section considers the bidirectional LMS algorithm over time-varying flat-fading channels with a result of near-optimal tracking performance and computationally efficient structure. Having introduced the channel model and the bidirectional LMS algorithm, the associated tracking behavior is analyzed for Rayleigh fading channel with nonrational power spectrum. As a result, a step-size dependent steady-state MSE expression is obtained together with a framework for optimal step-size selection. The numerical results for the analytical MSE expression and the experimental ones show a very good match. These results verify the near-optimal tracking performance of the bidirectional LMS algorithm which is significantly better than that of the unidirectional LMS algorithm and is very similar to that of the optimal Wiener filter. The effect of independent selection of the forward and the backward step-size values are also investigated and the best strategy is argued to be the equal selection. This analysis is novel in the sense that although there are several estimators as an application of the bidirectional processing in communication systems, none of them present such a theoretical analysis that reveals the basis of the appealing performance.

Effects of imperfect knowledge for the Doppler spread and SNR together with the imperfect initial value are also investigated. The tracking and optimal step-size selection analysis is then revisited for a Rayleigh fading channel with a temporal correlation characterized by a 1-st order AR process. As a result, more compact form of the steady-state MSE expression is obtained with a high accuracy which eliminates the necessity to the frequency-domain energy computation using numerical integration methods. Finally, iterative estimation of unknown time-varying flat-fading channels is considered as a much realistic application employing the estimation algorithms under consideration and the superiority of the bidirectional LMS algorithm is again observed.

CHAPTER 4

THE BIDIRECTIONAL LMS ALGORITHM OVER FREQUENCY-SELECTIVE FADING CHANNELS

In Chapter 3, we have considered the estimation of time-varying flat-fading channels using the bidirectional LMS algorithm where both the receiver and the transmitter are employed with a single antenna. Although the flat-fading assumption is useful in understanding and analyzing the performance of the bidirectional LMS algorithm, real-life communication systems frequently experience frequency-selective fading. Along with the increasing demand for wireless services of high data rate, frequency selectivity appears naturally as a result of the fact that the bandwidth of the next generation waveforms become much higher than the coherence bandwidth of communication channels. We therefore address the problem of estimation and tracking of frequency-selective fading channels in particular with the bidirectional LMS algorithm.

In frequency-selective fading channels, different replicas of the transmitted signal arrive at the receiver with various delays which are comparable to the symbol period. As a result, the symbol to be detected at a time instant is interfered with previously transmitted symbols which is known as the intersymbol interference (ISI) problem. These channels are sometimes called multipath fading channels since the various replicas travel through different paths [102]. Such communication channels are usually modeled in a vector form as opposed to the scalar random variable for flat-fading channels. As a result, any algorithm considered in this chapter should estimate or, for adaptive algorithms, update the random vectors representing time-varying multipath fading channels.

The analysis of the bidirectional LMS algorithm in frequency-selective channels is somewhat different from the one given for flat-fading channels. The reason behind this fact is that

the memory present in frequency-selective channels makes the well-known independence assumption inapplicable to the problem in hand [65]. In the following, a modified steady-state MSE expression for frequency-selective channels is given together with an optimal step-size analysis as before. In order to investigate real-life scenarios, a coded frequency-selective time-varying Rayleigh fading channel is considered together with the iterative channel estimation approach using the bidirectional LMS algorithm.

4.1 System Model for Frequency-Selective Fading Channels

We consider a frequency-selective, i.e., multipath, fading communication channel which is varying with time according to some temporal autocorrelation, and is represented by a set of complex fading coefficient vectors $\{\mathbf{f}_k\}_{k=1}^L$ with $\mathbf{f}_k = [f_{k,0}\, f_{k,1}\, \ldots\, f_{k,M-1}]^T$, where L is the observation length and M is the number of taps of the channel. We assume the uniform power delay profile for the average powers of the channel taps unless otherwise stated. We also present some results in Section 4.6 associated with a different power delay profile to justify the generality of this consideration. We therefore assume that $\{f_{k,l}\}_{l=0}^{M-1}$ is an uncorrelated set with each element $f_{k,l}$ has unit energy. In addition, the set $\{f_{k,l}\}_{k=1}^L$ is correlated according to the autocorrelation of the fading model under consideration. The fading coefficients are also assumed to be known neither at the transmitter nor the receiver.

A set of independent and identically distributed information symbols $\{a_k\}_{k=1}^L$ are chosen from a finite alphabet $\mathcal{A}$ with symbol energy $E_s = E\{|a_k|^2\}$ and are transmitted through the multipath fading channel under consideration. The corresponding discrete-time equivalent complex baseband channel model is then given as

$$y_k \quad = \quad \sum_{l=0}^{M-1} f_{k,l}\, a_{k-l}\, +\, n_k \tag{4.1}$$

$$= \quad \mathbf{f}_k^T\, \mathbf{a}_k\, +\, n_k \tag{4.2}$$

where y_k is the observation symbol, $\mathbf{a}_k = [a_k\, a_{k-1}\, \ldots\, a_{k-M+1}]^T$ is the input vector and n_k is a sample from a circularly symmetric white complex Gaussian process with zero-mean and variance N_0. We also assume perfect timing information for the transmitted blocks and individual symbols, and no frequency offset, as before.

101

The received SNR is given as

$$
\gamma_r = \frac{E\left\{\left|\sum_{l=0}^{M-1} f_{k,l}\, a_{k-l}\right|^2\right\}}{E\{|v_k|^2\}} = \frac{\sum_{l=0}^{M-1}\sum_{l'=0}^{M-1} E\left\{f_{k,l} f_{k,l'}^*\right\} \overbrace{E\left\{a_{k-l} a_{k-l'}^*\right\}}^{E_s\,\delta_{ll'}}}{N_0} \tag{4.3}
$$

$$
= \frac{E_s \sum_{l=0}^{M-1} E\left\{|f_{k,l}|^2\right\}}{N_0} = \frac{M\,E_s}{N_0}, \tag{4.4}
$$

where (4.4) makes use of the fact that the input sequence $\{a_k\}_{k=1}^{L}$ is assumed to be uncorrelated. The information symbol SNR is accordingly given as $\gamma = \gamma_r/R$ where R is the overall transmission rate of the communication system.

4.2 Estimation Algorithms for Frequency-Selective Fading Channels

In this section, we will revisit the channel estimation algorithms considered in the previous chapter and make some necessary modifications to be employed over the frequency-selective channel model given in Section 4.1. As an important difference, this section deals with the estimation of channel vectors as opposed to the scalar coefficient estimation considered in Section 3.2.

4.2.1 The MMSE Channel Estimation

Let us consider the MMSE estimation for a frequency-selective channel given as

$$
\hat{\mathbf{f}} = \mathbf{W}\,\mathbf{y} \tag{4.5}
$$

where $\hat{\mathbf{f}} = \left[\hat{\mathbf{f}}_1^T \ \ldots \ \hat{\mathbf{f}}_L^T\right]^T$ is the vector of estimates of fading coefficients, $\mathbf{W}$ is the estimation filter to be optimized and $\mathbf{y} = [y_1 \ \ldots \ y_L]^T$ is the observation vector including the desired observations to be employed.

Note that (4.5) could be modified such that the overall estimator is derived for a single fading vector $\hat{\mathbf{f}}_k$ instead of the complete set of unknown fading vectors represented by $\hat{\mathbf{f}}$. In addition, less number of observations could be employed in (4.5) instead of all the available ones. These two simplifications will result in a size reduction for the estimation matrix $\mathbf{W}$.

The optimal MMSE filter $\mathbf{W}$ could be found by using the Wiener-Hopf's solution given as

$$\mathbf{W} = \mathbf{P}_{fy}\,\mathbf{R}_{yy}^{-1} \tag{4.6}$$

where $\mathbf{P}_{fy} = E\left\{\mathbf{f}\,\mathbf{y}^H\right\}$ is the cross-correlation matrix and $\mathbf{R}_{yy} = E\left\{\mathbf{y}\,\mathbf{y}^H\right\}$ is the autocorrelation matrix of the observations in use. The details of the derivation of (4.6) is given Appendix B.1.

In order to present a fair complexity analysis, we assume that K observation symbols are used in order to estimate any of the $M \times 1$ fading vectors for each time epoch separately. With this assumption, the pre-filtering stage given in (4.6) requires a matrix inversion of complexity $O(K^3)$ and a matrix multiplication of complexity $O(M K^2)$ in order to find the optimal MMSE filter. In the filtering stage, desired estimate is computed according to (4.5) which requires $M(K-1)$ complex additions and $M K$ complex multiplications.

4.2.2   The Unidirectional and The Bidirectional LMS Algorithms

The conventional unidirectional LMS algorithm is given for time-varying frequency-selective channels as follows [57]

$$\hat{\mathbf{f}}_{k+1} = \hat{\mathbf{f}}_k + 2\mu\, e_k\, \mathbf{a}_k \tag{4.7}$$

where μ is the step-size of the adaptation and e_k is the estimation error defined as

$$e_k = y_k - \hat{\mathbf{f}}_k^T\, \mathbf{a}_k . \tag{4.8}$$

The unidirectional LMS algorithm is observed to need $M+1$ complex additions and $2(M+1)$ complex multiplications in order to estimate a single fading vector without any extra computational complexity.

In order to present the bidirectional LMS algorithm, let us define $\hat{f}_{k,l}^f$ and $\hat{f}_{k,l}^b$ to be the estimates of the l-th tap of the channel in the forward and the backward directions, respectively. The algorithm is then given as

$$\hat{\mathbf{f}}_{k+1}^f \;=\; \hat{\mathbf{f}}_k^f + 2\mu\, e_k^f\, \mathbf{a}_k \tag{4.9}$$

$$\hat{\mathbf{f}}_{k-1}^b \;=\; \hat{\mathbf{f}}_k^b + 2\mu\, e_k^b\, \mathbf{a}_k \tag{4.10}$$

where $\hat{\mathbf{f}}_k^f = \left[\hat{f}_{k,0}^f\ \hat{f}_{k,1}^f\ \ldots\ \hat{f}_{k,M-1}^f\right]^T$ and $\hat{\mathbf{f}}_k^b = \left[\hat{f}_{k,0}^b\ \hat{f}_{k,1}^b\ \ldots\ \hat{f}_{k,M-1}^b\right]^T$ are the estimates of the channel vector $\mathbf{f}_k$ in the forward and the backward directions, respectively, μ is the associated

step-size of the adaptations, e_k^f and e_k^b are the forward and the backward errors defined as

$$e_k^f \;=\; y_k - (\hat{\mathbf{f}}_k^f)^T \, \mathbf{a}_k \qquad (4.11)$$

$$e_k^b \;=\; y_k - (\hat{\mathbf{f}}_k^b)^T \, \mathbf{a}_k. \qquad (4.12)$$

The final estimate $\hat{\mathbf{f}}_k$ of the fading vector are again chosen to be the arithmetic average of the forward and the backward estimates given as

$$\hat{\mathbf{f}}_k = \frac{\hat{\mathbf{f}}_k^f + \hat{\mathbf{f}}_k^b}{2}. \qquad (4.13)$$

The overall complexity of the bidirectional LMS algorithm is now contributed by $(3M + 2)$ complex additions and $(5M + 4)$ the complex multiplications which is much smaller than the optimal Wiener filter. As a final note, the proper estimates of the transmitted symbols should be replaced with $\mathbf{a}_k$ in above equations if we deal with a real-life communication system rather than a system identification type problem, as before.

4.3 Tracking Performance of the Bidirectional LMS over Frequency-Selective Channels

In this section, we evaluate the tracking performance of the bidirectional LMS algorithm over a frequency-selective channel by making use of the results presented in Section 3.3 and [61, 63]. We concentrate only on the tracking quality of the bidirectional LMS algorithm rather than the detection performance for the transmitted symbols which will be the subject of the next section. We therefore assume that the overall system is operating in the training mode such that the transmitted symbols $\{a_k\}_{k=1}^{L}$ are known a priori at the receiver. In addition, since we are interested in the tracking performance of the algorithm, the system is assumed to be at the steady-state condition. We remind that the transition behavior of the algorithm is related to the overall convergence rate which is beyond the scope of this work.

In order to analyze the tracking performance of the bidirectional LMS algorithm over a time-varying frequency-selective channel, we will derive a theoretical step-size dependent steady-state MSE expression which benefits from the previous analysis performed in Section 3.3 for a scalar case.

The error performance surface, or equivalently the MSE, for this problem is given as

$$
\begin{aligned}
J_{MSE,k} &= E\left\{|e_k|^2\right\} & \text{(4.14)} \\
&= E\left\{\left|y_k - \hat{\mathbf{f}}_k^T \mathbf{a}_k\right|^2\right\} & \text{(4.15)} \\
&= E\left\{|n_k|^2\right\} + E\left\{\left|(\mathbf{f}_k - \hat{\mathbf{f}}_k)^T \mathbf{a}_k\right|^2\right\} & \text{(4.16)} \\
&= J_{min} + (\mathbf{f}_k - \hat{\mathbf{f}}_k)^T \underbrace{E\left\{\mathbf{a}_k \mathbf{a}_k^H\right\}}_{=E_s \mathbf{I}} (\mathbf{f}_k - \hat{\mathbf{f}}_k) & \text{(4.17)} \\
&= J_{min} + \underbrace{E_s \left\|\mathbf{f}_k - \hat{\mathbf{f}}_k\right\|^2}_{J_{ex,k}} & \text{(4.18)}
\end{aligned}
$$

where e_k is the overall tracking error, J_{min} is the minimum achievable MSE due to the presence of additive noise and is equal to N_0, and $J_{ex,k}$ is defined as the excess MSE due to the noisy gradient estimation and the time variation [61]. At the steady state, we may express the average MSE as follows

$$
\begin{aligned}
J_{MSE} &= E\left\{J_{MSE,k}\right\} & \text{(4.19)} \\
&= J_{min} + E_s \underbrace{E\left\{\left\|\mathbf{f}_k - \hat{\mathbf{f}}_k\right\|^2\right\}}_{J_{MSIE}} & \text{(4.20)}
\end{aligned}
$$

where J_{MSIE} is defined as the mean square identification error, and is related to the average excess MSE, i.e., $J_{ex} = E\{J_{ex,k}\}$, as $J_{MSIE} = J_{ex}/E_s$ [63]. From this overview, one could observe that J_{MSIE} is the basic expression to be derived in order to characterize J_{MSE}.

In the above expressions, all the expectations are assumed to be over the ensemble of [61] in which a set of independent input symbols are transmitted over the same time-varying frequency-selective channel, and the corresponding distinct observations are provided to the bidirectional LMS algorithm to obtain a set of estimates $\{\hat{\mathbf{f}}_k\}_{k=1}^L$. Under this scenario, J_{MSIE} could be decomposed as follows

$$
\begin{aligned}
J_{MSIE} &= E\left\{\left\|\hat{\mathbf{f}}_k - \mathbf{f}_k\right\|^2\right\} = E\left\{\left\|\left(\hat{\mathbf{f}}_k - E\{\hat{\mathbf{f}}_k\}\right) + \left(E\{\hat{\mathbf{f}}_k\} - \mathbf{f}_k\right)\right\|^2\right\} & \text{(4.21)} \\
&= E\left\{\left\|\hat{\mathbf{f}}_k - E\{\hat{\mathbf{f}}_k\}\right\|^2\right\} + E\left\{\left\|E\{\hat{\mathbf{f}}_k\} - \mathbf{f}_k\right\|^2\right\} \\
&\quad + 2\,\mathrm{Re}\left\{E\left\{\left(\hat{\mathbf{f}}_k - E\{\hat{\mathbf{f}}_k\}\right)\left(E\{\hat{\mathbf{f}}_k\} - \mathbf{f}_k\right)^*\right\}\right\}. & \text{(4.22)}
\end{aligned}
$$

Because the time-varying frequency-selective fading channel is assumed to be static across the ensemble, we have $E\{\mathbf{f}_k\} = \mathbf{f}_k$ and so that the last term in (4.22) could be simplified as

105

in [61] as follows

$$E\left\{\left(\hat{\mathbf{f}}_k - E\{\hat{\mathbf{f}}_k\}\right)\left(E\{\hat{\mathbf{f}}_k\} - \mathbf{f}_k\right)^*\right\} = E\left\{\hat{\mathbf{f}}_k E\{\hat{\mathbf{f}}_k\}^*\right\} - E\left\{\hat{\mathbf{f}}_k \mathbf{f}_k^*\right\} - E\left\{E\{\hat{\mathbf{f}}_k\} E\{\hat{\mathbf{f}}_k\}^*\right\}$$

$$+ E\left\{E\{\hat{\mathbf{f}}_k\} \mathbf{f}_k^*\right\} \tag{4.23}$$

$$= \left\|E\{\hat{\mathbf{f}}_k\}\right\|^2 - E\{\hat{\mathbf{f}}_k\} \mathbf{f}_k^* - \left\|E\{\hat{\mathbf{f}}_k\}\right\|^2 + E\{\hat{\mathbf{f}}_k\} \mathbf{f}_k^* \tag{4.24}$$

$$= 0 \tag{4.25}$$

Therefore, we could simply ignore the last term in (4.22), and J_{MSIE} of the bidirectional LMS in time-varying frequency-selective channels is then expressed as a sum of two terms which are called the self-noise (J_{self}) and the lag (J_{lag}) components [61], and is given as

$$J_{MSIE} = E\left\{\left\|\hat{\mathbf{f}}_k - \mathbf{f}_k\right\|^2\right\} \tag{4.26}$$

$$= \underbrace{E\left\{\left\|\hat{\mathbf{f}}_k - E\{\hat{\mathbf{f}}_k\}\right\|^2\right\}}_{J_{self}} + \underbrace{E\left\{\left\|E\{\hat{\mathbf{f}}_k\} - \mathbf{f}_k\right\|^2\right\}}_{J_{lag}} \tag{4.27}$$

$$= J_{self} + J_{lag}. \tag{4.28}$$

As before, any deviation of the fading coefficient estimate, $\hat{\mathbf{f}}_k$, from the ensemble mean, $E\{\hat{\mathbf{f}}_k\}$, contributes to the self-noise part while differences between the ensemble mean and the unknown coefficient $\mathbf{f}_k$ result in the lag part [61]. Equivalently, from a different perspective, J_{self} arises from the noisy gradient estimation of the error performance surface whereas J_{lag} is just due to time variation. In this section, we will separately derive the steady-state expressions for J_{self} and J_{lag} parts in order to determine a final expression for the steady-state MSE.

4.3.1 Derivation of the Self-Noise Component (J_{self})

In this section, we will derive the self-noise component which arises from the error in estimation of the gradient of the error performance surface, or equivalently the MSE, given in (4.14) associated with the system identification problem under consideration. While deriving J_{self} expression, any time variation is ignored to be considered later in the lag component and the focus will be only on the effect of the gradient estimation error, as is done in [57, 61].

The true gradients of the error performance surface in the forward and the backward directions

are given as

$$\boldsymbol{\nabla}_k^f \quad = \quad 2E_s(\hat{\mathbf{f}}_k^f - \mathbf{f}_k) \tag{4.29}$$

$$\boldsymbol{\nabla}_k^b \quad = \quad 2E_s(\hat{\mathbf{f}}_k^b - \mathbf{f}_k), \tag{4.30}$$

which make use of the results of Section 3.3.2. Since we concentrate on the error in gradient estimation, we model the noisy gradient estimates in the forward and the backward directions, respectively, as follows

$$\hat{\boldsymbol{\nabla}}_k^f = \boldsymbol{\nabla}_k^f + \boldsymbol{\epsilon}_k^f \quad = \quad 2E_s(\hat{\mathbf{f}}_k^f - \mathbf{f}_k) + \boldsymbol{\epsilon}_k^f \tag{4.31}$$

$$\hat{\boldsymbol{\nabla}}_k^b = \boldsymbol{\nabla}_k^b + \boldsymbol{\epsilon}_k^b \quad = \quad 2E_s(\hat{\mathbf{f}}_k^b - \mathbf{f}_k) + \boldsymbol{\epsilon}_k^b, \tag{4.32}$$

where $\boldsymbol{\epsilon}_k^f$ and $\boldsymbol{\epsilon}_k^b$ are the associated error vectors of the gradient estimation in the forward and the backward directions, respectively [57]. We assume that $\boldsymbol{\epsilon}_k^f$ and $\boldsymbol{\epsilon}_k^b$ are composed of uncorrelated elements which are modeled to be zero-mean complex Gaussian random variables. In order to incorporate the effect of gradient estimation error into the forward and the backward adaptations, we express the conventional LMS adaptations as follows

$$\hat{\mathbf{f}}_{k+1}^f \quad = \quad \hat{\mathbf{f}}_k^f - \mu\,\hat{\boldsymbol{\nabla}}_k^f \tag{4.33}$$

$$\hat{\mathbf{f}}_{k-1}^b \quad = \quad \hat{\mathbf{f}}_k^b - \mu\,\hat{\boldsymbol{\nabla}}_k^b \tag{4.34}$$

where $\hat{\boldsymbol{\nabla}}_k^f = -2\,e_k^f\mathbf{a}_k$ and $\hat{\boldsymbol{\nabla}}_k^b = -2\,e_k^b\mathbf{a}_k$ for the conventional LMS algorithm with the associated errors e_k^f and e_k^f given in (4.11)-(4.12) [102]. Since we are interested in the effect of noisy gradient estimation only, we prefer to express the adaptations given in (4.33)-(4.34) using the gradient estimates in (4.31)-(4.32) as follows

$$\hat{\mathbf{f}}_{k+1}^f \quad = \quad \hat{\mathbf{f}}_k^f - 2\mu E_s\,(\hat{\mathbf{f}}_k^f - \mathbf{f}_k) - \mu\,\boldsymbol{\epsilon}_k^f \tag{4.35}$$

$$\hat{\mathbf{f}}_{k-1}^b \quad = \quad \hat{\mathbf{f}}_k^b - 2\mu E_s\,(\hat{\mathbf{f}}_k^b - \mathbf{f}_k) - \mu\,\boldsymbol{\epsilon}_k^b. \tag{4.36}$$

Before going into further detail, we define the forward and the backward tap-weight tracking errors as $\mathbf{v}_k^f = \hat{\mathbf{f}}_k^f - \mathbf{f}_k$ and $\mathbf{v}_k^b = \hat{\mathbf{f}}_k^b - \mathbf{f}_k$, respectively. In order to express the adaptations given in (4.35)-(4.36) in terms of tap-weight tracking errors, we first subtract $\mathbf{f}_{k+1}$ from both sides and then replace $\mathbf{f}_{k+1}$ with $\mathbf{f}_k$ at the right side which is a result of the time invariance assumption as follows

$$\mathbf{v}_{k+1}^f \quad = \quad (1 - 2\,\mu E_s)\mathbf{v}_k^f - \mu\,\boldsymbol{\epsilon}_k^f \tag{4.37}$$

$$\mathbf{v}_{k-1}^b \quad = \quad (1 - 2\,\mu E_s)\mathbf{v}_k^b - \mu\,\boldsymbol{\epsilon}_k^b. \tag{4.38}$$

The next step is to further elaborate the self-noise expression given in (4.27). In order to get rid of the inner expectation $E\{\hat{\mathbf{f}}_k\}$ as a part of the self-noise expression, we take expectations of (4.35)-(4.36) as follows

$$\underbrace{E\{\hat{\mathbf{f}}_{k+1}^f\}}_{=E\{\hat{\mathbf{f}}_k^f\}} = E\{\hat{\mathbf{f}}_k^f\} - 2\mu E_s\left(E\{\hat{\mathbf{f}}_k^f\} - \underbrace{E\{\mathbf{f}_k\}}_{=\mathbf{f}_k}\right) - \mu\,\underbrace{E\{\boldsymbol{\epsilon}_k^f\}}_{=0} \tag{4.39}$$

$$\underbrace{E\{\hat{\mathbf{f}}_{k-1}^b\}}_{=E\{\hat{\mathbf{f}}_k^b\}} = E\{\hat{\mathbf{f}}_k^b\} - 2\mu E_s\left(E\{\hat{\mathbf{f}}_k^b\} - \underbrace{E\{\mathbf{f}_k\}}_{=\mathbf{f}_k}\right) - \mu\,\underbrace{E\{\boldsymbol{\epsilon}_k^b\}}_{=0}, \tag{4.40}$$

where $E\{\hat{\mathbf{f}}_{k+1}^f\} = E\{\hat{\mathbf{f}}_k^f\}$ and $E\{\hat{\mathbf{f}}_{k-1}^b\} = E\{\hat{\mathbf{f}}_k^b\}$ follows from the time invariance assumption, $E\{\mathbf{f}_k\} = \mathbf{f}_k$ is a consequence of the fact that $\mathbf{f}_k$ is common across the ensemble, and $E\{\boldsymbol{\epsilon}_k^f\} = E\{\boldsymbol{\epsilon}_k^b\} = 0$ by definition [57]. After some straightforward steps in (4.39)-(4.40), we have $E\{\hat{\mathbf{f}}_k^f\} = \mathbf{f}_k$ and $E\{\hat{\mathbf{f}}_k^b\} = \mathbf{f}_k$, and $E\{\hat{\mathbf{f}}_k\}$ is therefore found to be

$$E\{\hat{\mathbf{f}}_k\} = \frac{E\{\hat{\mathbf{f}}_k^f\} + E\{\hat{\mathbf{f}}_k^b\}}{2} = \frac{\mathbf{f}_k + \mathbf{f}_k}{2} = \mathbf{f}_k \tag{4.41}$$

with the help of (4.13). As a result of these findings, the self-noise expression given in (4.27) becomes

$$J_{self} = E\left\{\|\hat{\mathbf{f}}_k - E\{\hat{\mathbf{f}}_k\}\|^2\right\} = E\left\{\|\hat{\mathbf{f}}_k - \mathbf{f}_k\|^2\right\} = E\left\{\|\mathbf{v}_k\|^2\right\} \tag{4.42}$$

where $\mathbf{v}_k = \hat{\mathbf{f}}_k - \mathbf{f}_k$ is the overall tap-weight tracking error which is given as

$$\mathbf{v}_k = \hat{\mathbf{f}}_k - \mathbf{f}_k \tag{4.43}$$

$$= \frac{\hat{\mathbf{f}}_k^f + \hat{\mathbf{f}}_k^b}{2} - \mathbf{f}_k = \frac{(\hat{\mathbf{f}}_k^f - \mathbf{f}_k) + (\hat{\mathbf{f}}_k^b - \mathbf{f}_k)}{2} = \frac{\mathbf{v}_k^f + \mathbf{v}_k^b}{2}. \tag{4.44}$$

The self-noise defined in (4.42) could therefore be evaluated as

$$J_{self} = E\left\{\|\mathbf{v}_k\|^2\right\} \tag{4.45}$$

$$= E\left\{\left(\frac{\mathbf{v}_k^f + \mathbf{v}_k^b}{2}\right)^H \left(\frac{\mathbf{v}_k^f + \mathbf{v}_k^b}{2}\right)\right\} \tag{4.46}$$

$$= \frac{E\left\{\|\mathbf{v}_k^f\|^2\right\}}{4} + \frac{E\left\{\|\mathbf{v}_k^b\|^2\right\}}{4} + \frac{\mathrm{Re}\left\{E\left\{\left(\mathbf{v}_k^f\right)^H\left(\mathbf{v}_k^b\right)\right\}\right\}}{2}. \tag{4.47}$$

Note that, the computation of $E\{\|\mathbf{v}_k^f\|^2\}$ and $E\{\|\mathbf{v}_k^b\|^2\}$ for frequency-selective channels are different from those for flat-fading channels. The reason for this difference is that the well-known independence assumption employed for the analysis over flat-fading channels is not

valid for frequency-selective channels since $\mathbf{a}_k$'s are correlated for consecutive values of k as a result of the memory of the channel at hand. In [65, 64], an iterative expression is given for the mean-square energy of the tap-weight tracking error which could be expressed at the steady-state as follows

$$E\left\{\|\mathbf{v}_{k,i}\|^2\right\} = \frac{\mu M E_s}{E_s - \mu\left[(M-1)E_s + E_4\right]} J_{min} \tag{4.48}$$

where $E_4 = E\{|a_k|^4\}$. We could further simplify (4.48) assuming BPSK signalling as follows

$$E\left\{\|\mathbf{v}_{k,i}\|^2\right\} = \frac{\mu M}{1 - \mu M} J_{min}. \tag{4.49}$$

Note that $\mathbf{v}_{k,i}$ in (4.48) and (4.49) stands for the i-th element of the tap-weight tracking error for the unidirectional LMS algorithm and is therefore different from (4.45). Note also that (4.49) with $M = 1$ simplifies to the same result presented previously for flat-fading case.

The expectation in the last term of (4.47) could be further elaborated with the help of the modified adaptations (3.38)-(3.39) as follows

$$\begin{aligned}
E\left\{\left(\mathbf{v}_k^f\right)^H\left(\mathbf{v}_k^b\right)\right\} &= E\left\{\left[(1 - 2\mu E_s)\mathbf{v}_{k-1}^f - \mu\,\boldsymbol{\epsilon}_{k-1}^f\right]^H\left[(1 - 2\mu E_s)\mathbf{v}_{k+1}^b - \mu\,\boldsymbol{\epsilon}_{k+1}^b\right]\right\} \tag{4.50} \\
&= (1 - 2\mu E_s)^2 E\left\{\left(\mathbf{v}_{k-1}^f\right)^H \mathbf{v}_{k+1}^b\right\} - \mu\,(1 - 2\mu E_s)E\left\{\left(\mathbf{v}_{k-1}^f\right)^H \boldsymbol{\epsilon}_{k+1}^b\right\} \\
&\quad - \mu\,(1 - 2\mu E_s)E\left\{\left(\boldsymbol{\epsilon}_{k-1}^f\right)^H \mathbf{v}_{k+1}^b\right\} + \mu^2 E\left\{\left(\boldsymbol{\epsilon}_{k-1}^f\right)^H \boldsymbol{\epsilon}_{k+1}^b\right\} \tag{4.51} \\
&= (1 - 2\mu E_s)^2 E\left\{\left(\mathbf{v}_{k-1}^f\right)^H\left(\mathbf{v}_{k+1}^b\right)\right\} \tag{4.52}
\end{aligned}$$

where (4.52) makes use of the assumptions that the elements of $\boldsymbol{\epsilon}_k^f$ and $\boldsymbol{\epsilon}_k^b$ are zero-mean random variables which are mutually uncorrelated from each other and from the elements of $\mathbf{v}_k^f$ and $\mathbf{v}_k^b$, which follows directly from [57]. After sufficient number of iterations, (4.52) becomes

$$E\left\{\left(\mathbf{v}_k^f\right)^H\left(\mathbf{v}_k^b\right)\right\} = (1 - 2\mu E_s)^L E\left\{\left(\mathbf{v}_0^f\right)^H\left(\mathbf{v}_L^b\right)\right\} \tag{4.53}$$

The result obtained in (4.53) could be ignored safely since $|1 - 2\mu E_s| < 1$ is the stability condition of the conventional LMS algorithm, and therefore $(1 - 2\mu E_s)^L \ll 1$. Consequently, the self-noise expression becomes

$$J_{self} = \frac{\mu M^2 E_s}{2\left(E_s - \mu\left[(M-1)E_s + E_4\right]\right)} J_{min}, \tag{4.54}$$

109

which is observed to depend on the step-size μ, the minimum achievable MSE which is equal to the noise variance, the number of taps, i.e., M, of the frequency-selective channel, and second and fourth moments of the input signal. The self-noise part for the bidirectional LMS algorithm is also observed to be half that of the conventional LMS algorithm as before.

4.3.2 Derivation of the Lag Component (J_{lag})

As stated before, the lag component of the MSE arises from the time varying nature of the unknown system. Since we have considered the effect of noisy gradient estimation in the self-noise derivation, we assume perfect gradient estimation in this particular case and focus only on time variation [61, 63]. When the gradient is assumed to be estimated without any noise, the adaptive processes turn to be original steepest descent algorithm as follows

$$
\begin{aligned}
\hat{\mathbf{f}}_{k+1}^f &= \hat{\mathbf{f}}_k^f - \mu\,\nabla_k^f = (1 - 2\mu E_s)\hat{\mathbf{f}}_k^f + 2\mu E_s\,\mathbf{f}_k & (4.55) \\
\hat{\mathbf{f}}_{k-1}^b &= \hat{\mathbf{f}}_k^b - \mu\,\nabla_k^b = (1 - 2\mu E_s)\hat{\mathbf{f}}_k^b + 2\mu E_s\,\mathbf{f}_k. & (4.56)
\end{aligned}
$$

where the true gradients ∇_k^f and ∇_k^b are given in (4.29)-(4.30). In order to cope with the time variation in a better way, we prefer to translate the adaptations into the frequency domain, as in [61]. To this end, z-transform of (4.55) and (4.56) are computed, and the results are then rearranged as follows

$$
\begin{aligned}
\hat{\mathbf{f}}^f(z) &= Z\left\{\hat{\mathbf{f}}_k^f\right\} = \frac{1 - \beta}{z - \beta}\,\mathbf{f}(z) & (4.57) \\
\hat{\mathbf{f}}^b(z) &= Z\left\{\hat{\mathbf{f}}_k^b\right\} = \frac{1 - \beta}{z^{-1} - \beta}\,\mathbf{f}(z) & (4.58)
\end{aligned}
$$

where $\beta = 1 - 2\mu E_s$ is the geometric ratio of the adjustments, as before, $Z\{.\}$ stands for the elementwise z-transform of its vector argument and $\mathbf{f}(z) = Z\{\mathbf{f}_k\}$. The z-transform of $\hat{\mathbf{f}}_k$ is then found through (4.13) as follows

$$
\hat{\mathbf{f}}(z) = \frac{\hat{\mathbf{f}}^f(z) + \hat{\mathbf{f}}^b(z)}{2} = \frac{1}{2}\left(\frac{1 - \beta}{z - \beta} + \frac{1 - \beta}{z^{-1} - \beta}\right)\mathbf{f}(z), \tag{4.59}
$$

and the z-transform of the tap-weight tracking error accordingly becomes

$$
\hat{\mathbf{f}}(z) - \mathbf{f}(z) = H(z)\,\mathbf{f}(z). \tag{4.60}
$$

In (4.60), $H(z)$ is the transfer function for the bidirectional LMS algorithm which is independent of the channel characteristics to be estimated, and is given as

$$H(z) = -\frac{1+\beta}{2\beta} + \frac{1-\beta}{2\beta}\left(\frac{1}{1-\beta z^{-1}} - \frac{1}{1-\frac{1}{\beta}z^{-1}}\right), \tag{4.61}$$

which is equal to the one found in Section 3.3.2.

The next step is to further elaborate the lag expression given in (4.27) in order to get rid of the inner expectations $E\{\hat{\mathbf{f}}_k\}$. Due to the reasons explained in Section 3.3.2, we know that $E\{\hat{\mathbf{f}}_k\} = \hat{\mathbf{f}}_k$ under the assumptions made during the lag derivation. Consequently, the lag expression given in (4.27) becomes

$$\begin{aligned} J_{lag} &= E\left\{\left\|E\{\hat{\mathbf{f}}_k\} - \mathbf{f}_k\right\|^2\right\} = E\left\{\left\|\hat{\mathbf{f}}_k - \mathbf{f}_k\right\|^2\right\} \tag{4.62} \\ &= \sum_{l=0}^{M-1} E\left\{\left|\hat{f}_{k,l} - f_{k,l}\right|^2\right\} \tag{4.63} \end{aligned}$$

One of the consequences of (4.63) is that the J_{lag} could be interpreted as the sum of mean-square energies present in the element-wise tracking errors, i.e., $\hat{f}_{k,l} - f_{k,l}$ for $l = 0, 1, \ldots, M-1$. Due to the symmetry, the mean-square energy in the tracking error is therefore evaluated in the frequency domain as follows [63]

$$J_{lag} = \frac{M}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S(w)\, dw \tag{4.64}$$

with a priori known $S(w)$ which is the power spectrum of the frequency-selective fading channel.

As a result, the final expression for the steady state MSIE is given for an M-tap frequency-selective fading channel as

$$\begin{aligned} J_{MSIE} &= J_{self} + J_{lag} \\ &= \frac{\mu M^2 E_s}{2\left(E_s - \mu\left[(M-1)E_s + E_4\right]\right)} J_{min} + \frac{M}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S(w)\, dw, \tag{4.65} \end{aligned}$$

and the steady-state MSE is therefore found to be

$$\begin{aligned} J_{MSE} &= J_{min} + E_s J_{MSIE} \\ &= \left(1 + \frac{\mu M^2 E_s^2}{2\left(E_s - \mu\left[(M-1)E_s + E_4\right]\right)}\right) J_{min} + \frac{M E_s}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{jw})\right|^2 S(w)\, dw. \tag{4.66} \end{aligned}$$

The optimal choice for the step-size value, μ_{opt}, is again of interest to characterize the minimum achievable MSE at the steady-state. In order to derive μ_{opt} theoretically, we express (4.66) in terms of only β and take derivative of the resulting expression with respect to β as follows

$$
\begin{aligned}
\frac{\partial J_{MSE}}{\partial \beta} &= \frac{\partial}{\partial \beta} \left\{ \left(1 + \frac{(1-\beta)M^2 E_s^2}{2\left(2E_s^2 - (1-\beta)\left[(M-1)E_s + E_4\right]\right)} \right) J_{min} + \frac{E_s M}{2\pi} \int_{-\pi}^{\pi} \left| H(e^{jw}) \right|^2 S(w)\, dw \right\} \\
&= -\frac{E_s^4 M^2}{\left(2E_s^2 - (1-\beta)\left[(M-1)E_s + E_4\right]\right)^2} J_{min} + \frac{E_s M}{\pi} \int_{-\pi}^{\pi} H(e^{jw}) \frac{\partial H(e^{jw})}{\partial \beta} S(w)\, dw
\end{aligned}
\tag{4.67}
$$

where $\partial H(e^{jw})/\partial \beta$ is the same as before which is given as

$$
\frac{\partial H(e^{jw})}{\partial \beta} = -\frac{(1 - \cos w)(1 - \beta^2 - 2\beta + 2\cos w)}{(1 + \beta^2 - 2\beta \cos w)^2}.
\tag{4.68}
$$

The optimal geometric ratio β_{opt} could then be evaluated numerically using (4.67) and (4.68) as follows

$$
\left. \frac{\partial J_{MSE}}{\partial \beta} \right|_{\beta=\beta_{opt}} = 0
\tag{4.69}
$$

and the optimal step-size μ_{opt} could be found as $\mu_{opt} = (1 - \beta_{opt})/2E_s$.

4.3.3 Numerical Results

In this section, we will present numerical results in order to verify the MSE derivation for the bidirectional LMS algorithm operating at the steady-state over frequency-selective Rayleigh fading channels. The superiority of the bidirectional LMS algorithm in tracking the unknown communication channel as compared to the conventional unidirectional LMS and the optimal MMSE filter will also be demonstrated through these results. To this end, we compare the theoretical MSIE of the bidirectional LMS algorithm obtained through the numerical computation of (4.65) with the experimental MSIE results, which are obtained through Monte Carlo simulations, for both the bidirectional and the unidirectional LMS algorithms together with the optimal MMSE filter. We prefer to compare the MSIE values normalized with respect to the number of channel taps, i.e., $\overline{J_{MSIE}}$, which is defined as follows

$$
\overline{J_{MSIE}} = \frac{J_{MSIE}}{M}
\tag{4.70}
$$

where M was previously defined as the number of taps of the frequency-selective channel.

The properties of the frequency-selective channel under consideration is given in Section 4.1. Without any loss of generality, we assume that all taps of the frequency-selective channel experience independent Rayleigh fading with the normalized Doppler $f_d T_s = 0.01$ and the Jakes' power spectrum [98], unless otherwise stated. We also provide some examples for the double-Gaussian ans the Gaussian spectrums to discuss the generality of the previous results. The number of taps of the channel is chosen to be $M = \{2, 4\}$ throughout the simulations where a uniform power delay profile is assumed unless otherwise stated. In each Monte Carlo run, a set of $L = 100$ information symbols are chosen independently from the BPSK alphabet $\mathcal{A} = \{-1, +1\}$ with the symbol energy $E_s = 1$.
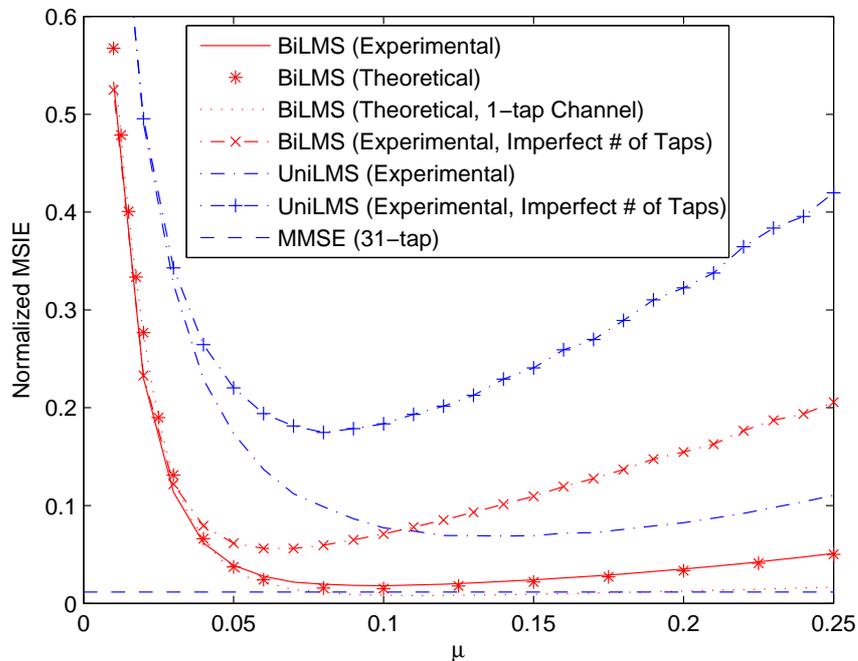
Figure 4.1: Theoretical and experimental normalized MSIE for BiLMS for varying step-size at SNR=10 dB over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. The experimental MSIE for UniLMS and a 31-tap optimal MMSE together with the theoretical MSIE associated with 1-tap channel are also provided.

In Fig. 4.1, the experimental normalized MSIE results for the bidirectional and the unidirectional LMS algorithms and the 31-tap optimal MMSE filter are presented together with the theoretical normalized MSIE for the bidirectional LMS algorithm over a 2-tap frequency se-

lective Rayleigh fading channel at $\gamma = 10$ dB SNR. We observe that the bidirectional LMS algorithm produces much lower normalized MSIE than the unidirectional LMS algorithm does for any choice of the step-size, and the minima associated with the normalized MSIE of the bidirectional LMS algorithm is very close to the optimal MMSE filter with 31 taps. This conclusion verifies the fact that tracking performance of the bidirectional LMS algorithm is much better than that of the unidirectional LMS algorithm and has near-optimal behavior as compared to that of the optimal Wiener filter. For comparison purposes, the results associated with the term $J_{self} = \mu/(2(1-\mu))J_{min}$ which is based on the approach presented in [61] and corresponds to a 1-tap channel with BPSK modulation is also provided which shows a significant deviation from the experimental data. The results for an interesting case in which the receiver employs both the unidirectional and the bidirectional LMS algorithm with an imperfect knowledge of number of channel taps are also provided. We observe that both algorithms yield a degraded results when the number of channel taps is imperfectly by the receiver taken to be 1 over a 2-tap multipath fading channel.
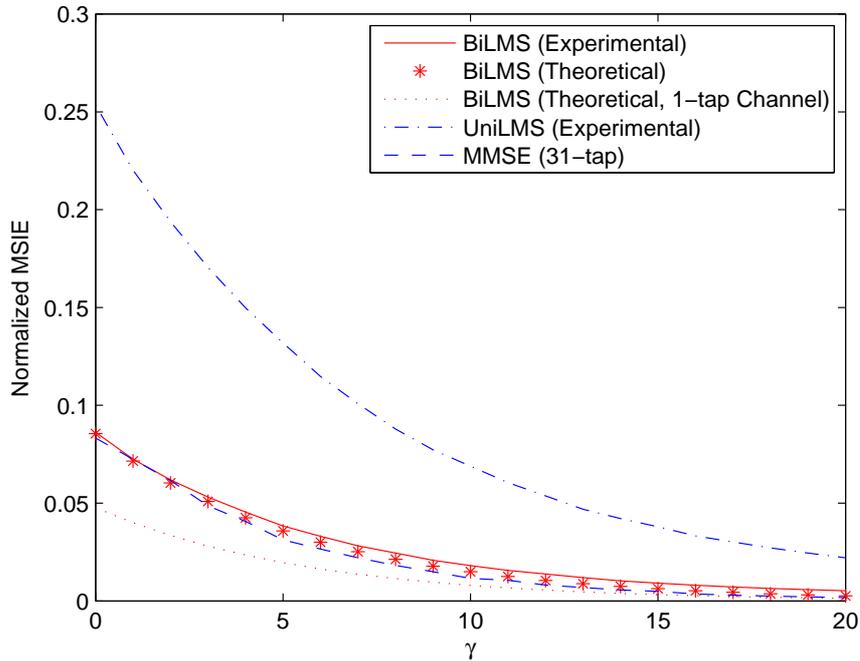
Figure 4.2: Theoretical and experimental normalized MSIE for BiLMS for optimal step-size over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. The experimental MSIE for UniLMS and a 31-tap optimal MMSE together with the theoretical MSIE associated with 1-tap channel are also provided.

114

We present similar results in Fig. 4.2, but for varying SNR and by using optimal step-size values for both the bidirectional and the unidirectional LMS algorithms. We observe that the theoretical MSIE results of the bidirectional LMS exhibit a good match to the experimental ones both of which are very close to that of the 31-tap optimal MMSE filtering for each SNR level when optimal values for the step-size are used. We also present the experimental MSIE for the optimal MMSE filters in Fig. 4.3 with various number of taps. By comparing Fig. 4.2 and Fig. 4.3, we observe that the MMSE filter with a number of taps smaller than 31 achieve a worse tracking performance than the bidirectional LMS especially at low SNR regime. Hence, the bidirectional LMS could be considered to be superior over the MMSE filter in the sense that the former achieves a very similar tracking performance to the latter at an extremely low computational complexity.
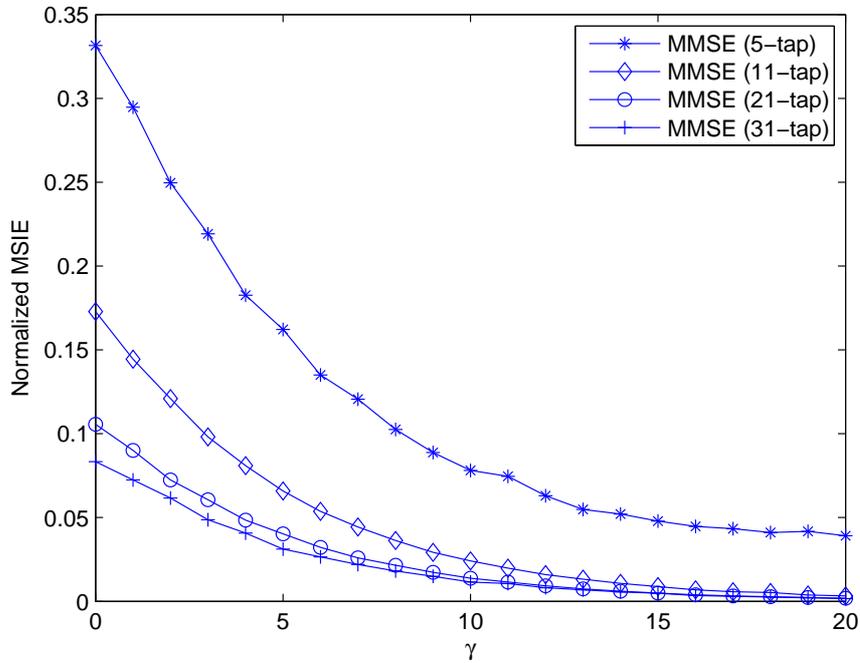
Figure 4.3: Experimental normalized MSIE for MMSE with various number of taps over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$.

In the subsequent figures, we present the performance results over a 4-tap Rayleigh fading frequency-selective channel with $f_d T_s = 0.01$. In Fig. 4.4, we observe that the theoretical results for the bidirectional LMS algorithm again closely follow the experimental ones for any choice of the step-size. Together with the results presented in Fig. 4.4, we also observe

that the tracking performance of the bidirectional LMS algorithm is much better than that of the unidirectional LMS algorithm and is very close to the that of the optimal MMSE filter with 31-tap.
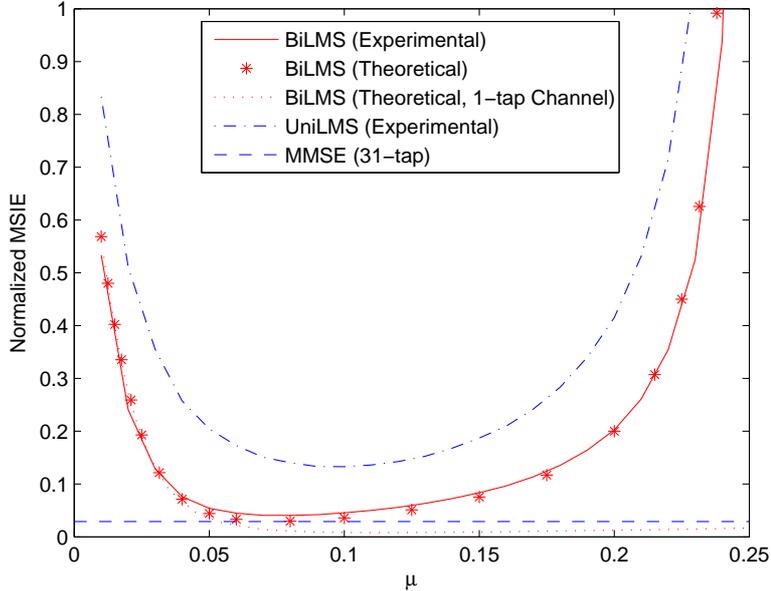
Figure 4.4: Theoretical and experimental normalized MSIE for BiLMS for varying step-size at SNR=10 dB over a 4-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. The experimental MSIE for UniLMS and a 31-tap optimal MMSE together with the theoretical MSIE associated with 1-tap channel are also provided.
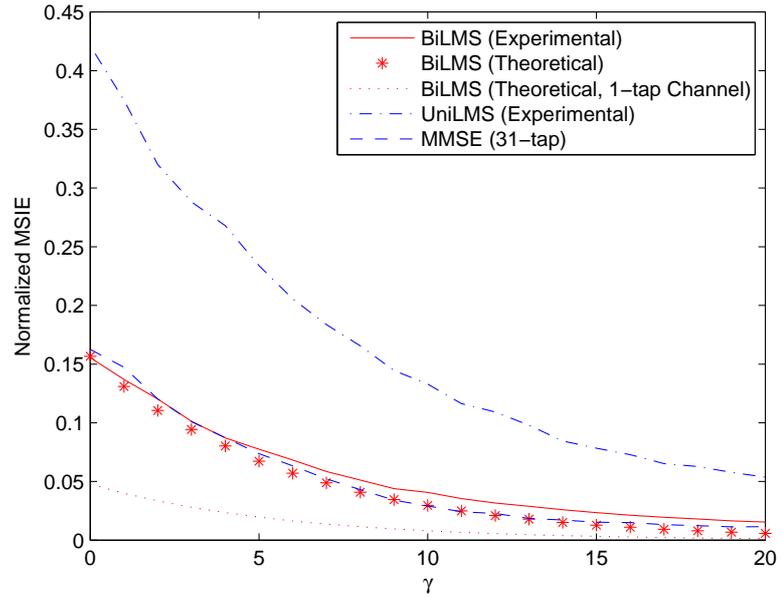
Figure 4.5: Theoretical and experimental normalized MSIE for BiLMS for optimal step-size over a 4-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. The experimental MSIE for UniLMS and a 31-tap optimal MMSE together with the theoretical MSIE associated with 1-tap channel are also provided.
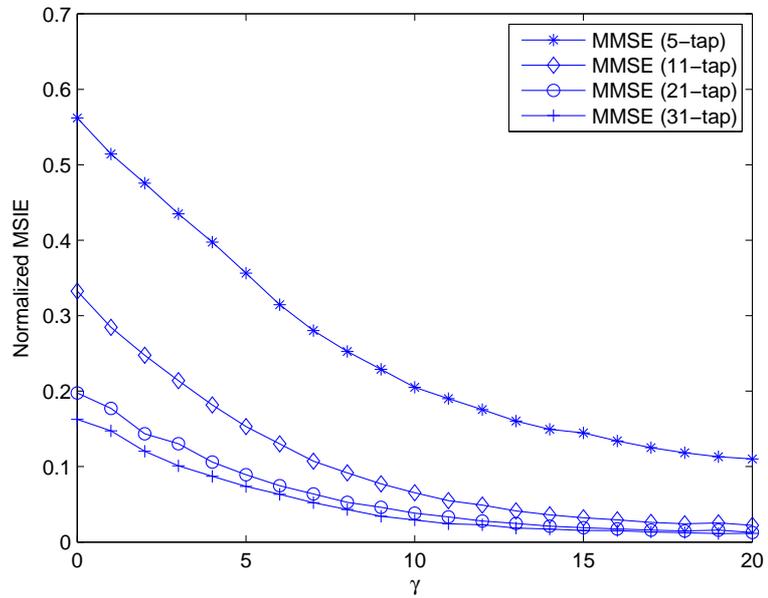
Figure 4.6: Experimental normalized MSIE for MMSE with various number of taps over a 4-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$.

We now investigate the effectiveness of the theoretical optimal steps-size values, i.e., μ_{opt}'s, for the bidirectional LMS algorithm computed according to (4.67) and (4.68). We present the resulting theoretical values for μ_{opt} together with the associated experimental ones for a 2-tap and 4-tap Rayleigh fading frequency-selective channels with $f_d T_s = 0.01$ in Table 4.1 and Table 4.2, respectively. We observe from these results that the theoretical μ_{opt} values are very close to the experimental results for a variety of SNR and number of channel tap choices. This result is believed to have a significant practical importance from system design point of view since it eliminates the necessity of excessive experiments to find μ_{opt} for a genie-aided scenario.

Table 4.1: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for a 2-tap Rayleigh Fading ISI Channel with $f_d T_s = 0.01$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.060	0.060	0.070	0.080	0.090	0.100	0.100	0.120	0.120
Theoretical	0.056	0.062	0.069	0.076	0.084	0.092	0.100	0.109	0.119

Table 4.2: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for a 4-tap Rayleigh Fading ISI Channel with $f_d T_s = 0.01$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.040	0.050	0.050	0.060	0.070	0.080	0.090	0.110	0.110
Theoretical	0.044	0.050	0.055	0.061	0.067	0.074	0.080	0.087	0.095

We also investigate the performance of the bidirectional LMS algorithm for nonuniform power delay profiles. Towards this end, we use the 3-tap Proakis-B channel [102] for which the impulse response of the average magnitudes is given as follows

$$\mathbf{h} = \{0.407, 0.815, 0.407\} \tag{4.71}$$
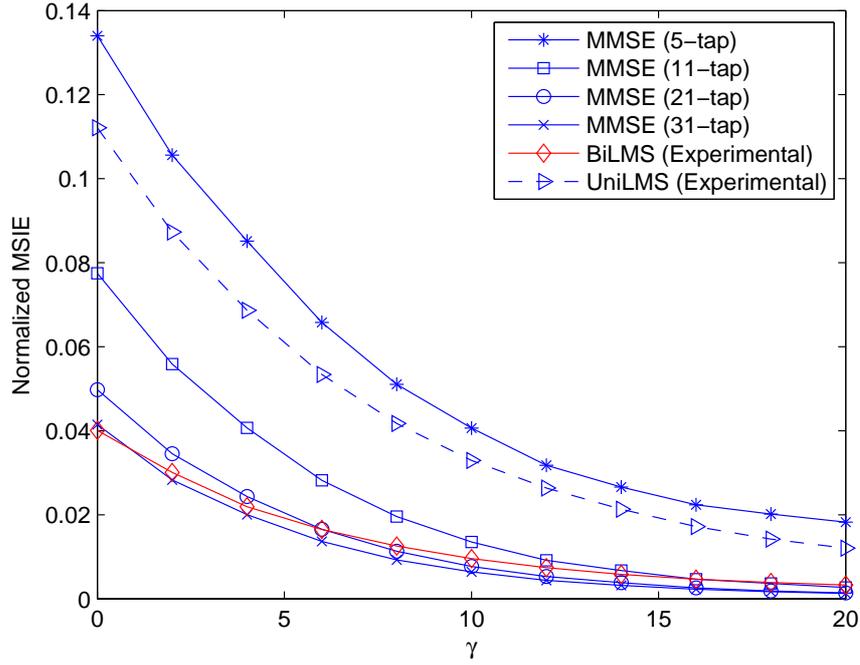
Figure 4.7: Experimental normalized MSIE for BiLMS and UniLMS with optimal step-size and MMSE filter with various taps over a 3-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. A nonuniform power delay profile of Proakis-B channel is used.

The overall fading coefficients for the time-varying channel taps are obtained accordingly as follows

$$f'_{k,l} = h_l f_{k,l} \tag{4.72}$$

for $l = 0, 1, \ldots, M - 1$ where h_l is the l-th element of $\mathbf{h}$ and $f_{k,l}$ is a unit-energy sample from a Rayleigh fading realization with $f_d T_s = 0.01$. Keeping the other system parameters the same, Fig. 4.7 presents the experimental normalized MSIE results for the bidirectional and the unidirectional LMS algorithms both of which operate with the optimal step-size values together with the MMSE filter with various number of taps. We observe that the bidirectional LMS algorithm is again superior over the unidirectional LMS algorithm at any SNR value, and has a similar tracking performance with the 31-tap MMSE filter. We also note that, even the 21-tap MMSE filter has a worse error performance than the bidirectional LMS algorithm at low SNR regime.

Finally, we consider the MSIE results in Fig.4.8-4.11 for 2-tap and 3-tap multipath fading channels with the double-Gaussian and the Gaussian spectrums characterized by the same

parameters given in Section 3.3.3. From these results, the bidirectional LMS algorithm is observed to have the same advantage of near-optimal performance. As a difference, we observe that there are some deviations in the theoretical results from the experimental ones for the lag part which represents the time-varying nature of the underlying channel. Note that, such deviations do not significantly change the optimal step-size value and the minimum achievable MSIE. Note also that, fading channels with AR type autocorrelation are discussed in Section 3.7.1 to show a similar behavior when channel dynamics are changing very fast which is the case here.
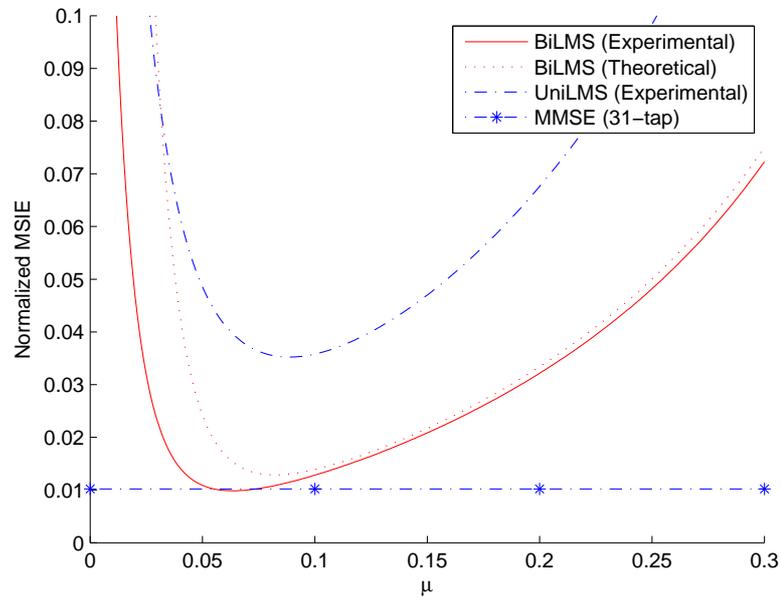
Figure 4.8: MSIE for UniLMS, BiLMS and 31-tap MMSE for a 2-tap ISI channel with the double-Gaussian spectrum characterized by $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz at $\gamma = 5$ dB.
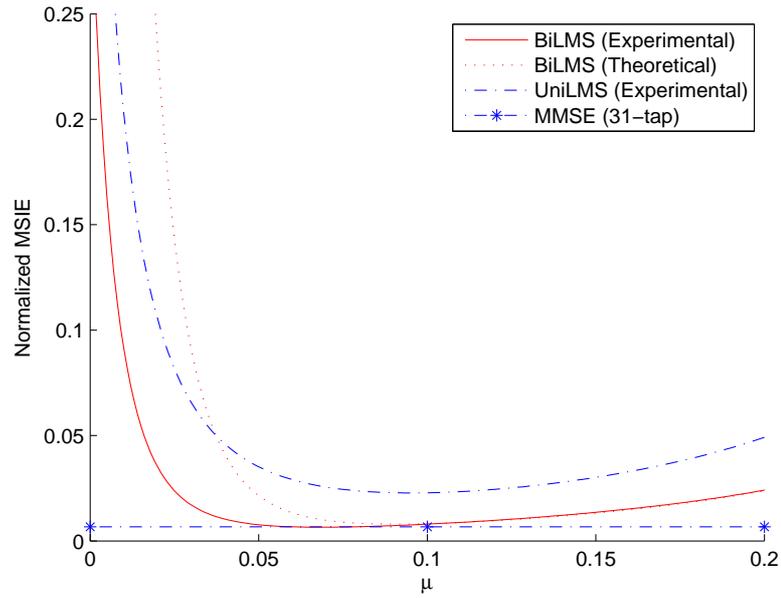
Figure 4.9: MSIE for UniLMS, BiLMS and 31-tap MMSE for a 3-tap ISI channel with the double-Gaussian spectrum characterized by $(C_a, C_b) = (0.5, 1)$, $(f_a, f_b) = (40, -50)$ Hz and $(\sigma_a, \sigma_b) = (30, 20)$ Hz at $\gamma = 15$ dB.
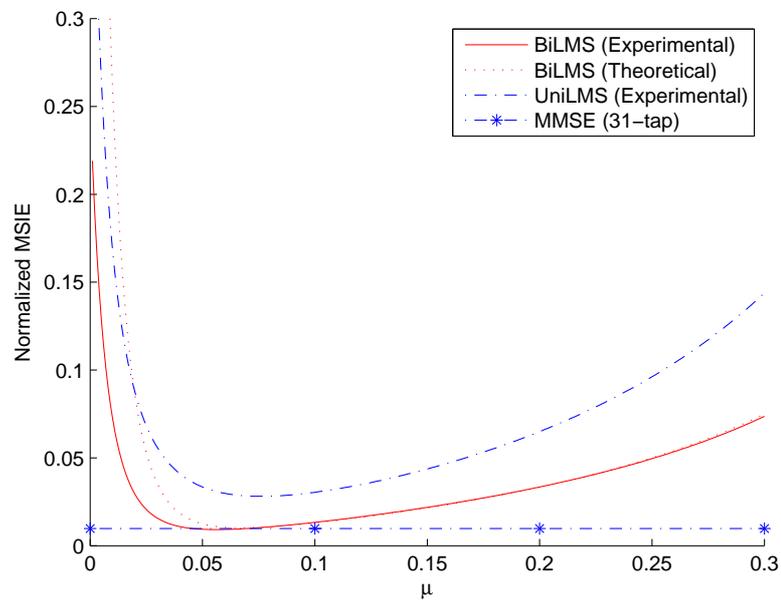
Figure 4.10: MSIE for UniLMS, BiLMS and 31-tap MMSE for a 2-tap ISI channel with the Gaussian spectrum characterized by $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz at $\gamma = 5$ dB.
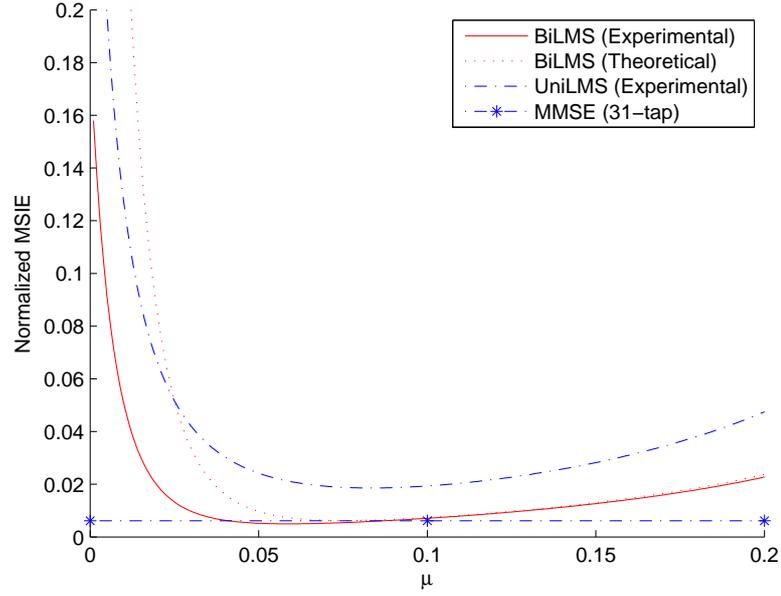
Figure 4.11: MSIE for UniLMS, BiLMS and 31-tap MMSE for a 3-tap ISI channel with the Gaussian spectrum characterized by $(C_a, C_b) = (1, 0)$, $(f_a, f_b) = (0, 0)$ Hz and $(\sigma_a, \sigma_b) = (40, 0)$ Hz at $\gamma = 15$ dB.

4.4 Effect of Imperfect Doppler and SNR Information

In this section, we generalize the results presented in Section 3.5 to the multipath fading channels with the system model given in Section 4.1. As a difference, we make use of the estimates of the independent channel paths in order to improve statistical efficiency. The LS cost function for the Doppler spread estimation accordingly becomes

$$F(f_d) = \frac{1}{Q} \frac{1}{M} \sum_{q=1}^{Q} \sum_{m=0}^{M-1} \sum_{l=1}^{L_T-1} \left| \frac{\hat{K}_{q,m}(l)}{\hat{K}_{q,m}(0)} - \frac{r(l; f_d)}{r(0; f_d)} \right|^2 \tag{4.73}$$

where $\hat{K}_{q,m}(l)$ is

$$\hat{K}_{q,m}(l) = \frac{1}{L_T - l} \sum_{k=0}^{L_T-l} \tilde{f}_{k,m}^q (\tilde{f}_{k+l,m}^q)^* . \tag{4.74}$$

We now present the associated numerical results over a 2-tap Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ for a number of $L = 200$ independent and identically

distributed BPSK symbols. In Fig. 4.12, the robustness of the algorithms under consideration to the imperfect Doppler estimate is depicted at $\gamma = 10$ dB. The performances of the Doppler and the SNR estimation algorithms are also presented in Fig. 4.13 for which we assume that $\hat{f}_d \in [0, 500]$ Hz. Since we incorporate the statistics available through the independent channel taps as an important difference from the flat-fading case, employing multiple frames is observed to achieve no significant performance improvement over the single frame case and we therefore choose $Q = 1$. The observations are very similar to that of the flat-fading case.
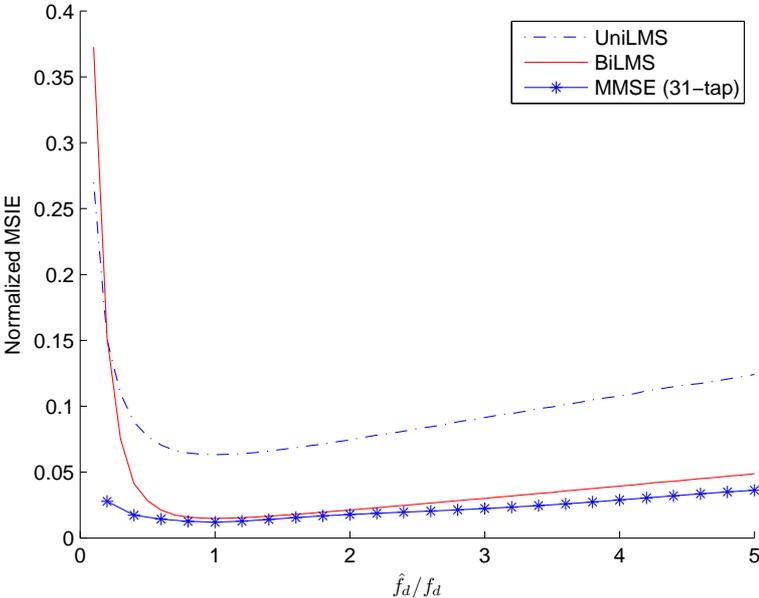
Figure 4.12: MSIE for UniLMS, BiLMS and 31-tap MMSE over a 2-tap Rayleigh fading channel for imperfect Doppler spread estimate, $\hat{f}_d$, where $\gamma = 10$ dB, $f_d = 100$ Hz and $T_s = 0.1$ ms.
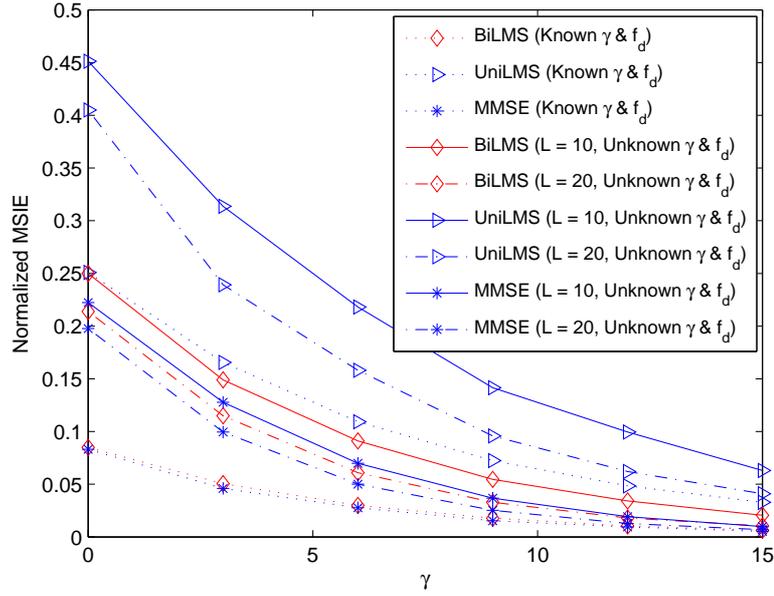
Figure 4.13: MSIE for UniLMS, BiLMS and 31-tap MMSE for known and estimated Doppler and SNR over a 2-tap Rayleigh fading channel with Jakes' spectrum where $f_d = 100$ Hz and $T_s = 0.1$ ms.

4.5 Effect of Imperfect Initialization

In this section, we generalize the results on the imperfect initialization presented in Section 3.6 to the frequency-selective fading channels. We again use a training block which consists of L_T independent and identically distributed BPSK symbols in order to provide an initial value for the fading vector at the beginning of the data block. The observations are received according to (4.2) with the signal variance of σ^2. We consider the LS initialization given as

$$\hat{\mathbf{f}} = \mathbf{A}^H \left(\mathbf{A}\mathbf{A}^H \right)^{-1} \mathbf{y}, \tag{4.75}$$

as well as the zero initialization for the frequency-selective fading channels. In (4.75), $\hat{\mathbf{f}} = [\hat{\mathbf{f}}_1^T \ldots \hat{\mathbf{f}}_{L_T}^T]^T$ is the estimate of the fading vectors, $\mathbf{A} = \text{diag}\{\mathbf{a}_1^T, \ldots \mathbf{a}_{L_T}^T\}$ is the pilot matrix and $\mathbf{y} = [y_1 \ldots y_L]^T$ is the observation vector.

The Monte Carlo results for the bidirectional LMS algorithm with the aforementioned initialization methods for various data block lengths and SNR are presented in Fig.3.30-3.32. The channel is a 2-tap Rayleigh fading with Jakes' spectrum and $f_d T_s = 0.01$. We again observe

124

that zero initialization is sufficient for most of the blocklengths and that LS initialization with $L_T = 2$ achieves a satisfactory performance.
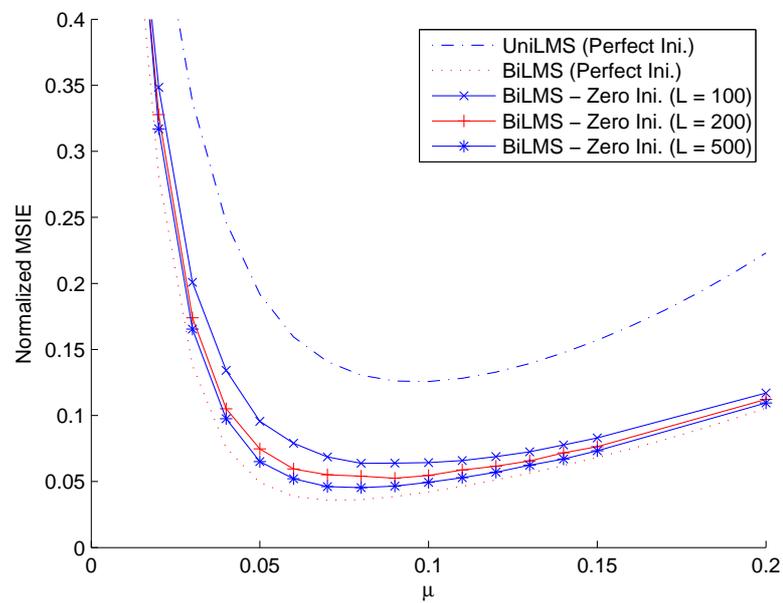
Figure 4.14: MSIE for BiLMS with zero and LS initializations together with perfectly initialized UniLMS over 2-tap Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ at $\gamma = 5$ dB.
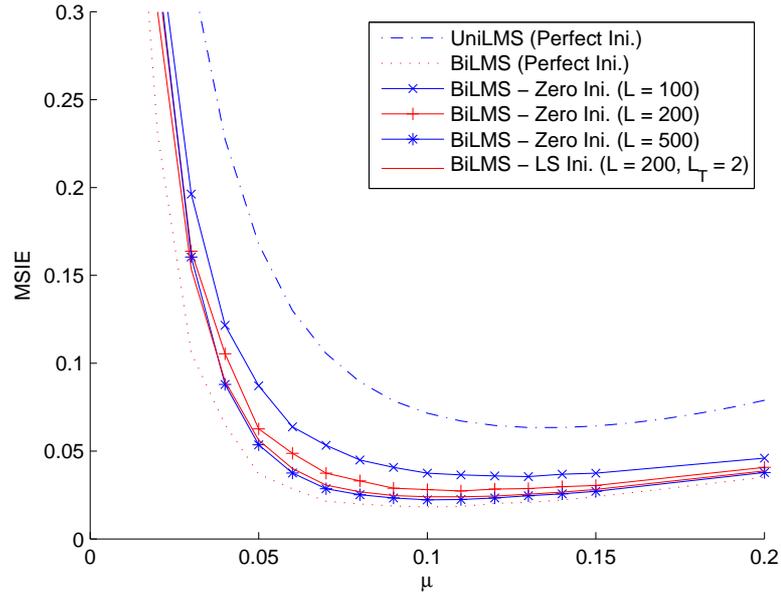
Figure 4.15: MSIE for BiLMS with zero and LS initializations together with perfectly initialized UniLMS over 2-tap Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ at $\gamma = 10$ dB.
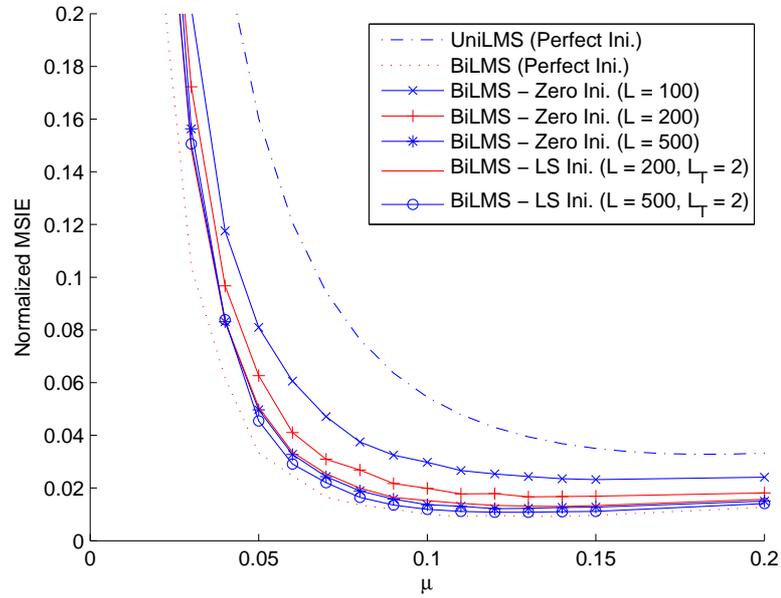
Figure 4.16: MSIE for BiLMS with zero and LS initializations together with perfectly initialized UniLMS over 2-tap Rayleigh fading channel with Jakes' spectrum and $f_d T_s = 0.01$ at $\gamma = 15$ dB.

For frequency-selective fading channels, we consider the Modified CRB (MCRB) in order to determine a lower bound on the initialization error. The MCRB is presented in [117] for scalar parameter estimation problems and then generalized in [118] to the vector case. It is much easier to compute the MCRB than the CRB in the presence of the nuisance parameters which are the transmitted symbols in our case. In [117], it is also shown that the MCRB approaches the CRB in many cases of interest, especially at high SNR region which is given in [119] as $\gamma > 0$ dB.

The MCRB for the frequency-selective fading channels under consideration is given as

$$E\left\{\left|\hat{f}_{i,j} - f_{i,j}\right|^2\right\} \geq \left(\mathbf{J}^{-1}\right)_{mm} \tag{4.76}$$

where $m = i(M-1) + j$ is the modified index. The *modified* Fisher's information matrix (MFIM) for this particular case is defined as

$$J_{mm'} = E_{\mathbf{a}}\left\{E_{\mathbf{y},\mathbf{f}}\left\{\frac{\partial \ln p(\mathbf{y},\mathbf{f}\,|\,\mathbf{a})}{\partial f_{i,j}}\frac{\partial \ln p(\mathbf{y},\mathbf{f}\,|\,\mathbf{a})}{\partial f_{i',j'}}\right\}\right\} \tag{4.77}$$

$$= -E_{\mathbf{a}}\left\{E_{\mathbf{y},\mathbf{f}}\left\{\frac{\partial^2 \ln p(\mathbf{y},\mathbf{f}\,|\,\mathbf{a})}{\partial f_{i,j}\,\partial f_{i',j'}}\right\}\right\} \tag{4.78}$$

for which the following expectations could be readily obtained as before

$$J_{mm'} = \underbrace{-E_{\mathbf{a}}\left\{E_{\mathbf{f}}\left\{E_{\mathbf{y}|\mathbf{f}}\left\{\frac{\partial^2 \ln p(\mathbf{y}\,|\,\mathbf{f},\mathbf{a})}{\partial f_{i,j}\,\partial f_{i',j'}}\right\}\right\}\right\}}_{J^1_{mm'}}\underbrace{-E_{\mathbf{f}}\left\{\frac{\partial^2 \ln p(\mathbf{f})}{\partial f_{i,j}\,\partial f_{i',j'}}\right\}}_{J^2_{mm'}} \tag{4.79}$$

$$= J^1_{mm'} + J^2_{mm'}. \tag{4.80}$$

Consider the following probability density function

$$p(y_k\,|\,\mathbf{f}_k,\mathbf{a}_k) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(y_k - \mathbf{f}_k^T\mathbf{a}_k\right)^2}{2\sigma^2}\right\} \tag{4.81}$$

by which we could obtained the joint probability density function as follows

$$p(\mathbf{y}\,|\,\mathbf{f},\mathbf{a}) = \prod_{k=1}^{L_T} p(y_k\,|\,\mathbf{f}_k,\mathbf{a}_k) = (2\pi\sigma^2)^{-\frac{L_T}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{L_T}p(y_k\,|\,\mathbf{f}_k,\mathbf{a}_k)\right\}. \tag{4.82}$$

The associated derivatives are then computed as follows

$$\frac{\partial \ln p(\mathbf{y}\,|\,\mathbf{f},\mathbf{a})}{\partial f_{i,j}} = \frac{a_{i-j}}{\sigma^2}(y_i - \mathbf{f}_i^T\mathbf{a}_i) \tag{4.83}$$

$$\frac{\partial^2 \ln p(\mathbf{y}\,|\,\mathbf{f},\mathbf{a})}{\partial f_{i,j}\,\partial f_{i',j'}} = -\frac{a_{i-j}\,a_{i'-j'}}{\sigma^2}\delta_{ii'}. \tag{4.84}$$

127

Therefore, $J^1_{mm'}$ is obtained as

$$J^1_{mm'} = -E_{\mathbf{a}} \left\{ E_{\mathbf{f}} \left\{ E_{\mathbf{y}|\mathbf{f}} \left\{ -\frac{a_{i-j} \, a_{i'-j'}}{\sigma^2} \delta_{ii'} \right\} \right\} \right\} = \frac{1}{\sigma^2} \delta_{ii'} \, \delta_{jj'} \,. \tag{4.85}$$

From the results presented in Section 3.6, we obtain

$$J^2_{mm'} = -E_{\mathbf{f}} \left\{ \frac{\partial^2 \ln p(\mathbf{f})}{\partial f_{i,j} \, \partial f_{i',j'}} \right\} = \left(\mathbf{R}_f^{-1} \right)_{mm'} \,. \tag{4.86}$$

As a result, MCRB for the frequency selective fading channels under consideration is found to be

$$J_{mm'} = \frac{1}{\sigma^2} \delta_{ii'} \, \delta_{jj'} + \left(\mathbf{R}_f^{-1} \right)_{mm'} \,. \tag{4.87}$$

By comparing the expressions in (3.140) and (4.87) with $M = 1$, we observe that the MCRB expression is equivalent to that of the CRB except for the term including the following integration

$$I_1 = -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{\int_{-\infty}^{\infty} \frac{u^2 \exp\left\{ -\frac{\sigma^2}{2} u^2 \right\}}{\cosh(u \, v)} du}_{I_2} \exp\left\{ -\frac{1 + \sigma^2}{2\sigma^2} v^2 \right\} dv \tag{4.88}$$

where I_2 is shown in [119] to go to 0 for the moderate and high SNR values which fulfills the equivalence of the MCRB and the exact CRB bounds for these SNR regions. In order to verify this result, we compute the MCRB according to (4.87) for the flat-fading channel specified in Section 3.6.2 and obtain the same results associated with the CRB.

In Table 4.3, we present the MSIE for LS initialization and the associated MCRB with $L_T = 2$ over the 2-tap Rayleigh fading channel specified in this section. In order to improve the performance of the LS estimation, we assume that the channel is not changing during $L_T = 2$ symbol intervals although it is changing very slowly during this time slot. Note again that, zero initialization yields $J_{MSIE} = 1$ as argued before, and that the LS initialization with $L_T = 2$ results in a satisfactory performance in terms of initialization. In addition, the associated MCRB guarantees that much lower MSIE values are also possible for $L_T = 2$. As a final remark, all SNR computation is according to (4.4).

Table 4.3: MCRB and MSIE for LS Initialization over 2-tap Rayleigh Fading with $f_d T_s = 0.01$ and $L_T = 2$.

SNR	5 dB	10 dB	15 dB
MCRB	0.1370	0.0481	0.0160
LS	0.9444	0.3004	0.0961

4.6   Iterative Channel Estimation for Frequency-Selective Channels

This section considers the iterative channel estimation method with the estimation algorithms under consideration in a more realistic communication scenario with a coded time-varying frequency-selective channel. Although some the following parts are similar to those presented in Section 3.8, we would like to explicitly overview the transceiver model and the estimation algorithms leaving some details to the previous sections to provide a better understanding.

4.6.1   Transmitter and Receiver Models

In this section, we make use of the equivalent discrete-time complex baseband channel model given in (4.1) for time-varying frequency selective channels. As a difference, the transmitted symbols $\{a_k\}_{k=1}^{L}$ are assumed not to be chosen independently from a finite alphabet $\mathcal{A}$ any more and are generated by the concatenated structure including channel encoder, interleaver, modulator and the PSAM block. The mechanism of the PSAM block is now modified according to the channel characteristics such that it inserts not a single pilot symbol but a short block of pilot symbols periodically, and the iterative receiver is updated accordingly.
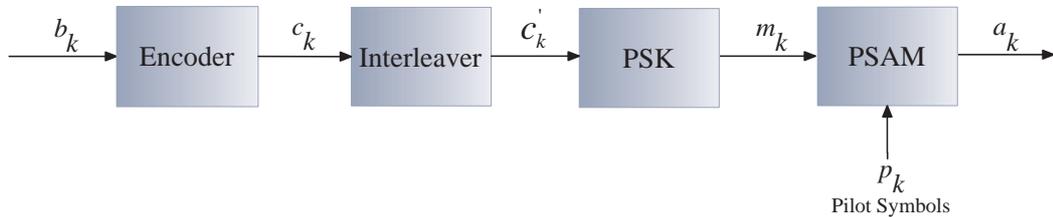
Figure 4.17: Transmitter model including channel encoder, interleaver, PSK modulator and PSAM block.

The transmitter block diagram under consideration is shown in Fig. 4.17 which is similar to

Fig. 3.41 except that some blocks operate in a different fashion details of which is explained later in this section. In this model, a set of binary information symbols $\{b_k\}_{k=1}^{L_d}$ are generated randomly from the binary set $\{0, 1\}$ in an independent fashion, and are encoded by a channel code of rate R_c. The set of coded symbols $\{c_k\}_{k=1}^{L_c}$ with $L_c = \frac{L_d}{R_c}$ are then interleaved to increase the resistance against the burst errors. The resulting interleaved coded symbols $\{c'_k\}_{k=1}^{L_c}$ are then modulated using a PSK modulator, as before, with a finite modulation alphabet $\mathcal{A}$. The symbols $\{m_k\}_{k=1}^{L_m}$ at the output of the modulator with $L_m = \frac{L_d}{R_c \log_2 |\mathcal{A}|}$ are then multiplexed with a set of pilot symbols $\{p_k\}_{k=1}^{ML_p}$ which are known a priori at the receiver and are chosen from the same alphabet $\mathcal{A}$. The resulting set of symbols $\{a_k\}_{k=1}^{L}$ are then transmitted through the time-varying multipath fading channel given in (4.1).

The PSAM mechanism for this particular case operates such that the modulated symbols are first split into groups of $M_p - M$ symbols prior to the multiplexing, which is called the pilot block, where M is the number of resolvable multipaths present in the channel. Then, M pilot symbols are inserted into each of these pilot blocks after the $\left(\frac{M_p+1}{2} - M\right)$-th individual element as shown in Fig. 4.18. Note that this transmission scheme enables channel estimation unit to employ the symbol vectors $\left\{\mathbf{a}_{p(i)}\right\}_{i=1}^{L_p}$ with $p(i) = (i - 1)M_p + \left(\frac{M_p+1}{2} - M\right)$ since elements of $\mathbf{a}_{p(i)}$ are the a priori known pilot symbols such that $\mathbf{a}_{p(i)} = \left[a_{p(i)} \, a_{p(i)-1} \, \cdots \, a_{p(i)-M+1}\right]^T$.
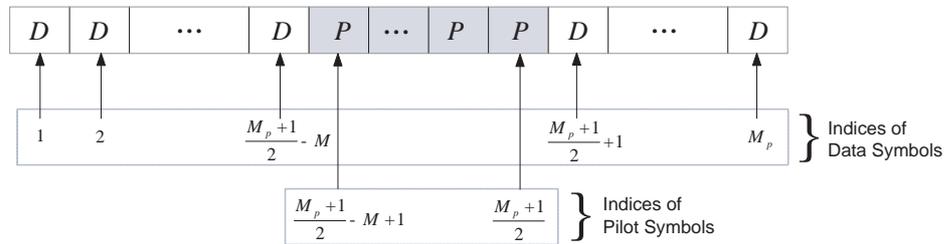
Figure 4.18: The structure of a single group of M_p symbols in a transmitted PSAM block for frequency-selective fading channels.

In this modified PSAM transmission scheme, the necessary number of pilot symbols are ML_p where L_p is computed to be

$$L_p = \left\lceil \frac{L_m - \left(\frac{M_p+1}{2} - M\right)}{M_p - M} \right\rceil + 1 = \left\lceil \frac{\frac{L_d}{R_c \log_2 |\mathcal{A}|} - \left(\frac{M_p+1}{2} - M\right)}{M_p - M} \right\rceil + 1. \qquad (4.89)$$

130

As a result, the final data sequence $\{a_k\}_{k=1}^L$ to be transmitted is given as follows

$$a_k = \begin{cases} m_{k_m(k)}, & k \neq (i-1)M_p + \frac{M_p+1}{2} - (j-1), \ i = 1,2,\ldots,L_p, \ j = 1,2,\ldots,M; \\ p_{k_p(k)}, & k = (i-1)M_p + \frac{M_p+1}{2} - (j-1), \ i = 1,2,\ldots,L_p, \ j = 1,2,\ldots,M. \end{cases}$$

(4.90)

where the indices $k_m(k)$ and $k_p(k)$ are given to be

$$k_m(k) = k - \left(\left\lfloor \frac{\left| k - \left(\frac{M_p+1}{2} - M \right) \right|}{M_p} \right\rfloor + 1 \right) M,$$

(4.91)

$$k_p(k) = k - \left(\frac{M_p+1}{2} - M \right) - \left\lfloor \frac{\left| k - \left(\frac{M_p+1}{2} - M \right) \right|}{M_p} \right\rfloor (M_p - M).$$

(4.92)

The associated frame length L is then computed accordingly to be

$$L = \frac{L_d}{R_c \log_2 |\mathcal{A}|} + L_p = \frac{L_d}{R_c \log_2 |\mathcal{A}|} + \left(\left\lfloor \frac{\frac{L_d}{R_c \log_2 |\mathcal{A}|} - \left(\frac{M_p+1}{2} - M \right)}{M_p - M} \right\rfloor + 1 \right) M.$$

(4.93)

We also define the set of indices associated with the pilot symbols as follows

$$P_p = \left\{ (i-1)M_p + \frac{M_p+1}{2} - (j-1) \right\}_{(i,j)=(1,1)}^{(L_p,M)}$$

(4.94)

$$= \left\{ \underbrace{\frac{M_p+1}{2} - M + 1, \ldots, \frac{M_p+1}{2}}_{\text{1-st pilot block}}, \ldots, \right.$$

(4.95)

$$\left. \ldots, \underbrace{\frac{(2L_p-1)M_p+1}{2} - M + 1, \ldots, \frac{(2L_p-1)M_p+1}{2}}_{L_p\text{-th pilot block}} \right\}.$$

(4.96)

Because the optimal design of the patterns or the values for the pilot symbols is again beyond the scope of this work, we choose the pilot symbols to be equal to one of the elements of the modulation alphabet $\mathcal{A}$ without any loss of generality.

We employ the similar receiver structure as in the flat-fading case which is shown in Fig. 4.19. In this iterative receiver, the observations are first sent to the channel estimation unit and the resulting estimate is employed in LLR computation. The resulting LLR sequence is deinterleaved and pushed to the soft decoder. The log-MAP algorithm is employed in the soft decoder to produce the soft estimates of both the original data symbols and the coded symbols. The initial channel estimate is obtained by making use of the observations and the a priori known pilot symbols only. The quality of the channel estimate is improved by making use of the soft

131
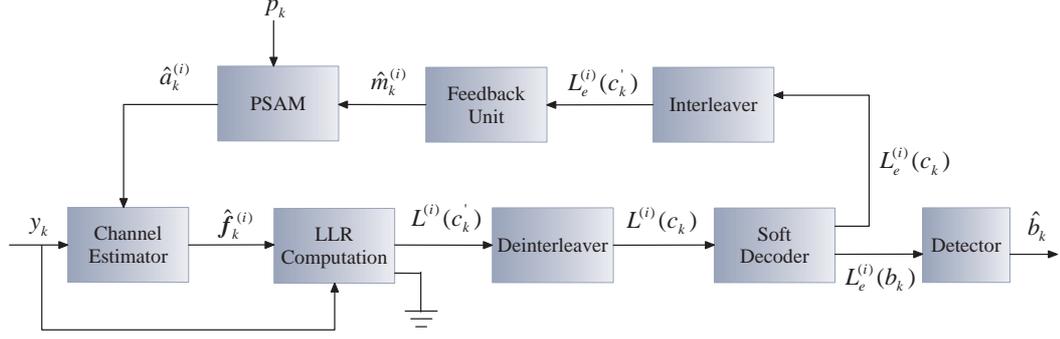
Figure 4.19: Receiver model for a time-varying frequency-selective fading channel with iterative channel estimation.

information on the coded symbols through iterations, as before. We again assume that LLR computation and feedback units are capable of symbol-to-bit level LLR conversion and vice versa, if necessary.

The input LLR of the symbol a_k at i-th estimation iteration is given as

$$L^{(i)}(a_k) = \log \frac{P\left(a_k = 1 \,\middle|\, \mathbf{y}, \hat{\mathbf{f}}^{(i)}\right)}{P\left(a_k = -1 \,\middle|\, \mathbf{y}, \hat{\mathbf{f}}^{(i)}\right)} \tag{4.97}$$

$$= \log \frac{P\left(\mathbf{y}, a_k = 1 \,\middle|\, \hat{\mathbf{f}}^{(i)}\right)}{P\left(\mathbf{y}, a_k = -1 \,\middle|\, \hat{\mathbf{f}}^{(i)}\right)} \tag{4.98}$$

which is a suboptimal approach as stated in Chapter 3, and where $\hat{\mathbf{f}}^{(i)} = \left[\left(\hat{\mathbf{f}}_k^{(i)}\right)^T \dots, \left(\hat{\mathbf{f}}_{k+M-1}^{(i)}\right)^T\right]^T$ is the set of estimates of the unknown multipath fading channel at the i-th iteration and $\mathbf{y} = [\, y_k \dots, y_{k+M-1} \,]^T$ is the set of observations involving the contribution of a_k. $L^{(i)}(a_k)$ could be further elaborated as follows

$$L^{(i)}(a_k) = \log \frac{\sum_{\mathbf{a}_k' : a_k = 1} P\left(\mathbf{y}, \mathbf{a}_k' \,\middle|\, \hat{\mathbf{f}}^{(i)}\right)}{\sum_{\mathbf{a}_k' : a_k = -1} P\left(\mathbf{y}, \mathbf{a}_k' \,\middle|\, \hat{\mathbf{f}}^{(i)}\right)} \tag{4.99}$$

$$= \log \frac{\sum_{\mathbf{a}_k' : a_k = 1} P\left(\mathbf{y} \,\middle|\, \hat{\mathbf{f}}^{(i)}, \mathbf{a}_k'\right)}{\sum_{\mathbf{a}_k' : a_k = -1} P\left(\mathbf{y} \,\middle|\, \hat{\mathbf{f}}^{(i)}, \mathbf{a}_k'\right)} \tag{4.100}$$

where the constraint $\mathbf{a}_k' : a_k$ is defined as the set $\mathbf{a}_k' = [\, a_{k-M+1} \dots, a_0, \dots, a_{k+M-1}]$ in which

a_k takes the specified value. Since y_k's are uncorrelated for a given channel estimate, an equivalent expression is obtained as follows

$$L^{(i)}(a_k) = \log \frac{\displaystyle\sum_{\mathbf{a}_k':a_k=1} \prod_{n=k}^{k+M-1} P\left(y_n \left| \hat{\mathbf{f}}_n^{(i)}, \mathbf{a}_n \right.\right)}{\displaystyle\sum_{\mathbf{a}_k':a_k=-1} \prod_{n=k}^{k+M-1} P\left(y_n \left| \hat{\mathbf{f}}_n^{(i)}, \mathbf{a}_n \right.\right)} \qquad (4.101)$$

where $\mathbf{a}_n$ is previously defined as $\mathbf{a}_n = [a_n \ldots, a_{n-M+1}]$. Using the channel model given in (4.1), the final form of the corresponding LLR becomes

$$L^{(i)}(a_k) = \log \frac{\displaystyle\sum_{\mathbf{a}_k':a_k=1} \prod_{n=k}^{k+M-1} \exp\left\{-\frac{\left|y_n - \left(\hat{\mathbf{f}}_n^{(i)}\right)^T \mathbf{a}_n\right|^2}{N_0}\right\}}{\displaystyle\sum_{\mathbf{a}_k':a_k=-1} \prod_{n=k}^{k+M-1} \exp\left\{-\frac{\left|y_n - \left(\hat{\mathbf{f}}_n^{(i)}\right)^T \mathbf{a}_n\right|^2}{N_0}\right\}}. \qquad (4.102)$$

Note that LLR of the coded symbols c_k's and the modulated symbols m_k's are equivalent due to the BPSK assumption for this particular case, and we therefore do not need any bit-to-symbol level LLR conversion or vice versa. The rest of the operation is the same as in Section 3.8.1.

4.6.2 Channel Estimation Algorithms for Frequency-Selective Fading Channels

4.6.2.1 MMSE Channel Estimation

In this section, we will overview the MMSE channel estimation adopted to frequency-selective channels leaving details to Appendix B.2. In the initial estimation iteration, the estimator uses only the pilot symbols and the corresponding observations as follows

$$\hat{\mathbf{f}}^{(1)} = \mathbf{W}^{(1)} \mathbf{y}_p \qquad (4.103)$$

where $\mathbf{y}_p = \left[y_{p(1)} \ldots y_{p(L_p)}\right]^T$ is the observation vector including all available observations associated with the pilot symbols. The optimal MMSE filter is computed according to

$$\mathbf{W}^{(1)} = \mathbf{P}_p \mathbf{A}_p^H \left(\mathbf{A}_p \mathbf{R}_p \mathbf{A}_p^H + N_0 \mathbf{I}\right)^{-1} \qquad (4.104)$$

where $\mathbf{P}_p = E\left\{\mathbf{f}\mathbf{f}_p^H\right\}$ and $\mathbf{R}_p = E\left\{\mathbf{f}_p\mathbf{f}_p^H\right\}$ are the cross-correlation and the autocorrelation matrices, $\mathbf{f}_p = \left[\mathbf{f}_{p(1)} \ldots \mathbf{f}_{p(L_p)}\right]^T$ is the fading vector corresponding to the pilot indices and $\mathbf{A}_p$ is the pilot matrix given as

$$\mathbf{A}_p = \begin{bmatrix} \mathbf{a}_{p(1)}^T & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{p(2)}^T & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{a}_{p(L_p)}^T \end{bmatrix}. \tag{4.105}$$

In the subsequent iterations, the desired estimate at the i-th iteration is obtained according to

$$\hat{\mathbf{f}}^{(i)} = \mathbf{W}^{(i)}\mathbf{y} \tag{4.106}$$

where $i > 1$. For this case, the MMSE filter at the i-th iteration is given as

$$\mathbf{W}^{(i)} = \mathbf{R}_f \left(\hat{\mathbf{A}}^{(i)}\right)^H \left(\hat{\mathbf{A}}^{(i)}\mathbf{R}_f\left(\hat{\mathbf{A}}^{(i)}\right)^H + N_0\mathbf{I}\right)^{-1} \tag{4.107}$$

where $\mathbf{R}_f = E\left\{\mathbf{f}\mathbf{f}^H\right\}$ is the autocorrelation matrix and $\hat{\mathbf{A}}^{(i)}$ is the data matrix given as

$$\hat{\mathbf{A}}^{(i)} = \begin{bmatrix} \left(\hat{\mathbf{a}}_1^{(i)}\right)^T & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \left(\hat{\mathbf{a}}_2^{(i)}\right)^T & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \left(\hat{\mathbf{a}}_L^{(i)}\right)^T \end{bmatrix}. \tag{4.108}$$

4.6.2.2 Unidirectional LMS Channel Estimation

Following the result of Section 4.2.2, the conventional unidirectional LMS algorithm adopted to the iterative channel estimation problem for a time-varying frequency-selective fading channel is given for the i-th iteration as follows

$$\hat{\mathbf{f}}_{k+1}^{(i)} = \hat{\mathbf{f}}_k^{(i)} + 2\mu e_k^{(i)}\hat{\mathbf{a}}_k^{(i)} \tag{4.109}$$

where $i > 1$ so that we employ the algorithm after the initial channel estimation iteration, as before, and μ is the step-size value of the adaptation. The error term $e_k^{(i)}$ in (4.109) is given as

$$e_k^{(i)} = y_k - \left(\hat{\mathbf{f}}_k^{(i)}\right)^T \hat{\mathbf{a}}_k^{(i)}. \tag{4.110}$$

4.6.2.3 Bidirectional LMS Channel Estimation

As described in Section 4.2.2, the forward and the backward adaptations of the bidirectional LMS algorithm adopted to the iterative channel estimation problem for a time-varying frequency-selective fading channel are given for the i-th iteration as follows

$$\hat{\mathbf{f}}_{k+1}^{f,(i)} \quad = \quad \hat{\mathbf{f}}_k^{f,(i)} + 2\mu \, e_k^{f,(i)} \, \hat{\mathbf{a}}_k^{(i)} \tag{4.111}$$

$$\hat{\mathbf{f}}_{k-1}^{b,(i)} \quad = \quad \hat{\mathbf{f}}_k^{b,(i)} + 2\mu \, e_k^{b,(i)} \, \hat{\mathbf{a}}_k^{(i)} \tag{4.112}$$

where $i > 1$ as before, μ is the common step-size value and $e_k^{f,(i)}$ and $e_k^{b,(i)}$ are the associated error terms given as

$$e_k^{f,(i)} \quad = \quad y_k - \left(\hat{\mathbf{f}}_k^{f,(i)} \right)^T \hat{\mathbf{a}}_k^{(i)} \tag{4.113}$$

$$e_k^{b,(i)} \quad = \quad y_k - \left(\hat{\mathbf{f}}_k^{b,(i)} \right)^T \hat{\mathbf{a}}_k^{(i)}. \tag{4.114}$$

The final estimate $\hat{\mathbf{f}}_k^{(i)}$ is again given to be

$$\hat{\mathbf{f}}_k^{(i)} = \frac{\hat{\mathbf{f}}_k^{f,(i)} + \hat{\mathbf{f}}_k^{b,(i)}}{2}. \tag{4.115}$$

4.6.3 Numerical Results

We now explore the performances of the channel estimation algorithms presented in Section 4.6.2 over a time-varying frequency selective channel with the transmitter and receiver models introduced in Section 4.6.1. A sufficient number of Monte Carlo simulations are performed to produce statistically significant results. As usual, Rayleigh fading with Jakes' model is considered for the temporal variation of each of the channel taps which are mutually uncorrelated. We assume BPSK alphabet throughout the simulations except for some examples using QPSK symbols with some special setting.

At the transmitter, a set of $L_d = 98$ symbols are chosen from the set $\{0, 1\}$ in an independent and identical fashion. A rate $R_c = 1/2$ convolutional encoder with generator $(1, 5/7)_8$ is used together with 2 termination bits, and a number of $L_c = 200$ coded symbols are produced accordingly. After passing through a random interleaver, the coded symbols are modulated using a BPSK modulator without any loss of generality. Once the PSAM block inserts the necessary pilot symbols into the modulated stream, the final set of symbols are transmitted

through the time-varying frequency-selective fading channel under consideration. The pilot symbol spacing is chosen to be $M_p = 11$ throughout the simulations which results in a number of $ML_p = 2 \times 22 = 44$ pilot symbols for each of the transmitted block of $L = 244$ symbols. The pilot overhead percentage is therefore 0.1803 and the overall transmission rate, i.e., $R = L_d/L$, becomes 0.4016 with these settings.

At the receiver, the feedback of the estimates of the coded symbols is in the form of soft-decisions to achieve a better error performance. The number of channel estimation iterations is chosen to be 3 beyond which all the algorithms are observed to saturate. In addition, both the unidirectional and bidirectional LMS algorithms use optimal step-size values which are determined in trial and error basis for each SNR level. The details of the receiver is explained in Section 4.6.1.
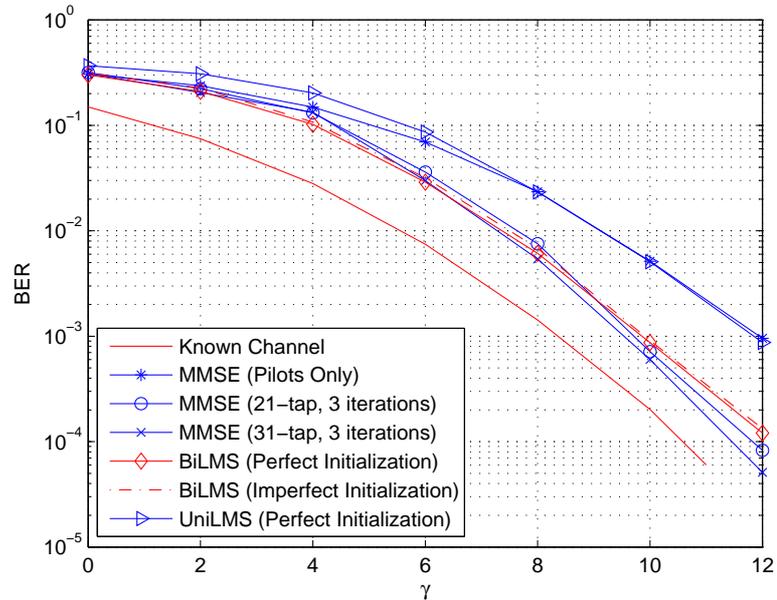
Figure 4.20: BER for BiLMS, UniLMS and MMSE with $M_p = 11$ over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. All results for BiLMS and UniLMS are associated with the 3-rd estimation iteration.

We evaluate the performance of the estimation algorithms under consideration through BER and normalized MSIE statistics with various Doppler frequency choices. Fig. 4.20 presents the BER results for the bidirectional LMS algorithm, unidirectional LMS algorithm and MMSE filter with 21 and 31 taps after the 3-rd channel estimation iteration. The BER performance associated with the initial channel estimate for which MMSE filter is used together

with all the available pilot symbols is also added to provide understanding the power of the iterative channel estimation idea. We make a number of observations for this particular case as follows. First of all, the BER performance of the bidirectional LMS algorithm is very close to that of the MMSE filter after 3 channel estimation iterations. We observe that the BER performance of the optimal MMSE filter employing only the pilot symbols is off the known channel bound by almost 4 dB at BER= 10^{-3}. The iterative estimation of the unknown channel enables the bidirectional LMS algorithm to fill this gap by 2 dB so that the associated BER performance is only 2 dB off the known channel case.
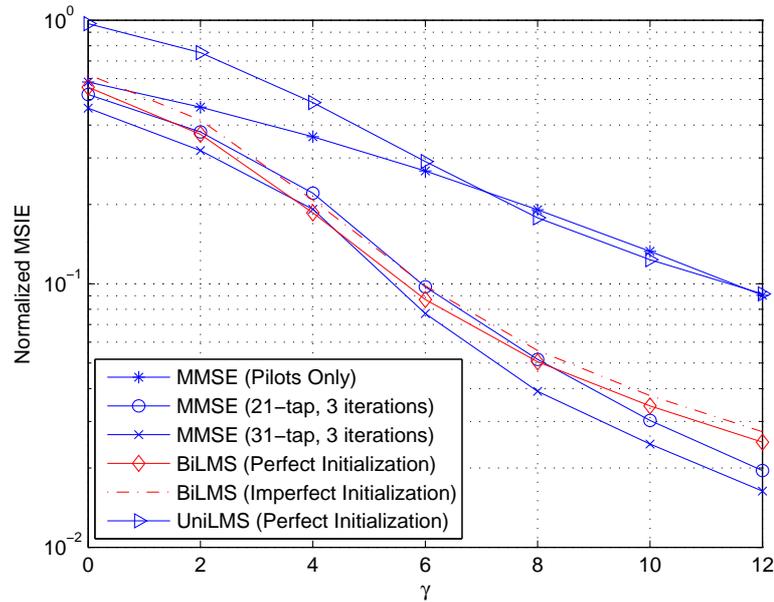
Figure 4.21: Normalized MSIE for BiLMS, UniLMS and MMSE with $M_p = 11$ over a 2-tap Rayleigh fading ISI channel with $f_dT_s = 0.01$. All results for BiLMS and UniLMS are associated with the 3-rd estimation iteration.

The effect of imperfect initialization on the bidirectional LMS algorithm is also investigated for practical purposes by employing estimates of the channel coefficients from the previous estimation iteration for the initialization instead of the actual values. The associated results in Fig. 4.20 show that no degradation occurs in the performance of the bidirectional LMS algorithm as a result of imperfect initialization. Finally, we observe that the unidirectional LMS algorithm proposes no further improvement over the initial channel estimation.

In Fig. 4.21, we present the performance of the algorithms under consideration through the normalized MSIE results. We first determine the statistics associated with the optimal MMSE

filtering which uses all the available pilots only as a reference level, and observe that the bidirectional LMS algorithm with both perfect and imperfect initialization together with the MMSE filter lay below the reference level for all SNR choices which implies improvement in channel estimation quality through iterations unlike the case for unidirectional LMS algorithm.
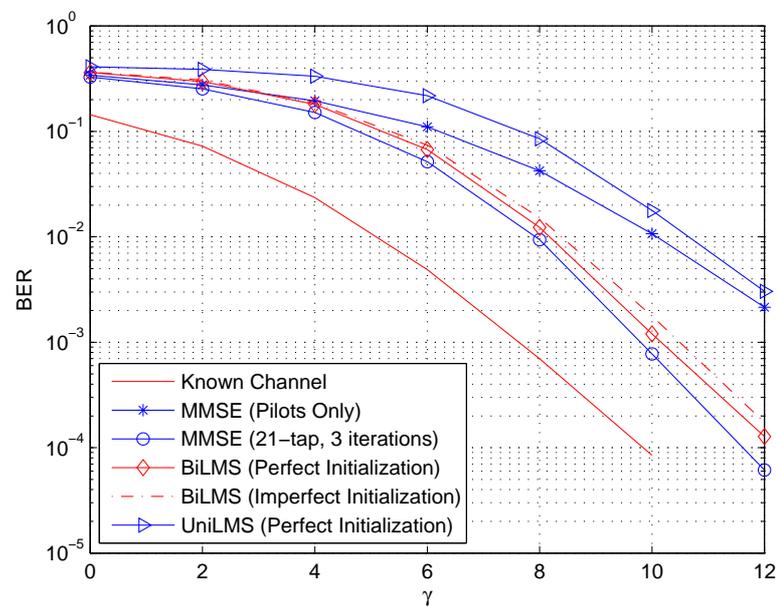
Figure 4.22: BER for BiLMS, UniLMS and MMSE with $M_p = 11$ over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.02$. All results for BiLMS and UniLMS are associated with the 3-rd estimation iteration.
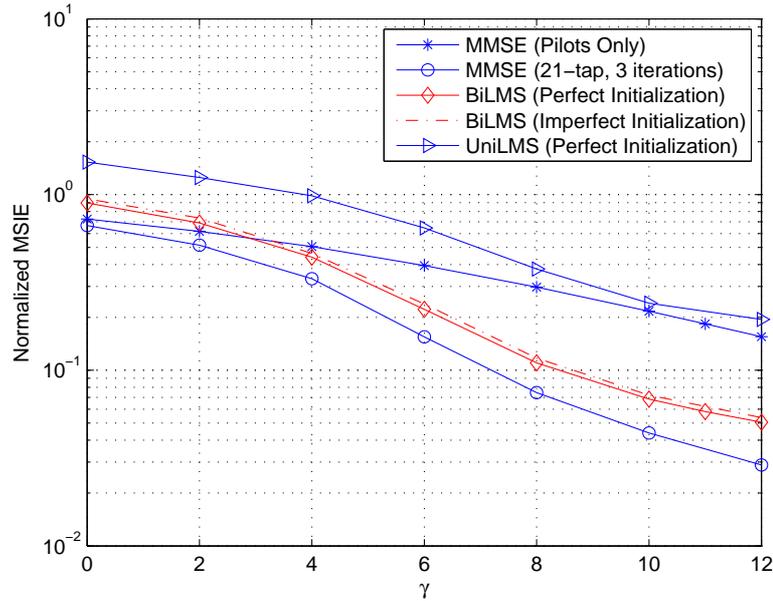
Figure 4.23: Normalized MSIE for BiLMS, UniLMS and MMSE with $M_p = 11$ over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.02$. All results for BiLMS and UniLMS are associated with the 3-rd estimation iteration.

The results for BER and MSE statistics for a much faster Rayleigh fading channel with $f_d T_s = 0.02$ are presented in Fig. 4.22- 4.23. We observe that the 21-tap MMSE filter achieves a BER performance which is off the known channel bound by 2 dB after 3 channel estimation iterations. The BER performance of bidirectional LMS algorithm is somewhat worse than the 21-tap MMSE especially at a high SNR regime for this challenging fast time-varying scenario, but nevertheless is better than that of the MMSE filter using the pilots only, i.e., initial channel estimation, by more than 2 dB. The unidirectional LMS algorithm performs so bad that the corresponding BER performance with the soft estimates of the coded symbols is off even the pilots only case. These results could also be observed in Fig. 4.23.

In Fig. 4.24-4.25, we also present the results for QPSK modulation with the constellation given in Fig. 3.43. Keeping all the other system parameters the same, we assume a 2-tap Rayleigh fading channel with Jakes' spectrum and $fdTs = 0.01$. We observe that the performance of MMSE filter with 11-tap is worse than the bidirectional LMS algorithm and one should use a 21-tap filter to fill this especially at high SNR region. We also observe that the unidirectional LMS algorithm with the soft estimates now always performs worse than the pilots only case, unlike the results for BPSK.
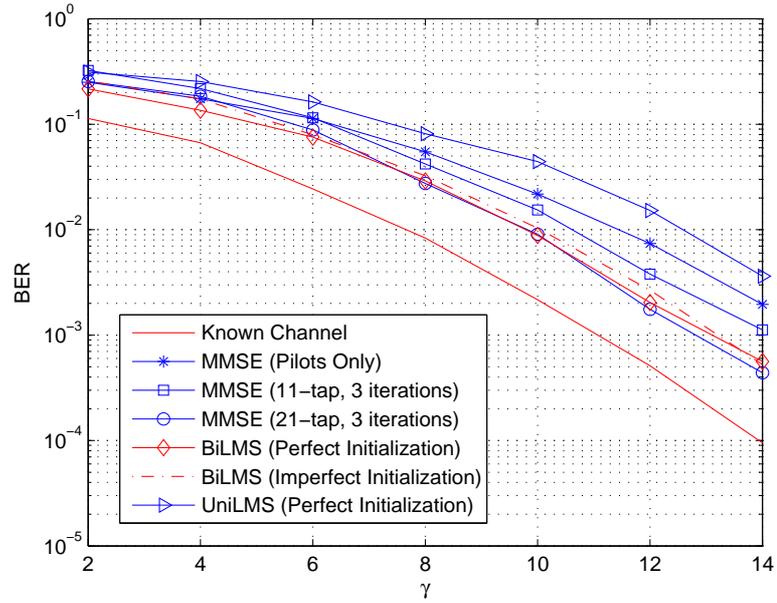
Figure 4.24: BER for BiLMS, UniLMS and MMSE for QPSK modulation with $M_p = 11$ over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. All results for BiLMS and UniLMS are associated with the 3-rd estimation iteration.
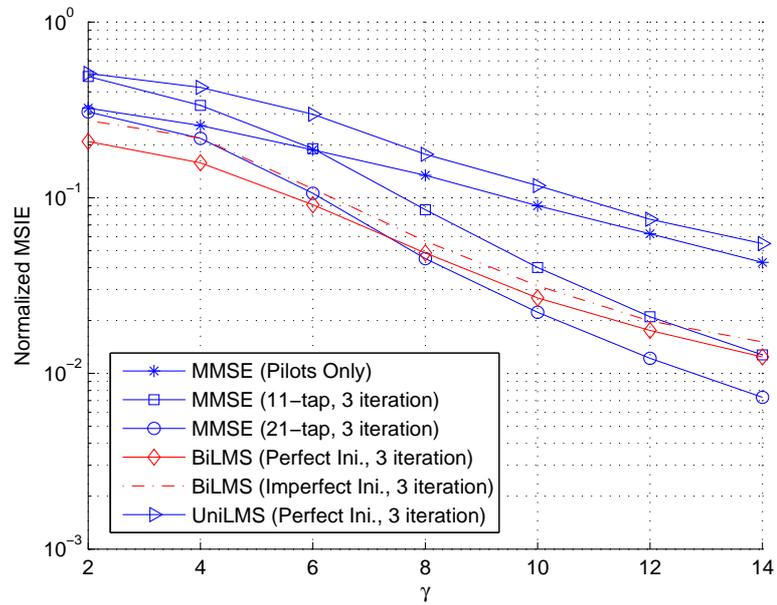
Figure 4.25: Normalized MSIE for BiLMS, UniLMS and MMSE for QPSK modulation with $M_p = 11$ over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. All results for BiLMS and UniLMS are associated with the 3-rd estimation iteration.

4.7   Conclusion

Estimation of time-varying frequency-selective channels are considered in this chapter which is indeed a generalization of the flat-fading case. Because the channel to be estimated is represented by a vector at each time epoch, the estimation algorithms with sufficiently low complexity is of significant importance. The bidirectional LMS algorithm is therefore revisited and investigated over frequency-selective channels both experimentally and analytically. The results for both the coded and the uncoded communication systems indicate the superiority of the bidirectional LMS algorithm both in terms of its good tracking performance as compared to the optimal Wiener filter and its low-complexity structure similar to the unidirectional LMS algorithm.

CHAPTER 5

THE BIDIRECTIONAL LMS ALGORITHM FOR MULTI-INPUT MULTI-OUTPUT CHANNELS

MIMO technology where both the transmitter and the receiver is equipped with multiple antennas has attracted much attention for over a decade with its performance-enhancing capabilities. The MIMO technology offers to use the spatial dimension properly to achieve the diversity and/or the multiplexing gain and mitigate the adverse effects of fading. Very high data rates are possible through the use of multiple antennas at both sides of the communication link. As such, the MIMO technology is believed to be one of the key ingredients of the next generation wireless standards [120, 121].

In this chapter, we consider the estimation and tracking of time-varying MIMO communication channels particularly by using the bidirectional LMS algorithm. Indeed, this chapter is a generalization of the results in Chapter 3 and Chapter 4. The considered channel model is flat-fading so that each of the subchannels to be estimated is a scalar random variable as in the case for Chapter 3. Besides, since the symbols transmitted from different antennas arrive on top of each other at a receiver antenna, the associated subchannels should be estimated jointly which results in a vector operation as in the case studied in Chapter 4.

5.1   System Model for Flat-Fading MIMO Channels

We consider a flat-fading MIMO communication channel which is represented at a time epoch k by a matrix $\mathbf{H}_k$ of size $N \times M$ where M and N are the number of transmitter and the receiver antennas, respectively. Because we are dealing with the channel estimation problem, we assume that the channel matrix $\mathbf{H}_k$ is known neither at the transmitter nor at the receiver.

The (m, n)-th element of the channel matrix $\mathbf{H}_k$ is indicated by $h_{nm,k}$ which represents the subchannel between the m-th transmitter and the n-th receiver antennas, and is assumed to have a unit energy. The subchannels are assumed to have flat-fading characteristics with the same statistics. It is also assumed that there is no correlation between the subchannels so that we have the following expression

$$E\left\{ h_{nm,k}\, h^*_{n'm',l} \right\} \;=\; \delta_{nn'}\, \delta_{mm'}\, r(k-l) \tag{5.1}$$

where $r(\cdot)$ is the autocorrelation function of the underlying fading channel.

An information symbol $a_{m,k}$ with a symbol energy $E_s = E\{|a_{m,k}|^2\}$ is chosen independently from a finite alphabet $\mathcal{A}$ as before, and is sent through the transmitter antenna m. The corresponding discrete-time equivalent complex baseband channel model is then given at a time epoch k as follows

$$y_{n,k} \;=\; \sum_{m=1}^{M} h_{nm,k}\, a_{m,k} \;+\; n_{n,k}, \tag{5.2}$$

$$\;=\; \mathbf{h}^T_{n,k}\, \mathbf{a}_k \;+\; n_{n,k} \tag{5.3}$$

for $n = 1, 2, \ldots, N$, where $y_{n,k}$ is the observation symbol at the n-th receiver antenna, $\mathbf{a}_k = [a_{1,k}\, a_{2,k}\, \ldots\, a_{M,k}]^T$ is the transmitted symbol vector, $\mathbf{h}^T_{n,k} = [h_{n1,k}\, h_{n2,k}\, \ldots\, h_{nM,k}]$ is the n-th row of $\mathbf{H}_k$, and $n_{n,k}$ is a sample from a circularly symmetric white complex Gaussian process with zero-mean and variance N_0. As could be inferred from (5.2)-(5.3), we assume perfect information on the block and symbol timings as well as the carrier frequency. The channel relation could also be expressed in a more compact form as in

$$\mathbf{y}_k \;=\; \mathbf{H}_k\, \mathbf{a}_k \;+\; \mathbf{n}_k \tag{5.4}$$

where $\mathbf{y}_k = [y_{1,k}\, y_{2,k}\, \ldots\, y_{N,k}]^T$ and $\mathbf{n}_k = [n_{1,k}\, n_{2,k}\, \ldots\, n_{N,k}]^T$.

The received SNR for each receiver antenna is then given as

$$\gamma_r \;=\; \frac{E\left\{ \left| \sum_{m=1}^{M} h_{nm,k}\, a_{m,k} \right|^2 \right\}}{E\{|n_{n,k}|^2\}} \;=\; \frac{\sum_{m=1}^{M} \sum_{m'=1}^{M} \overbrace{E\left\{ h_{nm,k}\, h^*_{nm',k} \right\}}^{\delta_{mm'}\, r(0)}\; \overbrace{E\left\{ a_{m,k}\, a^*_{m',k} \right\}}^{E_s\, \delta_{ll'}}}{N_0} \tag{5.5}$$

$$\;=\; \frac{E_s \sum_{m=1}^{M} E\left\{ |h_{nm,k}|^2 \right\}}{N_0} \;=\; \frac{M\, E_s}{N_0}, \tag{5.6}$$

where (5.6) makes use of the result of (5.1), and the fact that the set $\{a_{m,k}\}_{m=1}^{M}$ is assumed to be uncorrelated. The information symbol SNR is accordingly given as $\gamma = \gamma_r/R$ where R is the overall transmission rate of the communication system.

5.2 Estimation Algorithms for Flat-Fading MIMO Channels

In this section, the channel estimation algorithms under consideration is revisited to be modified according to the needs for the flat-fading MIMO channel model given in Section 5.1.

5.2.1 The MMSE Channel Estimation

In order to estimate a complex subchannel coefficient $h_{nm,k}$, one should consider many and possibly all correlated values of this fading coefficient during a transmission block. As a result, the set $\{h_{nm,k}\}_{k=1}^{L}$ should be estimated jointly. Furthermore, since each observation at a single receiver antenna has a partial information on all subchannels terminating at that antenna according to (5.2), the set $\{h_{nm,k}\}_{k=1}^{L}$, $m = 1, \ldots, M$, should be estimated jointly for optimal operation. As a result, the optimal Wiener filter should be derived to estimate the set of vectors $\{\mathbf{h}_{n,k}\}_{k=1}^{L}$, or equivalently $\mathbf{h}_n = \left[\mathbf{h}_{n,1}^T \ldots \mathbf{h}_{n,L}^T\right]^T$, jointly which may also be inferred from (5.2).

The corresponding MMSE estimator is then given as follows

$$\hat{\mathbf{h}}_n = \mathbf{W}_n\,\mathbf{y}_n \tag{5.7}$$

for $n = 1, 2, \ldots, N$, where $\mathbf{W}_n$ is the estimation filter to be optimized and $\mathbf{y}_n = [y_{n,1} \ldots y_{n,L}]^T$ is the concatenated observation vector associated with the n-th receiver antenna.

We also note that (5.7) could be modified such that overall estimator is derived for a single subchannel vector $\hat{\mathbf{h}}_{n,k}$ instead of the complete set of unknown fading vectors $\hat{\mathbf{h}}_n$, and/or less number of observation vectors could be employed instead of all the available observation vectors $\mathbf{y}_n$. These two choices will hopefully reduce the associated computational complexity.

The optimal MMSE filter is computed through the Wiener-Hopf's equations which are obtained by following the similar steps explained in Appendix B.1 and are given as

$$\mathbf{W}_n = \mathbf{P}_{hy,n}\,\mathbf{R}_{yy,n}^{-1} \tag{5.8}$$

144

where $\mathbf{P}_{hy,n} = E\left\{\mathbf{h}_n \mathbf{y}_n^H\right\}$ and $\mathbf{R}_{yy,n} = E\left\{\mathbf{y}_n \mathbf{y}_n^H\right\}$ are the cross-correlation and the autocorrelation matrices, respectively. Indeed, this is the same result presented in [122] which is obtained by deriving the maximum *a posteriori* (MAP) estimate of the unknown channel vector.

Assuming K observation symbols to be employed in estimating $\mathbf{h}_{n,k}$, the pre-filtering stage given in (5.8) requires a matrix inversion of complexity $O(K^3)$ and a matrix multiplication of complexity $O(MK^2)$ while the filtering stage in (5.7) needs $M(K-1)$ complex additions and MK complex multiplications. In order to estimate the complete channel matrix $\mathbf{H}_k$, the required computations are repeated N times.

5.2.2 The Unidirectional and The Bidirectional LMS Algorithms

The conventional LMS algorithm over a flat-fading MIMO channels is given as [57]

$$\hat{\mathbf{h}}_{n,k+1} = \hat{\mathbf{h}}_{n,k} + 2\mu\, e_{n,k}\, \mathbf{a}_k \qquad (5.9)$$

for $n = 1, 2, \ldots, N$, where μ is the associated step-size and $e_{n,k}$ is the estimation error associated with the n-th receiver antenna defined as

$$e_{n,k} = y_{n,k} - \hat{\mathbf{h}}_{n,k}^T \mathbf{a}_k. \qquad (5.10)$$

The conventional LMS algorithm just needs $N(M+1)$ complex additions and $2N(M+1)$ complex multiplications in order to estimate the channel matrix.

Let us define $\hat{h}_{nm,k}^f$ and $\hat{h}_{nm,k}^b$ to be the estimates of the fading coefficient $h_{nm,k}$ in the forward and the backward directions, respectively, which are represented in vector form as $\hat{\mathbf{h}}_{n,k}^f = \left[\hat{h}_{n1,k}^f\ \hat{h}_{n2,k}^f\ \ldots\ \hat{h}_{nM,k}^f\right]^T$ and $\hat{\mathbf{h}}_{n,k}^b = \left[\hat{h}_{n1,k}^b\ \hat{h}_{n2,k}^b\ \ldots\ \hat{h}_{nM,k}^b\right]^T$. The bidirectional LMS algorithm is then given as

$$\hat{\mathbf{h}}_{n,k+1}^f = \hat{\mathbf{h}}_{n,k}^f + 2\mu\, e_{n,k}^f\, \mathbf{a}_k \qquad (5.11)$$

$$\hat{\mathbf{h}}_{n,k-1}^b = \hat{\mathbf{h}}_{n,k}^b + 2\mu\, e_{n,k}^b\, \mathbf{a}_k \qquad (5.12)$$

for $n = 1, 2, \ldots, N$, where μ is the associated step-size of the adaptations, $e_{n,k}^f$ and $e_{n,k}^b$ are the forward and the backward errors associated with the n-th receiver antenna which is defined

as

$$e_{n,k}^{f} = y_{n,k} - (\hat{\mathbf{h}}_{n,k}^{f})^{T} \mathbf{a}_{k} \qquad (5.13)$$

$$e_{n,k}^{b} = y_{n,k} - (\hat{\mathbf{h}}_{n,k}^{b})^{T} \mathbf{a}_{k}. \qquad (5.14)$$

The final estimate is again chosen to be the arithmetic average of the forward and the backward estimates as follows

$$\hat{\mathbf{h}}_{n,k} = \frac{\hat{\mathbf{h}}_{n,k}^{f} + \hat{\mathbf{h}}_{n,k}^{b}}{2}, \qquad (5.15)$$

for $n = 1, 2, \ldots, N$.

The bidirectional LMS algorithm with this setting requires $N(3M + 2)$ complex additions and $N(5M + 4)$ complex multiplications. As a result, the overall complexity of the bidirectional LMS algorithm is close to the conventional unidirectional LMS algorithm as compared to that for the optimal MMSE filter.

5.3 Tracking Performance of Bidirectional LMS over MIMO Flat-Fading Channels

In this section, we evaluate the tracking performance of the bidirectional LMS algorithm for a flat-fading MIMO channel. As we discussed in the previous section, the update equations of the bidirectional LMS algorithm for a flat-fading MIMO channel is very similar to that for the single antenna frequency-selective fading channels considered in Chapter 4. We therefore benefit from the results of [61, 63] and Section 4.3 as much as possible, and remove some similar intermediate derivation steps.

We again concentrate only on the tracking quality of the bidirectional LMS algorithm at the steady-state and assume that the overall system is operating in the training mode where $\{a_{m,k}\}_{m=1}^{M}$ are known a priori at the receiver. The corresponding error performance surface, or

equivalently the MSE, is given for a flat-fading MIMO channel as follows

$$J_{MSE,k} = E\left\{\|\mathbf{e}_k\|^2\right\} \tag{5.16}$$

$$= E\left\{\|\mathbf{y}_k - \hat{\mathbf{H}}_k\,\mathbf{a}_k\|^2\right\} \tag{5.17}$$

$$= E\left\{\|(\mathbf{H}_k - \hat{\mathbf{H}}_k)^T\,\mathbf{a}_k + \mathbf{n}_k\|^2\right\} \tag{5.18}$$

$$= E\left\{\|\mathbf{n}_k\|^2\right\} + E\left\{\|(\mathbf{H}_k - \hat{\mathbf{H}}_k)^T\,\mathbf{a}_k\|^2\right\} \tag{5.19}$$

$$= \underbrace{\sum_{n=1}^{N} E\left\{|n_{n,k}|^2\right\}}_{N\,J_{min}} + \underbrace{E_s \sum_{n=1}^{N} \|\mathbf{h}_{n,k} - \hat{\mathbf{h}}_{n,k}\|^2}_{J_{ex,k}} \tag{5.20}$$

where $\mathbf{e}_k$ is the overall tracking error, $J_{ex,k}$ is the excess MSE, and the minimum achievable MSE due to the presence of additive noise is qual to $N\,J_{min}$ with $J_{min} = N_0$, as before. We may express the average MSE at the steady-state as follows

$$J_{MSE} = E\{J_{MSE,k}\} \tag{5.21}$$

$$= N\,J_{min} + E_s \sum_{n=1}^{N} \underbrace{E\left\{\|\mathbf{h}_{n,k} - \hat{\mathbf{h}}_{n,k}\|^2\right\}}_{J_{MSIE,n}} \tag{5.22}$$

where $J_{MSIE,n}$ is the MSIE associated with the estimation of the n-th column of the channel matrix $\mathbf{H}_k$. Due to the statistical symmetry, $J_{MSIE,n}$ is common for any choice of $n = 1, 2, \ldots, N$. As a result, the average MSIE becomes

$$J_{MSE} = N\,J_{min} + E_s \sum_{n=1}^{N} J_{MSIE,n} \tag{5.23}$$

$$= N\,J_{min} + E_s\,N\,J_{MSIE} \tag{5.24}$$

where J_{MSIE} is redefined here to be the MSIE in estimating any columns of $\mathbf{H}_k$.

We note that, $J_{MSIE} = E\left\{\|\mathbf{h}_{n,k} - \hat{\mathbf{h}}_{n,k}\|^2\right\}$ is derived in Section 4.3 for frequency-selective channels which is exactly the same for our case. Therefore, using the results of Section 4.3, J_{MSIE} for a flat-fading $N \times M$ MIMO channel is given as

$$J_{MSIE} = \frac{\mu M^2 N E_s}{2\left(E_s - \mu\left[(M-1)E_s + E_4\right]\right)} J_{min} + \frac{MN}{2\pi} \int_{-\pi}^{\pi} |H(e^{jw})|^2\, S(w)\,dw, \tag{5.25}$$

and the steady-state MSE is found to be

$$J_{MSE} = \left(1 + \frac{\mu M^2 N E_s^2}{2\left(E_s - \mu\left[(M-1)E_s + E_4\right]\right)}\right) J_{min} + \frac{MNE_s}{2\pi} \int_{-\pi}^{\pi} |H(e^{jw})|^2\, S(w)\,dw \tag{5.26}$$

where $S(w)$ is the power spectral density of the underlying fading model, and the transfer function $H(e^{jw})$ is given as

$$H(z) = -\frac{1+\beta}{2\beta} + \frac{1-\beta}{2\beta}\left(\frac{1}{1-\beta e^{-jw}} - \frac{1}{1-\frac{1}{\beta}e^{-jw}}\right). \tag{5.27}$$

Note that since the correlation property of the input vectors $\mathbf{a}_k$ are different for the frequency-selective and MIMO channels, the above expressions are treated to be approximate. Nevertheless, the associated results given in the next section is observed to be satisfactory. This consideration may be further investigated as a subject of a future work.

The optimal choice for the step-size value, μ_{opt}, is again of interest to characterize the minimum achievable MSE at the steady-state. In order to derive μ_{opt} theoretically, we express (5.26) in terms of only β and take derivative of the resulting expression with respect to β as follows

$$\frac{\partial J_{MSE}}{\partial \beta} = \frac{\partial}{\partial \beta}\left\{\left(1 + \frac{(1-\beta)M^2NE_s^2}{2\left(2E_s^2 - (1-\beta)\left[(M-1)E_s + E_4\right]\right)}\right)J_{min} + \frac{E_s\,MN}{2\pi}\int_{-\pi}^{\pi}\left|H(e^{jw})\right|^2 S(w)\,dw\right\}$$

$$= -\frac{E_s^4 M^2 N}{\left(2E_s^2 - (1-\beta)\left[(M-1)E_s + E_4\right]\right)^2}\,J_{min} + \frac{E_s\,MN}{\pi}\int_{-\pi}^{\pi}H(e^{jw})\frac{\partial H(e^{jw})}{\partial \beta}S(w)\,dw \tag{5.28}$$

where $\partial H(e^{jw})/\partial \beta$ is the same as before which is given as

$$\frac{\partial H(e^{jw})}{\partial \beta} = -\frac{(1-\cos w)(1-\beta^2 - 2\beta + 2\cos w)}{(1+\beta^2 - 2\beta\cos w)^2}. \tag{5.29}$$

The optimal geometric ratio β_{opt} could then be evaluated numerically using (5.28) and (5.29) as follows

$$\left.\frac{\partial J_{MSE}}{\partial \beta}\right|_{\beta=\beta_{opt}} = 0 \tag{5.30}$$

and the optimal step-size μ_{opt} could be found as $\mu_{opt} = (1 - \beta_{opt})/2E_s$.

5.3.1   Numerical Results

In this section, we investigate the accuracy of the MSE derivation for the bidirectional LMS algorithm operating at the steady-state over a time-varying flat-fading MIMO channel. We

compare the normalized MSIE results for the bidirectional LMS algorithm, unidirectional LMS algorithm and MMSE filter which are obtained through Monte Carlo simulations together with the theoretical normalized MSIE computed according to (5.25).

As a brief overview, we assume a time-varying MIMO system equipped with $M = 2$ transmitter and $N = 4$ receiver antennas where each subchannel between a transmitter and a receiver antenna pair experience a Rayleigh fading with Jakes' model and $f_d T_s = 0.01$ in a spatially uncorrelated fashion. In each Monte Carlo run, a set of $L = 100$ information symbols are chosen independently from the BPSK alphabet $\mathcal{A} = \{-1, +1\}$ with the symbol energy $E_s = 1$. In this case, the normalization is performed on MSIE with respect to number of the both the transmitters and receivers as follows

$$\overline{J_{MSIE}} = \frac{J_{MSIE}}{MN}. \tag{5.31}$$
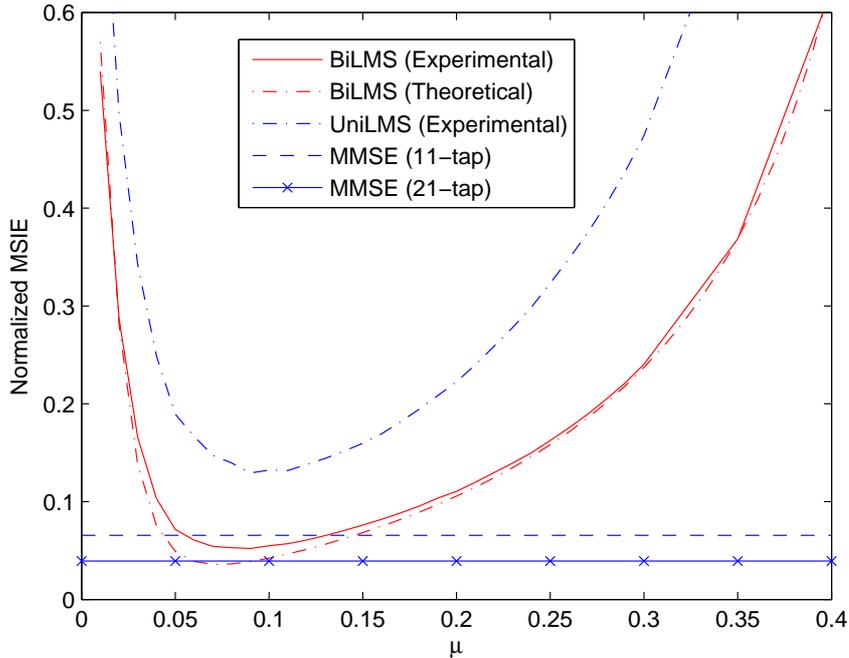
Figure 5.1: Theoretical and experimental normalized MSIE for BiLMS for varying step-size at SNR=5 dB over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$. The experimental MSIE for UniLMS together with a 11-tap and 21-tap optimal MMSE filters are also provided.

In the figures in this subsection, we present the experimental normalized MSIE for the bidirectional LMS algorithm, the unidirectional LMS algorithm and MMSE filter with various

number of taps together with the theoretical normalized MSIE for the bidirectional LMS algorithm. In Fig. 5.1 and Fig. 5.2, these MSIE performances are depicted for varying step-size at $\gamma = 5$ dB and $\gamma = 10$ dB SNR, respectively. We observe that the theoretical results for the bidirectional LMS algorithm have a good match to the experimental ones except in the neighborhood of the associated minimum points for both $\gamma = 5$ dB and $\gamma = 10$ dB cases. Indeed, such a conclusion may be expected since there are $2 \times 4 = 8$ subchannels to be estimated which naturally amplifies the small deviations of the theoretical results from the experimental ones as compared to a single channel estimation problem in single-antenna flat-fading case. We also observe that the normalized MSIE of the bidirectional LMS algorithm is better than that of the unidirectional LMS algorithm and is very close to the MMSE filter. The associated result for the 11-tap MMSE filter is worse than that of the bidirectional LMS algorithm at $\gamma = 5$ dB and better than it $\gamma = 10$ dB while the 21-tap MMSE filter has a superiority over the bidirectional LMS algorithm for both of the SNR choices.
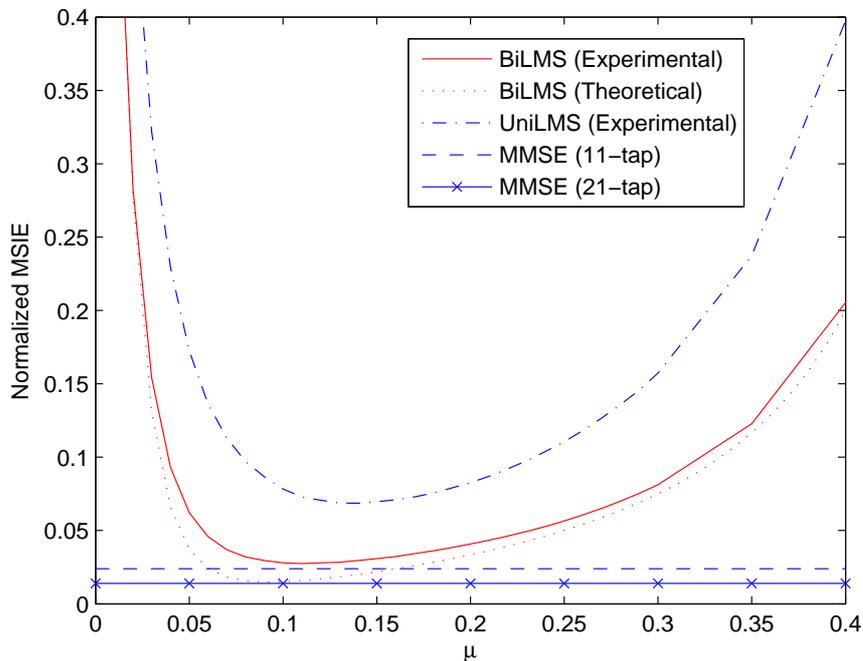
Figure 5.2: Theoretical and experimental normalized MSIE for BiLMS for varying step-size at SNR=10 dB over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$. The experimental MSIE for UniLMS together with a 11-tap and 21-tap optimal MMSE filters are also provided.
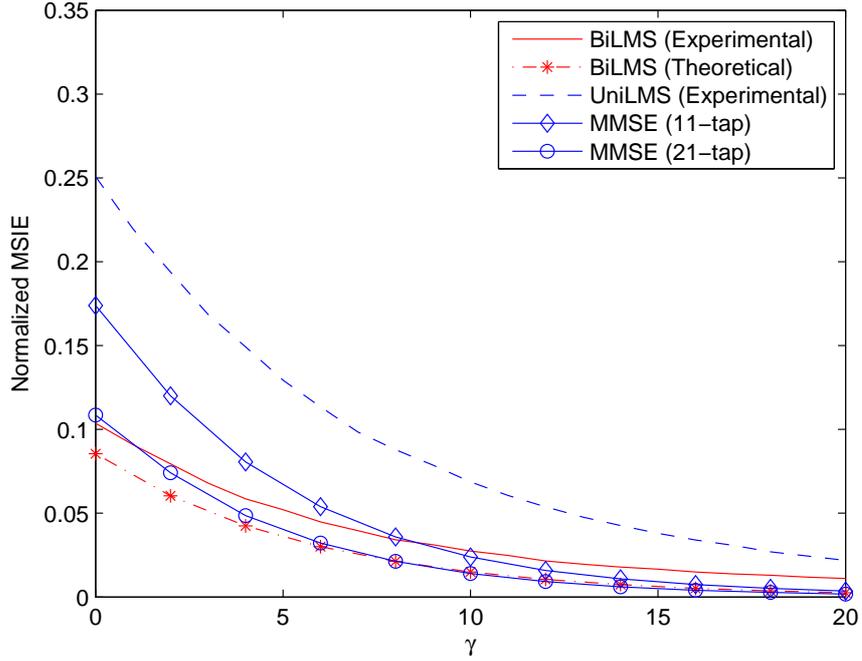
Figure 5.3: Theoretical and experimental normalized MSIE for BiLMS for optimal step-size and varying SNR over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$. The experimental MSIE for UniLMS together with a 11-tap and 21-tap optimal MMSE filters are also provided.

Similar conclusions could also be made from Fig. 5.3 where the same statistics are now depicted for varying SNR and optimal step-size choices for each SNR level. In Fig. 5.3, we observe that the 11-tap MMSE filter starts achieving a better MSIE performance than the bidirectional LMS algorithm after an SNR threshold of approximately $\gamma = 8$ dB. In Fig. 5.4, we also provide a comparison between normalized MSIE values for the MMSE filters with various number of taps in order to complete the picture.

We now investigate the effectiveness of the theoretical optimal steps-size values, i.e., μ_{opt}'s, computed according to (4.67) and (4.68). In Table 5.1, the resulting theoretical values for μ_{opt} together with the associated experimental ones are presented for a 2×4 MIMO Rayleigh fading channel with $f_d T_s = 0.01$. We observe from these results that the theoretical μ_{opt} values are very close to the experimental results for small SNR values while a deviation occurs for high SNR values which is obviously due to the deviations of the theoretical results from the experimental ones explained before in this section.
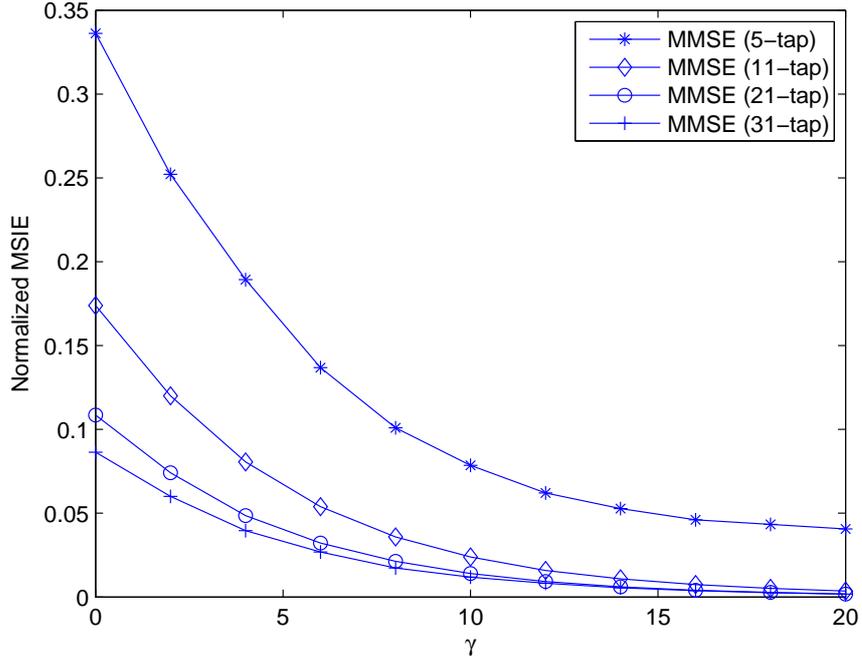
Figure 5.4: Experimental normalized MSIE for MMSE with various number of taps over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$.

Table 5.1: Theoretical and Experimental Optimal Step-Size (μ_{opt}) Values for a 2×4 MIMO Channel with $f_d T_s = 0.01$

SNR	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB	12 dB	14 dB	16 dB
Experimental	0.060	0.070	0.080	0.090	0.100	0.110	0.120	0.140	0.160
Theoretical	0.056	0.062	0.069	0.076	0.084	0.092	0.100	0.109	0.119

5.4 Iterative Channel Estimation for Flat-Fading MIMO Channels

5.4.1 Transmitter and Receiver Models

In this section, the equivalent discrete-time complex baseband channel model given in (5.2) is employed for the estimation of time-varying flat-fading MIMO channels with the transmitter block in Fig. 5.5. In this model, a set of binary information symbols $\{b_k\}_{k=1}^{L_d}$ are first generated randomly in an independent fashion by using the binary set $\{0, 1\}$, and are encoded by a channel code of rate R_c. The set of coded symbols $\{c_k\}_{k=1}^{L_c}$ are then interleaved to combat with the burst errors and the interleaved set $\{c_k'\}_{k=1}^{L_c}$ is produced where $L_c = \frac{L_d}{R_c}$. The interleaved coded symbols are then pushed to the serial-to-parallel (S/P) block such that consecutive symbols
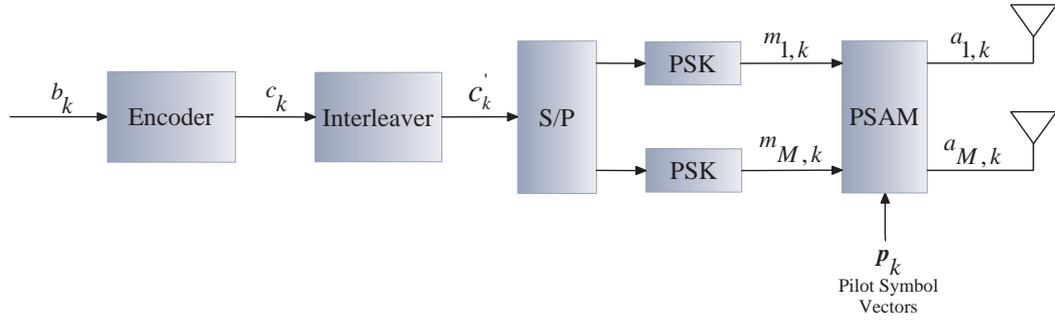
Figure 5.5: Transmitter model including channel encoder, interleaver, multiplexer, a set of PSK modulators and PSAM blocks.

are mapped to different transmitter antennas sequentially as shown in Fig. 5.6. The set of coded symbols associated with each transmitter are then modulated using a PSK modulator, as before, with a finite modulation alphabet $\mathcal{A}$. At each transmitter antenna branch, the associated set of modulated symbols $\{m_{m,k}\}_{k=1}^{L_m}$ with $L_m = \frac{L_d}{M R_c \log_2 |\mathcal{A}|}$ are multiplexed with a set of a priori known pilot symbols $\{p_{m,k}\}_{k=1}^{L_p}$ chosen from the same alphabet $\mathcal{A}$. The resulting set of symbols $\{a_{m,k}\}_{k=1}^{L}$ are then transmitted through the time-varying flat-fading subchannels according to (5.2).
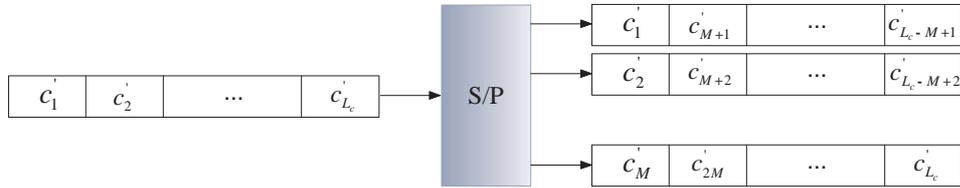
Figure 5.6: The serial-to-parallel (S/P) conversion at MIMO transmitter.

Note that the PSAM mechanism for this particular case operates as in Section 3.8.1 such that the modulated symbols at each transmitter branch are first split into the group of $M_p - 1$ symbols and a single pilot symbol is then inserted into the center of each of these groups. Because the pilot symbols are used at the same locations for all the transmitter branches, they are represented by vectors as $\{\mathbf{p}_k\}_{k=1}^{L_p}$ prior to multiplexing where $\mathbf{p}_k = [p_{1,k} \, p_{2,k} \, \ldots \, p_{M,k}]^T$. This transmission scheme implies that the transmitted symbol vector $\mathbf{a}_k$ for a given time epoch k is either completely composed of the unknown data symbols or the a priori known pilot symbols.

The necessary number of pilot symbols are determined to be $M L_p$ for this multi-antenna

153

scheme where L_p is given using the result of (3.180) as

$$L_p = \left\lfloor \frac{L_m - \frac{M_p-1}{2}}{M_p - 1} \right\rfloor + 1 = \left\lfloor \frac{\frac{L_d}{M R_c \log_2 |\mathcal{A}|} - \frac{M_p-1}{2}}{M_p - 1} \right\rfloor + 1, \tag{5.32}$$

and the associated frame length L is found to be

$$L = \frac{L_d}{M R_c \log_2 |\mathcal{A}|} + L_p = \frac{L_d}{M R_c \log_2 |\mathcal{A}|} + \left\lfloor \frac{\frac{L_d}{M R_c \log_2 |\mathcal{A}|} - \frac{M_p-1}{2}}{M_p - 1} \right\rfloor + 1. \tag{5.33}$$

Consequently, the final sequence of vectors $\{\mathbf{a}_k\}_{k=1}^L$ to be transmitted is expressed as

$$\mathbf{a}_k = \begin{cases} \mathbf{m}_{k - \left\lfloor \frac{k-(M_p+1)/2}{M_p} \right\rfloor}, & k \neq (i-1) M_p + \frac{M_p+1}{2}, \ i = 1, 2, \ldots, L_p; \\ \mathbf{p}_{\frac{k-(M_p+1)/2}{M_p}}, & k = (i-1) M_p + \frac{M_p+1}{2}, \ i = 1, 2, \ldots, L_p. \end{cases} \tag{5.34}$$

The set of indices associated with the pilot symbol vectors in a transmitted block are the same as (3.183) which is given by

$$P_p = \left\{ k_p(i) \right\}_{i=1}^{L_p} = \left\{ \frac{M_p + 1}{2}, \frac{3M_p + 1}{2}, \ldots, \frac{(2L_p - 1)M_p + 1}{2} \right\} \tag{5.35}$$

where $k_p(\cdot)$ specifies the indices of the pilot symbol vectors defined as

$$k_p(i) = (i-1) M_p + \frac{M_p + 1}{2}. \tag{5.36}$$
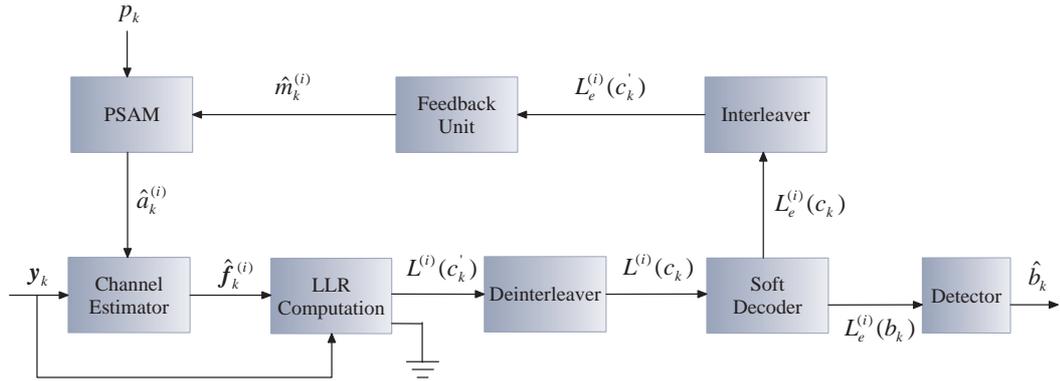
Figure 5.7: Receiver model for a time-varying flat-fading MIMO system with iterative channel estimation.

The receiver structure under consideration is given in Fig. 5.7. Leaving some operational details to Section 4.6.1, we assume BPSK modulation without any loss of generality. In order

154

to compute the input LLRs of the transmitted symbols, we benefit from the results of [123] where a block-fading channel model is considered instead of a time-varying multi-antenna channel as is the choice for this particular case. We emphasize the point that the LLR computation presented below is admittedly not optimal in the sense that it does no take into account the imperfection of the associated channel estimate as is argued in [124]. Nevertheless, deriving optimal receivers such as the one given in [124] is beyond the scope of this work, and the one presented here is believed to work well enough from our point of view.

The input LLR of the symbol $a_{k,m}$ associated with the m-th transmitter antenna and k-th symbol time at the i-th estimation iteration is then given as

$$L^{(i)}(a_{k,m}) = \log \frac{P\left(a_{k,m} = 1 \,\middle|\, \mathbf{y}_k, \hat{\mathbf{H}}_k^{(i)}\right)}{P\left(a_{k,m} = -1 \,\middle|\, \mathbf{y}_k, \hat{\mathbf{H}}_k^{(i)}\right)} \tag{5.37}$$

$$= \log \frac{P\left(\mathbf{y}_k, a_{k,m} = 1 \,\middle|\, \hat{\mathbf{H}}_k^{(i)}\right)}{P\left(\mathbf{y}_k, a_{k,m} = -1 \,\middle|\, \hat{\mathbf{H}}_k^{(i)}\right)} \tag{5.38}$$

which is a suboptimal approach as stated in Section 4.6.1, and where $\hat{\mathbf{H}}_k^{(i)}$ is the estimate of the channel matrix $\mathbf{H}_k$ at the i-th iteration. $L^{(i)}(a_{k,m})$ could be further elaborated as follows

$$L^{(i)}(a_{k,m}) = \log \frac{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=1} P\left(\mathbf{y}_k, \mathbf{a}_k \,\middle|\, \hat{\mathbf{H}}_k^{(i)}\right)}{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=-1} P\left(\mathbf{y}_k, \mathbf{a}_k \,\middle|\, \hat{\mathbf{H}}_k^{(i)}\right)} \tag{5.39}$$

$$= \log \frac{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=1} P\left(\mathbf{y}_k \,\middle|\, \hat{\mathbf{H}}_k^{(i)}, \mathbf{a}_k\right)}{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=-1} P\left(\mathbf{y}_k \,\middle|\, \hat{\mathbf{H}}_k^{(i)}, \mathbf{a}_k\right)} \tag{5.40}$$

where the constraint $\mathbf{a}_k : a_{m,k}$ refers to the set of $\mathbf{a}_k$'s whose m-th element $a_{m,k}$ takes the specified value. Since $y_{n,k}$'s are uncorrelated for a given channel matrix at a given time epoch k, we may obtain the following equivalent expression

$$L^{(i)}(a_{k,m}) = \log \frac{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=1} \prod_{n=1}^{N} P\left(y_{n,k} \,\middle|\, \hat{\mathbf{h}}_{n,k}^{(i)}, \mathbf{a}_k\right)}{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=-1} \prod_{n=1}^{N} P\left(y_{n,k} \,\middle|\, \hat{\mathbf{h}}_{n,k}^{(i)}, \mathbf{a}_k\right)}, \tag{5.41}$$

and it becomes with the help of (5.3) as follows:

$$L^{(i)}(a_{k,m}) = \log \frac{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=1} \prod_{n=1}^{N} \exp\left\{-\frac{\left|y_{n,k} - \left(\hat{\mathbf{h}}_{n,k}^{(i)}\right)^T \mathbf{a}_k\right|^2}{N_0}\right\}}{\displaystyle\sum_{\mathbf{a}_k : a_{k,m}=-1} \prod_{n=1}^{N} \exp\left\{-\frac{\left|y_{n,k} - \left(\hat{\mathbf{h}}_{n,k}^{(i)}\right)^T \mathbf{a}_k\right|^2}{N_0}\right\}}. \tag{5.42}$$

Note that LLR of the coded symbols c_k's and the modulated symbols m_k's are equivalent due to the BPSK assumption and we therefore do not need any bit-to-symbol level LLR conversion or vice versa. The rest of the operation is the same as in Section 3.8.1.

5.4.2 Channel Estimation Algorithms for Flat-Fading MIMO Channels

In this section, we overview the channel estimation algorithms for time-varying flat-fading MIMO channels. Because the formulations presented here are very similar to those given in Section 4.6.2, we will present some important results only.

5.4.2.1 MMSE Channel Estimation

An initial estimate for estimate of $\mathbf{h}_n$ at the i-th iteration is given as follows

$$\hat{\mathbf{h}}_n^{(1)} = \mathbf{W}_n^{(1)} \mathbf{y}_{n,p} \tag{5.43}$$

where $\mathbf{y}_{n,p} = \left[y_{n,k_p(1)} \cdots y_{n,k_p(L_p)}\right]^T$ is the observation vector received at the n-th antenna and is associated with all the available pilot symbol vectors. Defining $\mathbf{h}_{n,p} = \left[\mathbf{h}_{n,k_p(1)}^T, \ldots, \mathbf{h}_{n,k_p(L_p)}^T\right]^T$ to be the subchannel vector associated with all the available pilots, the desired Wiener-Hopf's equations are given by (4.104) as follows

$$\mathbf{W}^{(1)} = \mathbf{P}_p \mathbf{A}_p^H \left(\mathbf{A}_p \mathbf{R}_p \mathbf{A}_p^H + N_0 \mathbf{I}\right)^{-1} \tag{5.44}$$

where the subscript n is removed since the $\mathbf{P}_p = E\left\{\mathbf{h}_n \mathbf{h}_{n,p}^H\right\}$ and $\mathbf{R}_p = E\left\{\mathbf{h}_{n,p} \mathbf{h}_{n,p}^H\right\}$ do not depend on the receiver antenna choice. Using (5.1), any element of $\mathbf{P}_p$ and $\mathbf{R}_p$ could be

expressed as

$$E\left\{h_{nm,i} \, h_{n'm',k_p(j)}\right\} = \delta_{nn'} \, \delta_{mm'} \, r\left(i - k_p(j)\right)$$

$$E\left\{h_{nm,k_p(i)} \, h_{n'm',k_p(j)}\right\} = \delta_{nn'} \, \delta_{mm'} \, r\left(k_p(i) - k_p(j)\right)$$

which clarifies the above claim.

In the subsequent estimation iterations, estimate of $\mathbf{h}_n$ at the i-th iteration is given as

$$\hat{\mathbf{h}}_n^{(i)} = \mathbf{W}_n^{(i)} \, \mathbf{y}_n \tag{5.45}$$

for $i > 1$ where the associated MMSE filter is given by (B.16) as follows

$$\mathbf{W}^{(i)} = \mathbf{R}_{hh} \left(\hat{\mathbf{A}}^{(i)}\right)^H \left(\hat{\mathbf{A}}^{(i)} \, \mathbf{R}_{hh} \left(\hat{\mathbf{A}}^{(i)}\right)^H + N_0 \, \mathbf{I}\right)^{-1}. \tag{5.46}$$

where $\mathbf{R}_{hh} = E\left\{\mathbf{h}_n \, \mathbf{h}_n^H\right\}$ is the the correlation matrix which does not depend on the receiver antenna index n and $\hat{\mathbf{A}}^{(i)}$ is the data matrix introduced in Section 4.6.2.1.

5.4.2.2 Unidirectional LMS Channel Estimation

Following the result of Section 5.2.2, the conventional unidirectional LMS algorithm adopted to the iterative channel estimation problem for a time-varying flat-fading MIMO channel is given for the i-th iteration as follows

$$\hat{\mathbf{h}}_{n,k+1}^{(i)} = \hat{\mathbf{h}}_{n,k}^{(i)} + 2\mu \, e_{n,k}^{(i)} \, \hat{\mathbf{a}}_k^{(i)} \tag{5.47}$$

for $n = 1, 2, \ldots, N$, where $i > 1$ so that we employ the algorithm after the initial channel estimation iteration as before, and μ is the step-size value of the adaptation. The error term $e_{n,k}^{(i)}$ in (5.47) is given as

$$e_{n,k}^{(i)} = y_{n,k} - \left(\hat{\mathbf{h}}_{n,k}^{(i)}\right)^T \hat{\mathbf{a}}_k^{(i)}. \tag{5.48}$$

5.4.2.3 Bidirectional LMS Channel Estimation

As described in Section 5.2.2, the forward and the backward adaptations of the bidirectional LMS algorithm adopted to the iterative channel estimation problem for a time-varying flat-

157

fading MIMO channel are given for the i-th iteration as follows

$$\hat{\mathbf{h}}_{n,k+1}^{f,(i)} \quad = \quad \hat{\mathbf{h}}_{n,k}^{f,(i)} + 2\mu\, e_{n,k}^{f,(i)}\, \hat{\mathbf{a}}_k^{(i)} \tag{5.49}$$

$$\hat{\mathbf{h}}_{n,k-1}^{b,(i)} \quad = \quad \hat{\mathbf{h}}_{n,k}^{b,(i)} + 2\mu\, e_{n,k}^{b,(i)}\, \hat{\mathbf{a}}_k^{(i)} \tag{5.50}$$

for $n = 1, 2, \ldots, N$, where $i > 1$ as before, μ is the common step-size value and $e_{n,k}^{f,(i)}$ and $e_{n,k}^{b,(i)}$ are the associated error terms given as

$$e_{n,k}^{f,(i)} \quad = \quad y_{n,k} - \left(\hat{\mathbf{h}}_{n,k}^{f,(i)}\right)^T \hat{\mathbf{a}}_k^{(i)} \tag{5.51}$$

$$e_{n,k}^{b,(i)} \quad = \quad y_{n,k} - \left(\hat{\mathbf{h}}_{n,k}^{b,(i)}\right)^T \hat{\mathbf{a}}_k^{(i)}. \tag{5.52}$$

The final estimate $\hat{\mathbf{h}}_{n,k}^{(i)}$ is found to be

$$\hat{\mathbf{h}}_{n,k}^{(i)} = \frac{\hat{\mathbf{h}}_{n,k}^{f,(i)} + \hat{\mathbf{h}}_{n,k}^{b,(i)}}{2}. \tag{5.53}$$

5.4.3   Numerical Results

In this section, we evaluate the performances of the channel estimation algorithms given in Section 5.4.2 over a time-varying flat-fading MIMO channel with the transmitter and receiver models introduced in Section 5.4.1. Rayleigh fading with Jakes' model is assumed for each of the uncorrelated subchannels with the temporal autocorrelation given in (3.69).

We consider a MIMO system equipped with 2 transmitter and 4 receiver antennas. The speed of each subchannel variation is determined by the common normalized maximum Doppler frequency which is chosen to be $f_d T_s = 0.01$ and $f_d T_s = 0.02$ throughout the simulations. A set of $L_d = 98$ symbols are chosen from the set $\{0, 1\}$ in an independent and identical fashion at the transmitter. A rate $R_c = 1/2$ convolutional encoder with generator $(1, 5/7)_8$ is used, as before, together with 2 termination bits for each of the transmitted block, and a number of $L_c = 200$ coded symbols are produced accordingly. After passing through a random interleaver, the coded symbols are distributed to the transmitter antennas such that the consecutive symbols are associated with different antennas. The resulting coded sequences for each of the transmitter antennas are then modulated using a BPSK modulator, and necessary pilots are inserted into the sequence prior to transmission. The pilot symbol spacing is chosen to be $M_p = 11$ throughout the simulations which results in a number of $ML_p = 2\times10 = 20$ for

each of the transmitted block of length $L = 220$. The pilot overhead percentage is therefore 0.1 and the overall transmission rate, i.e., $R = L_d/L$, becomes 0.4455.

We prefer to transmit orthogonal pilot vectors chosen from the Alamouti set [125] as suggested in [126]. In Fig. 5.8, the effect of various pilot sequence choices are presented together with MMSE filter. A nonorthogonal pilot sequence using +1 only is observed to be extremely bad, whereas the Alamouti set result in a better error performance than the random BPSK sequence where symbols are selected randomly from the set $\{-1, 1\}$ for the latter.
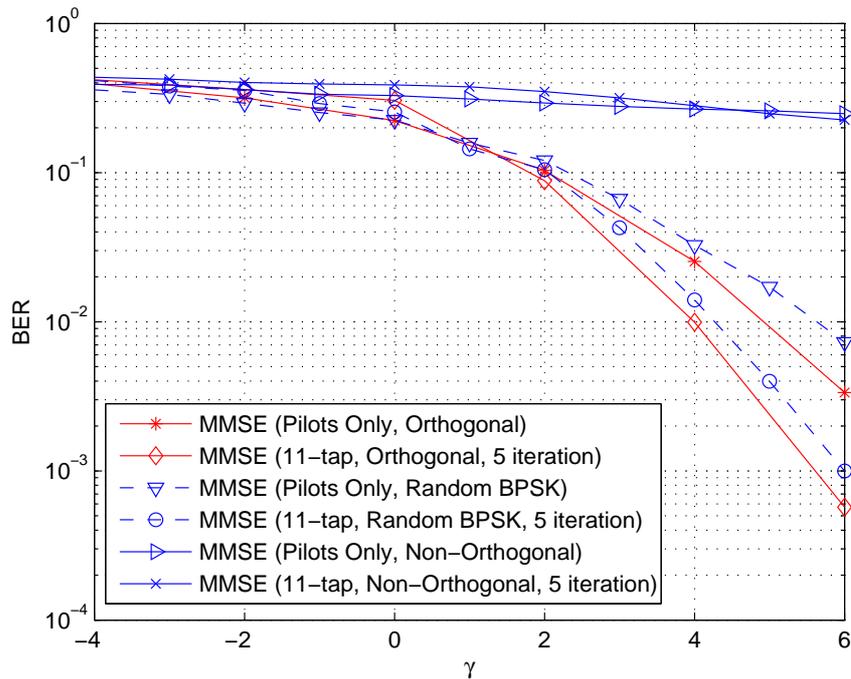
Figure 5.8: BER for MMSE with $M_p = 11$ over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$. Orthogonal pilots are chosen from the Alamouti set, and all pilots are 1 for non-orthogonal case.

At the receiver, soft-decision feedback is considered where the number of channel estimation iterations is 5 for which a good convergence behavior is observed. As usual, sufficient number of independent Monte Carlo trials are performed to produce statistically significant results. In addition, both the unidirectional and bidirectional LMS algorithms use optimal step-size values which are determined in trial and error basis for each SNR level. The details of the receiver operations are explained in Section 5.4.1.
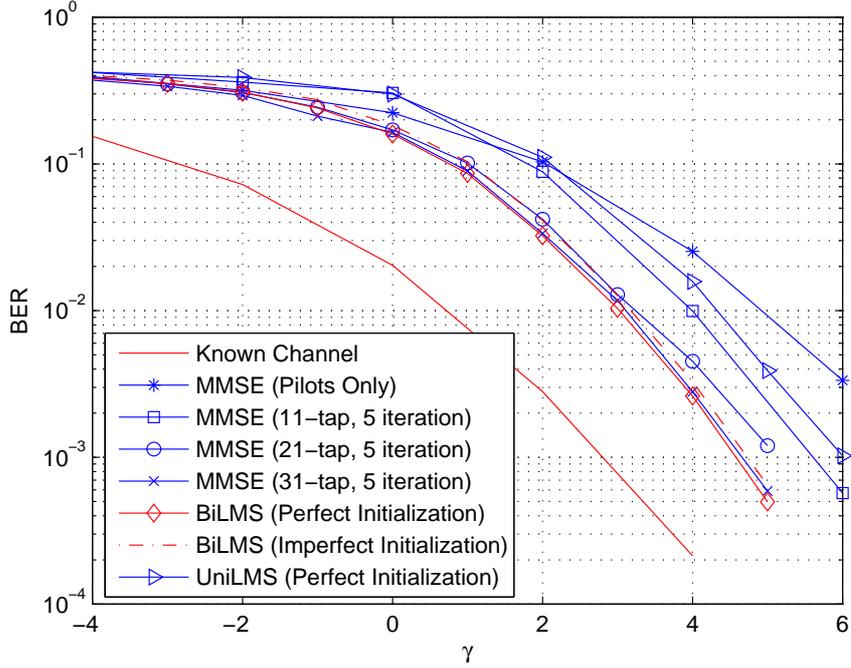
Figure 5.9: BER for BiLMS, UniLMS and MMSE with $M_p = 11$ over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$. All results for BiLMS and UniLMS are associated with the 5-th estimation iteration.

In Fig. 5.9, we depict BER results for the bidirectional LMS algorithm, the unidirectional LMS algorithm and the MMSE filter with 11, 21 and 31 taps associated with the 5-th channel estimation iteration. The BER performance of the MMSE filter after the 1-st channel estimation iteration for which only all the available pilot symbols are employed is also provided together with the known channel bound. We observe that, the bidirectional LMS algorithm achieves a BER performance which is much better than that of the 11-tap MMSE filter and is very close to that of the 21-tap and 31-tap MMSE filters. We also observe that the BER of the bidirectional LMS algorithm is off the known channel by only 2 dB whereas this gap is 4 dB for the initial channel estimation with the MMSE filter using all available pilots. In addition, imperfect initialization for the bidirectional LMS algorithm in which the estimates of channel coefficients from the last iteration is used to initialize the algorithm is observed not to cause any performance degradation. Finally, the unidirectional LMS algorithm is observed fill this gap from the known channel only 1 dB.

160
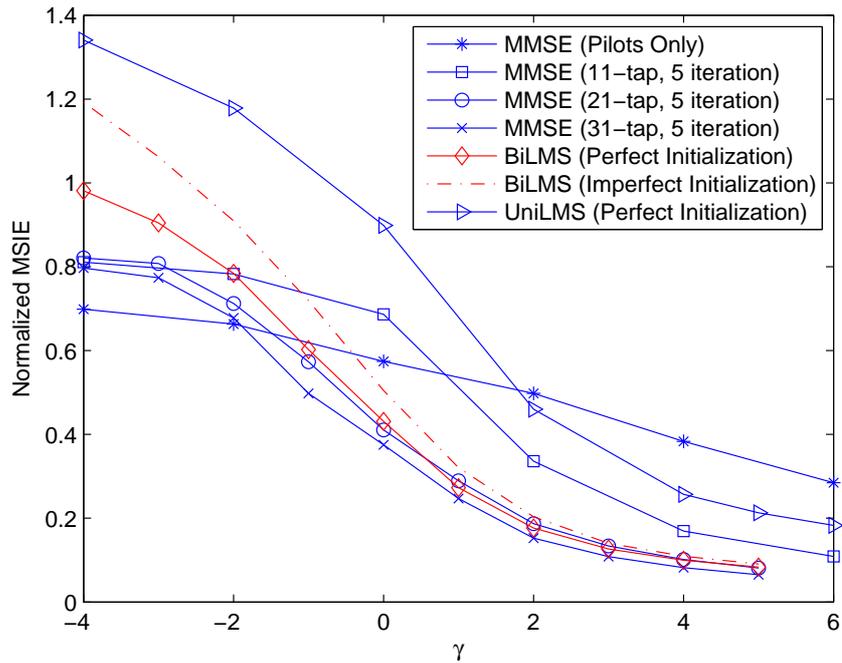
Figure 5.10: Normalized MSIE for BiLMS, UniLMS and MMSE with $M_p = 11$ over a Rayleigh fading MIMO channel with $f_d T_s = 0.01$ and $(M, N) = (2, 4)$. All results for BiLMS and UniLMS are associated with the 5-th estimation iteration.

We also depict the normalized MSIE statistics in Fig. 5.10 corresponding to the Monte Carlo simulations presented through BER results in the previous figure. We observe that the normalized MSIE of the bidirectional LMS algorithm is much better than that of the 11-tap MMSE filter and is close to the 21-tap and 31-tap MMSE filters, and that imperfect initialization does not cause a significant performance degradation, as before. The normalized MSIE of the 11-tap MMSE filter is observed to be better than the reference statistics corresponding to the initial channel estimation after $\gamma = 1$ dB SNR whereas this threshold is $\gamma = 2$ dB for the unidirectional LMS algorithm and $\gamma = -1$ dB for the rest of the algorithms. Consequently, these observations absolutely coincide with those inferred from Fig. 5.9.
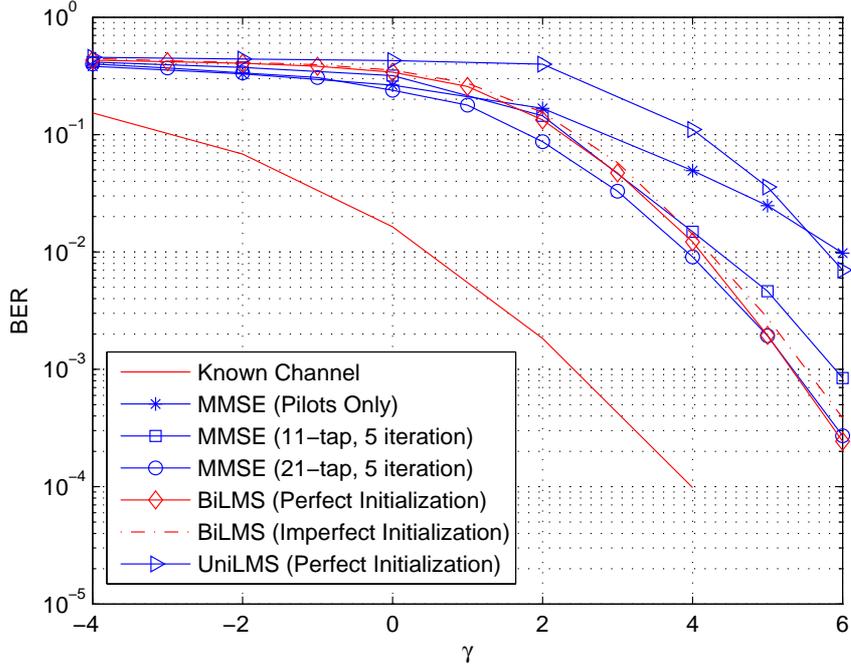
Figure 5.11: BER for BiLMS, UniLMS and MMSE with $M_p = 11$ over a Rayleigh fading MIMO channel with $f_d T_s = 0.02$ and $(M, N) = (2, 4)$. All results for BiLMS and UniLMS are associated with the 5-th estimation iteration.

In Fig. 5.11 and Fig. 5.12, we depict the BER and normalized MSIE results for a faster Rayleigh fading channel with $f_d T_s = 0.02$. For this case, the bidirectional LMS algorithm is observed to achieve a similar performance to that of the 21-tap MMSE filter and a better performance than that of the 11-tap MMSE filter at high SNR. The unidirectional LMS algorithm is observed to degrade the performance associated with the initial channel estimate except for a small region at very high SNR. The results show that the BER corresponding to the initial channel estimate is off the known channel bound by as large as 6 dB, and that this gap is only 3 dB for the bidirectional LMS algorithm and 21-tap MMSE filter after 5 estimation iteration. In this sense, the bidirectional LMS algorithm is observed to achieve almost 1 dB SNR improvement over the 11-tap MMSE filter, and is very robust to the imperfect initialization with the channel estimates from the previous estimation iteration. After these common conclusions for both figures, we also observe in Fig. 5.12 that the bidirectional LMS algorithm and the 21-tap MMSE filter achieve a better MSIE results than that of the reference level associated with the initial channel estimate after almost $\gamma = 1$ dB SNR whereas this threshold is as large as $\gamma = 5$ dB for the unidirectional LMS algorithm.
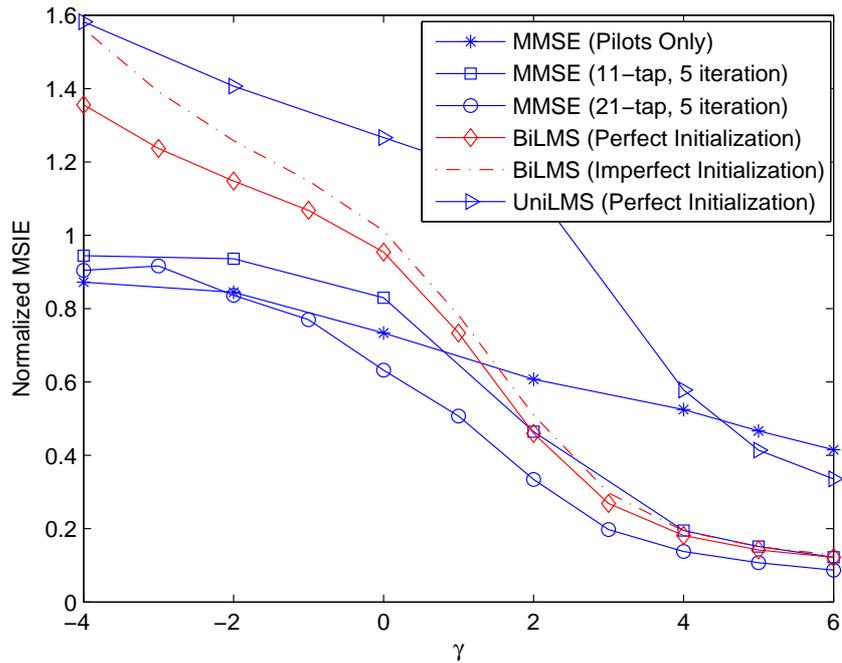
Figure 5.12: Normalized MSIE for BiLMS, UniLMS and MMSE with $M_p = 11$ over a Rayleigh fading MIMO channel with $f_d T_s = 0.02$ and $(M, N) = (2, 4)$. All results for BiLMS and UniLMS are associated with the 5-th estimation iteration.

5.5 Conclusion

A time-varying flat-fading MIMO communication channel is considered in this chapter with the purpose of searching for the efficient ways of channel estimation. The channel under consideration has some similarities with both the flat- and frequency-selective fading channels and indeed a kind of generalization for both of these previous channel models. Because the channel to be estimated is in the form of a matrix varying with time, the resulting complexity associated with the optimal Wiener filter is extremely high. The bidirectional LMS algorithm is shown to achieve a close tracking performance to the optimal Wiener filter in both coded and uncoded environments at a very low level of complexity which is comparable to that of the conventional unidirectional LMS algorithm. The tracking behavior is analyzed and the associated optimal step-size is derived. A step-size dependent steady-state MSE expression is obtained together with the optimal step-size expressions with a satisfactory accuracy.

CHAPTER 6

CONCLUSION

6.1 Contribution

This work considers estimation of time-varying communication channels with low-complexity and near-optimal algorithms. Making use of the bidirectional signal processing techniques present in the literature, the bidirectional LMS algorithm is considered as an extension of the well-known conventional unidirectional LMS algorithm. This algorithm is shown to achieve similar MSE statistics as that of the optimal Wiener filter under some specific scenarios as well as the BER and BLER statistics in both genie-aided and coded communication systems closer to real-life scenarios. Beside this good tracking ability, the complexity increase as a result of a bidirectional processing is shown to be still at practical levels as compared to both the conventional LMS and the optimal Wiener filter. In addition, the bidirectional LMS algorithm is also proved to be robust to the imperfect initialization which improves its practical value.

One of the major contributions of this work is the steady-state tracking analysis of the bidirectional LMS algorithm. The analysis derives a step-size dependent steady-state MSE expression which is valid for any kind of channel characteristics with a known power spectrum. Therefore, the proposed analysis is valid, for example, for the Rayleigh fading with a well-known nonrational Jakes' spectrum as well as a simple rational spectrum characterized in the time-domain by an AR process. In addition, much simpler form of the steady-state MSE expression is obtained for AR channels by transferring the frequency-domain computations into the time-domain. The optimal selection of the step-size value for the bidirectional LMS algorithm is also presented using the result of the steady-state MSE analysis. The numerical

evaluations show a very good match between the theoretical and experimental results for a variety of time-varying channel choices including single-antenna and multi-antenna structures with flat and frequency-selective fading. The analysis also shows that the best performance is achieved when the forward and the backward step-size values are equal.

In performance analysis of the bidirectional algorithm, a special attention is devoted to frequency-selective fading channels since the so-called independence assumption is not valid due to the memory of the channel of interest. The steady-state MSE analysis is therefore modified for frequency-selective channels to take into account the corresponding vector structure. The associated comparisons between theoretical and experimental results exhibit a good match for frequency-selective channels, as well.

Finally, we have realized that although there are various studies present in the literature on the bidirectional estimation strategies in communication systems, none of them provide a theoretical analysis about the underlying estimators. As such, our work not only presents a bidirectional extension of the LMS algorithm with a near-optimal performance at a practical level of complexity but also novel in the sense that it provides a detailed tracking analysis for the algorithm resulting in a step-size dependent steady-state MSE expression with high accuracy.

6.2 Future Work

In this work, we assume a packet-based transmission scheme with relatively short block-lengths. The transient phase of the bidirectional LMS algorithm is therefore skipped by using initial conditions on the channel estimate with varying accuracy, i.e., imperfect or genie-aided. Indeed, the transient behavior of the bidirectional LMS algorithm including the rate of convergence and the associated MSE value has a great importance for some practical applications including equalization as well as channel estimation. This analysis would be useful for both flat and frequency-selective fading channels. The work on the analysis of the conventional LMS algorithm in transient phase with/without the independence assumption is very limited even in the signal processing literature, and therefore there is a good opportunity to elaborate and apply these results to the communication area.

Some other extensions of this work is to investigate the applicability of the present results

to some other adaptive algorithms, e.g, the conventional RLS or extended RLS (ERLS). An interesting application of the bidirectional LMS algorithm is the channel estimation problem for both quasi-static and time-varying OFDM communication. Remembering the two-dimensional Wiener filter is employed in such systems to exploit the correlation both in the frequency and time domains, the analysis of the bidirectional LMS operated in a similar fashion seems really challenging.

REFERENCES

[1] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 4 edition, September 2001. ISBN 0130901261.

[2] Y.C. Lim and C.C. Ko. Forward-backward LMS adaptive line enhancer. *Circuits and Systems, IEEE Transactions on*, 37(7):936–940, July 1990.

[3] Hen-Geul Yeh. An efficient implementation of forward-backward LMS adaptive filters. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, 44(4):332–336, April 1997.

[4] A. Anastasopoulos and K.M. Chugg. Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty. *Communications, IEEE Transactions on*, 48(10):1638–1649, October 2000.

[5] R. Raheli, A. Polydoros, and C.-K. Tzou. The principle of per-survivor processing: A general approach to approximate and adaptive MLSE. In *Proc. of IEEE GLOBECOM'91*, volume 2, December 1991.

[6] R. Raheli, A. Polydoros, and Ching-Kae Tzou. Per-survivor Processing: A general approach to MLSE in uncertain environments. *Communications, IEEE Transactions on*, 43(234):354–364, Feb/Mar/Apr 1995.

[7] J.K. Tugnait, Lang Tong, and Zhi Ding. Single-user channel estimation and equalization. *Signal Processing Magazine, IEEE*, 17(3):16–28, May 2000.

[8] Raymond Steele and Lajos Hanzo. *Mobile Radio Communications*. Wiley-IEEE Press, 2 edition, May 1999. ISBN 978-0-471-97806-0.

[9] K.-D. Kammeyer, V. Kuhn, and T. Petermann. Blind and nonblind turbo estimation for fast fading GSM channels. *Selected Areas in Communications, IEEE Journal on*, 19 (9):1718–1728, September 2001.

[10] Lang Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, October 1998.

[11] J.H. Lodge and M.L. Moher. Time diversity for mobile satellite channels using trellis coded modulations. In *Proc. of IEEE Global Telecommunications*, volume 3, 1987.

[12] S. Sampei and T. Sunaga. Rayleigh fading compensation method for 16QAM in digital land mobile radio channels. In *Proc. of IEEE 39th Vehicular Technology Conference*, volume 2, pages 640–646, 1-3 1989.

[13] M.L. Moher and J.H. Lodge. TCMP - A modulation and coding strategy for Rician fading channels. *Selected Areas in Communications, IEEE Journal on*, 7(9):1347–1355, December 1989.

[14] J.K. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *Vehicular Technology, IEEE Transactions on*, 40(4):686–693, November 1991.

[15] J.K. Cavers. Pilot symbol assisted modulation and differential detection in fading and delay spread. *Communications, IEEE Transactions on*, 43(7):2206–2212, July 1995.

[16] P. Hoeher and F. Tufvesson. Channel estimation with superimposed pilot sequence. In *Proc. of GLOBECOM'99*, volume 4, pages 2162–2166, 1999.

[17] B. Farhang-Boroujeny. Pilot-based channel identification: Proposal for semi-blind identification of communication channels. *Electronics Letters*, 31(13):1044–1046, 22 1995.

[18] F. Mazzenga. Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence. *Broadcasting, IEEE Transactions on*, 46(2):170–176, June 2000.

[19] J.K. Tugnait and Xiaohong Meng. On superimposed training for channel estimation: Performance analysis, training power allocation, and frame synchronization. *Signal Processing, IEEE Transactions on*, 54(2):752–765, February 2006.

[20] A.G. Orozco-Lugo, M.M. Lara, and D.C. McLernon. Channel estimation using implicit training. *Signal Processing, IEEE Transactions on*, 52(1):240–254, January 2004.

[21] F. Tsuzuki and T. Ohtsuki. Channel estimation with selective superimposed pilot sequences under fast fading environments. In *Proc. of IEEE VTC'04 Fall*, volume 1, pages 62–66, 26-29 2004.

[22] J.K. Tugnait and Shuangchi He. Doubly-selective channel estimation using data-dependent superimposed training and exponential basis models. *Wireless Communications, IEEE Transactions on*, 6(11):3877–3883, November 2007.

[23] Shuangchi He, J.K. Tugnait, and Xiaohong Meng. On superimposed training for MIMO channel estimation and symbol detection. *Signal Processing, IEEE Transactions on*, 55(6):3007–3021, June 2007.

[24] Xiaohong Meng, J.K. Tugnait, and Shuangchi He. Iterative joint channel estimation and data detection using superimposed training: Algorithms and performance analysis. *Vehicular Technology, IEEE Transactions on*, 56(4):1873–1880, July 2007.

[25] A. Goljahani, N. Benvenuto, S. Tomasin, and L. Vangelista. Superimposed sequence versus pilot aided channel estimations for next generation DVB-T systems. *Broadcasting, IEEE Transactions on*, 55(1):140–144, March 2009.

[26] Min Dong, Lang Tong, and B.M. Sadler. Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels. *Signal Processing, IEEE Transactions on*, 52(5):1403–1418, May 2004.

[27] Min Dong and Lang Tong. Optimal design and placement of pilot symbols for channel estimation. *Signal Processing, IEEE Transactions on*, 50(12):3055–3069, December 2002.

[28] G. Tong Zhou, M. Viberg, and T. McKelvey. Superimposed periodic pilots for blind channel estimation. In *Proc. of 35th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 653–657, 2001.

[29] J.K. Tugnait and Weilin Luo. On channel estimation using superimposed training and first-order statistics. *Communications Letters, IEEE*, 7(9):413–415, September 2003.

[30] Lang Tong, B.M. Sadler, and Min Dong. Pilot-assisted wireless transmissions: General model, design criteria, and signal processing. *Signal Processing Magazine, IEEE*, 21 (6):12–25, November 2004.

[31] D. Boss, K.-D. Kammeyer, and T. Petermann. Is blind channel estimation feasible in mobile communication systems? A study based on GSM. *Selected Areas in Communications, IEEE Journal on*, 16(8):1479–1492, October 1998.

[32] J.-J. Werner, Jian Yang, D.D. Harman, and G.A. Dumont. Blind equalization for broadband access. *Communications Magazine, IEEE*, 37(4):87–93, April 1999.

[33] Y. Sato. A method of self-recovering equalization for multilevel amplitude-modulation systems. *Communications, IEEE Transactions on*, 23(6):679–682, June 1975.

[34] D. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *Communications, IEEE Transactions on*, 28(11):1867–1875, November 1980.

[35] J. Treichler and B. Agee. A new approach to multipath correction of constant modulus signals. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 31(2):459–472, April 1983.

[36] Hui Liu, Guanghan Xu, Lang Tong, and Thomas Kailath. Recent developments in blind channel equalization: From cyclostationarity to subspaces. *Signal Processing*, 50(1-2):83–99, 1996.

[37] Ruey-Wen Liu. Blind signal processing: An introduction. In *Proc. of 1996 IEEE International Symposium on Circuits and Systems (ISCAS'96)*, volume 2, pages 81–84, 12–15 1996.

[38] B. Muquet, M. de Courville, and P. Duhamel. Subspace-based blind and semi-blind channel estimation for OFDM systems. *Signal Processing, IEEE Transactions on*, 50 (7):1699–1712, July 2002.

[39] Husheng Li and H.V. Poor. Performance analysis of semiblind channel estimation in long-code DS-CDMA systems. *Signal Processing, IEEE Transactions on*, 54(9):3383–3399, September 2006.

[40] G.G. Raleigh and J.M. Cioffi. Spatio-temporal coding for wireless communication. *Communications, IEEE Transactions on*, 46(3):357–366, March 1998.

[41] Thomas L. Marzetta. BLAST training: Estimating channel characteristics for high capacity space-time wireless. In *Proc. of 37th Annual Allerton Conference*, pages 958–966, 1999.

[42] Thomas Kailath, Ali H. Sayed, and Babak Hassibi. *Linear Estimation*. Prentice Hall, 1 edition, April 2000. ISBN 0130224642.

[43] Qinfang Sun, D.C. Cox, A. Lozano, and H.C. Huang. Training-based channel estimation for continuous flat fading BLAST. In *Proc. of ICC'02*, volume 1, pages 325–329, April–May 2002.

[44] Qinfang Sun, D.C. Cox, H.C. Huang, and A. Lozano. Estimation of continuous flat fading MIMO channels. *Wireless Communications, IEEE Transactions on*, 1(4):549–553, October 2002.

[45] C. Komninakis, C. Fragouli, A.H. Sayed, and R.D. Wesel. Multi-input multi-output fading channel tracking and equalization using Kalman estimation. *Signal Processing, IEEE Transactions on*, 50(5):1065–1076, May 2002.

[46] Haidong Zhu, B. Farhang-Boroujeny, and C. Schlegel. Pilot embedding for joint channel estimation and data detection in MIMO communication systems. *Communications Letters, IEEE*, 7(1):30–32, January 2003.

[47] J.H. Kotecha and A.M. Sayeed. Transmit signal design for optimal estimation of correlated MIMO channels. *Signal Processing, IEEE Transactions on*, 52(2):546–557, February 2004.

[48] O. Simeone and U. Spagnolini. Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels. *Signal Processing, IEEE Transactions on*, 52(11):3265–3277, November 2004.

[49] O. Besson and P. Stoica. On parameter estimation of MIMO flat-fading channels with frequency offsets. *Signal Processing, IEEE Transactions on*, 51(3):602–613, March 2003.

[50] D.K.C. So and R.S. Cheng. Iterative EM receiver for space-time coded systems in MIMO frequency-selective fading channels with channel gain and order estimation. *Wireless Communications, IEEE Transactions on*, 3(6):1928–1935, November 2004.

[51] A. Dakdouki and J. Thompson. Channel estimation for multiple-antenna wireless communications. *Antennas and Propagation Magazine, IEEE*, 51(3):194–204, June 2009.

[52] J. Baltersee, G. Fock, and H. Meyr. Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving. *Selected Areas in Communications, IEEE Journal on*, 19(12):2358–2368, December 2001.

[53] T. Yoo and A. Goldsmith. Capacity and power allocation for fading MIMO channels with channel estimation error. *Information Theory, IEEE Transactions on*, 52(5):2203–2214, May 2006.

[54] B. Hassibi and B.M. Hochwald. How much training is needed in multiple-antenna wireless links? *Information Theory, IEEE Transactions on*, 49(4):951–963, April 2003.

[55] Xiaoli Ma, Liuqing Yang, and G.B. Giannakis. Optimal training for MIMO frequency-selective fading channels. In *Proc. of Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, volume 2, pages 1107–1111, November 2002.

[56] Hlaing Minn and N. Al-Dhahir. Optimal training signals for MIMO OFDM channel estimation. *Wireless Communications, IEEE Transactions on*, 5(5):1158–1168, May 2006.

[57] Bernard Widrow and Samuel Stearns. *Adaptive Signal Processing*. Prentice Hall, March 1985. ISBN 0130040290.

[58] Bernard Widrow and Marcian E. Hoff. Adaptive switching circuits. In *Proc. IRE WESCON Conf. Rec.*, pages 96–104, 1960.

[59] Bernard Widrow. Thinking about thinking: The discovery of the LMS algorithm. *Signal Processing Magazine, IEEE*, 22(1):100–106, January 2005.

[60] J. R. Zeidler. Performance analysis of LMS adaptive prediction filters. *Proceedings of the IEEE*, 78(12):1781–1806, December 1990.

[61] B. Widrow, J.M. McCool, M.G. Larimore, and Jr. Johnson, C.R. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceedings of the IEEE*, 64(8):1151–1162, August 1976.

[62] Bernard Widrow and Max Kamenetsky. Statistical efficiency of adaptive algorithms. *Neural Networks*, 16(5-6):735–744, 2003.

[63] Jingdong Lin, J.G. Proakis, Fuyun Ling, and H. Lev-Ari. Optimal tracking of time-varying channels: A frequency domain approach for known and new algorithms. *Selected Areas in Communications, IEEE Journal on*, 13(1):141–154, January 1995.

[64] W.A. Gardner. Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Processing*, 6(2):113–133, 1984.

[65] S.C. Douglas and Weimin Pan. Exact expectation analysis of the LMS adaptive filter. *Signal Processing, IEEE Transactions on*, 43(12):2863–2871, December 1995.

[66] H.J. Butterweck. A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption. In *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'95)*, volume 2, pages 1404–1407, May 1995.

[67] M. Rupp and H.-J. Butterweck. Overcoming the independence assumption in LMS filtering. In *Proc. of Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 1, pages 607–611, November 2003.

[68] E. Eweda. Tracking analysis of the normalized LMS algorithm without the independence and small step size assumptions. In *Proc. of Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on*, pages 129–134, December 2009.

[69] N.R. Yousef and A.H. Sayed. A unified approach to the steady-state and tracking analyses of adaptive filters. *Signal Processing, IEEE Transactions on*, 49(2):314–324, February 2001.

[70] T.Y. Al-Naffouri and A.H. Sayed. Transient analysis of adaptive filters with error nonlinearities. *Signal Processing, IEEE Transactions on*, 51(3):653–663, March 2003.

[71] T.Y. Al-Naffouri and A.H. Sayed. Transient analysis of data-normalized adaptive filters. *Signal Processing, IEEE Transactions on*, 51(3):639–652, March 2003.

[72] P. Wei, J. R. Zeidler, J. Han, and W. H. Ku. Comparative tracking performance of the LMS and RLS algorithms for chirped signals. *IEEE Transactions on Signal Processing*, 50:1602–1610, July 2002.

[73] S. Haykin, A. Sayed, J. R. Zeidler, P. Wei, and P. Yee. Tracking of linear time-variant systems by extended RLS algorithms. *IEEE Transactions on Signal Processing*, 45: 1118–1128, May 1997.

[74] B. C. Banister and J. R. Zeidler. Tracking performance of the RLS algorithm applied to an antenna array in a realistic fading environment. *IEEE Transactions on Signal Processing*, 50:1037–1050, May 2002.

[75] M. Reuter and J. R. Zeidler. Non-linear effects in LMS adaptive equalizers. *IEEE Transactions on Signal Processing*, 47(6):1570–1579, June 1999.

[76] A. Batra, J. R. Zeidler, and A. A. (Louis) Beex. A two-stage approach for improving the convergence of least-mean-square decision-feedback adaptive equalizers in the presence of severe narrowband interference. *EURASIP Journal on Advances in Signal Processing*, 2008, 2008. Article ID 390102, 13 pages, 2008. doi:10.1155/2008/390102.

[77] K. J. Quirk, L. B. Milstein, and J. R. Zeidler. A performance bound for the LMS estimator. *IEEE Transactions on Information Theory*, 46:1150–1158, May 2000.

[78] J. Han, J. R. Zeidler, and W. H. Ku. Nonlinear effects of the LMS adaptive predictor for chirped input signals. *EURASIP Journal on Applied Signal Processing, Special Issue on Nonlinear Signal and Image Processing*, 2002(1):21–29, January 2002.

[79] T. Ikuma, A. A. (Louis) Beex, and J. R. Zeidler. Non-Wiener mean weight behavior of LMS transversal equalizers with sinusoidal interference. *IEEE Transactions on Signal Processing*, 56(9):4521–4525, September 2008.

[80] Kuor-Hsin Chang and C.N. Georghiades. Iterative joint sequence and channel estimation for fast time-varying intersymbol interference channels. In *Proc. of IEEE ICC'95*, volume 1, pages 357–361, Seattle, WA, June 1995.

[81] H.E. Gamal, M.K. Khairy, and E. Geraniotis. Iterative decoding and channel estimation of DS/CDMA over slow Rayleigh fading channels. In *IEEE PIMRC*, volume 3, pages 1299–1303, Boston, MA, September 1998.

[82] Magnus Sandell, Carlo Luschi, Paul Strauch, and Ran Yan. Iterative channel estimation using soft decision feedback. In *Proc. of IEEE GLOBECOM'98*, volume 6, pages 3728–3733, Sydney, November 1998.

[83] Yingjiu Xu, Hsuan-Jung Su, and E. Geraniotis. Pilot symbol assisted QAM with iterative filtering and turbo decoding over Rayleigh flat-fading channels. In *Proc. of MILCOM'99*, volume 1, pages 86–91, 1999.

[84] Matthew C. Valenti and Brian D. Woerner. Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels. *Selected Areas in Communications, IEEE Journal on*, 19:1697–1705, September 2001.

[85] N. Nefedov, M. Pukkila, R. Visoz, and A.O. Berthet. Iterative data detection and channel estimation for advanced TDMA systems. *Wireless Communications, IEEE Transactions on*, 51(2):141–144, February 2003.

[86] M. Tuchler, R. Otnes, and A. Schmidbauer. Performance of soft iterative channel estimation in turbo equalization. In *Proc. of IEEE ICC'02*, volume 3, pages 1858–1862, New York, NY, May 2002.

[87] R. Otnes and M. Tuchler. On iterative equalization, estimation, and decoding. In *Proc. of IEEE ICC'03*, volume 4, pages 2958–2962, May 2003.

[88] R. Otnes and M. Tuchler. Iterative channel estimation for turbo equalization of time-varying frequency-selective channels. *Wireless Communications, IEEE Transactions on*, 3(6):1918–1923, November 2004.

[89] T. Abe and T. Matsumoto. Space-time turbo equalization in frequency-selective MIMO channels. *Vehicular Technology, IEEE Transactions on*, 52(3):469–475, May 2003.

[90] L.M. Davis, I.B. Collings, and P. Hoeher. Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels. *Communications, IEEE Transactions on*, 49(12):2106–2114, December 2001.

[91] Joonbeom Kim, Gordon L. Stuber, and Ye Li. Low-complexity iterative channel estimation for serially concatenated systems over frequency-nonselective Rayleigh fading channels. *Wireless Communications, IEEE Transactions on*, 6(2):438–442, February 2007.

[92] Hong Liu and P. Schniter. Iterative frequency-domain channel estimation and equalization for single-carrier transmissions without cyclic-prefix. *Wireless Communications, IEEE Transactions on*, 7(10):3686–3691, October 2008.

[93] Seongwook Song, A.C. Singer, and Koeng-Mo Sung. Soft input channel estimation for turbo equalization. *Signal Processing, IEEE Transactions on*, 52(10):2885–2894, October 2004.

[94] A. Hansson and T. Aulin. Generalized APP detection of continuous phase modulation over unknown ISI channels. *Communications, IEEE Transactions on*, 53(10):1615–1619, October 2005.

[95] Qinghua Guo, Li Ping, and Defeng Huang. A low-complexity iterative channel estimation and detection technique for doubly selective channels. *Wireless Communications, IEEE Transactions on*, 8(8):4340–4349, August 2009.

[96] H.-A. Loeliger. An introduction to factor graphs. *Signal Processing Magazine, IEEE*, 21(1):28–41, January 2004.

[97] H.-A. Loeliger, J. Dauwels, Junli Hu, S. Korl, Li Ping, and F.R. Kschischang. The factor graph approach to model-based signal processing. *Proceedings of the IEEE*, 95 (6):1295–1322, June 2007.

[98] W. C. Jakes. *Microwave Mobile Communications*. John Wiley and Sons, 1974.

[99] Yavuz Yapıcı and Ali Özgür Yılmaz. Joint channel estimation and decoding with low-complexity iterative structures in time-varying fading channels. In *Proc. of PIMRC'09*, pages 1943–1947, Tokyo, Japan, September 13-16, 2009.

[100] Yavuz Yapıcı and Ali Özgür Yılmaz. Low-complexity iterative channel estimation and tracking for time-varying multi-antenna systems. In *Proc. of PIMRC'09*, pages 1317–1321, Tokyo, Japan, September 13-16, 2009.

[101] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. Wiley, March 1996. ISBN 0471594318.

[102] John Proakis. *Digital Communications*. McGraw-Hill, 2000.

[103] Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing*. Prentice Hall, 1975.

[104] C. Watterson, J. Juroshek, and W. Bensema. Experimental confirmation of an hf channel model. *Communication Technology, IEEE Transactions on*, 18(6):792–803, December 1970.

[105] Cihan Tepedelenlioğlu, Ali Abdi, Georgios B. Giannakis, and Mostafa Kaveh. Estimation of doppler spread and signal strength in mobile communications with applications to handoff and adaptive transmission. *Wireless Communications and Mobile Computing*, 1(2):221–242, 2001.

[106] L. Krasny, H. Arslan, D. Koilpillai, and S. Chennakeshu. Doppler spread estimation in mobile radio systems. *Communications Letters, IEEE*, 5(5):197–199, May 2001.

[107] U. Vilaipornsawai and H. Leib. Joint data detection and channel estimation for fading unknown time-varying doppler environments. *Communications, IEEE Transactions on*, 58(8):2277–2291, August 2010.

[108] H. Abeida. Data-aided snr estimation in time-variant rayleigh fading channels. *Signal Processing, IEEE Transactions on*, 58(11):5496–5507, November 2010.

[109] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley-Interscience, 1 edition, September 2001. ISBN 0471095176.

[110] R.A. Ziegler and J.M. Cioffi. Estimation of time-varying digital radio channels. *Vehicular Technology, IEEE Transactions on*, 41(2):134–151, May 1992.

[111] S. Benedetto, A. D. Divsalar, B. G. Montorsi, and F. Pollara B. Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding. *Information Theory, IEEE Transactions on*, 44:909–926, May 1998.

[112] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara. A soft-input soft-output APP module for iterative decoding of concatenated codes. *Communications Letters, IEEE*, 1:22–24, January 1997.

[113] Sergio Benedetto, Dariush Divsalar, Guido Montorsi, and Fabrizio Pollara. Analysis, design, and iterative decoding of double serially concatenated codes with interleavers. *Selected Areas in Communications, IEEE Journal on*, 16:231–244, 1998.

[114] Shu Lin and Daniel J. Costello. *Error Control Coding, Second Edition*. Prentice Hall, 2004.

[115] M. Tuchler, A.C. Singer, and R. Koetter. Minimum mean squared error equalization using a priori information. *Signal Processing, IEEE Transactions on*, 50(3):673–683, mar 2002.

[116] Y. Li, Jr. Cimini, L.J., and N.R. Sollenberger. Robust channel estimation for OFDM systems with rapid dispersive fading channels. *Communications, IEEE Transactions on*, 46(7):902–915, July 1998.

[117] A.N. D'Andrea, U. Mengali, and R. Reggiannini. The modified cramer-rao bound and its application to synchronization problems. *Communications, IEEE Transactions on*, 42(234):1391–1399, feb/mar/apr 1994.

[118] F. Gini, R. Reggiannini, and U. Mengali. The modified cramer-rao bound in vector parameter estimation. *Communications, IEEE Transactions on*, 46(1):52–60, January 1998.

[119] J.-P. Delmas. Closed-form expressions of the exact cramer-rao bound for parameter estimation of bpsk, msk, or qpsk waveforms. *Signal Processing Letters, IEEE*, 15: 405–408, 2008.

[120] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005.

[121] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H. Vincent Poor. *MIMO Wireless Communications*. Cambridge University Press, 2007.

[122] Xinmin Deng, A.M. Haimovich, and J. Garcia-Frias. Decision directed iterative channel estimation for MIMO systems. In *Proc. of IEEE ICC'03*, volume 4, pages 2326–23294, May 2003.

[123] A. Stefanov and T.M. Duman. Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations. *Selected Areas in Communications, IEEE Journal on*, 19(5):958–968, May 2001.

[124] G. Taricco and E. Biglieri. Space-time decoding with imperfect channel estimation. *Wireless Communications, IEEE Transactions on*, 4(4):1874–1888, July 2005.

[125] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *Selected Areas in Communications, IEEE Journal on*, 16(8):1451–1458, October 1998.

[126] Jiann-Ching Guey, M.P. Fitz, M.R. Bell, and Wen-Yi Kuo. Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels. *Communications, IEEE Transactions on*, 47(4):527–537, April 1999.

[127] A. Hjorungnes, D. Gesbert, and D.P. Palomar. Unified theory of complex-valued matrix differentiation. In *Proc. of ICASSP'07*, volume 3, pages III–345–III–348, April 15–20, 2007.

[128] A. Hjorungnes and D. Gesbert. Complex-valued matrix differentiation: Techniques and key results. *Signal Processing, IEEE Transactions on*, 55(6):2740–2746, June 2007.

Appendix A

Derivation of the MMSE Filter for Flat-Fading Channels

A.1 The MMSE Filter for Perfectly Known Transmitted Symbols

The optimal coefficients for the MMSE filter are the ones which minimize the mean-square error given as

$$J_k = E\left\{|e_k|^2\right\},\tag{A.1}$$

where e_k is the estimation error defined as

$$
\begin{aligned}
e_k &= f_k - \hat{f}_k \\
&= f_k - \sum_{n=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,n}\, y_{k-n}.
\end{aligned}\tag{A.2}
$$

This minimization problem is solved by taking derivative of (A.1) with respect to the complex conjugate of the individual filter taps as follows

$$
\begin{aligned}
\frac{\partial J_k}{\partial w_{k,n}^*} &= \frac{\partial}{\partial w_{k,n}^*} E\left\{|e_k|^2\right\} \\
&= E\left\{ e_k \frac{\partial e_k^*}{\partial w_{k,n}^*} \right\} \\
&= E\left\{ e_k \frac{\partial}{\partial w_{k,n}^*} \left(f_k - \sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,l}\, y_{k-l} \right)^* \right\} \\
&= -E\left\{ e_k\, y_{k-n}^* \right\}.
\end{aligned}\tag{A.3}
$$

The problem to be solved now simplifies to

$$E\left\{ e_k\, y_{k-n}^* \right\} = 0, \qquad n = -\left\lfloor \frac{K}{2} \right\rfloor, \dots, \left\lfloor \frac{K}{2} \right\rfloor,\tag{A.4}$$

which is known to be the orthogonality principle [1, 101]. We may further elaborate the expectation in (A.4) as follows

$$E\left\{e_k\, y_{k-n}^*\right\} \;=\; E\left\{\left(f_k - \sum_{l=-\lfloor K/2\rfloor}^{\lfloor K/2\rfloor} w_{k,l}\, y_{k-l}\right) y_{k-n}^*\right\}$$

$$\;=\; E\left\{f_k\, y_{k-n}^*\right\} - E\left\{\sum_{l=-\lfloor K/2\rfloor}^{\lfloor K/2\rfloor} w_{k,l}\, y_{k-l}\, y_{k-n}^*\right\}. \tag{A.5}$$

By using (A.5) in (A.4), we obtain

$$\sum_{l=-\lfloor K/2\rfloor}^{\lfloor K/2\rfloor} w_{k,l}\, E\left\{y_{k-l}\, y_{k-n}^*\right\} = E\left\{f_k\, y_{k-n}^*\right\}. \tag{A.6}$$

Defining $r_{yy}(\cdot)$ to be the autocorrelation of the observations, and $r_{fy}(\cdot)$ to be the cross-correlation between the fading coefficients to be estimated and the observations, (A.6) becomes the well-known linear representation for the Wiener-Hopf's equations given as

$$\sum_{l=-\lfloor K/2\rfloor}^{\lfloor K/2\rfloor} w_{k,l}\, r_{yy}(n-l) = r_{fy}(n), \quad n = -\left\lfloor\frac{K}{2}\right\rfloor, \ldots, \left\lfloor\frac{K}{2}\right\rfloor \tag{A.7}$$

where $r_{yy}(n-l) = E\left\{y_{k-l}\, y_{k-n}^*\right\}$ and $r_{fy}(n) = E\left\{f_k\, y_{k-n}^*\right\}$.

A.2 The MMSE Filter for Iterative Channel Estimation

A.2.1 Initial Estimation Iteration

In order to obtain a global MMSE filter which is independent of the transmitted symbols, we modify the channel model given in (3.1) by multiplying it with the complex conjugate of the transmitted symbols as follows

$$y_k' \;=\; a_k^*\, y_k$$

$$\;=\; f_k\, |a_k|^2 + a_k^*\, n_k$$

$$\;=\; f_k + n_k', \tag{A.8}$$

where (A.8) makes use of the fact that $|a_k|^2 = 1$, e.g., PSK modulation, and $n_k' = a_k^*\, n_k$ is the modified additive white Gaussian noise with exactly the same statistics as n_k since

$$E\left\{n_k'\right\} \;=\; a_k\, E\{n_k\} = 0$$

$$E\left\{|n_k'|^2\right\} \;=\; |a_k|^2 E\left\{|n_k|^2\right\} = E\left\{|n_k|^2\right\}.$$

As a result, the right hand side of (A.8) is observed to be independent of a_k which is the phenomena that leads to the global estimator. The Wiener-Hopf's equations corresponding to this particular case are given as

$$\sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,l}\, r_{y'y'}\left((n-l)M_p\right) = r_{fy'}(k,n) \tag{A.9}$$

for $n = -\lfloor K/2 \rfloor, \dots, \lfloor K/2 \rfloor$. In (A.9), the autocorrelation function of the modified observations is given as follows

$$
\begin{aligned}
r_{y'y'}\left((n-l)M_p\right) &= E\left\{ y'_{n_p(k)-lM_p}\left(y'_{n_p(k)-nM_p}\right)^* \right\} \\
&= E\left\{ \left(f_{n_p(k)-lM_p} + n'_{n_p(k)-lM_p}\right)\left(f_{n_p(k)-nM_p} + n'_{n_p(k)-nM_p}\right)^* \right\} \\
&= r_f\left((n-l)M_p\right) + N_0\,\delta_{nl} \tag{A.10}
\end{aligned}
$$

where $r_f(\cdot)$ is the autocorrelation of the fading coefficients and the cross-correlation function is

$$
\begin{aligned}
r_{fy'}(k,n) &= E\left\{ f_k\, y'^*_{n_p(k)-nM_p} \right\} \\
&= E\left\{ f_k\left(f_{n_p(k)-nM_p} + n'_{n_p(k)-nM_p}\right)^* \right\} \\
&= r_f\left(k - n_p(k) + nM_p\right). \tag{A.11}
\end{aligned}
$$

Consequently, the final form of the Wiener-Hopf's equations given in (A.9) become

$$\sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w_{k,l}\left(r_f\left((n-l)M_p\right) + N_0\,\delta_{nl}\right) = r_f\left(k - n_p(k) + nM_p\right) \tag{A.12}$$

for $n = -\lfloor K/2 \rfloor, \dots, \lfloor K/2 \rfloor$.

This formulation should be modified at both edges of the observation block to properly use the available pilot symbols. At the left edge of the observation block, if

$$n_p(k) - \left\lfloor \frac{K}{2} \right\rfloor M_p \le 0, \tag{A.13}$$

the set of indices of the pilot symbols to be used is then given as

$$
\begin{aligned}
P_L &= \left\{ (i-1)M_p + \frac{M_p+1}{2} \right\}_{i=1}^{K} \\
&= \left\{ \frac{M_p+1}{2}, \frac{3M_p+1}{2}, \dots, \frac{(2K-1)M_p+1}{2} \right\}. \tag{A.14}
\end{aligned}
$$

and (3.196) therefore becomes

$$\hat{f}_k^{(1)} = \sum_{n=0}^{K-1} w_{k,n}\, y'_{\frac{(2n+1)M_p+1}{2}} \,. \tag{A.15}$$

The associated Wiener-Hopf's equations are obtained by the orthogonality principle as

$$
\begin{aligned}
E\left\{ e_k \left(y'_{\frac{(2n+1)M_p+1}{2}} \right)^* \right\} &= E\left\{ \left(f_k - \hat{f}_k^{(1)} \right) y'_{\frac{(2n+1)M_p+1}{2}} \right\} \\
&= E\left\{ \left(f_k - \sum_{l=0}^{K-1} w_{k,l}\, y'_{\frac{(2l+1)M_p+1}{2}} \right) \left(y'_{\frac{(2n+1)M_p+1}{2}} \right)^* \right\}
\end{aligned}
$$

which yields

$$\sum_{l=0}^{K-1} w_{k,l}\, E\left\{ y'_{\frac{(2l+1)M_p+1}{2}} \left(y'_{\frac{(2n+1)M_p+1}{2}} \right)^* \right\} = E\left\{ f_k \left(y'_{\frac{(2n+1)M_p+1}{2}} \right)^* \right\}$$

or equivalently

$$\sum_{l=0}^{K-1} w_{k,l} \left(r_f\big((l-n)M_p\big) + N_0 \delta_{ln} \right) = r_f\left(k - \frac{(2n+1)M_p+1}{2} \right) \tag{A.16}$$

for $n = 0, 1, \ldots, K-1$.

Similarly, at the right side of the observation block, if

$$n_p(k) + \left\lfloor \frac{K}{2} \right\rfloor M_p > L, \tag{A.17}$$

the set of indices of the pilot symbols to be used is then given as

$$
\begin{aligned}
P_R &= \left\{ (i-1)M_p + \frac{M_p+1}{2} \right\}_{i=L_p-K+1}^{L_p} \\
&= \left\{ \frac{(2(L_p-K)+1)M_p+1}{2}, \ldots, \frac{(2L_p-1)M_p+1}{2} \right\},
\end{aligned} \tag{A.18}
$$

and (3.196) therefore becomes

$$\hat{f}_k^{(1)} = \sum_{n=0}^{K-1} w_{k,n}\, y'_{\frac{(2(n+L_p-K)+1)M_p+1}{2}} \,. \tag{A.19}$$

179

The associated tap weights could be evaluated using (A.16) as follows

$$\sum_{l=0}^{K-1} w_{k,l} \left(r_f \big((l-n)M_p \big) + N_0 \delta_{ln} \right) = r_f \left(k - \frac{(2(n+L_p-K)+1)M_p+1}{2} \right) \qquad \text{(A.20)}$$

for $n = 0, 1, \ldots, K-1$.

As a final remark, we observe that the optimal filter coefficients at both edges of the transmitted block depend on the transmitted symbols and has to be recomputed for each time epoch k in the specified interval.

A.2.2   Subsequent Estimation Iterations

The final expression for the Wiener-Hopf's equations for subsequent channel estimation iterations is a modified version of (3.6). Let us consider the autocorrelation function computed in a genie-aided fashion as follows

$$\begin{aligned} r_{yy}(n-l) &= E\left\{ y_{k-l}\, (y_{k-n})^* \right\} \\ &= a_{k-l}\, r_f(n-l)\, a^*_{k-n} + N_0\, \delta_{nl}, \end{aligned} \qquad \text{(A.21)}$$

and the cross-correlation function given as

$$\begin{aligned} r_{fy}(n) &= E\left\{ f_k\, y^*_{k-n} \right\} \\ &= r_f(n)\, a^*_{k-n}. \end{aligned} \qquad \text{(A.22)}$$

Because we do not have the actual values of the data symbols and only have the corresponding estimates $\hat{a}^{(i)}_k$, (A.21) and (A.22) are modified to include these estimates and the final Wiener-Hopf's equations for the i-th iteration therefore become

$$\sum_{l=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} w^{(i)}_{k,l} \left\{ \hat{a}^{(i)}_{k-l}\, r_f(n-l)\, (\hat{a}^{(i)}_{k-n})^* + N_0\, \delta_{nl} \right\} = r_f(n)\, (\hat{a}^{(i)}_{k-n})^*, \qquad \text{(A.23)}$$

for $n = -\lfloor K/2 \rfloor, \ldots, \lfloor K/2 \rfloor$.

At the left edge of the block, the set of indices of the observations to be employed in channel estimation for $k - \left\lfloor \frac{K}{2} \right\rfloor \le 0$ is given as

$$P_L = \{1, 2, \ldots, K\}, \qquad \text{(A.24)}$$

and the estimation operation in (3.201) becomes

$$\hat{f}_k^{(i)} \quad = \quad \sum_{n=0}^{K-1} w_{k,n}^{(i)} \, y_{n+1}. \tag{A.25}$$

Similarly, the set of indices at the right edge of the block for $k + \left\lfloor \frac{K}{2} \right\rfloor > L$ is

$$P_R \quad = \quad \{L - K + 1, \, L - K + 2, \, \dots \, , L\}, \tag{A.26}$$

and the desired estimates are computed according to

$$\hat{f}_k^{(i)} \quad = \quad \sum_{n=0}^{K-1} w_{k,n}^{(i)} \, y_{n+L-K+1}. \tag{A.27}$$

At both left and right edges of the block, coefficients of the desired MMSE filter are computed by using (A.23) with adequate indices given by (A.24) and (A.26), respectively.

Appendix B

Derivation of the MMSE Filter for Frequency-Selective Fading Channels

B.1 The MMSE Filter for Perfectly Known Transmitted Symbols

Let us consider the estimation error vector associated with the estimation problem introduced in (4.5) given as

$$\mathbf{e} = \mathbf{f} - \hat{\mathbf{f}} \tag{B.1}$$

which results in the following MSE

$$J \quad = \quad E\left\{\|\mathbf{e}\|^2\right\} = E\left\{\mathbf{e}^H \mathbf{e}\right\} \tag{B.2}$$

$$= \quad E\left\{\mathrm{tr}\left(\mathbf{e}\,\mathbf{e}^H\right)\right\}. \tag{B.3}$$

Because the ultimate goal of the MMSE estimator is to minimize the associated MSE, the derivative of (B.3) has to be computed with respect to the complex-valued matrix $\mathbf{W}$ to find the global minima as follows

$$\frac{\partial J}{\partial \mathbf{W}} \quad = \quad \frac{\partial}{\partial \mathbf{W}} E\left\{\mathrm{tr}\left(\left(\mathbf{f} - \hat{\mathbf{f}}\right)\left(\mathbf{f} - \hat{\mathbf{f}}\right)^H\right)\right\}$$

$$= \quad E\left\{\frac{\partial}{\partial \mathbf{W}} \mathrm{tr}\left((\mathbf{f} - \mathbf{W}\mathbf{y})\,(\mathbf{f} - \mathbf{W}\mathbf{y})\,)^H\right\}$$

$$= \quad E\left\{\frac{\partial}{\partial \mathbf{W}} \mathrm{tr}\left(\mathbf{f}\mathbf{f}^H - \mathbf{f}\mathbf{y}^H\mathbf{W}^H - \mathbf{W}\mathbf{y}\mathbf{f}^H + \mathbf{W}\mathbf{y}\mathbf{y}^H\mathbf{W}^H\right)\right\}. \tag{B.4}$$

Using the results of [127, 128] on derivative computation with respect to the complex-valued

multi-dimensional input parameters, the desired derivative could be evaluated as follows

$$
\begin{aligned}
\frac{\partial J}{\partial \mathbf{W}} &= E\left\{-\frac{\partial}{\partial \mathbf{W}} \operatorname{tr}\left(\mathbf{f}\,\mathbf{y}^H \mathbf{W}^H\right) - \frac{\partial}{\partial \mathbf{W}} \operatorname{tr}\left(\mathbf{W}\,\mathbf{y}\,\mathbf{f}^H\right) + \frac{\partial}{\partial \mathbf{W}} \operatorname{tr}\left(\mathbf{W}\,\mathbf{y}\,\mathbf{y}^H \mathbf{W}^H\right)\right\} \\
&= E\left\{-\mathbf{f}^*\,\mathbf{y}^T + \mathbf{W}^*\left(\mathbf{y}\,\mathbf{y}^H\right)^T\right\} \\
&= -\left(E\left\{\mathbf{f}\,\mathbf{y}^H\right\}\right)^* + \left(\mathbf{W}E\left\{\mathbf{y}\,\mathbf{y}^H\right\}\right)^* \tag{B.5}
\end{aligned}
$$

where we have used the following identities from [128]

$$
\begin{aligned}
\frac{\partial \operatorname{tr}\left(\mathbf{A}\,\mathbf{X}^H\right)}{\partial \mathbf{X}} &= \mathbf{0} \\
\frac{\partial \operatorname{tr}\left(\mathbf{A}\,\mathbf{X}\right)}{\partial \mathbf{X}} &= \mathbf{A}^T \\
\frac{\partial \operatorname{tr}\left(\mathbf{X}\,\mathbf{A}\right)}{\partial \mathbf{X}} &= \mathbf{A}^T \\
\frac{\partial \operatorname{tr}\left(\mathbf{X}\,\mathbf{A}\,\mathbf{X}^H\right)}{\partial \mathbf{X}} &= \mathbf{X}^*\,\mathbf{A}^T.
\end{aligned}
$$

The optimal MMSE filter is the root of (B.5) and is therefore given as

$$
\mathbf{W} = \mathbf{P}_{fy}\,\mathbf{R}_{yy}^{-1} \tag{B.6}
$$

where $\mathbf{P}_{fy} = E\left\{\mathbf{f}\,\mathbf{y}^H\right\}$ and $\mathbf{R}_{yy} = E\left\{\mathbf{y}\,\mathbf{y}^H\right\}$.

B.2   The MMSE Filter for Iterative Channel Estimation

The initial estimation matrix $\mathbf{W}^{(1)}$ could be computed using (4.6) together with the modified cross-correlation and the autocorrelation matrices given as

$$
\begin{aligned}
\mathbf{P}_{fy} &= E\left\{\mathbf{f}\,\mathbf{y}_p^H\right\} \tag{B.7} \\
\mathbf{R}_{yy} &= E\left\{\mathbf{y}_p\,\mathbf{y}_p^H\right\}. \tag{B.8}
\end{aligned}
$$

In order to compute (B.7)-(B.8), we first express the observation vector $\mathbf{y}_p$ as follows

$$
\begin{aligned}
\mathbf{y}_p &= \underbrace{\begin{bmatrix} \mathbf{a}_{p(1)}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_{p(2)}^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{a}_{p(L_p)}^T \end{bmatrix}}_{\mathbf{A}_p} \underbrace{\begin{bmatrix} \mathbf{f}_{p(1)} \\ \mathbf{f}_{p(2)} \\ \vdots \\ \mathbf{f}_{p(L_p)} \end{bmatrix}}_{\mathbf{f}_p} + \underbrace{\begin{bmatrix} \mathbf{n}_{p(1)} \\ \mathbf{n}_{p(2)} \\ \vdots \\ \mathbf{n}_{p(L_p)} \end{bmatrix}}_{\mathbf{n}_p} \\
&= \mathbf{A}_p\,\mathbf{f}_p + \mathbf{n}_p \tag{B.9}
\end{aligned}
$$

Inserting (B.9) into (B.7)-(B.8) yields

$$\begin{aligned} \mathbf{P}_{fy} &= E\left\{ \mathbf{f} \left(\mathbf{A}_p \mathbf{f}_p + \mathbf{n}_p \right)^H \right\} \\ &= E\left\{ \mathbf{f} \mathbf{f}_p^H \right\} \mathbf{A}_p^H \end{aligned}$$

(B.10)

$$\begin{aligned} \mathbf{R}_{yy} &= E\left\{ \left(\mathbf{A}_p \mathbf{f}_p + \mathbf{n}_p \right) \left(\mathbf{A}_p \mathbf{f}_p + \mathbf{n}_p \right)^H \right\} \\ &= \mathbf{A}_p E\left\{ \mathbf{f}_p \mathbf{f}_p^H \right\} \mathbf{A}_p^H + N_0 \mathbf{I}. \end{aligned}$$

(B.11)

In order to obtain the final expression for the initial estimation matrix, (B.10) and (B.11) are incorporated into (4.6) as follows

$$\mathbf{W}^{(1)} = \mathbf{P}_p \mathbf{A}_p^H \left(\mathbf{A}_p \mathbf{R}_p \mathbf{A}_p^H + N_0 \mathbf{I} \right)^{-1}$$

(B.12)

where $\mathbf{P}_p = E\left\{ \mathbf{f} \mathbf{f}_p^H \right\}$ and $\mathbf{R}_p = E\left\{ \mathbf{f}_p \mathbf{f}_p^H \right\}$.

In order to compute $\mathbf{P}_{fy}$ and $\mathbf{R}_{yy}$ in the subsequent estimation iterations, we assume

$$\mathbf{y} \approx \underbrace{\begin{bmatrix} \left(\hat{\mathbf{a}}_1^{(i)}\right)^T & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \left(\hat{\mathbf{a}}_2^{(i)}\right)^T & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \left(\hat{\mathbf{a}}_L^{(i)}\right)^T \end{bmatrix}}_{\hat{\mathbf{A}}^{(i)}} \underbrace{\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_L \end{bmatrix}}_{\mathbf{f}} + \underbrace{\begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_L \end{bmatrix}}_{\mathbf{n}}$$

$$= \hat{\mathbf{A}}^{(i)} \mathbf{f} + \mathbf{n},$$

(B.13)

and then perform computations as follows

$$\begin{aligned} \mathbf{P}_{fy} &= E\left\{ \mathbf{f} \mathbf{y}^H \right\} \\ &= E\left\{ \mathbf{f} \mathbf{f}^H \right\} \left(\hat{\mathbf{A}}^{(i)} \right)^H \end{aligned}$$

(B.14)

$$\begin{aligned} \mathbf{R}_{yy} &= E\left\{ \mathbf{y} \mathbf{y}^H \right\} \\ &= \hat{\mathbf{A}}^{(i)} E\left\{ \mathbf{f} \mathbf{f}^H \right\} \left(\hat{\mathbf{A}}^{(i)} \right)^H + N_0 \mathbf{I}. \end{aligned}$$

(B.15)

where $\mathbf{R}_f = E\left\{ \mathbf{f} \mathbf{f}^H \right\}$. The corresponding MMSE filter at the i-th iteration is then given as

$$\mathbf{W}^{(i)} = \mathbf{R}_f \left(\hat{\mathbf{A}}^{(i)} \right)^H \left(\hat{\mathbf{A}}^{(i)} \mathbf{R}_f \left(\hat{\mathbf{A}}^{(i)} \right)^H + N_0 \mathbf{I} \right)^{-1}.$$

(B.16)

VITA

Yavuz Yapıcı was born in Zonguldak, Turkey in 1980. He received B.S. and M.S. degrees both in Department of Electrical and Electronics Engineering from Bilkent University in 2002 and 2005, respectively, with full scholarship. He has been working for Defense Technologies Engineering (STM A.Ş.) as a consultant since 2008.

His professional research is primarily on the communication systems with a special emphasis on the wireless technologies. Symbol detection and parameter estimation problems for highly mobile MIMO-OFDM systems, cooperative communication technologies and underwater acoustic (UWA) communication systems are some of the principle research subjects that are interest to him.

His publications are as follows:

1. Y. Yapıcı and A. Ö. Yılmaz, "A Bidirectional LMS Algorithm and Associated MSE Analysis over Time-Varying Channels", *to be submitted to IEEE Transactions on Signal Processing*.

2. Y. Yapıcı and A. Ö. Yılmaz, "Low-complexity iterative channel estimation and tracking for time-varying multi-antenna systems", in *Proc. of IEEE PIMRC'09*, Tokyo, Japan, Sep. 13-16, 2009.

3. Y. Yapıcı and A. Ö. Yılmaz, "Joint channel estimation and decoding with low-complexity iterative structures in time-varying fading channels", in *Proc. of IEEE PIMRC'09*, Tokyo, Japan, Sep. 13-16, 2009.

4. Y. Yapıcı and A. Ö. Yılmaz, "Turbo-like codes over block-fading and time-varying narrowband fading channels", in *Proc. of IEEE VTC'09*, Barcelona, Spain, Apr. 26-29, 2009.

5. E. Arıkan and Y. Yapıcı, "An Improvement of the V-BLAST Symbol Detection Algorithm for MIMO Channels", in *Proc. of XI International Symposium on Problems of*

185