Semi-blind Spatial-temporal Equalization for Short Burst Wireless Communications

by

Kelvin Kar-Kin Au

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science Graduate Department of Electrical and Computer Engineering University of Toronto

© Copyright by Kelvin Kar-Kin Au. 2000
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-53423-5
Semi-blind Spatial-temporal Equalization for Short Burst Wireless Communications

Kelvin Kar-Kin Au
Master of Applied Science, 2000
Graduate Department of Electrical and Computer Engineering
University of Toronto

Abstract

In short burst wireless communications, a training sequence is incorporated in each burst for the receiver to adjust the equalizer coefficients. However, when the amount of training symbols is less than the spatial-temporal equalizer tap weights, conventional least-square technique may not provide good MSE performance. Blind methods, on the other hand, may not achieve equalization in a short burst. Semi-blind algorithms have been proposed as alternative solutions, yet these algorithms may still exhibit undesirable local minima.

Two algorithms are proposed to address the problem. The first method attempts to modify the regularized semi-blind algorithm proposed previously by Kuzminskiy et al. However simulation results show that the MSE performance of the semi-blind algorithm is not superior than the original one under the short burst scenario. A convex cost with training symbols as the equalizer constraint is then proposed to avoid cost-dependent local minima completely. Furthermore, comparison with the regularized semi-blind algorithm suggests that this proposed algorithm achieves a lower MSE performance in the case of non-constant modulus signals such as 16-QAM signals.
Acknowledgements

I would like to thank my supervisor Professor Dimitrios Hatzinakos for his direction, advice and support throughout the course of my thesis.

Special thanks to my family and friends. They are a continuous source of support during the two years of my graduate studies. I would also like to thank my good friend Charlton Wang. He has offered me encouragements and kept me in perspective despite the many things that happened during the course of these two years.
Contents

Abstract ii

Acknowledgements iii

List of Tables vii

List of Figures x

1 Introduction 1
  1.1 Wireless Communications .................................. 1
  1.2 Current Methods for Equalization ............................. 3
  1.3 Short Burst Scenario ...................................... 4
  1.4 Objective of the Thesis ..................................... 5
    1.4.1 A Note about Notation ................................ 8

2 Blind and Semi-blind Equalization Algorithms 9
  2.1 Spatial-temporal Filter ................................... 9
  2.2 Spatial-temporal Signal Model for Short Burst Communications .... 11
  2.3 Conventional Training-based Method ........................ 13
  2.4 Constant Modulus Cost for Blind Equalization ................. 15
    2.4.1 CM $p - q$ Cost .................................. 15
5 Conclusions and Future Research

5.1 Concluding Remarks

5.2 Direction of Future Research

61

62

A Gradient of the Semi-blind Crime Cost

64

B Hessian of the Semi-blind Crime Cost

66

C Gradient of the Semi-blind Convex Cost

68

D Recursive Algorithms

70

D.1 Gauss-Newton Method

70

D.2 Recursive Gradient Descent Method

71

Bibliography

72
List of Tables

4.1 A summary of the semi-blind convex algorithm. .................. 52
List of Figures

1.1 A diagram of several mobiles transmitting to a basestation receiver. . . . 3
1.2 An example of a burst structure: a GSM burst. . . . . . . . . . . . . . 5
2.1 Signal model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
2.2 Contour plot of the CM 2-2 cost for a two-tap equalizer. . . . . . . 17
2.3 3D plot of the CM 2-2 cost for a two-tap equalizer. . . . . . . . . . 18
2.4 Contour plot of the CM 2-2 cost for a 16-QAM signal. . . . . . . . . 21
2.5 3D plot of the CM 2-2 cost for a 16-QAM signal. . . . . . . . . . . . 21
2.6 Contour plot of the semi-blind regularized cost for a two-tap equalizer. . 26
2.7 3D plot of the semi-blind regularized cost for a two-tap equalizer. . . . 26
3.1 MSE vs. \( N \), plot of semi-blind regularized and semi-blind CRIMO algorithms for 4-QAM signals (\( N = 150, K = 3, \rho = 1, \alpha = 0.5, M_\delta = 7 \)). . . 37
3.2 Plot of the number of iterations for semi-blind regularized and semi-blind CRIMO algorithms when the stopping condition is reached for 4-QAM signals (\( N = 150, K = 3, \rho = 1, \alpha = 0.5, M_\delta = 7 \)). . . . . . . . . . 38
3.3 MSE vs. \( N \), plot of semi-blind regularized and semi-blind CRIMO algorithms for 16-QAM signals (\( N = 150, K = 3, \rho = 1, \alpha = 0.5, M_\delta = 7 \)). . 38
3.4 Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for 16-QAM signals \((N = 150, K = 3, \rho = 1, \alpha = 0.5, M_\delta = 7)\).

3.5 MSE vs. \(N_t\) plot of semi-blind regularized and semi-blind CRIMNO algorithms when \(K = 1\) (16-QAM signals. \(N = 150, \rho = 1, \alpha = 0.5, M_\delta = 7\)).

3.6 Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for \(K = 1\) (16-QAM signals. \(N = 150, \rho = 1, \alpha = 0.5, M_\delta = 7\)).

3.7 MSE vs. \(N_t\) plot of semi-blind regularized and semi-blind CRIMNO algorithms for a 64-QAM signal \((K = 1, N = 150, \rho = 1, \alpha = 0.5, M_\delta = 7)\).

3.8 Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for a 64-QAM signal \((K = 1, N = 150, \rho = 1, \alpha = 0.5, M_\delta = 7)\).

3.9 Signal constellation at the equalizer output for a 64-QAM signal using the semi-blind CRIMNO algorithm when \(N_t = 22\) for a realization. It demonstrates that the eye is not opened. The MSE is 0.0513.

3.10 MSE vs. \(N_t\) plot of the semi-blind CRIMNO algorithms for different \(M_\delta\) \((K = 3, N = 150, \rho = 1, \alpha = 0.5, 16\text{-QAM signals})\).

3.11 Plot of the number of iterations for the semi-blind CRIMNO algorithms when the stopping condition is reached for different \(M_\delta\) \((K = 3, N = 150, \rho = 1, \alpha = 0.5, 16\text{-QAM signals})\).

4.1 Plot of MSE vs. \(N_t\) for the semi-blind convex algorithm using different \(p\) \((K = 3, 16\text{-QAM signals. }N = 200)\).

4.2 4-QAM case: MSE vs. \(N_t\) for the pure LS, convex blind, convex semi-blind and regularized semi-blind algorithms \((K = 3, N = 150)\).
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>16-QAM case: MSE vs. ( N_t ) for the pure LS, convex blind, convex semi-blind and regularized semi-blind algorithms (( K = 3, N = 200 )).</td>
</tr>
<tr>
<td>4.4</td>
<td>Constellation for the desired user before equalization (( N = 200, 16\text{-QAM} )).</td>
</tr>
<tr>
<td>4.5</td>
<td>Constellation for the desired user after equalization using the proposed convex semi-blind algorithm (( N = 200, 16\text{-QAM} )).</td>
</tr>
<tr>
<td>4.6</td>
<td>Plot of MSE vs. ( N_t ) when ( K = 1 ) for the convex semi-blind and regularized semi-blind algorithms (( N = 200, 16\text{-QAM signals} )).</td>
</tr>
<tr>
<td>4.7</td>
<td>Plot of MSE vs. ( N_t ) when ( N = 150 ) for the convex semi-blind and regularized semi-blind algorithms (( K = 3, 16\text{-QAM signals} )).</td>
</tr>
<tr>
<td>4.8</td>
<td>Plot of MSE vs. ( N_t ) when ( N = 300 ) for the convex semi-blind and regularized semi-blind algorithms (( K = 3, 16\text{-QAM signals} )).</td>
</tr>
<tr>
<td>4.9</td>
<td>Plot of MSE vs. ( N_t ) with ( N = 150, 200, 300 ) for the convex semi-blind algorithm (( K = 3, 16\text{-QAM signals} )).</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Wireless Communications

Wireless communications have emerged in recent years as one of the fastest growing areas in communications, surpassing the traditional wireline systems. Ever since the first deployment of commercial wireless phone systems, there is an increasing demand for higher capacity and better quality. The first generation mobile systems utilize analog modulation techniques similar to that of the radio. In the late 1980s and early 1990s, the traditional analog systems were gradually replaced by second generation digital systems that offer higher capacity and quality. Digital modulation schemes are employed in such systems and multiple access techniques such as Time Division Multiple Access (TDMA) and more recently, Code Division Multiple Access (CDMA) are used. With the finalization of the third generation standard, called the International Mobile Telephone 2000 (IMT 2000), the objective is to bring voice and data together so that people can use their mobile phone for a variety of purposes, such as voice communications, receive video and surf the Internet [1]. Such systems will certainly demand even greater capacity than the current wireless standards offer. Moreover, the quality of voice and data transmission
are paramount. These certainly provide many challenges and opportunities for improving signal processing algorithms in wireless communications.

In the 3rd-generation wireless systems and beyond, there are a number of techniques which shall provide performance improvement over current systems. Two of them, namely the use of antenna array and possible utilization of higher signal constellations, are closely related to the signal processing aspect of system design. They offer opportunities for improvements in signal processing algorithms. In [1], an overview of wireless technologies from 1st-generation to 3rd generation is given. In migrating towards the 3rd-generation, backward compatibility is a key so that service providers can gradually deploy new systems and avoid the need to switch resources (such as bandwidth and hardware configurations) overnight. The FRAMES Multiple Access (FMA) proposal is designed exactly for this purpose [4]. Many of the system parameters such as frame length or chip rate (for CDMA systems) are either kept the same or increased in multiples of the current values.

In wireless systems, signals are transmitted to the atmosphere which has buildings and other structures in the environment. This scenario results in the receiver receiving multiple copies of the transmitted signal with different delays and attenuations upon travelling different paths due to reflection. This multipath propagation causes intersymbol interference (ISI). Also, in a cellular system using TDMA, each user in a cell is allocated a time slot at a certain carrier frequency. Each user’s slot contains both training and information symbols. Since frequency spectrum can be regarded as a scarce resource in wireless communications, one frequency allocated in a cell will be re-used in another cell some distance apart. This is to avoid too strong an interference. Nonetheless, when several users transmitting at the same frequency from different cells arrive at the receiver (for example, the basestation receiver), the received signal is a sum of the desired user’s signal and other users (interference) plus noise. The interferers cause co-channel inter-
1.2. Current Methods for Equalization

There are three main types of signal processing schemes that can be used for equalization in wireless communications. They are methods which require training sequences, blind algorithms and more recently semi-blind algorithms.

Training-based methods are widely used in current receivers for wireless communications due to its simplicity. Known symbols are sent by the basestations or cellular phones to the destination receivers. At the receivers, the received symbols are than compared to the known symbols. An equalizer is used to minimize the mean square error (MSE) between the received symbols and desired symbols. There are many algorithms for the
minimization of the MSE cost. Most popular methods are the Least Mean Square (LMS), Method of Least Square (LS) and Recursive Least Square (RLS), to name but a few [5]. In recent years, many blind algorithms emerge in the literature as an alternative way to process the received signal. Blind algorithms do not require training sequences. Instead, they utilize information such as the statistics of the signal constellations and the property of the digital modulation (e.g., constant envelope) to achieve equalization [8]. One of the most popular blind cost functions is the constant modulus (CM) cost [9, 12]. However the CM cost function exhibits local minima which may result in "wrong solutions" to the problem of equalization. Another drawback is the slow convergence of algorithms using the CM cost. These points will be discussed in detail later in the thesis. More recently, another class of algorithms called "semi-blind" are proposed [10][14]–[23]. Semi-blind methods, as the name suggests, utilize the existence of known symbols and data symbols in the received signal for equalization. Simulation results have shown that this type of methods is applicable in situations where only a small block of data is available with few training symbols.

1.3 Short Burst Scenario

In short burst communications, a block of a few hundred symbols is transmitted and received at a time (example systems such as GSM [2] and PACS [3]). Each user's data block includes a training sequence which is used for equalization at the receiver (or for channel estimation). In this thesis, we assume that the position of the training sequence is known at the receiver. The duration of the block of data is short compared to the channel coherence time. Therefore, the channel does not vary significantly over the duration of the burst and can be regarded as stationary. For example, in GSM systems [2] the duration of the time slot for a user is 0.577 ms which is much smaller than a typical
value of channel coherence time of approximately 5 ms (when the velocity of the mobile is around 100 km/h) [25]. A GSM data burst contains approximately 150 symbols [2] with 26 training symbols in the middle of the burst. Figure 1.2 shows a diagram of a GSM burst.

![Diagram of a GSM burst](image)

Figure 1.2: An example of a burst structure: a GSM burst.

At the receiver, adaptive or recursive equalization algorithms can be employed. Adaptive algorithms are suitable in the case when the channel is varying over the duration of the received data. For channels which are assumed to be stationary, the availability of a block of data can be used as the input to batch-type algorithms. For instance, the training symbols in a GSM burst is used at the receiver for estimating the channel using the method of Least Square [2].

### 1.4 Objective of the Thesis

Conventional equalization algorithms using training sequences are simple to implement. However, the need to transmit long training sequences repeatedly represents a system overhead and effectively reduces the information rate. Blind equalization algorithms such as the constant modulus algorithm, on the other hand, experience slow convergence and the possibility of converging to local minima. Simulations have shown that equalization
is successful when around 10,000 symbols are available in the burst of data. Otherwise, the resultant MSE is still very high [14] when only a few hundred symbols are available. Moreover, for the 3rd generation systems and beyond, higher signal constellations will be used. In this case, the CM cost will have excess error even upon equalization (as will be discussed later).

A regularized semi-blind algorithm was proposed in [14] which combined the LS and CM 1-2 costs. The ability to successfully equalize the channels with a spatial-temporal filter was demonstrated when a small block of data is received under a stationary environment. This algorithm offers the possibility of reducing the number of training symbols in a short burst of data. However, local minima inherent to the cost exist.

The objective of this thesis is as follows: we shall outline and examine the behaviour of CM-type blind and semi-blind algorithms and propose ways to reduce or even eliminate the problems described earlier, that is, 1) to eliminate or reduce the possibility of convergence to local minima of a cost function and 2) to achieve better equalizer performance in terms of the MSE between the output and the transmitted symbols under a short burst of data for constant and non-constant modulus signals (such as 16-QAM) in the case when the number of training symbols is fewer than the number of spatial-temporal equalizer coefficients.

The contributions of the thesis are as follows:

1. In Chapter 2, we examine the behaviour of the CM cost surface and the regularized semi-blind cost that combines the LS and CM 1-2 costs [14]. We show that, by means of a simulation, the estimated cost surface of the regularized semi-blind cost exhibits local minima. This is due to the fact that the CM 1-2 cost has cost-dependent local minima.

2. In Chapter 3, a modification to the regularized semi-blind algorithm is proposed by
including additional penalizing terms on the statistical (in)dependence of the equalizer outputs. A block implementation of the algorithm using the Gauss-Newton method for the short burst scenario is formulated. However, simulation results demonstrate that the proposed modification does not provide superior MSE performance over the regularized semi-blind algorithm in the short burst environment with few training symbols for the case of constant and non-constant modulus signals.

3. In Chapter 4, a convex cost for blind equalization is adopted with a proposed training sequence constraint. The use of a convex cost eliminates local minima. A block implementation of the semi-blind algorithm using a gradient descent recursion is used to obtain the equalizer coefficients. We conclude that the convex semi-blind algorithm provides a better MSE performance for non-constant modulus signals such as 16-QAM than the semi-blind regularized algorithm when only a short burst of data is available. However, the latter method works better for 4-QAM signals. Both algorithms outperform the pure LS solution in the case when the number of training symbols is fewer than the spatial-temporal equalizer coefficients.

The organization of the thesis is as follows: in Chapter 2, we shall describe in more detail existing methods on equalization with emphasis on blind CM and semi-blind algorithms. An overview of spatial-temporal equalization is also discussed in relation to the effect it has on blind CM cost surfaces. The signal model that is used throughout this thesis is also described. In Chapter 3, we shall propose a modification to the regularized semi-blind cost with extra penalizing terms and discuss the simulation results. Chapter 4 studies another proposed semi-blind convex cost which offers some promising results. Finally, we shall provide some concluding remarks and directions of future research in Chapter 5.
1.4.1 A Note about Notation

Throughout this work, we shall use bold capital letter (eg. $A$) to denote a matrix. All vectors are denoted by bold lowercase letter (eg. $a$). Unless otherwise stated, all vectors are column vectors.
Chapter 2

Blind and Semi-blind Equalization Algorithms

2.1 Spatial-temporal Filter

Before we move on to discuss various blind and semi-blind equalization algorithms, an overview of spatial-temporal filtering is given in this section. As mentioned in the introductory chapter, antenna array will be adopted gradually in the next generation wireless systems. Although having an antenna array in a mobile terminal (such as a cellular phone) seems to be impossible, installing them at the basestations is a good way to improve performance of the systems.

There are primarily three advantages in using an antenna array: 1) it provides spatial diversity at the receiver, 2) antenna beamforming is possible and 3) from a system modelling point of view, using an antenna array is equivalent to fractionally-spaced sampling at the receiver and this has shown to affect the CM cost minima.

When the sensors of the antenna array are spaced out sufficiently, the received signal can be thought of as independent and uncorrelated with each other. Therefore we
2.1. *Spatial-temporal Filter*

can exploit this spatial diversity to provide a better signal-to-noise ratio (SNR) at the receiver [7]. Also if one of the sensors receives the signal which is deeply faded in its amplitude, the signal that arrives in another sensor may not be in a deep fade. Hence, it is advantageous to use an antenna array at the receiver.

The second advantage mentioned before is that antenna beamforming is possible. With the use of spatial tap weights at the receiver, the array of antenna is able to form a beam towards a particular direction and null out signals coming from other angles. This is particularly useful in situation where the desired signal is arriving at the basestation with other interferers. Since these signals will be arriving at different angles, the antenna is able to form the beam to suppress the interferers. Moreover, in situation where a direct path from the desired user is present, the antenna may be able to direct the beam towards the user and null the other multiple copies of the desired signal due to multipath propagation. In this case, an equalizer may not even be necessary at the receiver. In this thesis, we do not deal with the issue of antenna beamforming.

The third advantage of using an antenna array can be viewed from a system modelling point of view. The main result is as follows [8, 24, 25, 26]: If $T$ is the symbol duration, a system that is $T/P$-spaced is equivalent to a system with $P T$-spaced sensors. That means we can model a $T/P$-spaced channel with $P T$-spaced subchannels. Hence this multirate and multichannel models are "interchangeable" in this case. Therefore, spatial diversity at the receiver can be used to improve its performance in cases where excess bandwidth is not available. This technique has been proposed in [43] for GSM system when fractionally-spaced sampling is not feasible due to a small excess bandwidth. We shall discuss in Section 2.4.2 the effect of using antenna arrays on the CM cost surface.

In this thesis, a generic *multiple-input multiple-output (MIMO)* model is used. Multiple-input refers to multiple users in the signal model while multiple-output refers to the use of an antenna array at the receiver. The evaluation of the advantages and performance
gain by using an antenna array is not within the scope of this work.

2.2 Spatial-temporal Signal Model for Short Burst Communications

In this section, the spatial-temporal model is defined. This signal model incorporates what we have discussed so far: the presence of ISI and CCI and the use of spatial-temporal filter at the receiver for equalization. We assume that there are \( K \) users in the model. One of the user is the signal of interest. Without loss of generality, we shall denote the first user to be the desired signal. The remaining \( K - 1 \) signals are coming from nearby co-channel cells. At the base station receiver, an antenna array of \( M \) sensors is employed. Figure 2.1 depicts the scenario under study.

![Signal model diagram](image)

Figure 2.1: Signal model.

The following assumptions are made:
1. The data is processed in a burst of $N$ symbols which are assumed to be received under a stationary environment. This assumption is valid when the duration of the block of received data is less than the channel coherence time (e.g. in the GSM scenario described in Chapter 1).

2. There are $N_t$ training symbols in each burst and the starting position of the training sequence is $N$, which is assumed to be known.

3. The transmitted signals undergo linear channels which are assumed to be FIR of length $N_c$. This assumption is valid when we have a finite delay spread.

The channels $c_{ij}$'s capture the effect from the transmitter filter to the receiver before the equalizer. Equalization is necessary when the delay spread is larger than the symbol duration. The received signal at the $j$-th sensor is given by:

$$ y_j(n) = \sum_{i=1}^{K} c_{ij}^H x_i(n) + v_j(n) $$

for $i = 1, \ldots, K, j = 1, \ldots, M$

$$ c_{ij} = [c_{ij}(0) \ldots c_{ij}(N_c - 1)]^T. $$

$$ x_i(n) = [x_i(n) \ldots x_i(n - N_c + 1)]^T. $$

where $H$ denotes the conjugate transpose of a matrix and each $c_{ij}(n)$ is a complex Gaussian random variable whose amplitude (Rayleigh distributed) does not change over the duration of the burst. The noise $v_j(n)$ is a complex circularly symmetric additive white Gaussian noise of variance $\sigma_n^2$.

Recall that the first user is the desired signal. The equalizer output for the signal of interest is given by:

$$ z_1(n) = w^H y(n), $$

(2.4)
where \( y(n) = [y_1^T(n), \ldots, y_l^T(n)]^T \) and each \( y_j(n) = [y_j(n), \ldots, y_j(n - N_w + 1)]^T \). The spatial-temporal equalizer taps are \( w = [w_1^T, \ldots, w_l^T]^T \) and each \( w_j = [w_{j1}, \ldots, w_{jN_w}]^T \). The vector \( w \) has a dimension of \( MN_w \times 1 \).

### 2.3 Conventional Training-based Method

Training based methods are the most widely used methods in wireless communications. The training sequence is known at the receiver and the receiver utilizes this known information as the desired response. The channel parameters and the estimated signal can be obtained by the classical Wiener filtering problem \([5]\). The coefficients of the filter are adjusted to minimize the MSE between the output and the desired signal (in this case, the training data):

\[
\hat{w} = \min_w E \left[ |w^H y(n) - x_1(n - d)|^2 \right]
\]

where \( \hat{w} \) is a column vector of filter coefficients, \( y \) is a column vector of inputs to the filter, \( x_1(n - d) \) is the desired response for some delay \( d \) and \( H \) denotes the conjugate transpose of a vector. The solution to the above is given by the \textit{Wiener-Hopf Equation}:

\[
\hat{w} = R^{-1} p
\]

where \( R = E[yy^H] \) is the autocorrelation matrix of the input to the filter and \( p = E[yx_1^*(n - d)] \) is the cross-correlation vector between the input and the desired signal.

There are many variations in the implementation of the above cost. Least Mean Square \([6]\) is the most popular adaptive algorithm in minimizing the MSE cost. In the context of short burst communications where a burst of data is available at the receiver under a stationary environment, we can adopt the block approach. The method of least square (LS) is an off-line algorithm that replaces the expectation operator with a time
2.3. Conventional Training-based Method

average. The autocorrelation matrix is replaced by a time-average autocorrelation matrix over the length of the training sequence and is given by:

$$\hat{R} = \frac{1}{N_t} \sum_{n=n_s}^{n_s+N_t-1} y(n)y^H(n)$$  \hspace{1cm} (2.7)

where \(n_s\) is the position of the starting training sequence and \(N_t\) is the length of the training. Similarly, the cross-correlation is given by:

$$\hat{p} = \frac{1}{N_t} \sum_{n=n_s}^{n_s+N_t-1} y(n)x^*_1(n-d).$$  \hspace{1cm} (2.8)

The LS solution is thus given by:

$$w = \hat{R}^{-1}\hat{p}.$$  \hspace{1cm} (2.9)

In applying the above time averaging, we should note that the resulting autocorrelation matrix \(\hat{R}\) is a square matrix of size \(M.N \times M.N\) and is full rank. Otherwise, one will not be able to invert \(\hat{R}\). This is the case when the number of training symbols is fewer than the length of the received filter. When this situation arises, a technique called pseudo-inverse can be used to obtain a solution [5]. This idea will be elaborated more in Chapter 4. Moreover, it will be shown by simulation in Chapter 4 that the performance of the method of LS is poor under this condition.

Training based methods are popular because they are well understood and relatively easy to implement. The desired signal (i.e. training sequence) is always available at the receiver. However, one of the problems is that incorporating a long training sequence in each burst of data is a severe overhead. Increasing the length of a training sequence will mean a decrease in the capacity of the system. There are some recent developments in modifying existing training algorithms in order to provide a better estimation of the input autocorrelation and the cross-correlation between the input and the desired signal. Readers are encouraged to review [10] and references therein for a more in-depth look at the issue.
2.4 Constant Modulus Cost for Blind Equalization

2.4.1 CM $p – q$ Cost

As opposed to training-based methods, blind equalization does not utilize any known symbols. Therefore, if a blind equalization algorithm is employed at the receiver, a transmitter does not need to incorporate a training sequence when sending a signal. However, we can look at the situation in another perspective: if a signal being sent contains some known symbols, the receiver will not be able to utilize this information since it will process the signal "blindly".

One of the most popular algorithms for blind equalization is the family of Constant Modulus Algorithm (CMA) based on the Constant Modulus (CM) cost. The CM cost penalizes the deviations of the magnitude of the output of the equalizer from a constant. The CM $p – q$ cost is given by:

$$J(w) = E(|z(n)|^p - R_p)^q.$$  \hspace{1cm} (2.10)

where $w$ is a vector of equalizer coefficients. $z(n)$ is the output of the equalizer. $R_p$ is a constant given by:

$$R_p = \frac{E|a_n|^{2p}}{E|a_n|^p}. \quad a_n \in \text{alphabets of the signal constellation.}$$ \hspace{1cm} (2.11)

There is a family of CM cost functions with the index $p, q \in \{1, 2\}$ [31]. Among the four CM costs, the CM 1-2 and CM 2-2 are the most widely analyzed and used. The algorithm associated with the CM 1-2 cost is the Sato Algorithm [13] and that of the CM 2-2 cost is the Godard Algorithm [12]. We shall define the gradient of a complex vector as [5]:

$$\nabla_{w^*} = \frac{\partial}{\partial w^*}.$$ \hspace{1cm} (2.12)
If the output of the equalizer is \( z(n) = w^H y(n) \), the gradient of the CM 1-2 cost is given by:

\[
\nabla_w \cdot J(w) = E \left( y(n)z^*(n)(1 - R_1 \frac{1}{|z(n)|}) \right).
\]

(2.13)

The gradient of the CM 2-2 cost is:

\[
\nabla_w \cdot J(w) = 2E \left( y(n)z^*(n)(|z(n)|^2 - R_2) \right).
\]

(2.14)

It can be seen at once that the gradient expressions for the CM 1-2 and CM 2-2 costs are not easy to evaluate. If we were to find the stationary points of the costs, the gradient expressions should be set to be equal to zero and the equalizer vector \( w \) will be solved as in the case of the MSE cost (2.5). However, it is not a simple task to find an expression for \( w \) such as the one given in (2.6) for the MSE cost. This is the reason why CM family of costs is difficult to analyse.

### 2.4.2 Behaviour of the CM 1 – 2, 2 – 2 Cost Surface

There are a number of papers that study the behaviour of CM cost functions [8] [27]-[32] [38]. CM \( p - q \) cost surfaces are not quadratic (unlike the MSE cost). The stationary points (or equilibrium points) are the locations on the cost surface where the gradient is zero. When the Hessian matrices of the cost at these locations are positive definite, they are minima of the cost surface. But not all of these minima minimize the CM \( p - q \) cost. Those that minimize the cost are called *global minima* while the others are called *local minima*. Ding et al. have provided a lot of analyses on the CM \( p - q \) cost surfaces in particular the CM 1-2 and CM 2-2 functions. In our discussions, we shall mostly concentrate on these two costs.

Equilibrium points include minima, maxima and saddle points. CM 2-2 cost has been shown to have saddle points [27]. The all-zero equalizer setting \( (w = 0) \) is a maximum
point [27]. In the case of the CM 1-2 cost, \( w = 0 \) is is a pointy maximum [32]. The gradient (2.13) is undefined. One can also easily verify that the all-zero equalizer setting is indeed an equilibrium point for the CM 2-2 cost by putting \( w = 0 \) into (2.14). Figure 2.2 shows a contour plot and Figure 2.3 shows the 3-D plot of the CM 2-2 cost for a two-tap equalizer \( w = [w_1, w_2]^T \) using an analytic CM expression derived in [8]. In the case of a two-tap equalizer, there are two global minima, namely \([0.9, -0.4]^T\) and \([-0.9, 0.4]^T\). It is denoted by 'x' in Fig. 2.2. A pair of global minima is due to the fact that blind equalization using the CM \( p - q \) cost has a phase ambiguity [27]. In a real channel, the ambiguity is \( \pm 1 \). In addition to the two global minima, there are two local minima and a maximum point at \( w = [0, 0]^T \).

Figure 2.2: Contour plot of the CM 2-2 cost for a two-tap equalizer.

Equilibria in a CM cost surface can be classified into two categories: Algorithm Dependent Equilibria (ADE) and Channel Dependent Equilibria (CDE) [38]. This classification comes from the fact that when the gradient of the CM cost is equated to zero, there are two types of solutions: ADE are the ones that satisfy the trivial nullspace and CDE are
Figure 2.3: 3D plot of the CM 2-2 cost for a two-tap equalizer.

The ones that satisfy the non-trivial nullspace of the channel convolution matrix given in [38]. Nullspace analysis of the CM 1-2 and 2-2 costs was first given in [28]. The purpose of the following discussion is to lay the groundwork so that a relationship between the use of antenna array and its effect on CM minima, as alluded to in Section 2.1, can be established. Suppose we have the following:

\begin{align*}
  z(n) &= \mathbf{w}^H \mathbf{y}(n). \\
  y(n) &= \mathbf{c}^H \mathbf{x}(n).
\end{align*}

\tag{2.15}
\tag{2.16}

where \( y(n) \) and \( z(n) \) are the equalizer input and output at time \( n \) respectively, \( y(n) \) is the vector of received signal and \( x(n) \) is the vector of transmitted signal, \( \mathbf{w} \) and \( \mathbf{c} \) are the equalizer and channel vectors respectively. We can express \( z(n) \) to be the convolution of \( x(n) \) and the overall system response \( h(n) = \mathbf{c}(n) \ast \mathbf{w}(n) \). The channel convolution matrix is \( \mathbf{C} \). In [28, 38], it was shown that the gradient of a CM cost can be expressed
in terms of the transmitted symbols \( x(n) \) and the overall system response \( h \)

\[
\nabla_w J(w) = \nabla_h J(h)
\]

\( = C\phi(h). \)

(2.17)

where \( \phi(h) \) is the portion of the gradient expression after making substitution in terms of \( C \) and \( h \) [38]. The gradient expression \( C\phi(h) = 0 \) when \( \phi(h) \in \text{Null}(C) \), where \( \text{Null}(C) \) is the nullspace of \( C \). A complete nullspace analysis is given in [38].

For the case of ADE (\( \phi(h) = 0 \)), it is easier to analyse than that of CDE (\( \phi(h) \neq 0 \)) where non-trivial nullspace is involved [38]. The question one can ask is: Is it possible to avoid CDE at all? If one can guarantee that \( \phi(h) = 0 \) are the only equilibria, then the class of CDE can be eliminated altogether. In order to make sure \( C\phi(h) = 0 \) only when \( \phi(h) = 0 \), \( C \) has to be full rank. The condition for which \( C \) is full rank is called the length and zero condition [8, 38]. The result is that if an antenna array is used and the sub-channels share no common zeros, \( C \) will be full rank. This full rank condition cannot be satisfied if symbol-rate sampling (or single antenna) is used since the sub-channels in this case collapse to be one and it always has a common zero (i.e. the zero of the channel itself) [8].

If one can eliminate CDE by using an antenna array, the remaining ADE depend on the choice of the CM cost function. Ding et al. have shown that cost-dependent local minima exist for CM 1-2 cost [32] but not for the CM 2-2 [29] when using fractionally-spaced equalizer or antenna array. Therefore, despite the use of spatial-temporal equalizer, using a CM 1-2 cost function may result in convergence to local minima in blind algorithms such as the gradient descent method. This result is useful later in this thesis when discussing the regularized semi-blind algorithms using the CM 1-2 cost.

It is important to note that if the full rank condition mentioned earlier is violated, even the use of CM 2-2 cost will not guarantee the elimination of local minima which
fall into the class of CDE [38]. In practical situations, it is not possible to know the channels exactly. Thus, at the receiver, one cannot guarantee that only global minima exist without knowledge of the channel conditions. Therefore, in a more general scenario where we make no assumption about the channels, both CM 1-2 and CM 2-2 will exhibit local minima.

2.4.3 CM Cost for Non-Constant Modulus Signal

In the study of blind equalization, an analogy is usually drawn between the MSE cost and the CM cost [8]. Two important points which are helpful in relation to the objective of this thesis as mentioned in Chapter 1 are as follows:

1. The global minima of the CM cost surface is very similar to that of the MSE cost with a phase shift for constant modulus signals. By minimizing the CM cost, we can achieve perfect equalization in a noiseless environment for constant modulus signals [27, 8].

2. For non-constant modulus signals, the CM cost surface is lifted up. Figure 2.4 shows the contour plot of the CM 2-2 cost when the input signal is 16-QAM. For non-constant modulus signal, the constant $R_p$ in (2.11) will no longer be 1 and $|z(n)|^p$ is not 1 for all symbols. The lifted-up cost surface under 16-QAM is demonstrated in Fig. 2.5 when compared to Fig. 2.3 under the same scale. If we quantify the error by calculating the equalization outputs and that of the original signal, the resulting MSE is higher [8].
2.4. Constant Modulus Cost for Blind Equalization

Figure 2.4: Contour plot of the CM 2-2 cost for a 16-QAM signal.

Figure 2.5: 3D plot of the CM 2-2 cost for a 16-QAM signal.
2.5 Semi-blind Equalization Methods

Semi-blind methods come into play when we consider the following situation: Training sequences are incorporated into the transmitted signals for current and future wireless systems for classification of users in a multi-user environment and adjustment of equalizer coefficients. A logical question to ask is: Why can't we use a combination of training and blind methods at the receiver for equalization? By doing so, we are able to utilize both the known and unknown information in the received data to help determining the equalizer tap-weights.

2.5.1 Combination of Training and Blind Cost Functions

One of the approaches to semi-blind processing is to incorporate into the cost function both the training and blind criteria [14, 15]:

$$\hat{w} = \min_w \{a_1 J_{tr}(w) + a_2 J_{bl}(w)\}.$$  \hspace{1cm} (2.18)

where $J_{tr}(w)$ is the cost function of the training method. $J_{bl}(w)$ is that of the blind method and $a_1, a_2$ are some weighting factors.

In [14] the author combines the LS and CM 1-2 costs to form a semi-blind cost for equalization. The spatial-temporal equalizer coefficients are chosen by:

$$\hat{w} = \min_w \left( \frac{1}{N_t} \sum_{n=n_s}^{n_s+N_t-1} |w^H y(n) - x_1(n-d)|^2 + \rho \frac{1}{N} \sum_{n=1}^{N} (|w^H y(n)| - 1)^2 \right)$$ \hspace{1cm} (2.19)

where $n_s$ is the position of the starting training sequence, $d$ is some delay, $N_t$ is the length of the training sequence, $N$ is the length of the burst and $\rho$ is a weighting factor, $0 < \rho < \infty$. Our work tries to improve on this semi-blind method. Therefore, we shall discuss this method in more detail in Section 2.6.

Another semi-blind method by B. Ng et al. [15] is similar to the approach described above. However, instead of using the constant modulus property, this method imposes a
“structural constraint” in addition to the MSE criterion. The structural constraint comes from the fact that by exploiting the known transmit filter, the resulting tap-weights $w$ will lie in the nullspace of such structure. Using the notation from [15], $w$ satisfies:

$$U^HRw = 0$$  \hspace{1cm} (2.20)

where $R$ is the time-averaged autocorrelation matrix of the input to the equalizer. $U$ is the basis that spans the nullspace of $G^H$. The matrix $G$ is related to the sampled pulse shaping filter. The paper also shows that there is a family of $w$ that satisfies the above equation. Given such a structural constraint, the filter weighting is chosen by:

$$\hat{w} = \min_w \left( \sum_{k=1}^{N_1} |w^Hy(n) - x_1(n - d)|^2 + \alpha |U^HRw|^2 \right)$$  \hspace{1cm} (2.21)

where $y(n)$ is the input signal to the filter. Note also that when $\alpha \to \infty$, we obtain another form of blind constraint based on the structure of the system.

### 2.5.2 Enlarging Training Sequence with Data Symbols

From the above discussion on methods combining both training and blind criteria, it is evident that some minimum length of training data must be available. Results from [14] show that given a propagation environment, the aforementioned semi-blind method with LS and CM costs does not attain a small enough MSE if the training data is too short. Viewing the problem from another perspective, given a certain length of the training sequence, it might still be insufficient under some adverse propagation channels. This brings us to a modification of the semi-blind algorithm. The objective is to enlarge the training data by considering all possible combinations that can be formed from the finite alphabet in the enlarged portion. This “Training-like” [16] approach can be used as a stand-alone algorithm in which case the enlarged sequence will be used in any training-based methods. Moreover, this enlarged training sequence can also be used to improve
on the semi-blind methods mentioned in the previous section. Simulation results in [16] show the improvement in MSE before and after enlarging the training data. However, the computational complexity grows exponentially with the length of the “training-like” sequence.

2.5.3 Other Methods

There are other semi-blind methods which appear in literature and we shall briefly describe them below.

In [18, 19] the authors suggest another approach to properly initialize certain iterative blind algorithms. Proper initialization in blind algorithms has been discussed in [8] to avoid convergence to local minima. Sometimes a different initialization of the tap-weights might allow the algorithms to converge differently. In these papers, the authors make use of the available training data sequence to initialize their algorithms before switching to blind algorithms. Hence they can be categorized as semi-blind methods also.

There are also semi-blind maximum-likelihood methods [20, 21, 22] especially in the context of channel estimations. Maximum-likelihood approaches usually are computationally demanding but there are algorithms [20] to enable them to be more efficient.

2.6 Regularized Semi-blind Algorithm

In this section, the regularized semi-blind method mentioned earlier is described in more detail. The semi-blind algorithm in [14] tried to regularize the standard LS solution with the CM 1-2 cost to provide a better estimation of the equalizer coefficients in the case when the number of training symbols is fewer than that of the coefficients, $N_t < N_\omega M$. 
The algorithm minimizes the cost

\[ J(w) = \frac{1}{N_t} \sum_{n=N_s}^{N_t+N_t-1} |z_1(n) - x_1(n-d)|^2 + \frac{1}{N} \sum_{n=1}^{N} (|z_1(n)| - R_1)^2 \]  

(2.22)

where \( n_s \) is the position of the starting training sequence, \( d \) is some delay. \( N_t \) is the length of the training sequence. \( N \) is the length of the burst. \( R_1 = E|a_n|^2/E|a_n| \) with \( a_n \) being the alphabets in a signal constellation and \( 0 < \rho < \infty \) is a weighting factor. The first term is the summation over the training sequence while the second term is the summation over the whole block of data. We can view (2.22) as taking into account the constant modulus property of the signal in many communication systems. From simulation results in [14], the weighting factor \( \rho = 1 \) is a good value to use. Since CM 1-2 cost is used in the semi-blind algorithm, local minima inherent to the cost exist despite the use of antenna array in the equalization. This property was discussed in Section 2.4.2.

As a demonstration of local minima, Fig. 2.6 shows the contour plot of the regularized semi-blind cost for a two-tap equalizer. This is an estimation of the cost in the presence of one training symbol since we want to visualize the case when \( N_t < MN_w \). The channel used is \( h = [1.0.5]^T \). The transmitted signal is 4-QAM. The global minimum is marked by ‘x’. A 3-D plot of the cost is given in Fig. 2.7. A local minimum is clearly seen from the figure.

A Gauss-Newton recursive method is used to obtain the spatial-temporal equalizer coefficients. The ability to successfully equalize the channel was demonstrated and thus offer the possibility of reducing the number of training symbols used in a block of data. In particular, the MSE performance of the algorithm for 4-QAM signals when the number of training symbols is fewer than the spatial-temporal equalizer taps \( (N_t < MN_w) \) is promising. The gradient and Hessian of the semi-blind regularized cost function were obtained in [14] for 4-QAM where \( R_1 = 1 \). We shall obtain the expressions for non-constant modulus signal in Chapter 3 and demonstrate through simulations in Chapter
Figure 2.6: Contour plot of the semi-blind regularized cost for a two-tap equalizer.

Figure 2.7: 3D plot of the semi-blind regularized cost for a two-tap equalizer.
3 and 4 that the regularized semi-blind algorithm does not provide good steady-state MSE performance for non-constant modulus signals such as 16-QAM under a short burst environment.

In the semi-blind regularized algorithm, the CM 1-2 cost is used. The reason behind the choice is that in using the Gauss-Newton method, the Hessian matrices of the cost need to be computed (see Appendix D for a brief description of the Gauss-Newton method). The Hessian expression for the CM 1-2 cost is independent of the iteration index $k$ as shown in Equation (8.3). Therefore, it can be pre-computed at the start of the recursion without the need to update the Hessian expression subsequently. However, the Hessian expression for the CM 2-2 cost depends on the equalizer vector $w$ which needs to be updated in every iteration during the Gauss-Newton algorithm [17].
Chapter 3

A Semi-blind Algorithm with Additional Penalizing Terms

3.1 Overview

As discussed in the previous chapter, the existence of local minima in the family of CM cost functions and the regularized semi-blind cost function is a disadvantage. Moreover, we also want to improve the MSE performance for the regularized semi-blind method when \( N_t < M N_w \) for constant and non-constant modulus signals under a finite amount of data. In this chapter, we shall try to address the second problem, that is, improving the MSE performance. We shall propose a modification to the regularized semi-blind cost function by adding extra terms that penalize the statistical (in)dependence of the output symbols. The rationale behind this approach is that by putting more penalizing criteria, one might be able to achieve a lower MSE at the equalizer output under a short burst of data. However, as we shall show by simulations later in the chapter, this approach only offers a minor improvement on the MSE.
3.2 CM Cost with Additional Terms

In the blind equalization context when no symbols are known at the receiver, there have been a number of papers written on the use of additional penalizing terms to the family of CM cost functions. They mainly fall into two categories. The first category is specifically used in a multi-user scenario and is called the multi-user constant modulus algorithm (MU-CMA) [34, 35, 36, 37]. In the second case, additional terms penalizing the statistical (in)dependence of the past outputs of the equalizer are used in conjunction with the CM cost. Since this involves memory at the output, the equalization algorithm is termed *Criterion with Memory Nonlinearity* or *CRIMNO* in short [33].

In the first case, the MU-CMA has been used to equalize and separate a mixture of signals at the receiver. The cost function is given by the following:

\[
J(w) = E \sum_{j=1}^{K} (|z_j(n)|^2 - 1)^2 + 2 \sum_{l,n=1,l\neq n}^{K} \sum_{\delta=\delta_1}^{\delta_2} |r_{ln}(\delta)|^2.
\]  

(3.1)

where \( K \) is the number of users, \( z_j(n) \) is the equalizer output for the \( j \)-th user and \( r_{ln}(\delta) \) is the cross-correlation between the \( l \)-th and \( n \)-th users. Notice that the cost function is a sum of CM 2-2 costs and the second sum is penalizing the statistical dependence between users. These additional terms allow the multi-user algorithm to converge to different users' signals and avoid the possibility of converging to the same signal at the outputs of a multi-user equalizer. However, when training symbols are present in the received data and moreover, if we assume that the co-channel users (from other cells) have different training sequences, this MU-CMA is not particularly useful. By incorporating a short training sequence, the semi-blind algorithm that combines the CM and LS costs is able to obtain the desired user's signal when the global minimum is reached.

With this in mind, we shall look at the second category of this modified CM function, CRIMNO. The CRIMNO cost is given in [33] with specific modification to the CM 2-2
3.3. Semi-blind CRIMNO Algorithm

cost. The generalized CRIMNO cost for CM \( p - 2 \) is:

\[
J(w) = \alpha_0 E(|z_1(n)|^p - R_p)^2 + \alpha_1 E(z_1(n)z_1^*(n - 1))^2 + \ldots + \alpha_M E(z_1(n)z_1^*(n - M))^2.
\]

where the first term is the CM \( p - 2 \) cost, the rest of the terms are “memory terms” and \( R_p \) is given in (2.11). Again, without loss of generality, \( z_1(n) \) is the output of the desired user. The \( \alpha_i \)'s are weights to the memory terms.

It was shown by simulations that the blind CRIMNO function changes the shape of the pure CM cost in a small way. However, local minima still exist [33]. Nevertheless, simulation results in [33] showed that the adaptive CRIMNO algorithm achieves a lower MSE in steady state than that of the CM 2-2 algorithm for 64-QAM signals.

3.3 Semi-blind CRIMNO Algorithm

Having equipped ourselves with the idea of penalizing the statistical independence of the output of the equalizer, we shall propose the semi-blind CRIMNO cost in this section. A block implementation of the semi-blind algorithm specifically for a short burst scenario is formulated. As a proposed alternative to the regularized semi-blind cost, the semi-blind CRIMNO cost is given by:

\[
J(w) = \frac{1}{N_1} \sum_{N_i} |z_1(n) - x_1(n - d)|^2 + \frac{\rho}{N} \sum_N (|z_1(n)| - R_1)^2 + \alpha \sum_{\delta=1}^{M_\delta} |\check{r}(\delta)|^2.
\]

where

\[
R_1 = \frac{E[a_n|^2]}{E[|a_n|^2]}, \quad a_n \in \text{alphabets of the signal constellation.} \tag{3.4}
\]

\[
\check{r}(\delta) = \frac{1}{N} \sum_N z_1(n)z_1^*(n - \delta) - \beta(\delta) \tag{3.5}
\]

is the time-averaged auto-correlation penalizing term over the received symbols and \( M_\delta \) is the amount of “memory”. This is a general formulation of \( \check{r}(\delta) \) for transmitted symbols.
with known auto-correlation $\beta(\delta)$. For zero mean i.i.d. transmitted symbols, $\beta(\delta)$ will be zero and $\hat{r}(\delta)$ is just the time-averaged autocorrelation. The summation over $\delta$ signifies the extent to which we penalize the "memory" of the outputs of the equalizer. The weighting factors $\alpha$ and $\rho$ are positive real constants. If $\rho$ is chosen to be a large factor, the CM 1-2 cost dominates and the behaviour of the cost leans towards that of a pure blind CM 1-2 cost function. As in the regularized semi-blind cost, $\rho$ can be between 0 and $\infty$. We shall follow the choice of $\rho = 1$ in the semi-blind regularized algorithm discussed earlier in Chapter 2. The weighting factor $\alpha$, on the other hand, cannot be made too large because penalizing the statistical (in)dependence of the outputs is not enough to equalize the channel.

3.4 Implementation

The spatial-temporal equalizers can be obtained by the Gauss-Newton method as in the case of the regularized semi-blind method. The recursive formula for tap weights $w$ is given by:

$$w^{(k+1)} = w^{(k)} - (\hat{H}^{(k)})^{-1} \hat{G}^{(k)}.$$  \hspace{1cm} (3.6)

where the superscript denotes the stage of the recursion. The matrix $\hat{H}$ is the time-averaged Hessian and $\hat{G}$ is the time-averaged gradient of the cost (3.3). From the cost function, it can be seen that the gradient is the sum of the gradient of the LS. CM 1-2 and the rest of the autocorrelation terms. For a detailed derivation of both the gradient and Hessian, please refer to Appendix A and B. The algorithm is initialized the same way as the regularized semi-blind algorithm in Chapter 2 and [14] which is restated here:

$$w^{(0)} = (\hat{R} + \sigma I)^{-1} \hat{p}.$$  \hspace{1cm} (3.7)
where $\sigma$ is a small positive real constant. The addition of $\sigma I$ to $\hat{R}$ before matrix inversion is a technique that can be used when $\hat{R}$ is singular [5, 11].

In order to simplify the expressions of the gradient and Hessian of the autocorrelation terms below, $\beta(\delta)$ is assumed to be zero. However, the derivation in Appendix A and B keeps the term $\beta(\delta)$ for completeness. The gradient, $\dot{G}^{(k)}$, is given by:

$$
\dot{G}^{(k)} = \dot{G}^{(k)}_{LS} + \rho \dot{G}^{(k)}_{CM} + \alpha \sum_{\delta} \dot{G}^{(k)}_{r}(\delta).
$$

(3.8)

$$
\dot{G}^{(k)}_{LS} = \hat{R}w^{(k)} - \hat{p}.
$$

(3.9)

$$
\dot{G}^{(k)}_{CM} = [\hat{R}_{ab}(0) - R_{1}\hat{R}_{ab}(w^{(k)})]w^{(k)}.
$$

(3.10)

$$
\dot{G}^{(k)}_{r}(\delta) = \hat{R}_{ab}(\delta)w^{(k)}w^{(k)}H\hat{R}_{ab}(\delta)w^{(k)} + \hat{R}_{ab}(\delta)w^{(k)}w^{(k)}H\hat{R}_{ab}(\delta)w^{(k)}.
$$

(3.11)

where the $\dot{G}^{(k)}_{LS}$, $\dot{G}^{(k)}_{CM}$ and $\dot{G}^{(k)}_{r}(\delta)$ are the time-averaged gradient estimates for the LS, CM 1-2 and the autocorrelation terms respectively. There are 4 matrices from the above equations that need to be defined, namely, $\hat{R}$, $\hat{p}$, $\hat{R}_{ab}(\delta)$ and $\hat{R}_{ab}(w^{(k)})$:

$$
\hat{R} = \frac{1}{N_t} \sum_{N_t} y(n)y^{H}(n).
$$

(3.12)

$$
\hat{p} = \frac{1}{N_t} \sum_{N_t} x_1^{*}(n-d)y(n).
$$

(3.13)

$$
\hat{R}_{ab}(\delta) = \frac{1}{N} \sum_{N} y(n)y^{H}(n-\delta).
$$

(3.14)

$$
\hat{R}_{ab}(w^{(k)}) = \frac{1}{N} \sum_{N} \frac{y(n)y^{H}(n)}{|w^{(k)}Hy(n)|}.
$$

(3.15)

Of these 4 matrices, $\hat{R}$ and $\hat{p}$ are averaged over the length of the training sequence, $\hat{R}_{ab}(\delta)$ and $\hat{R}_{ab}(w^{(k)})$ are averaged over the entire burst. Apart from $\hat{R}_{ab}(w^{(k)})$ which changes from iteration to iteration, all the other 3 matrices can be pre-computed at the beginning of the algorithm. The constant $R_1$ is given in (3.4). Since the signal constellation is known, $R_1$ can be computed at the beginning of the algorithm as well.
3.5. Simulation Results and Discussions

We shall now move onto the computation of the Hessian $\hat{\mathbf{H}}^{(k)}$ (Appendix B shows a more detailed derivation):

$$\hat{\mathbf{H}}^{(k)} = \hat{\mathbf{H}}_{LS} + \rho \hat{\mathbf{H}}_{CM} + \alpha \sum \hat{\mathbf{H}}^{(k)}_r(\delta),$$

$$\hat{\mathbf{H}}_{LS} = \hat{\mathbf{R}}.$$  

$$\hat{\mathbf{H}}_{CM} = \hat{\mathbf{R}}_{ab}(0).$$

$$\hat{\mathbf{H}}^{(k)}_r(\delta) = w^{(k)H}\hat{\mathbf{R}}_{ab}(\delta)w^{(k)} + \hat{\mathbf{R}}_{ab}(\delta)w^{(k)H}\hat{\mathbf{R}}_{ab}(\delta)$$

$$\quad + \hat{\mathbf{R}}_{ab}(\delta)w^{(k)H}\hat{\mathbf{R}}_{ab}(\delta)$$

where $\hat{\mathbf{H}}_{LS}$, $\hat{\mathbf{H}}_{CM}$ and $\hat{\mathbf{H}}^{(k)}_r(\delta)$ are the Hessians of the LS, CM 1-2 and autocorrelation terms respectively. $\hat{\mathbf{R}}$ and $\hat{\mathbf{R}}_{ab}(\delta)$ are given in (3.12) and (3.14) respectively. An observation of the three terms reveal that the $\hat{\mathbf{H}}_{LS}$ and $\hat{\mathbf{H}}_{CM}$ are independent of the iteration index $k$. Thus the two matrices can be pre-computed but not for the Hessian $\hat{\mathbf{H}}^{(k)}_r(\delta)$. Therefore at each iteration of the Gauss-Newton algorithm, we need to update the $\hat{\mathbf{H}}^{(k)}_r(\delta)$ for different $\delta$. Because of the need to compute at every iteration some of the terms in the gradient and Hessian expressions, the computational complexity increases.

3.5 Simulation Results and Discussions

In this section, we provide simulation results for the proposed semi-blind CRIMNO algorithm against that of the regularized semi-blind algorithm. Simulations are done measuring the mean square error (MSE) of the equalizer outputs over the received burst of data of $N$ symbols. The number of iterations of the algorithms is also measured. The Gauss-Newton algorithm stops when the following condition is satisfied:

$$\frac{||w^{(k+1)} - w^{(k)}||^2}{||w^{(k)}||^2} < \epsilon.$$  

(3.20)
where $\epsilon$ is a small real positive number. In our simulations, $\epsilon$ is $10^{-5}$. The signals go through their respective channels which are modeled as 3 taps. This is the case when the delay spread is around 3 to 4 symbols. The receiver has 4 antennas with 6 taps on each antenna. The semi-blind CRIMNO algorithm has the following parameters: $\rho = 1$, $\alpha = 0.5$ and the amount of memory $M_\delta = 7$. The MSE performance when $N_t < M \cdot N_w$ is evaluated for both the proposed method and the original regularized algorithm. All simulations are run 40 times in order to generate the plots in an average sense. The algorithm is initialized using (3.7) with $\sigma = 0.005$. The received data is a burst of $N = 150$ symbols.

The first simulation is performed under the following scenario. There are 3 users ($K = 3$) with 2 of them from nearby co-channel cells. They are transmitting in 4-QAM. At the basestation receiver, an antenna array of 4 sensors is used. The signal-to-noise ratio (SNR) at the receiver is 30 dB and the signal-to-interference ratio (SIR) is 3 dB. Figure 3.1 shows the MSE vs. $N_t$ plot from $N_t = 4$ to 22 for 4-QAM signals. This is the region where the number of training symbols ($N_t$) is fewer than the number of equalizer coefficients ($M \cdot N_w$). The LS solution is not unique in this region and its performance is poor. We shall defer this discussion later in Chapter 4. It is apparent from Fig. 3.1 that the semi-blind CRIMNO method does not outperform the regularized semi-blind method. Figure 3.2 shows the number of iterations required by both algorithms to reach the stopping condition (3.20). Again, no significant improvement is demonstrated by the proposed method.

Figure 3.3 shows the MSE over training length from $N_t = 4$ to $N_t = 22$ for 16-QAM signals. We can see that the MSE performance in the 16-QAM case is only marginally better than that of the regularized semi-blind algorithm. In Fig. 3.4, the number of iterations for both algorithms are plotted against $N_t$ when the stopping condition (3.20) is reached. Again, we observe only a slight performance improvement over the regularized
3.5. Simulation Results and Discussions

The region of improvement is when longer training symbols are available in the received data.

In the next simulation, we assume that there are no interference from the nearby co-channel cells (i.e. $K = 1$ only). The received signal is only corrupted by additive white Gaussian noise only. The plot of MSE vs. length of training sequence is shown in Fig. 3.5. The performance improvement of the proposed semi-blind CRIMNO algorithm is only slightly better than that of the regularized semi-blind counterpart under this single user environment. Figure 3.6 shows the number of iterations vs. training. The number of iterations required for the semi-blind CRIMNO algorithm and the regularized semi-blind algorithm are similar.

Figure 3.7 and Figure 3.8 shows the MSE plot and number of iterations plot respectively under the single user environment for 64-QAM signals. Figure 3.9 shows the signal constellation at the output of the spatial-temporal equalizer when $N_t = 22$ using the semi-blind CRIMNO algorithm. The MSE is 0.0513 and it is clear from the figure that the eye is not open. Hence under a burst of 150 symbols, both algorithms cannot equalize the channel successfully by opening the eye of the constellation sufficiently.

In the last simulation, we investigate the effect on changing the amount of “memory” $M_d$. Figure 3.10 shows the plot of MSE vs. $N_t$ and Fig. 3.11 shows the number of iterations when $M_d = 3.5.7$. There are no substantial differences in the performance when different number of memory terms is used.

It is apparent that the use of the extra terms on penalizing the statistical independence of the equalizer output does not provide a tremendous improvement in MSE in situation where only a small block of data is available. Moreover, the number of iterations for the semi-blind CRIMNO algorithm is not reduced by a large amount. The reason for plotting the iterations between the two algorithms is the following: if the MSE performance is not superior for the proposed method, an improvement in the number of iterations required
3.5. Simulation Results and Discussions

to reach the stopping condition maybe an advantage. However, as demonstrated from the simulations, the proposed algorithm does not offer significant performance improvement.

From a computational complexity point of view, equation (3.11) for the gradient of the extra terms and equation (3.19) for the Hessian of the extra terms require a lot of matrix multiplications. When compared to the regularized semi-blind algorithm, the Hessian of the cost as given in [14] can be pre-computed and is independent of the stage of recursions (i.e. independent of $k$ — the iteration index). This is also evident by looking at the Hessian matrices (3.17) and (3.18) of the LS and CM 1-2 costs respectively. The Hessian expressions of the extra terms in the proposed algorithm not only contain more matrix multiplications but they also need to be updated iteration by iteration. In light of the extra computational complexity, we find that the use of this proposed semi-blind modification can only be justified if the performance improvement over the other method is significant either in terms of the final MSE achieved or the average number of iterations required by the algorithm when the stopping condition (3.20) is attained. Obviously, in this case the extra computational cost does not justify the use of such method. In fact, if one observes the signal constellation at the equalizer output and the MSE reached, both methods (the proposed one and the regularized method) fail to achieve equalization for non-constant modulus signals under the short burst environment.
Figure 3.1: MSE vs. $N_t$ plot of semi-blind regularized and semi-blind CRIMNO algorithms for 4-QAM signals ($N = 150$, $K = 3$, $\rho = 1$, $\alpha = 0.5$, $M_\delta = 7$).
3.5. Simulation Results and Discussions

Figure 3.2: Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for 4-QAM signals ($N = 150$, $K = 3$, $\rho = 1$, $\alpha = 0.5$, $M_\delta = 7$).

Figure 3.3: MSE vs. $N_t$ plot of semi-blind regularized and semi-blind CRIMNO algorithms for 16-QAM signals ($N = 150$, $K = 3$, $\rho = 1$, $\alpha = 0.5$, $M_\delta = 7$).
3.5. Simulation Results and Discussions

Figure 3.4: Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for 16-QAM signals ($N = 150$, $K = 3$, $\rho = 1$, $\alpha = 0.5$, $M_{\delta} = 7$).

Figure 3.5: MSE vs. $N_t$ plot of semi-blind regularized and semi-blind CRIMNO algorithms when $K = 1$ (16-QAM signals. $N = 150$, $\rho = 1$, $\alpha = 0.5$, $M_{\delta} = 7$).
Figure 3.6: Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for $K = 1$ (16-QAM signals, $N = 150$, $\rho = 1$, $\alpha = 0.5$, $M_\delta = 7$).

Figure 3.7: MSE vs. $N_t$ plot of semi-blind regularized and semi-blind CRIMNO algorithms for a 64-QAM signal ($K = 1$, $N = 150$, $\rho = 1$, $\alpha = 0.5$, $M_\delta = 7$).
3.5. Simulation Results and Discussions

Figure 3.8: Plot of the number of iterations for semi-blind regularized and semi-blind CRIMNO algorithms when the stopping condition is reached for a 64-QAM signal ($K = 1$, $N = 150$, $\rho = 1$, $\alpha = 0.5$, $M_\delta = 7$).

Figure 3.9: Signal constellation at the equalizer output for a 64-QAM signal using the semi-blind CRIMNO algorithm when $N_t = 22$ for a realization. It demonstrates that the eye is not opened. The MSE is $0.0513$. 
Figure 3.10: MSE vs. $N_t$ plot of the semi-blind CRIMNO algorithms for different $M_d$ ($K = 3. N = 150. \rho = 1. \alpha = 0.5. 16$-QAM signals).

Figure 3.11: Plot of the number of iterations for the semi-blind CRIMNO algorithms when the stopping condition is reached for different $M_d$ ($K = 3. N = 150. \rho = 1. \alpha = 0.5. 16$-QAM signals).
Chapter 4

A Convex Semi-blind Algorithm

In this chapter, a convex semi-blind method is proposed [44]. The idea is to look for an alternative criterion to minimize in the hope that the performance will be promising under non-constant modulus signals. It is apparent from our study so far that the CM cost function, even with the help of a short length of training sequence, does not provide an acceptable MSE level in the short burst environment for non-constant modulus signals.

We first start with a discussion of the LS solution when it is not unique (i.e. the spatial-temporal autocorrelation matrix $\hat{R}$ loses rank). This non-uniqueness of the LS solution provides the groundwork in the subsequent development of adopting a convex cost with a training sequence constraint.

4.1 LS Solution with Few Training Symbols

When a burst of known symbols (training) is received, the method of least square can be used to obtain the spatial-temporal equalizer coefficients. The following equation is satisfied:

$$\hat{R}w = \hat{p},$$

(4.1)
where $\hat{R}$ is the time-averaged spatial-temporal autocorrelation matrix

$$\hat{R} = \frac{1}{N_t} \sum_{n=N_s}^{N_s+N_t-1} y(n)y(n)^H.$$  

(4.2)

and $\hat{p}$ is the time-averaged spatial-temporal cross-correlation matrix

$$\hat{p} = \frac{1}{N_t} \sum_{n=N_s}^{N_s+N_t-1} x_1^*(n-d)y(n)$$  

(4.3)

for some delay $d$. The first user is again assumed to be the desired signal in the multi-user environment.

We can write the above Equation (4.2) and (4.3) in matrix form. Define the matrices containing the received symbols at the i-th antenna to be:

$$Y_i = \begin{bmatrix} y_i(N_s) & y_i(N_s+1) & \cdots & y_i(N_s+N_t-1) \\
                        y_i(N_s-1) & \cdots & \cdots & \vdots \\
                        \vdots & \vdots & \vdots & \vdots \\
                        y_i(N_s-N_w-1) & \cdots & \cdots & y_i(N_s+N_t-N_w) \end{bmatrix}.$$  

(4.4)

$$Y = [Y_1, \ldots, Y_M]^T.$$  

(4.5)

where $Y$ is an $MN_w \times N_t$ matrix. Equation (4.2) can be expressed as

$$\hat{R} = \frac{1}{N_t} YY^H.$$  

(4.6)

Similarly, we can define a vector

$$\hat{x}_1 = [x_1(N_s-d), \ldots, x_1(N_s+N_t-1-d)]^T.$$  

(4.7)

and Equation (4.3) can be expressed as

$$\hat{p} = \frac{1}{N_t} Y \hat{x}_1^*.$$  

(4.8)
Hence (4.1) can be written as

$$
\frac{1}{N_t} Y Y^H w = \frac{1}{N_t} Y \hat{x}_i^*.
$$

(4.9)

$$
Y^H w = \hat{x}_i^*.
$$

(4.10)

This is a system of $N_t$ equations with $M N_w$ unknowns. For the solution $w$ to be unique, we need at least as many equations as the number of unknowns, i.e. $N_t \geq M N_w$. However, if $N_t < M N_w$, the system of equations is underdetermined and we have infinite number of solutions.

Therefore, from (4.1) and (4.9), if the number of training symbols is fewer than the number of spatial-temporal equalizer coefficients $M N_w$, the equalizer is given by

$$
w = \hat{R}^+ \hat{p} + \sum_{i=1}^{N_{ns}} \nu_i U_i.
$$

(4.11)

where $\hat{R}^+$ is the pseudo-inverse of $\hat{R}$ which we shall define later. $U_i$'s are a set of orthonormal basis of the null space of $\hat{R}$. $N_{ns}$ is the number of vector $U_i$'s that spans the null space of $\hat{R}$ and $\nu_i$'s are a set of coefficients. Equation (4.11) can be expressed compactly as:

$$
w = \hat{R}^+ \hat{p} + Uv.
$$

(4.12)

where $U = [U_1, U_2, \ldots, U_{N_{ns}}]$ and $v = [\nu_1, \nu_2, \ldots, \nu_{N_{ns}}]^T$.

Using the Singular Value Decomposition (SVD) theorem [5], the data matrix can be written as

$$
Y = \tilde{U} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \tilde{V}^H.
$$

(4.13)

where $\tilde{U}, \tilde{V}$ are unitary matrices and $\Sigma$ is a diagonal matrix containing the non-zero eigenvalues of $Y$ in decreasing order. If we absorb the scalar $1/N_t$ into the matrix
expressions. the time-average spatial-temporal correlation matrix is (using (4.13)):

\[ \mathbf{\hat{R}} = \mathbf{Y} \mathbf{Y}^H \]

\[ = \hat{U} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \hat{V}^H \hat{V} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \hat{U}^H \]

\[ = \hat{U} \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} \hat{U}^H \]

\[ = [\mathbf{U}_A \mathbf{U}_B] \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}_A^H \\ \mathbf{U}_B^H \end{pmatrix} \]

\[ = \mathbf{U}_A \Sigma^2 \mathbf{U}_A^H . \]

Hence the pseudo-inverse of \( \mathbf{\hat{R}} \), in our case, is defined as [5]:

\[ \mathbf{\hat{R}}^+ = (\mathbf{U}_A \Sigma^2 \mathbf{U}_A^H)^{-1} = \mathbf{U}_A \Sigma^{-2} \mathbf{U}_A^H \]

(4.15)

and \( \mathbf{\hat{R}}^+ \mathbf{p} \) is the particular solution of (4.1). Since

\[ \hat{U}^H \mathbf{Y} \mathbf{Y}^H \hat{U} = \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} . \]

(4.16)

\[ \mathbf{U}_B^H \mathbf{Y} \mathbf{Y}^H \mathbf{U}_B = 0. \]

(4.17)

\[ \mathbf{Y}^H \mathbf{U}_B = 0. \]

(4.18)

\( \mathbf{U}_B \) is the null-space of \( \mathbf{Y}^H \) and forms the columns of \( \mathbf{U} \) given in Equation (4.12).
4.2 Semi-blind Equalization Based on a Convex Cost Function

4.2.1 Background

Since cost-dependent local minima exist in the regularized semi-blind algorithm, there are two ways to avoid convergence to such minima: 1) devising a good initialization strategy of equalizer tap weights or 2) choosing alternative cost functions that are convex. In this Chapter, we are primarily interested in adopting a convex cost function in the problem of semi-blind equalization.

In [39] (and references therein), a convex cost function based on the $l_\infty$ norm of an equalizer output was proposed in the context of blind equalization. The idea comes from the fact that the opening of the eye of the signal constellation is characterized by the intersymbol interference (ISI). Suppose the combined channel-equalizer response is $c^* w = h$. The eye is opened when the magnitude of $h(\delta)$ for some delay $\delta$ dominates the rest of the coefficients. $\sum |h(i)|$. This is closely related to the $l_1$ norm of the combined channel-equalizer response. In practice, we can never know the channel response explicitly. An equivalent but more useful formulation is using the $l_\infty$ norm of the equalizer output [39, 40, 41]. The equivalence between the $l_1$ norm of the combined channel-equalizer response and the $l_\infty$ norm of the output was shown in [42]. In [40], the following cost is proposed:

$$J(w) = ||\text{Re}(z(n))||_\infty = \max |\text{Re}(z(n))|$$

(4.19)

with the constraint

$$\text{Re}(w_{jk}) + \text{Im}(w_{jk}) = 1.$$  

(4.20)

where this constraint anchors one of the spatial-temporal equalizer coefficients $w_{jk}$ and $z(n)$ is the equalizer output of the signal of interest.
The $l_\infty$ norm can be expressed as the following:

$$
\| \text{Re}(z(n)) \|_\infty = \| \text{Re}(h(n) \ast x(n)) \|_\infty
= \| \sum_k \text{Re}(h(k))\text{Re}(x(n-k)) - \text{Im}(h(k))\text{Im}(x(n-k)) \|_\infty. \tag{4.21}
$$

From [40], it was shown that for square-type constellation

$$
\| \text{Re}(z(n)) \|_\infty = \max(\text{Re}(x(n))) \left[ \sum_k |\text{Re}(h(n))| + |\text{Im}(h(n))| \right]
= \| \text{Im}(z(n)) \|_\infty. \tag{4.22}
$$

The cost function $J(w)$ is convex because [40]

$$
J(\lambda w_1 + (1 - \lambda)w_2) = K \left[ \sum_k |\text{Re}(c(n) \ast (\lambda w_1(n) + (1 - \lambda)w_2(n)))| 
+ |\text{Im}(c(n) \ast (\lambda w_1(n) + (1 - \lambda)w_2(n)))| \right]
= K \left[ \sum_k |\text{Re}(c(n) \ast \lambda w_1(n)) + \text{Re}(c(n) \ast (1 - \lambda)w_2(n))| 
+ |\text{Im}(c(n) \ast \lambda w_1(n)) + \text{Im}(c(n) \ast (1 - \lambda)w_2(n))| \right]
\leq K \left[ \lambda \sum_k |\text{Re}(c(n) \ast w_1(n))| + |\text{Im}(c(n) \ast w_1(n))| 
+ (1 - \lambda) \sum_k |\text{Re}(c(n) \ast w_2(n))| + |\text{Im}(c(n) \ast w_2(n))| \right]
= \lambda J(w_1) + (1 - \lambda)J(w_2). \quad \text{for} \ 0 < \lambda < 1. \quad K = \max(\text{Re}(x(n))). \tag{4.23}
$$

Two remarks about (4.19) and (4.20) are in order:

1. The cost (4.19) is appropriate for square-type constellations such as 4-QAM, 16-QAM etc. where $\max\{\text{Re}(a_n)\} = \max\{\text{Im}(a_n)\}$.

2. The constraint (4.20) anchors one of the equalizer taps. This is needed to avoid the all-zero equalizer coefficients which is a valid but trivial minimum to this type of convex cost function.
4.2.2 Convex Cost with Training Constraint

In this section, we propose a linear constraint to be used for the convex cost (4.19). We call it semi-blind because the linear constraint makes use of the small amount of known symbols present in the received burst of data. The idea was essentially discussed in the Section 4.1. When the number of training symbols is fewer than the spatial-temporal equalizer coefficients, the solution of the LS problem can be expressed as (and restated here):

\[ \hat{R}w = \hat{p}. \]  
\[ w = \hat{R}^+\hat{p} + Uv. \]  

Equation (4.24) can be viewed as a constraint on the equalizer and can be adopted to replace the tap-anchoring technique. Hence (4.19) and (4.24) describe our semi-blind convex cost.

There are several properties of this semi-blind cost:

1. The semi-blind constraint (4.24) can be thought of as a generalization of the single-tap anchoring technique.

2. Because of the use of a linear constraint, convexity of the cost (4.19) is still preserved [40, 41].

3. Convexity of the cost (4.19) is established in a doubly infinite equalizer (ideal) setting and also in a finitely parameterized equalizer (practical) setting [40]. Therefore, using an FIR equalizer maintains convexity unlike the CM function.

4. As in the case of the blind convex method, this kind of equalization technique leaves an unknown gain at the equalizer output [41]. This is because the anchoring of a tap or several taps to some predefined values changes the relative gain of the
equalizer response to the “true” one. Hence an automatic gain control (AGC) is needed to scale the output. This can be done with the knowledge of the known signal constellation.

4.3 Implementation

Since $l_{\infty}$ norm cannot be implemented in practice, we approximate the $l_{\infty}$ norm with $l_p$ norm for large $p$:

$$J(w) = ||\text{Re}(z(n))||_{\infty} + ||\text{Im}(z(n))||_{\infty}$$

$$\simeq \lim_{p \to \infty} ||\text{Re}(z(n))||_p + ||\text{Im}(z(n))||_p$$

$$\simeq (E|\text{Re}(z(n))|^p)^{\frac{1}{p}} + (E|\text{Im}(z(n))|^p)^{\frac{1}{p}}$$

(4.26)

for large $p$.

We minimize both the $l_p$ norm of the real and imaginary part of the signal instead of just either one as in (4.19). This is done in [40] to speed up convergence since the cost is now subjected to twice the parameter variations than before. Convexity is preserved in approximating the $l_{\infty}$ norm with an $l_p$ norm [41]. Equation (4.26) is a sum of two convex functions. Therefore convexity is again preserved. In actual implementation, we can minimize the cost

$$J(w) = E|\text{Re}(z(n))|^p + E|\text{Im}(z(n))|^p$$

(4.27)

to simplify computation [40]. Substituting (4.25) in (4.27) and taking the gradient with respect to $v^\ast$, we obtain

$$G = \nabla_{v^\ast} J(v) = E\left\{ pU^H y(n) \left( |\text{Re}(z(n))|^{p-2} \text{Re}(z(n)) \right. \right.$$ 

$$\left. - j |\text{Im}(z(n))|^{p-2} \text{Im}(z(n)) \right\}.$$
A detailed derivation is given in Appendix C.

The received data is processed in a burst of $N$ symbols. A recursive method based on the gradient descent is used to obtain the spatial-temporal equalizer coefficients. The algorithm is given by:

$$v^{(k+1)} = v^{(k)} - \mu \hat{G}^{(k)}$$  \hspace{1cm} (4.29)

where $v^{(k+1)}$ denotes the vector $v$ at the $k$-th recursion. $\mu$ is a small step size and $\hat{G}^{(k)}$ is an estimate of the gradient (4.28) at the $k$-th recursion. We use this recursive method instead of the Gauss-Newton method due to the simplicity of implementation. As one can deduce from (4.28), the Hessian expression will be quite complicated and the use of Gauss-Newton method will be disadvantageous in this respect. The gradient estimate $\hat{G}^{(k)}$ is obtained by averaging over the burst:

$$\hat{G}^{(k)} = \frac{1}{N} \sum_{n=1}^{N} \left\{ p U^H y(n) \left( |\text{Re}(z^{(k)}(n))|^{p-2} \text{Re}(z^{(k)}(n)) \\ - j|\text{Im}(z^{(k)}(n))|^{p-2} \text{Im}(z^{(k)}(n)) \right) \right\}.$$  \hspace{1cm} (4.30)

The algorithm is initialized with $v^{(0)} = 0$. Such initialization is equivalent to setting the equalizer with $\hat{R}^+\hat{p}$ (i.e. the particular LS solution in (4.25)). Then $w^{(k)} = \hat{R}^+\hat{p} + Uv^{(k)}$.

Table 4.1 provides a summary of the algorithm.

### 4.4 Simulations and Discussions

In this section, we shall provide some simulation results on the performance of the proposed semi-blind algorithm. Three users' signals $(K = 3)$ are impinging on a receiver with four sensors $(M = 4)$. The first user is the desired signal and the other 2 users are interferers from other co-channel cells. We shall assume that the SNR of the desired
4.4. Simulations and Discussions

1. Choose the power $p$ and step-size $\mu$

2. Compute $\hat{R}^+ \hat{p} U$

3. $v^{(0)} = 0$

4. for $k = 1$ to iterations

   • $w^{(k)} = \hat{R}^+ \hat{p} + U v^{(k)}$
   
   • $z^{(k)}(n) = w^{(k)H} y(n)$
   
   • Compute $\hat{G}^{(k)}$ from (4.30)
   
   • Compute $v^{(k+1)}$ from (4.29)

5. Compute the equalizer output $z$ from the final $w$

6. Adjust the gain of the equalizer output by an AGC if necessary

Table 4.1: A summary of the semi-blind convex algorithm.
signal at the receiver is 30 dB. The signal-to-interference ratio (SIR) is 3 dB in our simulations. The signals go through their respective channels which are modeled as 3 taps. This is the case when the delay spread is around 3 – 4 symbol periods. At the receiver, each sensor has an equalizer of length 6. Hence the spatial-temporal equalizer has a total of 24 coefficients.

When implementing the semi-blind algorithm (4.27), the choice of the exponent $p$ has to be determined. Figure 4.1 shows a plot of the MSE achieved using different $p$’s for 16-QAM signals. The MSE is lower when a larger $p$ is used. However, a compromise has to be struck. Using a small $p$ does not approximate (4.27) well. In fact, when $p = 2$ (i.e. penalizing the output energy of the equalizer) the cost may not be an appropriate one to minimize [39]. Using too large a $p$, on the other hand, might have numerical problems in the recursion at the initial stage when the noise and ISI are severe. It seems that from Fig. 4.1, larger $p$ gives a lower MSE. The pure blind convex algorithm in [40] uses $p = 12$. We shall also use this value of $p$ in subsequent simulations. The step size $\mu$ for the recursive algorithm is 0.001. The step size is obtained by trial-and-error. The performance measure is the mean square error (MSE) of the output. We shall compare the MSE among the convex semi-blind, regularized semi-blind and pure LS algorithms in the case where $N_t < M N_w$. The blind algorithm with tap-anchoring constraint (4.20) is also implemented using a recursion similar to (4.29) but in terms of $w$. The blind case (which does not take into account of known symbols present in the burst) fails to converge under this scenario for both 4-QAM and 16-QAM (Fig. (4.2) and Fig. (4.3)).

An AGC is used at the output for the convex semi-blind algorithm so that the comparison is meaningful. The AGC adjusts the gain by [33]:

$$A = \left( \frac{E|a_n|^2}{|z(n)|^2} \right)^{\frac{1}{2}},$$

(4.31)

where $a_n$ is the alphabets in the constellation and $|z(n)|^2$ is the average over the burst.
The term $E|a_n|^2$ can be pre-computed since the constellation is known. This is, in fact, the variance of the constellation and in our simulations, we set $E|a_n|^2 = 1$.

4.4.1 Constant and non-constant modulus signals

Figure (4.2) shows the MSE vs. $N_t$ for the case of 4-QAM signals. The MSE is that of the desired user. The burst has 150 symbols. The LS curve indicates the MSE if we are only using the training sequence to compute the equalizer coefficients. It is also an indication of the MSE before passing through the semi-blind algorithms since we initialize the algorithms using the LS solution. The regularized semi-blind algorithm is implemented as in [14]. Our convex semi-blind algorithm runs for 500 recursions. The MSE plot is obtained by averaging over 40 runs of bursts of 150 symbols. The regularized semi-blind algorithm achieves smaller MSE in this scenario than that of the convex semi-blind algorithm.

The next simulation is on 16-QAM signals. In this case the MSE vs. $N_t$ plot (Fig. (4.3)) is obtained by averaging 40 runs of bursts of 200 symbols. We can see that in this scenario, it has a smaller MSE starting from $N_t = 12$ in a burst than the regularized semi-blind algorithm. The regularized semi-blind method does not perform as good as in the case of 4-QAM signals. If we can tolerate an MSE of no more than, say, 0.05, then the regularized semi-blind method will fail in this case while the convex semi-blind method is suitable for $N_t > 16$ in a burst. In Fig. (4.4) and Fig. (4.5), the constellations of the desired signal before and after equalization using the proposed semi-blind algorithm are shown for a realization ($N_t = 18$). The MSE after equalization is 0.022. This higher constellation is proposed in the 3rd generation wireless standard when higher data rate is needed.
4.4.2 Single user with noise only

Our next experiment is to investigate the performance of the proposed algorithm in the single user environment. The significance is that it can be thought of as the extreme case when SIR is infinite or very low interference. The semi-blind convex algorithm achieves a very small MSE (below 0.02) starting from $N_c = 6$ as shown in Fig. 4.6. The regularized semi-blind method, on the other hand, does not seem to improve much in a single user case. Therefore, if the interference can be reduced at the receiver, the convex semi-blind method offers a promising result.

4.4.3 Varying the length of received burst

The performance of the proposed algorithm is evaluated with different lengths of the received data. The number of users is again 3. SNR and SIR are kept the same as before. Figure 4.7 and 4.8 show the plots of MSE when $N = 150$ and $N = 300$ symbols respectively (Figure 4.3 corresponds to $N = 200$). The performance improvement of the proposed semi-blind convex algorithm over the regularized semi-blind algorithm is more pronounced with decreasing number of received symbols. When $N = 300$, the two algorithms provide similar MSE. An interesting observation is that the semi-blind regularized algorithm actually outperforms the semi-blind convex algorithm from $N_c = 4$ to 12. Nonetheless, both algorithms achieve a lower MSE with increasing $N$ as it provides a better estimate of the gradient (and Hessian for the regularized semi-blind algorithm). Figure 4.9 shows all three MSE curves ($N = 150, 200, 300$) for the convex semi-blind algorithm.
4.4. Simulations and Discussions

Figure 4.1: Plot of MSE vs. $N$, for the semi-blind convex algorithm using different $p$ ($K = 3$, 16-QAM signals. $N = 200$).

Figure 4.2: 4-QAM case: MSE vs. $N$, for the pure LS, convex blind, convex semi-blind and regularized semi-blind algorithms ($K = 3$, $N = 150$).
4.4. Simulations and Discussions

Figure 4.3: 16-QAM case: MSE vs. $N_t$ for the pure LS, convex blind, convex semi-blind and regularized semi-blind algorithms ($K = 3, \mathcal{N} = 200$).

Figure 4.4: Constellation for the desired user before equalization ($\mathcal{N} = 200, 16$-QAM).
4.4. Simulations and Discussions

Figure 4.5: Constellation for the desired user after equalization using the proposed convex semi-blind algorithm ($N = 200, 16$-QAM).

Figure 4.6: Plot of MSE vs. $N_t$ when $K = 1$ for the convex semi-blind and regularized semi-blind algorithms ($N = 200, 16$-QAM signals).
Figure 4.7: Plot of MSE vs. $N_t$ when $N = 150$ for the convex semi-blind and regularized semi-blind algorithms ($K = 3$, 16-QAM signals).

Figure 4.8: Plot of MSE vs. $N_t$ when $N = 300$ for the convex semi-blind and regularized semi-blind algorithms ($K = 3$, 16-QAM signals).
Figure 4.9: Plot of MSE vs. \( N_t \) with \( N = 150, 200, 300 \) for the convex semi-blind algorithm (\( K = 3 \), 16-QAM signals).
Chapter 5

Conclusions and Future Research

5.1 Concluding Remarks

In this thesis, we have studied the problems that arise when using CM type of cost function in equalization. The regularized semi-blind cost inherits the problems of the family of CM functions: it does not perform well under non-constant modulus signals in a short burst environment and it has local minima.

Two alternative methods are proposed in order to overcome the above challenges in a short burst environment where a block of data is received under quasi-stationary condition. The first method is to incorporate extra terms to penalize the statistical independence (or dependence) of the spatial-temporal equalizer outputs. An offline algorithm using the Gauss-Newton method is developed to process the received block of data. The main purpose is to examine whether a reduction in MSE is possible in a short burst environment with few training symbols by having additional penalizing criteria. However, simulation results show that the proposed algorithm only offers a slight reduction in MSE for non-constant modulus signal in the short burst environment. Moreover, the computational complexity is higher due to the additional penalizing terms. However,
the most important observation is that the proposed semi-blind CRIMNO and the regularized semi-blind algorithms fail to achieve equalization in a small block of data for non-constant modulus signals such as 16-QAM.

In the second method, a convex cost with training constraint is proposed which eliminates the undesirable local minima. The MSE performance of the regularized semi-blind method is better than that of the convex semi-blind counterpart in the 4-QAM case. However, the proposed method is able to achieve a smaller MSE in the 16-QAM scenario where the signal constellation is non-constant modulus.

5.2 Direction of Future Research

CM type of cost functions are quite difficult to analyze in general. A more theoretical study of the semi-blind CRIMNO cost may provide some insight into ways of improving its performance, especially in the situation when only a small block of data is available for batch processing. In the regularized semi-blind and the semi-blind CRIMNO algorithms, CM 1-2 cost is used due to the simplicity of the Hessian expression. However, the use of CM 2-2 cost in these algorithms can also be investigated.

In our system modelling, we do not simulate signals arriving at the receiver with different angles. Antenna beamforming (one of the benefits mentioned in Chapter 1 on the use of antenna array) can be employed to track the desired user and direct the beam towards it. In light of the good performance demonstrated by the convex semi-blind method when the SIR is infinite, the inclusion of interference suppression techniques at the receiver will certainly offer a tremendous improvement in lowering the resultant MSE. This will allow the system to further reduce the number of training symbols in a burst of data.

As a final remark, blind equalization using the constant modulus criterion was first
proposed in the early 80's. Only until the mid 80's [30] to the early 90's [27] did more thorough analyses on the behaviour of the CM cost appear in literature. Semi-blind equalization methods are still at its infancy. Much work still needs to be done in both the theoretical analysis and actual implementation of semi-blind methods.
Appendix A

Gradient of the Semi-blind CRIMNO Cost

The semi-blind CRIMNO cost is given by (3.3) and is restated here:

\[ J(w) = \frac{1}{N_t} \sum_{N_t} |z_1(n) - x_1(n - d)|^2 + \frac{\rho}{N} \sum_N (|z_1(n)| - R_1)^2 + \alpha \sum_{\delta} |\tilde{r}(\delta)|^2. \quad (A.1) \]

where the first summation term is the LS cost, the second summation term is the CM 1-2 cost and the third summation includes the extra penalizing terms. Let the gradients be \( \hat{G}_{LS} \), \( \hat{G}_{CM} \) and \( \hat{G}_r(\delta) \), then

\[
\hat{G}_{LS} = \frac{1}{N_t} \frac{\partial}{\partial w} \sum_{N_t} (w^H y(n) - x_1(n - d))^H (w^H y(n) - x_1(n - d))
\]

\[
= \frac{1}{N_t} \frac{\partial}{\partial w} \sum_{N_t} (y^H (n) w w^H y(n) - |x_1(n - d)|^2 - \]

\[ x_1^H (n - d) w^H y(n) - y^H (n) w x_1(n - d)) \]

\[
= \frac{1}{N_t} \sum_{N_t} (y(n) y^H (n) w - y(n) x_1^*(n - d))
\]

\[ = \hat{R} w - \hat{p}. \quad (A.2) \]
where $\hat{R}$ and $\hat{p}$ are given by (3.12) and (3.13) respectively. For the gradient of the CM 1-2 cost:

$$
\hat{G}_{CM} = \frac{1}{N} \frac{\partial}{\partial w} \sum_N \left( (|z_1(n)|^2)^{1/2} - R_1 \right)^2 
$$

$$
= \frac{1}{N} \sum_N 2(|z_1(n)| - R_1) \frac{1}{2|z_1(n)|} y(n)y^H(n)w 
$$

$$
= \frac{1}{N} \sum_N \left( y(n)y^H(n) - R_1 \frac{y(n)y^H(n)}{|w^H(y(n))|} \right) w 
$$

$$
= \left( \hat{R}_{ab}(0) - R_1 \hat{R}_{bb}(w) \right) w. \tag{A.3}
$$

where $\hat{R}_{ab}(0)$ and $\hat{R}_{bb}(w)$ are given by (3.14) and (3.15) respectively. The gradient of the penalizing term at $\delta$ is:

$$
\hat{G}_r(\delta) = \frac{\partial}{\partial w} \left( \frac{1}{N} \sum_N w^H y(n)y^H(n - \delta) w - \beta(\delta) \right)^H 
$$

$$
= \left( \frac{1}{N} \sum_N w^H y(n)y^H(n - \delta) w - \beta(\delta) \right) \left( \frac{1}{N} \sum_N (y(n)y^H(n - \delta) w \right) 
$$

$$
+ \left( \frac{1}{N} \sum_N y(n - \delta)y^H(n) w \right) \left( \frac{1}{N} \sum_N (w^H y(n)y^H(n - \delta) w - \beta(\delta) \right) 
$$

$$
= \left( \hat{R}_{ab}(\delta) w \right) \left( w^H \hat{R}_{ab}^H(\delta) w - \beta(\delta) \right) + \left( \hat{R}_{bb}(\delta) w \right) \left( w^H \hat{R}_{ab}(\delta) w - \beta(\delta) \right). \tag{A.4}
$$

where $\hat{R}_{ab}(\delta)$ is defined in (3.14).

The gradient expressions in Chapter 3 are modified to reflect the gradients obtained at the k-th stage of the recursive algorithm. In the recursive algorithm, the equalizer weight vector $w$ is replaced by $w^{(k)}$. 
Appendix B

Hessian of the Semi-blind Crime Cost

Let the Hessians of the LS, CM 1-2 and the extra terms be $\hat{H}_{LS}$, $\hat{H}_{CM}$ and $\hat{H}(\delta)$ respectively. The Hessian is evaluated as:

$$\hat{H} = \nabla_w \hat{G}^H.$$  \hspace{2cm} (B.1)

Using (B.1), the Hessian for the LS term is:

$$\hat{H}_{LS} = \frac{\partial}{\partial w^*}(\hat{R}w - \hat{p})^H$$
$$= \hat{R}.$$  \hspace{2cm} (B.2)

where $\hat{R}$ is given by (3.12). The Hessian of the CM 1-2 term is given by:

$$\hat{H}_{CM} = \frac{\partial}{\partial w^*} \left( (\hat{R}_{ab}(0) - R_1 \hat{R}_{ab}(w))w \right)^H$$
$$= \hat{R}_{ab}(0).$$  \hspace{2cm} (B.3)
where $\hat{R}_{ab}(0)$ is given by (3.14). The Hessian for each of the "memory" term for a given $\delta$ is given by:

$$
\hat{H}_r(\delta) = \frac{\partial}{\partial w^*} \left\{ \left( \hat{R}_{ab}(\delta) w \right) \left( w^H \hat{R}_{ab}^H(\delta) w - \beta(\delta) \right) \\
+ \left( w^H \hat{R}_{ab}^H(\delta) w - \beta(\delta) \right) \left( w^H \hat{R}_{ab}^H(\delta) \right) \right\}^H
$$

$$
= \frac{\partial}{\partial w^*} \left\{ \left( w^H \hat{R}_{ab}(\delta) w - \beta(\delta) \right) \left( w^H \hat{R}_{ab}^H(\delta) \right) \\
+ \left( w^H \hat{R}_{ab}^H(\delta) w - \beta(\delta) \right) \left( w^H \hat{R}_{ab}^H(\delta) \right) \right\}
$$

(B.4)

$$
= \left( w^H \hat{R}_{ab}(\delta) w - \beta(\delta) \right) \hat{R}_{ab}^H(\delta) + \hat{R}_{ab}(\delta) w w^H \hat{R}_{ab}^H(\delta)
+ \left( w^H \hat{R}_{ab}^H(\delta) w - \beta(\delta) \right) \hat{R}_{ab}(\delta) + \hat{R}_{ab}^H(\delta) w w^H \hat{R}_{ab}(\delta).
$$

These Hessian expressions are modified in Chapter 3 to reflect the Hessians obtained by the recursive algorithm at the k-th stage. If the known autocorrelation of the transmitted symbols is zero (i.e. $\beta(\delta) = 0$). $\hat{H}_r(\delta)$ will be reduced to (3.19) for the k-th stage.
Appendix C

Gradient of the Semi-blind Convex Cost

Let $J_1(w) = E|\text{Re}(z(n))|^p$ and $J_2(w) = E|\text{Im}(z(n))|^p$. Since $w = \hat{R}^+\hat{p} + Uv$.

\begin{align*}
J_1(w) &= J_1(v) = E(|\text{Re}(z(n))|^2)^{\frac{p}{2}} \\
\nabla_v J_1(v) &= 2 \frac{\partial}{\partial v^*} J_1(v) = EP|\text{Re}(z(n))|^{p-2} \frac{\partial}{\partial v^*} |\text{Re}(z(n))|^2. \quad (C.1)
\end{align*}

The partial derivative is evaluated as follows:

\begin{equation}
\frac{\partial}{\partial v^*} |\text{Re}(z(n))|^2 = \frac{1}{4} \frac{\partial}{\partial v^*} [2z^H(n)z(n) + z^T(n)z(n) + z^H(n)z^*(n)] \quad (C.3)
\end{equation}

where

\begin{align*}
z(n) &= w^H y(n) \\
&= (\hat{R}^+\hat{p} + Uv)^H y(n) \quad (C.4) \\
&= A + (Uv)^H y(n).
\end{align*}
5.2. Direction of Future Research

Substituting (C.4) into (C.3) and expanding the terms, one will obtain the following after many steps:

\[
\frac{\partial}{\partial v^*} |\text{Re}(z(n))|^2 = \frac{1}{2} U^H y(n) \left( y^H(n) U v + y^T(n) U^* v^* + A^* + A \right) \\
= \frac{1}{2} U^H y(n) \left( 2\text{Re}((Uv)^H y(n)) + 2\text{Re}(A) \right) \\
= U^H y(n) \left( \text{Re}((Uv)^H y(n)) + A \right) \\
= U^H y(n) \text{Re}(z(n))
\]

Therefore, the gradient of \( J_1(v) \) is:

\[
\nabla_v \cdot J_1(v) = E \left( \rho |\text{Re}(z(n))|^{p-2} \text{Re}(z(n)) U^H y(n) \right).
\]

(C.6)

Similarly, the gradient of \( J_2(v) \) is:

\[
\nabla_v \cdot J_2(v) = E \left( -j \rho |\text{Im}(z(n))|^{p-2} \text{Im}(z(n)) U^H y(n) \right)
\]

(C.7)

Hence the gradient of the sum of \( J_1(v) \) and \( J_2(v) \) is:

\[
\nabla_v \cdot J(v) = E \left\{ \rho U^H y(n) \left( |\text{Re}(z(n))|^{p-2} \text{Re}(z(n)) \right) \\
- j |\text{Im}(z(n))|^{p-2} \text{Im}(z(n)) \right\}.
\]

(C.8)
Appendix D

Recursive Algorithms

Two recursive methods are used in the thesis for determining the equalizer coefficients. They are the Gauss-Newton method and a recursive algorithm based on the gradient descent. In short burst communications where a block of data is received under a stationary environment, the recursive algorithms use the same block of data to update the equalizer tap weights.

D.1 Gauss-Newton Method

The Gauss-Newton method [45] is given by:

\[ w^{(k+1)} = w^{(k)} - (\hat{H}^{(k)})^{-1}\hat{G}^{(k)} \]  \hspace{1cm} (D.1)

where \( w \) is the vector of tap weights, the superscript \( (k) \) indicates the stage of the recursion, \( \hat{H}^{(k)} \) is the Hessian estimate of the cost function at the \( k \)-th iteration and \( \hat{G}^{(k)} \) is the gradient estimate at the \( k \)-th iteration. In general, \( \hat{H}^{(k)} \) and \( \hat{G}^{(k)} \) need to be computed at each iteration as the tap weights are updated. Gauss-Newton method may accelerate the search on the optimum tap weights than the gradient descent method [45]. However, there are two main drawbacks: the need to compute the Hessian matrix of
the cost and the need to compute matrix inverses. These increase the computational complexity of the Gauss-Newton method.

D.2 Recursive Gradient Descent Method

The recursive gradient descent method is given by:

\[ w^{(k+1)} = w^{(k)} - \mu \hat{G}^{(k)} \]  

(D.2)

where the only new quantity here is the step-size \( \mu \). The step-size is a small positive real number. It is usually obtained by trial-and-error. The recursive gradient descent method is easier to implement than the Gauss-Newton method. There is no need to compute the Hessian matrix. This is an advantage in cases where the expression for the Hessian is too complicated. The disadvantage is that it is slower than the Gauss-Newton method [45].
Bibliography


