School of Engineering & Design

Electronic & Computer Engineering

NOVEL ALGORITHMS IN WIRELESS CDMA SYSTEMS FOR ESTIMATION & KERNEL BASED EQUALIZATION

A thesis partial submitted for the degree of Doctor of Philosophy

DIMITRIOS VLACHOS

Brunel University

2012

Table of Contents

Figure list

Table list

Acknowledgments

This research project is dedicated, to my grandfather, Gianni Athanasoulia and to my Dear friend and thesis supervisor Dr Marios Hadjinicolaou.

Furthermore, I would like to thank my 2nd supervisor Prof Hamed Al-Raweshidy. Special thanks my friend Dr Anastasios Gravalos for his help while accomplishing this thesis and Prof Luiz Wrobel.

ABSTRACT

A powerful technique is presented for joint blind channel estimation and carrier offset method for code- division multiple access (CDMA) communication systems. The new technique combines singular value decomposition (SVD) analysis with carrier offset parameter.

Current blind methods sustain a high computational complexity as they require the computation of a large SVD twice, and they are sensitive to accurate knowledge of the noise subspace rank. The proposed method overcomes both problems by computing the SVD only once.

Extensive simulations using MatLab demonstrate the robustness of the proposed scheme and its performance is comparable to other existing SVD techniques with significant lower computational as much as 70% cost because it does not require knowledge of the rank of the noise sub-space.

Also a kernel based equalization for CDMA communication systems is proposed, designed and simulated using MatLab. The proposed method in CDMA systems overcomes all other methods.

List of Acronyms

APSM	Adaptive Project Sub-gradient Method
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
FIR	Finite Impulse Response
LSE	Least Square Estimation
MSE	Mean Square Error
POCS	Projection onto Convex Sets
RKHS	Reproducing Kernel Hilbert Space
SNR	Signal to Noise Ratio
SVD	Singular Value Decomposition
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 Overview

Wireless communications, will be the leading way of communicating. One of the most important wireless systems nowadays, is the mobile networks, where wireless channels are used for the communication, between the mobile phones and the base stations.

Of course, for the communication between the mobile phone and the base station, many different techniques and algorithms have been deployed, so as that communication to be possible.

Since we talk about wireless communication, between the mobile phone and the base station, we need a channel access method or a multiple access method, which allows several terminals connected to the same physical medium to transmit over it and to share its capacity.

Three principal types of multiple access schemes are used in modern digital radio systems. These are *Frequency Division Multiple Access* (FDMA), *Time Division Multiple Access* (TDMA) and CDMA.

In FDMA, the bandwidth of the available spectrum is divided into separate channels, each individual channel frequency being allocated to a different active remote station for transmission.

In TDMA, the same spectrum channel frequency is shared by all the active remote stations, but each is only permitted to transmit in short bursts of time (slots), thus sharing the channel between all the remote stations by dividing it over time (hence time division).

In a CDMA system all users occupy the same frequency, and there are separated from each by means of a special code. Each user is assigned a code applied as a secondary modulation, which is used to transform user's signal into spread-spectrum-coded version of the user's data stream. The receiver then uses the same spreading code to transform the spread-spectrum signal back into the original user's data stream.

Most of GSM systems today use TDMA, and some a hybrid version of TDMA with FDMA. These methods though have some major drawbacks such as the co channel interference, for this reason guard periods between the TDMA channels are used and guard bands for the FDMA channels are used.

Both systems have an additional drawback when used for mobile communication. If for example they are used for digital transmission of voice, since a user's voice contains large pause periods, FDMA and TDMA perform poorly. This is because frequency bands and time slots continue to be allotted to the user even in pause periods. Thus, such a multiple access system limits the number of active users that simultaneously share the communication channel.

An alternative multiple access system consists in allowing more than one users to share exactly the same channel with the use of direct-sequence spread spectrum waveforms.

According to this method, users are assigned different signature waveforms (or codes), and each transmitter sends its data stream by modulating its own signature waveform as in a single-user digital communication system.

This approach is known as code division multiple access (CDMA), and permits users to access randomly the communication channel, at the same time and occupy the same frequency band. It is the signature waveforms that facilitate demodulation and signal separation at the receiver.

The mobile users in CDMA systems are assigned a wave signature, which is used for signal transmission. These signatures due to their orthogonality, allow different users' signals to occupy the same time and frequency. The mobile user receives the signal which is transmitted by the base station antenna. The mobile user must be able to detect and allocate the information which is designated for him, and isolate it from the rest of the received signal. In order to be accomplished the aforementioned; the knowledge of the composite signature is a necessity.

Due to the multipath effect that the channel introduces, the duration of the signature is increased because of the convolution with the channel's impulse response. The result of this convolution will be referred from now on as composite signature. So in order to be possible the detection, an estimation of the unknown channel must be first evaluated.

Apart from the channel a second parameter which we take under consideration is the carrier offset estimation. Generally in the wireless receivers of a telecommunication system, a sinusoidal signal is generated from a local oscillator, which has to be multiplied with the received signal, so as to be converted from the Radio Frequency zone to the Intermediate Frequency or baseband zone for further processing.

When the transmitter is the Base Station and receivers the mobile phones, because of the different Local Oscillators in every mobile there will be a residual carrier after the above multiplication which we call carrier offset.

The most recent joint blind channel and carrier offset estimation methods in synchronous CDMA systems; they are based in the modelling of the problem made by [1]. Taking into consideration only the multipath effect they proceed to a channel estimation technique through a technique which is based on the analysis of the signal and noise subspace.

While [2] proceed to a joint blind channel and carrier offset estimation through the solution of an eigenvalue polynomial problem. Finally [3] convert [2] to a generalized eigenvalue problem.

The main characteristic of the above techniques is that they are based on the signature samples of the mobile receiver which stay unaffected from the intersymbol interference and at the same time the noise subspace dimension must be known [39].

By having this information they perform a Singular Value Decomposition Analysis – SVD, to a matrix of a large dimension, so as to take, a base for the signal and noise subspace, after that taking advantage of the perpendicular placement to each other the two subspaces, they perform a second SVD analysis, in order to obtain a joint blind channel and carrier offset estimation.

On the contrary, [3] examines the problem of blind channel estimation without considering the carrier offset parameter. By replacing the first SVD analysis [5] and using a matrix raised to a power, the determination of the base of the two subspaces is possible.

The main advantage of this technique is that it does not require the knowledge of the rank of the noise space, while all composite signature samples are taken into account, making the solution for this problem more realistic.

In this thesis the above techniques will be thoroughly examined, by demonstrating analytically the problem modeling, and the subspace decomposition method technique.

Afterwards, the methodology of [4] will be used for the joint channel and carrier offset estimation for CDMA systems problem, but we will produce an alternative solution to the problem with lower complexity in comparison with any other method exists up to date.

In addition, we will examine the algorithms which are based on kernel. These algorithms were recently developed in the machine learning scientific area, in few words this area is not based on a set of a predefined method but learns relations by itself from the incoming data. These algorithms were initially used for the solution of two classes' classification problem.

In general, an algorithm which is based on kernels constitutes one non-linear version of a linear algorithm, with the data to be processed so as to be picturized initially to a space of a larger dimension.

This representation has as a target, the non-linear formations which show the data initially to be vanished in the new space. Like if we have two non-linear separable classes, then by transferring the classification problem to a larger dimension space, then we can accomplish their linear separation.

The classification algorithms through kernel like for example the SVM, it was only applied to occasions where the training data where known in advance. The desirable algorithms though are those which their training is on-line.

Thus, the training data need not to be known in advance to the system, but to come with the passage of time. NORMA constitutes such an algorithm which is based to the technique of stochastic gradient descent. On the contrary, APSM converts the classification problem to the finding of a spot which belong to the cut of a group of convex sets in a Hilbert space.

1.2 Motivation

The original motivation of this research arose from the fact that no previous work is reported in the literature regarding the joint blind channel and carrier offset estimation using the power method and the kernel based classification was never used before for CDMA systems and compared with the rest of the classification techniques.

In the last ten years, subspace based channel estimation algorithms have been developed for and applied to various vector channels.

One of the earliest suggestions of applying the subspace method to channel estimation problems can be traced back to the work by Moulines et al. in 1995, which focuses on identifying time dispersive channel (modelled as an FIR filter) in Time Division Multiple Access (TDMA) system with oversampling in time and/or space domain by using subspace methods.

With the popularity of CDMA communication systems, several works on the estimation of multipath channels in CDMA system by subspace methods have been reported in 1996. Among them, Liu and Xu's work in deserves better observation. In this work, the authors study the identifiably problem of subspace channel estimation for the first time. Furthermore, they supply a closed form expression of the asymptotic performance of their estimator by using a first-order perturbation analysis.

Since that time, several blind subspace channel estimation methods have been proposed and applied to different scenarios, such as: SIMO channels, frequency selective fading channels in DS-CDMA systems and MC-CDMA systems, multiple receiver antennae and multiple transmitter antennae channels in CDMA systems, multi-carrier channels, etc.

While these algorithms were developed separately for certain specific transmission scenarios, the similarities among them indicate that there must exist some common features of the underlying system models, which provide for the feasibility of the subspace channel estimation. Nevertheless, so far these common features have not been studied in the literature.

1.3 Scope of the Thesis

The scope of this thesis is to investigate the joint blind channel and carrier offset estimation methods and compares them with the proposed method. The new method combines singular value decomposition (SVD) analysis with carrier offset parameter.

While existing blind methods suffer from high computational complexity as it is required the computation of a large SVD not only one but twice, plus it is sensitive to accurate knowledge of the noise subspace rank. The proposed method overcomes both problems by computing the SVD only once.

Extensive simulations demonstrate the robustness of the proposed scheme and its performance is comparable to other existing SVD techniques with significant lower computational cost as much as 67%, because it does not require knowledge of the rank of the noise sub-space. In addition, a kernel based equalization for CDMA communication systems is proposed.

The proposed method in CDMA systems overcomes all other methods. Extensive simulations demonstrate the robustness of the proposed scheme and its performance superiority to other existing CDMA channel classification algorithms.

1.4 Contribution to Knowledge

The contributions of the work presented in the thesis can be categorised as follows:

1. The estimation of the joint blind channel and carrier offset using the power method for a first time in literature and proving its superiority with simulations comparisons to the other methods. Additionally, showing its computational efficiency in comparison with the rest of the methods in the literature.

2. The Kernel Based estimation that is for the first time implemented in the literature for CDMA systems, and proving with the model analysis and the simulation results, its superiority in comparison with the other classification models.

1.5 Thesis Outline

The main body of the thesis is divided into eleven chapters, the contents which are outlined in the following:

In Chapter 2 is presented the data modeling of the system.

In Chapter 3 we model the problem according to [1], based on the CDMA baseband signal which is designated for one user. Also, we examine the Inter-Symbol-

Interference (ISI) due to the multipath effects that the channel introduces, while the samples of the user's composite signature that remain unaffected from ISI are isolated. Then we define the received signal from the (mobile) user of interested which constitutes the overlapping of P signals which are being transmitted from the base station, each one of them are being designated for one of the P users of the system.

In Chapter 4 we show the subspace analysis and channel estimation method from the way the data vectors are being received.

In Chapter 5 the carrier offset parameter is being introduced, subspace analysis is being performed and we implement the joint estimation with the channel according to the techniques based on [2] and [3]. Techniques [2] and [3] are being compared as well.

In Chapter 6 we conclude our work and we suggest future research based on our work. The published papers resulting from this research are attached in the Appendix B. Finally in Appendix A we have added the code used for the simulations which were performed with Matlab.

In Chapter 6 it is presented of two linear separable classes through the support system machine, as well as in the case where the classes are not linear seperable.

In Chapter 7 giving the basic elements which refer to the projections on convex sets (POCS[1]) on a Hilbert space.

In Chapter 8 we present APSM and its performance is compared with NORMA and PERCEPTRON.

In Chapter 9 is modeled the problem of channel equation as a classification problem and we use and we compare the above algorithms, on a linear and on a non-linear channel.

In Chapter 10 we examine a downlink of a CDMA system and we apply the algorithms for the recovery of the information of the user of interest.

CHAPTER 2

DATA MODELLING

2.1 Introduction

We consider the estimation of channel parameters for code-division multiple access (CDMA) communication systems operating over channels with either single or multiple propagation paths. The multiuser channel estimation problem is decomposed into a series of single user problems through a subspace-based approach. By exploiting the eigen-structure of the received signal's sample correlation matrix, the observation space can be partitioned into a signal subspace and a noise subspace without prior knowledge of the unknown parameters. The channel estimate is formed by projecting a given user's spreading waveform into the estimated noise subspace and then either minimizing the likelihood or minimizing the Euclidean norm of this projection. Both of these approaches yield algorithms which are near-far resistant and do not require a preamble.

2.2 Review Stage for single user (baseband signal)

Let us consider a CDMA channel that is shared by *P* simultaneous users. The notations used are tabulated in Table 1.

Table 2.1: Notations

Symbol	Description
S_n	The sequence of the information symbols
L_c	Number of bits in the chip code
c	User’s code

N	Number of received vectors
R_{xx}	Autocorrelation matrix
h	Channel
φ	Phase of the carrier
L	Length of the channel h
P	Number of users
T_s	Symbol interval
T_c	Chip interval

Each user is assigned a signature waveform $w(t)$ with duration $T_s = L_c T_c$, where T_s is the symbol interval. A signature waveform may be expressed at the transmitter as

$$\mathrm{w_s(t)} = \sum_{k=1}^{L_c} \mathrm{c_k}\, p(t - kT_c) \qquad 0 \le t \le T_s \tag{1}$$

and the transmitted waveform $y(t)$ may be expressed as

$$y(t) = \sum_{\mathrm{n}=-\infty}^{\infty} s_n \mathrm{w}(t - nT_s) \tag{2}$$

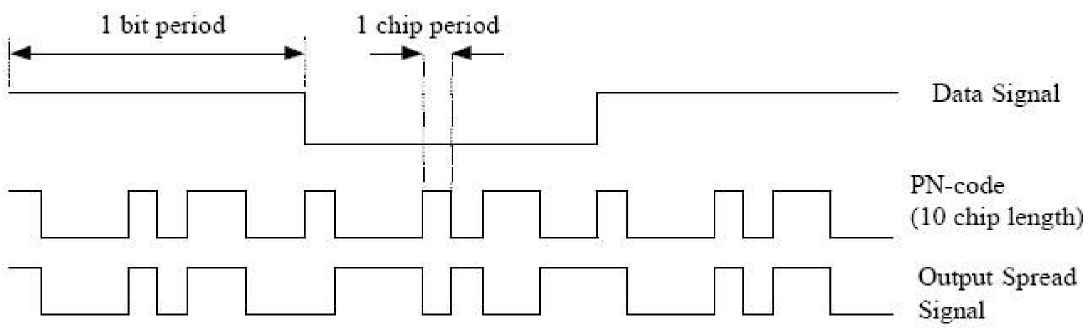


Figure 2.1: Baseband CDMA signal for BPSK modulation

According to the sample of the y(t) with sampling rate $R_c = \dfrac{1}{T_c}$ (3) (chip rate), we get for each symbol L_c (4) samples. The samples resulting from the sampling of a particular symbol are:

$$y(nT_s + T_c),\ y(nT_s + 2T_c), \ldots,\ y(nT_s + L_cT_c),\ \text{with}\ \ T_s = L_cT_c \tag{5}$$

Inserting (1) into (2) and the result is inserted in (5) so we get (6)

In (6) because of (1), (2) and (5) we get: (nT_s+iT_c- kT_c-nT_s) = iT_c-kT_c = (i-k) Tc

$$y(nT_s + iT_c) = \sum_{k=1}^{L_c} s_n c_k\, p(nT_s + iT_c - kT_c - nT_s) = \sum_{k=1}^{L_c} s_n c_k\, p((i-k)T_c) \tag{6}$$

Where, $n = 1, \ldots . N$ can be any symbol in the transmitted data sequence from the base station. And $i = 1, \ldots .., P$ is any user of interest amongst P users which are receiving the information transmitted from the servicing base station.

Where

$$p((i-k)T_c) = 1 \qquad\qquad 0 \le t \le T_c \tag{7}$$

Then

$$\left\{\begin{array}{c} y(nT_s + T_c) = s_n c_1 \\ y(nT_s + 2T_c) = s_n c_2 \\ \vdots \\ y(nT_s + L_cT_c) = s_n c_{L_c} \end{array}\right. \Rightarrow y(n) = \begin{pmatrix} c_1 \\ \vdots \\ c_{L_c} \end{pmatrix} s_n \tag{8}$$

Due to the multiple spread, the received signal it contains the original signal (symbol) that follow the direct path from the transmitter to the receiver and also from its reaction (due to physical and technical obstructions) due to the multipath following different delayed paths [15].

This effect introduces inter-symbol interference (ISI) and therefore it is increased while the data rate is increased. In order to calculate the performance of the mobile communication systems it is convenient to introduce a magnitude for the channel spreading in time domain, known as multipath delay spread.

This delay concerns the measurement for the time space that is intermediate between the first received signal and the last received delayed signal.

We assume that the channel has a duration length LT_c that is multiple to the chip interval. Therefore, we may assume that L<< L_c when the maximum delay introduced by the channel is too low compared to the symbol duration T_s. The impulse response of the channel is

$$h(t) = \sum_{i=1}^{L_d} a_i p(t - \tau_i) \tag{9}$$

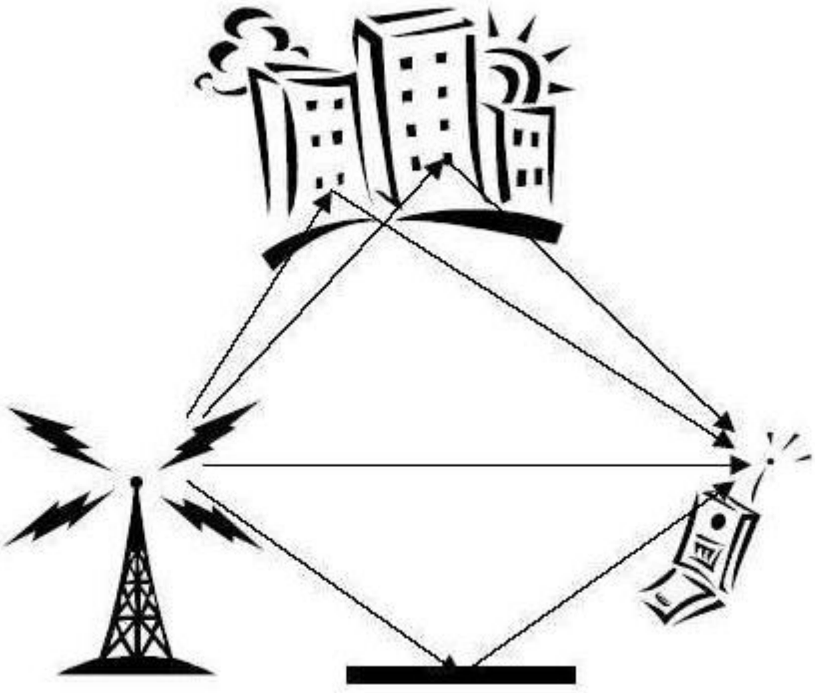


Figure 2.2: The multipath effect

where a_i is the channel's complex gain, τ_i the i-th path delay and L_d is the number of delay paths. The $h(t)$ has a sampling chip rate

$$R_c = \frac{1}{T_c} \tag{10}$$

Therefore by using L samples to configure the channel as an FIR filter with L sample coefficients

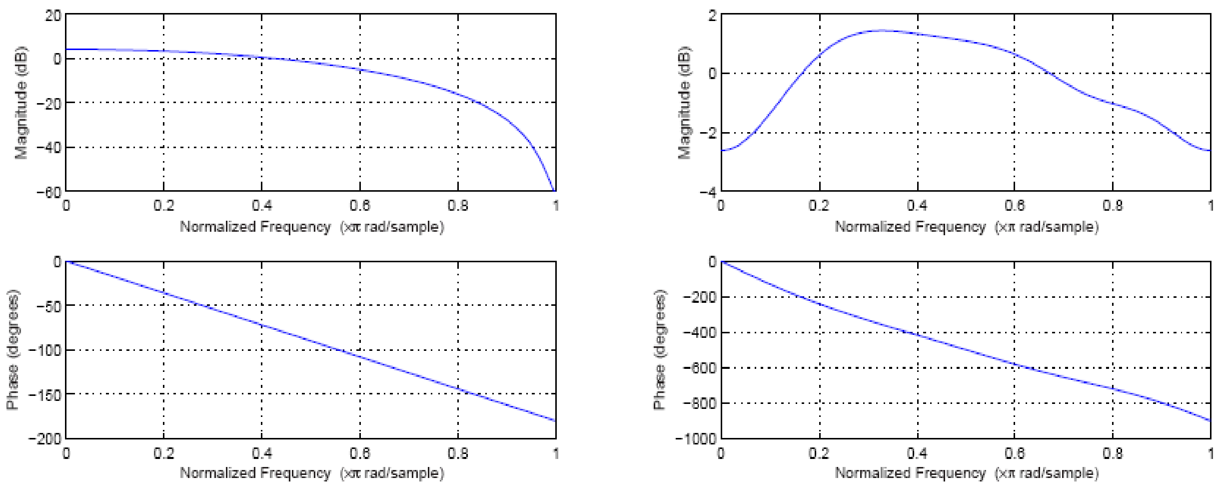
$$\mathbf{h} = (h_1 \ldots h_L)^T \tag{11}$$

that introduce a channel vector. We have selected two time – invariant ISI channels for evaluating the performance of the system [25]. The models are shown in Table 2.2.

Table 2.2: Coefficients of PROAKIS Channels

Channel	Channel coefficients
weak	0.04 -0.05 0.07 -0.21 -0.5 0.72 0.36 0.21 0.03 0.07
strong	0.407 0.815 0.407

Their characteristics with respect to magnitude-frequency and phase-frequency are shown in the figure below.



(a) Weak channel (b) Strong channel

Figure 2.3: Simulation of Proakis channels

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{L_c+L-1} \end{pmatrix} = \begin{pmatrix} c_1 & 0 & \cdots & 0 \\ c_2 & c_1 & \ddots & \vdots \\ \vdots & c_2 & \cdots & 0 \\ c_{L_C} & \vdots & \ddots & c_1 \\ 0 & c_{L_C} & \cdots & c_2 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{L_C} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_L \end{pmatrix} \Rightarrow w = C\mathbf{h} \tag{12}$$

where C is a matrix that consists of a Toeplitz matrix with matrix dimensions (L_c+L-1)xL. Due to the growth of the signature duration during the sampling of the n-th symbol we have an interference at the (n-1)-th symbol. Figure 2.4 shows the second symbol sampling is:

$$\begin{pmatrix} y(T_s+T_c) \\ y(T_s+2T_c) \\ y(T_s+3T_c) \\ \vdots \\ y(T_s+(L-1)T_c) \\ y(T_s+LT_c) \\ \vdots \\ y(T_s+L_cT_c) \end{pmatrix} = \begin{pmatrix} s_2w_1+s_1w_{L_c+1} \\ s_2w_2+s_1w_{L_c+2} \\ s_2w_3+s_1w_{L_c+3} \\ \vdots \\ s_2w_{L-1}+s_1w_{L_c+L-1} \\ s_2w_L+0 \\ \vdots \\ s_2w_{L_c}+0 \end{pmatrix} \tag{13}$$

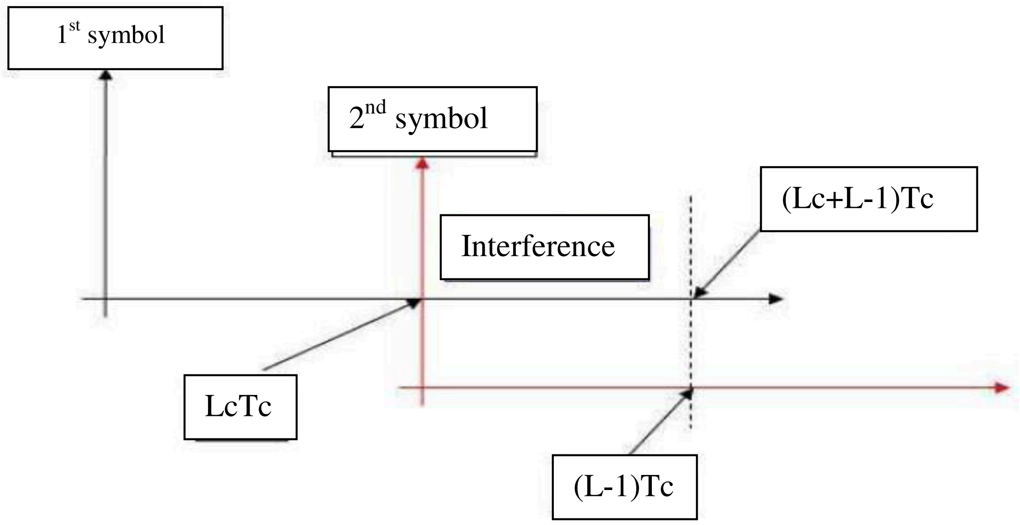


Figure 2.4: The first two symbols

Therefore, the n-th sample symbol is

$$y_n = \begin{pmatrix} w_1 & w_{L_c+1} \\ w_2 & w_{L_c+2} \\ \vdots & \vdots \\ w_{L-1} & w_{L_c+L-1} \\ w_L & 0 \\ \vdots & \vdots \\ w_{L_c} & 0 \end{pmatrix} \begin{pmatrix} s_n \\ s_{n-1} \end{pmatrix} \tag{14}$$

The part from the above equation y_n that remains unaffected from the ISI is:

$$x_n = \begin{pmatrix} w_L \\ \vdots \\ w_{L_c} \end{pmatrix} s_n \Rightarrow x_n = \overline{\mathrm{W}} s_n \tag{15}$$

For the samples ω_i, with i=L, …, L_c of the composite signature for each user will remain unaffected from the ISI are

$$w_i = \sum_{k=1}^{L} h_k c(i-k+1), \quad i = L,\dots,L_c$$

$$\Rightarrow \begin{pmatrix} w_L \\ \vdots \\ w_{L_c} \end{pmatrix} = \begin{pmatrix} c_L & \cdots & c_1 \\ c_{L+1} & \cdots & c_2 \\ \vdots & \ddots & \vdots \\ c_{L_c} & \cdots & c_{L_c-L+1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_L \end{pmatrix} \tag{16}$$

$$\Rightarrow \overline{\mathbf{W}} = \mathbf{Ch}$$

with C to be Toeplitz matrix with dimensions (L_c-L+1)L.

2.3 Baseband Signal for P Users

We assume that the number of users is P. For each user i, the sample vector of the n-th symbol that are independent from the ISI are:

$$x_i(n) = \begin{pmatrix} w_t(L) \\ \\ w_t(L_c) \end{pmatrix} s_i(n) \tag{17}$$

$$\Rightarrow \mathbf{x_i(n)} = \overline{\mathbf{W_i}} s_i(n), \quad \overline{\mathbf{W_i}} = \mathbf{C_i h_i} \tag{18}$$

the data vector that is received by the receiver it consists of the vector summation $x_i(n)$, with i=1…P.

$$\tilde{x}_n = \sum_{i=1}^{P} x_i(n) = \sum_{i=1}^{P} \overline{w}_i s_i(n) \tag{19}$$

where x_n it consists of L_c-L+1 samples of the n-th symbol and P users of the system. In addition we assume that there are N vectors of x_n, with n=1…N from N symbol samples at the receiver.

$$\begin{cases} \tilde{x}_1 = \bar{w}_1 s_1(1) + \bar{w}_2 s_2(1) + \cdots + \bar{w}_P s_P(1) \\ \tilde{x}_2 = \bar{w}_1 s_1(1) + \bar{w}_2 s_2(1) + \cdots + \bar{w}_P s_P(2) \\ \qquad \vdots \\ \tilde{x}_N = \bar{w}_1 s_1(1) + \bar{w}_2 s_2(1) + \cdots + \bar{w}_P s_P(N) \end{cases} \tag{20}$$

$$\Rightarrow \begin{pmatrix} \tilde{x}_1 & \tilde{x}_2 & \dots & \tilde{x}_N \end{pmatrix} = \begin{pmatrix} \bar{w}_1 & \bar{w}_2 & \dots & \bar{w}_P \end{pmatrix} \begin{pmatrix} s_1(1) & s_1(2) & \cdots & s_1(N) \\ s_2(1) & s_2(2) & \cdots & s_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ s_P(1) & s_P(1) & \cdots & s_P(N) \end{pmatrix}$$

$$\Rightarrow \mathbf{X}_{(L_c-L+1)\times N} = \mathbf{W}_{(L_c-L+1)\times P} \mathbf{S}_{P\times N} \tag{21}$$

2.4 Summary

Since the signal subspace is determined solely by the users' spreading waveforms and not their relative amplitudes, subspace-based methods are both near- far resistant and well suited for fading channels. It is also interesting to note that when calculating the estimate for given user, no knowledge of other users' spreading waveforms is necessary. Thus, the algorithm can be used for both multi-user or single user estimation. Although we limited our work to channels with multipath spreads of less than half the symbol period, longer delays could easily be accommodated by increasing the length of the observation vectors.

CHAPTER 3

SUBSPACE ANALYSIS AND CHANNEL ESTIMATION

3.1 Introduction

In this chapter, the performance of a subspace based channel estimation algorithm is investigated in a Code Division Multiple Access (CDMA) communication system. It is analyzed within two cases. The first case is absence of noise, while the second is calculated with the existence of noise.

3.2 Absence of noise

In the previous chapter was presented that:

$$\mathrm{X}_{(L_c-L+1)N} = \mathrm{W}_{(L_c-L+1)P}\,\mathrm{S}_{PxN} \tag{22}$$

Next the following assumptions are made:

1. Assuming that the P columns of matrix $\mathrm{W}_{(L_c-L+1)P}$, which are the signature vectors w_i; $i = 1,\ldots,P$ of the users are linearly independent. Thus the columns of W will constitute a base of the signal subspace.

2. Assuming that the P lines of matrix S_{PxN}, thus the users' symbols are linearly independent. Therefore, $rank(\mathrm{W}) = rank(\mathrm{S}) = P$.

The column i =1, . . . ,P of matrix X constitutes a linear combination of P linearly independent columns of W with coefficients the elements of i column of S. Because $P < N$ we will only have P linear independent columns to X, so $rank(X) = P$. Performing an SVD analysis at matrix $X_{(Lc-L+1)xN}$ we get:

$$X_{(Lc-L+1)xN} = U_{(Lc-L+1)x(Lc-L+1)} \Sigma_{(Lc-L+1)} V^{H}_{NxN} \tag{23}$$

The matrices U and V are orthogonal. The columns u_i of U constitute the right eigenvectors of matrix X and they result from the eigenvectors of matrix $X^H X$ [13][23].

Accordingly the columns v_i of V represent the left eigenvectors of matrix X and they result from eigenvectors of matrix $\Rightarrow X = U1(\Sigma V1^{H})$. Indeed, since $rank(X) = P$ the above relationship can be written as:

$$X_{(Lc-L+1)xN} = \left(U_{1(Lc-L+1)xP} U_{2(Lc-L+1)x(Lc-L+1)} \right) \begin{pmatrix} \Sigma_{PxP} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V^{H}_{1PxN} \\ V^{H}_{2(N-P)xN} \end{pmatrix}$$

$$\Rightarrow X_{(Lc-L+1)xN} = U_{1(Lc-L+1)xP} B_{PxN} \tag{24}$$

The columns of U_i corresponds to the non-zero and different eigenvalues of matrix X, so they are linear independent. From the last equation it can be observed that the column i=1,…,P of matrix X constitute a linear combination of P independent columns of U_2 with coefficients the elements of I column of matrix B. For that reason the columns of $U_2 \perp W$ they constitute a base of the signal subspace [33].

Due to the orthoganality of U it is $U_1 \perp U_2$. That is U_2 is perpendicular to the base of U_i of the signal subspace, so it is vertical and to the matrix W which also constitutes a different base of the signal subspace of: $U_2 \perp W$ .Therefore, the orthogonal complement of the signal subspace constitutes a noise subspace which is generated from the columns of matrix $U_2^H w_i = 0$. In fact, the columns of U_2 correspond to the zero eigenvalues of X, they will not compromise the only base [51].

So, by exploiting the orthogonality of the two subspaces, we get:

$$w_i = C_i h_i \quad , \quad \Rightarrow U_2^H C_i h_i = 0 \tag{25}$$

$$(Lc - L + 1) - P \geq L \Rightarrow P \leq Lc - 2L + 1 , \text{ i=1,...,P} \tag{26}$$

The above system is constituted from (L_c -L + 1)-P equations and L unknows $h_i(1) \ldots h_i(L)$ which are calculated for each one of the i=1,…,P users. Estimating the channel vector h_i of user I, we can then to recover the signature of $\overline{w_i}$. It is observed that in order our system to have a solution the following must apply:

$$(Lc - L + 1) - P \geq L \Rightarrow P \leq Lc - 2L + 1 \tag{27}$$

We end the analysis by simulating the procedure for a CDMA system with BPSK modulation and P=10 users, with spreading gain Lc=32, N=80 the received data vectors of the user of interested, to the strong and to the weak channel, without the presence of noise. We notice that the estimation is very close to the actual one, if we eliminate the mistake which occur in the sign, which is unavoidable.

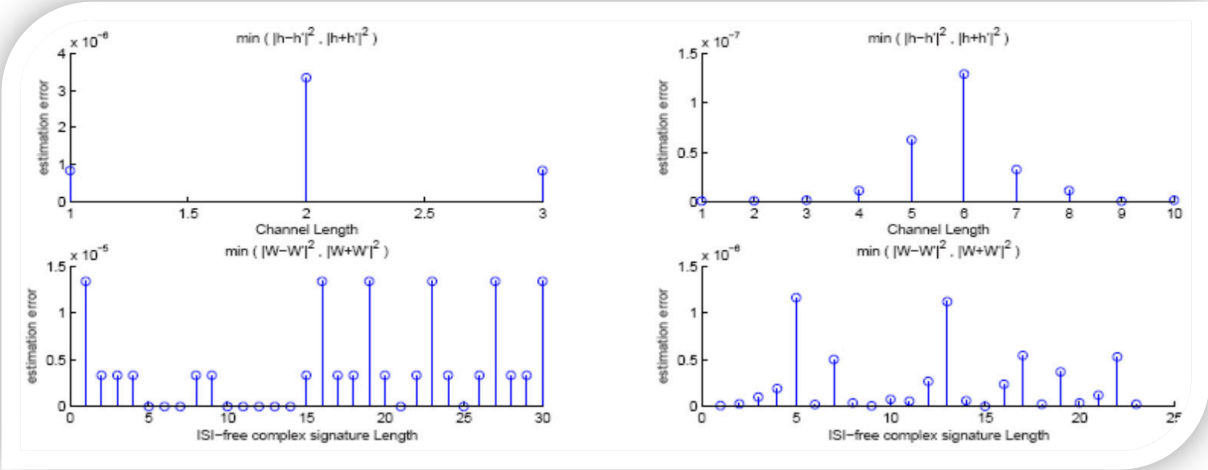


Figure 3.1: Channel and composite signature estimation with noise absence

3.3 Noise Presence

In a telecommunication channel there are various noise sources that can degrade the received signal. As an example the noise that is induced at the received antenna element or at thermal noise and the noise that is produced at the pre-amplified stage of the receiver. At the receiver's input, all the above produced noises can be modeled as a complementary AWGN signal, that is statistically independent from the desirable signal. The power spectral density of the AWGN can be received from an analytical order or from experimental measurements. In this section, we introduce the noise parameter at the received data. It will be examined the behavior of the subspace method analysis as we analyzed before [35][38].

We assume that the signal according to the user of interest i, and the additive noise as a wide sense stationary, with stochastic procedures are independent to the time (where only the first two aptitudes, therefore the mean value and the autocorrelation). The autocorrelation matrix $R_{xx} \epsilon R^{mxm}$ for a stationary stochastic procedure x is defined as

$$\mathbf{R}_{xx} = E\left\{(x-\mu)(x-\mu)^T\right\} \tag{28}$$

where μ is the mean value of the vector procedure. For zero mean values procedures the above Eq. (28) becomes

$$\mathbf{R}_{xx} = E\left\{xx^T\right\} \tag{29}$$

$$\Rightarrow \mathbf{R}_{xx} = E\left\{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}\begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix}\right\} \tag{30}$$

$$\Rightarrow \mathbf{R}_{xx} = E\left\{\begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_m \\ x_1x_2 & x_2^2 & \cdots & x_2x_m \\ \vdots & \vdots & \ddots & \vdots \\ x_mx_1 & x_mx_2 & \cdots & x_m^2 \end{pmatrix}\right\} \tag{31}$$

The $R_{xx}(1,1)$ element is the E{} that is equal to the dispersion σ^2 were the first element of all probable vectors x. Due to the fact that this procedure is stationary

$$\mathbf{R}_{xx}(1,1) = \mathbf{R}_{xx}(2,2) = \ldots = \mathbf{R}_{xx}(m,m) = \sigma^2 \tag{32}$$

and therefore the main diagonal of R_{xx} is equal to the procedure's dispersion. The $R_{xx}(1,2)$ element is the E{} and it consists of the hetero-correlation of the first and second vector of x. Therefore we observe that the elements

$$\mathbf{R}_{xx}(1,2) = \mathbf{R}_{xx}(2,3) = \ldots = \mathbf{R}_{xx}(m-1,m) \tag{33}$$

consists of the hetero-correlation of a delayed sample procedure. Therefore due to the fact that is stationary will be

$$\begin{aligned} \mathbf{R}_{xx}(1,2) = \mathbf{R}_{xx}(2,3) = \ldots = \mathbf{R}_{xx}(m-1,m) = \\ \mathbf{R}_{xx}(2,1) = \mathbf{R}_{xx}(3,2) = \ldots = \mathbf{R}_{xx}(m,m-1) \end{aligned} \tag{34}$$

That means that the elements of the first upper and lower diagonal are equal. Using the same concept, for the j elements of upper and lower diagonals are equal and represent the hetero-correlation procedure of its delayed version of j samples.

In a procedure that the neighbor samples are correlated which means that is altering slower in time, the diagonals of the R_{xx} are decreasing constantly when draw away from the main diagonal. In addition all the neighbor samples are not correlated, therefore changing faster in time, the diagonals of R_{xx} are decreasing very fast when the elements are draw away from the main diagonal. As an example is the white noise that each sample correlates only by itself [48]. Everything is random and uncorrelated. Therefore the autocorrelation matrix of this process is

$$\mathbf{R}_{xx} = \begin{pmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}_{m\times m} \tag{35}$$

The autocorrelation matrix properties are Hermitian, Toeplitz and positive semi-defined. Therefore the eigenvalues will be. Let det[R_{xx}] to be the determinant. Then the eigenvalues solutions λ_i, with i=1,…, N of the N-th order equation is

$$\det\left[\mathbf{R}_{xx} - \lambda \mathbf{I}\right] = 0 \tag{36}$$

and the respective eigenvectors of u_i column will satisfy

$$\mathbf{R}_{xx}\mathbf{u}_i = \lambda \mathbf{u}_1 \tag{37}$$

For a white noise procedure implies that all eigenvalues, $\lambda_1= \lambda_2=... \lambda_N=\sigma^2$ are equal to u_1, $1\le i\le N$ and can be an arbitrary vector. If the eigenvalues are distinctive then the eigenvectors are linearly independent and originate a base in R^N.

Hence, the n-th data vector that is received is

$$\tilde{x}_n = \sum_{i=1}^{P} x_i(n) = \sum_{i=1}^{P} \bar{\mathbf{w}}_i s_i(n) \tag{38}$$

$$\Rightarrow \tilde{x}_n = (\bar{\mathbf{w}}_1 \bar{\mathbf{w}}_2 \cdots \bar{\mathbf{w}}_P) \begin{pmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_P(n) \end{pmatrix} \tag{39}$$

$$\Rightarrow \tilde{x}_n = \tilde{\mathbf{w}}_{(L_c - L+1)\times P} s_n \tag{40}$$

Let $n_{(Lc-L+1)x1}$ the noise vector that is added at the data x_n. Then the autocorrelation matrix of the received vector x_n+n is

$$\mathbf{R}_{xx} = E\left[\left(\tilde{x}_n + n\right)\left(\tilde{x}_n + n\right)^H\right] = E\left[\left(\tilde{\mathbf{w}}_n + n\right)\left(\tilde{\mathbf{w}}_n + n\right)^H\right]$$

$$\Rightarrow \mathbf{R}_{xx} = \tilde{\mathbf{w}}E\left[\mathbf{s}_n\mathbf{s}_n^H\right]\tilde{\mathbf{w}}^H + \tilde{\mathbf{w}}E\left[\mathbf{s}_n\mathbf{n}^H\right]\tilde{\mathbf{w}}^H + \tilde{\mathbf{w}}E\left[\mathbf{n}\mathbf{s}_n^H\right]\tilde{\mathbf{w}}^H + \tilde{\mathbf{w}}E\left[\mathbf{n}\mathbf{n}^H\right]\tilde{\mathbf{w}}^H$$

$$\mathrm{E}\left[s_n n^H\right] = \mathrm{E}\left[n s_n^H\right] = 0 \tag{41}$$

If it is assumed that all noise samples and symbols are uncorrelated then and (41) becomes:

$$\mathbf{R}_{xx} = \tilde{\mathbf{w}}\mathbf{R}_{ss}\tilde{\mathbf{w}}^H + \mathbf{R}_{nn} \tag{42}$$

Also, it is assumed that all s_n samples are independent and therefore uncorrelated, are following the same distribution. Hence

$$\mathbf{R}_{ss} = \begin{pmatrix} \sigma_{\mathrm{s}_1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\mathrm{s}_\mathrm{n}}^2 \end{pmatrix}_{P\times P} \tag{43}$$

and for auto-correlated noise matrix

$$\mathbf{R}_{nn} = \begin{pmatrix} \sigma_\mathrm{n}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_\mathrm{n}^2 \end{pmatrix}_{(L_c-L+1)\times(L_c-L+1)} = \sigma_\mathrm{n}^2\mathbf{I} \tag{44}$$

Finally, the autocorrelation matrix of the n-th vector that is received including noise is

$$\mathbf{R}_{\tilde{x}\tilde{x}} = \tilde{\mathbf{w}} \left(\begin{array}{ccc|ccc} \sigma_{\mathrm{s}_1}^2 + \sigma_\mathrm{n}^2 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \sigma_{\mathrm{s}_\mathrm{n}}^2 + \sigma_\mathrm{n}^2 & & & \\ \hline & & & \sigma_\mathrm{n}^2 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & \sigma_\mathrm{n}^2 \end{array}\right) \tilde{\mathbf{w}}^H \tag{45}$$

$$\Rightarrow \mathbf{R}_{\tilde{x}\tilde{x}} = \left(\mathbf{V}_1\ \mathbf{V}_2\right)\left(\begin{array}{ccc|ccc} \sigma_{s_1}^2+\sigma_n^2 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \sigma_{s_n}^2+\sigma_n^2 & & & \\ \hline & & & \sigma_n^2 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & \sigma_n^2 \end{array}\right)\begin{pmatrix}\mathbf{V}_1^H \\ \mathbf{V}_2^H\end{pmatrix} \tag{46}$$

Therefore we can analyze the R_{xx} eigenvalues in order to provide the eigenvectors V_2^H that represent the smallest eigenvalues σ^2 that consist a base of a noise subspace. In addition we calculate the channel vector h_i for each user i=1,…,P as

$$\begin{aligned} &\mathbf{V}_2^H \bar{\mathbf{w}}_i = 0 \ \ ,\bar{\mathbf{w}}_i = \mathbf{C}_i \mathbf{h}_i \\ &\Rightarrow \mathbf{V}_2^H \mathbf{C}_i \mathbf{h}_i = 0 \quad ,i = 1,\ldots,P \end{aligned} \tag{47}$$

In reality, it is, an estimation of the R_{xx}, with a base of a finite number of received vectors x_n, n=1,…, N

$$\hat{\mathbf{R}}_{\tilde{x}\tilde{x}} = \frac{1}{N}\sum_{i=1}^{N} \tilde{x}_n \tilde{x}_n^H = \frac{1}{N}\mathbf{X}\mathbf{X}^H,\ \mathbf{X} = \left(\tilde{x}_1 \ldots \tilde{x}_N\right) \tag{48}$$

With $X = \begin{pmatrix} \urcorner & \urcorner \\ x_1 \cdots & x_N \end{pmatrix}$ .using this form we can trace small changes in the original autocorrelation matrix as time passing, when the change in the procedure is small and specified in the duration of the N samples [23]. Also, the smallest eigenvalues is not equal to zero but equal σ^2_n. Therefore using Eq. (47)

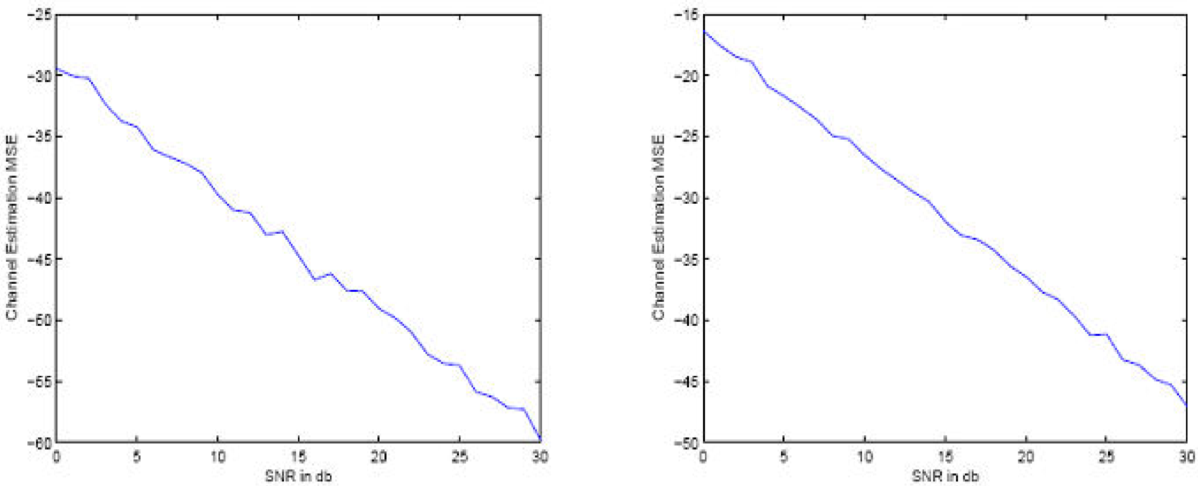
$$\begin{aligned} &\mathbf{V}_2^H \mathbf{C}_i \mathbf{h}_i \approx 0 \Rightarrow \left(\mathbf{V}_2^H \mathbf{C}_i\right)^H \mathbf{V}_2^H \mathbf{C}_i \mathbf{h}_i \approx 0 \\ &\Rightarrow \left(\mathbf{C}_i^H \mathbf{V}_2 \mathbf{V}_2^H \mathbf{C}_i\right)\mathbf{h}_i \approx 0 \end{aligned} \tag{49}$$

and h_i represents the eigenvector that corresponds to the smallest eigenvalues of the matrix $(C_i^H V_2 V_2^H C_i)$. (50)

Hence, the simulation results using the above procedure for the same CDMA system for 10 users including noise and SNR from 0 dB to 30 dB are presented. The mean square error (MSE) is

$$\mathrm{MSE} = E\left\{ \min\left(\left\| \frac{h}{\|h\|} - \frac{h}{\|h\|} \right\|^2 , \left\| \frac{h}{\|h\|} + \frac{h}{\|h\|} \right\|^2 \right) \right\} \tag{51}$$

where E{.} is the stochastic average that approaches the arithmetic average of 100 independent simulations. Finally, all the figures that follow, the MSE is in dB, 10log10(MSE). In both cases when the SNR is increased the MSE is decreased



(a) Strong Channel (b) Weak channel

Figure 3.2: MSE channel estimation in conjunction with SNR

3.4 Summary

A blind channel identification method for MultiCarrier CDMA systems has been presented for both cases, absence or not of noise. First it estimated channel and composite signature with noise absence. Second, the method exploits the orthogonality between the signal and noise subspaces of the incoming signal. It also has been investigated the performance of the method: using a perturbation technique since in telecommunication nothing cannot be solved exactly without having errors at the receiver, we derived an analytical approximate expression of the estimation MSE, where the units of the MSE are the same as the quantity being estimated. The optimal solution is then perturbed. Computer simulations have revealed the high accuracy of the analytical approximation carried out.

CHAPTER 4

ESTIMATION OF THE CARRIER OFFSET

4.1 Introduction

The demodulators which are used at the receivers are classified either as accordant or absonant. Depending whether they use or not one signal carrier which ideally must have the same phase and frequency with the transmitter carrier, so as to be the receiver able to demodulate the received signal.

Usually, the phase and the frequency are recovered from the received signal by using a phase locked loop (PLL) which is using a local oscillator. The recovered can vary from the transmitter's carrier due to phase noise which can be ought to, i.e. frequency slipping of the oscillator and due to the dynamic characteristics and of the transitional behaviour of the PLL.

4.2 System Model

The recovered carrier is expressed as

$$u(t) = V_o[1 + a(t)] \cos(\omega_o t + \varphi_j(t) + dt^2/2) \tag{52}$$

where d (longhead slipping) represents the result due to the caducity of the oscillator, $a(t)$ is the amplitude noise and $\varphi_j(t)$ appoints the phase noise and V_o the initial voltage.

Often, the phase noise is usually entered to the model of a system of transmission as follows:

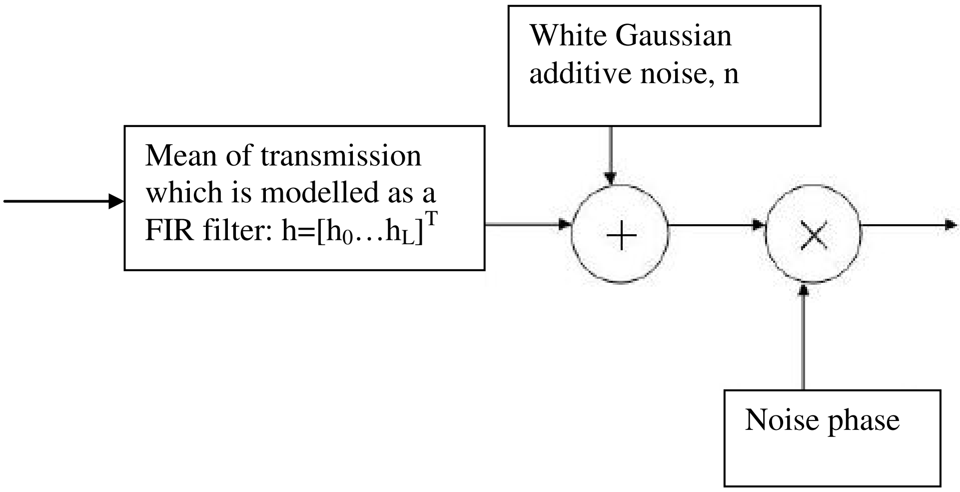


Figure 4.1: Equivalent channel model for baseband transmission

The phase noise or carrier offset $\varphi_j(t)$, as we call it from now on, is constituted from components and random noise. For example, changes in temperature and voltage and changes in the impedance output of the oscillator are some components to name.

In this chapter we will introduce the carrier offset of the received CDMA signal and we will perform a joint channel and carrier offset estimation. Now we will assume that the signal received from one user is as follows:

$$y(t) = \sum_{l=1}^{Ld} a_l x(t-\tau_l) e^{j\omega(t-\tau_l)} \qquad (53)$$

Where L_d is the number of paths, α_l and τ_l are the attenuation and the delay which are introduced from path l, x(t) is the CDMA baseband signal of user i, and ω the residual of the carrier due to the not perfect synchronisation between the transmitter and receiver (mobile terminal).

For the signature and the baseband signal of user, as it was presented in chapter 2, we have accordingly:

$$\mathrm{w(t)}=\sum_{k=1}^{L_c}\mathrm{c_k}\, p(t-kT_c) \tag{54}$$

$$y(t)=\sum_{\mathrm{n=-\infty}}^{\infty}\ \sum_{k=1}^{Lc}\ \sum_{l=1}^{Ld} a_l s_n c_k p(t-\tau_l-nT_s-kTc)e^{j\omega(t-\tau_l)} \tag{55}$$

According to (54) and (55), (56) becomes:

$$y(t)=\sum_{\mathrm{n=-\infty}}^{\infty}\ \sum_{k=1}^{Lc}\ \sum_{l=1}^{Ld} a_l s_n c_k p(t-\tau_l-nT_s-kTc)e^{j\omega(t-\tau_l)} \tag{56}$$

By setting:

$$h(t)=\sum_{l=1}^{Ld} a_l p(t-\tau_l)e^{-j\omega\tau_l} \tag{57}$$

(56) becomes:

$$y(t)=\sum_{\mathrm{n=-\infty}}^{\infty} s_n e^{j\omega t}\sum_{k=1}^{Lc} c_k h(t-nT_s-kTc) \tag{58}$$

We sampling y(t) with sampling rate R_c = $1/T_c$ (chip rate), by receiving from each symbol L_c samples. The samples which come from the sampling of the n-th symbol is:

$$y(nTs+iTc)=s_n e^{j\omega(nT_s+iT_c)}\sum_{k=1}^{Lc} c_k h(nTs+iTc-nTs-KTc)$$

$$\Rightarrow y(nLcTs+iTc)=s_n e^{j\omega nTs}\sum_{k=1}^{Lc} c_k h((i-k)Tc)e^{j\omega iTc} \tag{59}$$

$$w_i=\sum_{k=1}^{Lc} h_k c(i-k+1)e^{j\varphi i} \tag{60}$$

With φ=ωTc and Tc=LcTc. For the samples w_i, i=L, …, L_c of the composite signature which remain unaffected from the ISI are:

$$w_i = \sum_{k=1}^{Lc} h_k c(i-k+1) e^{j\varphi i}\,, \quad \text{i=L,…,L}_c$$

$$\Rightarrow \begin{pmatrix} w_L \\ \vdots \\ w_{Lc} \end{pmatrix} = \begin{pmatrix} e^{jL\phi} & & & \\ & e^{j(L+1)\phi} & & \\ & & \ddots & \\ & & & e^{jLc\phi} \end{pmatrix} \begin{pmatrix} c_L & \cdots & c_1 \\ c_{L+1} & \cdots & c_2 \\ \vdots & & \vdots \\ c_{Lc} & \cdots & c_{Lc-L+1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_L \end{pmatrix}$$

$$\Rightarrow \overline{w} = Z_{(Lc-L+1)x(Lc-L+1)} C_{(Lc-L+1)xL} h_{Lx1} \qquad (61)$$

For matrix Z we can write the following:

$$Z = \begin{pmatrix} e^{j[\phi+(L-1)\phi]} & & & \\ & e^{j[2\phi+(L-1)\phi]} & & \\ & & \ddots & \\ & & & e^{j[(Lc-L+1)\phi+(L-1)\phi]} \end{pmatrix}$$

$$Z = \begin{pmatrix} e^{j[\phi+(L-1)\phi]} & & & \\ & e^{j[2\phi+(L-1)\phi]} & & \\ & & \ddots & \\ & & & e^{j[(Lc-L+1)\phi+(L-1)\phi]} \end{pmatrix} = e^{j\theta} \begin{pmatrix} e^{j\phi} & & & \\ & e^{j2\phi} & & \\ & & \ddots & \\ & & & e^{j(Lc-L+1)\phi} \end{pmatrix} \qquad (62)$$

The L_c-L+1 samples from the n-th symbol which remain unaffected from the ISI

$$\begin{pmatrix} y(nLc+L) \\ y(nLc+L+1) \\ \vdots \\ y(nLc+Lc) \end{pmatrix} = s_n e^{jLc\phi n} \begin{pmatrix} w_L \\ w_{L+1} \\ \vdots \\ w_{Lc} \end{pmatrix} = s_n e^{jLc\phi n} \overline{w} \qquad (63)$$

Supposing that we got P users. If we assume for every user i, the vector with samples of n-th symbol which are free of ISI are:

$$x_i(n) = \begin{pmatrix} w_i(L) \\ \vdots \\ w_i(Lc) \end{pmatrix} s_i(n) e^{jnLc\phi_i} \tag{64}$$

With matrix $\overline{w_i}$ for user i to be: $\begin{pmatrix} w_i(L) \\ w_i(L+1) \\ \vdots \\ w_i(Lc) \end{pmatrix} =$

$$e^{j\theta_i} \begin{pmatrix} e^{j\phi} & & & \\ & e^{j2\phi_i} & & \\ & & \ddots & \\ & & & e^{j(Lc-L+1)\phi} \end{pmatrix} \begin{pmatrix} c_i(L) & c_i(L-1) & \cdots & c_i(1) \\ c_i(L+1) & c_i(L) & \cdots & c_i(2) \\ \vdots & & \ddots & \vdots \\ c_i(Lc) & c_i(Lc-1) & \cdots & c_i(Lc-L+1) \end{pmatrix} \begin{pmatrix} h_i(1) \\ h_i(2) \\ \vdots \\ h_i(L) \end{pmatrix}$$

$$\Rightarrow \overline{w_i} = Z_i C_i h_i \tag{65}$$

It must be mentioned that in the case where we assume as a receiver the mobile phone of the user of interest the carrier offset its common everywhere, so instead of φ_i we have φ.

The data vectors which are received will constitute the sum of vectors of

$x_i(n)$, i=1…P.

$$x_n = \sum_{i=1}^{P} x_i(n) = \sum_{i=1}^{P} \overline{w_i} s_i(n) e^{jnLc\phi i}$$
(66)

So x_n will be comprised from the Lc-L+1 samples of the n-th symbol, which is formed from the overlapping of n-th symbols of P users.

Next we will adopt that we got N vectors x_n, n=1…N which can form the sampling of N symbols obtained at the receiver.

$$\begin{cases} x_1 = \overline{w_1} s_1(1) e^{j1Lc\phi_1} + \overline{w_2} s_2(1) e^{j1Lc\phi_2} + \ldots + \overline{w_p} s_p(1) e^{j1Lc\phi_p} \\ x_2 = \overline{w_1} s_1(2) e^{j2Lc\phi_1} + \overline{w_2} s_2(2) e^{j2Lc\phi_2} + \ldots + \overline{w_p} s_p(2) e^{j2Lc\phi_p} \\ \vdots \\ x_1 = \overline{w_1} s_1(N) e^{jNLc\phi_1} + \overline{w_2} s_2(N) e^{jNLc\phi_2} + \ldots + \overline{w_p} s_p(N) e^{jNLc\phi_p} \end{cases}$$

$$\Rightarrow \left(x_1 x_2 \ldots x_N \right) =$$

$$\left(\overline{w_1} \overline{w_2} \cdots \overline{w_p} \right) \begin{pmatrix} s_1(1) e^{j1Lc\phi_1} s_1(2) e^{j2Lc\phi_1} \ldots s_1(N) e^{jNLc\phi_1} \\ s_2(1) e^{j1Lc\phi_2} s_2(2) e^{j2Lc\phi_2} \ldots s_2(N) e^{jNLc\phi_2} \\ \vdots \\ s_p(1) e^{j1Lc\phi_p} s_p(2) e^{j2Lc\phi_p} \ldots s_p(N) e^{jNLc\phi_p} \end{pmatrix}$$

$$\Rightarrow X_{(Lc-L+1)} = W_{(Lc-L+1)} S_{PxN} \quad (67)$$

Considering the noise matrix $N_{(Lc-L+1)xN}$ then:

$$\Rightarrow X_{(Lc-L+1)} = W_{(Lc-L+1)} S_{PxN} + N_{(Lc-L+1)xN} \quad (68)$$

In previous chapter we saw that we can perform eigenvalue analysis of matrix R_{xx} which constitutes an estimation of the autocorrelation matrix of X, so as to find the eigenvectors V_2^H which correspond to the smallest eigenvalues and represent a base of the noise subspace [26][28]. Thereafter, we calculate the channel h_i vector for every user i=1,…,P like before:

$$V^H_{2(Lc-L+1-P)x(Lc-L+1)}\overline{w_i} \approx 0, \quad \overline{w_i} = Z_i C_i h_i$$

$$\Rightarrow V_2^H Z_i C_i h_i \approx 0 \tag{69}$$

The above system of equations for every user, is not anymore linear because of matrix Z_i. We focus now on the user of interest and we ignore the receiver i [62]. We set K = Lc-l+1 and z = $e^{j\varphi}$. If q_i the column of i of matrix V_2^H and c_i^T the row of i of matrix C, the above system can be written as:

$$\Rightarrow \left(q_1 q_2 \ldots q_K\right)_{(Lc-L)xK} e^{j\theta} \begin{pmatrix} z & & & \\ & z^1 & & \\ & & \ddots & \\ & & & z^K \end{pmatrix}_{KxK} \begin{pmatrix} c_2^T \\ c_1^T \\ \vdots \\ c_K^T \end{pmatrix}_{KxL} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_L \end{pmatrix}_{Lx1} \approx 0$$

$$\Rightarrow e^{j\theta}\left(\underbrace{q_1 c_1^T}_{g_1} z + \underbrace{q_2 c_2^T}_{g_2} z^2 + \ldots + \underbrace{q_K c_K^T}_{g_K} z^K \right) h \approx 0$$

$$\Rightarrow e^{-j\theta} e^{j\theta}\left(g_1 z + g_1 z^2 + \ldots + g_1 z^K\right) h \approx 0$$

$$\Rightarrow \left(\sum_{k=1}^{K} g_k z^k \right)_{(Lc-L)xL} h_{Lx1} \approx 0 \Rightarrow Q(z) h \approx 0 \tag{70}$$

$$\Rightarrow \left(Q^H(z) Q(z)\right) h \approx 0$$

We observe that we end up to a polynomial eigenvalue problem. When $z_0 = e^{j\phi_0}$, where φ_0 is the real carrier offset, then the matrix $Q(z_0)$ must be invertible, so as the channel vector h to be defined univocally.

Thus, the real carrier offset will correspond to the smallest eigenvalue (which will be very close to zero) of matrix $Q(z_0)$ and the channel vector will constitute the corresponding vector. Because the z_0 is unique, we have:

- For z=z0 the smallest eigenvalue approaches zero
- For $z \neq z_0$ the smallest eigenvalue is not zero

Summing, the above steps are the following:

1. SVD analysis to the autocorrelation matrix so as to get the eigenvectors V_2^H which correspond to the smallest eigen-values and they constitute a base of the noise subspace.
2. Sampling of φ at the range [-0.1 0.1], where we assume fluctuates and for every value we keep the smallest eigenvalue of matrix Q. The estimation of carrier offset will be that φ(vector) which gives the smallest eigenvalue and the channel estimation will be the eigenvector which corresponds to it.

$$\phi : -0.0163$$
$$\hat{\phi} : -0.0165$$
$$h = 0.4070\ 0.8150\ 0.4070$$
$$\hat{h} = -0.4123\ \ -0.8118\ \ -0.4131$$

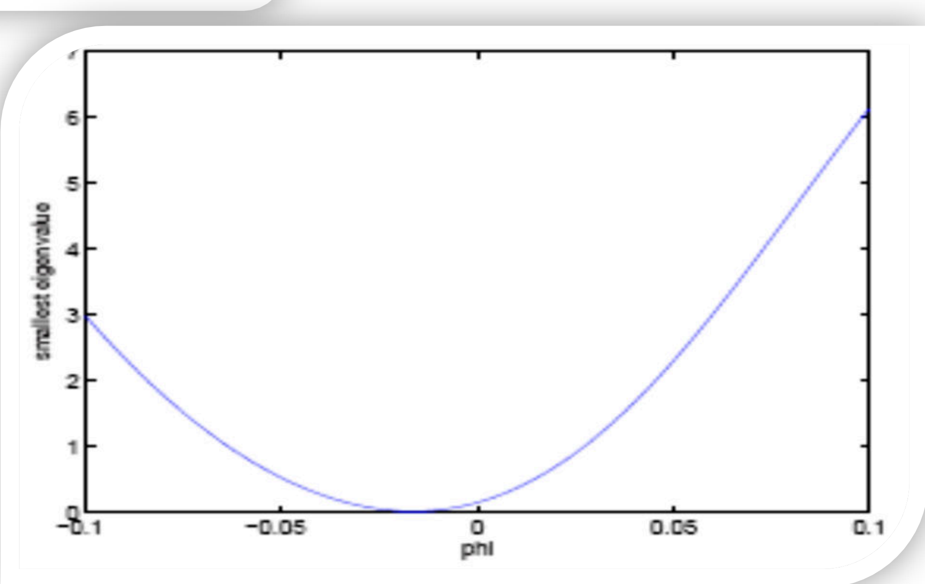


Figure 4.2: The smallest eigenvlaues in function with the samples φ(vector). Strong channel – SNR=10dB

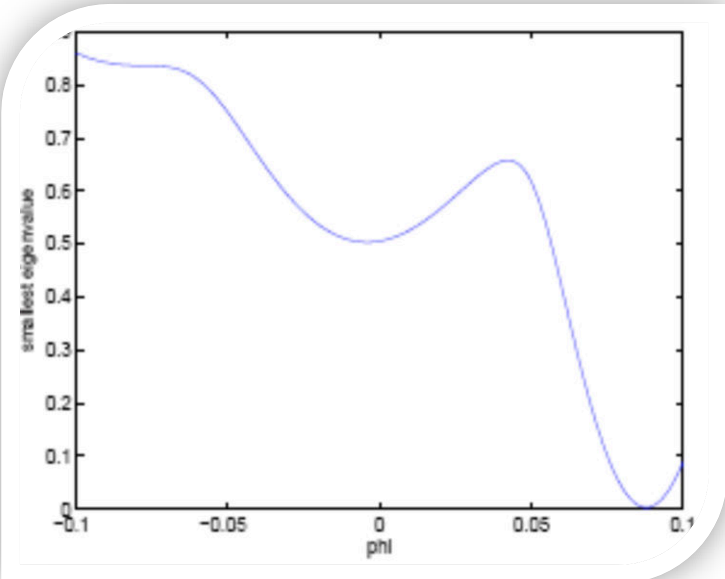


Figure 4.3: The smallest eigenvlaues in function with the samples φ(vector). Weak channel – SNR=10dB

Table 4.1 : φ vectors and their smallest eigenvalues

$$\phi : 0.0877$$
$$\hat{\phi} : 0.0880$$

h	$\hat{h}$
0.0400	−0.0483
−0.0500	0.0306
0.0700	− 0.0735
−0.2100	0.2106
−0.5000	0.4615
0.7200	− 0.7049
0.3600	− 0.3628
0.2100	− 0.2087
0.0300	− 0.0215
0.0700	− 0.0816

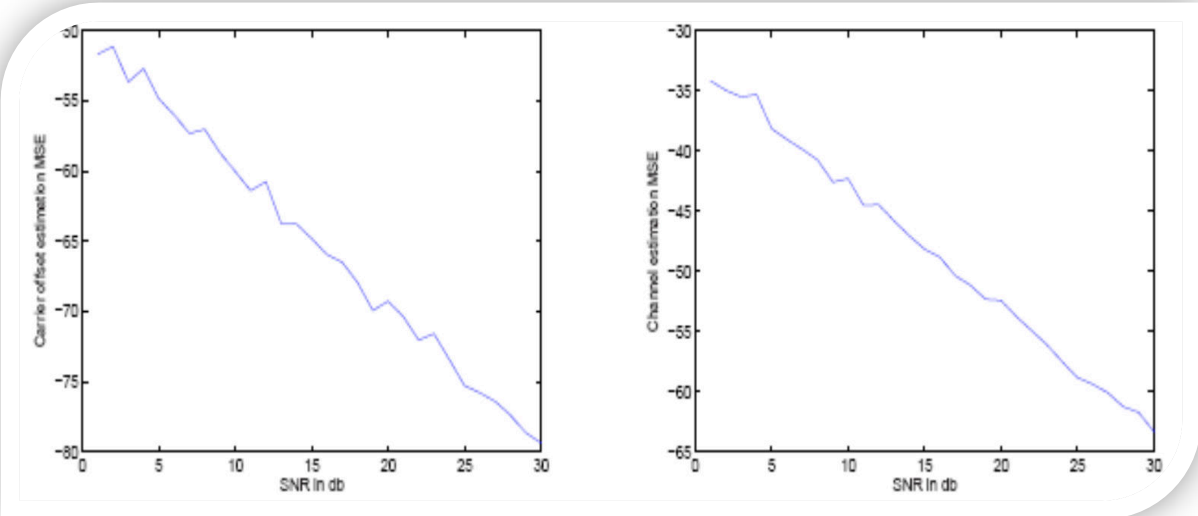
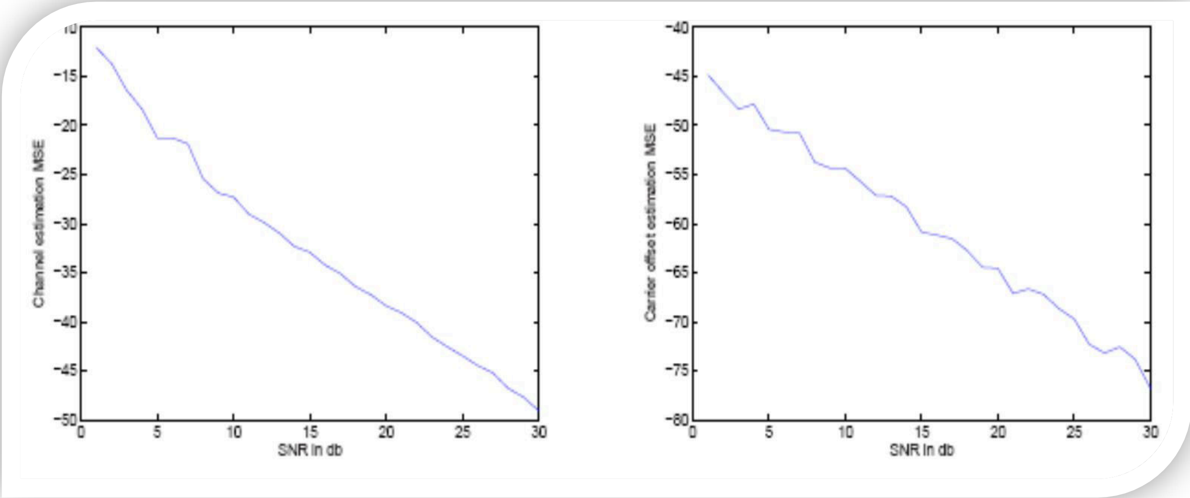


Figure 4.4: Strong channel

(a) MSE carrier offset estimation in function with SNR

(b) MSE channel vector estimation in function with SNR



(a) MSE carrier offset estimation in function with SNR

(b) MSE channel vector estimation in function with SNR

Figure 4.5: Weak channel

At the above figures is presented the MSE of the carrier offset estimation and the MSE of channel vector estimation against SNR for both strong and weak channels.

At the strong channel the MSE of carrier offset estimation is getting a smaller value as the SNR increases than in does in the weak channel.

The opposite though happens with the MSE of the channel vector estimation, where the weak channel achieves lower values as the SNR increases in comparison with the strong channel.

So, in the weak channel there is a better estimation of the channel vector, and in the strong channel there is a better estimation of the carrier offset.

4.3 Reduction to the generalized eigenvalue problem

A different way to encounter the problem is to reduction the polynomial eigenvalue problem to a generalized eigenvalue problem. From (70) we got:

$$\left(g_K z^k + g_{K-1} z^{k-1} + \ldots + g_2 z^2 + g_1 z\right) h = 0$$

$$\Rightarrow g_K z \underbrace{z^{K-1} h}_{x_{1_{Lx1}}} + g_K z \underbrace{z^{K-1} h}_{x_{1_{Lx1}}} + g_K z \underbrace{z^{K-2} h}_{x_{2_{Lx1}}} + \ldots + g_2 \underbrace{z^2 h}_{x_{K-2_{Lx1}}} + g_1 \underbrace{z h}_{x_{K-1_{Lx1}}} = 0$$

$$\Rightarrow g_K z\, x_1 + g_{K-1} x_1 + g_{K-2} x_1 + \ldots + g_K x_{K-2} + g_K x_{K-1} = 0$$

$$\Rightarrow \left(zX + Y\right) x = 0$$

$$\Rightarrow Yx = (-z) Xx \Rightarrow Ax = \lambda Bx \tag{71}$$

We end up to a generalized eigenvalue problem, with matrix zX+Y to have the same eigenvalues as matrix Q(z) if we examine the matrices dimensions. The above matrix equation leads to the following system of equations:

$$\begin{cases} (g_K + g_K) z\, x_1 + g_{K-2} z\, x_2 + \ldots + g_{K-1} z\, x_1 = 0_{(Lc-L)x1} \\ -I\, x_1 + z\, I\, x_2 = 0_{Lx1} \\ -I\, x_2 + z\, I\, x_3 = 0_{Lx1} \\ \vdots \\ -I\, x_{K-2} + z\, I\, x_{K-1} = 0_{Lx1} \end{cases} \tag{72}$$

So in order to appoint the computations, the dimensions of the matrices L and X will be as follows:

$$X = \begin{pmatrix} g_{K(Lc-L)xL} & 0_{(Lc-L)xL} & \cdots & 0_{(Lc-L)xL} \\ 0_{LxL} & I_{LxL} & & 0_{LxL} \\ \vdots & & \ddots & \vdots \\ 0_{LxL} & \cdots & 0_{LxL} & I_{LxL} \end{pmatrix}_{[(Lc-L)+(K-2)L]x(K-1)L} \tag{73}$$

$$Y = \begin{pmatrix} g_{K-1(Lc-L)xL} & g_{K-2(Lc-L)xL} & \cdots & g_{1(Lc-L)xL} \\ -I_{LxL} & 0_{LxL} & & 0_{LxL} \\ \vdots & & \ddots & \vdots \\ 0_{LxL} & \cdots & -I_{LxL} & 0_{LxL} \end{pmatrix}_{[(Lc-L)+(K-2)L]x(K-1)L} \tag{74}$$

We observe that only when Lc=2L the matrices X and Y are square matrices and we can to solve the generalized eigenvalue problem, with the estimation of the carrier offset which corresponds to the smallest eigenvalue, with the carrier offset estimation to correspond to the smallest eigenvalue and the channel estimation to come from the corresponding eigenvector choosing the first L elements and dividing with $-z^{K-1}$. Otherwise we continue from (74) as follows:

$$(zX+Y)^H(zX+Y)x \approx 0 \Rightarrow (z^{-1}X^H+Y^H)(zX+Y)x \approx 0$$

$$\Rightarrow (z^{-1}z\ X^H + z^{-1}X^HY + zY^HX + Y^HY)x \approx 0$$

$$\Rightarrow (z^2Y^HX\ + z\ (X^HX+Y^HY\ )\ + X^HY)x \approx 0$$

$$\Rightarrow Q_{(K-1)Lx(K-1)L}\,x \approx 0 \tag{75}$$

The estimation is performed as before, by sampling on φ in the range [-0.1 0.1] and choosing the φ(vector) which corresponds to the smallest eigenvalue of Q each time.

But in this situation the dimension of Q is much larger, which makes the SVD analysis time-consuming [62][36].

4.4 Approximation of the Z matrix with the help of the Taylor expansion

Closing this chapter we will present a way of encountering the problem which is based to the Z_i matrix approach through Taylor series.

Setting D=diag{1,2,…,K} is:

$$Z_i = I + j\,\phi_i D + \frac{(j\phi_i)^2}{2!} D^2 + \ldots + \frac{(j\phi_i)^n}{n!} D^n + \ldots \tag{76}$$

Using the first two terms of Taylor series () becomes:

$$V_2^H (I + \phi_i D) C_i h_i \approx 0$$
(77)

So the wanted φ_i(vector) and hi(vector) are those for which stands:

$$\hat{\phi}_i,\ \hat{h}_i = \arg \min_{\hat{\phi}_i,\ \hat{h}_i} \left\| V_2^H I + j\phi D) C_i h_i \right\|^2$$

$$\Rightarrow \hat{\phi}_i,\ \hat{h}_i = \arg \min_{\hat{\phi}_i,\ \hat{h}_i} J\ (\phi_i, h_i) \tag{78}$$

$$\Rightarrow \hat{\phi}_i,\ \hat{h}_i = \arg \min_{\hat{\phi}_i,\ \hat{h}_i} J\ (\phi_i, h_i) \hat{\phi}_i,\ \hat{h}_i = \arg \min_{\hat{\phi}_i,\ \hat{h}_i} \left\| \underbrace{V_2^H C_i}_{A_i} h_i + j\phi \underbrace{V_2^H D C_i}_{B_i} h_i \right\|^2$$

For the cost function J(φ_i, h_i) we have:

$$J\ (\phi_i, h_i) = (A_i h_i + j\phi_i B_i h_i)^H (A_i h_i + j\ \phi_i B_i h_i) \tag{79}$$

$$\Rightarrow J\ (\phi_i, h_i) = h_i^H A_i^H + A_i h_i + j\phi_i h_i^H (A_i B_i - B_i^H A_i) h_i + \phi_i^2 h_i^H B_i^H B_i h_i$$

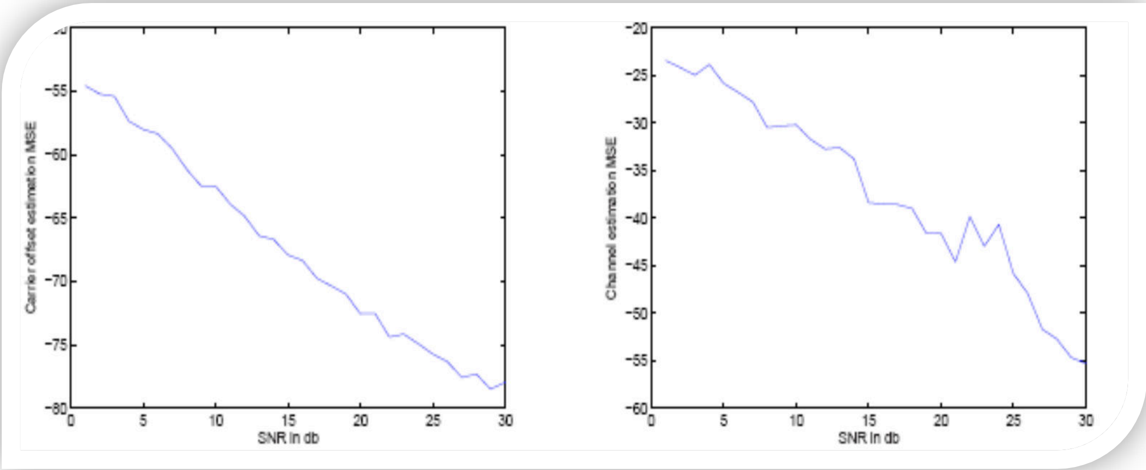
$$\Rightarrow \frac{\vartheta J(\phi_i,\ h_i)}{\vartheta\ h_i} = A_i^H A_i h_i + j\phi_i (A_i B_i - B_i^H A_i) h_i + \phi_i^2 B_i^H B_i h_i \tag{80}$$

We observe that (80) leads to a polynomial problem, which will redact to a generalized eigenvalue problem Mx=λx. From (80) setting g_i=φh, then we have:

$$A_i^H A_i h_i + j(A_i B_i - B_i^H A_i) g_i + \phi_i^2 B_i^H B_i g_i \ 0$$

$$\Rightarrow -(B_i^H B_i)^{-1} A_i^H A_i - j(B_i^H B_i)^{-1} (\ A_i B_i - B_i^H A_i)\ g_i - \phi_i g_i \approx 0 \tag{81}$$

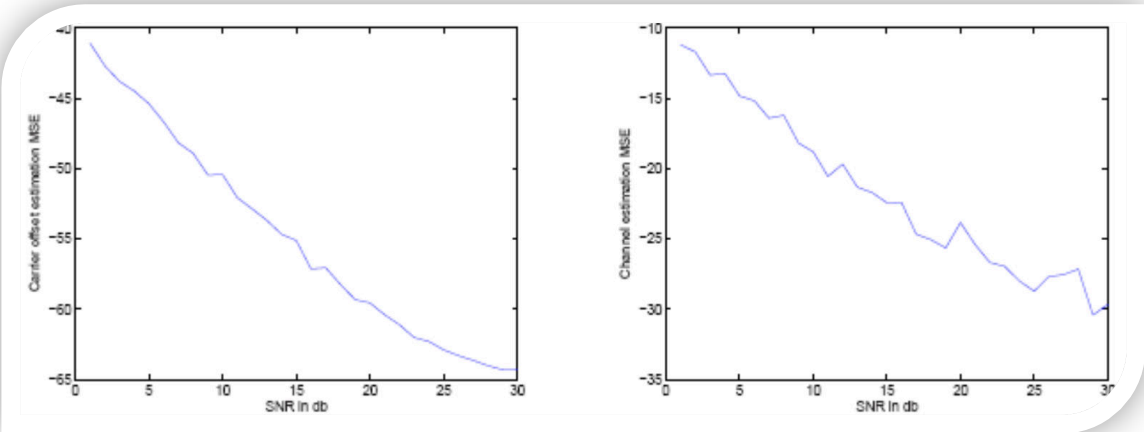
We end up to a generalized eigenvalue problem for the carrier offset estimation to constitute the smallest eigenvalue of matrix M and the channel vector estimation, the first L samples of the corresponding eigenvector. From the figures which will follow we observe the Taylor method approach to give a better carrier offset estimation in the strong channel (Fig (4.6)) and Fig (4.7)). We could use more coefficients in Taylor technique with a result to have a better estimation. This would lead again to a generalized eigenvalue problem with a dimension that would made SVD analysis prohibitive.



(a) MSE carrier offset estimation in function with SNR

(b) MSE channel vector estimation in function with SNR

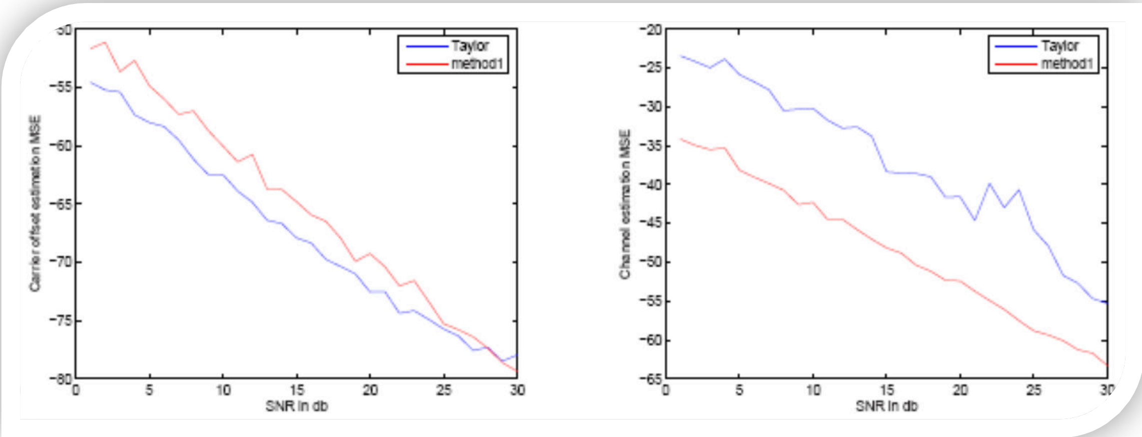
Figure 4.6: Taylor approach -Strong channel



(a) MSE carrier offset estimation in function with SNR

(b) MSE channel vector estimation in function with SNR

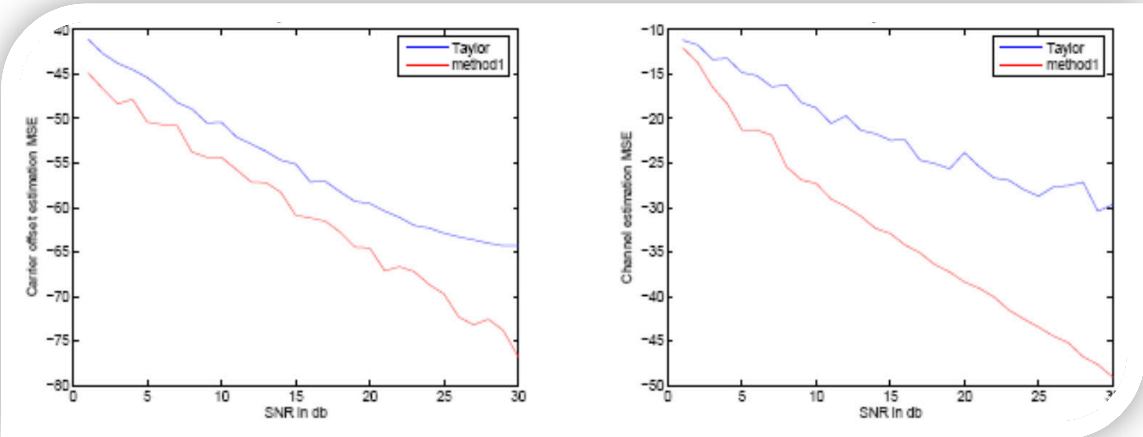
Figure 4.7: Taylor approach – Weak channel



(a) MSE carrier offset estimation in function with SNR

(b) MSE channel vector estimation in function with SNR

Figure 4.8: Comparison of the two techniques on the strong channel



(a) MSE carrier offset estimation in function with SNR

(b) MSE channel vector estimation in function with SNR

Figure 4.9: Comparison of the two techniques on the weak channel

4.5 Summary

In this chapter we introduced the carrier offset of the received CDMA signal and it was performed a joint channel and carrier offset estimation. It is demonstrated the MSE of carrier offset and channel error in function with the SNR for the two channels, the weak and the strong. The result was that with weak channel there is a better estimation of channel vector and with the strong channel there is better carrier offset estimation. Also it was performed a reduction to the generalized eigenvalue problem. The estimation is performed by sampling on φ in the range [-0.1 0.1] and choosing the φ(vector) which corresponds to the smallest eigenvalue of Q each time. But in this situation the dimension of Q is much larger, which makes the SVD analysis time-consuming. Closing this chapter we present a way of encountering of the problem which is based to the Z_i matrix approaching through Taylor series.

CHAPTER 5

RAISE POWER METHOD

5.1 Introduction

The estimation techniques that we analyzed before are based on two steps:

1. SVD analysis on a large matrix so as to take a base of the noise subspace and a base of the received signal subspace.

2. SVD analysis on a smaller dimension matrix for the channel and carrier offset estimation.

Basic drawback of these methods is the need for knowledge of the noise subspace. Taking into account only the ISI free symbols we end up that the signal subspace has dimension P, thus the numbers of users, while the noise subspace has a dimension of Lc-L+1-P. If we want to include all the samples of the composite signature, then the dimensions of the subspaces will not ne so obvious.

In this paragraph will we work with all the sample of the composite signature, and then we will see that we can replace the first SVD analysis with the computation of a power of a matrix, while will not need the knowledge of the noise subspace.

5.2 The Raised Power method

The (L_c + L – 1) samples from the nth symbol that are received at the receiver are given by:

$$y\left(nL_c+i\right)=s_n e^{jL_c\phi n}\underbrace{\sum_{k=1}^{L}h_k c_{i-k+1}e^{j\phi i}}_{\omega_i}\quad ,i=1,\ldots L_c+L-1$$

$$\Rightarrow\begin{pmatrix} y\left(nL_c+1\right)\\ y\left(nL_c+2\right)\\ \vdots\\ y\left(nL_c+\left[L_c+L-1\right]\right)\end{pmatrix}=\begin{pmatrix} w_1\\ w_1\\ \vdots\\ w_{L_c+L-1}\end{pmatrix}s_n e^{jL_c\phi n} \tag{83}$$

Where the samples of the composite signature are as follows:

$$\begin{pmatrix} w_1\\ w_1\\ \vdots\\ w_{L_c+L-1}\end{pmatrix}=\begin{pmatrix} e^{j\phi} & & & \\ & e^{j2\phi} & & \\ & & \ddots & \\ & & & e^{j(L_c+L-1)\phi}\end{pmatrix}\begin{pmatrix} c_1 & 0 & \cdots & 0\\ c_2 & c_1 & \ddots & \vdots\\ \vdots & c_2 & \cdots & 0\\ c_{L_c} & \vdots & \ddots & c_1\\ 0 & c_{L_c} & \cdots & c_2\\ \vdots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & c_{L_c}\end{pmatrix}\begin{pmatrix} h_1\\ h_2\\ \vdots\\ h_L\end{pmatrix}$$

$$\Rightarrow W_{(L_c+L-1)\times 1}=Z_{(L_c+L-1)\times(L_c+L-1)}C_{(L_c+L-1)\times L}h_{L\times 1} \tag{84}$$

We now focus now to the user of interest (Zi=Z, Ci=C, Hi=h). We assume that the user obtains, with its receiver N vectors of data, each one of these vectors constitute the overlapping of the symbols of the P users. Taking into account and the noise vectors which lay on the matrix $N_{(L_c+L-1)\times N}$, the N vectors which are obtained at the receiver will be as (55).

$$X_{(L_c+L-1)\times N}=W_{(L_c+L-1)\times P}S_{P\times N}+N_{(L_c+L-1)\times N} \tag{85}$$

Performing an SVD analysis now on the autocorrelation matrix of (3), then according to the analysis of the previous paragraph we would end up with:

$$\Rightarrow \mathbf{R}_{xx} = \left(\mathbf{U}_s\ \mathbf{U}_n\right)\left(\begin{array}{ccc|ccc} \sigma_{s_1}^2+\sigma_n^2 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \sigma_{s_n}^2+\sigma_n^2 & & & \\ \hline & & & \sigma_n^2 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & \sigma_n^2 \end{array}\right)\begin{pmatrix}\mathbf{U}_s^H \\ \mathbf{U}_n^H\end{pmatrix}$$

(86)

Exploiting the orthogonality between signal and noise subspace we would have:

$$\begin{aligned} &\mathbf{U}_n^H\mathbf{w}_i = 0 \Rightarrow \mathbf{U}_n^H\mathbf{ZCh} = 0 \\ &\Rightarrow \left(\mathbf{U}_n^H\mathbf{ZC}\right)^H\left(\mathbf{U}_n^H\mathbf{ZC}\right)\mathbf{h} = 0 \\ &\Rightarrow \left(\mathbf{C}^H\mathbf{Z}^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{ZC}\right)\mathbf{h} = 0 \end{aligned} \tag{87}$$

With matrix $U_nU_n^H$ to constitute the orthogonal projection matrix to the noise subspace, using equation (87) we have:

$$\Rightarrow \mathbf{R}_{xx}^{-1} = \left(\mathbf{U}_s\ \mathbf{U}_n\right)\left(\begin{array}{ccc|ccc} \frac{1}{\sigma_{s_1}^2+\sigma_n^2} & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \frac{1}{\sigma_{s_n}^2+\sigma_n^2} & & & \\ \hline & & & \frac{1}{\sigma_n^2} & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & \frac{1}{\sigma_n^2} \end{array}\right)\begin{pmatrix}\mathbf{U}_s^H \\ \mathbf{U}_n^H\end{pmatrix} \tag{88}$$

$$\Rightarrow \sigma_{\mathrm{n}}^2 \mathbf{R}_{xx}^{-1} = \left(\mathbf{U}_s\ \mathbf{U}_n\right) \left(\begin{array}{ccc|ccc} \frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}_1}^2+\sigma_{\mathrm{n}}^2} & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}_{\mathrm{n}}}^2+\sigma_{\mathrm{n}}^2} & & & \\ \hline & & & 1 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & 1 \end{array}\right) \begin{pmatrix}\mathbf{U}_s^H \\ \mathbf{U}_n^H\end{pmatrix} \tag{59}$$

$$\Rightarrow \left(\sigma_{\mathrm{n}}^2 \mathbf{R}_{xx}^{-1}\right)^k = \left(\mathbf{U}_s\ \mathbf{U}_n\right) \left(\begin{array}{ccc|ccc} \left(\frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}_1}^2+\sigma_{\mathrm{n}}^2}\right)^k & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \left(\frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}_{\mathrm{n}}}^2+\sigma_{\mathrm{n}}^2}\right)^k & & & \\ \hline & & & 1 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & 1 \end{array}\right) \begin{pmatrix}\mathbf{U}_s^H \\ \mathbf{U}_n^H\end{pmatrix} \tag{60}$$

$$\Rightarrow \lim_{k\to\infty}\left(\sigma_{\mathrm{n}}^2 \mathbf{R}_{xx}^{-1}\right)^k = \left(\mathbf{U}_s\ \mathbf{U}_n\right) \left(\begin{array}{ccc|ccc} 0 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & 0 & & & \\ \hline & & & 1 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & 1 \end{array}\right) \begin{pmatrix}\mathbf{U}_s^H \\ \mathbf{U}_n^H\end{pmatrix} \tag{89}$$

Finally,

$$\Rightarrow \lim_{k\to\infty}\left(\sigma_{\mathrm{n}}^2 \mathbf{R}_{xx}^{-1}\right)^k = \mathbf{U}_n\ \mathbf{U}_n^H \tag{90}$$

From equation (90) it can be observed that by raising the inverse of the autocorrelation matrix to a power we can approach the projection matrix $U_n U_n^H$, where the degree of the power does not need to be higher than three (k=3). So from (87) and (90) we get:

$$
\begin{aligned}
&\left(\mathbf{C}^H \mathbf{Z}^H \left(\sigma_n^2 \mathbf{R}_{xx}^{-1}\right)^k \mathbf{ZC}\right)\mathbf{h} = 0 \quad ,k = 1,2,3 \\
&\Rightarrow Q_{L\times L}\mathrm{h}=0
\end{aligned}
\tag{91}
$$

We end up to the known problem of eigenvalues but without performing an SVD analysis to the autocorrelation matrix, so as to define a base for the noise subspace. We must to mention here, that this is performed without taking into account only the inter-symbol-interference free vectors which belong to the composite signature of the transmitter [96][99]. In addition, it is important to be mentioned that our method is not concerned with the dimension of the noise subspace. So now, from the moment that we have to our service an estimation of the autocorrelation matrix (91) becomes $Q_{L\times L}\mathrm{h} \approx 0$ [45].

So the steps required performing the joint blind channel estimation and carrier offset power method are:

1. Calculate the autocorrelation matrix raised to the power k where k=1, 2 and 3.
2. Sampling of φ in space [-0.1 0.1], where it is assumed to fluctuate, the smallest values of φ are stored in matrix Q. The estimation of the carrier offset it would be that φ which gives the smallest eigenvalues. The estimation of the channel is the eigenvector which corresponds to that eigenvalues.

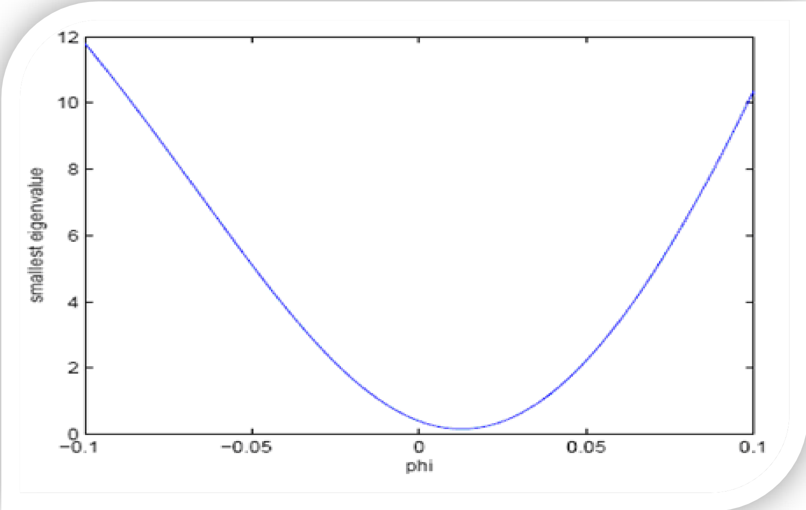
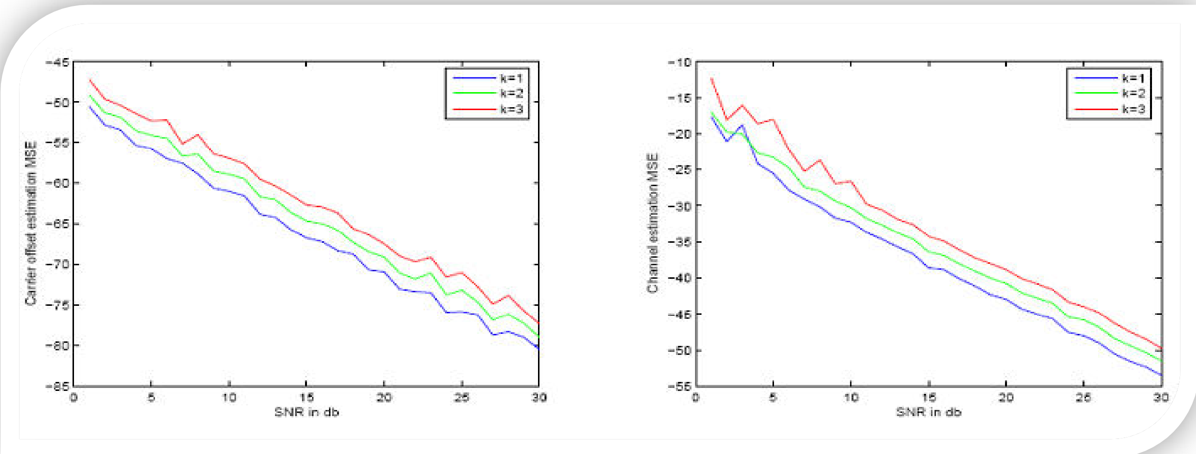


Figure.5.1: The smallest eigenvalues in conjunction with samples (phi) φ - Strong channel with k=1.

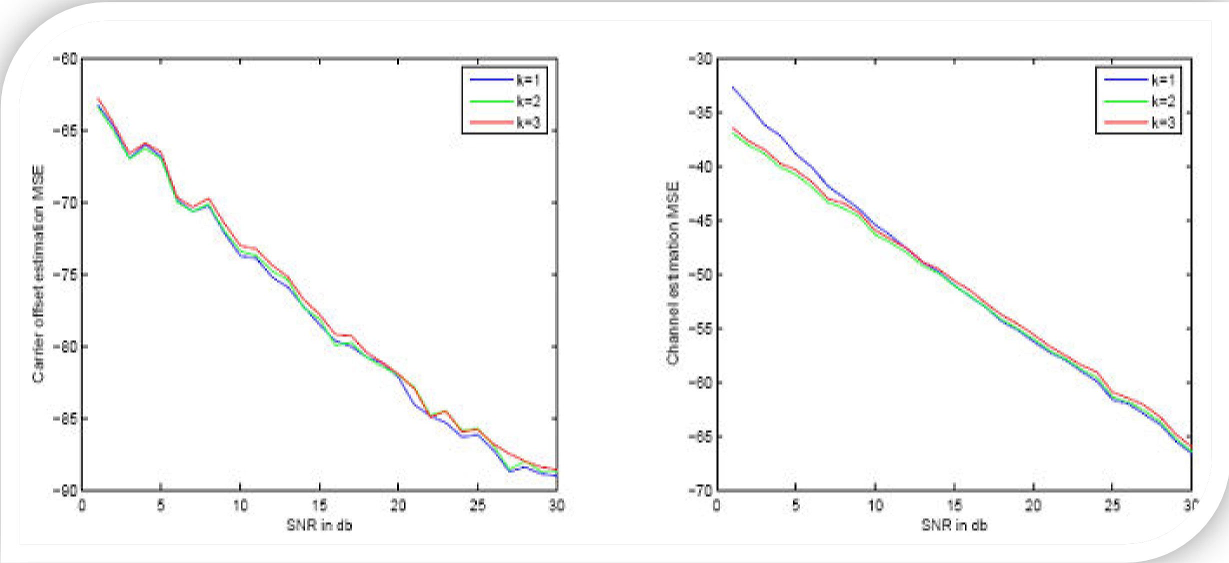
$$\phi : 0.0122$$
$$\hat{\phi} : 0.0126$$
$$h = 0.4070 \ 0.8150 \ 0.4070$$
$$\hat{h} = -0.4081 \ \ -0.8181 \ \ -0.4050$$

Then we calculate the MSE (mean square average) error of the estimation of the carrier offset and of the channel for an (signal to noise ratio) for various SNR values using N=100, N=1000 and N=10000 symbols for the generation of the autocorrelation matrix [20][22]. We observe that only when the arithmetic average is big enough (N>>) the rule of big numbers applies, by acquiring better results for k>1.



(a) MSE carrier offset estimation against SNR (b) MSE Channel vector estimation against SNR

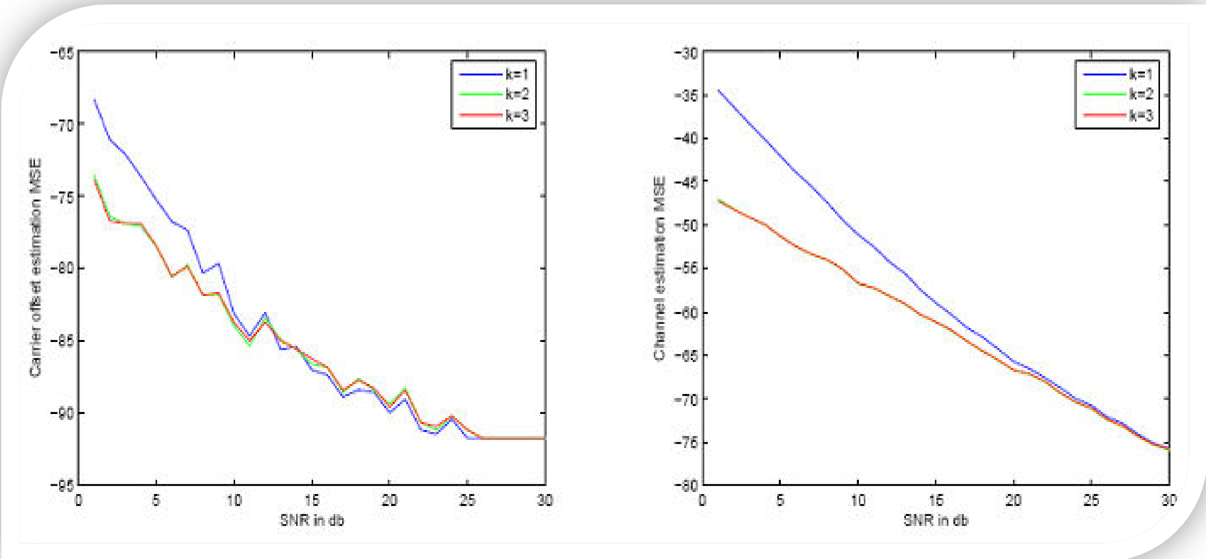
Figure 5.2: Formulation of autocorrelation matrix, strong channel with N=100 symbols



(a) MSE carrier offset estimation in conjuction with SNR

(b) MSE channel vector estimation in in conjuction with SNR

Figure 5.3: Formulation of autocorrelation matrix, strong channel with N=1000 symbols



(a) MSE carrier offset estimation in conjunction with SNR

(b) MSE channel vector estimation in conjunction with SNR

Figure 5.4: Formulation of autocorrelation matrix, strong channel with N=10000 symbols

Concluding we simulate the power raising method with that one of the subspace decomposition and then to the salvation of a polynomial eigenvalue problem in environment with SNR=10dBB and SNR=20dB, in conjunction with the number of symbols being received.

The autocorrelation matrix is being estimated in advance as follows:

$$\mathbf{R}_{\tilde{x}\tilde{x}}(n) = \lambda \mathbf{R}_{\tilde{x}\tilde{x}}\left(n-1\right) + x(n)x^{H}(n) \tag{92}$$

With λ = 0.997 to correspond to a $\frac{1}{1-\lambda} = 333.333$ samples window, while the inverse autocorrelation matrix which is used in the power method is:

$$\mathbf{R}_{\tilde{x}\tilde{x}}^{-1}\left(n\right) = \frac{1}{\lambda}\left[\mathbf{R}_{\tilde{x}\tilde{x}}^{-1}\left(n-1\right) - \frac{\mathbf{R}_{\tilde{x}\tilde{x}}^{-1}\left(n-1\right)x(n)x^{H}(n)\mathbf{R}_{\tilde{x}\tilde{x}}^{-1}\left(n-1\right)}{\lambda + x^{H}(n)\mathbf{R}_{\tilde{x}\tilde{x}}^{-1}\left(n-1\right)x(n)}\right] \tag{93}$$

With initial value:

$$\mathbf{R}_{\tilde{x}\tilde{x}}^{-1}\left(0\right) = \delta\mathbf{I} \quad , \delta = \frac{100}{\sigma_n^2} \tag{94}$$

5.3 Simulation Results

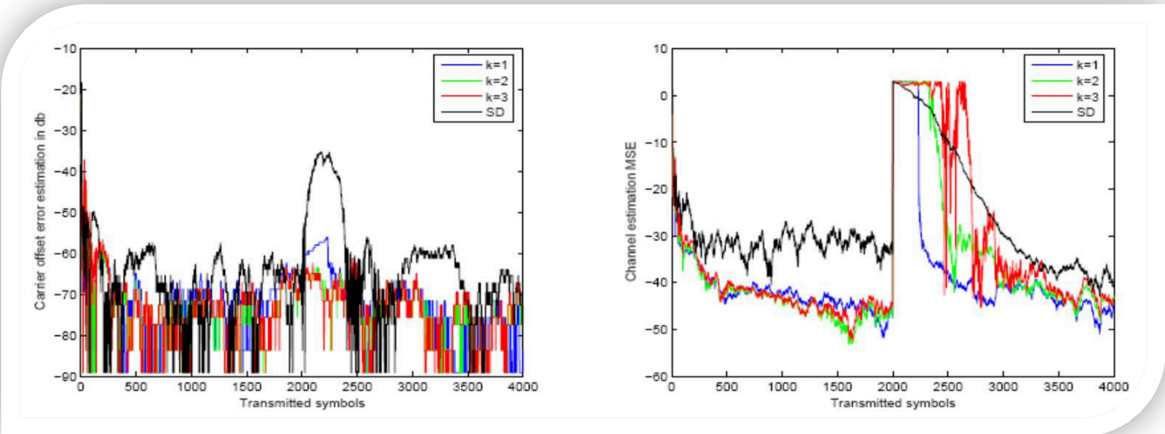
It is examined 4000 symbols windows at the receiver, by performing either a channel change at symbol 2000 or a carrier offset change at symbol 2000. Initially the error in all methods is the same and with the passing of symbols the power method converges first and downgrades the error estimation level faster.

It can be observed from all the simulation windows, that the power method converges faster to lower estimation levels compared to Subspace Decomposition (SD) methods. After the channel change or the carrier offset changes at 2000 symbol, for k=1 the method downgrades faster the estimation error requiring approximately 400 symbols, for k=2 approx. 500 and for k=3 approx. 700. It can be observed from the simulation windows that the raised power method is better compared to the SD method.. In order to illustrate the performance of our method in this paper we use

BPSK modulation with number of users P=10, spreading gain L_c=32, and N=100 the received number of data vectors at the receiver. The SNR is set at 10dB and 20dB.

It can be observed the behavior of our method using k=1, 2 and 3 in comparison with SD. We can clearly see that after a small number of transmitted symbols (100) our method has lower carrier offset estimation error in dB than the SD method. Moreover, when we have a change of the serving channel chosen to be at symbol 2000, our method performs better immediately after the change since the carrier offset estimation error is at lower levels for the power method in comparison with the SD method. When we trigger a change in the carrier offset (φ), the proposed method performs better again since the channel estimation error is at lower levels before and after the change in phase (φ) occurred at transmitted symbol 2000.

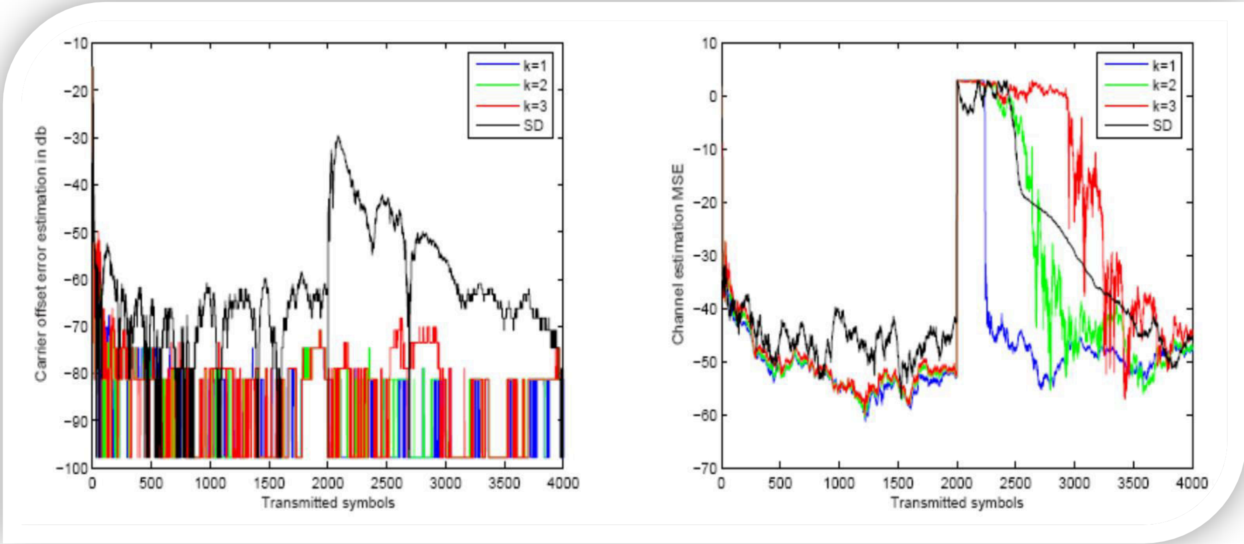
It is worth mentioning that when one parameter changes then estimation of the other parameter is being affected too, despite of the fact that the value of this parameter may not change. The power method performs better to this phenomenon in comparison with the subspace analysis method. Finally, as it was expected for SNR=20dB we get smaller errors for estimation in comparison with SNR=10dB.



(a) Error in carrier offset estimation in conjunction with symbols.

(b) Error in channel vector estimation received in conjunction with received symbols.

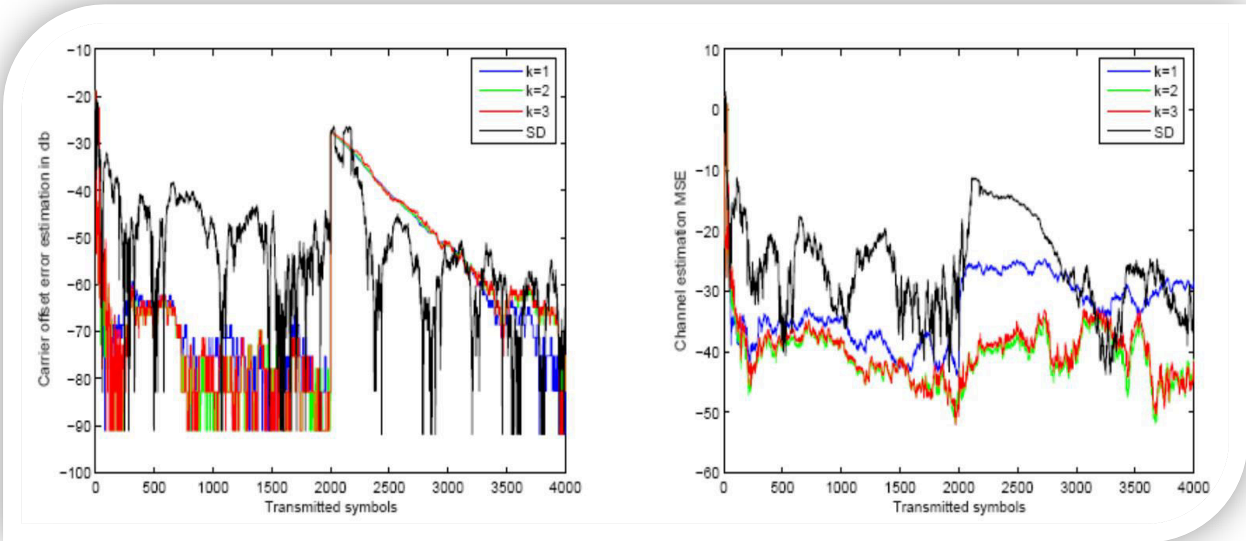
Figure 5.5: Channel change – SNR=10dB



(a) Error in carrier offset estimation in conjunction conjunction with received symbols

(b) Error in channel vector estimation in conjunction with received symbols.

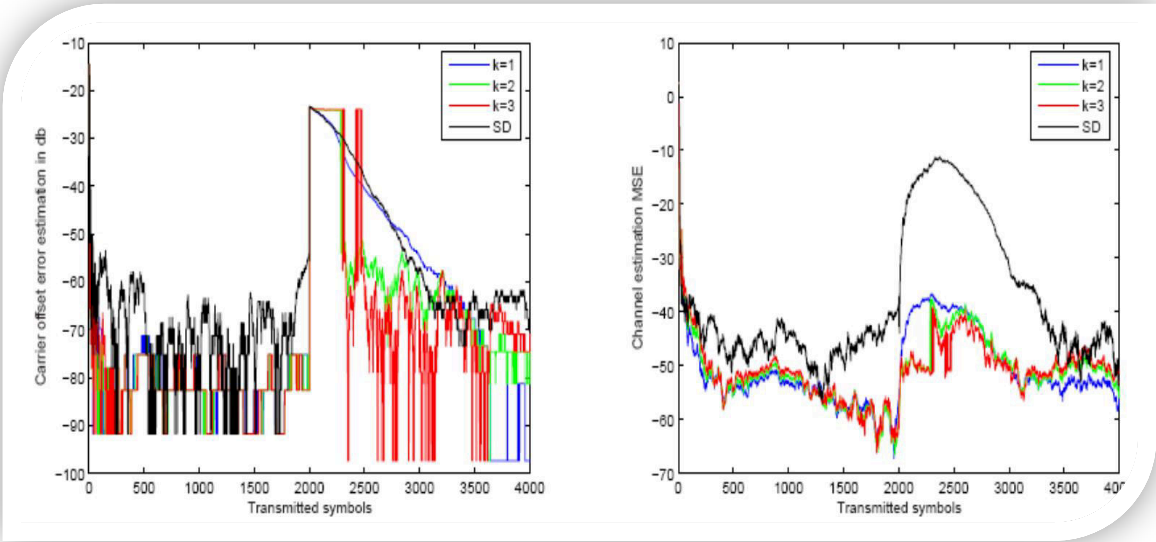
Figure 5.6: Channel change – SNR=20dB



(a) Error in carrier offset estimation in conjunction conjunction with received symbols.

(b) Error in channel vector estimation in conjunction with received symbols

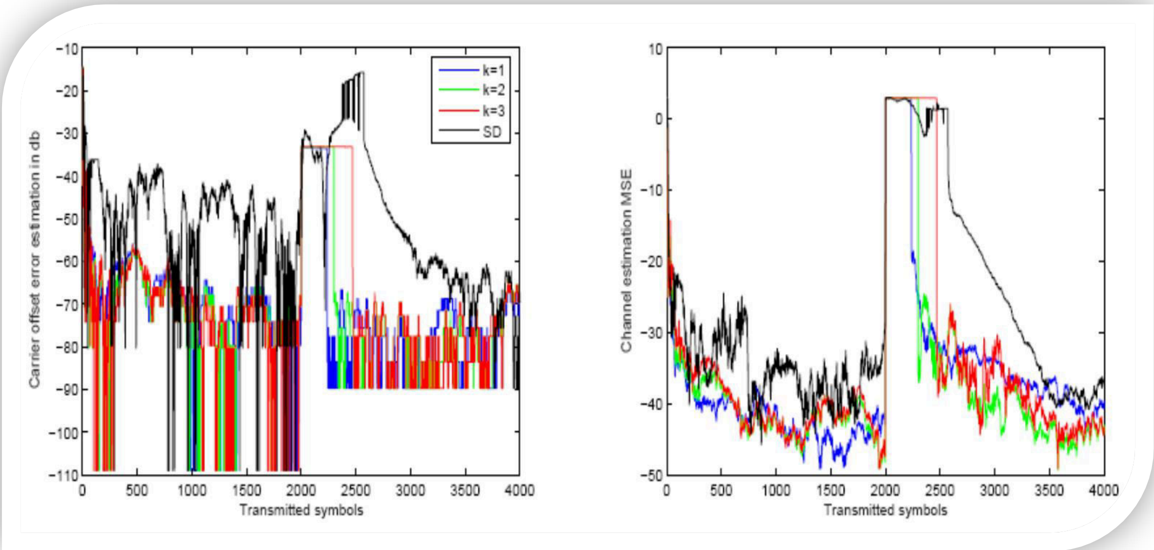
Figure 5.7: Change in the carrier offset parameter – SNR=10dB



(a) Error in carrier offset estimation in conjunction with received symbols .

(b) Error in channel vector estimation in conjunction with received symbols.

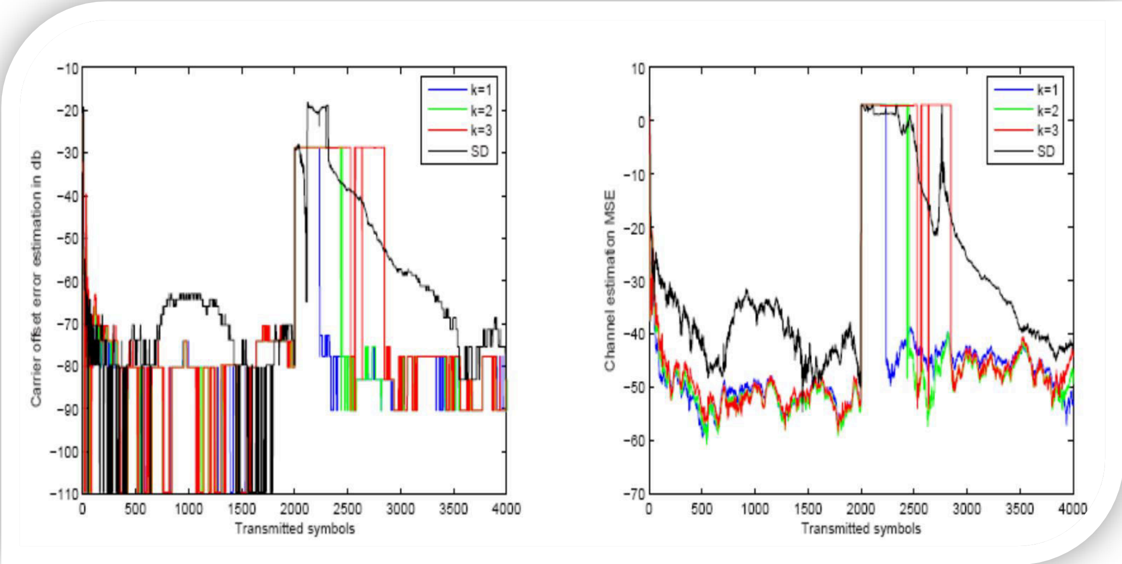
Figure 5.8: Change in the carrier offset parameter – SNR=20dB



(a) Error in carrier offset estimation in conjunction conjunction with received symbol

(b) Error in channel vector estimation in conjunction with received symbols.

Figure 5.9: Change in both the channel and the carrier offset parameter – SNR=10dB



(a) Error in carrier offset estimation in conjunction with received symbols.

(b) Error in channel vector estimation in conjunction with received symbols.

Figure 5.10: Change in both the channel and the carrier offset parameter – SNR=20dB

5.4 Summary

In this part of the thesis we presented a novel method for joint channel and carrier offset estimation for CDMA communication systems. Our method is based on a two-step methodology including the offset carrier parameter in the power method presented in [4] which has the advantage of reducing a two-step SVD analysis to a single step. Also, the performance of this method is independent of the knowledge of the signal subspace rank whereas the approaches in [1], [2], [3] are sensitive to correct knowledge of this parameter. As a result our technique performed better compared to other existing techniques at a significantly lower computational cost.

CHAPTER 6

SUPPORT VECTOR MACHINES

6.1 Introduction

In this chapter we will first examine the problem of classification of two linearly separable classes and then we will look into the situation where the prototypes for classification belong to two non linear separable classes.

6.2 Linearly separable classes

We assume a set of N prototypes $x_i \in R^n; i = 1,\ldots M,$ which is divided in two classes C1 and C2 which are linearly separable. We will examine the problem of classification of these two classes. If M<n, thus the number of components is bigger than from the number of prototypes, then there is always a hyperlevel which separates them [83].

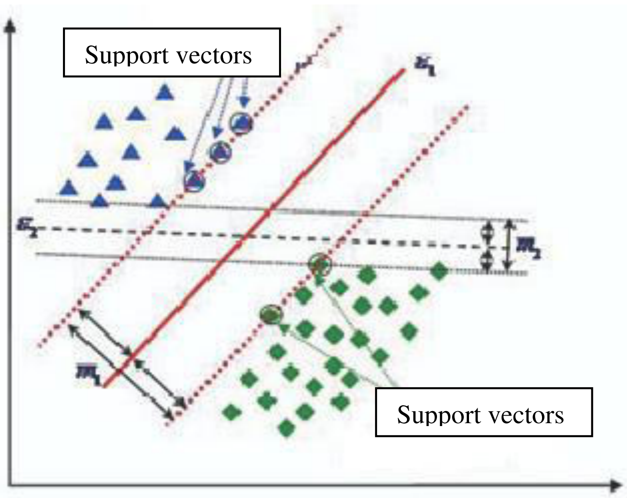


Figure 6.1: Support vectors

On figure 6.1, ε1 and ε2 are two hyperplanes which are separable in two classes and they constitute equivalent solutions of the simple perception. ε1 however constitutes what absolute separates the two classes, according to the maximum margin criterion [83].

The hyperplane to be far of equal distances from the closest prototypes of the two classes [83].

This margin to be the maximum possible.

The prototypes of both classes which equal-distant from the margin of the dividing hyperplane are called support vectors. The equation of the dividing plane discriminant function is as follows:

$$g(x) = w^T x + w_o = 0 \tag{96}$$

The w defines the address of the hyperplane, while w_o defines the exact position on space. Our target is to seek that direction which gives us the maximum possible margin between the two classes.

So given that the two classes are linearly separable there is a hyperplane $w^T x + w_o = 0$ and one positive ρ>0 so as to:

$$w^T x + w_0 \geq 1 \text{ , when } \frac{w^T x_{S1} + \omega_0}{\|w\|} + \frac{w^T x_{S2} + \omega_0}{\|w\|} = \frac{|1|}{\|w\|} + \frac{|-1|}{\|w\|} = \frac{2}{\|w\|}$$

$$w^T x + w_0 \leq 1 \text{ , when } \qquad z = \frac{|g(x)|}{\|w\|} = \frac{|w^T x + w_0|}{\|w\|} \tag{97}$$

When the x prototypes constitute support vectors then the equations will apply. The distance of a point from line ε1 (the plane of interest in this situation) is:

$$z = \frac{|g(x)|}{\|w\|} = \frac{|w^T x + w_0|}{\|w\|} \tag{98}$$

Based on the above observation, if x_{s1} a support vector of class C1 and x_{s2} a support vector of class C2, then the margin between the two classes will be as follows:

$$\frac{w^T x_{s1} + \omega_0}{\|w\|} + \frac{w^T x_{s2} + \omega_0}{\|w\|} = \frac{|1|}{\|w\|} + \frac{|-1|}{\|w\|} = \frac{2}{\|w\|} \tag{99}$$

For every xi, i=1,…,N we the corresponding label (class indicator) yi (exits targets as in the simple perception case) will be as follows:

$$\forall x_i \in C_1, w^T x_i + \omega_0 \geq 1 \ , \ y_i = 1$$

$$y_i\left(w^T x_i + \omega_0\right) \geq 1 \ , \ i=1,...,N \tag{100}$$

Equally we can write: $y_i\left(w^T x_i + \omega_0\right) \geq 1 \ , \ i=1,...,N$. We end up to the following optimization problem:

minimize $\lambda_i \geq 0$

subject to $y_i\left(w^T x_i + \omega_0\right) \geq 1 \ , \ i=1,...,N$ (101)

The cost function which is to be minimized J(w) is strict convex from the moment its Hessian matrix is positively defined. Moreover, the inequality constraints are constituted from linear functions. These observations guarantee that the local minimum will be simultaneously be overall and unique.

We define the Lagrangian equation:

$$L\left(w,\omega_0,\underline{\lambda}\right)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{N}\lambda_i[\ y_i(\ w^T x_i+\omega_0)\ -1\ )\ ] \tag{102}$$

The Karush-Kuhn-Tucker conditions are as follows:

$$\vartheta_{\omega_0} L\left(w,\omega_0,\underline{\lambda}\right)=0 \Rightarrow \sum_{i=1}^{N}\lambda_i y_i=0 \tag{103}$$

$$\vartheta_W\, L\left(w,\omega_0,\underline{\lambda}\right)=0 \Rightarrow w\ =\sum_{i=1}^{N}\lambda_i y_i x_i \tag{104}$$

$\lambda_i \geq 0$, active constrain (λι>0) we have when xi is a support vector and not active (λi=0), when is not

$N_s \leq N$

The λi can be positive, when xi is support vector or zero in the case where is not. Therefore, w will compromise a linear combination of $\lambda_i \neq 0$ prototypes which correspond to $\lambda_i \neq 0$ and they constitute the classes support vectors.

From the moment they constitute the cost function is convex, and the sum of the feasible solutions constitutes a convex set, we can encounter the problem via the Langrange duality by formulating it to a Wolfe dual representation form, using the following conditions:

$$\sum_{i=1}^{N} \lambda_i y_i = 0 \tag{105}$$

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i \tag{106}$$

The Lagrangian function will be as follows:

$$L\left(w, \omega_0, \underline{\lambda}\right) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i [\, y_i (\, w^T x_i + \omega_0) \, - 1 \,) \,]$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i y_i w^T x_i + \sum_{i=1}^{N} \lambda_i$$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i y_i w^T x_i + \sum_{i=1}^{N} \lambda_i$$

$$= \frac{1}{2}\|w\|^2 = \frac{1}{2}[w^T w \,]$$

$$= \frac{1}{2}[\sum_{i=1}^{N} \lambda_i y_i x_i^T][\sum_{j=1}^{N} \lambda_j y_j x_j]$$

$$= \frac{1}{2}\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j$$

$$\sum_{i=1}^{N} \lambda_i y_i w^T x_i = \sum_{i=1}^{N} \lambda_i y_i [\, \sum_{j=1}^{N} \lambda_j y_j x_j^T] x_i$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j \tag{106}$$

By combining the above, we end up to the following dual problem of optimization:

$$\max_{\lambda} L(w, \omega_0, \underline{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j \tag{107}$$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0$

$$x \in \boldsymbol{R}^n \rightarrow \phi(x) \in \boldsymbol{R}^k, k >> n \text{ , i=1,…,N} \qquad (108)$$

We see that the prototypes are coming to the problem in the form of internal product and the cost function its not anymore depends from the prototypes dimension, like we had before, where w will have the same dimension with the prototypes. After the calculation of λι, the w and w0 they will be as follows:

$$w = \sum_{i=1}^{N} \lambda_i y_i x_i \,, \quad \omega_0 = -\frac{\sum_{i=1}^{N} \lambda_i w^T x_i}{\sum_{i=1}^{N} \lambda_i} \qquad (109)$$

6.3 Non Linear Separable Classes

We observe now the situation where the prototypes for classification belong to two non linear separable classes. The Cover theorem declares that such a space can be transformed to a new space, where the prototyped are now linearly separable with a big probability, given that the following applies:

1. the transformation is non-linear

2. the dimension of the new space is large enough

Supposing $x \in R^n$ a prototype to the original dimension space n, which can belong to the class C1 or C2 which are not linear separable. We suppose the k non linear functions:

$\Phi j(.) : Rn \rightarrow R$, j=1,…,k (110)

Which define the depiction:

$$x \in R^n \rightarrow \phi(x) \in R^k, k >> n \quad (111)$$

where:

$$\varphi(x)=[\varphi_1(x), \varphi_2(x),\ldots, \varphi_k(x)]^T \quad (112)$$

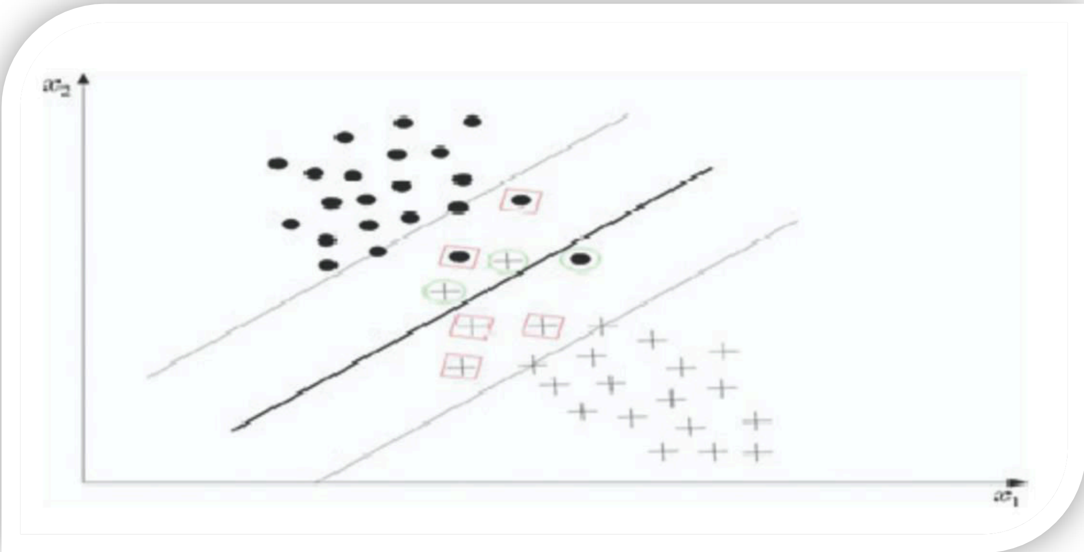


Figure 6.2 : Class Separation

Our target now its to see, if there exists a suitable value for k and the $\varphi_j(.)$ so as the classes C1 and C2 to be linearly separable at the k-dimensional space which is defined by $\varphi(x)$. Thus, we are looking for a k-dimension space where we can construct a hyperplane $w \in R^k$ so as:

$$w^T \phi(x) + w_0 \geq 0, \forall x \in C_1$$
$$w^T \phi(x) + w_0 \leq 0, \forall x \in C_2 \quad (113)$$

If it is supposed that the initial space n the two classes are separated from a non linear hyperplane g(x)=0, then the above relationships constitute an approach of the non linear function g(x) and the linear combination of $\varphi_j(x)$:

$$g(x) = w_0 + \sum_{i=1}^{k} w_i \phi_i(x) \tag{114}$$

That end to the classical problem of function approach through a class interopolation function. From the moment φ_j (.) are defined the problem is transformed to a linear classifier, where the estimation of w and w_0 is demanded.

Setting that $\phi_0(x) = 1 \forall x$ we end up to:

$$g(x) = \sum_{i=0}^{k} w_i \phi_i(x) \Rightarrow g(x) = w^T \phi(x) \tag{115}$$

Therefore the new space dimension of the separable hyperplane w will be:

$$w = \sum_{i=1}^{N} \lambda_i y_i \phi(x_i) \tag{116}$$

Hence the equation of the separable hyperplane becomes:

$$g(x) = \sum_{i=0}^{N} \lambda_i y_i \phi^T \phi\ (x) \tag{117}$$

The term $\varphi^T(x_i)\varphi(x)$ represents an internal product to a space of larger dimension. The inner-product kernel is defined as follows:

$$K(x\ , x_i) = \phi^T(x)\phi\ (x_i) = \sum_{j=0}^{k} \phi_j(x\ )\phi_j(x_i)\ ,\ \text{i=1,…,N} \tag{118}$$

and it constitutes a symmetrical function, thus $K(x,x_i)=K(x_i,x)$.

The eq.(118) declares that the inner-product of the prototypes in the new bigger dimension space that have been visualized, its expressed as a function of their inner product to a space of smaller dimension.

The space of larger dimension is known as Reproducing Kernel Hilbert Space (RKHS).

Coming back to the problem of optimal separating hyperplane, and finding the appropriate kernel, that will indirectly express the representation to a larger dimension space, the Wolfe dual problem becomes:

$$\max_{\lambda} L(w, w_0, \underset{-}{\lambda}) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j K(x_i x_j) \tag{119}$$

Subject to $\sum_{i=1}^{N} \lambda_i y_i = 0$

$\lambda_i \geq 0, \ i = 1, \ldots, N$

and after the finding of λ_i the classification of the unknown prototype x will be as follows:

$$g(x) = \sum_{i=1}^{Ns} \lambda_i y_i K(x_i, x) + w_0 > 0, \ \forall x \in C_1 \tag{120}$$

$$g(x) = \sum_{i=1}^{Ns} \lambda_i y_i K(x_i, x) + w_0 < 0, \ \forall x \in C_2 \tag{121}$$

where Ns is the volume of support vectors, since for these corresponding λ_i is not zero.

Closing we have to mention that the SVM optimization problem can be written and as a normalization problem:

$$\min_{f,b} \frac{1}{n} \sum_{i=1}^{n} L(y_i(f(x_i) + b)) + \lambda \|f\|^2 \tag{122}$$

where $f(x_i)=w^T\varphi(x_i)$. $\|f\| := \|w\|$ and L is the loss function. The role of this normalization parameter λ is to limit the space of feasible solutions. By choosing

L(z)=max(0,1-z) we get a soft margin SVM . with $L(z)=max(0,1-z)^2$ we have a 2-norm soft margin SVM while for $L(z)=(1-z)^2$ we get the least squares SVM.

6.4 Summary

Closing the chapter it was demonstrated, that the linearly classes and the prototypes are coming to the problem in the form of internal product and the cost function is not depend on the prototypes dimension, where w will have the same dimension with the prototypes.

As far as the problem of optimal separating hyperplane, and having find the appropriate kernel, which will indirectly express the representation to a larger dimension space we present the Wolfe dual problem. Finally, it has to be mentioned that the SVM optimization problem can be written and as a normalization problem, as it was discussed.

CHAPTER 7

PROJECTIONS ONTO CONVEX SETS

7.1 Introduction

In the following pages it is examined a projection of a close convex set in a Hilbert Space. Referred to that we will look at the modelling through convex sets, afterwards, having secure the closeness and convexity of a set, the next step will be to define the projection operator. At last we mention projections onto convex sets algorithms

7.2 Projection of Convex Set in Hilbert Space

Supposing C a close convex set in a Hilbert space H $\forall x \in H \exists$ a unique

$x^{*} = P_c x$, which is close to x, thus:

$$\left\| x - x^{*} \right\| = \min_{y \in C} \left\| x - y \right\| \tag{123}$$

The x^* constitutes the projection of $x \in H$ to a closed convex set $C \subseteq H$ and is given as follows:

$$x^{*} = P_c x \tag{124}$$

with P_c to constitute the projection operator on C.

Let it be W a closed convex set in a Hilbert H space. If $W^{\perp}$ is the orthogonal complement of W, then $H = W + W^{\perp}$ and $W \cap W^{\perp} = \{0\}$ which means $\forall x \in H$ we have:

$$x = x1 + x2\,,\ \ x_1 \in W \text{ and } x_2 \in W^{\perp} \tag{125}$$

The x1 constitutes the projection of x on W, $x_1 = P_w x$, while x2 constitutes the projection of x on $W^{\perp}, x_2 = P_w \perp x$, with $P_{w^{\perp}} = I - P_w$. For a convex set on a Hilbert space, the projector operator is defined uniquely. An expansion which constitutes a relaxed operator:

$$Tc = I + \lambda(Pc - I),\ \ \lambda \in (0,2) \tag{126}$$

With $\lambda=1$ we get $T_c=P_c$. Moreover,

$$\begin{aligned} Tcx &= [I+\lambda(P_c-I)]x \\ &= x+\lambda(P_c-I)x \\ &= (1-\lambda)x+\lambda P_c x \end{aligned}$$

If $x \in C$, then $P_c x=x$ and $T_c x=x$. If $x \notin C$ then $\|T_c x - y\| < \|x - y\|$, thus T_c brings x through projection closer to group C [104].

Basic POCS theory:

Let it be $C_1, C_2, \ldots, C_m$ closed convex sets in a Hilbert space and $C_0 = \bigcap_{i=1}^{m} C_i \neq 0$ their section which will also constitute a convex set. Setting $T = T_{c_m} T_{c_{m-1}} \ldots T_{c_1}$ the synthesis of all relaxed operators.

$\forall x \in H$ and $\forall \lambda_i \in (0,2), i = 1,...,m$ the sequence $\{T^n x\}$ converge weakly to a point of C_0. If anyone of $C_1, C_2, \ldots, C_m$ constitutes a closed subspace, then we have a strong convergence to point Pc_0x thus the projection of $x \in H$ on C_0.

Modelling through Convex Sets:

Let it be the linear equation system Ax=b, $A \in R^{mxn}$, $b \in R^m$. We can write it in the form:

$$\begin{aligned} \langle \mathrm{a}_1, x \rangle &= b_1 \\ \langle \mathrm{a}_2, x \rangle &= b_2 \\ &\vdots \\ \langle \mathrm{a}_m, x \rangle &= b_m \end{aligned}$$

Supposing the vector y verifies equation $\langle a_i, x \rangle = b_i$. The sum of possible solutions of this equation is defined as follows:

$$C_i = \{ y : \langle \mathrm{a}_i, y \rangle = b_i \} \tag{127}$$

For the m equations of the system they are formed m groups of possible solutions C_1, $C_2,\ldots,C_m$. The solution y* of the system will satisfy the m equations, so it will be at the trace of the sum of possible solutions:

$$y^* \in C_0 \bigcap_{i=1}^{m} C_i \tag{128}$$

If the system has no solution, then $C_0 \neq 0$, otherwise C_0 will contain the unique solution or an infinity number of elements which constitute a solution to the system.

When $C_0 \neq 0$ in order to find a solution to the system we apply POCS. Initially though we have to secure that C_i are close convex sets.

The C_i will be convex if for any 2 elements of y_1 and y_2, their convex combination $y_3=\alpha y1+(1-\alpha)y_2$, $\alpha \in [0,1]$ also constitutes element of C_i . a

$$\begin{aligned} \langle \mathrm{a}_i, y3 \rangle &= \langle \mathrm{a}_i, ay_1 + (1-a)y_2 \rangle \\ &= \langle \mathrm{a}_i, ay_1 + \mathrm{a}_i, (1-a)y_2 \rangle \\ &= a\langle \mathrm{a}_i, y_i \rangle + (1-a)\langle \mathrm{a}_i, y_2 \rangle \\ &= ab_i + (1-a)b_i \\ &= b_i \end{aligned} \tag{129}$$

So, $y_3 \in C_i$ and the C_i constitute a convex set.

The Ci will be closed if the limit y*of the converging sequence {yk} which is contained in Ci, also occurs in Ci. From the inequality Schwartz we have:

$$\left|\langle \mathrm{a}_i, y_k - y^* \rangle\right| \leq \|\mathrm{a}_i\| \|y_k - y^*\|$$

Due to convergence that will be:

$$\lim_{k\to\infty} \left\| y_k - y^* \right\| = 0$$

so:

$$\langle \mathrm{a}_i, y_k - y^* \rangle = 0 \Rightarrow \langle \mathrm{a}_i, y_k \rangle = \langle \mathrm{a}_i, y^* \rangle = b_i \tag{130}$$

Having guaranteed the closeness and convexity of Ci, the next step is to define the projection operator Pc_i on Ci. The projection $Pc_i x$ of a vector x on Ci constitutes the vector $y \in C_i$, which minimizes the distance $\|y - x\|$.

So we have the following optimization problem:

minimize $\|y-x\|^2$
subject to $\langle a_i, y\rangle = b_i$

We define the Lagrangian function:

$$
\begin{aligned}
&L(y,\lambda) = \|y-x^2\| + \lambda(\langle a_i, y\rangle - b_i)\\
&= (y-x)^T (y-x) + \lambda(\langle a_i, y\rangle - b_i)\\
&= y^T y - y^T x - x^T y + x^T x + \lambda(\langle a_i, y\rangle - b_i)
\end{aligned}
$$

$$
\begin{aligned}
&\partial_y L(y,\lambda) = 0 \Rightarrow 2y - 2x + \lambda a_i = 0\\
&\partial_y L(y,\lambda) = 0 \Rightarrow a_i^T y - b_i = 0 \Rightarrow \frac{\lambda}{2} = \frac{\langle a_i, x\rangle - b_i}{\|a_i\|^2}
\end{aligned}
$$

Combining the above relationships comes up:

$$
y \equiv P_{c_i} x = x - \frac{\langle a_i, x\rangle - b_i}{\|a_i\|^2} a_i \tag{131}
$$

So starting from an initial value x_0, the repetition

$$
x_{n+1} = P_{c_m} \ldots P_{c_2} P_{c_1} x_n \tag{132}
$$

converges to a point of the total of possible solutions, which will constitute the solution to the system. When the repetition approaches this point, let it be x^*, it would be:

$$
x^* = P_{c_1} x^* = P_{c_1} x^* = \ldots = P_{c_m} x^* \tag{133}
$$

meaning that $x^* \in C_0$. Otherwise, $C_0 \neq 0$, the system has no solution and the algorithm is not converging.

POCS Algorithm :

Supposing $C_1, C_2, \ldots, C_m$ m convex sets on a Hilbert space and $C_0 = \bigcap_{i=1}^{m} C_i \neq 0$ their trace, which will constitute also a convex set. Let it be Pc_i the projection operator on group C_i.

$\forall x_0 \in H$ and $w_i \geq 0$ such as $\sum_{i=1}^{m} w_i = 1$, the sequence $\{x_n\}$ which is created from:

$x_{n+1} = \sum_{i=1}^{m} w_i P_{c_i} x_n$, (convex combination)

it converges weakly to a point of C_0. A more generalized morph of the above equation is as follows:

$$x_{n+1} = x_n + \lambda \sum_{i=1}^{m} (w_i P_{c_i} x_n - x_n), \quad 0<\lambda<2. \qquad (134)$$

In every repetition the x_n is projected to all Ci and then we have a sum with these projections weights, which occur in parallel [10]. On the contrary, on the repetition $x_{n+1} = P_{c_m} \ldots P_{c_2} P_{c_1} x_n$ the x_n is projected to the groups sequentially, thus first at C1: $P_{c_1} x_n$ after at C2: $P_{c_2}(P_{c_1} x_n)$ etc. If we let λ to vary we end up to the following repetitive form:

$$x_{n+1} = x_n + \lambda_n \sum_{i=1}^{m} (w_i P_{c_i} x_n - x_n) \qquad (135)$$

where $\varepsilon < \lambda_n < 2 - \varepsilon$ for every $0 < \varepsilon < 1$. The definition of λ_n from repetition to repetition can be determined as follows (Pierra):

$$\varepsilon < \lambda_n < L_n$$
$$0 < \varepsilon < 1$$

$$L_n = \frac{\sum_{i=m}^{m} w_i \left\| P_{c_i} x_n - x_n \right\|^2}{\left\| \sum_{i=1}^{m} w_i P_{c_i} x_n - x_n \right\|^2} \qquad (136)$$

CHAPTER 8

ONLINE KERNEL-BASED CLASSIFICATION

8.1 Introduction

In this chapter we look at the classification through kernels, as well as online classification through adaptive project sub-gradient method (APSM). Furthermore there is a numerical example in last pages of the chapter.

8.2 Kernel Classification and APSM method

We consider the classification problem of a group of X vectors of R^n. The X is represented to a space of larger dimension H which constitutes a Reproducing Kernel Hilbert Space (RKHS). This representation is performed through kernel function $K : R^n x R^n \mapsto R$ which defines the representation $R^n \mapsto H$:

$$x \mapsto \phi(x) \equiv K(x,.) \tag{137}$$

With φ(.) to define the representation of R^n to H. The K(x,.) like any one point f of H, $\forall x \in R^n$ constitutes a function from $R^n \mapsto R$. Indeed at RHKS the following attribute is satisfied:

$$\langle (f, K(x,.) \rangle = f(x), \forall f \in H, \forall x \in R^n \tag{138}$$

resulting to:

$$\|K(x,.)\|^2 = \langle K(x,.), K(x,.) \rangle = K(x,x) \tag{139}$$

Supposing $C \subset H$ a convex set. Given that $f \in H$ we seek for the optimised way

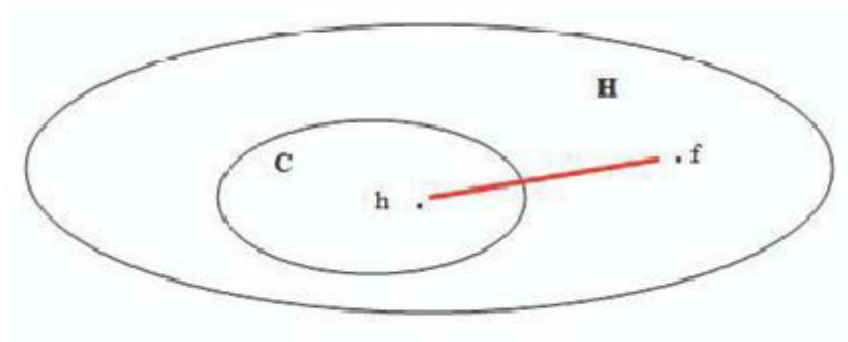


Figure 8.1: Convex projection

so as to move to a point C. It is enough to project f on C at the point where the distance is minimized:

$$\left\| f - P_c(f) \right\| = \min_h \left\| f - h \right\| \tag{140}$$

In order to find the projection operator Pc(.) on C it only requires to calculate vector h which minimizes the distance $\left\| f - h \right\|$. So, we have the following optimization problem:

$\min_h \left\| f - h \right\|^2$

subject to $\langle \mathrm{a}, h \rangle = \xi$

as we saw in chapter 7 the solution is:

$$P_c(f) \equiv h = f + \frac{\xi - \langle \mathrm{a}, h \rangle}{\left\| \mathrm{a} \right\|^2} \mathrm{a} \tag{141}$$

when $f \in C$, then Pc(f)=f, so the above equation becomes:

$$P_c(f) = f + \frac{\max\left\{0, \xi - \langle \mathrm{a}, h \rangle\right\}}{\left\| \mathrm{a} \right\|^2} \mathrm{a}, \forall f \in H \tag{142}$$

Classification through Kernels:

Supposing group $X = \{x_1, x_2, ...\} \subset R^n$ that is constituted from vectors is coming from two classes. $\forall x_i$ we define the label y_i which is $\{\pm 1\}$, according to the class it belong. Two pairs are created: {(x1,y1),(x1,y2),…}. Our target is given a margin $\rho \geq 0$ to define the function f(.), $(f \in H)$ and the offset $b \in R$ such as:

$$y(f(x)+b) \geq \rho, \forall (x,y) \tag{143}$$

The unknowns f(.) and b are assumed as components of a vector u=[f(.),b], which belongs to HxR constitutes a space with all possible classifiers. This vector constitutes and the wanted classifier while HxR constitutes the space with all the possible classifiers. Supplying this space with inner-product of the form:

$$\langle u_1, u_2 \rangle = \langle f_1(.), f_2(.) \rangle + b_1 b_2 \tag{144}$$

This is transformed to a Hilbert space. The wanted classifier will be located in the group:

$$C = \{u \in HxR : y(f(x)+b) \geq \rho\} \tag{145}$$

Eq.(145) is convex, from the moment that is defined as a positive half-plane. Combining equations (143), (144) and (145) gives:

$$\begin{aligned}
& y(f(x)+b) \geq \rho \Rightarrow yf(x) + yb \geq \rho \\
& \Rightarrow y\langle f(x), K(x,.) \rangle + yb \geq \rho \\
& \Rightarrow \langle [f(x),b],[yk(x,.),y] \rangle \geq \rho \\
& \Rightarrow \langle [f(x),b], y[k(x,.),1] \rangle \geq \rho \\
& \Rightarrow \left\langle u, \tilde{u} \right\rangle \geq \rho
\end{aligned} \tag{146}$$

Bases on the last equation, the group of the wanted classifiers can be written as:

$$C = \{u \in HxR : \langle u, u \rangle \geq \rho \tag{147}$$

It is observed that C constitutes a half-plane. Starting from an arbitrary $u \in HxR$, the best way of moving to a classifier it is through a projector.

The desired projection

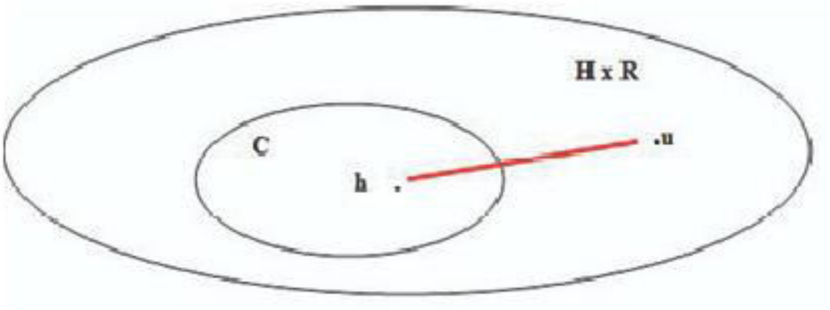


Figure 8.2: Projector (u) through classifier

constitutes a solution to the problem:

$$\min_h \|h - u\|^2$$

$$\text{subject to } \langle u, u \rangle \geq \rho \tag{148}$$

and it is

$$P_c(u) = u + \frac{\max\{0, \rho - \langle u, u \rangle\}}{\|u\|^2} u, \forall u \in HxR \tag{149}$$

Moreover, it is:

$$\left\|u\right\|^2 = \left\langle u, u \right\rangle$$

$$\begin{aligned}
&= \left\langle y[K(x,.),1], y[K(x,.),1] \right\rangle \\
&= \left\langle yK(x,.), yK(x,.) \right\rangle + y^2 \\
&= y^2[\left\langle K(x,.), K(x,.) \right\rangle + 1], y^2 = (\pm 1)^2 = 1 \\
&= K(x,x) + 1
\end{aligned} \tag{150}$$

Finally, the projection will be as follows:

$$P_c(u) = u + \frac{\max\{0, \rho - y(f(x)+b)\}}{K(x,x)+1} y[K(x,.),1], \forall u \in HxR \tag{151}$$

It is important at this point to define the following observation. At classification problems, the given margin ρ gives the loss function as it is used widely.

$$L(y(f(x)+b)) = \max\{0, \rho - y(f(x)+b)\} \tag{152}$$

If $y(f(x)+b) \geq \rho$ then we got the right classification and the margin is satisfied thus the vector which is to be classified lies at the class which belong and outside the margin of the two classes.

If $y(f(x)+b) \leq \rho$ then we have again right classification but the margin is not satisfied. At this situation the vector that is classified, lies within the margin but from the side of the class which belongs to, it is classified correctly.

Finally, if $y(f(x)+b)<0$ then we have a wrong classification. So, the desired classifier follows the above loss function as it is minimized.

The distance between an arbitrary $u \in HxR$ and the demanded classifier on C is:

$$D(u,P_c u)) = \|u - P_c(u)\| \tag{153}$$
$$= \left\| u - u + \frac{\max\{0, \rho - y(f(x)+b}{K(x,x)+1} y[K(x,.),1] \right\|$$
$$= \|ay[K(x,.),1\|, a = \frac{\max\{0, \rho - y(f(x)+b)\}}{K(x,x)+1}$$
$$= \langle ay[K(x,.),1], ay[K(x,.),1] \rangle^{1/2}$$
$$= (a^2 y^2 \langle K(x,.), K(x,.) \rangle + a^2 y^2)^{1/2}, y^2 = 1$$
$$= (a^2 K(x,x) + a^2)^{1/2}$$
$$= (aK(x,x)+1)^{1/2} \tag{154}$$

and the distance has finally as follows:

$$D(u,P_c(u)) = \frac{\max\{0, \rho - y(f(x)+b)\}}{\sqrt{K(x,x)+1}} = \frac{L(y(f(x)+b))}{\sqrt{K(x,x)+1}} \tag{155}$$

Therefore, the minimization of the distance $D(u,P_c(u))$ equals with the minimization of the loss function $L(y(f(x)+b))$.

Online Classification through adaptive Projected Subgradient Method (APSM)

Supposing now that the sequence of pairs {(xi,yi)} with the data which come from the two classes with their labels, as well as the sequence of the margins $\{\rho_i\}$. Every pair with the corresponding margin defines the half plane.

$$C_i = \{u \in HxR : y_i(f(x_i)+b) \geq \rho_i\} \tag{156}$$

Thus the group that all the classifiers belong, achieve the margin {ρ_i} for given (x_i,y_i). As the data is coming a sequence of groups {Ci} is formulated, so the demanded classifier will lie on their trace, is not constituting the empty group.

The projection is not performed sequentially in every half-plane but parallel on a group which is appointed from index set I_n defining which (x_i,y_i) will be processed at the time point n. starting from an arbitrary u_0=[0,b_0], the following sequence is created:

$$u_{n+1} = u_n + \mu_n \sum_{i \in I_n} w_i P_{c_i}(u_n) - u_n$$

Combining (155) and (156) results in:

$$u_{n+1} = u_n + \mu_n \sum_{i \in I_n} w_i \left(u_n + \frac{\max\{0, \rho_i - y_i(f_n(x_i) + b_n)\}}{K(x_i, x_i) + 1} y_i [K(x_i,.), 1]\right) - u_n$$
$$= u_n + \mu_n \left(u_n \sum_{i \in I_n} w_i + \sum_{i \in I_n} w_i \frac{\max\{0, \rho_i - y_i(f_n(x_i) + b_n)\}}{K(x_i, x_i) + 1} y_i [K(x_i,.), 1]\right) - u_n)$$
$$\overset{1}{=} u_n + \mu_n \sum_{i \in I_n} w_i \frac{\max\{0, \rho_i - y_i(f_n(x_i) + b_n)\}}{K(x_i, x_i) + 1} y_i [K(x,.), 1]$$
$$\Rightarrow [f_{n+1}(.), b_{n+1}] = [f_n(.), b_n] + \mu_n \sum_{i \in I_n} w_i \frac{\max\{0, \rho_i - y_i(f_n(x_i) + b_n)\}}{K(x_i, x_i) + 1} y_i [K(x,.), 1] \quad (157)$$

with the coefficient $\mu_n \in [0, 2M_n]$ where:

$$M_n = \begin{cases} \dfrac{\sum_{j \in I_n} w_j \left\| P_{c_j}^{(n)}(u_n) - u_n \right\|}{\left\| \sum_{j \in I_n} w_j P_{c_j}^{(n)}(u_n) - u_n \right\|^2} & \text{, when } u_n \notin \bigcap_{j \in I_n} P_{c_j}^{(n)} \\ 1, otherwise & \end{cases} \quad (158)$$

If we assume that every timing moment we process only the current I_n=n. With this simplification we will get a closed expression for the classifier at the time moment n in function with the samples which have been received so far.

$$[f(.),b_0]=[0,b_0]$$

$$[f_1(.),b_0]=[0,b]+\mu_0\frac{\max\{0,\rho_0-y_0b_0\}}{K(x_0,x_0)+1}y_0[K(x_0,.),1]$$

$$=[a_0K(x_0,.),b_0+a_0],\ \ a_0=\mu_0\frac{y_0\max\{0,\rho_0-y_0b_0\}}{K(x_0,x_0)+1}$$

$$[f_2(.),b_2]=[f_1(.),b_1]+\mu_1\frac{\max\{0,\rho_1-y_1(f_1(x_1)+b_1)\}}{K(x_0,x_0)+1}y_1[K(x_1,x_1)+1]$$

$$=[a_0K(x_0,.)+a_1K(x_1,.),b_0+a_0+a_1],\quad a_1=\mu_1\frac{y_1\max\{0,\rho_1-y_1(f_1(x_1)+b_1)\}}{K(x_{1,}x_1)+1}$$

$$\vdots$$

$$[f_n(.),b_n]=[\sum_{i=0}^{n-1}a_iK(x_i,.),b_0+\sum_{i=0}^{n-1}a_i]$$

$$a_i=\mu_i\frac{y_i\max\{0,\rho_1-y_1(f_i(x_i)+b_i)\}}{K(x_i,x_i)+1}$$

$$=\mu_i\frac{y_i\max\{0,\rho_i-y_i(\sum_{j=0}^{i-1}a_jK(x_j,x_i)+b_0+\sum_{j=0}^{i=1}a_j)\}}{K(x_i,x_i)+1},i=1,\ldots \tag{159}$$

When we have the processing of a number of samples at the same time moment, the classifier at timing moment n will be as follows

$$[f_n(.),b_n]=[\sum_{i=0}^{n-1}\sum_{j\in I_n}a_j^{(i)}K(x_j,.),b_0+\sum_{i=0}^{n-1}\sum_{j\in I_n}a_j^{(i)}]$$

$$a_j^{(i)}=\mu_iw_j\frac{y_{j\max\{0,\rho_j^{(i)}-y_i(f_i(x_j)+b_i)\}}}{K(x_i+x_j)+1},i=1\ldots n-1 \tag{160}$$

Adaptive Selection of margin

Setting $C=\{u \in HxR : y(f(x)+b \geq r\}$ as the group with all the classifiers which fall into the margin $r \geq 0$ for some pairs (x,y). For $\gamma \geq 1$ and for $\rho=\gamma r$ we have the group $C=\{u \in HxR : y(f(x)+b) \geq \rho\} \subset C$. If we project on C will get a C classifier,

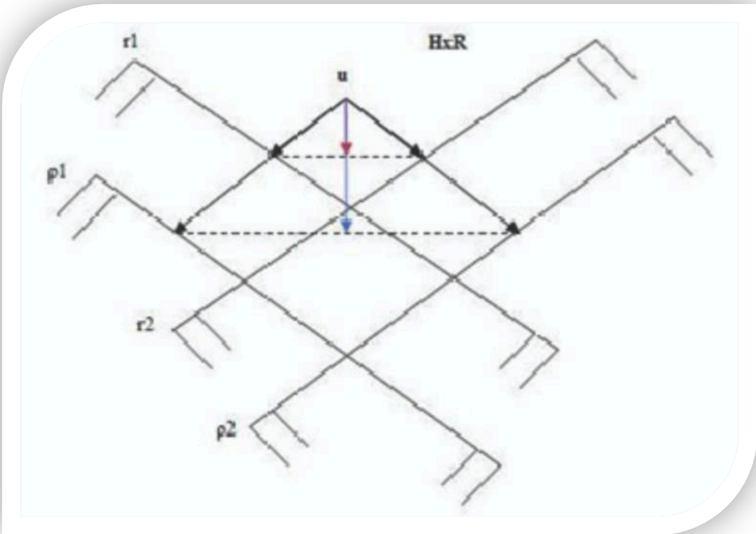


Figure 8.3: Projector (u)

moving though deeper into this. On the above figure we observe that the projection $\sum_{j=1}^{2} w_j P_{c_j}(u)$ (blue colour) brings us closer at the cut of groups C_{r1} and C_{r2}

than the projection $\sum_{j=1}^{2} w_j P_{c_j}(u)$ (red colour).

The parameters ρ_n which define the half-planes that is projected on, constitute multiples of r_n. The basic idea of selection has as follows. If the current estimation u_n hit on the margin r_n, then and the next estimation u_{n+1} its very probable to do the same thing, so we can increase ρ_n to a slightly bigger value ρ_{n+1}. On the contrary, if the current estimation does not hit on r_n, then the ρ_n decreases to a smaller value ρ_{n+1} so as to the next estimation to have bigger chance to hit on the margin $r_{n+1}=r_n$.

The changes of ρ_n are defined by the linear parametric model of $v(\theta-\gamma r)+\gamma r$ and θ where $\theta \in R$ and v is an adequately positive minimized inclination. According to this any possible increase of θ is followed by an increase of ρ and vice-versa.

8.3 Numeric Example

We assume two classes, each one of them constitutes from data which come from the mix of two 2-D Gaussian allocations with equal weights. For the first class we have

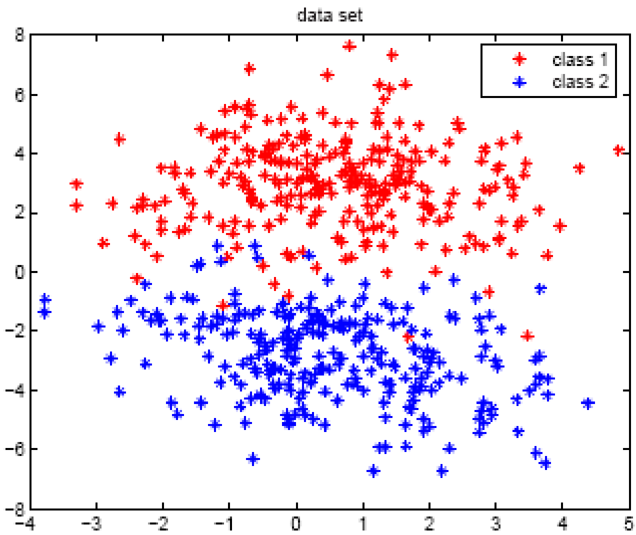


Figure 8.4: Gaussian Classes

The Gaussian mean values and the co-dispersion matrices:

$$\mu_{11} = [0 \quad 3]^2 \qquad \mu_{12} = [1 \quad 3]^t$$

$$\sum_{11} \begin{bmatrix} 5 & 2.5 \\ 2.5 & 5 \end{bmatrix} \qquad \sum_{12} \begin{bmatrix} 5 & -1.5 \\ -1.5 & 5 \end{bmatrix}$$

And for the second:

$$\mu_{21} = \begin{bmatrix} 0 & -3 \end{bmatrix}^t \qquad \mu_{22} = \begin{bmatrix} 1 & -3 \end{bmatrix}^t$$

$$\sum_{21} \begin{bmatrix} 5 & -2.5 \\ -2.5 & 5 \end{bmatrix} \qquad \sum_{22} \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$$

As kernel it was used the $K(x, y) = e^{\frac{(x-y)^t (x-y)}{\sigma^2}}$

Suppose a group of 400 samples which is used and for training and for control as well, under the meaning that the classifier which results at the time moment n is being controlled on the samples which have been received until then.

The curves which come up correspond to the accumulated classification errors. Thus, the errors which have been measured for the current classifier at the time moment n and the errors that have occurred at the time moment n-1.

For NORMA and Perceptron the rate of learning is $\eta = 1/\sqrt{n}$. The APSM4 constitutes a parallel implementation of APSM with the index offset to be In={n,n+1,n+2,n+3}, thus it is processing 4 samples simultaneously.

Moreover, the relaxation parameter for the APSM is μ_n=1, while for APSM4 is μ_n= 1.9M_n. In order the curves to be smoothing out we repeat the experiment 100 times and then we get the average values of the results.

We observe that the dispersion σ^2 of the kernel function as well as the inclination ν plays an important role to the performance of the algorithms.

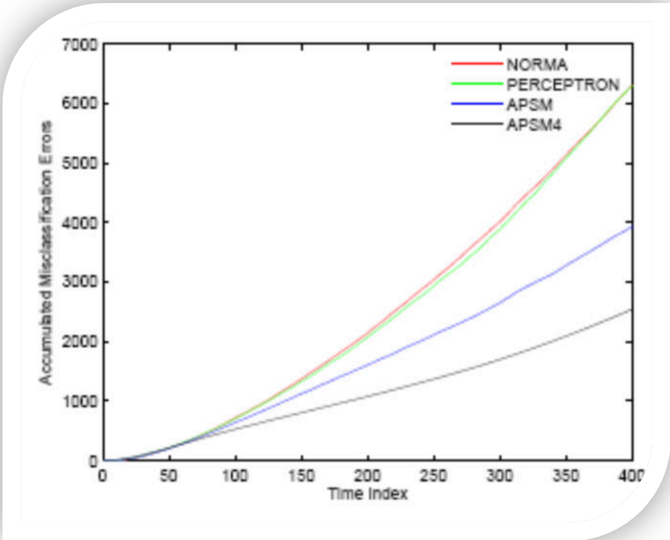


Figure 8.5: σ^2=0.1, ν_{NORMA}=0.01, γ=3, $\delta_\theta = 10^{-2}$, ν_{APSM}=0.1

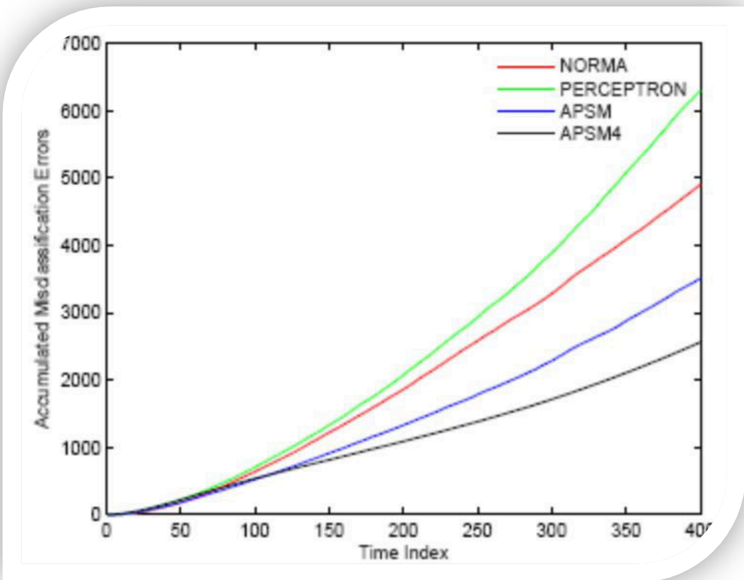


Figure 8.6: σ^2=0.1, ν_{NORMA}=0.5, γ=1, $\delta_\theta = 10^{-1}$, ν_{APSM}=10^{-3}

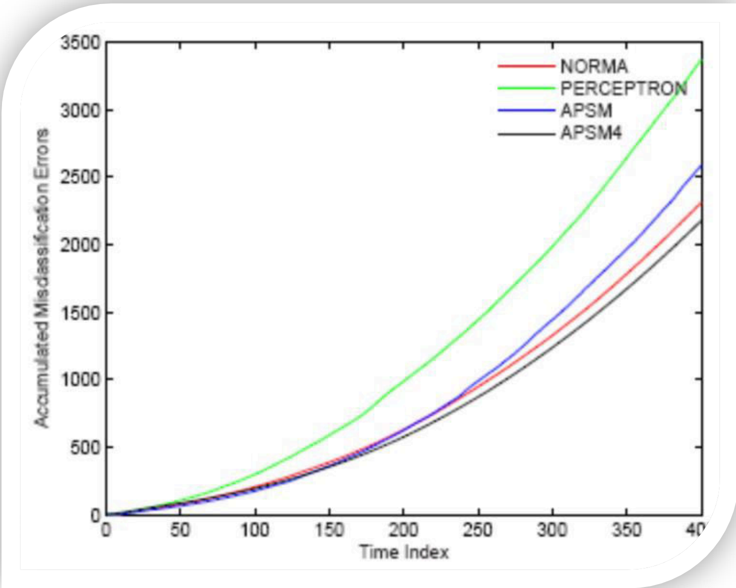


Figure 8.7: σ^2=0.8, ν_{NORMA}=0.5, γ=1, $\delta_\theta = 10^{-1}$, ν_{APSM}=10^{-3}

We separate the group of sample in two parts. We use 400 in total samples for the training and every 25 samples we examine the classifier, which has arisen till then, on 100 different samples.

Next the curves are presented which come up from the NORMA for learning rate $\eta = 1/\sqrt{n}$ and different values of inclination ν.

We observe for ν=0.5 we get a better algorithm behaviour, for the specific always group of samples which came up from the mixture of the Gaussians described above.

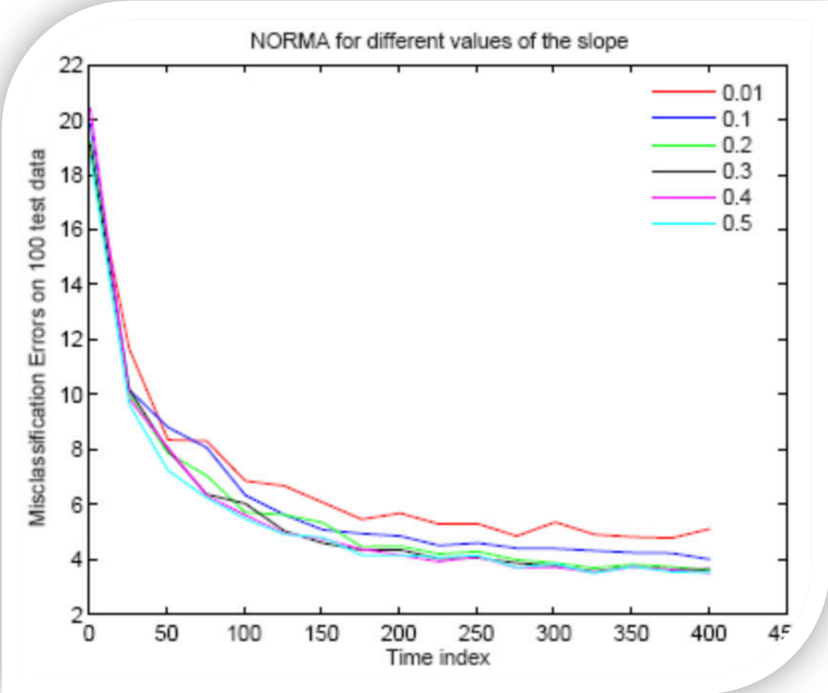


Figure 8.8 : σ^2=0.8

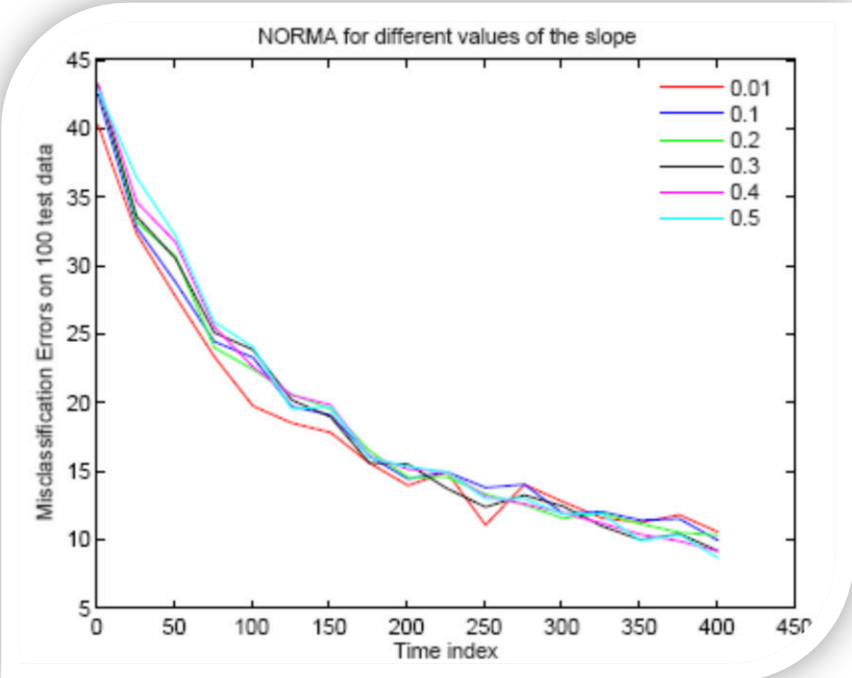


Figure 8.9 : σ^2=0.1

Similarly as with APSM we study the behaviour of the random values of ν inclination and of parameter θ on the parametric model which controls the changes of the margin. We observe that the inclination is small so the margin changes will be also small, the algorithm behaves better.

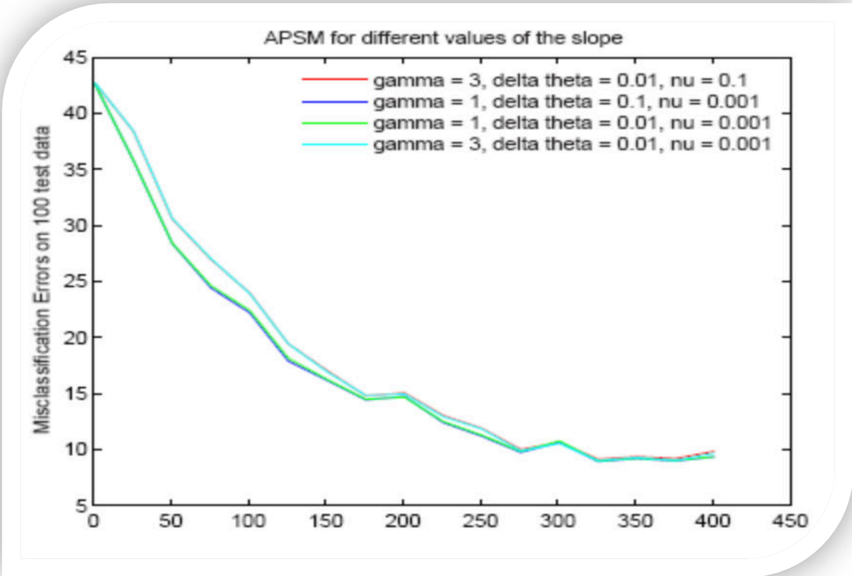


Figure 8.10: σ^2=0.1

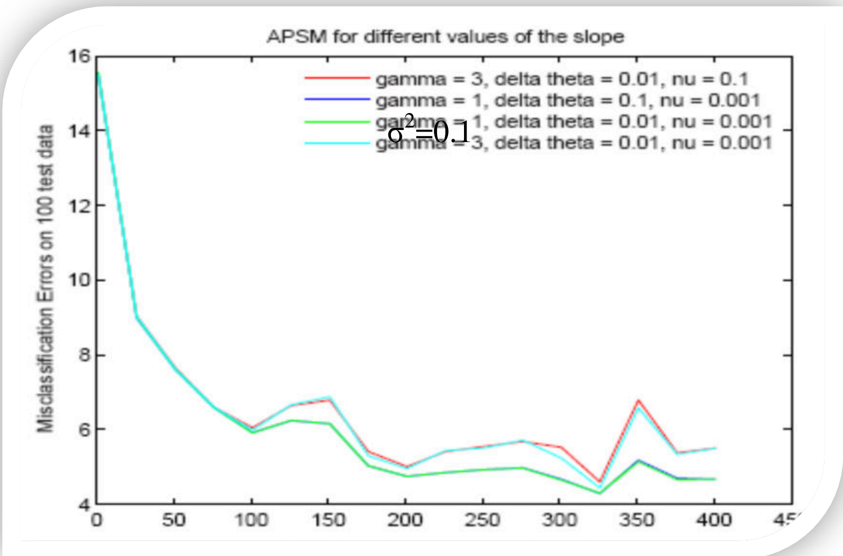


Figure 8.11: σ^2=0.8

Finally, after the first 400 symbols we perform the below change on the Gaussian allocations. For the first class the Gaussian have Gaussian values and matrices of co-dispersion.

$$\mu_{11} = [6 \quad 6]^t \qquad \mu_{12} = [5 \quad 6]^t$$

$$\Sigma_{11}\begin{bmatrix} 6 & 1.5 \\ 1.5 & 6 \end{bmatrix} \qquad \Sigma_{12}\begin{bmatrix} 6 & -0.5 \\ -0.5 & 6 \end{bmatrix}$$

And for the second:

$$\mu_{21} = [-2 \quad 6]^t \qquad \mu_{22} = [-2 \quad 5]^t$$

$$\Sigma_{21}\begin{bmatrix} 6 & -1.5 \\ -1.5 & 6 \end{bmatrix} \qquad \Sigma_{22}\begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$$

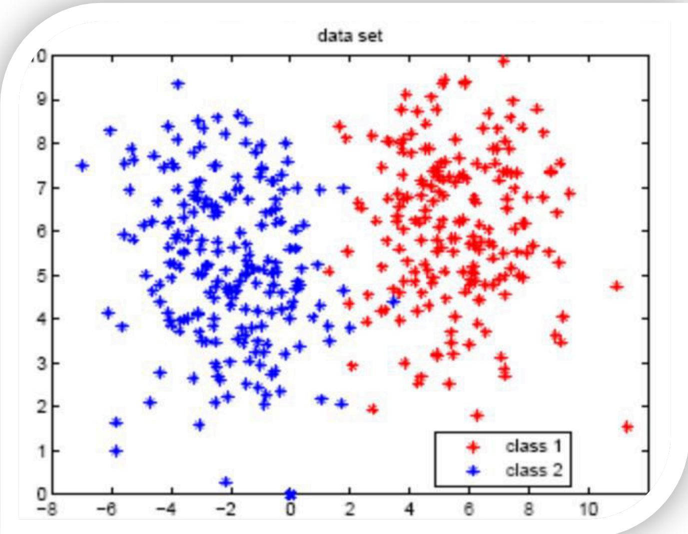


Figure 8.12: class 1 and class 2

So for the figures that follow we have σ^2=0.1, γ=1, $\delta_\theta = 10^{-2}$, ν_{APSM}=10^{-3}, μ_n=1, while the ν_{NORMA} values are changing and M_n. We observe that when we have the parallel processing of the four samples (APSM4) so as the algorithm to be able to give the best behaviour and to follow successfully the change in sample dispersion, the μ_n must be as close to $2M_n$ as possible.

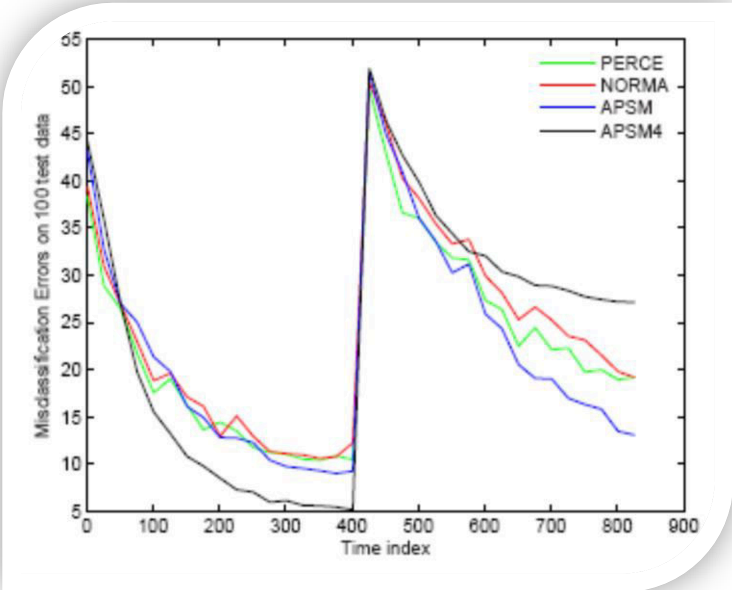


Figure 8.13 : ν_{NORMA}=0.01, 1.9M

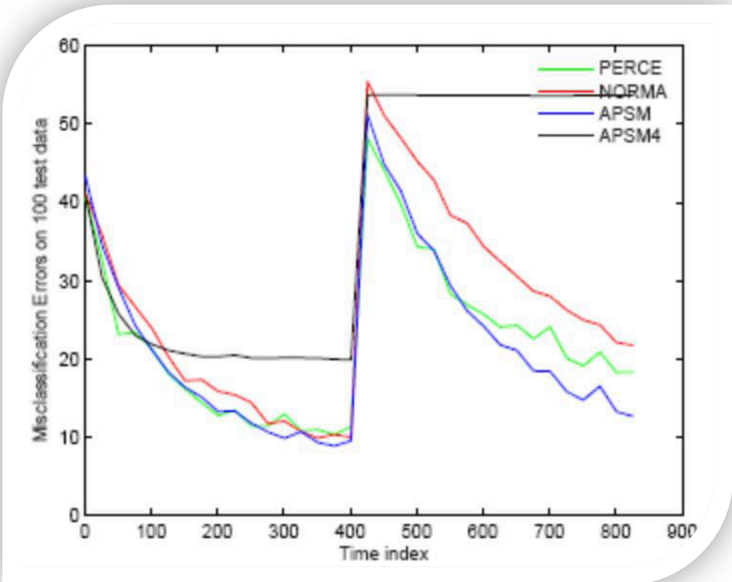


Figure 8.14 : ν_{NORMA}=0.5, 1

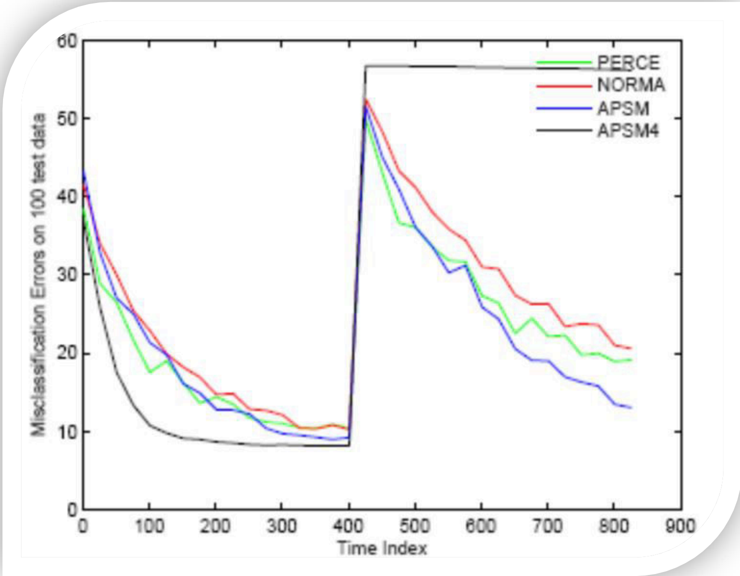


Figure 8.15: ν_{NORMA}=0.1, M_n

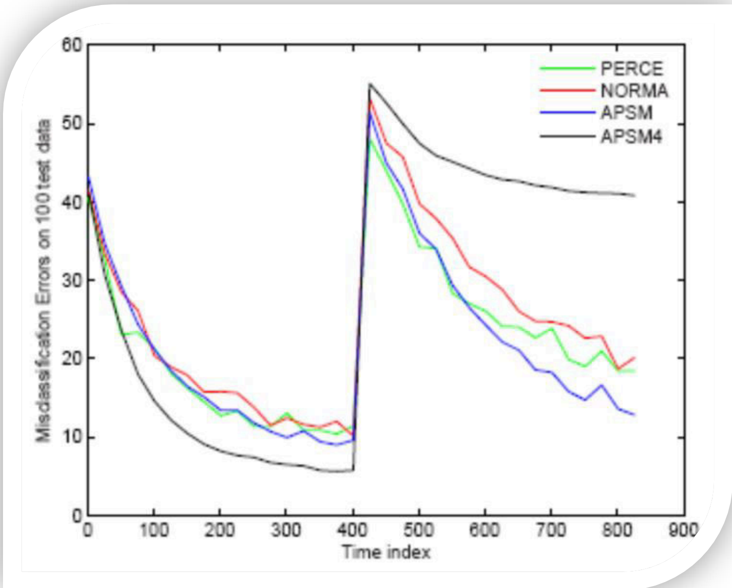


Figure 8.16 : ν_{NORMA}=0.05, $1.5M_n$

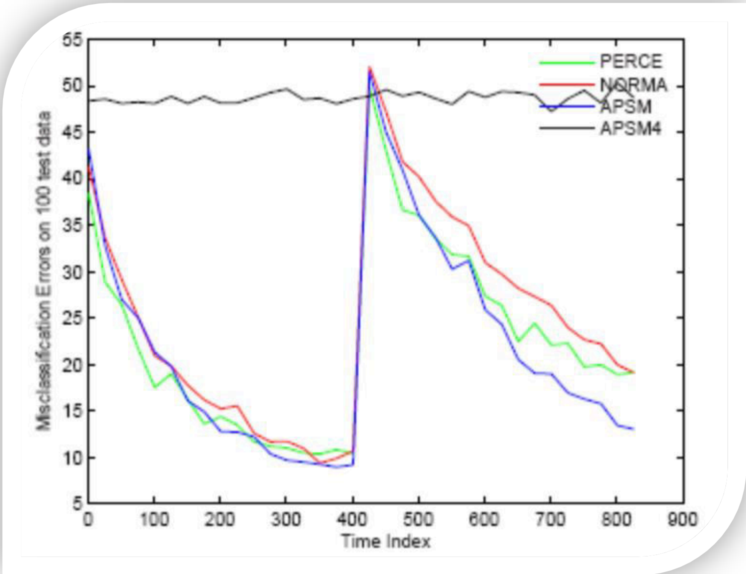


Figure 8.17 : ν_{NORMA}=0.01, 2.3M_n

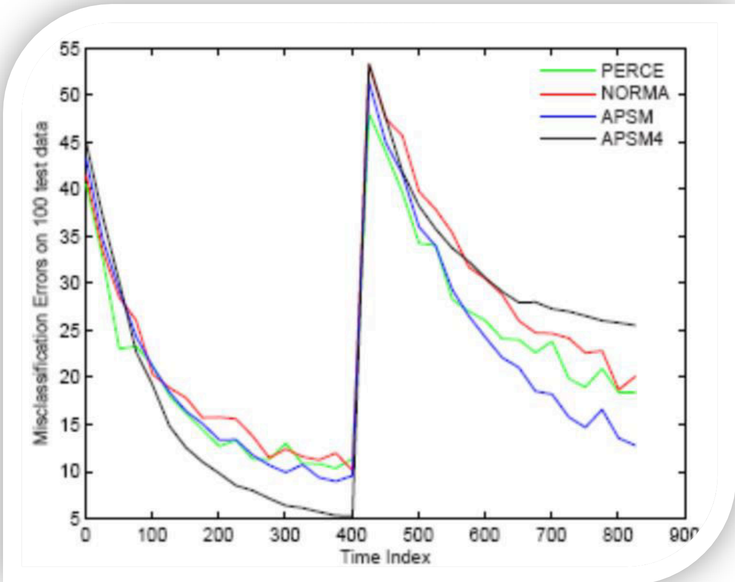


Figure 8.18 : ν_{NORMA}=0.05, 1.95M_n

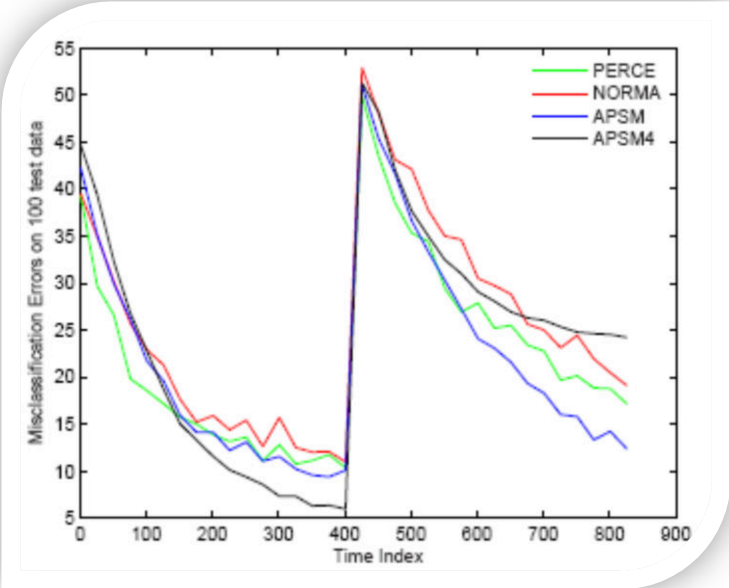


Figure 8.19 : ν_{NORMA}=0.05, 1.8M_n

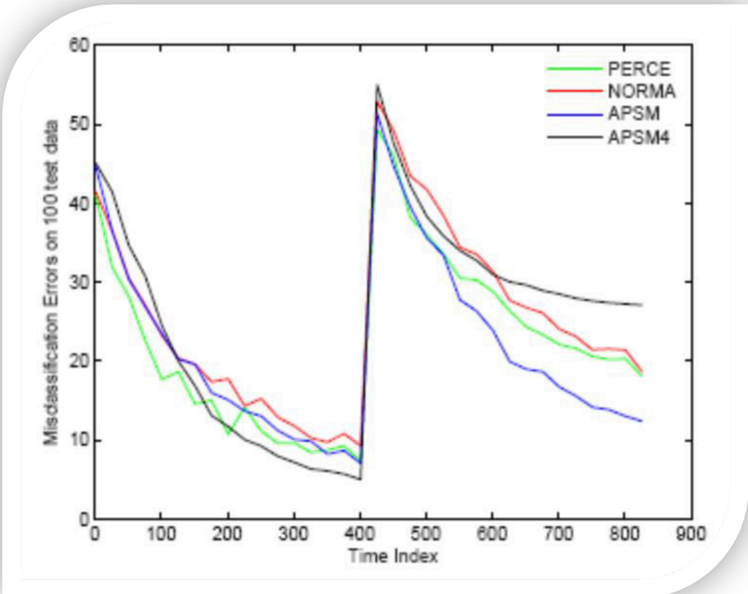


Figure 8.20 : Change in classifiers sequence

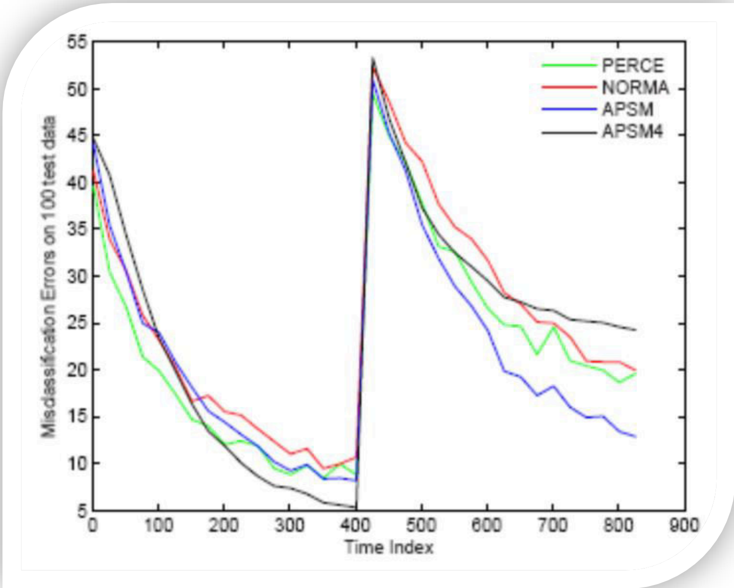


Figure 8.21: More weight to the running sample

Weight change

8.4 Summary

Looking at the figures above we conclude that although it understands the change in samples dispersion the MSE delays to fall when change occurs at time index 400 and onwards. At figure (8.20) we changed the sequence with which the change in dispersion takes place and we got again the same results, so the data is not responsible for that.

At figure (8.21) we changed the weights giving more weight to the running sample. We observe that the result it was slightly better. So this delay owed to the “memory” which is retained at the classifier’s coefficients from the previous dispersion.

CHAPTER 9

CHANNEL EQUALIZATION

9.1 Introduction

In this chapter first it is examined the optimal solution in a channel with noise. Afterwards there will be an implementation from which conclusions will be drawn in the attached figures.

9.2 Equalizer classification

Suppose there is a telecommunication system with BPSK data modulation. The information bits $b_k = \pm 1$ traverse through a linear channel, its exit is deformed from AWGN. The equalizer's target is the recovery of samples which were transmitted, on basis of the notices it takes from the channel's exit.

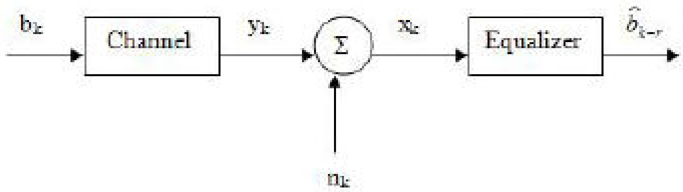


Figure 9.1 : System Equalizer

The samples which are received from the receiver are given from the following relationship:

$$x_k = f(b_k, b_{k-1}, \ldots, b_{k-l+1}) + n_k \qquad (161)$$

with the function f(.) to represent the impact of the channel and the n_k to constitute the noise sequence. The confluence of the channel to the whole deformation it is the intersymbol interference which expands to successively data symbols.

The equalizer constitutes the reverse system and it tries to confute the impact of the channel, supplying decisions $\hat{b}_k$ and the transmitting symbols b_k based on m successively symbols which received, $x_k = [x_k x_{k-1} \ldots x_{k-m+1}]^t$. Usually we do usage of a delay r so as to be secured with the possible no causality nature of the inversed system. Therefore, the functionality of the equalizer at the time moment k has to make a decision based on m recent observations for the symbol which was sent at time moment k-r.

Suppose a digital sequence which is transmitted through a channel with a transfer function:

$$H(z) = \sum_{i=0}^{l-1} h_i z^{-i} \qquad (162)$$

The volume of the possible sequences which we can have at the input of the channel, in the situation where the data symbols are binary, is $n_s = 2^{l+m-1}$. Suppose for example the linear channel: $H(z)=0.5+1.0z^{-1}$. The symbols sequence which is received from the equalizer will be as follows:

$$x_k = 0.5b_k + b_{k-1} + n_k \qquad (163)$$

For equalizer with length m=2 the observation vectors are transformed:

$$x_k = [x_k x_{k-1}]^t \ , \ k=1,\ldots,N \tag{164}$$

We observe that the values of vectors x_k depend from the values of three consecutive data symbols, specifically of b_k , b_{k-1}, b_{k-2}. Omitting the impact of noise, the possible values of the received symbols x_k , as and the possible input sequences at the channel are given in the following matrix:

b_k	b_{k-1}	b_{k-2}	x_k	x_{k-1}
1	1	1	1.5	1.5
1	1	−1	1.5	−0.5
−1	1	1	0.5	1.5
−1	1	−1	0.5	−0.5
1	−1	1	−0.5	0.5
1	−1	−1	−0.5	−1.5
−1	−1	1	−1.5	0.5
−1	−1	−1	−1.5	−1.5

The symbols which are received from the channel without the impact of noise we will call them channel conditions which will break in two classes:

$$\begin{aligned} \Omega^+ &= \{x^+ \,|b(k-r)=1\} \\ \Omega^- &= \{x^- \,|b(k-r)=-1\} \end{aligned} \tag{165}$$

The two classes which were formed contain information for the transfer function of the channel, the symbols statistics and the length of the equalizer. Given that the data symbols have equal possibility of occurrence, each one of the ns channel situations will have possibility of occurrence equals to $p_i = \dfrac{1}{n_s}$.

Due to noise though, the received observation vector x_k constitutes a random procedure with in condition density probability functions cantered on each channel's condition. Therefore the observations will form clouds around the above points.

For big values of the noise power, these clouds will be diffusive around these points, while for small values will be almost gathered around them. In the following figure it can be seen 1000 received symbols from the above channel for SNR=15dB and delay equals to r=1.

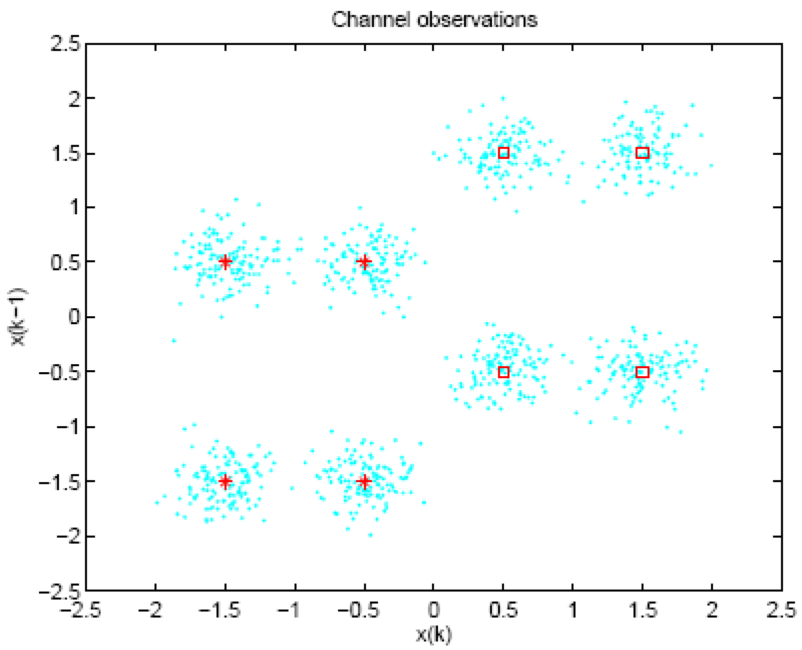


Figure 9.2: Noise clouds

The clouds with centres □ correspond to symbols $b_{k-1}=1$ while the clouds with centres $*$ correspond to symbols $b_{k-1} = -1$. Therefore, the right decision problem deduce to a dual class classification problem.

The conditional density possibility function of vector x_k which belong to the cloud which is formed around the channel condition $x_i^+, i = 1, \ldots, n_s^+$ is:

$$p(x_k \left| x_i^+ \right.) = \frac{1}{(2\pi)^{\frac{m}{2}} \left| \sum_i \right|^{\frac{m}{2}}} e^{-\frac{1}{2}(x_k - x_i^+)^t \left| \sum_i \right|^{-1} (x_k - x_i^+)} \tag{166}$$

But because the noise samples there among them independent so and uncorrelated, the matrix of co-dispersion is diagonal, $\Sigma = \sigma^2 I$ and it will be:

$$p(x_k \left| x_i^+ \right.) = \frac{1}{(2\pi\sigma)^{\frac{m}{2}}} e^{-\frac{\left\| x_k - x_i^+ \right\|^2}{2\sigma^2}} \quad , \; i = 1, \ldots, n_s^+ \tag{167}$$

Depending on the conditional density possibility function of vector x_k which belongs to the cloud which is transformed around the channel condition $x_i^-, i = 1, \ldots, n_s^-$ is:

$$p(x_k \left| x_i^- \right.) = \frac{1}{(2\pi\sigma)^{\frac{m}{2}}} e^{-\frac{\left\| x_k - x_i^- \right\|^2}{2\sigma^2}} \quad , \; i = 1, \ldots, n_s^- \tag{168}$$

Thereupon, conditional density probability function the x_k to belong in class Ω^+ is:

$$p(x_k \left| \Omega^+ \right.) = \sum_{i=1}^{n_s^+} p_i p(x_k \left| x_i^+ \right.), \, p_i = \frac{1}{n_s} \tag{169}$$

and correspondingly on Ω^- is:

$$p(x_k \left| \Omega^- \right.) = \sum_{i=1}^{n_s^-} p_i p(x_k \left| x_i^- \right.), \, p_i = \frac{1}{n_s} \tag{170}$$

According to the Bayes classification rule the classification will be done with base the bigger a posteriori possibility, given the observation x_k:

$$P(\Omega^{+} | x_k) >< P(\Omega^{-} | x_k) \tag{171}$$

Supposing the classes as the data symbol equal, we end up to the following separating level equation:

$$f(x_k) = \sum_{i=1}^{n_s^{+}} \frac{1}{(2\pi\sigma)^{\frac{m}{2}}} e^{-\frac{\|x_k - x_i^{-}\|^2}{2\sigma^2}} - \sum_{j=1}^{n_s^{-}} \frac{1}{(2\pi\sigma)^{\frac{m}{2}}} e^{-\frac{\|x_k - x_j^{-}\|^2}{2\sigma^2}} = 0 \tag{172}$$

And the classification is performed according its sign:

$$f(x_k) > 0, x_k \in \Omega^{+}$$

$$f(x_k) < 0, x_k \in \Omega^{-} \tag{173}$$

9.3 Implementation of NORMA, Perceptron, APSM

The above optimal solution depends on the noise power as from the desirable channel conditions. Moreover, it shows the same structure with the response of a two level RBF network.

$$f(x_k) = \sum_{i=1}^{n} w_i \phi\left(\frac{\|x_k - c_i\|^2}{\rho}\right) \tag{174}$$

Where w_i denotes the weights of a hidden level which in the case of equal probable symbols may be constant, c_i are the centres of the network which constitute the conditions of the channel. The parameter ρ is equal with the double of noise dispersion.

The non-linear function φ(.) constitutes a kernel function. But because the channel function in practise it is not known so as the RBF network to learn the optimal solution it will need an effective training on which it must detect the desirable conditions of the channel and place them on the centres of the network.

On the previous chapter with recursive way projecting on convex sets we end up to the classifier:

$$[f_n(.),b_n]=[\sum_{i=0}^{n} a_i K(x_i,.)\,,\; b_0+\sum_{i=0}^{n-1} a_i] \qquad (175)$$

which we will use for equalization. Here the training does not include the detection of channel conditions as well as the determination of weights a_i and of the offset. We use 800 in total symbols and at every 25 we examine the classifier, which has come up until then, on 100 new symbols.

For the first 400 symbols the channel function is $H(z)=0.5+1.0z^{-1}$ with delay r=1 and then the channel changes to $H(z)=1.0+0.8z^{-1}+0.5z^{-2}$ with delay r=0. The length of the classifier is m=2 and the SNR is 10dB.
For the implementation of the non-linear channel it is used the tanh(.) function after the exit of the channel symbols. For the APSM we have μ_n=1 and for the APSM5 1.9M_n.

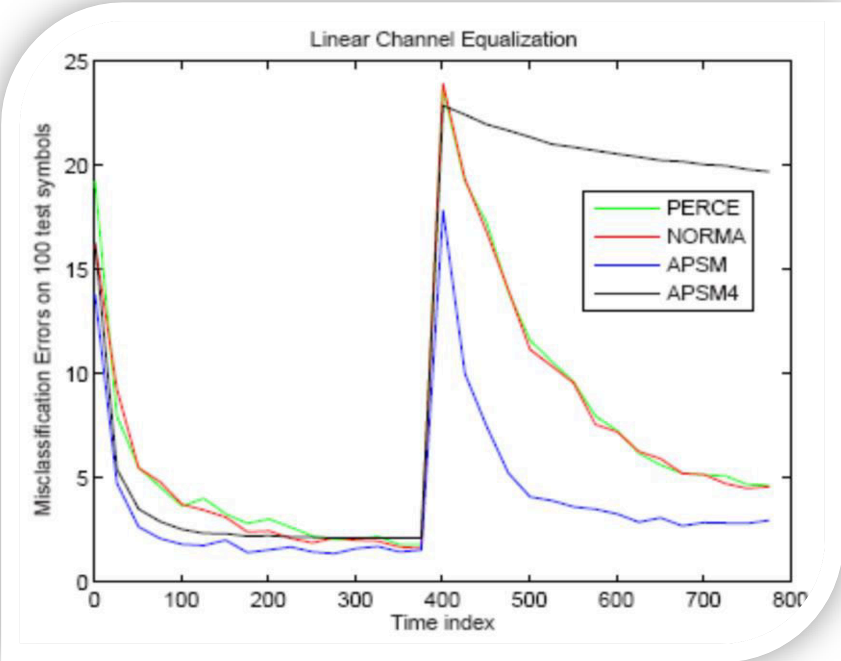


Figure 9.3: ν_{NORMA}=0.01, γ=1, $\delta_\theta = 10^{-2}$, ν_{APSM}=10^{-3}

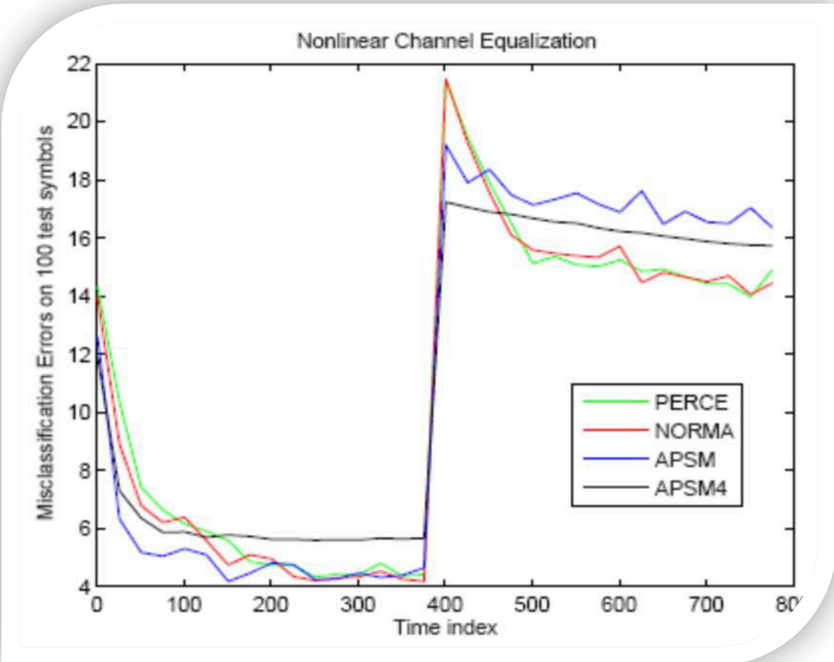


Figure 9.4: ν_{NORMA}=0.01, γ=1, $\delta_\theta = 10^{-2}$, ν_{APSM}=10^{-3}

9.4 Summary

Looking at the above graphs we conclude that the misclassification errors are less in the nonlinear channel equalization case. Although as expected the graph has a similar figure. The highest value is achieved in both cases by NORMA. It is worth noting that in the nonlinear case, after the 400 time index the misclassification errors are steady between 16 and 18 for all aspects.

CHAPTER 10

KERNEL BASED EQUALIZATION CDMA DOWNLINK SYSTEMS

10.1 Introduction

In the last chapter it is examined the equalization in general. In the pages of this chapter we are going further to the kernel based equalization CDMA downlink systems. So, first there will be a description and furthermore an analysis of the given problem. In last pages you can find a numerical example.

10.2 Kernel Equalization

In CDMA systems for each user an assignment of wave signature is allocated which is used to transmit its signal. These signatures they have the orthogonality property, which allow users to simultaneously occupy the same frequency band and time frame [38].

The receiver (mobile terminal) of the user in interest receives the signal, which is transmitted from the base station, and it must be in a position to detect the information, which is designated to him, and be able to isolate it from the rest of the signal, which

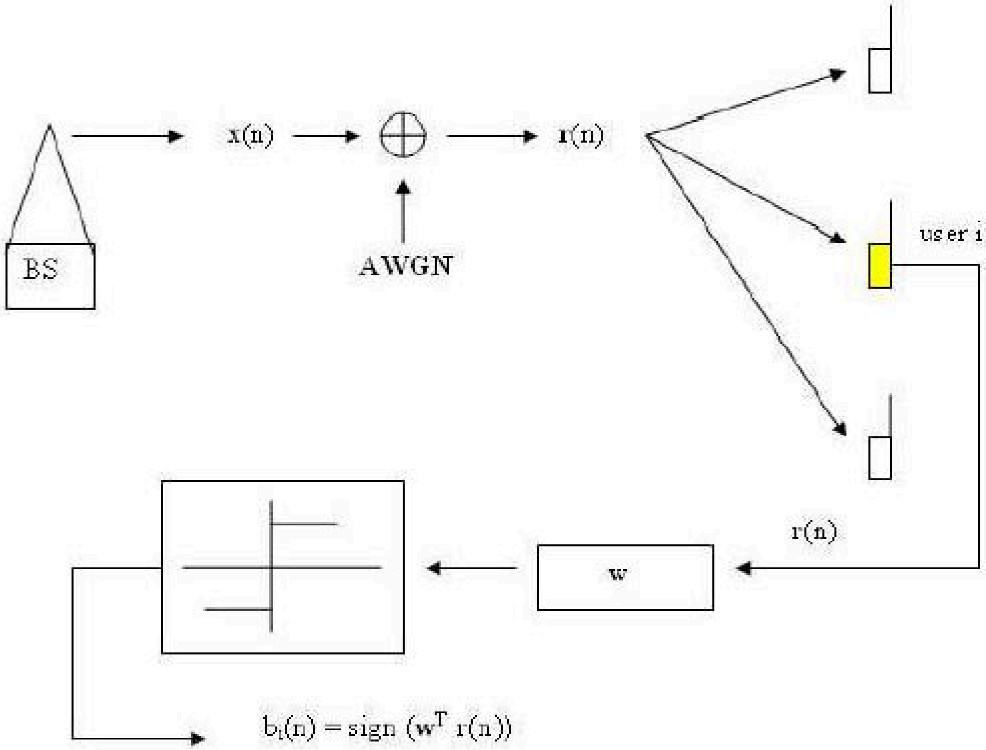


Figure 10.1: Downlink System

represents some kind of interference [74].

System Model

Suppose the downlink CDMA system with P users, the spreading gain L_c, $b_i(n) = \pm 1$ the BPSK symbol of user i=1,…,P at the time moment n=1,…,N and $s_i = [s_1 \ldots s_{L_c} \left| s_k = \pm 1 \right]^t$ the signature vector of user i.

At the time moment n the signal which transmits the base station will be as follows:

$$x(n) = s_1 b_1 + s_2 b_2 + \ldots + s_P b_P(n)$$

$$= [s_1 s_2 \ldots s_P] \begin{bmatrix} b_1(n) \\ b_2(n) \\ \vdots \\ b_P(n) \end{bmatrix}$$

$$= S_{L_c x P} b(n)_{P x 1}$$

For the matrix S which contains the users' signatures applies: $S^t S = I$

Initially we examine the case of an AWGN channel, so the received signal from the receivers at time moment n will be:

$$r(n) = Sb(n) + v(n) \tag{176}$$

where $v(n) \sim N(0, \sigma^2)$. Our target is for every user of interest i to appoint the equalizer w so as the symbol $b_i(n)$ which is appointed for this time moment n to be given from the equation:

$$b_i(n) = \text{sgn}(w^t r(n)) \tag{177}$$

MMSE Linear Detector

We seek the w_{opt} so as:

$$w_{opt} = \arg\min_{w} E\{(b_i(n) - w^t r(n))^2\} \tag{178}$$

We define the cost function:

$$J(w) = E\{(b_i^2(n) - w^t r(n))(b_i(n) - w^t r(n))^t\}$$

$$= E\{b_i^2(n)\} - 2E\{b_i(n) r^t(n)\} w + w^t E\{r(n) r^t(n)\} w \tag{179}$$

and its derivative is:

$$\nabla J(w) = -2E\{r(n) b_i(n)\} + 2E\{r(n) r^t(n)\} \tag{180}$$

Setting the derivative equals to 0 we come up with:

$$E\{r(n) r^t(n)\} w_{opt} = E\{r(n) b_i(n)\} \tag{181}$$

Supposing the data symbols are among them independent as well as the data symbols with the noise is uncorrelated, we have:

$$E\{r(n) r^t(n)\} = E\{(Sb(n) + v(n))(Sb(n) + v(n))^t\}$$

$$= SE\{b(n) b^t(n)\} S^t + 2E\{v(n) b^t(n)\} S^t + E\{v(n) v^t(n)\}$$

$$= SS^t + \sigma^2 I \tag{182}$$

and

$$E\{r(n) b_i(n)\} = E\{(Sb(n) + v(n)) b_i(n)\}$$

$$= SE\{b(n) b_i(n)\} + E\{v(n) b_i(n)\}$$

$$= s_i \tag{183}$$

Combining the above, the optimal linear equalizer for user I will be as follows:

$$w_{opt} = (SS^t + \sigma^2 I)^{-1} s_i \tag{184}$$

10.3 Adaptive Implementation

Suppose we want to find the $\vartheta^{*\in} R^n$ which minimizes the function:

$$\Psi(\vartheta) = E\{g(X,\vartheta)\} \tag{185}$$

If $\Psi(.)$ is convex then by starting from any ϑ_0 we will have the convergence to the minimum of $\Psi(.)$ following the direction of the steepest descent $(-\nabla\Psi)$:

$$\vartheta_{n+1} = \vartheta_n - \mu\nabla\Psi(\vartheta_n) \tag{186}$$

We suppose that we have under our disposal the independent sequence $\{x_1, x_2, \ldots, x_n\}$ with implementations of the random variable X.

In every step of the algorithm we replace the unknown $\nabla\Psi(\vartheta(n) = E\{\nabla g(X,\vartheta_n)\}$ with the noisy version of $\nabla g(x_{n+1}, \vartheta_n)$, so the stochastic gradient descent algorithm will be as follows:

$$\vartheta_{n+1} = \vartheta_n - \mu\nabla g(x_{n+1}, \vartheta_n) \tag{187}$$

The cost function which we supposed to be convex to w because is quadratic with matrix Hessian: $\nabla^2 J(w) = E\{r(n) r^t(n)\} \geq 0$ since the autocorrelation matrix has non negative eigenvalues [9] [61].

So as to get the stochastic gradient descent algorithm it is enough its gradient:

$$g(X,w) = (b_i(n) - w^t r(n))^2 \ , \ X = [b_i(n), r(n)] \qquad (188)$$

Finally the implementation is done with the known LMS:

$$w_n = w_{n-1} - \frac{1}{2}\mu\nabla g(x_n, w_{n-1}) = w_{n-1} - \mu(w_{n-1}^t r(n) - b_i(n))r(n) \qquad (189)$$

Which converge to w_{opt} which we calculated before.

10.4 Numerical Example

We suppose a CDMA system with P=3 users, with signature length Lc=8 and symbol modulation BPSK and AWGN channel with SNR = 10dB. We apply the MMSE equalizer with m=0.01. We send N=500 symbols and we control the equalizer every 25 symbols on T=100 news. The experiment is repeated 100 times and next we receive an average values.

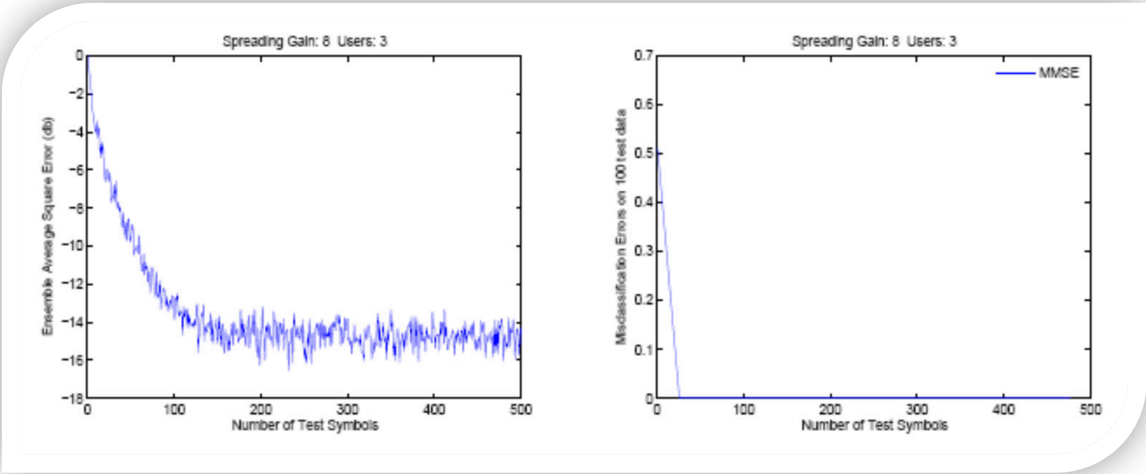


Figure 10.2: CDMA multiuser environment

In environments with many users typically are used large length codes, which can be designed to be of low correlation, but are not orthogonal. For example, if we assume an environment with 30 users on an AWGN channel with SNR=10dB.

On the figures below with the red colour it can be seen the BPSK symbols which are sent from the user of interest and with the blue colour the symbols that it receives, since it makes usage of its own signature. We observe that the bigger is the length of

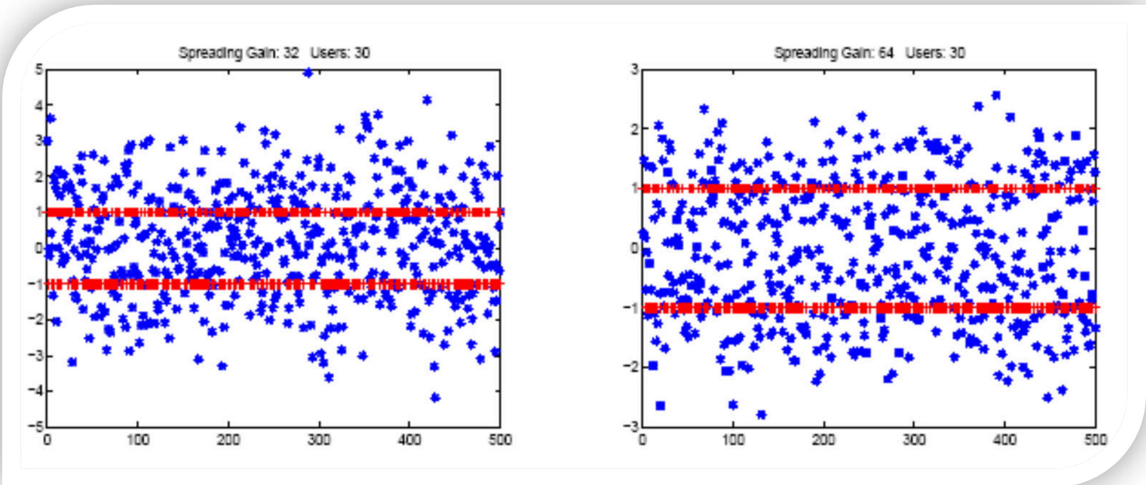


Figure 10.3: BPSK symbols sent

the signature it is starting and formed 2 linear separable classes around the BPSK symbols.

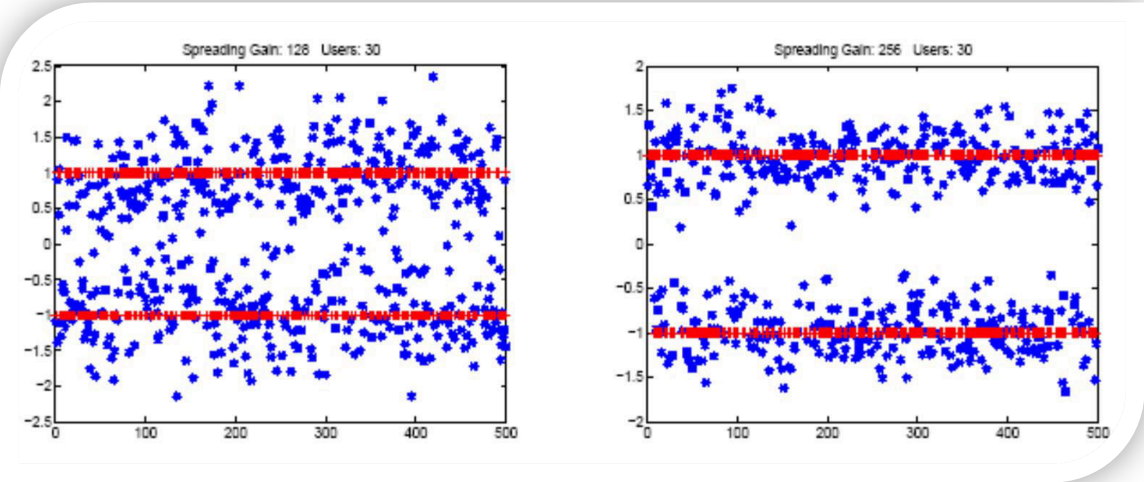


Figure 10.4: Spreading Gain over users

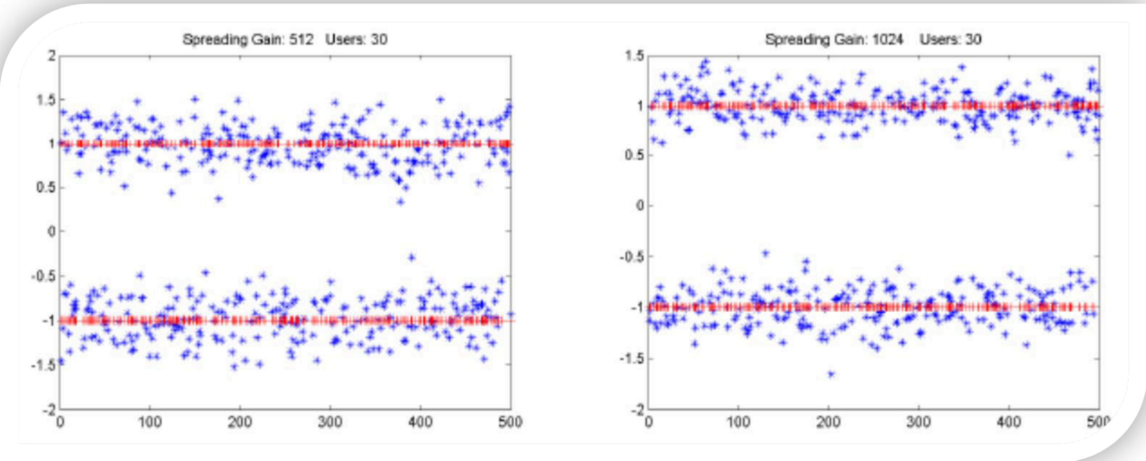


Figure 10.5: Spreading Gain over users

The increase in signature length has as a result the increase of the length of the received symbol r(n) at the receiver. At LMS the correction $\mu(w_{n-1}^{t}r(n)-b_i(n))r(n)$ Which is applied on w(n) at the n+1 repetition, is proportional of r(n). When the length of observation vectors is big, the algorithm appears gradient noise amplification.

Solution to the problem gives the normalized LMS (NLMS), which is based on the concept of minimal disturbance (Windrow and Lehr 1990), [95].

"in the light of a new input data, the parameters of an adaptive system should only be disturbed in a minimal fashion" [95].

The iterative algorithm now will be as follows:

$$w_n = w_{n-1} - \mu \frac{(w_{n-1}^t r(n) - b_i(n)) r(n)}{\left\| r(n) \right\|^2} \tag{190}$$

On the figure that it follows is presented the performance of the NLMS on a 30-users environment and AWGN channel with SNR=10dB, for various values of the signature length L_c.

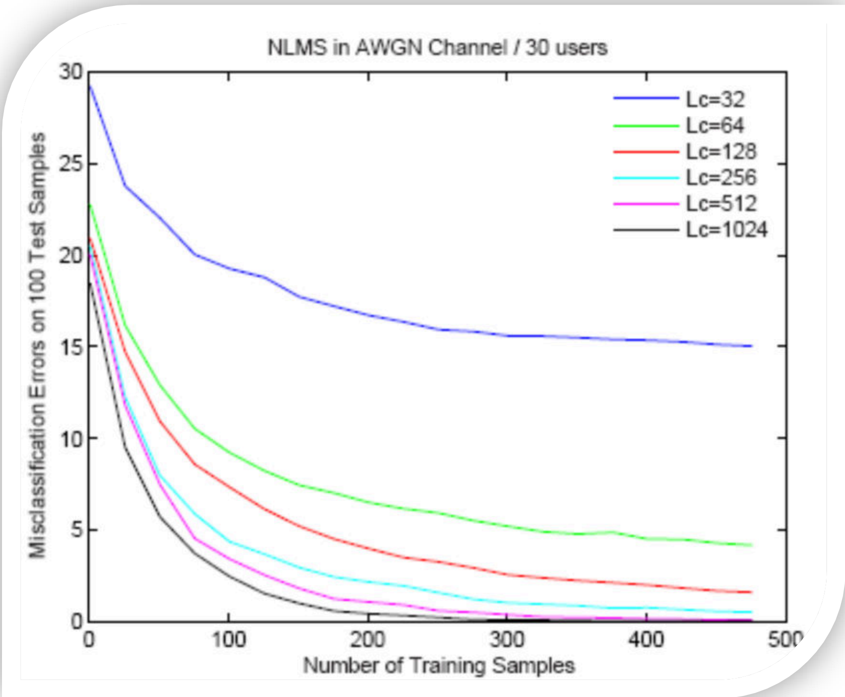


Figure 10.6: Test Samples over Signature Length

10.5 Multipath effect

From the above figure it can be seen the performance of the optimal linear equalizer is dependent from the signature length. The performance is improved for big signature length. In the case, when the channel is not AWGN, the multipath effect has as a result to increase the length of the received symbol Lc, which is the length of the signature, to Lc+L-1, due to the convolution with channel of length L.

This has as a result at the receiver of interest during the sampling to have projection between the nth and (n-1) symbol.

The system model will be now as follows:

$$r(n) = H\begin{bmatrix} Sb(n) \\ Sb(n-1) \end{bmatrix} + v(n), \tag{191}$$

where $b(n) = [b_1(n)b_2(n)\ldots b_p(n)]^t$ the vector with the symbols of P users at the time moment n, $S_{L_c xP}$ the matrix with the users signatures and $H_{L_c x 2L_c}$ the matrix Toeplitz of the channel:

$$H = \begin{bmatrix} h_0 & h_1 & \ldots & h_{L-1} & 0 & \ldots & 0 \\ 0 & h_0 & h_1 & \ldots & h_{L-1} & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & h_0 & h_1 & \ldots & h_{L-1} \end{bmatrix} \tag{192}$$

Next we use the following two channels:

$$h_i = \left[\frac{\sin(a_i)}{\sqrt{2}} \cos(a_i) \frac{\sin(a_i)}{\sqrt{2}} \right] i = 1,2 \tag{193}$$

with $a_i = 29.5^o$ and $a_2 = -35^o$.

The advantage of Norma, Perceptron, and APSM is that it is not required the knowledge of the channel length because the calculations of the inner-products is done between the received symbols ok kernel functions and not between the received symbol and the coefficients of the equalizer as done in NLMS.

Next we examine the algorithms in an environment of 30 users with signature length 64 and SNR = 10dB. The kernels that we use are linear, while the nonlinearity on the channel is achieved through a polynomial $x + 0.2x^2 - 0.1x^3$. The margin is constant. For the linear channel it has a value of 30 while for the non-linear a value of 10. We observe that the APSM and in the two occasions has a better performance.

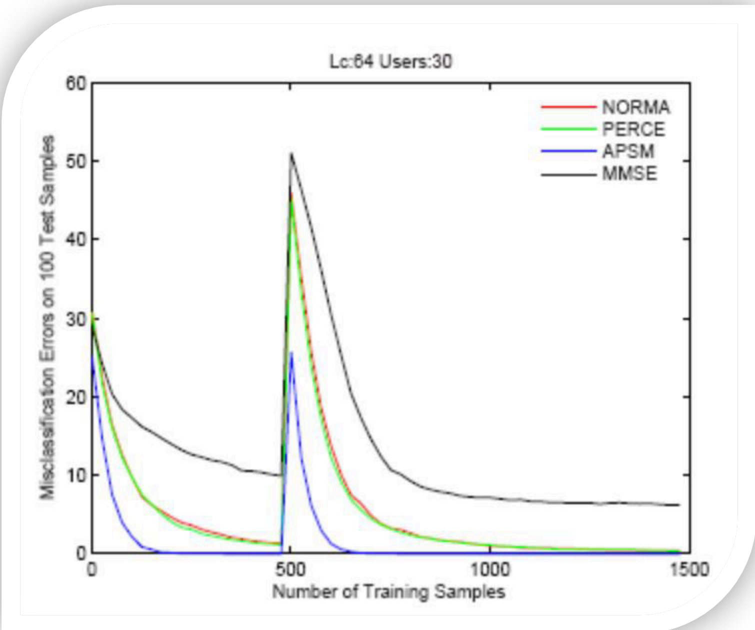


Figure 10.7: Linear Channel

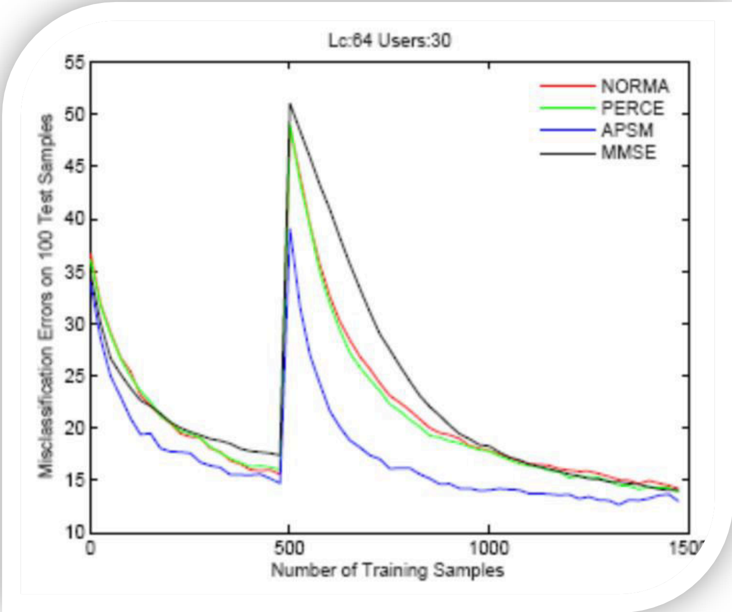


Figure 10.8: Non-Linear Channel

10.6 Summary

Looking at the graphs above we can drawn some conclusions. First of all the figures have in both cases a very similar shape. The values in all aspects have the same highest value, in the linear compared to the nonlinear case. The peak is achieved at the same number of training samples. It is worth mention that the lowest values are accomplished with the linear channel.

CHAPTER 11

CONCLUSIONS

CDMA is the new standard used in the 3rd generation mobile systems. Since multipath is known to degrade any modern communication system, so we were focused on developing channel estimation methods that could overcome this serious problem.

A novel joint blind channel estimation and carrier offset method for code division multiple access CDMA communication systems is proposed. The new method combines SVD analysis with carrier offset parameter.

While existing blind methods sustain from a high computational complexity as it is required the computation of a large SVD not only one but twice, plus it is sensitive to accurate knowledge of the noise subspace rank. The proposed method overcomes both problems by computing the SVD only once. Extensive simulations demonstrate the robustness of the proposed scheme and its performance is comparable to other existing SVD techniques with significant lower computational cost because it does not require knowledge of the rank of the noise sub-space.

We mainly worked with blind methodology that requires minimal a-priori knowledge and adaptive techniques that are characterized by low computational complexity. Our major contribution in this area consists in developing the power method so as to include for the first time the carrier offset combined with estimation of the channel. The proposed method is as much as 70% more efficient than any other method proposed in the literature.

Also, the Kernel based estimation which is for the first time implemented in the literature for CDMA systems. Therefore in this thesis it was managed to introduce for a first in literature a kernel based equalization for CDMA systems.

In addition, we show its superiority in comparison with the other classification models and proving that with the model analysis and the simulation results.

In general, this thesis proposed an advanced adaptive algorithm which outperforms all other existing algorithms. Moreover, it also incorporated the carrier offset estimation which is an inventible problem all time classic problem. Our method outperforms all other methods because is characterized by the lower computational complexity and the inclusion of the carrier offset for the first time in literature.

Furthermore, this thesis presented and modelled again for the first in literature how SVM can be used for CDMA system equalization.

FUTURE WORK

Apart from CDMA, there is another communication technology named OFDM which has been proposed as a standard for a number of modern digital services.

These two digital technologies differ, so they also differ in the way the channel vector, carrier offset and equalization can be done as we demonstrated for CDMA.

As no other method, in literature match ours for CDMA, the same case would be for OFDM. A novel and state of the art method for OFDM can also be modelled in the same fashion we showed for CDMA.

Up to day there was no blind channel method with carrier frequency offset estimation and equalization for CDMA. The same case again is for OFDM.

It would be of great interested to implement the same novel methods for OFDM for the first time in literature.

Moving further, another method to be considered for implementing the novel models is the Multi Carrier-CDMA which combines the advantages of both CDMA and OFMD.

All the above methods can also be expanded for MIMO structures MIMO CDMA or MIMO OFDM and finally ending to MIMO MC-CDMA.

All the above improvements modelled and suggested in this Thesis will give to the existing LTE (Long Term Evolution) networks even better performance. As LTE uses digital networks which include CDMA and OFDM technologies.

REFERENCES

[1] Hiu Lui and Guanghau Xi: “A Subspace Method for Signature Waveform Estimation in Synchronous CDMA Systems” *IEEE Trans. on Comms.,* vol.44, no.10, October 1996.

[2] Kemin Li and Hui Liu: “Joint Channel and Carrier Offset Estimation in CDMA Communications” *IEEE Trans. on Signal Processing,* vol.47, no.7, July 1999.

[3] Xenofon G. Dokopoulos and George V. Moustakides: “Blind Channel Estimation for Downlink CDMA Systems” *IEEE Intern. Conf . on Comms*. ICC’2003, Alaska, USA, May 2003.

[4] P. Loubaton and E. Moulinas, “On blind multiuser forward link channel estmation by the subspace method: Identifiabbility results,,” *IEEE Trans. on Signal Processing,* vol.48, pp. 2366-2376, Aug. 2000.

[5] Z. Xu and M. K. Tsatsanis, “Blind adaptive algorithms for minimu variance CDMA receivers,” *IEEE Trans. On Communications,* vol. 49, pp. 180-194, Jan 2001.

[6] Samir Attalah and Hongyi Fu: “*Joint Channel and Carrier Offset Estimation in a Multiuser CDMA System “IEEE Communications Letters,* vol.6, no.10*, Page(s): 428 – 430,* October 2002.

[7] Fu, H and Attallah, S “Fast blind channel and carrier offset estimation in multiuser CDMA,” *IEE Proceedings Communications,* vol. 151, issue 1, pp. 32-38, Feb. 2004.

[8] Samir Attallah; Thiagarajan, L.B.; Hongyi Fu; Ying-Chang Liang “Joint Channel and Carrier Offset Estimation for Synchronous Uplink CDMA Systems”
Vehicular Technology, IEEE Transactions on Volume 56, Issue 5, Sept. 2007
Page(s):2769 – 2774.

[9] X. G. Doukopoulos, and G. V. Moustakides, “Adaptive power techniques for blind channel estimation in CDMA systems”, IEEE Trans. Signal Processing, no. 3, pp. 1110-1120, Mar. 2005.

[10] Youngyi Yang Henry Stark. *Vector Space Projections*. John Wiley & Sons INC.

[11] Alexander J. Smola Jyrki Kivenen and Robert C. Williamson. Online learnig with kernels. In *TRANSACTIONS ON SIGNAL PROCESSING*, volume 52, pages 2165{2176. IEEE, August 2004.

[12] R. Badeau, G. Richard, and B. David, ”Sliding Window Adaptive SVD Algorithms,”
IEEE Transactions on Signal Processing, vol. 52, no. 1, pp. 1-10, Jan. 2004.
[13] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM Philadelphia, 2000.
[14] A. Benveniste, M. M´etivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*, Berlin: Springer, 1990.
[15] J.A.C. Bingham, “Multicarrier Modulation for Data Transmission: An Idea whose Time Has Come,” *IEEE Communications Magazine*, vol. 28, pp. 5-14, May 1990.
[16] H. B¨olcskei, R.W. Heath, and A.J. Paulraj, “Blind channel identification and equalization in OFDM-based multiantenna systems”, *IEEE Transactions on Signal Processing*, vol. 50, pp. 96-109, Jan. 2002.
[17] S. Buzzi, M. Lops, and H.V. Poor, “Blind Adaptive Joint Multiuser Detection and Equalization in Dispersive Differentially Encoded CDMA Channels”, *IEEE Transactions on Signal Processing*, vol. 51, pp. 1880-1893, July 2003.
[18] J.K. Cavers, *Mobile Channel Characteristics*, Kluwer Avademic Publishers, 2000.
[19] B. Champagne and Q.G. Liu, “Plane Rotation-Based EVD Updating Schemes for Efficient Subspace Tracking,”
[20] B. Champagne, “On the Asymptotic Convergence and Numerical Stability of the Proteus EVD Trackers,” *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 242-246, Jan. 2000.
[21] T. Chen, Y. Hua, and W.Y. Yan. “Global Convergence of Oja’s Subspace Algorithm for Principal Component Extraction,” *IEEE Transactions on Neural Networks*, vol. 9, no. 1, pp. 58-67, Jan. 1998.
[22] J. Choi, “A Moment-Based Blind Method for Identification of FIR Downlink DSCDMA Channels”, *IEEE Transactions on Signal Processing*, vol. 48, pp. 966-981, Apr. 2000.

[23] T. Chonavel, B. Champagne, and C. Riou, “Fast Adaptive Eigenvalue Decomposition:
A Maximum Likelihood Approach,” *Signal Processing*, vol. 83, pp. 307-324,
2003.
[24] P. Comon and G.H. Golub, “Tracking a Few Extreme Singular Values and Vectors
in Signal Processing,” Proceedings of IEEE, vol. 78, no. 8, Aug. 1990.
[25] C.E. Davila, “Efficient, High Performance, Subspace Tracking for Time-Domain Data,” *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3307-3315,
Dec. 2000.
[26] J. Dehaene, M. Moonen, and J. Vandewalle, “An Improved Stochastic Gradient Algorithm for Principal Component Analysis and Subspace Tracking,” *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2582-2586, Oct. 1997.
[27] J.P. Delmas, “Performances Analysis of a Givens Parametrized Adaptive Eigenspace Algorithm,” *Signal Processing*, vol. 68, pp. 87-105, 1998.
[28] J.P. Delmas and J.F. Cardoso, “Performance Analysis of an Adaptive Algorithm for Tracking Dominant Subspaces,” *IEEE Transactions on Signal Processing*, vol.
46, no. 11, pp. 3045-3057, Nov. 1998.
[29] R.D. DeGroat, “Noniterative Subspace Tracking,” *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 571-577, Mar. 1992.
[30] R.D. DeGroat, E.M. Dowling, H. Ye, and D.A. Linebarger, “Spherical Subspace Tracking for Efficient, High Performance Adaptive Signal Processing Applications,” *Signal Processing*, vol. 50, pp. 101-121, 1996.

[31] L. Deneire, P. Vandenameele, L. Perre, B.Gyselinckx, and M. Engels, “A Low-Complexity ML Channel Estimator for OFDM,” *IEEE Transactions on Communications*,
vol. 51, pp. 135-140, Feb. 2003.
[32] E.M. Dowling, L.P. Ammann, and R.D. DeGroat, “A TQR-Iteration Based Adaptive
SVD for Real Time Angle and Frequency Tracking,” *IEEE Transactions on Signal Processing*, vol. 42, no. 4, pp. 914-926, Apr. 1994.
[33] O. Edfors, M. Sandel, J.J.V. Beek, S.K. Wilson, and P.O. B¨orjesson, “OFDM Channel Estimation by Singular Value Decomposition,” *IEEE Transactions on*

Communications, vol. 46, pp. 931-939, July 1998.
[34] H.H. Fan and X. Li, "Linear Prediction Approach for Joint Blind Equalization and Blind Multiuser Detection in CDMA Systems", *IEEE Transactions on Signal Processing*, vol. 48, pp. 3134-3145, Nov. 2000.
[35] D.Z. Feng, Z. Bao, and X.D. Zhang, "A Bi-Iteration Instrumental Variable Noise-Subspace Tracking Algorithm," *Signal Processing*, vol. 81, pp. 2215-2221, 2001.
[36] Z. Fu and M. Dowling, "Conjugate Gradient Eigenstructure Tracking for Adaptive
Spectral Estimation," *IEEE Transactions on Signal Processing*, vol. 43, no.
5, pp. 1151-1160, May 1993.
[37] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd edn, The John Hopkins
University Press, 1996.
[38] T. Gustaffson, "Instrumental Variable Subspace Tracking Using Projection Approximation,"
IEEE Transactions on Signal Processing, vol. 46, no. 3, pp. 669-
681, Mar. 1998.
[39] T. Gustaffson and C.S. MacInnes, "A Class of Subspace Tracking Algorithms Based on Approximation of the Noise Subspace," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3231-3235, Nov. 2000.
[40] F. Hansen and P.Y. Yalamov, "Computing Symmetric Rank-Revealing Decompositions
via Triangular Factorization," *SIAM Journal on Matrix Analysis &*
[1] *Applications*, vol. 23, no. 2, pp. 443-458, 2001.
[20] B. Champagne, "On the Asymptotic Convergence and Numerical Stability of the Proteus EVD Trackers," *IEEE Transactions on Signal Processing*, vol. 48, no. 1,
pp. 242-246, Jan. 2000.
[21] T. Chen, Y. Hua, and W.Y. Yan. "Global Convergence of Oja's Subspace Algorithm
for Principal Component Extraction," *IEEE Transactions on Neural*
Networks, vol. 9, no. 1, pp. 58-67, Jan. 1998.
[22] J. Choi, "A Moment-Based Blind Method for Identification of FIR Downlink DSCDMA

Channels”, *IEEE Transactions on Signal Processing*, vol. 48, pp. 966-981,
Apr. 2000.

[23] T. Chonavel, B. Champagne, and C. Riou, “Fast Adaptive Eigenvalue Decomposition:
A Maximum Likelihood Approach,” *Signal Processing*, vol. 83, pp. 307-324,
2003.

[24] P. Comon and G.H. Golub, “Tracking a Few Extreme Singular Values and Vectors
in Signal Processing,” Proceedings of IEEE, vol. 78, no. 8, Aug. 1990.

[25] C.E. Davila, “Efficient, High Performance, Subspace Tracking for Time-Domain Data,” *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3307-3315,
Dec. 2000.

[26] J. Dehaene, M. Moonen, and J. Vandewalle, “An Improved Stochastic Gradient Algorithm for Principal Component Analysis and Subspace Tracking,” *IEEE Transactions on Signal Processing*, vol. 45, no. 10, pp. 2582-2586, Oct. 1997.

[27] J.P. Delmas, “Performances Analysis of a Givens Parametrized Adaptive Eigenspace Algorithm,” *Signal Processing*, vol. 68, pp. 87-105, 1998.

[28] J.P. Delmas and J.F. Cardoso, “Performance Analysis of an Adaptive Algorithm for Tracking Dominant Subspaces,” *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 3045-3057, Nov. 1998.

[29] R.D. DeGroat, “Noniterative Subspace Tracking,” *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 571-577, Mar. 1992.

[30] R.D. DeGroat, E.M. Dowling, H. Ye, and D.A. Linebarger, “Spherical Subspace Tracking for Efficient, High Performance Adaptive Signal Processing Applications,” *Signal Processing*, vol. 50, pp. 101-121, 1996.

[31] L. Deneire, P. Vandenameele, L. Perre, B.Gyselinckx, and M. Engels, “A Low-Complexity ML Channel Estimator for OFDM,” *IEEE Transactions on Communications*,
vol. 51, pp. 135-140, Feb. 2003.

[32] E.M. Dowling, L.P. Ammann, and R.D. DeGroat, “A TQR-Iteration Based Adaptive
SVD for Real Time Angle and Frequency Tracking,” *IEEE Transactions on Signal Processing*, vol. 42, no. 4, pp. 914-926, Apr. 1994.

[33] O. Edfors, M. Sandel, J.J.V. Beek, S.K. Wilson, and P.O. B¨orjesson, "OFDM Channel Estimation by Singular Value Decomposition," *IEEE Transactions on Communications*, vol. 46, pp. 931-939, July 1998.
[34] H.H. Fan and X. Li, "Linear Prediction Approach for Joint Blind Equalization and Blind Multiuser Detection in CDMA Systems", *IEEE Transactions on Signal Processing*, vol. 48, pp. 3134-3145, Nov. 2000.
[35] D.Z. Feng, Z. Bao, and X.D. Zhang, "A Bi-Iteration Instrumental Variable Noise-Subspace Tracking Algorithm," *Signal Processing*, vol. 81, pp. 2215-2221, 2001.
[36] Z. Fu and M. Dowling, "Conjugate Gradient Eigenstructure Tracking for Adaptive Spectral Estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1151-1160, May 1993.
[37] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd edn, The John Hopkins University Press, 1996.
[38] T. Gustaffson, "Instrumental Variable Subspace Tracking Using Projection Approximation," *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 669-681, Mar. 1998.
[39] T. Gustaffson and C.S. MacInnes, "A Class of Subspace Tracking Algorithms Based on Approximation of the Noise Subspace," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 3231-3235, Nov. 2000.
[40] F. Hansen and P.Y. Yalamov, "Computing Symmetric Rank-Revealing Decompositions via Triangular Factorization," *SIAM Journal on Matrix Analysis & Applications*, vol. 23, no. 2, pp. 443-458, 2001.
[41] S. Haykin, *Adaptive Filter Theory*, 4-th Edition, Prentice-Hall, Upper Saddle River, NJ, 2002.
[42] R.W. Heath and G.B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems", *IEEE Transactions on Signal Processing*, vol. 47, pp. 848-856, Mar. 1999.
[43] M. Honig, U. Madhow and S. Verd´u, "Blind Adaptive Multiuser Detection," *IEEE Transactions on Information Theory,* vol. 41, pp. 944-960, July 1995.
[44] S. Hosur, A.H. Tewfik, and D. Boley, "ULV and Generalized ULV Subspace

Tracking Adaptive Algorithms," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1282-1297, May 1998.

[45] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "A New Look at the Power Method for Fast Subspace Tracking," *Digital Signal Processing* 9, pp. 297-314, 1999.

[46] J.Y. Huang, I. Kontoyiannis, S.P. Meyn, "The ODE Method and Spectral Theory of Markov Operators," *Lect. Notes Contr. Inf.*, vol. 280, pp. 205-221, 2002.

[47] W.C. Jakes, *Microwave Mobile Communications*, AT&T, 1974.

[48] S. Jo, J. Choi, and Y.H. Lee, "Modified Leaky LMS Algorithm for Channel Estimation in DS-CDMA Systems", *IEEE Communications Letters*, vol. 6, pp. 202-204, May 2002.

[49] T. Kailath. *Linear Systems*, Prentice Hall, 1980.

[50] I. Karasalo, "Estimating the Covariance Matrix by Signal Subspace Averaging", *IEEE Transactions on ASSP*, vol. 34, no. 1, pp. 8-12, Feb. 1986.

[51] S. Lasaulce, P. Loubaton, and E. Moulines,"A Semi-Blind Channel Estimation Technique Based on Second-Order Blind Method for CDMA Systems", *IEEE Transactions on Signal Processing*, vol. 51, pp. 1894-1904, July 2003.

[52] H. Li, X. Lu, and G.B. Giannakis, "Capon Multiuser Receiver for CDMA Systems
with Space-Time Coding", *IEEE Transactions on Signal Processing*, vol. 50, pp. 1193-1204, May 2002.

[53] H. Li and R. Wang, "Filterbank-Based Blind Code Synchronization for DSCDMA
Systems in Multipath Fading Channels", *IEEE Transactions on Signal Processing*, vol. 51, pp. 160-171, Jan. 2003.

[54] K. Li and H. Liu, "Joint Channel and Carrier Offset Estimation in CDMA Communications",
IEEE Transactions on Signal Processing, vol. 47, pp. 1811-1822,
July 1999.

[55] W.C. Lim, B. Kannan and T.T. Tjhung, "Joint channel estimation and OFDM synchronization in multipath fading," *International Conference on Communications*, ICC'2004, Paris, France, 2004.

[56] C. Li and S. Roy, "Subspace-Based Blind Channel Estimation for OFDM by

Exploiting Virtual Carriers," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 141-150, Jan. 2003.
[57] T.H. Liu and J.M. Mendel, "Cumulant-Based Subspace Tracking," *Signal Processing*, vol. 76, pp. 237-252, 1999.
[58] W. Liu, W.Y. Yan, V. Sreeram, and K.L. Teo, "Global Convergence Analysis for the NIC Flow," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2422-2430, Oct. 2001.
[59] P. Loubaton and E. Moulines, "On Blind Multiuser Forward Link Channel Estimation by the Subspace Method: Identifiability Results," *IEEE Transactions on Signal Processing*, vol. 48, pp. 2366-2376, Aug. 2000.
[60] C.S. MacInnes, "Filtering and Frequency Estimation Using Perturbation Formulas," *IEEE Transactions on Signal Processing*, vol. 44, no. 1, pp. 139-142, Jan. 1996.
[61] G. Mathew, V.U. Reddy, and S. Dasgupta, "Adaptive Estimation of Eigensubspace," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 401-411, Feb. 1995.
[62] G. Mathew and V.U. Reddy, "A Quasi-Newton Adaptive Algorithm for Generalized Symmetric Eigenvalue Problem," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2413-2422, Oct. 1996.
[63] P. Strobach, "Fast Recursive Orthogonal Iteration Subspace Tracking Algorithms and Applications," *Signal Processing*, vol. 59, pp. 73-100, 1997.
[64] P. Strobach, "Bi-Iteration SVD Subspace Tracking Algorithms," *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1222-1240, May 1997.
[65] P. Strobach, "Bi-Iteration Recursive Instrumental Variable Subspace Tracking and Adaptive Filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2708-2725, Oct. 1998.
[66] P. Strobach, "Fast Orthogonal Iteration Adaptive Algorithms for the Generalized Symmetric Eigenproblem," *IEEE Transactions on Signal Processing*, vol. 46, no.

12, pp. 3345-3359, Dec. 1998.
[67] P. Strobach, "Bi-Iteration Multiple Invariance Subspace Tracking and Adaptive ESPRIT," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 442-456, Feb. 2000.
[68] P. Strobach, "Equirotational Stack Parameterization in Subspace Estimation and Tracking," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 712-722, Mar. 2000.
[69] P. Strobach, "Square-Root QR Inverse Iteration for Tracking the Minor Subspace," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 2994-2999, Nov. 2000.
[70] M.K. Tsatsanis and Z. Xu, "Performance Analysis of Minimum Variance CDMA Receivers," *IEEE Transactions on Signal Processing*, vol. 46, pp. 3014-3022, Nov. 1998.
[71] W. Utschick, "Tracking of Signal Subspace Projectors", *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 769-778, Apr. 2002.
[72] A.J. Van der Veen, "A Schur Method for Low-Rank Matrix Approximation," *SIAM Journal on Matrix Analysis & Applications*, vol. 17, no. 1, pp. 139-160, 1996.
[73] S. Verd´u, *Multiuser Detection*, Cambridge University Press, New York, 1998.
[74] X. Wang and H.V. Poor, "Blind Equalization and Multiuser Detection in Dispersive CDMA Channels," *IEEE Transactions on Communications*, vol. 46, pp. 91-103, Jan 1998.
[75] X. Wang and H.V. Poor, "Blind Multiuser Detection: a Subspace Approach" *IEEE Transactions on Information Theory*, vol. 44 , pp. 677-690, Mar. 1998.
[76] X. Wang and H.V. Poor, "Blind Joint Equalization and Multiuser Detection for DS-CDMA in Unknown Correlated Noise", *IEEE Transactions on Circuits ans Systems II*, vol. 46, pp. 886-895, July 1999.
[77] A. Weiss and B. Friedlander, "Synchronous DS-CDMA Downlink with Frequency Selective Fading", *IEEE Transactions on Signal Processing*, vol. 47, pp. 158-167, Jan. 1999.
[78] G. Xu and T. Kailath, "Fast Subspace Decomposition," *IEEE Transactions on

Signal Processing, vol. 42, no. 3, pp. 539-551, Mar. 1994.
[79] G. Xu and T. Kailath, "Fast Estimation of Principal Eigenspace Using Lanczos Algorithm," *SIAM Journal on Matrix Analysis & Applications*, vol. 15, no. 3, pp. 974-994, July 1994.
[80] Z. Xu and M.K. Tsatsanis, "Blind Adaptive Algorithms for Minimum Variance CDMA Receivers," *IEEE Transactions on Communications*, vol. 49, pp. 180-194, Jan. 2001.
[81] Z. Xu and P. Liu, "Kyrtosis Based Maximization/minimization Approach to Blind Equalization for DS/CDMA Systems in Unknown Multipath," IEEE ICASSP2002, Orlando, vol. III, pp. 2585-2588, 2002.
[82] Z. Xu, P. Liu and X. Wang, "Blind Multiuser Detection: From MOE to Subspace Methods," *IEEE*
[83] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995
[84] Cristianini, Nello; and Shawe-Taylor, John; *An Introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2000.
[85] Schölkopf, Bernhard; Burges, Christopher J. C.; and Smola, Alexander J. (editors); *Advances in Kernel Methods: Support Vector Learning*, MIT Press, Cambridge, MA, 1999.
[86] Boser, Bernhard E.; Guyon, Isabelle M.; and Vapnik, Vladimir N.; *A training algorithm for optimal margin classifiers*. In Haussler, David (editor); 5th Annual ACM Workshop on COLT, pages 144–152, Pittsburgh, PA, 1992. ACM Press
[87] Duan, Kai-Bo; and Keerthi, S. Sathiya (2005). "Which Is the Best Multiclass SVM Method? An Empirical Study". *Proceedings of the Sixth International Workshop on Multiple Classifier Systems*.
[88] Henry Stark, Youngyi Yang. "Vector Space Projections" John Wiley & Sons INC.
[89] Chen, S., Mulgrew, B. and Grant, P. M. (1993) A clustering technique for digital communications channel equalization using radial basis function networks. *IEEE Transactions on Neural Networks*, 4, (4), 570-579.
[90] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002

[91] Jyrki Kivinen, Alexander J. Smola, Robert C. Williamson: Online learning with kernels. IEEE Transactions on Signal Processing 52(8): 2165-2176 (2004).

[92] Contemporary Communication Systems Using Matlab - Proakis and Salehi

[93] CDMA with MATLAB and Simulink, Yukitoshi Sanada*, Keio University.*

[94] Matlab Programming for Engineers**,** Stephen J. Chapman.

[95] B. Widrow and M.A. Lehr, ``30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation," *Proceedings of the IEEE*, 78(9):1415-1442, September 1990.