Thèse

présentée pour obtenir le grade de docteur
de l'École Nationale Supérieure des Télécommunications
Spécialité : Signal et Images

Anahid SAFAVI

Contributions pour l'exploitation de la diversité en transmission et en réception dans les systèmes à antennes multiples

Soutenue le 6 Novembre 2003 devant le jury composé de :

Pierre Duhamel	Président
Jean-Pierre Delmas	Rapporteurs
David Gesbert	
Jean-Claude Belfiore	Examinateurs
Adel Belouchrani	
Karim Abed-Meraim	Directeur de thèse

To my family

Acknowledgements

Je tiens à remercier chaleureusement Karim ABED MERAIM, pour m'avoir encadrée pendant toute ma thèse et fait preuve tout du long de patience et d'attention.

Je souhaite remercier Henri MAITRE, pour m'avoir accueillie au sein du département Traitement du Signal et de l'Image.

Je remercie également David GESBERT et Jean-Pierre DELMAS, qui ont accepté d'être les rapporteurs de ma thèse et dont les corrections m'ont permis de l'enrichir.

Je remercie aussi Pierre DUHAMEL, pour avoir présidé le jury de ma thèse.

Je remercie Jean-Claude BELFIORE et Adel BELOUCHRANI pour avoir accepté d'être les examinateurs de ma thèse.

Merci à ma famille, qui a toujours cru en moi.

Merci à mes amis et collègues, qui m'ont toujours soutenue.

Merci à mon mari, qui m'a donnée l'énergie nécessaire pour achever ma thèse.

Résumé

Un des principaux défis pour la prochaine génération des systèmes de communication sans fil est d'utiliser des ressources limitées telles que la bande passante, d'une manière efficace pour fournir une qualité et une capacité qui soient suffisantes pour la gamme grandissante de services.

Il est bien connu que les réseaux d'antennes peuvent améliorer la performance et/ou le débit des systèmes de communication sans fil dans un environnement évanouissement. Les réseaux d'antennes peuvent être employés du côté de l'émetteur et/ou du côté du récepteur , créant de la *diversité en transmission* et/ou en *réception*.

L'objectif de cette thèse est d'étudier, de développer et de combiner des techniques de traitement du signal et/ou de codage afin d'exploiter la diversité en transmission et/ou en réception dans des canaux évanouissement.

Dans la première partie de notre thèse, consacrée au côté émetteur, nous traiterons particulièrement les aspects du codage spatio-temporel dans le systme MIMO (multi-entrées multi-sorties) et nous fournirons une nouvelle architecture spatio-temporelle adaptée au contexte CDMA (systèmes étalement de spectre). Nous examinerons jusqu'oú nous pourrons aller en termes de performance, quand nous utiliserons ces schémas de codage.

La plupart des contributions faites dans le contexte MIMO considèrent que le canal de propagation est connu du côté de l'émetteur et/ou du récepteur. Dans la réalité, le canal de propagation devrait être estimé au moins du côté du récepteur. Habituellement, dans les systèmes de communication, on utilise des séquences d'entraînement afin d'estimer le canal de propagation. Cependant, des résultats récents montrent que les systmes utilisant les séquences d'entraînement dans le contexte MIMO sont sous-optimaux, particulièrement faible SNR (rapport signal sur bruit). De plus, ils réduisent le flux des informations dans le système.

C'est pourquoi, dans la seconde partie de notre thèse, qui concerne le côté récepteur, nous nous intéresserons aux techniques d'identification aveugle, et en particulier aux méthodes sous-espace. Dans cette partie, nous considérerons les principaux inconvénients des algorithmes d'identification de type sous-espace, tels que le coût de calcul élevé et le manque de robustesse par rapport la surestimation d'ordre du canal de propagation. Dans ce cadre, nous présenterons plusieurs algorithmes, qui pallient aux limites de cette technique. Tous les algorithmes proposés ont été validés par des simulations.

Summary

One of the major challenges for the next generation of wireless communication systems is to utilize limited resources like spectrum, efficiently enough to provide sufficient quality and capacity for the whole range of services.

It is well known that multiple-element antenna arrays can improve the performance and/or data rate of wireless communication systems in a fading environment. Antenna arrays can be employed at the transmitter and/or receiver side creating *transmit* and/or *receive diversity*.

The aim of this PhD. thesis is to study, develop and combine efficient signal processing and/or coding techniques to exploit transmit and/or receive antenna diversity over wireless fading channels. Consequently, our document is divided in two parts: *transmitter side* and *receiver side*.

In the first part of our document devoted to transmitter side, we deal especially with Space-Time coding aspects in MIMO (Multi-Input Multi-Output) system and we provide a new Space-Time architecture adapted to CDMA (Coded Division Multiple Access) context. We will examine how far we can go in performance when we use these coding schemes.

Most of the contributions made in the MIMO context assume that the propagation channel is known at the transmitter and/or receiver side. In practice, the propagation channel might be estimated at least at the receiver side. In a traditional communication context a training sequence is used in order to identify the propagation channel. However, recent results show that the training-based schemes in MIMO context are suboptimal specially at low SNR (Signal to Noise Ratio). Moreover, they reduce the through put of system.

That is why we focus in the second part of our document -receiver side- on the practical blind identification techniques, specially subspace-based schemes. In this part, we deal with main drawbacks of the subspace identification algorithms such as high computational complexity and lack of robustness to order over-estimation of propagation channel. Herein, we give several algorithms that overcome limitations of this technique. All of these algorithms are validated by simulations.

Contents

Notations

Through out this report, small letters are used to denote scaler, complex or real variables. In order to denote real, complex or integer vectors we use small boldface letters and for real or complex matrices we use capital boldface letters.

$\Re(.)$	real part of complex variable or matrix
$\Im(.)$	imaginary part of complex variable or matrix
$E(.)$	statistical expectation
$\|\|.\|\|^2$	Euclidean (frobenious) norm
$\|.\|$	absolute value
$\boldsymbol{I}_n$	identity matrix of size n
$\mathbf{0}_n$	$n \times n$ matrix of zeros
$\mathbf{0}$	matrix of zeros with appropriate size
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
$\mathbf{A}^{\#}$	pseudo inverse of matrix $\mathbf{A}$
$\mathbf{A}^T$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	conjugate transpose of matrix $\mathbf{A}$
$\mathbf{A}_{i,j}$	element (i,j) of matrix $\mathbf{A}$
$\mathbf{A}_{i,:}$	i-th row of matrix $\mathbf{A}$
$\mathbf{A}_{:,i}$	i-th column of matrix $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	trace of matrix $\mathbf{A}$
$\det(\mathbf{A})$	determinant of matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{a}_1, \cdots, \mathbf{a}_k)$	a $k \times k$ (block) matrix with diagonal elements $\mathbf{a}_1, \cdots, \mathbf{a}_k$
$\mathrm{vec}(\mathbf{A})$	vector valued function of matrix $\mathbf{A}$ it can be written as : $\mathrm{vec}\big([\mathbf{a}_1, \cdots, \mathbf{a}_k]\big) = [\mathbf{a}_1^T, \cdots, \mathbf{a}_k^T]^T$
$\otimes$	Kronocker Product. Let $\mathbf{A}$ be a $n \times m$ and $\mathbf{B}$ be a $k \times l$ matrix $\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{1,1}\mathbf{B} & \mathbf{A}_{1,2}\mathbf{B} & \dots & \mathbf{A}_{1,m}\mathbf{B} \\ \mathbf{A}_{2,1}\mathbf{B} & \mathbf{A}_{2,2}\mathbf{B} & \dots & \mathbf{A}_{2,m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{n,1}\mathbf{B} & \mathbf{A}_{n,2}\mathbf{B} & \dots & \mathbf{A}_{n,m}\mathbf{B} \end{bmatrix}$ the resulted matrix is of dimension $nk \times ml$.

Abbreviations

AWGN	Additive White Gaussian Noise
BSI	Blind System Identification
BPSK	Binary Phase Shift Keying
BLAST	Bell Laboratories Layered Space Time
CMA	Constant Modulus Approach
CSI	Channel State Information
CDMA	Coded Division Multiple Access
DOA	Direction of Arrival
DTD	Delay Transmit Diversity
DAST	Digonalized Algebraic Space Time
HOS	Higher Order Statistics
i.i.d.	independent identically distributed
ISI	Inter Symbol Interfrence
LSF	Least Squares
LSF	Least Squares Fitting
MNS	Minimum Noise Subspace
MSE	Mean Squares Error
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Squares Error
OTD	Orthogonal Transmit Diversity
OMNS	Orthogonal Minimum Noise Subspace
PCS	Properly Connected Sequence
QAM	Quadratic Amplitude Modulation
SD	Sphere Decoder
SS	Spreading Sequence
SNR	Signal to Noise Ratio
SOS	Second Order Statistics
STC	Space Time Codes
STS	Space Time Spreading
SIMO	Single-Input Multiple-Output
SMNS	Symmetic Minimum Noise Subspace
STBC	Space Time Block Codes
SVD	Singular Value Decomposition

Chapter 1

Motivations

One of the major challenges for the next generation of wireless communication systems is to utilize limited resources i.e. spectrum, efficiently enough to provide sufficient quality and capacity for the mixtures of services.

It was well known that multiple-element antenna arrays can improve the performance and/or data rate of wireless communication systems in a fading environment [58]. Antenna arrays can be employed at the transmitter and/or receiver side creating *transmitter* and/or *receiver diversity*. However, several year later (in 90's) these techniques had not been fully exploited.

It has been shown later in [91], that the capacity of multiple-element antenna arrays in a fading environment grow at most linearly with the number of transmit antennas as long as the number of receiver antennas is greater than or equal to the number of transmit antennas [1]. Being sure of the capacity enhancement of multi-element antenna arrays was the first step, the second step was to find the appropriate architectures and a know-how to use this capacity in an optimal way.

Implementation of the first laboratory prototype of multiple-element antenna arrays in Bell laboratories operating at high spectral efficiency [46], might be seen as a starting point of a new challenge in the context of the wireless communications. The substantial performance obtained using multi-element antenna array, motivated the researchers to improve existing schemes or develop new architectures.

Since then a lot of research efforts have been made in the context of so-called Multiple-Input Multiple-Output (MIMO) systems (see [41] and references therein for an overview of MIMO systems).

[1] Under assumption the fading is uncorrelated

Main axes of research are focused on the information theoretic aspects of MIMO systems leading to different contributions on capacity evaluation, the modulation and coding/decoding aspects in MIMO system in order to propose high spectral efficiency schemes adapted to Multi-element systems and finally signal processing aspects which provide mainly the efficient way of extracting information at the receiver side and canceling the interference.

1.1 Objectives of this document

The aim of this Ph.D. thesis is to study, develop and combine efficient signal processing and/or coding techniques to exploit transmit and/or receive antenna diversity over wireless fading channels. This document is divided on two parts *transmitter side* and *receiver side*.

In the first part of this document, transmitter side, we deal especially with Space-Time coding aspects in MIMO systems and we provide a new Space-Time architecture adapted to CDMA context. We will examine how far we can go in performance when we use these coding schemes.

Most of the contributions made in the MIMO context, assume that the propagation channel is known at the transmitter and/or receiver side. In practice, the propagation channel might be estimated at least at the receiver side. In a traditional communication context a training sequence is used in order to identify the propagation channel. However, recent results [52] show that the training based schemes are suboptimal specially at low SNR (Signal to Noise Ratio). Moreover, training base schemes reduce the through put of system.

That is why we focus in the second part of this report on the practical blind identification techniques, specially subspace based schemes. In this part, we deal with main drawbacks of the subspace identification algorithms like computational complexity and robustness to order over estimation of propagation channel. Herein, we give several algorithms that overcome limitations of this technique.

1.2 Contributions

Major contributions of this report should be divided into two different parts:

1.2.1 Part I : Transmit diversity

In this part, we propose an approach which exploits the benefits of Space-Time codes on CDMA framework. Contrary to other contributions for CDMA systems in which only the inherent diversity in the system has been exploited, we propose a Space-Time coding scheme that achieves the maximum diversity gain and a good coding gain in a CDMA multi antenna context.

This scheme is obtained by a proper combination of available Spreading Sequences and Space-Time Algebraic codes over an arbitrary number of transmit antennas. We also propose a coding scheme to adapt the proposed architecture to the multi-user and multi-path case .

The receiver has one or several receive antennas and the decoding procedure consists of decorrelating all users first, then applying Maximum Likelihood Sphere Decoder (SD) to each user separately.

Under quasi-static fading assumptions, we propose a method of blind channel estimation based on the orthogonality between signal and noise subspaces.

1.2.2 Part II : Receive diversity

As mentioned before, we deal here with two important drawbacks of subspace based methods: computational complexity and channel order overestimation.

1.2.2.1 Minimum Noise Subspace (MNS)

Our contributions are based on the concept of Minimum Noise Subspace (MNS). MNS algorithm is a promising low computational cost subspace technique. The main advantage of the MNS method is that the large matrix eigendecomposition is avoided and the noise vectors are computed in a parallel scheme as the least eigenvectors of covariance matrices corresponding to an *appropriate set* of combinations of system output that form a basis of the rational noise subspace.

We propose two extensions of MNS method called (Orthogonal Minimum Noise Subspace) OMNS and (Symmetric Minimum Noise Subspace) SMNS. We propose also an efficient implementation technique to compute the noise subspace that converges rapidly in few iterations.

Symmetric Minimum Noise Subspace (SMNS)

Indeed, in the original MNS method, certain system outputs are used more than others, depending on the chosen *appropriate set* . This might

lead to poor estimation performance if the system outputs that are used most correspond to the 'worst system channels'. This raises the problem of the 'best' choice of *appropriate set* of outputs. In SMNS technique we guarantee a certain symmetry between the system outputs.

SMNS method has the advantage of better robustness and estimation accuracy at the cost of a slight increase of the computational cost in comparison with original MNS.

Orthogonal Minimum Noise Subspace (OMNS)

Here the noise subspace is formed through computation of noise vectors that correspond to an orthogonal set of noise polynomial vectors (Orthogonal basis of the rational noise subspace).

The orthogonality of the noise subspace in OMNS method improve quality of parameter estimation. OMNS method is a recursive method contrary to the MNS that can be implemented in a parallel scheme. The computational cost of OMNS method is much less than that of original subspace.

Efficient implementation

In all MNS-like methods we should compute the least eigenvector of the output covariance matrix. Even though, computing 1 single eigenvector costs $\mathcal{O}(n^2)$ (n being the size of considered matrices), we found out that existing algorithms for extracting minor subspaces or the minor eigenvectors are un-efficient and slowly convergent in comparison with those dedicated to principal subspaces or principal eigenvectors.

For this reason, we propose an adaptive implementation in which we use the power method in conjunction with a RLS-type algorithm to extract the principal eigenvectors of each of the considered matrices.

1.2.2.2 Channel order overestimation

We propose in this part two different categories of algorithms robust to order overestimation of propagation channel in a SIMO context. When an overestimation of channel degree occurs, a specific structural property over estimated channel can be derived. This structural property is used as a framework to obtain several algorithms :

Least Squares Fitting Approach

In this approach we minimize a least squares fitting cost function based on observation and the estimated channel equalizer expression without any additional assumption over input data. We propose here two methods called 'Non-linear' and 'Bilinear' optimization approach. We also propose an efficient way for initialization of both algorithms to guarantee the convergence of both algorithms in few iterations.

Constant Modulus Approach

In this approach we minimize a Constant Modulus Criterion, based on the observations and the estimated channel equalizer expression. This approach is valid only when the input sequence is of constant modulus.

1.3 Document organization

In **chapter 2** of this document we first recall the main concept of Space-Time codes together with their design criteria as first proposed in [90]. Construction and properties of Space-Time block codes from orthogonal design are then presented. A powerful alternative to Space-Time block codes called Diagonal Algebraic Space-Time (DAST) block codes are also described in this chapter.

In **Chapter 3** we review Space-Time architectures proposed in the context of CDMA for MIMO systems and we give a new Space-Time architecture for CDMA system. We give also the coding and decoding algorithm for the case of multi-user and multi path. At the end of this chapter we propose a blind identification algorithm in order to identify the coefficients of propagation channel.

In **chapter 4** we give several advantages concerning the use of Blind System Identification algorithms in a MIMO context. Later on in this chapter we focus on subspace methods and algorithms for MIMO context and we give briefly advantages and main limitations of subspace based methods in the MIMO context. In the rest of the document we propose several techniques to overcome these limitations.

Chapter 5 is devoted to the Minimum Noise Subspace (MNS) method and its extensions. MNS method is a computationally efficient version of Subspace method. We derive several extensions of this algorithm adapted to different scenarios. We will give here the exact computational complexity of each method and we propose an efficient way to implement the proposed algorithms.

Chapter 6 is devoted to identification algorithms robust to order overes-

timation. Here, we exploit the structural properties of subspace method to propose two categories of algorithms robust to order overestimation of channel degree.

Finally, We conclude the report in **chapter 7** and we propose several perspectives for further works.

Part I

Transmitter Side ...

Chapter 2

Exploiting transmit diversity

Achieving transmit diversity through the use of several transmitter antennas is a recent but important topic. There are several communication scenarios (like broadcasting) where only transmit diversity can be applied in the system. Moreover, for some applications, transmit diversity provides a substantial gain comparing with that of the receive diversity. Several transmit diversity strategies are already known [69].

Among other techniques we can mention that of the Bell Laboratories Layered Space-Time (BLAST) and Space-Time coding scenario. BLAST ([35]) transmits independent coded/uncoded bit-streams from each transmit antenna using the spatial dimension available in a multi-element antenna array, whereas Space-Time coding exploit both temporal and spatial diversity to construct high performance coding designs. Otherwise, in the Space-Time coding, error-correction coding and multiple-antenna techniques are combined to further improve the performance of system, operating at the maximum possible diversity advantage.

In this document, we address only the Space-Time coding schemes, more precisely we address new algebraic coding scheme derived in [23].

This chapter is organized as follows : First in 2.1 we give some historical perspectives which led to the Space-Time Code Concept. In section 2.2 we will briefly introduce Space-Time codes and give the design criteria of such codes in the multi-antenna context. Then, we will present the idea of Space-Time Block Codes (STBC) and the well known Space-Time Block Codes from orthogonal design. Section 2.3 is devoted to algebraic codes: a powerful tool to construct STBC codes. In this section we present a STBC scheme called Diagonal Algebraic Space-Time (DAST) codes, constructed based on algebraic

codes. Finally, before conclusion, we compare two schemes presented in 2.6 in terms of gain, decoding complexity and achievable data rate.

2.1 Historical perspectives

Although ST codes and their design criteria have been derived first in [90], the idea of exploiting transmit diversity using several transmission antenna in order to reduce the effect of multi-path fading was expressed and employed long time before [105, 104, 76].

For example in [104], the same data symbol is transmitted from two antennas simultaneously with a delay of one symbol interval. It has been shown that [85, 104] this scheme provides a diversity gain of two. This scheme is an example of delay diversity system operating over two transmit antennas. Looking this scheme from a coding point of view rather than a signal processing problem, Tarokh [90] recognized that the codeword transmitted over two transmit antennas is a repetition code with transmit diversity of two. Thereafter, he tried to find the channel codes which are better than repetition codes in term of performance. This leads to a new category of codes adapted to multiple-antenna systems called Space-Time codes.

In the other hand, during several years there are a lot of research activities made at coding community in order to construct signal constellations matched to the Rayleigh fading channel [42, 43, 15, 16, 17]. Although at the beginning the idea was to find signal constellations over Single Input Single Output (SISO) systems with good performance robust to fading effects, it has been shown later [23], that some of these constellations can be also adapted to the context of multiple-antennas providing significant coding and diversity gain.

2.2 Space-Time codes

Space-Time codes exploit both temporal and spatial diversity available in a Multiple Input Multiple Output (MIMO) system in order to construct coding designs which effectively mitigate fading effects and are able to send data over several transmission paths within the propagation channel. As mentioned before, the important difference between Space-Time codes and the other transmit diversity techniques which use multiple antenna at the transmitter side is that it offers an intrinsic *coding gain* thanks to the coding procedure added at the

transmitter side. Therefore, Space-Time code should be seen more than a simple transmit scheme which exploit space and time axes separately to access the transmission channel.

Let us consider a wireless communication system with n transmit and m receive antennas. Data is encoded using Space-Time codes before transmission. The encoded data is then split into n sub-streams. At each time slot t, n data samples are sent simultaneously over n transmit antennas. The received signal at each receiver is the superposition of the encoded data which is corrupted by additive noise and affected by Rayleigh Fading channel. Figure 2.1 represents principal idea behind Space-Time coding scheme.

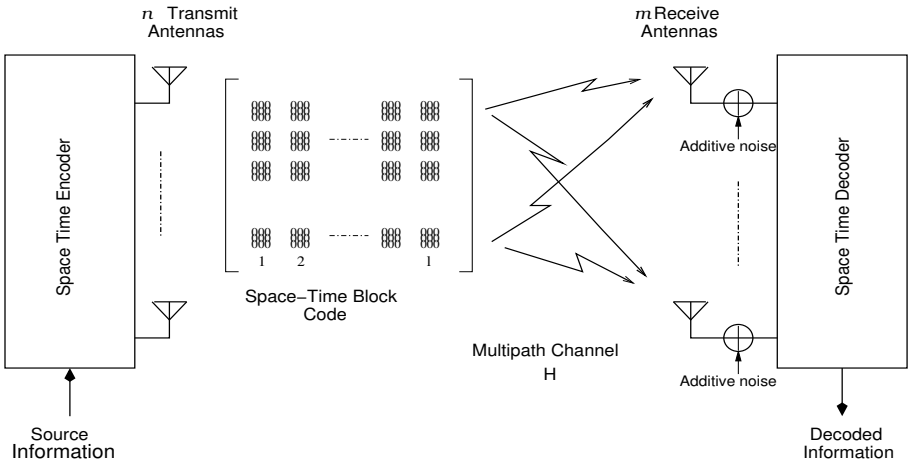

Figure 2.1: Space-Time coding scheme

2.2.1 Construction criteria

The performance of Space-Time codes is determined by measuring their *diversity gain* and *coding gain*. These two parameters are known as construction tools when designing Space-Time codes over Fading channels. Design criteria for multiple antennas are first established by Tarokh et al. in his quite well-known paper [90]. Here, we recall briefly the construction criteria as it was expressed at the original work in order to highlight the relationship between these parameters and the matrices constructed from code sequences sent over transmit antennas.

In a Multi-antenna system with the aforementioned characteristics, the error

probability of sending code sequence

$$\boldsymbol{x} = x_1^1 x_1^2 \cdots x_1^n x_2^1 x_2^2 \cdots x_2^n x_l^1 \cdots x_l^2 \cdots x_l^n$$

[1] and deciding in favor of

$$\boldsymbol{y} = y_1^1 y_1^2 \cdots y_1^n y_2^1 y_2^2 \cdots y_2^n y_l^1 \cdots y_l^2 \cdots y_l^n$$

is obtained by :

$$P(\boldsymbol{x} \rightarrow \boldsymbol{y} | h_{ij}, i = 1, 2, ..., n, j = 1, 2, 3, ...m) \leq \exp\ (-d^2(\boldsymbol{x}, \boldsymbol{y}) E_s / 4N_0) \quad (2.1)$$

where $d^2(\boldsymbol{x}, \boldsymbol{y})$ can be written as:

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{m} \sum_{t=1}^{l} \left| \sum_{i=1}^{n} h_{ij} (y_t^i - x_t^i) \right|^2$$

$\boldsymbol{x}$ and $\boldsymbol{y}$ denote two different code sequences and h_{ij} is the path gain between transmitter antenna i and the receiver antenna j. The channel is assumed quasi static flat fading [2]. i. e. path gains are assumed constant during a frame and vary from one frame to another. $N_0/2$ is the noise variance per dimension and E_s stands for the average energy per symbol of constellation. $d^2(\boldsymbol{x}, \boldsymbol{y})$ can be written in quadratic form as:

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{m} \boldsymbol{\Omega}_j \boldsymbol{A} \boldsymbol{\Omega}_j^{\dagger}$$

where $\boldsymbol{\Omega}_j = [h_{1j}, h_{2j}, ..., h_{nj}]$, $\boldsymbol{A} = \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y})^H$ and $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y})$ stands for:

$$\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y}) = \begin{bmatrix} x_1^1 - y_1^1 & x_2^1 - y_2^1 & \cdots & x_l^1 - y_l^1 \\ x_1^2 - y_1^2 & x_2^2 - y_2^2 & \cdots & x_l^2 - y_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n - y_1^n & x_2^n - y_2^n & \cdots & x_l^n - y_l^n \end{bmatrix}$$

Next, expressing $d^2(\boldsymbol{x}, \boldsymbol{y})$ in term of eigenvalues of $\boldsymbol{A}$, $(\lambda_i, i = 1, 2, ...n)$ and computing the upper bound of (2.1), we obtain [90]:

$$P(\boldsymbol{x} \rightarrow \boldsymbol{y}) \leq \left(\frac{1}{\prod_{i=1}^{n} (1 + \lambda_i E_s / 4N_0)} \right)^m \quad (2.2)$$

[1] x_t^i denotes the transmitted data at time slot t over transmit antenna i.

[2] Channel is time invariant over the observation period and its coefficients are modeled by independent samples of complex Gaussian random variables with zero complex mean and variance 0.5 per dimension. Therefore, the signals transmitted from different antennas undergo independent fades.

Let r denote the rank of $\boldsymbol{A}$ therefore $n-r$ eigenvalues of $\boldsymbol{A}$ are zero, then the equation (2.2) implies:

$$P(\boldsymbol{x} \rightarrow \boldsymbol{y}) \leq \left(\prod_{i=1}^{r} \lambda_i \right)^{-m} (E_s/4N_0)^{-rm} \tag{2.3}$$

Therefore, a diversity advantage of rm and a coding advantage of $(\lambda_1\lambda_2\lambda_3...\lambda_r)^{1/r}$ is achieved. $(\lambda_1\lambda_2\lambda_3...\lambda_r)$ is the absolute value of the sum of the determinants of all the principal $r \times r$ cofactors of $\boldsymbol{A}$.

Remark : The diversity gain is the power of Signal to Noise Ratio (SNR) in the denominator of Pairwise Error Probability expression and the coding gain is the gain obtained over an uncoded system which provides the same diversity.

Having Pairwise Error Probability expression in mind, design criteria for Space-Time codes over uncorrelated Rayleigh fading channel can be expressed as :

- *The Rank Criterion* : In order to obtain the maximum diversity mn, the matrix $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y})$ has to be full rank for any code word $\boldsymbol{x}$ and $\boldsymbol{y}$. If $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y})$ has minimum rank of r, then a diversity of rm is achieved. Diversity gain was also derived in [51].

- *The Determinant Criterion*: Suppose that the diversity advantage of rm is our target. The minimum of rth roots of the sum of determinants of all $r \times r$ principal cofactors of $\boldsymbol{A} = \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y})^{\dagger}$ taken over all distinct codewords $\boldsymbol{x}$ and $\boldsymbol{y}$ corresponds to the coding advantage. The design target is making this sum as large as possible. If a maximum diversity nm is our target, the minimum of the determinant of $\boldsymbol{A}$ taken over all distinct codewords must be maximized.

The above mentioned criteria are used in the following in order to design ST codes over flat fading channels. Using the appropriate expression of Pairwise Error Probability, the design criteria over fast fading channels also can be obtained [90].

In this document we limit our discussion to the ST codes designed for the flat fading channels. Even if, some of the proposed codes are also suitable for the fast fading channels.

In his original paper [90], Tarokh has proposed trellis codes adapted for multiple-antenna for coding and Viterbi decoder for decoding such constructed

codes. The proposed ST codes are optimal; that is they produce a good tradeoff between the transmission rate, diversity, trellis complexity, and constellation size. Moreover, it is mentioned that it is possible to design ST codes in order to obtain the highest rate, maintaining a fixed diversity advantage. Otherwise, if it is necessary, the ST codes used in this document can be exploited in order to construct some high date rate schemes [23].

Remark 1 : There is a tradeoff between data rate, diversity and coding gain and implementation complexity when designing ST codes. The choice of one scheme rather than another depends on the practical constraints and desired objectives. Here, we focus on systems which are not working at rates greater than the usual systems but operating at significantly lower signal-to-noise ratios. Proposed systems exploit the maximum transmit diversity available in a multi-antenna context and they guarantee a good coding gain with a good performance. More details for these kind of constructions will be given in next chapter.

Remark 2 : ST codes maintain their coding and diversity advantages in presence of channel estimation errors, mobility and multiple paths as shown in [89]. However, it is known that when the fading coefficients between channels are correlated the capacity and consequently the performance of system degrades dramatically [86, 13] and [24] [3].

2.2.2 Space-Time block codes from orthogonal design

ST trellis codes proposed in [90] have been designed by hand. Moreover, their coding/decoding procedure is computationally complex (ML Viterbi decoding for a 8-PSK constellation with 32 states). These considerations motivated the researchers to find coding schemes simple to construct and easy to decode.

The reduced complexity of block decoding techniques and the existence of a simple block scheme for two transmit antennas (Alamouti scheme [10]) motivated the research community to generalize such schemes.

Definition 1 *An ST block code associates with each information vector* $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ *an* $n \times l$ *matrix* $\boldsymbol{B}(\boldsymbol{x})$ *with elements* b_{jt}, $j = 1, 2, ..., n$ *and* $t = 1, 2, ..., l$ *such that the element* b_{jt} *is sent over transmit antenna* j *at time* t.

[3] Generally speaking, the systems using orthogonal structures are more robust to the effect of correlated fades. This fact is demonstrated via simulations in [14].

2.2.2.1 Construction and properties

Mathematical framework of orthogonal designs [75], has been used in [88, 87] to construct block ST codes.

In an ST block code from orthogonal design the aforementioned matrix $\boldsymbol{B}(\boldsymbol{x})$ is an orthogonal matrix.

The reader can refer to [88] for construction and decoding details for more information. Here, we just recall some properties of so called Space-Time Orthogonal design. The important advantages of orthogonal designs is summarized in the following :

- They provide maximum possible achievable rate [4] of 1 symbol/sec at full diversity order mn. We remember that n is the number of transmit antennas and m denotes the number of receive antennas.
- Maximum-Likelihood decoding procedure is extremely simple. Indeed, due to the orthogonal structure of these codes the ML decoding is possible only by linear combining at the receiver side.
- The orthogonal structure of these codes makes them more robust (comparing with trellis codes) when the channel fading coefficients are correlated [14].

Some disadvantages of these codes are:

- Real orthogonal designs exist only for a limited number of transmit antennas $n = 2, 4$ and 8.
- For complex orthogonal designs the maximum achievable rate for the fixed diversity mn depends on n and it can be 2/3 symbol/sec.

  The only complex orthogonal scheme with maximum rate of one is Alamouti scheme [10].
- To the author's knowledge, the number of complex orthogonal designs are also limited. They have been reported in [88] for only $n = 2$.

In the following we present an alternative scheme for exploiting the maximum diversity in the Multi-antenna context. We introduce first, the special coding/decoding aspects of these schemes, then we compare it to the previous one, i.e. the orthogonal design ST code.

[4] It has been shown that for a diversity order of mn it is possible to transmit 1 symbol per time slot and that is the maximum possible rate.

2.3 Algebraic codes : A powerful alternative

Among the existing techniques to build Space-Time block codes we can mention a powerful approach using algebraic number theory to construct full diversity ST codes. This theory has been used to construct appropriate modulation schemes well adapted to Rayleigh Fading channel based on rotated constellations.

Here, we recall the main properties of such a constellation and we give a brief review on the way these constellations have been constructed.

2.3.1 Rotated constellation

This idea has been first proposed by K. Boulle and J.-C. Belfiore in [15]. In this work they have proved that an n-dimensional constellation where all pairs of distinct symbols have all their coordinates distinct, leads to n-th order diversity. In other words a two-dimensional rotated QAM constellation as shown in figure 2.2 has a diversity order of 2 comparing with a classical QAM constellation with a transmit diversity of 1. Let's consider the above mentioned example: if one of

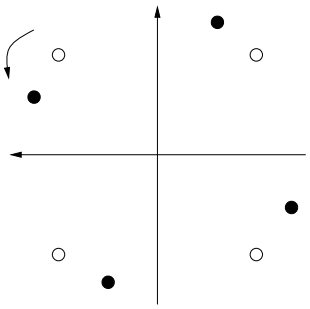

Figure 2.2: The transmission diversity of 2 for the rotated constellation

the symbols of the resulted rotated QAM constellation is affected by fading, the receiver can recover the transmitted symbol as all of the transmitted symbols have their coordinates different from each other. Therefore the performance of system can be enhanced using the appropriate rotated constellation over a fading channel. This can be shown mathematically using the expression of Cheroneff upper bound on the pairwise error probability. See [30, 15, 16] for more details . Later on, it has been mentioned out that the diversity order is not the only parameter which enhance performance of the system. For a given signal-to-noise ratio it is known that maximizing the *Minimum Product*

Distance (MPD) between two constellation points is another parameter which affects the system performance.

At the beginning, these constellations were only used to improve the performance of SISO system over Fading channels. M. O. Damen was the first who applied these constellations on the context of Multiple-antennas in [23] by imposing design criteria of 2.2.1 on algebraic codes. This idea was then generalized in [27] to propose new block ST practical architecture, where the rotated constellation are used to code information symbols.

A lot of research efforts have been made in coding community in order to find the best rotations in term of diversity and coding gain. We can mention here the work of [16] where a method for construction of best rotations for low dimension $n \leq 8$ has been derived. Best rotation means here the rotation with full transmit diversity and the maximum values of MPD. In the following, for $n \geq 8$ we consider the quasi-optimum rotations proposed in [11] which provide maximum diversity gain for certain number of transmit antennas and a good coding gain.

The latter have been recently generalized in [36] to provide full transmit diversity for any dimension.

2.3.1.1 Construction Procedure and Properties

We expressed that the idea behind the rotated constellation is to find rotations with maximum diversity gain and a good value of coding gain. Here we present first the exact definition of *diversity gain* and *minimum product distance* in this context, then we give the algorithm used in order to construct such rotations.

Diversity Order Diversity order is referred to as the minimum number of distinct components between two constellation points. It is also called signal space diversity in [16] . In this report we use the rotations constructed for $n = 2^q$ transmit antennas. These constellations provide maximum diversity order over fading channels.

Minimum product distance is a design parameter defined by

$$d_{d,min} \stackrel{\triangle}{=} \min_{\boldsymbol{y}=\boldsymbol{x}_1-\boldsymbol{x}_2, \boldsymbol{x}_1 \neq \boldsymbol{x}_2} \prod_{j=1}^{d} |y_j|$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ denote two distinct codewords and $\boldsymbol{y} = [y_1, ..., y_d]$.

Construction procedure For small dimensions of $d = 2, 3, 6$, the quasi-optimal rotation matrix with the best value of the minimum product distance is obtained using an iterative algorithm [16].

For larger values of $d = 2^q$ ($q \geq 3$), the quasi-optimal rotation matrix with a good value of the product distance is obtained using a general algorithm :

```
M=[];
for j=1:d
  M(j,:)=sqrt(2/d)*cos((pi/(4*d))*(4*j-1)*(2*(1:d)-1));
end
```

where the resulted matrix corresponds to a $d \times d$ rotation matrix. For example for $d = 2$ the optimal rotation matrix corresponds to :

$$\boldsymbol{M}_2 = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \text{ with } \theta = 1.0172 \,\text{radians}. \tag{2.4}$$

Remark : The optimal rotations are those that guarantee a maximum value of minimum product distance. The relation between these parameters and the design criteria established in 2.2.1 will be explained later when we describe the practical schemes.

2.3.2 Sphere Decoder

Sphere decoder is used in order to decode the rotated constellations described before. Here we give a brief review of this decoder and its properties.

Sphere decoder is a Maximum Likelihood decoder used to decode lattices[5]. in presence of AWGN noise [99] with a moderate complexity. It has been shown that each Multiple-antenna system has a lattice representation [23]. Therefore, Sphere Decoder can be used to decode multiple-antenna communication systems as well. Despite of Viterbi decoder, the complexity of the Sphere decoder is quasi-independent of constellation size which makes it interesting for high data

[5] An n-dimensional lattice Λ can be written as

$$\Lambda = \{\boldsymbol{y} | \boldsymbol{y} = \boldsymbol{G}\boldsymbol{x}, \boldsymbol{x} \in Z^m\}.$$

$\boldsymbol{G}$ is called the generator matrix and columns of $\boldsymbol{G}$ are linearly independent m-dimensional vectors called the basis of Λ. $\boldsymbol{x}$ represents the information vector. Therefore, each point of lattice $\boldsymbol{y}$ can be written as a linear combination of a basis vector (see [21] for more details on lattice representation).

rate communication systems. The Sphere decoder can easily be adapted to decode coded/uncoded communication system. For example this decoder has been used in the context of ST block codes [23, 27, 22]. In [18] Sphere Decoder is also used in the context of multi-carrier CDMA systems.

We will show in 2.3.2.1 that the problem of ML decoding of a lattice can be expressed geometrically as a problem of finding a basis of lattice with *short* vectors. This problem is an important problem in geometry of numbers and it has been investigated since the 80's. We believe that M. Pohst was the first who developed an efficient algorithm for this purpose [72]. Later on, some further improvement of the algorithm has been made and a first complexity analysis has been performed in [33]. Moreover in the same paper it has been pointed out that the Lenstra Lenstra Lovasz (LLL) algorithm proposed by [61] is the best numerical algorithm for complexity reduction of the proposed algorithm. The Sphere Decoder algorithm has been applied first to the communication context in [98, 99]. It has been highlighted that the so called Sphere Decoder algorithm is well adapted to decode multidimensional modulation schemes in presence of Fading. A Generalized Sphere Decoder specially adapted to Multiple-antenna system has been proposed in [25] which makes possible the ML decoding of Multiple-antenna system with arbitrary number of transmit and receive antennas. Sphere Decoder was initially proposed for real valued systems . The generalization of SD to complex valued systems has been made in [29].

The complexity issues concerning SD is still an open and interesting topic. Related works on this topic are discussed later in section 2.3.2.2.

2.3.2.1 Algorithm Description

Here, we recall the structure and principals of Sphere Decoder algorithm. For more details and a flew chart of algorithm the reader can see [99]. Let us consider a Multiple-antenna system. The received vector $\boldsymbol{r}$ can be written as:

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{\nu} \tag{2.5}$$

where $\boldsymbol{x}$ represents the information vector of size n and $\boldsymbol{M}$ is the lattice generator matrix. $\boldsymbol{\nu}$ denotes the real AWGN noise in the system [6]. The channel is assumed to be known at the receiver side. The Maximum Likelihood decoding

[6] Here we consider the case of real information vector and real noise. The cases of complex valued vectors and Fading channels are treated later.

is based on the minimization of the following equation:

$$\min_{\boldsymbol{x}} ||\boldsymbol{r} - \boldsymbol{M}\boldsymbol{x}||^2 \tag{2.6}$$

Geometrically, the lattice decoding algorithm searches for the points of lattice which are found in the Sphere of a radius $\sqrt{C}$ centered at the received point $\boldsymbol{r}$. Therefore, there is no need to make an exhaustive search over all lattice points. Moreover, if the radius of sphere is properly chosen one can limit the number of operations used in order to find the desired point in sphere. Equation (2.6) is equivalent to :

$$\begin{aligned} Q(\boldsymbol{x}) &= \min_{\boldsymbol{x'}} ||\boldsymbol{M}\boldsymbol{x'}||^2 && (2.7)\\ &= \min_{\boldsymbol{x'}} (\boldsymbol{x'}^H \boldsymbol{M}^H \boldsymbol{M}\boldsymbol{x'}) && (2.8)\\ &= \min_{\boldsymbol{x'}} (\boldsymbol{x'}^H \boldsymbol{Q}^H \boldsymbol{Q}\boldsymbol{x'}) \end{aligned}$$

where $\boldsymbol{x'} = \boldsymbol{x} - \boldsymbol{\rho}$ and $\boldsymbol{\rho} = \boldsymbol{M}^{\#}\boldsymbol{r}$. $\boldsymbol{Q}^H\boldsymbol{Q}$ is the result of cholesky factorization of $\boldsymbol{M}^H\boldsymbol{M}$. Therefore, $\boldsymbol{Q}$ is an upper triangular matrix. We denote the elements of $\boldsymbol{Q}$ by $q_{i,j}$ for $1 \leq i, j \leq n$.

Using the above mentioned transformation, sphere of radius $\sqrt{C}$ is converted into an ellipsoid centered at the origin of the new coordinate system defined by $\boldsymbol{x'}$. The problem of ML decoding can now be seen as a problem of finding the shortest vector belonging to the new lattice represented above.

The corresponding range of two last elements of $\boldsymbol{x}$, x_n and x_{n-1} can be expressed as:

$$\begin{aligned} \lceil -\frac{\sqrt{C}}{q_{n,n}} + \rho_n \rceil \quad &\leq x_n \leq \quad \lfloor \frac{\sqrt{C}}{q_{n,n}} + \rho_n \rfloor \\ \lceil -\frac{\sqrt{C} - |q_{n-1,n} x_n^{'}|}{q_{n-1,n-1}} + \rho_{n-1} \rceil \quad &\leq x_{n-1} \leq \quad \lfloor \frac{\sqrt{C} - |q_{n-1,n} x_n^{'}|}{q_{n-1,n-1}} \rfloor \end{aligned} \tag{2.9}$$

where $\lceil x \rceil$ and $\lfloor x \rfloor$ denote respectively the integer ceil and the floor functions and $|.|$ denotes the absolute value. The expression of x_{n-2} can be also obtained by the same procedure until all points inside the sphere are found.

More precisely, the decoding algorithm starts from the last component x_n and performs a bounded search over a fixed upper and lower bound defined in [33] in order to find the shortest vector of lattice.At each step, the initial radius of sphere $\sqrt{C}$ is changed in order to find a new candidate closest point to the received point. See [16, 53] for a complete description of the algorithm. The Sphere Decoder program is also available on [1].

Remark 1 : This representation can easily been extended to the case of complex constellation with complex AWGN noise (see [29] for more details). The idea is to separate the complex and the real part of the received vector, writing:

$$\begin{aligned} \boldsymbol{r'} &\triangleq [\Re(\boldsymbol{r}^T) \quad \Im(\boldsymbol{r}^T)]^T \\ &= \begin{bmatrix} \Re(\boldsymbol{M}) & \Im(\boldsymbol{M}) \\ -\Im(\boldsymbol{M}) & \Re(\boldsymbol{M}) \end{bmatrix} [\Re(\boldsymbol{x}^T) \quad \Im(\boldsymbol{x}^T)]^T + [\Re(\boldsymbol{\nu}^T) \quad \Im(\boldsymbol{\nu}^T)]^T \end{aligned}$$

In this case the original dimension of lattice n is extended to $2n$. Therefore, the scheme is computationally complex. In [55] a Sphere Decoder algorithm adapted to complex vectors has been proposed. This algorithm is computationally less complex than separating real and complex elements of received vector. However, an exact analysis of complexity is not made in this paper.

Remark 2 : When the channel is Rayleigh Fading the Generator matrix $\boldsymbol{M}$ contains also Fading coefficients. Therefore, the generator lattice is totally arbitrary. In this case choosing the initial sphere radius becomes more crucial and decoding procedure might be slower in this case. It seems that the choice of initial sphere radius in the fading environments is not considered separately in the related works on complexity [53].

2.3.2.2 Complexity issues

For rotated lattices, the search radius is set to $C = 1$, and the computational complexity is approximated by $O(n^6)$ [99]. Recent results show that the decoding of a lattice needs at most $O(n^{4.5})$ operations for low signal-to-noise ratio and $O(n^3)$ operations for high signal-to-noise ratio [28]. Complexity issues of SD algorithm have also been addressed in [53, 77] for real data and the case where the number of transmit and receive antennas are equal. In [19], a modified sphere decoding algorithm has been proposed where the performance is independent of the initial choice of radius Sphere.

2.4 Diagonal Algebraic Space-Time codes

Rotated constellation and Sphere Decoder have been already introduced in previous sections. In this section we will present Diagonal Algebraic Space-Time codes constructed using these tools. DAST codes are constructed by

combining the rotated constellations and Hadamard transforms of appropriate dimensions. We will describe here the construction and properties of these codes.

2.4.1 Construction and Properties

Let $\boldsymbol{M}_n$ be a rotation of dimension $n \times n$ and $\boldsymbol{\mathcal{H}}_n$ be the Hadamard transform of dimension $n = 2^q$, then DAST block code $\boldsymbol{\Xi}_n$, of dimension n is constructed by :

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{M}_n \boldsymbol{a} = [x_1 \; x_2 \; ... \; x_n]^T \\ \boldsymbol{\Xi}_n &= \begin{bmatrix} x_1\hbar_{11} & x_2\hbar_{12} & \cdots & x_n\hbar_{1n} \\ x_1\hbar_{21} & x_2\hbar_{22} & \cdots & x_n\hbar_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_1\hbar_{n1} & x_2\hbar_{n2} & \cdots & x_n\hbar_{nn} \end{bmatrix} \end{aligned} \tag{2.10}$$

Where $\mathbf{a} = [a_1 \; a_2 \; ... \; a_n]^T$ denotes the information vector and $\hbar_{ij}$ corresponds to the element (i, j) of the Hadamard transform $\boldsymbol{\mathcal{H}}_n$. The proposed DAST code spreads each component x_j of the rotated vector $\boldsymbol{x}$ by the Hadamard transform $\boldsymbol{\mathcal{H}}_n$. The spread component is then sent over n antennas at time instant j.

Diversity gain : The proposed block ST code has maximum transmit diversity gain of n for n transmit antenna [7]. We limit our designs to the case where $n = 2^q$. Using such constructed ST block codes with properly chosen algebraic rotations $\boldsymbol{M}_n$, we obtain the maximum achievable diversity order nm in the system when rotated constellation is used.

[7] See [27] for the proof.

Coding gain : The coding gain of DAST code can be obtained [8] by

$$\begin{aligned}\min_{\boldsymbol{y}\neq 0} \det\left(\boldsymbol{\Xi}_n^{\dagger}\boldsymbol{\Xi}_n\right) &= \min \det\left(\mathrm{diag}(y_1,y_2,...,y_n)^{\dagger}\boldsymbol{\mathcal{H}}_n^{\dagger}\boldsymbol{\mathcal{H}}_n\mathrm{diag}(y_1,y_2,...,y_n)\right)\\ &= \min \det\left(\boldsymbol{\mathcal{H}}_n^{\dagger}\boldsymbol{\mathcal{H}}_n \ \mathrm{diag}(|y_1|^2,|y_2|^2,...,|y_n|^2)\right)\\ &= \min \ n^n\prod_{j=1}^{n}|y_j|^2\end{aligned} \tag{2.11}$$

Having $\boldsymbol{\mathcal{H}}_n^{\dagger}\boldsymbol{\mathcal{H}}_n = n\boldsymbol{I}_n$ in mind, taking the minimum of n^{th} root of the above equation, yield to the coding gain :

$$\delta_n = n(d_{n,min})^{2/n}$$

where $d_{n,min}$ denotes the minimum product distance.

It is proved in [27] that the DAST block code $\boldsymbol{\Xi}_n$ of dimension n has a maximum coding gain over the ST block codes formed by the conjunction of the rotated constellation and linear transformations with entries from $\{+1,-1\}$. In the case of DAST codes with $n = 2^q$ there is only n orthogonal sequences with entries $\{+1,-1\}$ of length n. This makes the Hadamard transform the unique transform maximizing the coding gain. We can see later that replacing this transform by other matrices with entries coming from $\{+1,-1\}$ which are not necessary of length n permits us to construct other ST block codes.

Achievable data rate : Data rate of the DAST codes constructed here is 1 sym/sec/Hz.

Remark : The ST block codes designed here have full data rate of 1 symbol per time slot, and they achieve full modulation diversity nm. If necessary, we can also construct ST codes with higher data rates using the aforementioned tools. Decoding procedure of ST block codes constructed above is made by SD algorithm with moderate complexity 2.3.2.

2.5 Comparisons

The following table allows us to compare DAST codes and ST codes from Orthogonal designs.

[8] By definition, coding gain is computed over all distinct codewords $\boldsymbol{x}$ and $\boldsymbol{e}$, with $\boldsymbol{y} = \boldsymbol{x}-\boldsymbol{e} = \boldsymbol{M}_n(\boldsymbol{a}-\boldsymbol{b})$.

	DAST codes	Orthogonal codes
Transmit antennas	$n = 2^q$	$n = 2, 4$ and 8
Diversity gain	full diversity for n	full diversity for n
Coding gain	good coding gain regarding n	less than DAST codes
Decoding complexity	ML decoding with $O(n^3)$	linear ML decoding
Data rate	full rate for both real and complex const.	full rate for real, half rate for complex const.
System Performance	achieve better performance for $n > 2$	–

Table 2.1: Diagonal Algebraic ST codes versus ST codes from Orthogonal designs

More details can be found in [27]. However it is shown in [27] that for the same spectral efficiency, DAST block codes overcome ST codes from orthogonal design in term of performance for $n > 2$.

2.6 Conclusions

In this chapter we have provided a brief review on the advantages and targets of Space-Time codes. We have also introduced two well known STBC schemes called Space-Time codes from orthogonal design and DAST codes. Both of these schemes are compared in term of diversity and coding gain, decoding complexity, data rate and achievable performance. We have also proved that the DAST codes is a good performing coding scheme that achieves maximum diversity and a sufficiently large-valued coding gain in a MIMO context.

Other forms of DAST block codes can be obtained, replacing the Hadamard transform by any other transform with $\{+1, -1\}$ entries, even if the transform matrix is not square. This idea is used in the next chapter to introduce ST codes for Coded Division Multiple Access (CDMA) context. Motivated by the same idea, similar constructions can be obtained depending on the desired architecture.

Independent of ST block codes there are some topics that can be explored separately. For example, turbo equalization with algebraic Space-Time codes seems to be another interesting topic to be investigated. The need of a unified model for fading channel which takes into account different propagation parameters is also important in practice.

In this document, the focus is made on the physical layer aspects of multi-antenna systems. However, appropriate protocols, if possible, might be developed in order to support the data flow obtained in the physical layer.

Chapter 3

Transmit diversity scheme for CDMA system

In this chapter we propose a new transmit diversity scheme for multi path Fading channels [1]. At the beginning, in section 3.2, we highlight the need for transmit diversity in the CDMA context. Then in 3.3, we give a brief overview on existing schemes and their limitations. In section 3.4 we deal with a new approach which consists of using Space-Time codes on CDMA system. In Section 3.4.1 we present the system model. Section 3.4.2 treats the single user system; we give the ST codes construction, their properties and the decoding algorithm. The generalization to multiuser systems is given in Section 3.4.4. Multi path fading channel scenario is treated in Section 3.4.5. In Section 3.4.6 we tackle the problem of blind channel estimation in a multi-antenna quasi-static fading environment. Simulation results are given in Section 3.4.7. In Section 5.8 we discuss the obtained results.

3.1 Introduction to CDMA system

In this section we will give a brief review of CDMA system in order to highlight the main properties of these schemes that are used or mentioned later in this chapter. However, a large number of references on CDMA system exist. We can mention [2] and [44, 97] for further reading.

[1] The results concerning this work have already been published in : M. O.Damen, A. Safavi and K.Abed-Meraim, " On CDMA with Space-Time Codes over Multi-path Fading channel ", *IEEE Transactions on Wireless Communications*, Jan. 2003, pp. 11-19.

Access schemes For radio systems there are two resources, frequency and time. Division by frequency, so that each pair of communicators is allocated part of the spectrum for all of the time, results in Frequency Division Multiple Access (FDMA). Division by time, so that each pair of communicators is allocated all (or at least a large part) of the spectrum for part of the time results in Time Division Multiple Access (TDMA). In Code Division Multiple Access (CDMA), every communicator will be allocated the entire spectrum all of the time. CDMA uses codes to identify connections.

Spreading codes CDMA uses unique spreading codes to spread the baseband data before transmission. The signal is transmitted in a channel, which is below noise level. The receiver then uses a correlator to despread the wanted signal, which is passed through a narrow bandpass filter. Unwanted signals will not be despread and will not pass through the filter. Codes take the form of a carefully designed one/zero sequence produced at a much higher rate than that of the baseband data. The rate of a spreading code is referred to as chip rate rather than bit rate.

CDMA codes are not required to provide call security, but create a uniqueness to enable call identification. Codes should not correlate to other codes or time shifted version of itself. Spreading codes are noise like pseudo-random codes, channel codes are designed for maximum separation from each other and cell identification codes are balanced not to correlate to other codes of itself.

Spreading process CDMA uses Direct Sequence spreading, where spreading process is done by directly combining the baseband information to high chip rate binary code.

Multi path and RAKE receivers One of the main advantages of CDMA systems is the capability of using signals that arrive in the receivers with different time delays. This phenomenon is called multipath. FDMA and TDMA, which are narrow band systems, cannot discriminate between the multipath arrivals, and resort to equalization to mitigate the negative effects of multipath. Due to its wide bandwidth and rake receivers, CDMA uses the multipath signals and combines them to make an even stronger signal at the receivers. CDMA subscriber units use rake receivers. This is essentially a set of several receivers. One of the receivers (fingers) constantly searches for different multipaths and

feeds the information to the other three fingers. Each finger then demodulates the signal corresponding to a strong multipath. The results are then combined together to make the signal stronger.

3.2 Why using transmit diversity in CDMA ?

Applications and services described in 3-rd Generation (3G) wireless systems (based on CDMA) require significantly higher capacities than what is realized today. Multiple receive antennas can be used at the user terminal to increase the capacity of the CDMA system. Exploiting receive diversity in CDMA receivers is a well known topic [97]. However, since 98 there was almost no proposition exploiting transmit diversity in CDMA.

There are some scenarios where the receive diversity can not be implemented or it is not useful. Because for certain applications, like Internet, data flow is not balanced. This is the case when in the system down link, a considerable flow of data is sent to the user under a simple request.

Moreover, some practical considerations like as physical implementation of several antennas or complex processing units on the small handset is a big problem.

On the other hand, it is well known [91] that MIMO configuration, if it is well exploited, provide substantial improvement of the capacity.

Consequently, one possible solution to these limitations would be to introduce antennas at the transmitter side to obtain also transmit diversity. At the transmitter side (base station) it is possible to implement complex processing units and there is not space-limitation in order to implement several antennas. Moreover, the coding and diversity advantage obtained using transmit diversity is more than that of receive diversity.

3.3 Overview of related works

Motivated by previously mentioned factors, a lot of research efforts have been made in order to introduce transmit diversity in the CDMA communication systems. Here, we focus on transmit diversity techniques where the propagation channel is not known at the transmitter. These techniques are known as open loop transmit diversity techniques. We can mention several propositions:

Delay-Transmit-Diversity (DTD) In this proposition, the base station transmits the signal and the delayed version of the signal over two transmit antennas [103]. To the author's knowledge it is the first proposition attempting to combine multiple transmit antennas with CDMA system. This scheme exploits the full spatial diversity, but the sources are not allocated properly. Otherwise, the bandwidth and the number of spreading sequences used are two times more than a simple scheme when the transmit diversity is not used. Moreover, interference level due to the transmission scheme can not be neglected.

Orthogonal Transmit Diversity (OTD) The base station transmits orthogonally spread signals over multiple antennas [78]. This scheme provides large improvement over DTD scheme and it has been adopted by third generation CDMA systems in the U.S.A. [78]. In this scheme no-extra sources are needed in the transmitter side.

Space-Time Block Coded transmit diversity (STBC) This scheme is also known as the Alamouti scheme [10]. In this scheme two symbols are sent over two transmit antennas during two time slots. See [10] for more details and ML decoding procedure. This scheme is also included in the UMTS standards [3]. In this scheme spatial and temporal diversity advantage inherent in the system is well exploited. Moreover, a simple decoding procedure has been proposed for this scheme. But, extension of this scheme to more than two transmit antennas is not straightforward.

Space-Time Spreading (STS) Space-Time Spreading is proposed in [54]. Contrary to the STBC, each user's data is spread in a different fashion on each transmit antenna. STS schemes proposed for two transmit antennas and STBC have the same diversity advantage. This scheme is included in the IS-2000 standards [4]. It has been also extended to the case of more than two transmit antennas [56]. However, there are limitation on these schemes. When using more than two transmit antennas, the number of spreading sequences is larger than the number of transmit antennas.

3.4 CDMA with Space-Time codes : A new approach

Here, we propose a different approach which exploits the benefits of Space-Time codes, proposed in [90], over CDMA system. We include an appropriate coding

procedure at the transmitter side in order to exploit maximum possible diversity and coding gain available in the system[2]. More precisely, we consider a CDMA system where each user has several antennas and by a proper combination of different Spreading Sequences (SS) and full spatial diversity rotated constellations, the corresponding user satisfies the construction criteria of ST codes [90] and achieves full transmit diversity. The receiver has one or possibly several receive antennas and uses a linear multiuser detector based on the combination of linear decorrelation with respect to all users, and the application of the sphere decoder [29] to decode each user separately. Moreover, in a multi path fading channel scenario we present a method of blind channel estimation based on subspace decomposition.

3.4.1 System model

We consider a system of K users transmitting simultaneously and synchronously. Each user has p transmit antennas and uses p spreading sequences of length N; in the sequel we only consider $n \stackrel{\triangle}{=} Kp \leq N$ [3]. We assume that all the SS are linearly independent, and denoted by:

$$\boldsymbol{s}_j \quad = \quad [s_j(1), \ldots, s_j(N)]^T, \quad j = 1, \ldots, n. \tag{3.1}$$

where the superscript ^{T} denotes the transpose and $s_j(t)$, $t = 1, \ldots, N$, denotes the spreading coefficient at time t. The user i uses the sequences $\boldsymbol{s}_{(i-1)p+v}$, $i = 1, \ldots, K$, $v = 1, \ldots, p$. We also suppose that the receiver has m receiver antennas. The transmission is done by bursts of length l over a quasi-static fading channel. Between each transmitter antenna j and receiver antenna k there are L distinct paths, where the channel coefficients of each path are denoted here by h_{kj}^q, for $q = 1, \ldots, L$. In other words, $h_{kj}(t) = \sum_{q=1}^{L} h_{kj}^q \delta(t - \tau_{kj}^q)$, where τ_{kj}^q denotes the time delay for the q-th path. The antennas are sufficiently spaced so that h_{kj}^q are assumed to be mutually uncorrelated fading coefficients.

[2]Contrary to the previous schemes where the available inherent diversity in the system is exploited.

[3]In the case where multiple receivers are available, this limiting condition becomes $Kp \leq mN$ (where m is the number of receivers) [20].

3.4.2 Single-user single-path

3.4.2.1 Coding algorithm

Consider a single user multi-antenna CDMA scenario where the user wants to transmit the information symbols $a_1, a_2, \ldots, a_p$ by using the spreading sequences $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p$ over p transmit antennas. We propose the following scheme. First, we rotate the information symbol vector $\boldsymbol{a} = [a_1, \ldots, a_p]^T$ by the rotation $\boldsymbol{M}_p$ to obtain $\boldsymbol{x} = [x_1, \ldots, x_p]^T = \boldsymbol{M}_p\boldsymbol{a}$. Then we spread the symbol x_j by the SS $\boldsymbol{s}_j$ and send it over the transmit antenna $j = 1, \ldots, p$. We can write the proposed ST code at the word $\boldsymbol{x}$ as a $p \times N$ matrix $\boldsymbol{B}(\boldsymbol{x})$ with entries $b_{jt}, j = 1 \ldots p, t = 1 \ldots N$, where b_{jt} is sent at time t over the antenna j

$$\begin{aligned} \boldsymbol{B}(\boldsymbol{x}) &= \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x_p \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1 & \ldots & \boldsymbol{s}_p \end{bmatrix}^T \\ &= \operatorname{diag}(x_1, \ldots, x_p)\boldsymbol{S}^T. \end{aligned} \tag{3.2}$$

3.4.2.2 Example

Consider a single user system with two transmit antennas using the following SS of length 4 ($K = 1, p = 2, N = 4$)

$$\boldsymbol{S}^T = \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

The optimal 2×2 real rotation which maximizes $d_{2,min}$, $\boldsymbol{M}_2$ is given by [26]:

$$\boldsymbol{M}_2 = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \text{ with } \theta = 1.0172 \,\text{radians}.$$

Let $\boldsymbol{a} = [a_1, a_2]^T$ be the information symbol vector, then over two transmit antennas and 4 chip periods we transmit:

$$\boldsymbol{B}(\boldsymbol{x}) = \frac{1}{2}\begin{bmatrix} x_1 & -x_1 & x_1 & -x_1 \\ x_2 & x_2 & -x_2 & -x_2 \end{bmatrix}$$

with

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{M}_2 \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta)a_1 + sin(\theta)a_2 \\ -\sin(\theta)a_1 + \cos(\theta)a_2 \end{bmatrix}.$$

The i-th column of $\boldsymbol{B}(\boldsymbol{x})$ represents the transmitted signals at antenna 1 (first row) and antenna 2 (second row) at the i-th chip period.

3.4.2.3 Properties

Proposition 1 *The proposed ST code in (3.2) over p transmit antennas and N periods of time has a transmit diversity gain of p.*

Proof. Let $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{e} = \boldsymbol{M}_p(\boldsymbol{a} - \boldsymbol{b})$ such that $\boldsymbol{a} \neq \boldsymbol{b}$. Consider the ST code $\boldsymbol{B}(\boldsymbol{y})$ at the word $\boldsymbol{y}$. Since $\boldsymbol{M}_p$ generates a full spatial diversity constellation, one has $y_j \neq 0 \ \forall \ j = 1, \ldots, p$ and $\forall \boldsymbol{a} \neq \boldsymbol{b}$ in the considered constellation. It follows that the matrix $\mathrm{diag}(y_1, \ldots, y_p)$ has a rank p and thus the ST code matrix $\boldsymbol{B}$ has a rank p over all the distinct codewords since the SS are linearly independent. This implies that the ST code has a transmit diversity gain of p. □

Note that the coding gain depends on the cross-correlation of the SS, and it is equal to [90]:

$$\begin{aligned} \delta_p &\triangleq \min_{\boldsymbol{x} \neq \boldsymbol{e}} \left(\det\left(\boldsymbol{B}(\boldsymbol{y})\boldsymbol{B}^H(\boldsymbol{y})\right)\right)^{1/p} \\ &= \left(\det(\boldsymbol{S}^T\boldsymbol{S}) \min_{\boldsymbol{a} \neq \boldsymbol{b}} \prod_{j=1}^{p} |y_j|^2\right)^{1/p} \\ &= \left(\det(\boldsymbol{S}^T\boldsymbol{S})\right)^{1/p} (d_{p,min})^{2/p} \end{aligned} \tag{3.3}$$

when the SS are assumed to take real values, which is the case in this work. The data rate of the proposed ST code is p/N symbol/T_c , where T_c denotes the duration of a chip. During N chip periods the received signal at the k-th receiver antenna, $k = 1 \ldots m$, is given by

$$\boldsymbol{r}_k = \boldsymbol{S}\mathrm{diag}(\boldsymbol{h}_k)\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu}_k \tag{3.4}$$

where $\mathrm{diag}(\boldsymbol{h}_k) = \mathrm{diag}(h_{k1}, \ldots, h_{kp})$ is a diagonal matrix which has the fading coefficients of the k-th receiver antenna on its diagonal. The $N \times 1$ vector $\boldsymbol{\nu}_k$ represents the additive white Gaussian noise (AWGN), and it is assumed to have independent components with variance σ^2 per dimension, i.e., $E[\boldsymbol{\nu}_k\boldsymbol{\nu}_k^H] = 2\sigma^2\boldsymbol{I}_N$, with $\boldsymbol{I}_N$ the $N \times N$ identity matrix.

3.4.3 Decoding algorithm

We suppose here that the channel is known at the receiver. Channel estimation is addressed in Section 3.4.6. Let $\boldsymbol{S}^{\#} \triangleq (\boldsymbol{S}^T\boldsymbol{S})^{-1}\boldsymbol{S}^T$ denote the $p \times N$ pseudo-inverse matrix of $\boldsymbol{S}$ [47]. The decoding is performed as follows.

1. First, we multiply the received signal by $\boldsymbol{S}^{\#}$ at each receiver antenna

$$\boldsymbol{r}'_k = \boldsymbol{S}^{\#}\boldsymbol{r}_k = \mathrm{diag}(\boldsymbol{h}_k)\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu}'_k \tag{3.5}$$

where $\boldsymbol{\nu}'_k$ is a colored Gaussian noise with covariance matrix equal to $2\sigma^2\boldsymbol{S}^{\#}\boldsymbol{S}^{\#T}$. When the SS have good cross-correlation properties, $\boldsymbol{\nu}'_k$ can be approximated by a white Gaussian process. Otherwise, the signal $\boldsymbol{r}'_k$ needs to be whitened.

2. Next, we perform a channel phase inversion[4] at each receiver antenna, which can be expressed as

$$\begin{aligned}\boldsymbol{r}''_k &= \mathrm{diag}\left(e^{-i\phi_{k1}},\ldots,e^{-i\phi_{kp}}\right)\boldsymbol{r}'_k \\ &= \mathrm{diag}(|h_{k1}|,\ldots,|h_{kp}|)\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu}''_k,\end{aligned} \tag{3.6}$$

where $h_{kj} = |h_{kj}|e^{i\phi_{kj}}$, with $i = \sqrt{-1}$.

3. Finally, we collect the signals on all the receiver antennas

$$\begin{aligned}\boldsymbol{r} &= \left[\boldsymbol{r}''^{T}_1,\ldots,\boldsymbol{r}''^{\,T}_m\right]^T = \begin{bmatrix}\mathrm{diag}(|h_{11}|,\ldots,|h_{1p}|)\\ \vdots \\ \mathrm{diag}(|h_{m1}|,\ldots,|h_{mp}|)\end{bmatrix}\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu} \\ &= \boldsymbol{\mathcal{H}}\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu}\end{aligned} \tag{3.7}$$

where $\boldsymbol{\mathcal{H}}$ is an $mp \times p$ real-value matrix of rank p. In order to take advantage of all the available information on all the receiver antennas we propose the following detection procedure based on singular value decomposition (SVD) followed by sphere decoding [29]:

- $[\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V}] = \mathrm{SVD}(\boldsymbol{\mathcal{H}})$: $\boldsymbol{\mathcal{H}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, where $\boldsymbol{\Sigma}$ is a $p \times p$ diagonal matrix, $\boldsymbol{U}$ is an $mp \times p$ matrix and $\boldsymbol{V}$ is a $p \times p$ matrix with $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_p$.
- $\boldsymbol{r}' \longleftarrow \boldsymbol{U}^T\boldsymbol{r} = \boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu}'$, where $\boldsymbol{\nu}' = \boldsymbol{U}^T\boldsymbol{\nu}$.
- Apply the sphere decoder algorithm on the real and imaginary parts of $\boldsymbol{r}'$, in order to obtain transmitted symbol vector $\boldsymbol{a}$.

[4] Even in blind channel identification, it is assumed that either the channel phase is known or estimated by pilot symbols.

The sphere decoder algorithm reaches near-ML decoding performance with a complexity approximated by $O(p^6)$ arithmetical operations[5] [29], in the sense that it maximizes the least squares (LS) criterion

$$\min_{\boldsymbol{a}\in\mathcal{A}} \left\| \boldsymbol{r}' - \boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{M}_p\boldsymbol{a} \right\|^2 = \min_{\boldsymbol{a}\in\mathcal{A}} \left\| \boldsymbol{r} - \boldsymbol{\mathcal{H}}\boldsymbol{M}_p\boldsymbol{a} \right\|^2$$

where $\mathcal{A}$ denotes the set of all symbol vector values. This criterion coincides with the ML criterion when the additive noise is white Gaussian [74].

A suboptimal approach that avoids the computation of the SVD of $\mathcal{H}$ consists of using a coherent receiver according to:

- Compute the $p \times 1$ vector $\tilde{\boldsymbol{r}}$ as:

$$\tilde{\boldsymbol{r}} = \sum_{i=1}^{m} \boldsymbol{r}_i'' = \text{diag}\left(\sum_{i=1}^{m} |h_{i1}|, \ldots, \sum_{i=1}^{m} |h_{ip}| \right) \boldsymbol{M}_p\boldsymbol{a} + \tilde{\boldsymbol{\nu}}$$

  where $\tilde{\boldsymbol{\nu}}$ is a $p \times 1$ noise vector of covariance matrix $E[\tilde{\boldsymbol{\nu}}\tilde{\boldsymbol{\nu}}^H] = 2\sigma^2 m\boldsymbol{I}_p$.

- Apply the sphere decoder algorithm on the real and imaginary parts of $\tilde{\boldsymbol{r}}$, in order to obtain transmitted symbol vector $\boldsymbol{a}$.

3.4.4 Generalization to multi-user

Now we consider K users transmitting simultaneously and synchronously so that the user i uses the ST code expressed by (3.2) with the SS given by $\boldsymbol{s}_{(i-1)p+v}$, $i = 1, \ldots, K$ and $v = 1, \ldots, p$. All users use the same rotation matrix $\boldsymbol{M}_p$, and their information symbols are supposed to be independent and identically distributed. At each receiver antenna $k = 1, \ldots, m$, the received signal is given by

$$\boldsymbol{r}_k \;=\; [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n]\,\text{diag}(h_{k1}, \ldots, h_{kn})\boldsymbol{M}\boldsymbol{a} + \boldsymbol{\nu}_k \tag{3.8}$$

where $n \triangleq Kp$, and $\boldsymbol{a} = [a_1, \ldots, a_n]^T$ so that the user i transmits the information symbols $a_{(i-1)p+v}$, $v = 1, \ldots, p$. The $n \times n$ block diagonal matrix $\boldsymbol{M}$ has the matrix $\boldsymbol{M}_p$ on the $p \times p$ principal diagonal blocks and zeros elsewhere, i.e.,

[5] This complexity represents a rather "pessimistic" bound on the worst case complexity of the sphere decoder algorithm. Recent results show that the efficient implementation of the sphere decoder allows for much less complexity (roughly cubic in p), especially at moderate and high SNR [28].

$\boldsymbol{M} \triangleq \boldsymbol{I}_p \otimes \boldsymbol{M}_p$, with $\otimes$ denoting the Kronecker product.

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_p & \mathbf{0}_p & \dots & \mathbf{0}_p \\ \mathbf{0}_p & \boldsymbol{M}_p & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0}_p \\ \mathbf{0}_p & \mathbf{0}_p & \dots & \boldsymbol{M}_p \end{bmatrix}, \tag{3.9}$$

with $\mathbf{0}_p$ is a $p \times p$ matrix with all the entries equal to 0.

Comparing (3.8) and (3.4), one can represent the proposed multiuser ST-CDMA scheme as a single user with $n \triangleq Kp$ transmit antennas with a transmission rate of n/N symbol/Tc.

Proposition 2 *The proposed multiuser ST-CDMA scheme has a transmit diversity gain of p.*

Proof. This is a direct result when one represents the multiuser scheme by the single user ST code $\boldsymbol{B}(\boldsymbol{x})$ in (3.2) with $\boldsymbol{S} = [\boldsymbol{s}_1, \dots, \boldsymbol{s}_n]$ and $\boldsymbol{x} = \boldsymbol{M}\boldsymbol{a}$. □

One can also prove that the multiuser scheme has the same coding gain as that of the single user ST code (3.3) with the smallest determinant of the cross-correlation matrix. We have the additional property that the information symbols $\boldsymbol{a}$ are encoded by the matrix $\boldsymbol{M}$ separately by blocks of size p, which we exploit in order to simplify the decoding algorithm. Processing the received signal over m receive antennas as in subsection 3.4.3, yields

$$\boldsymbol{r} = \boldsymbol{\mathcal{H}}\boldsymbol{M}\boldsymbol{a} + \boldsymbol{\nu}$$

where $\boldsymbol{\nu}$ is an $mn \times 1$ complex Gaussian noise vector having the covariance matrix $2\sigma^2 \boldsymbol{S}^{\#}\boldsymbol{S}^{\#^T}$, and $\boldsymbol{\mathcal{H}}$ is an $mn \times n$ matrix given by

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathrm{diag}(|h_{11}|, \dots, |h_{1n}|) \\ \vdots \\ \mathrm{diag}(|h_{m1}|, \dots, |h_{mn}|) \end{bmatrix}. \tag{3.10}$$

Since $\boldsymbol{M}$ and $\boldsymbol{\mathcal{H}}$ are block diagonal, we can separate the different users; user i has its received symbols contained in $\boldsymbol{r}^i = \boldsymbol{r}_{(j-1)n+(i-1)p+v}$, $v = 1, \dots, p$, $j = 1, \dots, m$.

$$\boldsymbol{r}^i = \boldsymbol{\mathcal{H}}^i \boldsymbol{M}_p \boldsymbol{a}_i + \boldsymbol{\nu}^i \tag{3.11}$$

with $\boldsymbol{a}_i = [a_{(i-1)p+1}, \ldots, a_{ip}]^T$, and $\boldsymbol{\nu}^i$ is an $mp \times 1$ column vector contained in $\boldsymbol{\nu}$. $\boldsymbol{\mathcal{H}}^i$ is an $mp \times p$ matrix extracted from $\boldsymbol{\mathcal{H}}$ such that

$$\boldsymbol{\mathcal{H}}^i = \begin{bmatrix} \mathrm{diag}(|h_{1,(i-1)p+1}|, \ldots, |h_{1,ip}|) \\ \vdots \\ \mathrm{diag}(|h_{m,(i-1)p+1}|, \ldots, |h_{m,ip}|) \end{bmatrix}. \tag{3.12}$$

Therefore, the sphere decoder algorithm can be applied separately on each user in order to detect the transmitted information by the K users.

Remark 1: In most practical systems we have $N \gg p$ and $N \gg m$, so that the computational cost can be approximated by $O(N^2 n) + O(2Kp^6)$. This cost is dominated by $O(N^2 n)$ which is the cost of computing the decorrelation matrix $\boldsymbol{S}^{\#}$.

Remark 2: When choosing the SS with minimum cross-correlation performing the sphere decoder without whitening the noise yields small degradation in performance. In that case the noise covariance matrix after decorrelating the received signal is equal to $2\sigma^2 \boldsymbol{S}^{\#} \boldsymbol{S}^{\#^T} \approx 2\sigma^2 \boldsymbol{I}_n$ when the SS are almost orthogonal. This is especially true when $N \gg n$.

3.4.5 Generalization to multi-path fading channels

For simplicity's sake we only consider a single user system in this section and section 3.4.6. As shown in Section 3.4.4, we can always represent a multiuser system by a single user system, where the information symbols are encoded separately by blocks of size p. When sending the ST code in (3.2) over a multipath fading channel of L paths, with $L \ll N$, the inter-symbol interference (ISI) is neglected [64]. Therefore, the received signal at the k-th receiver antenna can be approximated as

$$\begin{aligned} \boldsymbol{r}_k &= \left[\tilde{\boldsymbol{S}}_1, \ldots, \tilde{\boldsymbol{S}}_p\right] \tilde{\boldsymbol{H}}_k \boldsymbol{M}_p \boldsymbol{a} + \boldsymbol{\nu}_k \\ &= \tilde{\boldsymbol{\mathcal{A}}}_k \boldsymbol{M}_p \boldsymbol{a} + \boldsymbol{\nu}_k \end{aligned} \tag{3.13}$$

where $\boldsymbol{\nu}_k$ is an $N\times 1$ AWGN with variance σ^2 per dimension, and $\tilde{\boldsymbol{H}}_k$ is an $Lp\times p$ matrix which represents the channel fading coefficients at the k-th receiver

$$\tilde{\boldsymbol{H}}_k \triangleq \begin{bmatrix} h_{k1}^1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h_{k1}^L & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & 0 & h_{kp}^1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & 0 & h_{kp}^L \end{bmatrix}. \tag{3.14}$$

The modified SS matrix is

$$\tilde{\boldsymbol{S}}_j = \begin{bmatrix} \boldsymbol{s}_j(1) & 0 & \dots & 0 \\ \boldsymbol{s}_j(2) & \boldsymbol{s}_j(1) & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \boldsymbol{s}_j(L) & \boldsymbol{s}_j(L-1) & \dots & \boldsymbol{s}_j(1) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{s}_j(N) & \boldsymbol{s}_j(N-1) & \dots & \boldsymbol{s}_j(N-L+1) \end{bmatrix}, \quad N\times L. \tag{3.15}$$

Processing the received signal over m receiver antennas yields

$$\boldsymbol{r} = \tilde{\boldsymbol{\mathcal{A}}}\boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu} \tag{3.16}$$

where $\tilde{\boldsymbol{\mathcal{A}}} = [\tilde{\boldsymbol{\mathcal{A}}}_1{}^T, \cdots, \tilde{\boldsymbol{\mathcal{A}}}_m{}^T]^T$. Inverting the $mN\times p$ complex matrix $\tilde{\boldsymbol{\mathcal{A}}}$ gives

$$\boldsymbol{r}' = \tilde{\boldsymbol{\mathcal{A}}}^{\#}\boldsymbol{r} = \boldsymbol{M}_p\boldsymbol{a} + \boldsymbol{\nu}'. \tag{3.17}$$

This matrix inversion results in a performance loss when compared to ML decoding performance, which depends on the conditioning of $\tilde{\boldsymbol{\mathcal{A}}}$. Nevertheless, in the context of a multiuser system, the matrix inversion of $\tilde{\boldsymbol{\mathcal{A}}}$ allows us to separately perform the sphere decoder on each user which reduces the computational cost from[6] $O((2Kp)^6)$ to $O(2Kp^6)$. The noise term $\boldsymbol{\nu}'$ is now a colored Gaussian noise with a covariance matrix $\boldsymbol{\Sigma} = \sigma^2\tilde{\boldsymbol{\mathcal{A}}}^{\#}\tilde{\boldsymbol{\mathcal{A}}}^{\#^H}$. Under this assumption, the ML criterion for estimating $\boldsymbol{a}$ leads to [73]

$$\left\|\boldsymbol{r}' - \boldsymbol{M}_p\boldsymbol{a}\right\|_{\boldsymbol{\Sigma}}^2 \triangleq (\boldsymbol{r}' - \boldsymbol{M}_p\boldsymbol{a})^H\ \boldsymbol{\Sigma}^{-1}\ (\boldsymbol{r}' - \boldsymbol{M}_p\boldsymbol{a}).$$

[6]Even at moderate and high SNR where the complexity of the sphere decoder can be cubic in the lattice dimension [28], it is useful to decorrelate the users first then to perform the sphere decoder because $O((2Kp)^3)$ can be much greater than $O(2Kp^3)$ for large K.

This is equivalent to first whitening the observation using the inverse square root matrix of $\mathbf{\Sigma}$, $\boldsymbol{r}'' = \mathbf{\Sigma}^{-1/2}\boldsymbol{r}'$, then computing $\min_{\boldsymbol{a}\in\mathcal{A}} \left\| \boldsymbol{r}'' - \mathbf{\Sigma}^{-1/2}\boldsymbol{M}_p\boldsymbol{a} \right\|^2$. Since we apply the sphere decoder algorithm on real-value vectors, we perform the whitening separately on the real and imaginary parts of $\boldsymbol{r}'$ as follows

$$\begin{aligned} \Re(\boldsymbol{r}') &= \boldsymbol{M}_p\Re(\boldsymbol{a}) + \Re(\boldsymbol{\nu}'),\ \Im(\boldsymbol{r}') = \boldsymbol{M}_p\Im(\boldsymbol{a}) + \Im(\boldsymbol{\nu}') \\ \boldsymbol{r}_1'' &\triangleq \mathbf{W}_r\Re(\boldsymbol{r}'), \quad \boldsymbol{r}_2'' \triangleq \mathbf{W}_i\Im(\boldsymbol{r}') \end{aligned}$$

where $\mathbf{W}_r\mathbf{W}_r^T = \mathrm{cov}(\Re(\boldsymbol{\nu}'))^{-1}$ and $\mathbf{W}_i\mathbf{W}_i^T = \mathrm{cov}(\Im(\boldsymbol{\nu}'))^{-1}$, such that $\mathrm{cov}(\cdot)$ denotes the covariance matrix.

We note that our space-time codes (3.2) used in a multiuser and multi path scenario do not explicitly take into account the diversity that can be obtained by exploiting the presence of multi path environment since we suppose that the channel state information (CSI) is not available at the transmitter. However, as is proved in [89], the achieved diversity in a multi path environment is at least equal to the achieved diversity without multi path. The additional gain in diversity in the presence of multi path that can be obtained in our scheme is merely due to the sphere decoder which exploits all the degrees of freedom offered by the channel (see Figs. 3.6, 3.10, and 3.12 in Section 3.4.7).

3.4.6 Blind channel estimation

In previous sections we supposed that the CSI (Channel State Information) is known at the receiver. Here we propose a blind channel identification scheme to estimate the channel parameters. Blind channel identification is based on exploiting the orthogonality between noise and signal subspaces $\boldsymbol{U}_n$ and $\boldsymbol{U}_s$ respectively. $\boldsymbol{U}_n$ and $\boldsymbol{U}_s$ are computed from the covariance matrix of $\boldsymbol{r}$ in (3.16) as follows

$$\begin{aligned} \boldsymbol{R} &\triangleq E[\boldsymbol{r}\boldsymbol{r}^H] = \bar{E}_s\tilde{\boldsymbol{\mathcal{A}}}\tilde{\boldsymbol{\mathcal{A}}}^H + 2\sigma^2\boldsymbol{I}_{mN} \\ &= \boldsymbol{U}\mathbf{\Lambda}\boldsymbol{U}^H = [\boldsymbol{U}_s \ \ \boldsymbol{U}_n] \begin{bmatrix} \mathbf{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_s^H \\ \boldsymbol{U}_n^H \end{bmatrix} \end{aligned} \tag{3.18}$$

where $\bar{E}_s$ is the average energy per symbol of the input signal[7], $\mathbf{\Lambda}_s = \mathrm{diag}(\lambda_s(1), \ldots, \lambda_s(p))$ with $\lambda_s(1) \geq \lambda_s(2) \geq \ \ldots \ \geq \lambda_s(p) > 2\sigma^2$, and $\mathbf{\Lambda}_n = 2\sigma^2\mathbf{I}_{mN-p}$. Letting Range($\boldsymbol{U}$) denotes the subspace generated by the column vectors of $\boldsymbol{U}$, and

[7]The input sequence is assumed to be i.i.d., i.e. $E[\boldsymbol{a}\boldsymbol{a}^H] = \bar{E}_s\boldsymbol{I}_p$.

noting that Range($\boldsymbol{U}_n$) is orthogonal to Range($\boldsymbol{U}_s$) (i.e., $\boldsymbol{U}_n^H \boldsymbol{U}_s = 0$), and that Range($\boldsymbol{U}_s$) = Range($\tilde{\boldsymbol{\mathcal{A}}}$) [20], one has

$$\boldsymbol{U}_n^H \tilde{\boldsymbol{\mathcal{A}}} = 0 \tag{3.19}$$

where $\tilde{\boldsymbol{\mathcal{A}}} \triangleq [\tilde{\boldsymbol{s}}_1, \cdots, \tilde{\boldsymbol{s}}_p]$. This is equivalent to

$$\boldsymbol{U}_n^H \tilde{\boldsymbol{s}}_j = 0, \ j = 1 \dots p. \tag{3.20}$$

The column vectors of $\tilde{\boldsymbol{\mathcal{A}}}$ are linear functions of the unknown channel parameters and can be written as [20]:

$$\tilde{\boldsymbol{s}}_j = (\boldsymbol{I}_m \otimes \boldsymbol{S}_j)\tilde{\boldsymbol{h}}_j$$

with $\tilde{\boldsymbol{h}}_j \triangleq [\tilde{h}_{1j}^T, \cdots, \tilde{h}_{mj}^T]^T$, and $\tilde{h}_{kj} \triangleq [h_{kj}^1, \dots, h_{kj}^L]^T$, $k = 1, \dots, m$. In practice, $\tilde{\boldsymbol{h}}_j$ is uniquely estimated (up to a scalar constant) [101] as the LS solution of (5)

$$\tilde{\boldsymbol{h}}_j = \arg \min_{\|\boldsymbol{h}\|=1} \|\boldsymbol{U}_n^H (\boldsymbol{I}_m \otimes \boldsymbol{S}_j)\boldsymbol{h}\|^2 = \boldsymbol{h}^H \boldsymbol{G}_j \boldsymbol{h} \tag{3.21}$$

where $\boldsymbol{G}_j = (\boldsymbol{I}_m \otimes \boldsymbol{S}_j^H)\boldsymbol{U}_n \boldsymbol{U}_n^H (\boldsymbol{I}_m \otimes \boldsymbol{S}_j)$ is the associated matrix of a positive quadratic form. The solution of (3.21) is given by the unit-norm least eigenvector of $\boldsymbol{G}_j$ (see [68] for more details). For comparison (see Fig. 3.12), we have also considered a two step estimation approach consisting of:

- Blind channel estimation using the above subspace method.
- Symbol detection (using SD or eventually an MMSE detector) then channel re-estimation (in order to refine the channel estimation and improve the performances) using a simple input-output least squares fitting.

3.4.7 Simulation results

In the simulations we use normalized QAM constellations with the average energy per symbol $\bar{E}_s = 1$. The channel matrix is fixed over N chip periods and is modeled as in Section 3.4.1. The received signal is corrupted by a complex AWGN process. The plots are versus the SNR per symbol for the symbol error rate or the SNR per bit for the bit error rate. We have chosen the Gold SS of length $N = 7$ and 31 normalized such that $\boldsymbol{s}_j^T \boldsymbol{s}_j = 1$ [31].

Fig. 3.1 shows the performance of the ST single user codes for $p = 2, 3, 4$, and 6 transmit antennas and one receiver antenna using 4-QAM constellation. The used SS have a length of $N = 7$ and the data rate is $\frac{2p}{N}$ bits/T_c. For

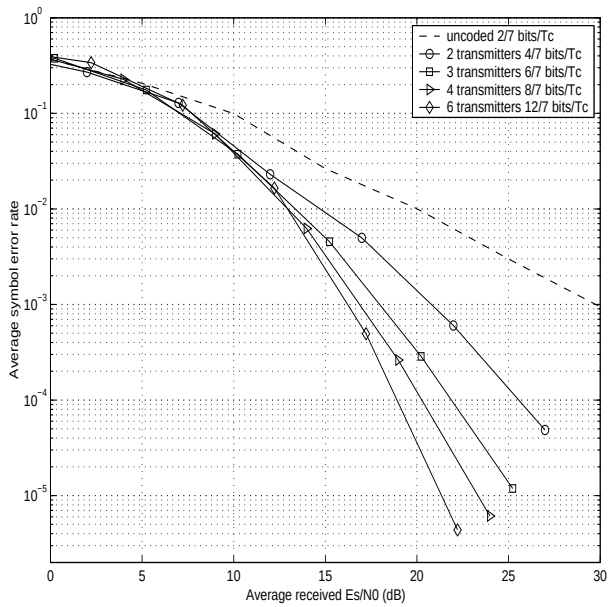

Figure 3.1: Single user, $N = 7$, $p = 2, 3, 4, 6$, $m = 1$, with 4-QAM.

comparison, we also present the performance of the uncoded 4-QAM over a Rayleigh fading channel. It is shown that even though the spectral efficiency increases when increasing p, an improvement in performance is also achieved since the diversity gain also increases with p. For example, at the symbol error probability of 10^{-4}, the ST code with 6 transmit antennas has a gain of 6 dB over the ST code with 2 transmit antennas. Moreover, the spectral efficiency of the former ST code is 3 times more than the latter with no bandwidth increase.

Fig. 3.2 shows the performance of the same scheme as in Fig. 3.1 but when using Gold SS of length $N = 31$. It is shown that the coding and the diversity gain are almost not changed when increasing N.

Fig. 3.3 shows the performance of a ST-CDMA scheme for $K = 1, 2, 3$, $p = 2$, $N = 7$, and $m = 1$ receiver antenna. We note that the proposed detector based on decorrelation and the sphere decoder of each user achieves near-optimum performance. It is shown, for example, that the 3-user system has almost 1 dB loss compared to the single user system.

Fig. 3.4 presents the performance of a 15-user ST-CDMA scheme with $p = 2$, $N = 31$ and $m = 1$ receiver antenna. The total data rate is $\frac{60}{31}$ bits/T_c, and the degradation of the error probability compared to that of the single user is about 1.5 dB.

In Fig. 3.5, we compare a CDMA system using 16-QAM constellation with-

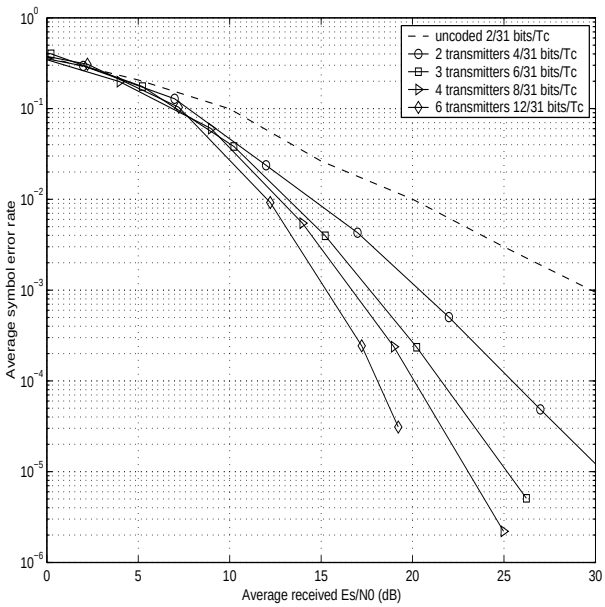

Figure 3.2: Single user, $N = 31$, $p = 2, 3, 4, 6$, $m = 1$, with 4-QAM.

out transmit diversity with two other ST coded systems using transmit diversity and keeping the same data rate ($\frac{4}{31}$ bits/T_c) for all the three systems. One system with two transmit antennas ($p = 2$) using 4-QAM constellation, another with four transmit antennas ($p = 4$) using BPSK modulation. We consider a single user with Gold SS of length $N = 31$ and one receiver antenna. It is shown that a significant gain in terms of bit error rate is obtained using transmit diversity at the same data rate.

Fig. 3.6 illustrates the additional diversity and coding gain obtained in the presence of multi path in a single user scheme with $p = 2$, $m = 1$ and Gold SS of length $N = 31$. The channel between each transmitter antenna and the receiver antenna is modeled by L independent fading coefficients. Different paths from different antennas are assumed to arrive with the same set of delays (see Section 3.4.5). Since $L \ll N$, the ISI can be neglected. Note that a significant performance gain is obtained in this multi path context when increasing the number of paths L.

Figs. 3.7 and 3.8 illustrate the performance of a single user system working at the data rate $\frac{8}{31}$ bits/T_c with two transmitter antennas ($p = 2$) for single path ($L = 1$) and multi path ($L = 3$) respectively. In both figures we use 4-QAM constellation and SS of length 31 and one receiver antenna. The curves denoted by dash-dot line present the system performance when different spreading se-

quences are assigned to different transmitter antennas without coding (i.e. that is equivalent to choosing $\boldsymbol{M}_p = \boldsymbol{I}_p$). The curves denoted by square shows the system performance when using the optimal rotation matrix $\boldsymbol{M}_p$ in conjunction with different spreading sequences. It is shown that a significant gain is obtained when using optimal rotation matrix $\boldsymbol{M}_p$ in both single path and multi path context.

Figs. 3.9 and 3.10 show the performance of a single user system with $p = 1, 2$ transmitter antennas and $m = 1, 2$ receiver antennas for $L = 1$ and $L = 3$, respectively. In both Figs. we use 4-QAM constellation. It is shown in Fig. 3.9, that for a symbol error probability of 10^{-2} in a system using two transmitter antennas, a gain of 5 dB is achieved over an uncoded system using one transmitter antenna. This gain is reduced to 1 dB in the multi path case ($L = 3$) as shown in Fig. 3.10.

In Fig. 3.11 we compare two systems with two receiver antennas with the same data rate of $\frac{4}{31}$ bits/T_c when using Gold SS of length $N = 31$. The first system uses 16-QAM constellation without transmit diversity, whereas the second system uses 4-QAM constellation over two transmit antennas. A gain of almost 6 dB is obtained for the second system at a bit error rate of 10^{-3}.

Finally, Fig. 3.12 shows the performance of the proposed blind channel estimation method in the presence of multi path. Under quasi-static fading assumptions, the channel coefficients are fixed over 100 symbol periods then are changed randomly. We consider a single user system with Gold SS of length $N = 31$, $p = 2$ transmitter antennas, $m = 1$ receiver antenna, $L = 3$ paths per antenna, and 4-QAM constellation. Due to channel estimation errors we note a performance loss when compared to perfect CSI. The performance loss (especially for high SNR) is quite significant and might result in losing the gain of ST code diversity.

3.4.8 Conclusions

We have studied in this chapter ST-CDMA multiuser systems, where we have introduced and analyzed a new method of exploiting the transmit diversity by the combination of spreading sequences and rotated constellations.

We have shown that a linear decorrelator followed by sphere decoding each user separately yields near-optimum performance thanks to the good cross-correlations properties of the used spreading sequences. This also helps us to maintain a good coding gain for the proposed ST codes over multi path fading

channels.

Under quasi-static fading assumptions, we have examined a method of blind channel estimation based on the orthogonality between signal and noise subspaces. It is shown that the performance loss in the system when we use blind channel identification might be quite significant and can be lead to a diversity loss. Therefore, some more reliable identification technique have to be considered.

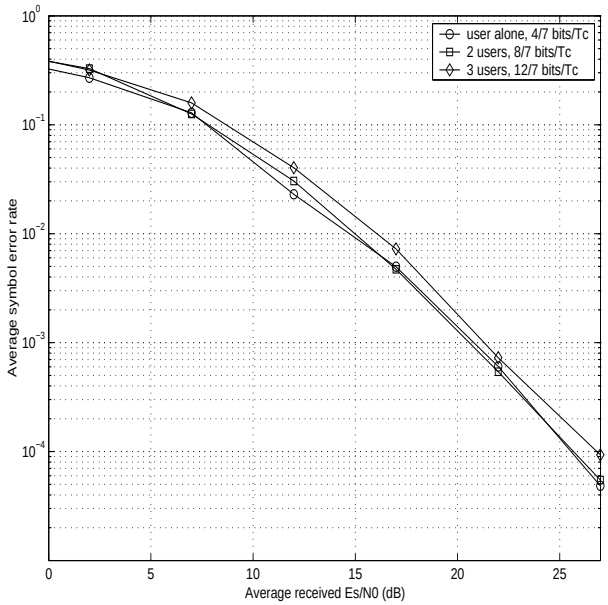

Figure 3.3: Multiuser ST-CDMA, $K = 1, 2, 3$, $N = 7$, $p = 2$, $m = 1$, with 4-QAM.

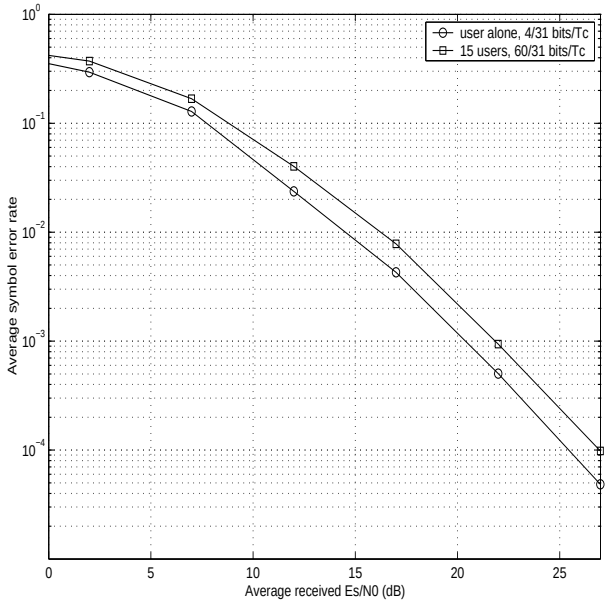

Figure 3.4: Multiuser ST-CDMA, $K = 1, 15$, $N = 31$, $p = 2$, $m = 1$, with 4-QAM.

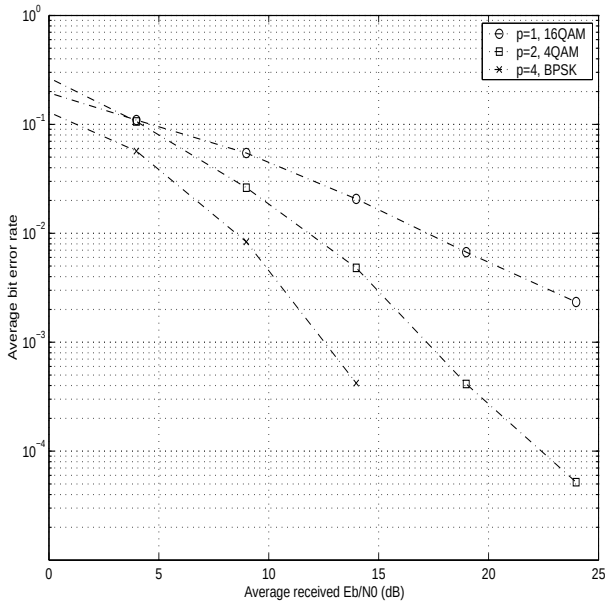

Figure 3.5: Single user, Single path, $N = 31$, $p = 1, 2, 4$, $m = 1$, data rate of $4/31$ bits/T_c.

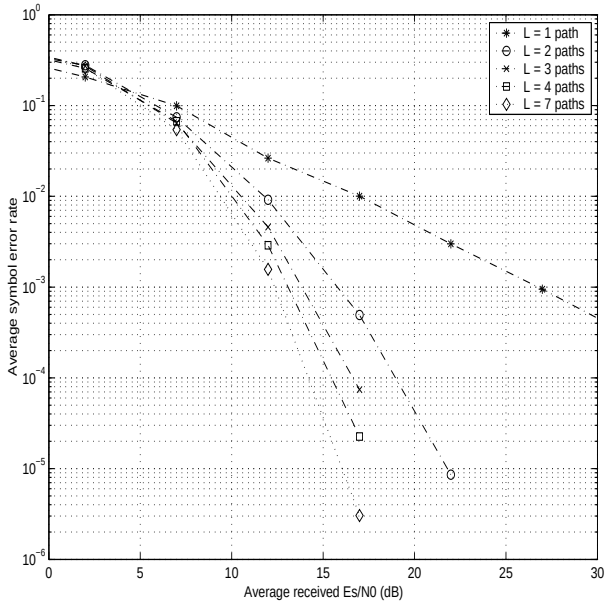

Figure 3.6: Single user, multi path $L = 1, 2, 3, 4, 7$, $N = 31$, $p = 2$, $m = 1$, with 4-QAM.

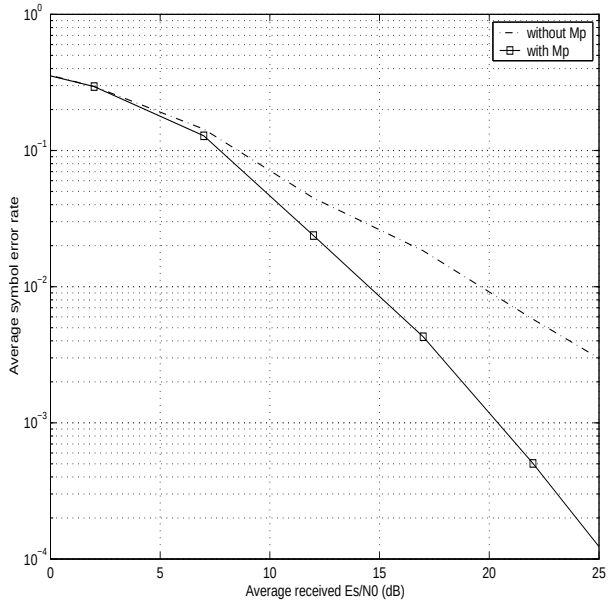

Figure 3.7: Single user, Single path, $N = 31$, $p = 2$, $m = 1$, with 4-QAM.

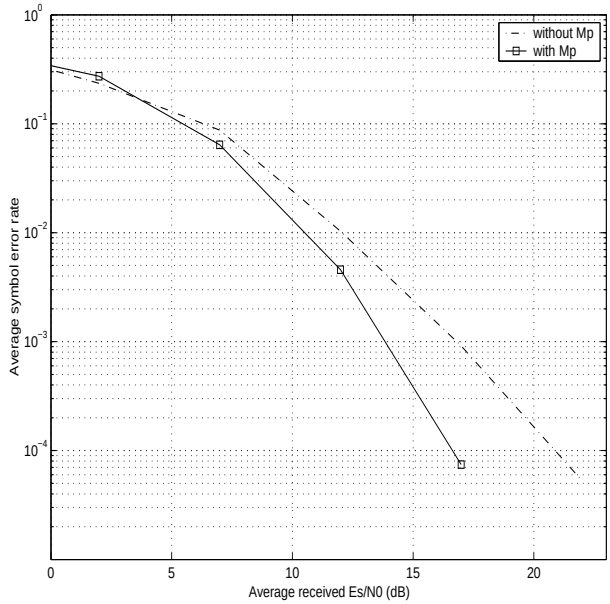

Figure 3.8: Single user, multi path $L = 3$, $N = 31$, $p = 2$, $m = 1$, with 4-QAM.

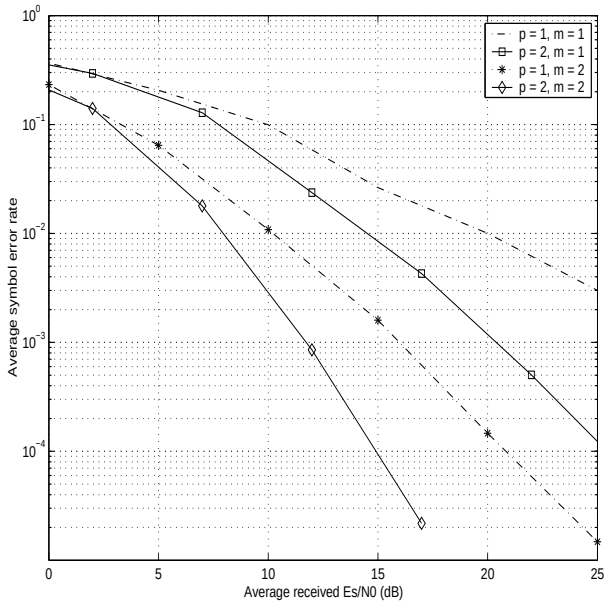

Figure 3.9: Single user, Single path, $p = 1, 2$, $N = 31$, $m = 1, 2$ with 4-QAM.

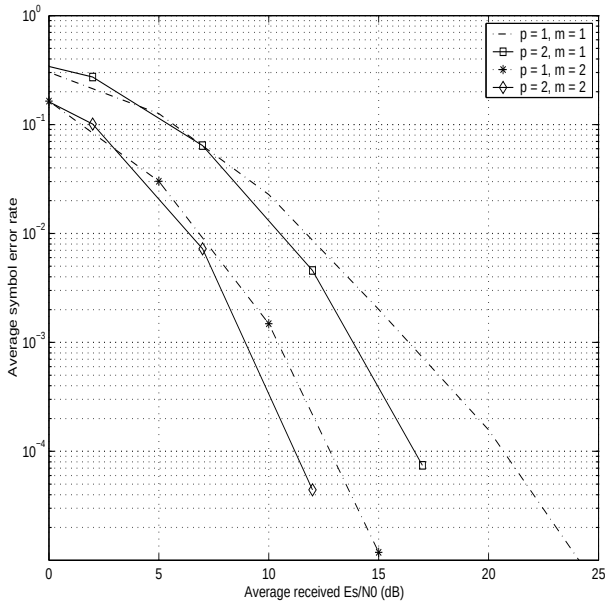

Figure 3.10: Single user, multi path $L = 3$, $N = 31$, $m = 1, 2$, $p = 1, 2$ with 4-QAM.

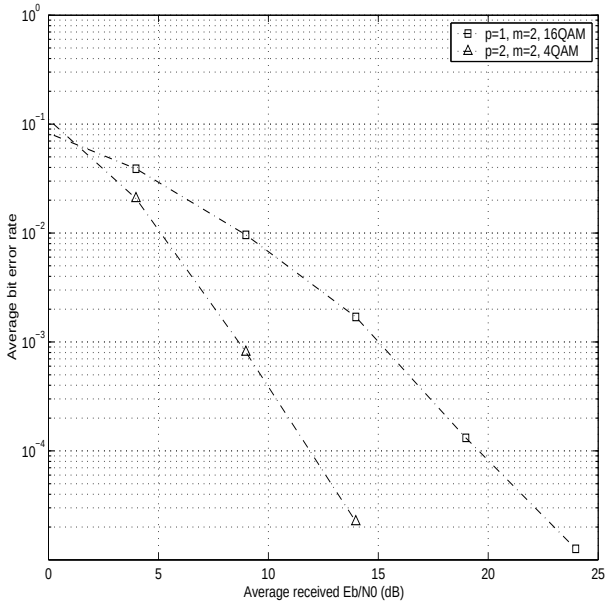

Figure 3.11: Single user, Single path, $N = 31$, $m = 2$, data rate of $4/31$ bits/T_c.

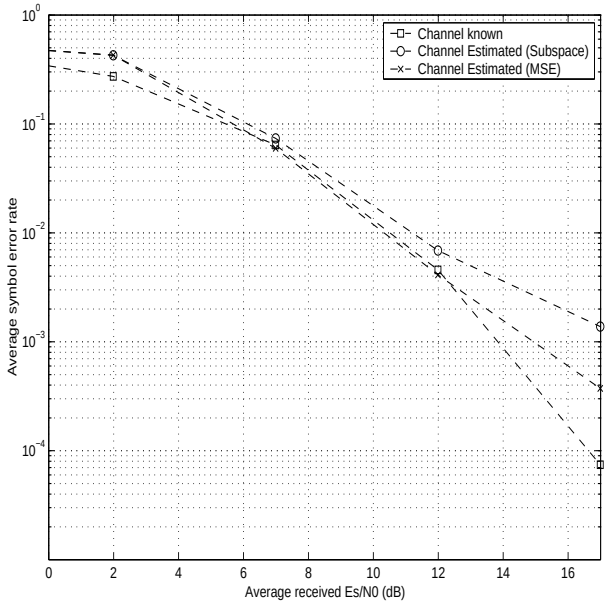

Figure 3.12: Blind channel estimation, single user, $L = 3$, $p = 2$, $m = 1$, $N = 31$, 4-QAM.

Part II

Receiver Side ...

Chapter 4

Exploiting receive diversity

Blind identification techniques are addressed in the rest of this document. These techniques are classified in the receive diversity part of this report as they use the spatial receive diversity to perform the channel identification.

In this chapter we review briefly the subspace based method for MIMO system and we introduce the general notations that we will use later in next chapters. This chapter is organized as follows:

In section 4.1, we highlight the motivation for Blind System Identification (BSI). In 4.2, we focus on subspace based methods and we discuss advantages and limitations of this method. Finally, in 4.3, we give the data model and review the main subspace algorithm.

We will discuss in detail in next chapters proposed techniques developed in order to overcome the limitations of subspace algorithm.

4.1 Motivation for Blind System Identification in MIMO context

Generally, in the communication context a training sequence is used in order to identify the propagation channel and to access the source signals. However, the use of training sequence limits the through put of channel, especially when the latter is time varying. For instance, in the case of MIMO systems $q \times p \times L$ [1] channel coefficients must be identified in the coherence interval.

Moreover, there is a power limitation for the case of MIMO systems: when the number of transmitter antennas increases less power can be devoted to an

[1] q and p denote respectively number of receivers and number of inputs. L represents memory of channels.

eventual training sequence per antenna and the channel identification becomes more difficult and sometimes impossible.

For more information and some information-theoretic approach on training based schemes over multi-antenna systems the reader can see [52], where it is shown that in the case of small SNR the training based schemes are suboptimal in sense of achieving the capacity available in the MIMO context.

For these reasons, Blind System Identification techniques, that estimate channel impulse response using only its second order output statistics, form a good alternative to training based schemes.

4.2 Subspace-based identification methods : Advantages and Limitations

Beside HOS (Higher-Order-Statistics) approaches, we can mention the methods developed in order to blindly identify the Single-Input Multiple-Output (SIMO) systems from the Second-Order-Statistics (SOS) of data ([94, 62] and [92]).

An important class of blind SOS-SIMO identification method is based on subspace decomposition ([68, 6]). This method has been first extended to the MIMO context in [9].

One of the important advantages of subspace method is its deterministic property. That is, the channel parameters can be recovered perfectly in the absence of noise, using only a finite set of data samples, without any statistical assumption on input data. Therefore, subspace method is promising for applications where only a few number of output data is available, or the input data is arbitrary.

Another attractive property of subspace based methods is that the channel estimates can be obtained from optimizing a quadratic cost function under a simple constraint. Therefore, a closed form expression for the channel estimate exists.

Despite their high estimation efficiency, subspace based identification methods have some drawbacks:

Computational complexity : Although the subspace method leads to a minimization of cost function, which is easy to resolve under some constraint, it remains computationally complex especially when there are a large number of sensors in the system. This is mainly due to the fact that the subspace

method requires the computationally expensive and non-parallelizable Eigen Value Decomposition (EVD) of observed outputs (or their covariance matrix). Therefore, finding some subspace based, low computational cost and parallelizable technique is a key challenge.

Among others, we can mention the work of [57] where Minimum Noise Subspace method has been introduced. The latter is a computationally efficient and paralleziable version of subspace method.

Chapter 5 is devoted to this problem, where we introduce the original Minimum Noise Subspace (MNS) method and we present new extensions and implementations of this method in the case of BSI. These methods are called Orthogonal Minimum Noise Subspace (OMNS) and Symmetric Minimum Noise Subspace (SMNS). They have been already presented in [80, 82]. Moreover, we propose a new algorithm to increase the convergence rate of MNS method [82]. A general framework of Minimum Noise Subspace algorithms for array processing applications will be presented later in [84].

Channel order estimation : A difficulty with the subspace method is its sensitivity to mis-specification of the channel order.

It is shown in [9] that, subspace method requires the exact prior knowledge or estimation of the channel order; otherwise it fails.

To address the problem of channel order estimation there are several approaches for SIMO communication system. Most of them are classified in [93]. We will review some of them in detail in chapter 6, where we propose a new approach and several algorithms in order to overcome this limitation. These algorithms are published in [83, 81]. A journal paper concerning this topic is also under preparation [79].

4.3 Subspace method - A review

The subspace identification method is based on the fact that the subspace spanned by a set of observation statistics derived from unknown parameter can be used to estimate that parameter.

Here in this section, we give a brief review of subspace method. For more information and details reader can refer to [6, 9].

Let $\mathbf{y}(n)$ be a q-variate discrete time stationary time series given by:

$$\begin{aligned} \mathbf{y}(n) &\triangleq [\mathbf{H}(z)]\mathbf{s}(n) + \mathbf{n}(n) \\ &= \sum_{k=0}^{M} \mathbf{H}(k)\mathbf{s}(n-k) + \mathbf{n}(n) \end{aligned} \tag{4.1}$$

$\mathbf{s}(n) = [s_1(n), ..., s_p(n)]^T$ is a p-dimensional unknown process and $\mathbf{n}(n)$ is an additive q-dimensional white noise, i.e. $E[\mathbf{n}(n)\mathbf{n}^*(n)] = \sigma^2\mathbf{I}_q$ with σ^2 unknown. $\mathbf{H}(z)$ is an unknown causal FIR $q \times p$ transfer function with $q > p$. It can be written as :

$$\begin{aligned} \mathbf{H}(z) &= \sum_{k=0}^{M} \mathbf{H}(k) z^{-k} \\ &\triangleq \begin{bmatrix} h_{1,1}(z) & \cdots & h_{1,p}(z) \\ \vdots & \ddots & \vdots \\ h_{q,1}(z) & \cdots & h_{q,p}(z) \end{bmatrix}. \end{aligned} \tag{4.2}$$

M denotes the polynomial degree of scalar transfer functions $h_{i,j}(z)$ with $1 \leq i \leq q$ and $1 \leq j \leq p$.

In the communication context, the input sequence $\mathbf{s}(n)$ denotes the transmitted symbols, and the unknown FIR transfer function $\mathbf{H}(z)$ models the propagation channel between sources and sensors.

We study here the estimation of $\mathbf{H}(z)$ from the observation $\mathbf{y}(n)$ under the following assumptions:

$$rank(\mathbf{H}(z)) = p \quad \text{for each } z \tag{4.3}$$

$$\mathbf{H}(z) \quad \text{is column-reduced} \tag{4.3}$$

These assumptions are realistic if the distance between sensors are large enough, with respect to the spatial variation of the channels so that the different propagation channels are independent.

It has been shown in [34], that $\mathbf{H}(z)$ is identifiable up to a constant unimodular $p \times p$ matrix from a finite number of auto-covariance coefficients. This is done by computing the covariance matrix of the output vector $\mathbf{y}(n)$.

Let $\boldsymbol{y}_i(n)$ be a vector of N successive samples corresponding to the i-th output of the system [80]. According to (4.1), it can be written as :

$$\begin{aligned} \boldsymbol{y}_i(n) &= [y_i(n), ..., y_i(n-N+1)]^T \\ &= \mathcal{T}_N(\mathbf{H}_{i,:})\bar{\mathbf{s}}(n) + \mathbf{n}_i(n) \end{aligned} \tag{4.3}$$

$\bar{\mathbf{s}}(n)$ denotes the vector of input samples, i.e.

$$\bar{\mathbf{s}}(n) = [\mathbf{s}_1^T(n), ..., \mathbf{s}_p^T(n)]^T \tag{4.3}$$

where,

$$\mathbf{s}_j(n) = [s_j(n), ..., s_j(n - N - M + 1)]^T \tag{4.3}$$

for $1 \leq j \leq p$, and

$$\mathbf{n}_i(n) = [n_i(n), ..., n_i(n - N + 1)]^T \tag{4.3}$$

$\mathcal{T}_N(\mathbf{H}_{i,:})$ is the $N \times p(N + M)$ block Sylvester matrix given by:

$$\mathcal{T}_N(\mathbf{H}_{i,:}) = [\mathcal{T}_N(h_{i,1}), ..., \mathcal{T}_N(h_{i,p})] \tag{4.3}$$

where, $\mathcal{T}_N(h_{i,j})$ denotes the $N \times (N + M)$ Sylvester matrix associated to $h_{i,j}$. It can be written as:

$$\mathcal{T}_N(h_{i,j}) = \begin{bmatrix} h_{i,j}(0) & \cdots & h_{i,j}(M) & & 0 \\ & \ddots & & \ddots & \\ 0 & & h_{i,j}(0) & \cdots & h_{i,j}(M) \end{bmatrix}$$

Considering all of the outputs of the system and putting them together to obtain $\bar{\boldsymbol{y}}(n)$, we have

$$\begin{aligned} \bar{\boldsymbol{y}}(n) &= [\boldsymbol{y}_1^T(n), ..., \boldsymbol{y}_q^T(n)]^T \\ &= \mathcal{T}_N(\mathbf{H})\bar{\mathbf{s}}(n) + \bar{\mathbf{n}}(n) \end{aligned} \tag{4.3}$$

with

$$\begin{aligned} \bar{\mathbf{n}}(n) &= [\mathbf{n}_1^T(n), ..., \mathbf{n}_q^T(n)]^T \\ \mathcal{T}_N(\mathbf{H}) &= [\mathcal{T}_N^T(\mathbf{H}_{1,:}), ..., \mathcal{T}_N^T(\mathbf{H}_{q,:})]^T \end{aligned} \tag{4.3}$$

$\mathcal{T}_N(\mathbf{H})$ is $qN \times p(N + M)$ generalized Sylvester matrix of order N associated to $\mathbf{H}(z)$. Let $\mathbf{R}_N$ be the $qN \times qN$ covariance matrix of $\bar{\boldsymbol{y}}(n)$:

$$\begin{aligned} \mathbf{R}_N &\triangleq E[\bar{\boldsymbol{y}}(n)\bar{\boldsymbol{y}}^*(n)] \\ &= \mathcal{T}_N(\mathbf{H})\bar{\mathbf{S}}\mathcal{T}_N^*(\mathbf{H}) + \sigma^2\mathbf{I}_{qN} \end{aligned} \tag{4.3}$$

where $\mathcal{S} \triangleq E[\bar{\mathbf{s}}(n)\bar{\mathbf{s}}^*(n)] > 0$. Under assumptions (4.3) and (4.3), and for

$$N > pM \tag{4.3}$$

$\mathcal{T}_N(\mathbf{H})$ has full column rank $p(N+M)$ [2]. In this case, the noise variance σ^2 is the smallest eigenvalue of $\mathbf{R}_N$. The eigenspace associated to σ^2 referred to as the *noise subspace* is of dimension $qN - p(N+M)$. The noise subspace is the orthogonal complement of Range $(\mathcal{T}_N(\mathbf{H}))$, the *signal subspace*.

The eigen-decomposition of $\mathbf{R}_N$ allows to identify the noise subspace Range$(\mathcal{T}_N(\mathbf{H}))^{\perp}$. Denote by $\mathbf{\Pi}_N$ the orthogonal projection matrix onto Range$(\mathcal{T}_N(\mathbf{H}))^{\perp}$. Using the orthogonality relation between noise and signal subspace we can write the following equation:

$$\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H}) = 0 \tag{4.3}$$

In order to estimate the propagation channel, we have to find all of the matrix polynomials $\mathbf{H}(z)$ satisfying (4.3). The subspace identification method is ultimately related at to following Theorem:

Theorem 1 ([34],[9],[50]) *: Suppose that (4.3), (4.3) and (4.3) hold. Let*

$$\boldsymbol{F}(z) = [\boldsymbol{f}_1(z), \cdots, \boldsymbol{f}_p(z)]$$

be a full-rank $q \times p$ polynomial matrix such that

$$deg(\boldsymbol{f}_1(z)) \leq \cdots \leq deg(\boldsymbol{f}_p(z)) \leq M \tag{4.3}$$

The matrix equation

$$\mathbf{\Pi}_N \mathcal{T}_N(\boldsymbol{F}) = 0 \tag{4.3}$$

does not admit any solution if there exists an index i, $1 \leq i \leq p$ such that $deg(\boldsymbol{f}_i(z)) < M$. If $deg(\boldsymbol{f}_i(z)) = M$ for all $1 \leq i \leq p$, then (1) admits solutions of the form

$$\mathbf{F}(z) = \mathbf{H}(z)\mathbf{R}$$

where $\mathbf{R}$ is a constant $p \times p$ invertible matrix.

Theorem 1 is the key stone of the subspace method. For instance, we limit us to the case where $\deg(\boldsymbol{f}_i(z)) = M_i$ with $M_i = M$ for all $1 \leq i \leq p$.

[2] For $N > pM$ we have

$$(p+1)N > p(N+M)$$

As $q > p$,

$$qN > p(N+M).$$

In practice the matrix equation (4.3) is solved in a least squares sense under a suitable constraint. It can be written as:

$$\begin{aligned}\hat{\mathbf{h}} &= \arg\min_{\mathbf{H}} \left\{ ||\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H})||^2 \right\} \\ &= \arg\min_{\mathbf{H}} \left\{ \text{ trace } \{(\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H}))^* (\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H}))\} \right\} \\ &= \arg\min_{\mathbf{H}} \left\{ \text{ vec}(\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H}))^* \text{ vec}(\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H})) \right\}\end{aligned} \tag{4.1}$$

where $\hat{\mathbf{h}}$ denotes the channel parameter vector defined as :

$$\hat{\mathbf{h}} \stackrel{\triangle}{=} \text{vec}([\mathbf{H}_1^T, \cdots, \mathbf{H}_q^T]^T) \tag{4.0}$$

with

$$\mathbf{H}_i = [\mathbf{H}_i^T(0), \cdots, \mathbf{H}_i^T(M)]^T \tag{4.0}$$

In the following we describe equations which are used to obtain the closed form solution of (1).

The generalized Sylvester $qN \times p(N+M)$ matrix $\mathcal{T}_N(\mathbf{H})$ depends linearly on its parameters. Therefore, we can write:

$$\begin{aligned}\text{vec}(\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H})) &= \sum_{i=1}^{q} \text{vec}\big(\mathbf{\Pi}_N^i \mathcal{T}_N(\mathbf{H}_{i,:})\big) \\ &= \sum_{i=1}^{q} \big(\boldsymbol{I}_p \otimes \ \mathcal{D}(\mathbf{\Pi}_N^i)\big)\mathbf{H}_i\end{aligned} \tag{4.-1}$$

$\mathcal{T}_N(\mathbf{H}_{i,:})$ denotes a $N \times p(N+M)$ matrix corresponding to the i-th block of $\mathcal{T}_N(\mathbf{H})$ and $\mathbf{\Pi}_N^i$ denotes the sub-matrices of dimension $qN \times N$ corresponding to the i-th block of $\mathbf{\Pi}_N$. Therefore, $\mathbf{\Pi}_N$ can be written as:

$$\mathbf{\Pi}_N = [\mathbf{\Pi}_N^1, \cdots, \mathbf{\Pi}_N^i, \cdots, \mathbf{\Pi}_N^q] \tag{4.-1}$$

where

$$\mathbf{\Pi}_N^i = \big[\boldsymbol{\pi}_{i,1}, \cdots, \boldsymbol{\pi}_{i,N}\big] \tag{4.-1}$$

$\boldsymbol{\pi}_{i,N}$ correspond to vectors of dimension qN. $\mathcal{D}(\mathbf{\Pi}_N^i)$, in (4.1), is a block-toeplitz matrix of dimension $qN(M+N) \times (M+1)$

$$\mathcal{D}(\mathbf{\Pi}_N^i) \triangleq \begin{bmatrix} \boldsymbol{\pi}_{i,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\pi}_{i,2} & \boldsymbol{\pi}_{i,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \boldsymbol{\pi}_{i,1} \\ \boldsymbol{\pi}_{i,N} & & & \boldsymbol{\pi}_{i,2} \\ \mathbf{0} & \boldsymbol{\pi}_{i,N} & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\pi}_{i,N} \end{bmatrix} \tag{4.-1}$$

Equation (4.1) can be written in matrix form as:

$$\mathrm{vec}(\mathbf{\Pi}_N \mathcal{T}_N(\mathbf{H})) = \underbrace{\begin{bmatrix} \left(\boldsymbol{I}_p \otimes \ \mathcal{D}(\mathbf{\Pi}_N^1)\right) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \left(\boldsymbol{I}_p \otimes \ \mathcal{D}(\mathbf{\Pi}_N^2)\right) & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \left(\boldsymbol{I}_p \otimes \ \mathcal{D}(\mathbf{\Pi}_N^q)\right) \end{bmatrix}}_{D(\mathbf{\Pi})} \mathbf{h} \tag{4.-1}$$

therefore the minimization of equation (4.4) can be written as a quadratic form:

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{H}} \left\{ \mathbf{h}^* \Big(\ D^*(\mathbf{\Pi}) \ D(\mathbf{\Pi}) \Big) \mathbf{h} \right\} \tag{4.-1}$$

The solution of (4.3) is the eigenvector associated to the least eigenvalue of $D(\mathbf{\Pi})^* \ D(\mathbf{\Pi})$.

Remark : Later, we will see that this form of arranging data as in (4.4) is more convenient when we want to separate different sub-systems in the case of MNS algorithm. It is important to know that, outputs of the system can be arranged by putting them in a vector of N successive samples of all output of the system like in [9]. However, the main properties of subspace method does not depend on the way of arranging the output data.

4.3.1 Subspace algorithm

Subspace algorithm can be summarized as follows:

- Choose a window length N such that $N > pM$ and estimate the $qN \times qN$ covariance matrix $\hat{\mathbf{R}}_N$ from the observations.

$$\hat{\mathbf{R}}_N = \frac{1}{T-N+1} \sum_{n=1}^{T-N+1} \big(\mathbf{y}(n)\mathbf{y}^*(n)\big) \tag{4.-1}$$

  where $T >> N$ denotes number of samples.

- Obtain the noise projector $\hat{\mathbf{\Pi}}_N$ onto the subspace spanned by the eigenvector associated to the $qN-p(N+M)$ smallest eigenvalues of the covariance matrice $\hat{\mathbf{R}}_N$.

- Find $\hat{\mathbf{H}}(z)$ (up to a constant non singular $p \times p$ matrix) as the eigenvector associated to the smallest eigenvalue of $qp(M+1) \times qp(M+1)$ matrix $D(\mathbf{\Pi})^*\ D(\mathbf{\Pi})$.

4.3.2 Discussions

In the following sections we discuss two aforementioned limitations of subspace method.

4.3.2.1 Computational complexity

Computational cost of subspace method is dominated by SVD computation of $qN \times qN$ covariance matrix $\hat{\mathbf{R}}_N$.

Generally speaking, SVD computation of a $qN \times qN$ matrix costs $\mathcal{O}(qN^3)$ flops. Moreover, this decomposition can not be implemented in a parallel way. This computational complexity is one of the main drawbacks of subspace method, especially for large sensors array systems.

Example : Let us consider a two-input ($p = 2$) three-output ($q = 3$) MIMO communication system with sub-channels of degree $M = 4$. The minimum size of temporal window N must be $N = pM+1 = 9$. Therefore, SVD computation of covariance matrix costs $\mathcal{O}(27^3)$.

In next chapter we will study and propose some computationally efficient subspace based methods called as Minimum Noise Subspace (MNS). We will see later that the computational complexity of the MNS-like algorithms are much less than that of the original subspace method. Moreover, they can be implemented in parallel scheme.

4.3.2.2 Order overestimation

In this chapter we suppose that the order of sub-channel M is known or correctly estimated. Based on this fact, the propagation channel can be estimated up to a constant $p \times p$ matrix as presented in Theorem 1.

Indeed, when the exact channel order is not known or correctly estimated the structural properties of matrix equation (1) are no more valid. Therefore, the subspace method fails.

In practice, the exact channel order value is not known and it depends on the application and the environment. One existing approach to estimate the channel order consists of applying the Minimum Description Length (MDL) criterion [102] to the covariance matrix. However, it is known that the MDL criterion, tends to overestimate the exact channel order.

When an overestimation of a channel order occurs the main subspace Theorem in 1 can be extended to the following one:

Theorem 2 ([34],[9],[50]) *: Suppose that (4.3), (4.3) and (4.3) hold. Let*

$$\boldsymbol{F}(z) = [\boldsymbol{f}_1(z), \cdots, \boldsymbol{f}_p(z)]$$

be a full-rank $q \times p$ polynomial matrix such that

$$deg(\boldsymbol{f}_1(z)) = \cdots = deg(\boldsymbol{f}_p(z)) = M' \geq M \tag{4.-1}$$

The matrix equation

$$\boldsymbol{\Pi}_N \mathcal{T}_N(\boldsymbol{F}) = 0 \tag{4.-1}$$

admits solutions of the form

$$\mathbf{F}(z) = \mathbf{H}(z)\mathbf{R}(z)$$

where $\mathbf{R}(z)$ is a polynomial $p \times p$ matrix of degree $M' - M$.

In chapter 6 we see that in the context of SIMO system solutions of equation (2) have a specific structural property. Using this property, we propose there, several algorithms robust to order overestimation of the propagation channel.

4.4 Conclusions

Subspace decomposition is known to be an important tool in array signal processing for several years. Moreover, we have seen in this chapter that using the

subspace method we can identify the MIMO propagation channel under several assumptions without loss in capacity.

However, Subspace method is computationally expensive to implement. That is why, we study and propose in the next chapter some computationally efficient and parallelizable subspace method referred to as the Minimum Noise Subspace (MNS) method.

Moreover, it is well known that subspace method fail to provide the exact solution when the degree of channel is not exactly estimated. Therefore, we have to improve existing subspace algorithm by using some order detection process or finding robust algorithms. That is the subject of chapter 6.

Chapter 5

Minimum Noise Subspace

5.1 Toward a computationally efficient subspace method

We have seen that existing subspace techniques despite their high estimation efficiency, are computationally intensive and difficult to implement in real time, especially for large sensor array systems. The main reason is that they require computationally expensive and non-parallelizable eigen value decomposition (EVD) to estimate the noise subspace[1].

It has been shown in [57, 7] that the full noise subspace computation of the system output covariance matrix is not necessary to yield the unique estimate of the channel responses. This gives rise to the concept of Minimum Noise Subspace (MNS). MNS technique is a promising technique for effective computation of the noise subspace. In this method, the noise vectors are computed in parallel scheme from $q - p$ sub-systems chosen so that the corresponding rational noise subspaces are of minimum dimension equal to one.

This chapter is devoted to MNS techniques and their extensions. We start by introducing, in section 5.2, the concept of rational Subspace and minimal polynomial bases which are necessary to introduce the MNS concept. MNS concept and algorithm is detailed in section 5.3. Several extensions of the main MNS concepts and their properties are presented in section 5.4. In section 5.5, we compare the computational complexity of proposed methods. In section 5.6, we present different algorithms for the block implementation of MNS method. In section 5.7, we compare the performance of different algorithms by simulation. Finally, in section 5.8 we conclude this chapter and we propose some perspectives.

[1] EVD of an $n \times n$ matrix costs $\mathcal{O}(n^3)$ flops.

5.2 Problem formulation

The system model used in this chapter is that of section 4.3.

Subspace identification method can be recast in a more general framework by resorting to the concept of rational subspaces. As we shall see below, one can express the signal and noise subspaces in the field of rational functions to get more insights into the subspace method.

5.2.1 Rational subspace and polynomial bases

A $qN \times 1$ vector $\mathbf{g} = [\mathbf{g}_1^T, \ldots, \mathbf{g}_q^T]^T$ (where each vector $\mathbf{g}_k = [g_k(0), \ldots, g_k(N-1)]^T$ has dimension $N \times 1$) belongs to the noise subspace of $\mathbf{R}_N$ if and only if $\mathbf{g}^* \mathcal{T}_N(\mathbf{H}) = 0$. The orthogonality condition is conveniently rewritten as:

$$\mathbf{g}^* \mathcal{T}_N(\mathbf{H}) = 0 \Longleftrightarrow \mathbf{g}^*(z)\mathbf{H}(z) = 0 \quad \text{for each } z \tag{5.0}$$

$$\mathbf{g}(z) = \begin{bmatrix} g_1(z) \\ \vdots \\ g_q(z) \end{bmatrix} \quad \text{and} \quad \mathbf{g}^*(z) = \begin{bmatrix} g_1^*(z) \\ \vdots \\ g_q^*(z) \end{bmatrix}$$

$$g_k(z) = \sum_{l=0}^{N-1} g_k(l) z^{-l} \quad \text{and} \quad g_k^*(z) = \sum_{l=0}^{N-1} g_k^*(l) z^{-l} \tag{5.0}$$

We denote by $\mathcal{C}^q(z)$ the set of all q-dimensional rational functions or, in other words, the q-dimensional vector space on the field of all scalar rational functions. Such a vector subspace is referred to as *rational space*. Let $\mathcal{S}$ be the p-dimensional rational subspace of $\mathcal{C}^q(z)$ spanned by the column vectors of $\mathbf{H}(z)$ ($\mathcal{S} = \text{Range}(\mathbf{H}(z))$. Let $\mathcal{B} \subset \mathcal{C}^q(z)$ denote the orthogonal complement of $\mathcal{S}$ (i.e., the subspace of all q-dimensional rational transfer functions $\mathbf{g}(z)$ satisfying $\mathbf{g}(z)\mathbf{f}^*(z) = 0$ for each $\mathbf{f}(z) \in \mathcal{S}$). It then follows that $\mathcal{B}$ has dimension $q - p$.

The subspace method can now be seen as a method of finding $\mathbf{H}(z)$ such that $\mathbf{H}(z) \perp \mathcal{B}$. However, $\mathcal{B}$ can be uniquely spanned by a basis of $q - p$, q-dimensional polynomial vectors. Therefore, to identify $\mathbf{H}(z)$, it suffices to find a *polynomial basis* $\mathbf{V}(z) = [\mathbf{v}_1(z), ..., \mathbf{v}_{q-p}(z)]$ of $\mathcal{B}$ and to express the orthogonality between $\mathbf{v}_i$ and $\mathbf{H}(z)$ for $i = 1, \ldots, q - p$, i.e.

$$\mathbf{v}_i^*(z)\mathbf{H}(z) = 0.$$

Theorem 3 [60, 34] *Let* $\mathbf{F}(z)$ *be a full column rank* $q \times p$ *polynomial matrix with* $deg(\mathbf{F}_{:,i}(z)) = M_i = M$ *for all* $1 \leq i \leq p$, *and* $Range(\mathbf{V}(z)) = \mathcal{B} = \text{range}(\mathbf{H}(z))^{\perp}$. *Then, the solution to*

$$\mathbf{V}^*(z)\mathbf{F}(z) = 0$$

is given by $\mathbf{F}(z) = \mathbf{H}(z)\mathbf{T}$, *where* $\mathbf{T}$ *is a constant* $p \times p$ *invertible matrix.*

Since $\mathbf{V}(z)$ contains the minimum number $q - p$ of noise polynomial vectors to span $\mathcal{B}$, the estimation method is called Minimum Noise Subspace. Later on, we will see that we can go beyond this assumption, under certain conditions.

5.3 Minimum Noise Subspace Concepts

In MNS method we can compute noise subspace in a parallel scheme using a special set of system outputs which form a Properly Connected Sequence (PCS). In the following, we will introduce this concept before starting algorithm description.

5.3.1 Properly Connected Sequence

Denote the q system outputs by a set of members $m_1, \cdots, m_q$. A combination of $p+1$ members $T = (m_{i_1}, \cdots, m_{i_{p+1}})$ is called a $(p+1)$-tuple.

Definition 2 *A sequence of* $q-p$ *tuples is said to be properly connected if each tuple in the sequence consists of* p *members shared by its preceding tuples and another member not shared by its preceding tuples.*

A properly connected sequence (PCS) is denoted by $S(p,q) = \{T_1, T_2, \cdots, T_{q-p}\}$ where

$$\begin{aligned} &T_i = (m_{i_1}, \cdots, m_{i_p}, m_{i_{p+1}}), \quad 1 \leq i \leq q-p \\ &\{m_{i_1}, \cdots, m_{i_p}\} \subset \{m_{j_k} |\ j < i,\ 1 \leq k \leq p+1\} \\ &m_{i_{p+1}} \notin \{m_{j_k} |\ \ j < i,\ 1 \leq k \leq p+1\}. \end{aligned}$$

Example : Two examples of PCS called PCS1 and PCS2 are given below in table 5.1 for a system with $q = 6$ outputs and $p = 2$ inputs

$T_1 = (m_1, m_2, m_3)$	$T_1 = (m_1, m_2, m_3)$
$T_2 = (m_1, m_2, m_4)$	$T_2 = (m_2, m_3, m_4)$
$T_3 = (m_1, m_2, m_5)$	$T_3 = (m_1, m_4, m_5)$
$T_4 = (m_1, m_2, m_6)$	$T_4 = (m_2, m_5, m_6)$
PCS1	PCS2

Table 5.1: Example of two different Properly Connected Sequences PCS1 and PCS2.

5.3.2 Parallel computation of MNS using PCS

Each tuple from $S(p, q)$ can be used, by following equations (4.4)-(4.4), to form a $(p+1)N \times (p+1)N$ covariance matrix $\boldsymbol{R}_N^i$ for $i = 1, \cdots, q-p$. It is clear that each $\boldsymbol{R}_N^i$ is a sub-matrix of the system output covariance matrix $\boldsymbol{R}_N$. Later on, in section 5.3.3, we will see in detail the computation of $\boldsymbol{R}_N^i$.

The least eigenvector of $\boldsymbol{R}_N^i$ is denoted by $\tilde{\mathbf{v}}_i = [\tilde{\mathbf{v}}_{i_1}^T, \cdots, \tilde{\mathbf{v}}_{i_{p+1}}^T]^T$. Each N-dimensional sub-vector $\tilde{\mathbf{v}}_{i_k}$ (where $\tilde{\mathbf{v}}_{i_k} = [\tilde{\mathbf{v}}_{i_k}(0), ..., \tilde{\mathbf{v}}_{i_k}(N-1)]^T$, for $k = 1, ..., p+1$) corresponds to a member in the i-th $(p+1)$-tuple of system outputs.

We will show later in this section that the set of obtained least eigenvectors $\{\tilde{\mathbf{v}}_1, \cdots, \tilde{\mathbf{v}}_{q-p}\}$ asymptotically determine $\mathbf{H}(z)$ up to a $p \times p$ constant matrix $\mathbf{T}$ under a mild additional assumption.

Let $\mathbf{H}_i(z)$ consist of $p+1$ rows of $\mathbf{H}(z)$, where each row corresponds to a member in the i-th $(p+1)$-tuple of system outputs and $\mathcal{T}_N(\mathbf{H}_i)$ be a generalized Sylvester matrix associated with $\mathbf{H}_i(z)$. Then similar to equation (5.2.1), we have :
For $i = 1, \cdots, q-p$

$$\tilde{\mathbf{v}}_i^* \mathcal{T}_N(\mathbf{H}_i) = 0 \qquad \text{or equivalently,} \qquad \tilde{\mathbf{v}}_i^*(z)\mathbf{H}_i(z) = 0 \tag{5.-3}$$

where

$$\tilde{\mathbf{v}}_i(z) = [\tilde{\mathbf{v}}_{i_1}^T(z), \cdots, \tilde{\mathbf{v}}_{i_{p+1}}^T(z)]^T \tag{5.-3}$$

with

$$\tilde{\mathbf{v}}_{i_j}(z) = \sum_{k=0}^{N-1} \tilde{\mathbf{v}}_{i_j}(k) z^{-k}. \tag{5.-3}$$

Then, a set of independent *"zero padded"* qN-dimensional noise vectors $\{\mathbf{v}_i\}_{1 \le i \le q-p}$ is computed from the set of least eigenvectors $\{\tilde{\mathbf{v}}_i\}_{1 \le i \le q-p}$ in a parallel scheme, in the following way:

For $k = 1, \cdots, q$

$$\mathbf{v}_{i,k} = \begin{cases} \tilde{\mathbf{v}}_{i_j} & \text{if} \quad k = i_j \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{5.-3}$$

where

$$\mathbf{v}_i = \begin{bmatrix} \mathbf{v}_{i,1} \\ \vdots \\ \mathbf{v}_{i,q} \end{bmatrix}. \tag{5.-3}$$

On the other hand zero padding correspond in adding zero at the index that does not belong to the considered output. Table 5.2 demonstrates the computation of independent noise vectors $\mathbf{v}_i$ corresponding to the set of least noise vectors $\tilde{\mathbf{v}}_i$, for two different PCS mentioned in the previous example (PCS 1 and PCS 2).

For example, let us consider Tuple T_1 demonstrated in Table 5.2. This tuple contains outputs 1, 2 and 3. Consequently, the corresponding zero padded vector $\mathbf{v}_1$ contains blocks of zero in the positions 4, 5 and 6 of the resulted vector (see the first column of the matrix corresponding to PCS1).

Remark : Provided none of the $p \times p$ sub matrices of $\mathbf{H}_i(z)$ for $1 \leq i \leq q-p$ is singular, each entry of $\tilde{\mathbf{v}}_i(z)$ is nonzero. Therefore, such constructed zero padded vectors $\mathbf{v}_i(z)$ belong to $\mathcal{B}$.

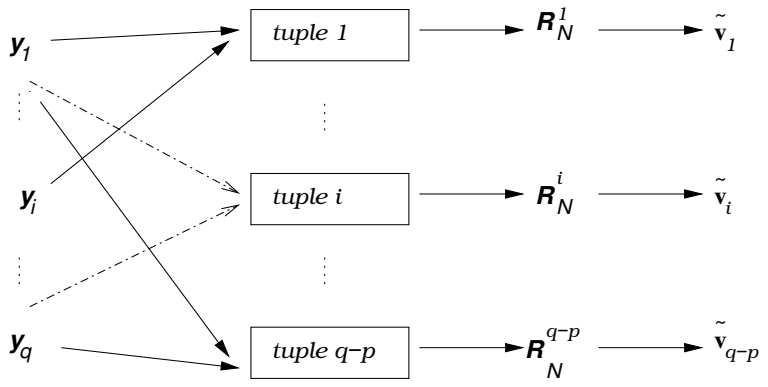

Figure 5.1: Computation of the Minimum Noise Subspace.

Figure 5.1 summarizes the computation of the Minimum Noise Subspace.

Therefore, we have the following Theorem:

$$\begin{array}{c} \begin{array}{cccc} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{array} \\ \left[\begin{array}{cccc} \tilde{\mathbf{v}}_{1_1} & \tilde{\mathbf{v}}_{2_1} & \tilde{\mathbf{v}}_{3_1} & \tilde{\mathbf{v}}_{4_1} \\ \tilde{\mathbf{v}}_{1_2} & \tilde{\mathbf{v}}_{2_2} & \tilde{\mathbf{v}}_{3_2} & \tilde{\mathbf{v}}_{4_2} \\ \tilde{\mathbf{v}}_{1_3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{v}}_{2_3} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{v}}_{3_3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{v}}_{4_3} \end{array}\right] \\ \text{PCS1.} \end{array} \qquad \begin{array}{c} \begin{array}{cccc} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{array} \\ \left[\begin{array}{cccc} \tilde{\mathbf{v}}_{1_1} & \mathbf{0} & \tilde{\mathbf{v}}_{3_1} & \mathbf{0} \\ \tilde{\mathbf{v}}_{1_2} & \tilde{\mathbf{v}}_{2_1} & \mathbf{0} & \tilde{\mathbf{v}}_{4_1} \\ \tilde{\mathbf{v}}_{1_3} & \tilde{\mathbf{v}}_{2_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{v}}_{2_3} & \tilde{\mathbf{v}}_{3_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{v}}_{3_3} & \tilde{\mathbf{v}}_{4_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{v}}_{4_3} \end{array}\right] \\ \text{PCS2.} \end{array}$$

Table 5.2: Computation of independent noise vectors $\mathbf{v}_i$ corresponding to the set of least noise vectors $\tilde{\mathbf{v}}_i$, for PCS1 and PCS2

Theorem 4 [7] *Assume none of the $p \times p$ sub matrices of $\mathbf{H}_i(z)$, $1 \leq i \leq q-p$, is singular, and $\mathbf{F}(z)$ is a full rank $q \times p$ polynomial matrix with $\deg(\mathbf{F}_{:,i}(z)) = M_i = M$, for all i with $1 \leq i \leq p$. Then, the solution to*

$$[\mathbf{v}_1, \cdots, \mathbf{v}_{q-p}]^* \mathcal{T}_N(\mathbf{F}) = 0 \tag{5.-3}$$

is given by

$$\mathbf{F}(z) = \mathbf{H}(z)\mathbf{T} \tag{5.-3}$$

where $\mathbf{T}$ is a non-singular constant $p \times p$ matrix.

5.3.3 MNS algorithm

Here we briefly recall the estimation procedure given in [7].

- Select a set of $q-p$ $(p+1)$-tuples from the q system outputs such that the set forms a PCS:

  $$S(p,q) = \{T_1, \cdots, T_{q-p}\}.$$

  Compute the covariance matrix $\boldsymbol{R}_N^i$ in parallel scheme for each tuple of outputs i where $i = 1, \cdots, q-p$.

  $$\boldsymbol{R}_N^i = \frac{1}{T-N+1} \sum_{n=1}^{T-N+1} \mathbf{y}_i(n)\mathbf{y}_i^*(n) \tag{5.-3}$$

  with T being the sample size and :

  $$\mathbf{y}_i(n) \triangleq [\boldsymbol{y}_{m_{i_1}}^T(n), \cdots, \boldsymbol{y}_{m_{i_{p+1}}}^T(n)]^T. \tag{5.-3}$$

- Compute a least eigenvector $\tilde{\mathbf{v}}_i$ of each $\boldsymbol{R}_N^i$. Note that the $(p+1)N \times 1$ vector $\tilde{\mathbf{v}}_i$ satisfies

$$\tilde{\mathbf{v}}_i^* \mathcal{T}_N(\mathbf{H}_i) = 0 \Longleftrightarrow \tilde{\mathbf{v}}_i^*(z)\mathbf{H}_i(z) = 0$$

where

$$\tilde{\mathbf{v}}_i(z) = [\tilde{\mathbf{v}}_{i_1}^T(z), \cdots, \tilde{\mathbf{v}}_{i_{p+1}}^T(z)]^T \tag{5.-3}$$

with

$$\tilde{\mathbf{v}}_{i_j}(z) = \sum_{k=0}^{N-1} \tilde{\mathbf{v}}_{i_j}(k) z^{-k}. \tag{5.-3}$$

- Following (5.3.3), construct the q-dimensional noise polynomial vectors $\mathbf{v}_i(z)$ corresponding to $\tilde{\mathbf{v}}_i(z)$:

For $k = 1, \cdots, q$

$$\mathbf{v}_i(k) = \begin{cases} \tilde{\mathbf{v}}_{i_j} & \text{if} \quad k = i_j \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{5.-3}$$

where

$$\mathbf{v}_i = \begin{bmatrix} \mathbf{v}_{i,1}^T \\ \vdots \\ \mathbf{v}_{i,q}^T \end{bmatrix}. \tag{5.-3}$$

- Estimate the channel transfer function $\mathbf{H}(z)$ (up to a constant $p \times p$ matrix) minimizing the quadratic criterion (in terms of the channel coefficients)

$$\hat{\mathbf{H}}(z) = \arg\min_{\mathbf{H}(z)} \|\mathbf{V}^* \mathcal{T}_N(\mathbf{H})\|^2 \tag{5.-3}$$

under some adequate constraint (see [7, 50] for more details).

Here, $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_{q-p}]$ is the noise vector matrix of dimension $qN \times (q-p)$ [2]and $\mathcal{T}_N(\mathbf{H})$ is the $qN \times p(N+M)$ block Sylvester matrix corresponding to $\mathbf{H}(z)$.

5.4 Extensions of MNS method

In this section we introduce two possible extensions of the MNS method corresponding to the Symmetric MNS (S-MNS) and Orthogonal MNS (O-MNS)

[2]In MNS algorithm the noise vectors $\mathbf{v}_1, \cdots, \mathbf{v}_{q-p}$ are not orthogonal.

methods. The former method exploits the parallel computation of q noise vectors and the latter is based on an iterative procedure to obtain an orthogonal basis of noise subspace.

5.4.1 Symmetric MNS - SMNS

The idea of SMNS is to use q noise vectors instead of $q-p$ as in the original MNS in such a way that all system outputs are used identically.

Indeed, in the original MNS method, certain system outputs are used more than others depending on the chosen PCS. This might lead to poor estimation performances if the system outputs that are used most correspond to the 'worst system channels'. This raises the problem of the 'best' choice of PCS.

Definition 3 *A sequence of q, $(p+1)$-tuple is said to be symmetric if all of the system outputs are used $p+1$ times.*

Example: Symmetric tuples are given below in table 5.3 for $q=6$ outputs and $p=2$ inputs.

$$
\begin{aligned}
T_1 &= (m_1, m_2, m_3)\\
T_2 &= (m_2, m_3, m_4)\\
T_3 &= (m_3, m_4, m_5)\\
T_4 &= (m_4, m_5, m_6)\\
T_5 &= (m_5, m_6, m_1)\\
T_6 &= (m_6, m_1, m_2)
\end{aligned}
$$

Table 5.3: Example of Symmetric tuple

Note that the first $q-p$ tuples correspond to a PCS and the last p tuples correspond to the additional redundancy we introduce to guarantee the use of all system outputs $p+1$ times. We guarantee here a certain symmetry between the system outputs. It is obvious that the computation of the set of q noise vectors can be made in parallel scheme as mentioned before. Computation of symmetric MNS is illustrated on figure 5.2.

5.4.2 Orthogonal MNS - OMNS

An alternative method to compute the noise polynomial basis $\mathbf{V}(z) = [\mathbf{v}_1(z), ..., \mathbf{v}_{q-p}(z)]$ is proposed in this section. The noise vectors are computed

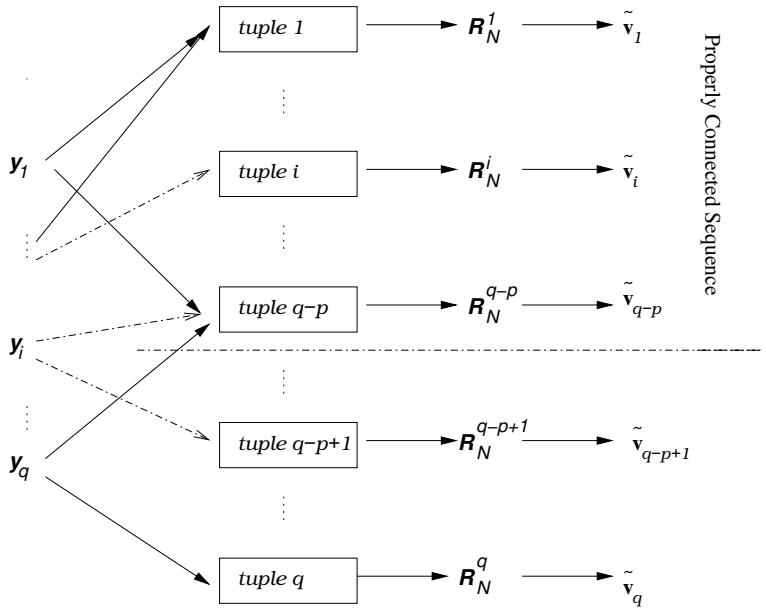

Figure 5.2: Computation of the Symmetric Minimum Noise Subspace.

- Recursively (contrary to standard MNS where noise vectors are computed in a parallel scheme).
- Using all system outputs (in standard MNS each vector is computed using only $p+1$ system outputs).
- In such a way to form an orthogonal basis of $\mathcal{B}$ (this is not the case in section 5.3.3), i.e.

$$\mathbf{v}_i^*(z)\mathbf{v}_j(z) = 0 \qquad \text{for} \quad i \neq j. \tag{5.-3}$$

At the i-th step, we compute a q-dimensional polynomial noise vector $\mathbf{v}_i(z)$ orthogonal to $\mathbf{H}(z)$ and to the previously computed q-dimensional polynomial noise vectors. Each noise vector is obtained by computing the least eigenvector of a $qN_i \times qN_i$, $(i = 1, ..., q-p)$ matrix which is a function of the channel outputs and the previously computed polynomial noise vectors. N_i is chosen in order to obtain a tall matrix at each step.

5.4.2.1 OMNS algorithm

More precisely, we have the following algorithm.

1. Initialization:

- Choose N_1 a window length such that $qN_1 > p(M+N_1)$ and estimate the covariance matrix $\mathbf{R}_{N_1}$ from the observations.
- Compute $\mathbf{v}_1$ as the least eigenvector of $\mathbf{R}_{N_1}$, the latter satisfies:

$$\mathbf{v}_1^* \mathcal{T}_{N_1}(\mathbf{H}) = 0 \Longleftrightarrow \mathbf{v}_1^*(z)\mathbf{H}(z) = 0 \tag{5.-3}$$

  where $\mathbf{v}_1(z)$ is constructed as shown in section 5.2.

2. for $i = 2, \ldots, q-p$:

- Choose N_i a window length such that:

$$qN_i > p(M+N_i) + \sum_{j=1}^{i-1}(N_j - 1) \tag{5.-3}$$

  [3] and then compute the matrix:

$$\mathbf{M}_i = \mathbf{R}_{N_i} + \sum_{j=1}^{i-1} \mathcal{T}_{N_i}(\mathbf{v}_j)\mathcal{T}_{N_i}^H(\mathbf{v}_j) \tag{5.-3}$$

- Compute $\mathbf{v}_i$ as the least eigenvector of $\mathbf{M}_i$. The latter satisfies:

$$\begin{aligned} \mathbf{v}_i^* \mathcal{T}_{N_i}(\mathbf{H}) &= 0 \\ \mathbf{v}_i^* \mathcal{T}_{N_i}(\mathbf{v}_j) &= 0 \quad \text{for} \quad j = 1, \ldots, i-1 \end{aligned} \tag{5.-3}$$

  or equivalently :

$$\begin{aligned} \mathbf{v}_i^*(z)\mathbf{H}(z) &= 0 \\ \mathbf{v}_i^*(z)\mathbf{v}_j(z) &= 0 \quad \text{for} \quad j = 1, \ldots, i-1 \end{aligned} \tag{5.-3}$$

3. Once the $(q-p)$ noise vectors are computed, estimate the channel matrix $\mathbf{H}(z)$ (up to a constant nonsingular $p \times p$ matrix) as:

$$\hat{\mathbf{H}}(z) = \arg\min_{\mathbf{H}(z)} \sum_{i=1}^{q-p} \|\mathbf{v}_i{}^* \mathcal{T}_{N_i}(\mathbf{H})\|^2 \tag{5.-3}$$

   The minimization in (3) is done under a suitable constraint as shown in [50].

[3] On the other hand at each iteration the Matrix containing the channel $\mathbf{H}(z)$ and the previously computed noise vectors $\mathbf{v}_i$ is a tall matrix.

5.5 Computational complexity

The computational cost of the MNS method is $\mathcal{O}((q-p)(p+1)^2N^2)$ flops comparing to $\mathcal{O}(q(p+1)^2N^2)$ flops for SMNS method and $\mathcal{O}(\sum_{i=1}^{q-p}(N_iq)^2)$ flops for OMNS method when it is $\mathcal{O}((qN)^3)$ for the subspace method.

In this computational costs we did not include the cost of covariance matrix estimations, i.e. the estimation of $E[\mathbf{y}(n+k)\mathbf{y}^*(n)]$ for $k=0,\ldots,M$ which is almost equal for all considered methods.

	Computational Complexity
SS method	$\mathcal{O}((qN)^3)$
MNS method	$\mathcal{O}((q-p)(p+1)^2N^2)$
OMNS method	$\mathcal{O}(\sum_{i=1}^{q-p}(N_iq)^2)$
SMNS method	$\mathcal{O}(q(p+1)^2N^2)$

Table 5.4: Computational complexity of SS, MNS, OMNS and SMNS

Therefore, MNS and SMNS methods have the least computational complexity. However, when $q >> (q-p)$ the SMNS method becomes computationally more expensive.

The above computational cost does not take into account the parallel structure of MNS and SMNS method, which is an additional advantage of these methods.

The OMNS method remains less complex than the original subspace method in term of computational complexity but it is more complex than MNS method. Table 5.5 provides some examples for the values of the window lengths used in MNS and OMNS in function of the system parameters q, p and M.

	Channel parameters	N_{OMNS}	$N_{MNS}=pM+1$
Fig. 5.6 (a)	10×1 and $M=2$	$(1,1,1,1,1,1,1,2,4)$	3
Fig. 5.6 (b)	10×2 and $M=2$	$(1,1,1,1,1,2,2,4)$	5
Fig. 5.5 (a)	4×1 and $M=2$	$(1,2,4)$	3
Fig. 5.5 (b)	4×2 and $M=2$	$(3,7)$	5

Table 5.5: Channel parameters and length of the processing window used in experiments.

5.6 Efficient implementation

As we have seen before, the main advantage of the MNS-Like methods is that the large matrix eigendecomposition is avoided and the noise vectors are computed in a parallel scheme as the least eigenvectors of covariance matrices corresponding to the $(p+1)$-tuple of system outputs.

Even-though, computing 1 single eigenvectors costs $\mathcal{O}(n^2)$ (n being the size of considered matrices), we found out that existing algorithms (e.g., [5, 65]) for extracting minor subspaces or the minor eigenvectors are un-efficient and slowly convergent in comparison with those dedicated to principal subspaces or principal eigenvectors.

For this reason, we propose here to compute from the beginning the inverse matrices using RLS-type technique followed by a power method to extract the principal eigenvectors of each of the considered matrices. Therefore, using Schur inversion lemma, the first step of MNS algorithm is replaced by the following:

For $t = 1, \cdots, T-N+1$

$$\begin{aligned} \mathbf{z}_i(t) &= \mathbf{P}_i(t-1)\mathbf{y}_i(t) \\ \mathbf{P}_i(t) &= \mathbf{P}_i(t-1) - \frac{\mathbf{z}_i(t)\mathbf{z}_i^*(t)}{1+\mathbf{y}_i^*(t)\mathbf{z}_i(t)} \end{aligned} \tag{5.-3}$$

where $\mathbf{P}_i$ represents the inverse matrix of $\boldsymbol{R}_N^i$. The initialization is done by choosing $\mathbf{P}_i(0) = \alpha\mathbf{I}$, α being a small positive scalar[4]. The least eigenvector of $\boldsymbol{R}_N^i$ becomes the principal eigenvector of $\mathbf{P}_i$ that can be computed using power iterations according to:

For $k \geq 1$

$$\begin{aligned} \mathbf{w}_i(k) &= \mathbf{P}_i\mathbf{w}_i(k-1) \\ \mathbf{w}_i(k) &:= \mathbf{w}_i(k)/\|\mathbf{w}_i(k)\| \end{aligned}$$

where $\mathbf{w}_i(k)$ represents the desired eigenvector estimate at the kth iteration. The initialization vector $\mathbf{w}_i(0)$ is chosen randomly.

Remark : To solve equation 3 in the single input case, one can use a similar RLS+power iteration algorithm applied to the quadratic form matrix of criterion 3. The latter can be rewritten as:

$$\|\mathcal{T}_N^*(\mathbf{H})\mathbf{V}_n\|^2 = \mathbf{H}^* \left(\sum_{i=1}^{q} \mathcal{D}_N(\mathbf{v}_i)^* \mathcal{D}_N(\mathbf{v}_i) \right) \mathbf{H} \tag{5.-5}$$

[4]Note that the eigenvectors of $\boldsymbol{R}_N^i$ coincide with those of $\boldsymbol{R}_N^i + \alpha\mathbf{I}$.

where

$$\mathcal{D}_N(\mathbf{v}_i) = [\mathcal{D}_N(\mathbf{v}_{i,1}), \cdots, \mathcal{D}_N(\mathbf{v}_{i,q})]$$

$$\mathcal{D}_N(\mathbf{v}_{i,j}) = \begin{bmatrix} v_{i,j}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ v_{i,j}(N-1) & & & v_{i,j}(0) \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & v_{i,j}(N-1) \end{bmatrix}.$$

However, in the multi-input case, the solution of 3 depends on the considered constraint and thus it cannot be computed necessarily as a least eigenvector of a given matrix [50].

Convergence rate : The average convergence rate of the RLS + power iteration method is high (typically 5 to 10 iterations are sufficient) and overcomes the one of the optimal step size gradient method in [65]. This has been observed in different scenarios (see figures 5.10-5.13) at low and high SNRs and for single and multiple ($p = 2$) input cases. Also, we have observed that the average convergence rate remains quite high (see figure 5.13) even if the noise subspace dimension is larger than one (this is the case if we choose $N > pM + 1$)[5].

5.7 Simulation results and discussions

In this section, first the performance of the SS method is compared with that of the MNS, SMNS and OMNS methods via several simulations. Then we compare the convergence rate of efficient implementation proposed in 5.6.

For the first part, we consider two different case with $p = 1$ and $p = 2$ inputs where each input sequence is an i.i.d., zero-mean, unit-variance QAM4 process.

All of MNS methods estimate the polynomial matrix $\mathbf{H}(z)$ up to a $p \times p$ constant matrix $\mathbf{Q}$. The output observation noise is a sequence of i.i.d., zero-mean, Gaussian variables and the number of samples is held constant ($T = 250$). For each experiment $N_r = 100$ independent Monte-Carlo runs are performed.

[5]Note that the convergence rate of power methods is exponential in terms of the ratio of the two largest eigenvalues (λ_2/λ_1) of $\mathbf{P}_i$.

The performance is measured by the mean-square-error (MSE) defined by:

$$MSE = [\sum_{r=1}^{N_r} \|\widehat{\mathbf{H}}_r \mathbf{Q}_r - \mathbf{H}\|^2 / N_r]^{\frac{1}{2}} \tag{5.-7}$$

Where $\widehat{\mathbf{H}}_r$ is an estimate of $\mathbf{H} \triangleq [\mathbf{H}^T(0), ..., \mathbf{H}^T(M)]^T$ at the r-th run, and $\mathbf{Q}_r$ is chosen so that $\|\widehat{\mathbf{H}}_r \mathbf{Q}_r - \mathbf{H}\|$ is minimum (This is to get rid of the constant matrix indeterminacy).

The SNR value in the simulations is equal to :

$$\text{SNR} = 10\log_{10} \frac{||\mathbf{H}||^2 \sigma_s^2}{q\sigma_n^2}.$$

In the figures the MSE of channel parameter estimates are plotted against the SNR, defined as the inverse noise power.

Channel Characteristics Two different type of MIMO channels have been used in our simulations :

The first one consists of using a physical propagation model and the second one consists of choosing random complex Gaussian channel coefficients that are changed for each monte-carlo run.

In the following we describe in detail the propagation model. In this model, the channel transfer function associated with the first output corresponds to the same impulse response. This can be written as (for the case of $p = 2$):

$$h_{1,1}(k) = h_{1,2}(k) = \sum_{t=0}^{L-1} \lambda_t g(kT_s - \tau_t) \tag{5.-7}$$

L denotes the number of paths and $g(t)$ is the raised cosine function with the roll-off factor equal to $1/2$. Then it is delayed and sampled at the rate of 270 kb/sec ($T_s = 3.7$ μs). The resulted channel impulse response is windowed such that the polynomial degree is M. τ_t denotes the delay and λ_t the attenuation. Attenuation is considered equal to 0 dB for λ_0 and -5 dB for all of the other paths and the delay τ_t is a multiple of path number. i.e. $\tau_t = t \times 3.2\mu$ sec (for $t = 0, \ldots, L-1$).

Other channel transfer functions are generated by assuming a plane propagation model of each path with corresponding electric angles uniformly distributed in $[0, \pi/2]$, for $l = 2, \cdots, q$

$$h_{k,l}(z) = \sum_{t=0}^{M} h_{1,l}(t) e^{jl\theta_{t,l}} z^{-t} \quad \text{with} \quad \theta_{t,l} \in [0, \pi/2] \tag{5.-7}$$

Figures 5.3-5.7 are plotted using the physical propagation channel. For figures 5.7-5.9 the channel coefficients are generated randomly, following a complex Gaussian distribution for each channel coefficient at each run. This helps us to compare the mean performance of MNS, SMNS, SS and OMNS methods.

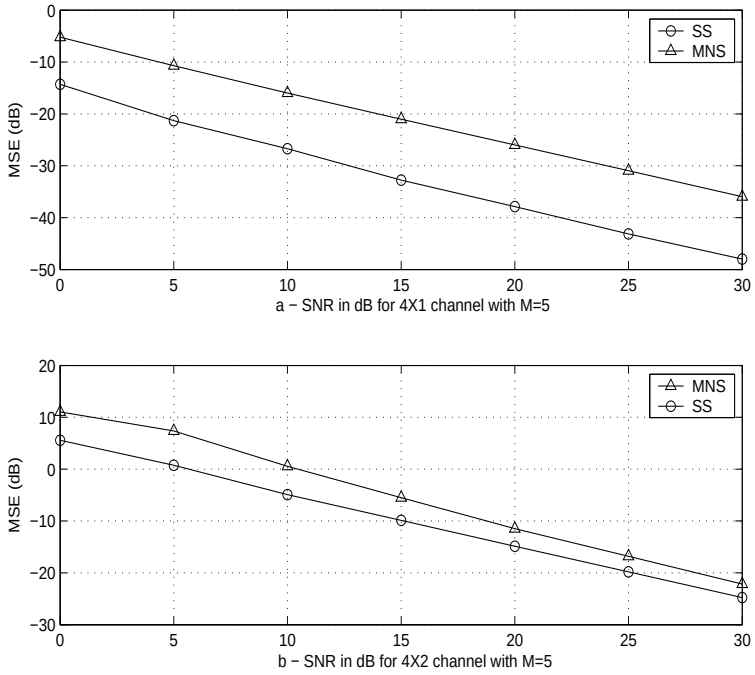

Figure 5.3: Performance of SS and MNS methods.

Figures 5.3 and 5.4 compare performance of MNS and standard SS method for different choices of channel parameters q, p and M. These figures show the fact that SS method has better performance than MNS method.

Indeed, in MNS noise polynomial vectors are obtained using only $p+1$ outputs for each of them (using $q-p$, $p+1$ tuples in parallel scheme), while in SS method each noise vector is computed from all the q system outputs (with $q > p$).

This fact leads to an improved (a more robust) channel estimation especially when the number of system outputs is much larger than the number of system inputs. However, the computational complexity of MNS method is much less than that of SS method. For example the computational complexity in figure 5.4 with a two-input ten-output (10×2) channel of degree $M = 2$ and $N = pM + 1 = 5$, is $\mathcal{O}((q-p)(p+1)^2N^2) = \mathcal{O}(1800)$ flops for MNS comparing with that of the SS which is almost $\mathcal{O}((qN)^3) = \mathcal{O}(125000)$ flops.

Figures 5.5 and 5.6 illustrate the performance of OMNS and MNS method

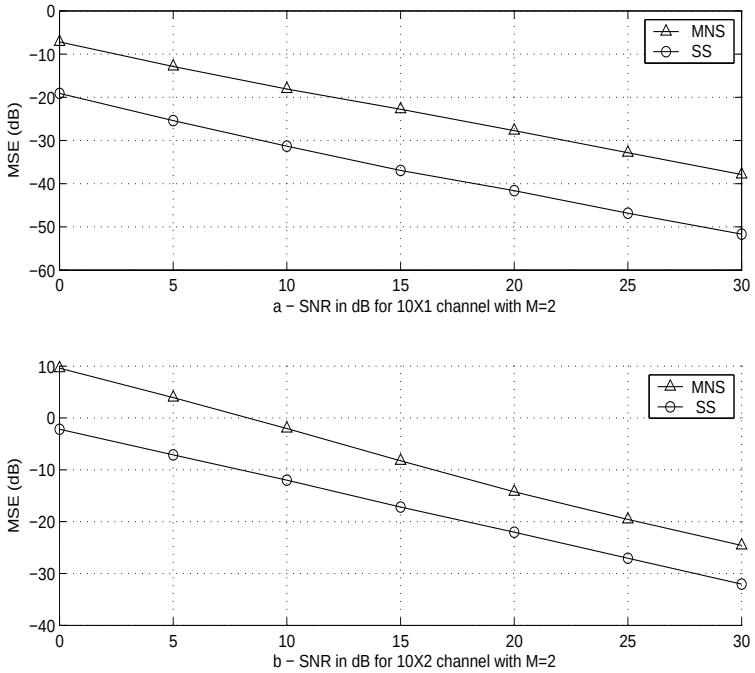

Figure 5.4: Performance of SS and MNS methods.

for different values of channel parameters q, p, M and window length reported also in Table 5.5.

As mentioned before, noise polynomial vectors in MNS method are obtained using only $p+1$ outputs for each of them, while in OMNS method each noise vector is computed from all the q system outputs in an iterative manner. This leads to an improved channel estimation especially when the number of system outputs is much larger than the number of system inputs. Furthermore, the orthogonality of noise subspace might improve the quality of the parameter estimation.

However, it is shown in Figure 5.5 that for $q-p$ small and large channel degree, the performance of OMNS is slightly deteriorates in comparison with that of MNS. This is possibly due (but need to be certified by a theoretical study) to the large window sizes that are needed to compute the OMNS basis leading to a large number of parameters (here the noise vector coefficients) to be estimated.

We have also seen in section 5.5 that the computational complexity of OMNS method is less than that of SS method. It is shown that the computational complexity of OMNS method is more than that of MNS.

For example, in figure 5.6 for a one-input ten-output propagation channel

(10×1) of degree $M = 2$, with the values of window size reported in Table 5.5, MNS method requires $\mathcal{O}((q-p)(p+1)^2N^2) = \mathcal{O}(2070)$ flops comparing with $\mathcal{O}(\sum_{i=1}^{q-p}(N_i q)^2) = \mathcal{O}(3600)$ flops required by OMNS method.

Figure 5.7 compares performance of SS, MNS and SMNS methods. It is shown that the SMNS method is more robust than the MNS method at the cost of a slight increase of the computational cost in comparison with MNS [6].

However, SMNS is a good alternative of MNS method because all different system outputs are used identically in order to compute the noise subspace. Moreover, for a fixed number of sensors q, the larger the number of sources is, the smaller is the number of noise vectors used by the MNS for channel identification. Even though, this is theoretically justified, the estimation performance is affected seriously as the parameter estimation problem becomes harder and thus requires 'more efforts' when the number of sources increases.

Figures 5.8 and 5.9 present the performance results for the same parameters as in figures 5.3 and 5.4 for random normalized channels. As we mentioned before the OMNS algorithm provide better results for the large number of sensors (large values of q).

Figures 5.10-5.13 show the convergence rate of the RLS+power iteration method and the optimal step size gradient method in [65] in different scenarios corresponding to the single input case with an SNR of 30dB, the single input case with an SNR of 10dB, the 2-input case with an SNR of 30dB, and the single input case when the noise subspace dimension is strictly larger than one, respectively. In all these scenarios, one can observe a relatively high convergence rate of the proposed algorithm.

5.8 Conclusions and further work

This chapter is devoted to the MNS (Minimum Noise Subspace) methods for the blind identification of the MIMO systems. These methods are computationally more efficient than the original subspace method where all of the noise vectors must be computed. Moreover, noise subspace can be computed in the MNS in a parallel scheme. We have also introduced some new extensions of this method called as OMNS and SMNS methods. Computational complexity of proposed algorithms are also evaluated.

[6] $\mathcal{O}((q-p)(p+1)^2N^2) = \mathcal{O}(900)$ flops for MNS and $\mathcal{O}(q(p+1)^2N^2) = \mathcal{O}(1350)$ flops for SMNS.

It has been shown via simulations that the OMNS method is more performant for large number of sensors and SMNS method works better because all of the system outputs are used.

We have also proposed an efficient RLS-like implementation of MNS algorithms to extract minor subspaces of covariance matrices.

MNS-Like algorithms are promising and some more topics can be studied later :

Beyond the minimum Under additional assumptions on the channel transfer functions, we can use less than $q-p$ noise vectors (recall that without any additional assumptions, $q-p$ is the minimum number of noise vectors needed to achieve unique channel identification). Just to present the basic ideas, let consider the case $p=1$ and assume that the channels are such that $h_i(z)$ and $h_j(z)$ are co-prime [7] for all $i \neq j$. In that case, only $\lceil \frac{q}{2} \rceil + 1$ noise vectors are sufficient for unique channel identification where $\lceil . \rceil$ represents the integer rounding towards plus infinity. In fact, we can observe that each of the following system output pairs: $(1,2),(3,4),\cdots,(q-1,q)$ allows a unique identification of the corresponding channel polynomials, i.e., $(\mathrm{h}_1(z),\mathrm{h}_2(z))$, $(\mathrm{h}_3(z),\mathrm{h}_4(z)),\cdots,(\mathrm{h}_{q-1}(z),\mathrm{h}_q(z))$ up to unknown scalar constants $\alpha_1,\alpha_2,\cdots,\alpha_{\lceil \frac{q}{2} \rceil}$, respectively. To get rid of these scalar constant indeterminacy, we need to use one extra relation that links all system outputs together.

The MNS-like algorithm developed in this chapter can be extended to other applications like as DOA estimation [8] and array calibration [66, 67].

The asymptotic performance analysis needs to be performed for the MIMO system identification in order to analyze the behavior of these methods and the effect of each parameter on their overall performance.

[7] This is a stronger assumption than that used previously when only *global* co-primeness is required, i.e. $h_1(z),h_2(z),\cdots,h_q(z)$ do not share common zeros.

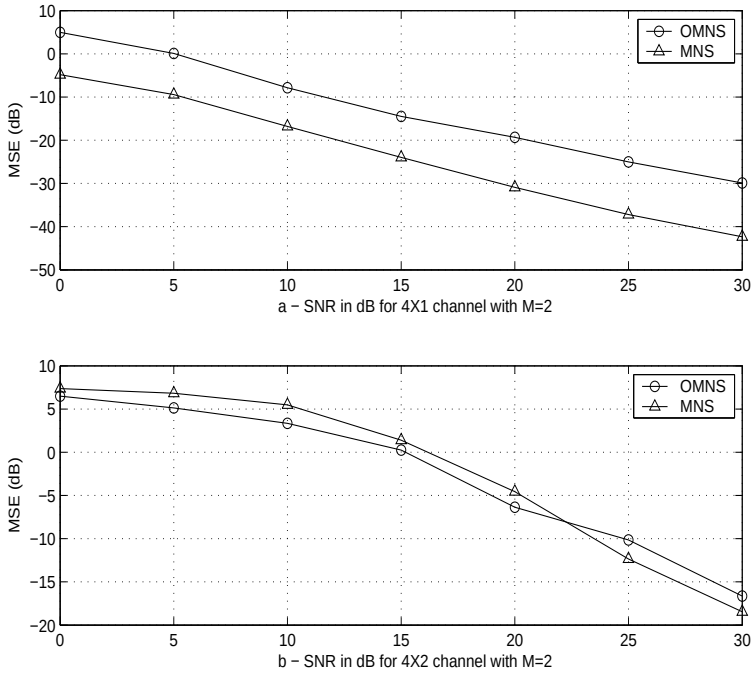

Figure 5.5: Performance of MNS and OMNS methods.

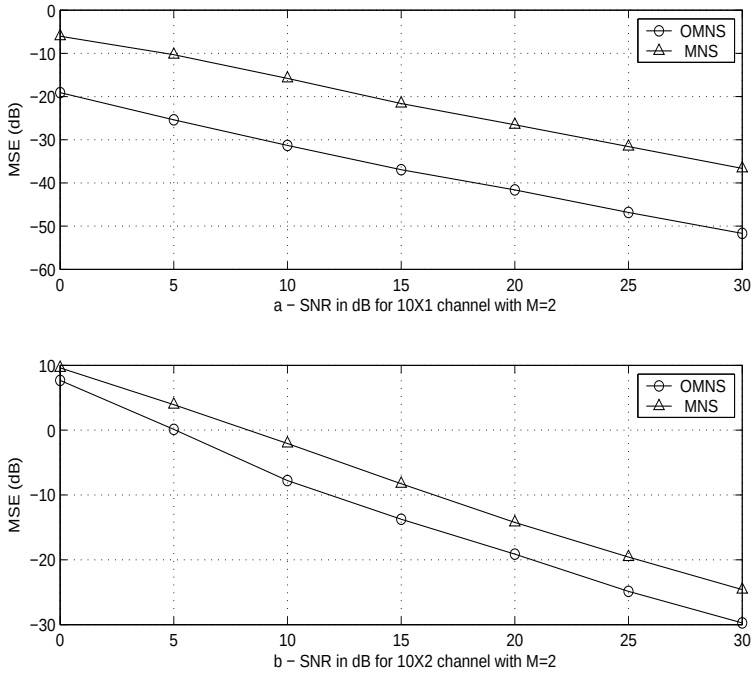

Figure 5.6: Performance of MNS and OMNS methods.

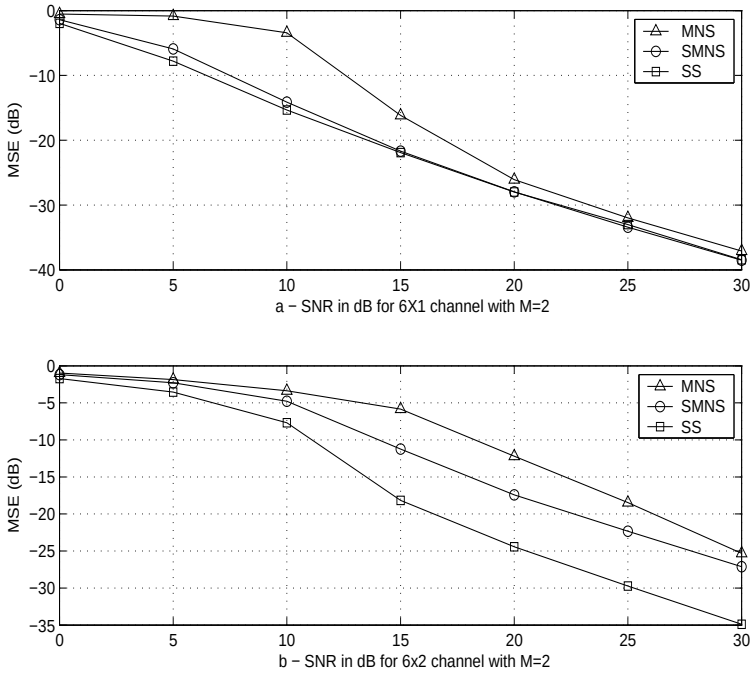

Figure 5.7: Performance of SS, MNS and SMNS methods.

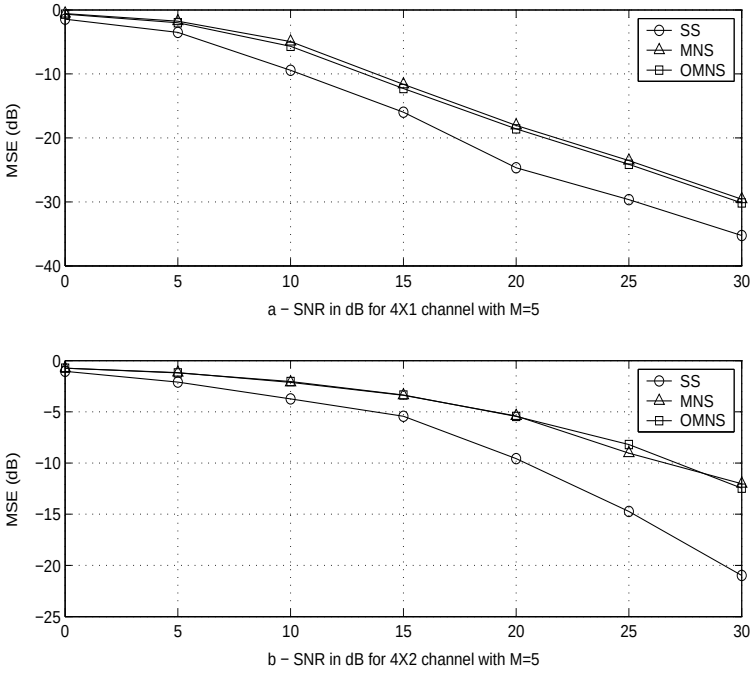

Figure 5.8: Performance of SS, MNS and OMNS methods.

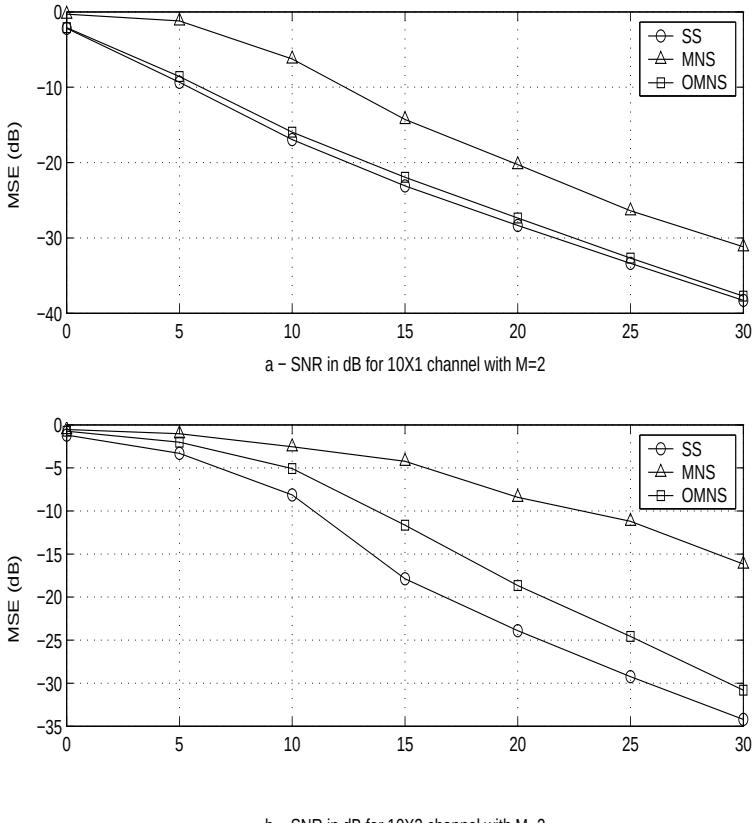

Figure 5.9: Performance of SS, MNS and OMNS methods.

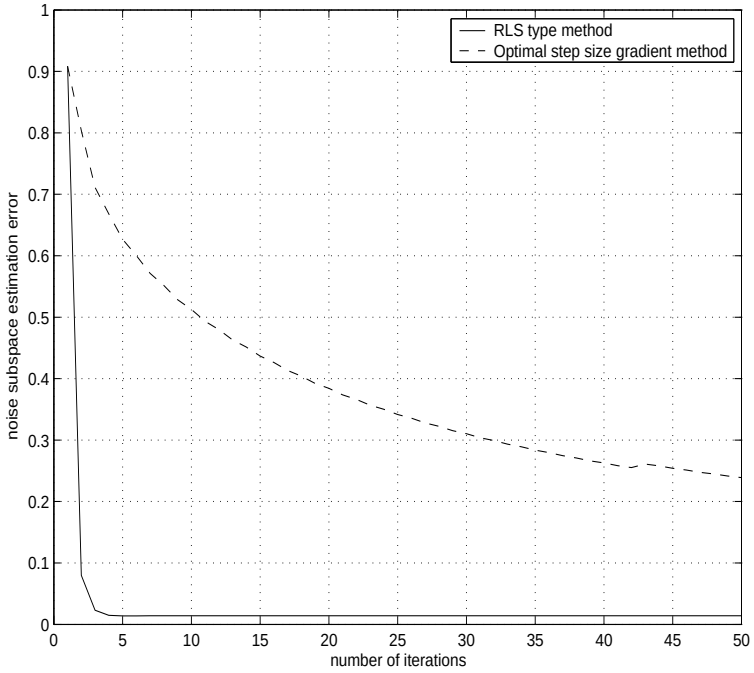

Figure 5.10: Convergence rate comparison for $p = 1$, $q = 2$, $N = pM + 1$ and $SNR = 30$ dB.

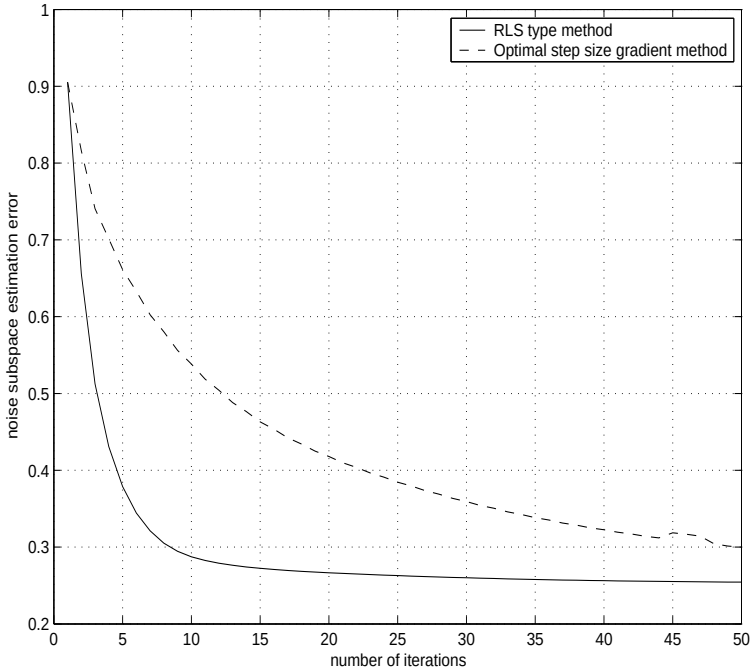

Figure 5.11: Convergence rate comparison for $p = 1$, $q = 2$, $N = pM + 1$ and $SNR = 10$ dB.

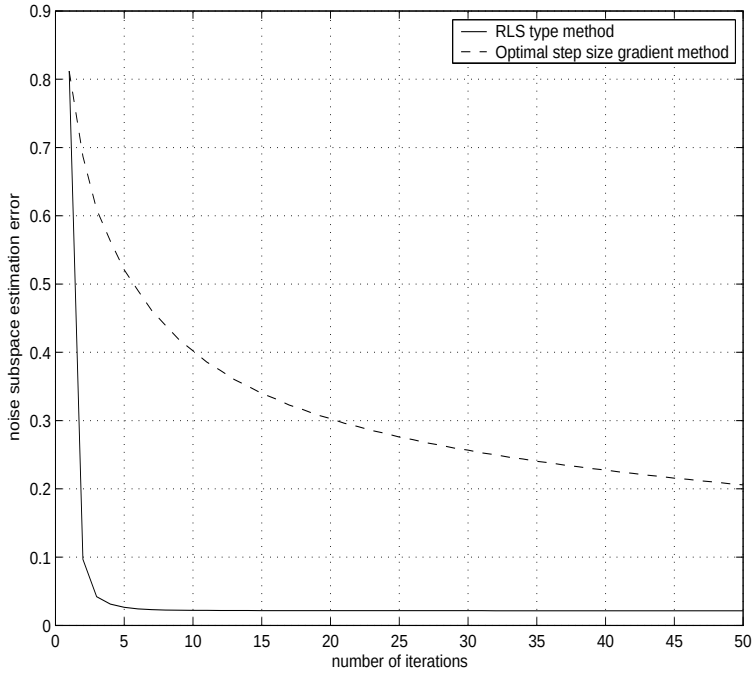

Figure 5.12: Convergence rate comparison for $p = 2$, $q = 3$, $N = pM + 1$ and $SNR = 30$ dB.

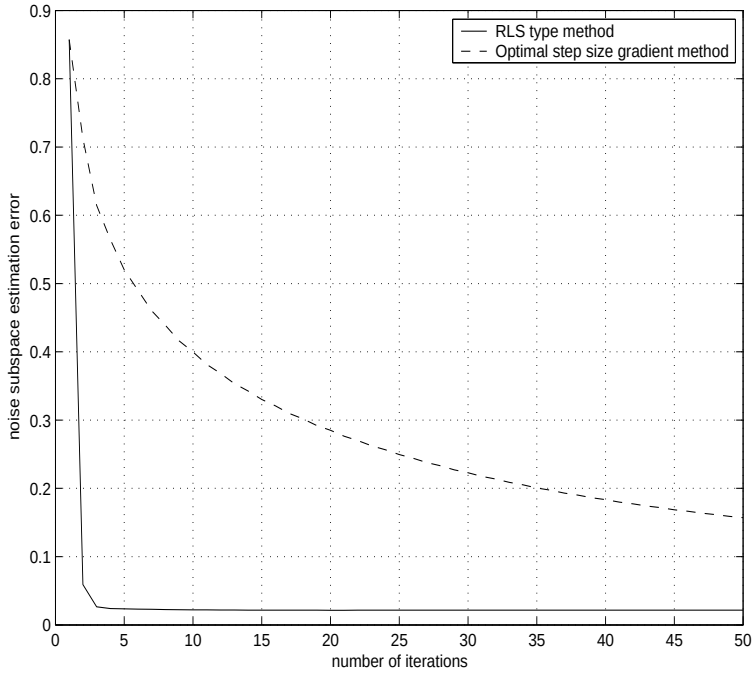

Figure 5.13: Convergence rate comparison for $p = 1$, $q = 2$, $N = pM + 2$ and $SNR = 30$ dB.

Chapter 6

Identification algorithms robust to channel order overestimation

6.1 Overcome a well-known limitation : channel order overestimation

It is shown that subspace method requires the exact prior knowledge of the channel order [9], otherwise it fails : When the system order is known in a (Single-Input Multi-Output) SIMO system, the channel can be estimated up to a constant scalar factor. When an overestimation of system order occurs, there is a linear space of all possible solutions.

In order to overcome with this limitation, and based on a simple observation that is detailed in 6.3, we propose two categories of algorithms robust to channel order overestimation. Both algorithms use side information on the input data; temporal deconvolution of the input sequence and constant modulus property, respectively.

Least Squares Fitting approach This approach, presented in section 6.4, consists of choosing the appropriate solution which corresponds to the best fit of the estimated channel and its relative equalized signal to the observed signal. Based on this idea, we propose two block algorithms for robust channel identification. Global convergence of the algorithms are guaranteed over a few number of iterations due to the efficient initialization of algorithms detailed in

section 6.4.3.

Constant Modulus approach The Constant Modulus (CM) approach is detailed in section 6.5. It consists of choosing the appropriate channel estimate (among the linear space of possible solutions of the SS criterion) which minimizes the Constant Modulus (CM) criterion of the equalized signal. Based on this idea and exploiting the relationship between Minimum Mean Square Error (MMSE) and CM equalizer, we propose a new algorithm for robust channel identification.

6.2 Overview of related works

For several years a lot of research efforts have been made in order to develop new channel identification techniques and algorithms robust to channel order overestimation.

Among other works we can mention a second order based approach in [38, 37, 39], where the authors propose a new algorithm robust to order overestimation based on the properties of the shifted version of the output correlation matrix.

Another second order based method has been proposed in [40]. In this approach, the authors propose an adaptive algorithm that minimizes an observation fitting cost function derived from observation and the expression of the linear equalizer of the channel.

The aforementioned algorithms propose new approaches to robust channel order overestimation in a general context using a least squares fitting or a correlation approach. However, the robustness problem of subspace criterion has been addressed in a few works.

In [48, 49], a robust blind channel identification approach for subspace algorithm has been proposed. In this approach the authors propose an adaptive order detection algorithm. In this proposition the subspace minimization criterion is solved under different minimization linear constraints over channel. Therefore, for an arbitrary order of channel, several possible subspace based estimates (under new linear constraints) and a measure of proximity of the estimates are computed. Then the best solution (the channel order) is the solution that maximize the measure of the proximity of solutions (see [49] for algorithm description.).

In [70] parametric robust subspace technique has been proposed. In this

approach the authors consider a specific pulse shape filter at transmitter and a specular model for the propagation channel. These assumptions lead to the minimization of a convex function over a few parameter for subspace criterion. It is shown in [70] that this approach gives consistent estimate of channel parameters when an overestimation of channel order occurred.

In practice, the exact channel order value is not known and it depends on the application and the environment. One existing approach to estimate the channel order consists of applying the Minimum Description Length (MDL) criterion [102] to the covariance matrix. However, it is known that the MDL criterion, tends to overestimate the exact channel order. That's why we focus on algorithms robust to order overestimation of channel.

In next section, we recall first the notations used in the rest of this chapter.

6.3 Notations and Main idea

As mentioned before the algorithms of this chapter are derived for the SIMO (Single-Input Multiple-Output) case [1]. In this section, we recall the system model and the related SIMO subspace method. Therefore, the system model presented in 4.3 can be rewritten as :

$$\begin{aligned} \mathbf{y}(n) &\triangleq [\mathbf{h}(z)]\mathbf{s}(n) + \mathbf{n}(n) \\ &= \sum_{k=0}^{M} \mathbf{h}(k)\mathbf{s}(n-k) + \mathbf{n}(n) \end{aligned} \tag{6.0}$$

where $\mathbf{h}(z) = \sum_{k=0}^{M} \boldsymbol{h}(k) z^{-k}$ is a $q \times 1$ polynomial transfer function modeling the channel and $\{\mathbf{n}(n)\}$ is a measurement noise. It is assumed that $\{\mathbf{n}(n)\}$ is white both temporally and spatially ($E(\mathbf{n}(n)\mathbf{n}^T(n)) = \sigma^2 \mathbf{I}_q$, where σ^2 is unknown), and is independent from the symbol sequence.

The general hypotheses over the channel, expressed in section 4.3, are reduced to the fact that $\mathbf{h}(z)$ must be of full-rank for each z. It can be written as :

$$\mathbf{h}(z) \neq 0 \text{ for each } z \qquad \deg(\mathbf{h}(z)) = M \tag{6.0}$$

it has been shown in [94], that $\mathbf{h}(z)$ and σ^2 are identifiable from a finite number

[1] The case of MIMO systems is more complex and might be developed later.

of auto covariance coefficients.

$$\begin{aligned}\mathbf{Y}(n) &= [\mathbf{y}^T(n), ..., \mathbf{y}^T(n-N+1)]^T \\ &= \mathcal{T}_N(\boldsymbol{h})\mathbf{S}_{M+N}(n) + \mathbf{N}(n) \qquad (6.0)\end{aligned}$$

where $\mathbf{S}_{M+N}(n) = [s(n), ..., s(n-N-M+1)]^T$ and $\mathbf{N}(n) = [\mathbf{n}^T(n), ..., \mathbf{n}^T(n-N+1)]^T$. N is a chosen processing window length and $\mathcal{T}_N(\boldsymbol{h})$ is a $qN \times (N+M)$ block Sylvester matrix associated to $\boldsymbol{h} \triangleq [\boldsymbol{h}^T(0), ..., \boldsymbol{h}^T(M)]^T$.

Remark : It is important to notice that the observation vector $\mathbf{Y}(n)$ is obtained by separating first all of the q-outputs of the system at desired time instant $t = n, \cdots, n-N+1$. This is different form the notations used in chapter 4 [2]. However, as mentioned before, the subspace properties remain unchanged using different notations.
The covariance matrix of $\mathbf{Y}(n)$ may be expressed as:

$$\begin{aligned}\boldsymbol{R}_N &= E(\mathbf{Y}(n)\mathbf{Y}^T(n)) \qquad (6.1)\\ &= \mathcal{T}_N(\boldsymbol{h})\mathcal{S}\mathcal{T}_N^T(\boldsymbol{h}) + \sigma^2\boldsymbol{I}_{qN} \qquad (6.2)\end{aligned}$$

where $\mathcal{S} \triangleq E(\mathbf{S}_{M+N}(n)\mathbf{S}_{M+N}^T(n))$.

The first term in the right hand side of (6.1) is singular as soon as $qN > (N+M)$ (this condition is assumed to hold throughout). In this case, the noise variance σ^2 is the smallest eigenvalue of $\boldsymbol{R}_N$. The eigenspace associated to σ^2 is referred to as the *noise subspace*. $\mathcal{S}$ is assumed to be positive-definite and thus the noise subspace is the orthogonal complement of Range($\mathcal{T}_N(\boldsymbol{h})$), the *signal subspace*.

The eigen-decomposition of $\mathbf{R}_N$ allows to identify the noise subspace Range$(\mathcal{T}_N(\boldsymbol{h}))^\perp$. Denote by Π_N the orthogonal projection matrix onto Range$(\mathcal{T}_N(\boldsymbol{h}))^\perp$.

In order to estimate the channel, as in (4.3), we have to characterize the set of all polynomials $\mathbf{h}'(z)$ of degree M', satisfying $\Pi_N\mathcal{T}_N(\boldsymbol{h}') = 0$. The following Theorem holds:

Theorem 5 *[34, 50] Assume (6.3) holds and $N \geq M$. Denote by Π_N the orthogonal projection matrix onto the noise subspace of $\mathbf{R}_N$. The matrix equation*

$$\Pi_N\mathcal{T}_N(\boldsymbol{h}') = 0 \qquad (6.2)$$

[2] In chapter 4, the observation vector is obtained over output number i, for all time instants $t = n, \cdots, n-N+1$.

does not admit any solution if $M' < M$ (where $M' = deg(\mathbf{h}'(z))$. If $M' \geq M$, the solutions of (5) are of the form $\mathbf{h}'(z) = r(z)\mathbf{h}(z)$ where $r(z)$ is an arbitrary scalar polynomial of degree $M' - M$.

Matrix equation (5) is solved in a least squares sense. Since the Sylvester matrix depends linearly on its parameters, the least squares criterion can be rewritten as a quadratic form

$$\mathrm{Tr}(\mathcal{T}_N^H(\boldsymbol{h}')\Pi_N\mathcal{T}_N(\boldsymbol{h}')) = \boldsymbol{h}'^H\mathcal{Q}_N\boldsymbol{h}' \tag{6.2}$$

$\mathcal{Q}_N$ denotes a $q(M'+1)\times q(M'+1)$ symmetric matrix. When the degree M of system is known or correctly estimated, $\mathbf{h}'(z)$ can be written as $\mathbf{h}'(z) = r\mathbf{h}(z)$ (where r is a complex scalar) according to Theorem 5. Therefore, the channel can be estimated up to a scale factor by minimizing in $\boldsymbol{h}'$ the above mentioned criterion under a suitable constraint. On the other hand, when $M' > M$, there are many solutions of the form $\mathbf{h}'(z) = r(z)\mathbf{h}(z)$. In this case, we have the following result:

Proposition 1 *Under the above mentioned conditions, matrix $\mathcal{Q}_N$ is singular with a null space of dimension $d = M' - M + 1$.*

Consequently, the linear space of solutions can be obtained using $L \geq d$ eigenvectors, represented here by matrix $\boldsymbol{U}_L$, associated to the L least eigen-values of $\mathcal{Q}_N$. Mathematically it can be expressed as

$$\boldsymbol{h}'_{\mathbf{v}} = \boldsymbol{U}_L\mathbf{v} \tag{6.2}$$

where $\boldsymbol{U}_L$ denotes a $q(M'+1)\times L$ matrix, $\mathbf{v}$ stands for an L-dimensional vector to be estimated and

$$\boldsymbol{h}'_{\mathbf{v}} = [\boldsymbol{h}'^T_{\mathbf{v}}(0), ..., \boldsymbol{h}'^T_{\mathbf{v}}(M')]^T.$$

6.4 Least Squares Fitting Approach

As mentioned before, the key idea of this section is to find the L-dimensional vector $\mathbf{v}$ which minimizes the distance between the estimated and the observed signal in a least squares sense. The observation fitting cost function can be written in two ways

- It can be expressed as a function of only $\mathbf{v}$. In this case, both channel and equalizer expression are written as a function of $\mathbf{v}$. The equalizer corresponds to the MMSE filter.

- It can be obtained as a function of vector $\mathbf{v}$ and an equalizer $\mathbf{w}$. A bilinear approach, as in [40], should be adopted to estimate two parameters alternatively.

Minimization of cost functions is performed using gradient-based or newton-based algorithms. The initial point of these algorithms are chosen in order to guarantee the convergence to the minimum in a few iterations. Efficient initialization for both algorithms is discussed later in section 6.4.3.

6.4.1 Non-linear Optimization using MMSE equalizer

We assume here that the source signal is temporally white: i.e. $E\big(\mathbf{s}(n)\mathbf{s}^*(m)\big) = \sigma_p^2\delta(n-m)$. We consider here an MMSE equalizer as a function of $\mathbf{v}$. In this case, vector $\mathbf{v}$ corresponding to the desired solution is estimated by minimizing the cost function given by :

$$\min_{\mathbf{v}} \mathcal{J}(\mathbf{v}) = \|\mathbf{Y}_{obs} - \mathcal{T}(\boldsymbol{h}'_{\mathbf{v}})\hat{\mathbf{S}}_{\mathbf{v}}\|^2 \tag{6.2}$$

where $\mathbf{Y}_{obs} = [\mathbf{y}^T(T-\tau), ..., \mathbf{y}^T(N)]^T$ is the block vector of observations (T being the sample size), with an equalization delay of $\tau = M' - 1$[3] and $\hat{\mathbf{S}}_{\mathbf{v}}$ is the estimated input sequence using the MMSE equalizer associated to $\boldsymbol{h}'_{\mathbf{v}}$ according to

$$\begin{aligned} \hat{\mathbf{S}}_{\mathbf{v}} &= \tilde{\mathbf{Y}}(\hat{\mathbf{R}}_N^{-1} \begin{bmatrix} \boldsymbol{h}'_{\mathbf{v}}(M') \\ \vdots \\ \boldsymbol{h}'_{\mathbf{v}}(0) \\ \mathbf{0}_{q(N-M'-1),1} \end{bmatrix}) \\ &= \tilde{\mathbf{Y}}\mathbf{M}\mathbf{v} \end{aligned} \tag{6.2}$$

where

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y}^T(T) \\ \vdots \\ \mathbf{Y}^T(N+\tau-M'+1) \end{bmatrix} \tag{6.2}$$

[3]The delay value can be chosen arbitrary.

and

$$\mathbf{M} = \hat{\mathbf{R}}_N^{-1} \begin{bmatrix} \boldsymbol{U}_{L,M'} \\ \vdots \\ \boldsymbol{U}_{L,0} \\ \mathbf{0}_{q(N-M'-1),L} \end{bmatrix}, \tag{6.2}$$

$\boldsymbol{U}_{L,i}$ denotes $q \times L$ matrix, corresponding to the i_th block of matrix $\boldsymbol{U}_L$ (i.e. $\boldsymbol{U}_L = [\boldsymbol{U}_{L,0}^T, ..., \boldsymbol{U}_{L,M'}^T]^T$). Using $\mathcal{T}(\boldsymbol{h}'_{\mathbf{v}}) = \sum_{i=1}^{L} \mathcal{T}_i v_i$, with

$$\mathcal{T}_i = \begin{bmatrix} \boldsymbol{U}_{L,0}(:,i) & \cdots & \boldsymbol{U}_{L,M'}(:,i) & & 0 \\ & \ddots & & \ddots & \\ 0 & & \boldsymbol{U}_{L,0}(:,i) & \cdots & \boldsymbol{U}_{L,M'}(:,i) \end{bmatrix}$$

the cost function (6.4.1) can be written as

$$\begin{aligned} \mathcal{J}(\mathbf{v}) &= \|\mathbf{Y}_{obs} - \sum_{i=1}^{L} \mathcal{T}_i \tilde{\mathbf{Y}} \mathbf{M} v_i \mathbf{v}\|^2 \\ &= \|\mathbf{Y}_{obs} - \mathbf{Q}(\mathbf{v} \otimes \mathbf{v})\|^2 \end{aligned} \tag{6.3}$$

with $\mathbf{Q} = [\mathcal{T}_1 \tilde{\mathbf{Y}} \mathbf{M}, ..., \mathcal{T}_L \tilde{\mathbf{Y}} \mathbf{M}]$ and $\otimes$ denotes the kronecker product.

6.4.1.1 Gradient Algorithm

Criterion $\mathcal{J}(\mathbf{v})$ is a 4th order polynomial function of $\mathbf{v}$ that is optimized via a gradient algorithm. Gradient algorithm is based on the first order development of the cost function. It is written as

$$\mathbf{v}^{(t+1)} = \mathbf{v}^{(t)} - \lambda_{opt} \nabla \mathcal{J}(\mathbf{v}^{(t)}) \tag{6.1}$$

t denotes the index number and λ_{opt} is the optimal step size. Gradient of $\mathcal{J}(\mathbf{v})$ is driven by

$$\nabla\ \mathcal{J}(\mathbf{v}) = \begin{bmatrix} \hat{\mathbf{S}}_{\mathbf{v}}^T \mathcal{T}_1^T (\mathbf{Y}_{obs} - \mathbf{Q}(\mathbf{v} \otimes \mathbf{v})) \\ \vdots \\ \hat{\mathbf{S}}_{\mathbf{v}}^T \mathcal{T}_L^T (\mathbf{Y}_{obs} - \mathbf{Q}(\mathbf{v} \otimes \mathbf{v})) \end{bmatrix} \tag{6.1}$$

6.4.1.2 Optimal step size selection

In order to determine optimal step size λ_{opt}, the cost function $\mathcal{J}(\mathbf{v}^{(t+1)})$, is approximated by a quadratic function $q(\lambda)$:

$$q(\lambda) = a\lambda^2 + b\lambda + c \tag{6.1}$$

Let us consider the cost function $\mathcal{J}(\mathbf{v})$ at iteration $t+1$. It can be written as

$$\mathcal{J}_u(\lambda) = \mathcal{J}(\mathbf{v}^{(t+1)}) = \mathcal{J}(\mathbf{v}^{(t)} - \lambda \nabla \mathcal{J}(\mathbf{v}^{(t)})) \tag{6.1}$$

Parameter c is obtained using the criterion value at the current point (which corresponds to $\lambda = 0$) and b is obtained by derivating the criterion at the same point

$$\begin{aligned} c &= q(0) = \mathcal{J}(\mathbf{v}^{(t)}) \\ b &= \frac{dq(\lambda)}{d\lambda}\Big|_{\lambda=0} = \frac{d\mathcal{J}_u(\lambda)}{d\lambda}\Big|_{\lambda=0} = -||\nabla \mathcal{J}(\mathbf{v}^{(t)})||^2 \end{aligned} \tag{6.2}$$

In order to obtain a, we impose that the quadratic function $q(\lambda)$ at the point $\lambda_1 = \frac{-2c}{b}$ has the same value as $\mathcal{J}_u(\lambda)$ (see [100] for more details). Therefore, we have

$$a = \frac{1}{\lambda_1^2}[\, \mathcal{J}(\mathbf{v}^{(t)} - \lambda_1 \nabla \mathcal{J}(\mathbf{v}^{(t)})) + c] \tag{6.0}$$

Parameters a, b and c are obtained using only one additional computation of criterion $\mathcal{J}(\mathbf{v})$ at each iteration. The optimal step size which minimizes $q(\lambda)$ is therefore determined as

$$\lambda_{opt} = \frac{-b}{2a}.$$

6.4.2 Bilinear Optimization

The bilinear approach consists of minimizing the cost function $\mathcal{J}$ alternatively over $\mathbf{v}$ and $\mathbf{w}$, as in [40][4]:

$$\min_{\mathbf{v},\mathbf{w}} \mathcal{J}(\mathbf{v}, \mathbf{w}) = \|\mathbf{Y}_{obs} - \mathcal{T}(\boldsymbol{h}'_{\mathbf{v}})\hat{\mathbf{S}}_{\mathbf{w}}\|^2 \tag{6.0}$$

where $\mathbf{w}$ is an 'equalizer' vector and $\hat{\mathbf{S}}_{\mathbf{w}}$ an estimate of the input signal. It is known that the equalizer vector belongs to signal subspace. Consequently, we choose here an equalizer of the form $\hat{\mathbf{U}}_{\mathbf{s}}\mathbf{w}$, where $\hat{\mathbf{U}}_{\mathbf{s}}$ represents the estimated signal subspace. Equation (6.4.2) can be written as a function of both $\mathbf{v}$ and $\mathbf{w}$ according to:

$$\begin{aligned} \mathcal{T}(\boldsymbol{h}'_{\mathbf{v}})\hat{\mathbf{S}}_{\mathbf{w}} &= (\mathcal{T}(\boldsymbol{h}'_{\mathbf{v}})\tilde{\mathbf{Y}}\hat{\mathbf{U}}_{\mathbf{s}})\mathbf{w} \\ &= \mathbf{A}\mathbf{w} \\ &= \mathbf{B}\mathbf{v} \end{aligned} \tag{6.1, 6.2}$$

[4]Note that in [40], the least squares fitting criterion was function of the channel parameter vector $\boldsymbol{h}$ and MMSE equalizer $\mathbf{w}$. While here it is function of the *reduced* parameter vector $\mathbf{v}$ and a *reduced* equalizer vector restrained to belong to the signal subspace.

with $\mathbf{B} = [\mathcal{T}_1\hat{\mathbf{S}}_{\mathbf{w}}, ..., \mathcal{T}_L\hat{\mathbf{S}}_{\mathbf{w}}]^T$.

6.4.2.1 Newton Algorithm

Newton algorithm is based on the second order development of the cost function. The bilinear newton algorithm over two parameters $\mathbf{v}$ and $\mathbf{w}$ can be expressed as

$$\begin{aligned} \mathbf{w}^{(t+1)} &= \mathbf{w}^{(t)} - \lambda_w(\nabla^2\mathcal{J}(\mathbf{w}^{(t)}))^{-1}\nabla\mathcal{J}(\mathbf{w}^{(t)}) \\ \mathbf{v}^{(t+1)} &= \mathbf{v}^{(t)} - \lambda_v(\nabla^2\mathcal{J}(\mathbf{v}^{(t)}))^{-1}\nabla\mathcal{J}(\mathbf{v}^{(t)}) \end{aligned} \tag{6.2}$$

λ_w and λ_v denote two *small* step-sizes. The gradient of the cost function can be written as

$$\begin{aligned} \nabla\mathcal{J}(\mathbf{w}) &= -2\mathbf{A}^T(\mathbf{Y}_{obs} - \mathbf{A}\mathbf{w}) \\ \nabla\mathcal{J}(\mathbf{v}) &= -2\mathbf{B}^T(\mathbf{Y}_{obs} - \mathbf{B}\mathbf{v}) \end{aligned} \tag{6.1}$$

and the Hessian of the cost function is given by

$$\begin{aligned} \nabla^2\mathcal{J}(\mathbf{w}) &= 2\mathbf{A}^T\mathbf{A} \\ \nabla^2\mathcal{J}(\mathbf{v}) &= 2\mathbf{B}^T\mathbf{B}. \end{aligned} \tag{6.0}$$

6.4.3 Efficient initialization

It is known that non-linear and bilinear algorithms do not converge to the global minimum of the cost function if an efficient initialization is not performed. For example, in [71] a way of detecting the nature of the minimum of such cost functions has been proposed, but there is no indication on the way we can efficiently initialize these algorithms. Here, we propose an efficient way of initialization that guarantees the convergence of the algorithm to the global minimum.

Initialization of proposed algorithms is based on the fact that the multichannel structure ensures the existence of linear finite-length zero-forcing equalizer under classical assumptions [68]. To well initialize the bilinear optimization algorithm we chose $\tau = 0$ and $\mathbf{w}_0$ according to:

$$\hat{\mathbf{U}}_{\mathbf{s}}\mathbf{w}_0 \approx \hat{\mathbf{R}}_N^{-1}\begin{bmatrix} \boldsymbol{h}(0) \\ \mathbf{0}_{q(N-1),1} \end{bmatrix} \Leftrightarrow \hat{\mathbf{R}}_N\hat{\mathbf{U}}_{\mathbf{s}}\mathbf{w}_0 \approx \begin{bmatrix} \boldsymbol{h}(0) \\ \mathbf{0}_{q(N-1),1} \end{bmatrix} \tag{6.-2}$$

Therefore:

$$\bar{\mathbf{U}}_{\mathbf{s}}\mathbf{\Lambda}_{\mathbf{s}}\mathbf{w}_0 = \mathbf{0}_{q(N-1),1} \tag{6.-2}$$

where $\bar{\mathbf{U}}_{\mathbf{s}}$ is the sub-matrix of $\mathbf{U}_{\mathbf{s}}$ given by its last $q(N-1)$ rows. In practice, $\mathbf{w}_0$ is estimated as the least eigen-vector of $(\mathbf{\Lambda}_{\mathbf{s}}\bar{\mathbf{U}}_{\mathbf{s}}^H\bar{\mathbf{U}}_{\mathbf{s}}\mathbf{\Lambda}_{\mathbf{s}})$. This choice of $\mathbf{w}_0$ corresponds to a good initial point as shown by the following proposition:

Proposition 2 *In the noiseless case and for $N > (M+1)$, vector $\mathbf{w}_0$ solution of (6.4.3) corresponds to a zero-forcing equalizer of delay $\tau = 0$, i.e. $(\mathbf{U}_{\mathbf{s}}\mathbf{w}_0)^T\mathcal{T}_N(\boldsymbol{h}) = [\beta, 0, ..., 0]$, where β is a given scalar.*

Proof. Let us consider the *complete* $qN \times 1$ equalizer of channel noted $\tilde{\mathbf{w}}_0 = (\mathbf{U}_{\mathbf{s}}\mathbf{w}_0)$ without loss of generality. Therefore, we have to prove the following equation :

$$\bar{\boldsymbol{R}}_N\tilde{\mathbf{w}}_0 = \mathbf{0}_{q(N-1),1} \Leftrightarrow (\tilde{\mathbf{w}}_0)^T\mathcal{T}_N(\boldsymbol{h}) = [\beta, 0, ..., 0] \tag{6.-2}$$

$\bar{\boldsymbol{R}}_N$ denotes the $q(N-1) \times qN$ matrix that corresponds to the last $q(N-1)$ rows of the $qN \times qN$ covariance matrix $\boldsymbol{R}_N$. Following equations (6.1) and (6.1), $\bar{\boldsymbol{R}}_N$ can be written as:

$$\bar{\boldsymbol{R}}_N = E(\mathbf{Y}(n-1)\mathbf{Y}^T(n)) \tag{6.-2}$$

$\mathbf{Y}(n-1)$ is the $qN \times 1$ output observation vector, as in (6.1). It can be written as :

$$\mathbf{Y}(n-1) = \mathcal{T}_N(\boldsymbol{h})\mathbf{S}_{M+N}(n-1) \tag{6.-2}$$

Having these expressions in mind, we will first demonstrate direct part of the equation (6.4.3). We have :

$$\begin{aligned} \bar{\boldsymbol{R}}_N\tilde{\mathbf{w}}_0 &= \mathbf{0}_{q(N-1),1} \\ E(\mathbf{Y}(n-1)\mathbf{Y}^T(n))\tilde{\mathbf{w}}_0 &= \mathbf{0}_{q(N-1),1} \\ \mathcal{T}_N(\boldsymbol{h})E\big(\mathbf{S}_{M+N}(n-1)\mathbf{S}_{M+N}(n)^T\big)\mathcal{T}_N(\boldsymbol{h})^T\tilde{\mathbf{w}}_0 &= \mathbf{0}_{q(N-1),1} \end{aligned} \tag{6.-3}$$

$\mathcal{T}_N(\boldsymbol{h})$ is full columns rank. Therefore we have :

$$E\big(\mathbf{S}_{M+N}(n-1)\mathbf{S}_{M+N}(n)^T\big)\mathcal{T}_N(\boldsymbol{h})^T\tilde{\mathbf{w}}_0 = \mathbf{0}_{q(N-1),1} \tag{6.-3}$$

$E\big(\mathbf{S}_{M+N}(n-1)\mathbf{S}_{M+N}(n)^T\big)$ can be written as:

$$E\big(\mathbf{S}_{M+N}(n-1)\mathbf{S}_{M+N}(n)^T\big) = \mathbf{J}_{M+N}$$

where $\mathbf{J}_{M+N}$ is a $(M+N)\times(M+N)$ shifted unitary matrix expressed as

$$\mathbf{J}_{M+N} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

therefore we rewrite the equation (6.4.3) as :

$$\mathbf{J}_{M+N}\mathcal{T}_N(\boldsymbol{h})^T\tilde{\mathbf{w}}_0 = \mathbf{0}_{q(N-1),1} \tag{6.-3}$$

therefore

$$\tilde{\mathbf{w}}_0^T\mathcal{T}_N(\boldsymbol{h}) = [\beta, 0, ..., 0] \tag{6.-3}$$

Reverse part of demonstration is proved in the following:

$$\begin{aligned} \tilde{\mathbf{w}}_0^T\mathcal{T}_N(\boldsymbol{h}) &= [\beta, 0, ..., 0] \\ \tilde{\mathbf{w}}_0^T\mathcal{T}_N(\boldsymbol{h})\mathbf{S}_{M+N}(n) &= \beta s(n) \\ \tilde{\mathbf{w}}_0^T\mathbf{Y}(n) &= \beta s(n) \end{aligned} \tag{6.-4}$$

multiplying both sides of previous equation by $\mathbf{Y}^T(n-1)$ and taking the expectation on both parts we have:

$$\tilde{\mathbf{w}}_0^T E(\mathbf{Y}(n)\mathbf{Y}^T(n-1)) = \beta E(\mathbf{Y}^T(n-1)s(n)) \tag{6.-4}$$

$\mathbf{Y}^T(n-1)$ depends only on time instants $n-1, n-2, ...$, therefore the second part of the equation (6.4.3) is equal to zero. Consequently we obtain:

$$\bar{\boldsymbol{R}}_N\tilde{\mathbf{w}}_0 = \mathbf{0}_{q(N-1),1} \tag{6.-4}$$

□.

Remark 1 :In order to launch the non-linear optimization algorithm, the initial point $\mathbf{v_0}$ is obtained from $\mathbf{w_0}$ in the same way, even if this equalizer is obtained basically for zero-delay equalizer.

Remark 2 : To enhance the performance of bilinear optimization algorithm,

at the first iteration we choose a delay value equal to zero, in order to estimate a good initial point. Then we change the delay value to $M' - 1$ for the second iteration to obtain a better equalizer quality.

6.5 Constant Modulus Approach

In this section, we mention first a well known property concerning the CM criterion. Godard was the first who observed that the mean square error performance of CMA is close to that of the MMSE equalizer [45]. For several years, a lot of research efforts was made to confirm or prove that, under several conditions, the CM minima remain in the vicinity of MSE minima for different choices of delay and sign (see [32, 95] and [96]). Among other works, we can cite an approach used in [59]. It consists of plotting the contour of CM cost function in equalizer space and compare the location of CM minima and MMSE equalizer. This approach confirms the above mentioned result, under certain conditions.

As mentioned before, the main idea of this section is to find the L-dimensional vector $\mathbf{v}$, minimizing the CM cost function under the CM assumption i.e. $\forall n \quad |s(n)| = C$. Where $C > 0$ is a given constant. More precisely, given the channel estimate in (6.3), we can express an MMSE equalizer vector as a function of $\mathbf{v}$ according to :

$$\begin{aligned} \mathbf{w} &= \hat{\mathbf{R}}_N^{-1} \begin{bmatrix} \boldsymbol{h}'_{\mathbf{v}}(M') \\ \vdots \\ \boldsymbol{h}'_{\mathbf{v}}(0) \\ \mathbf{0}_{q(N-M'-1),1} \end{bmatrix} \\ &= \mathbf{W}\mathbf{v} \end{aligned} \tag{6.-4}$$

Where

$$\mathbf{W} \triangleq \hat{\boldsymbol{R}}_N{}^{-1} \begin{bmatrix} \boldsymbol{U}_{L,M'} \\ \vdots \\ \boldsymbol{U}_{L,0} \\ \mathbf{0}_{q(N-M'-1),L} \end{bmatrix}$$

$\boldsymbol{U}_{L,i}$ denotes the $q \times L$ matrix corresponding to the $i_$th block of matrix $\boldsymbol{U}_L$ (i.e. $\boldsymbol{U}_L = [\boldsymbol{U}_{L,0}^T, ..., \boldsymbol{U}_{L,M'}^T]^T$) and $\mathbf{v}$ is a given L-dimensional vector. The desired vector $\mathbf{v}$ associated to the desired channel estimate (the desired channel

estimate corresponds to

$$\boldsymbol{h}'_{\mathbf{v}} = \alpha[\mathbf{0}^T_{qk,1},\ \boldsymbol{h}^T,\ \mathbf{0}^T_{q(M'-M-k),1}]^T$$

for a given scalar constant α and a positive integer k) is obtained by minimizing the following CM criterion:

$$\begin{aligned} \min\ \mathcal{J}(\mathbf{w}) &= \min E(|\mathbf{w}^T\mathbf{Y}(n)|^2 - r)^2 \\ &= \min_{\mathbf{v}} E(\ |\mathbf{v}^T\boldsymbol{Z}(n)|^2 - r\ )^2 \end{aligned} \tag{6.-4}$$

where $\boldsymbol{Z}(n) = \mathbf{W}^T\mathbf{Y}(n)$ and r represents the dispersion constant.

In order to minimize equation (6.-3), we constrain $\mathbf{v}$ to be of unit norm and we use a parameterization based on the following result [47][5] :

lemma 1 *Each unit norm row vector can be represented as the last row of an orthogonal matrix* $\mathbf{P}$ *given by:*

$$\mathbf{P} = \prod_{1\le p\le it\_nb} \Big(\prod_{1\le i\le L-1} \boldsymbol{P}(\theta_p^i) \Big) \tag{6.-4}$$

where θ_p^i *are a set of rotation angles in* $]-\pi/2, \pi/2]$ *and*

$$\boldsymbol{P}(\theta_p^i) = \begin{bmatrix} \boldsymbol{I}_{i-1} & & & \\ & \cos(\theta_p^i) & \dots & -\sin(\theta_p^i) \\ & \vdots & \boldsymbol{I}_{L-i-1} & \vdots \\ & \sin(\theta_p^i) & \dots & \cos(\theta_p^i) \end{bmatrix} \tag{6.-4}$$

Consequently, we propose a recursive minimization algorithm, where at each step, the cost function (6.-3) is written as a function of rotation angle θ_p^i :

$$\min_{\theta} \mathcal{J}(\theta_p^i) = E(\ |\mathbf{v}_0^T(\boldsymbol{P}(\theta_p^i))^T\boldsymbol{Z}(n)|^2 - r\ )^2 \tag{6.-4}$$

$\mathbf{v}_0^T$ is a row vector of length L with all components equal to zero except the last one which is equal to one. This choice of $\mathbf{v}_0^T$ permits us to select the last row of orthogonal matrix $\boldsymbol{P}(\theta_p^i)$.

At each iteration, the angle θ_p^i that minimizes the cost function (6.5) is computed. The algorithm is stopped when $\boldsymbol{P}(\theta_p^i)$ are close to identity matrix for all $1 \le i \le L-1$. More precisely, we have the following iterative process :

[5]Note that constraining $\mathbf{v}$ to be of unit norm is equivalent to constrain $||\boldsymbol{h}'_{\mathbf{v}}|| = 1$ since $\mathbf{U_L}$ is unitary.

1. Initialization [6]:
$$\mathbf{v_0}^T = [\underbrace{0, ..., 0}_{L-1}, 1] \tag{6.-4}$$

2. For $i = 1, 2, ..., L-1$ and the current iteration p, find the rotation which minimizes the cost function :
$$\theta_p^i = \text{argmin}\{\mathcal{J}(\theta_p^i)\} \tag{6.-4}$$
The minimization details are discussed below.

3. Compute the new values of $\boldsymbol{Z}$ and $\mathbf{v}$:
$$\begin{aligned} \boldsymbol{Z} &:= \boldsymbol{P}(\theta_p^i)^T \boldsymbol{Z} \\ \mathbf{v} &:= \boldsymbol{P}(\theta_p^i)\ \mathbf{v} \end{aligned}$$
where $\boldsymbol{Z} \triangleq [\boldsymbol{Z}(N), ..., \boldsymbol{Z}(T)]$ (T being the sample size).

4. If θ_p^i for all $1 \leq i \leq L-1$ are close to zero, then stop. Else, $p = p+1$ and go to step 2.

Here, we describe how the cost function $\mathcal{J}(\theta_p^i)$ is minimized. This cost function can be written as (we replace the expectation by time averaging):
$$\begin{aligned} \mathcal{J}(\theta_p^i) &= \sum_n \left(|-\sin(\theta_p^i) Z^i(n) + \cos(\theta_p^i) Z^L(n)|^2 - r \right)^2 \\ &= \sum_n \left(\mathbf{u}^T \boldsymbol{y}^i(n) + \alpha^i(n) \right)^2 \\ &= \mathbf{u}^T \underbrace{\left(\sum_n \boldsymbol{y}^i(n) \boldsymbol{y}^i(n)^T \right)}_{\mathbf{G}} \mathbf{u} + 2 \underbrace{\left(\sum_n \alpha^i(n) \boldsymbol{y}^i(n) \right)}_{\mathbf{g}^T} \mathbf{u} \end{aligned} \tag{6.-6}$$

where [7]
$$\mathbf{u} = \begin{bmatrix} \cos(2\ \theta_p^i) \\ \sin(2\ \theta_p^i) \end{bmatrix} \tag{6.-7}$$

and
$$\begin{aligned} \boldsymbol{y}^i(n) &= \begin{bmatrix} \left(Z^i(n)^2 - Z^L(n)^2 \right)/2 \\ -Z^i(n) Z^L(n) \end{bmatrix}, \\ \alpha^i(n) &= \frac{Z^i(n)^2 + Z^L(n)^2}{2} - r. \end{aligned} \tag{6.-7}$$

[6] This initialization corresponds to choosing at first the channel estimate of the standard subspace algorithm that is given by the least eigenvector of $\mathcal{Q}_N$, i.e. the last column vector of $\mathbf{U_L}$.

[7] In (6.-5), we omit a constant term independent from θ_p^i.

where $Z^i(n)$ denotes the i-th entry of $\boldsymbol{Z}(n)$. Therefore, minimizing $\mathcal{J}(\theta_p^i)$ versus θ_p^i is equivalent to minimizing the following equation subject to $||\mathbf{u}||^2 = 1$.

$$\min_{||\mathbf{u}||^2=1} \tilde{\mathcal{J}}(\mathbf{u}) = \min_{||\mathbf{u}||^2=1} (\mathbf{u}^T \mathbf{G} \mathbf{u} + 2\mathbf{g}^T \mathbf{u}) \tag{6.-7}$$

In order to minimize equation (6.5), we use the method of Lagrange multipliers [12]. By zeroing the gradient of (6.5), we obtain the following expression of $\mathbf{u}$:

$$\begin{aligned} \mathbf{u} &= -(\mathbf{G} + \lambda \boldsymbol{I})^{-1} \mathbf{g} \\ &= -\left[\frac{(\boldsymbol{u}_1^T \mathbf{g})}{(\lambda + \lambda_1)} \boldsymbol{u}_1 + \frac{(\boldsymbol{u}_2^T \mathbf{g})}{(\lambda + \lambda_2)} \boldsymbol{u}_2 \right] \end{aligned} \tag{6.-6}$$

where $\mathbf{G} = \lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T + \lambda_2 \boldsymbol{u}_2 \boldsymbol{u}_2^T$ is the eigen decomposition of $\mathbf{G}$. Using the fact that $||\mathbf{u}||^2 = 1$, λ must satisfy:

$$\left[\frac{(\boldsymbol{u}_1^T \mathbf{g})}{(\lambda + \lambda_1)} \right]^2 + \left[\frac{(\boldsymbol{u}_2^T \mathbf{g})}{(\lambda + \lambda_2)} \right]^2 = 1 \tag{6.-7}$$

This corresponds to a polynomial equation of degree 4. By solving this equation and choosing λ equal to its real root, we obtain $\mathbf{u}$ from equation (6.-6). Consequently, the corresponding θ_p^i can be found. If there are more than one real root to the above equation then, the solution which minimizes the cost function (6.5) is selected as the desired one.

Remark: In the case where the source signal is not of constant modulus, the minima of MMSE and CM criteria do not coincide (or at least are not close) [59] and thus the proposed algorithm fails to provide a consistent channel estimate.

6.5.1 Extension to complex case

For simplicity, we have considered previously the case where the signals and the channels are real-valued. In the complex case, the proposed algorithm remains essentially the same except for the fact that the rotation matrices are function of two angle parameters θ and α (i.e., $-\sin(\theta)$ and $\sin(\theta)$ are replaced by $-\sin(\theta)e^{j\alpha}$ and $\sin(\theta)e^{-j\alpha}$, respectively).

In that case, optimizing the CM cost function versus these angle parameters leads to:

$$\min_{||\overline{\mathbf{u}}||^2=1} \overline{\mathcal{J}}(\overline{\mathbf{u}}) = \min_{||\overline{\mathbf{u}}||^2=1} (\overline{\mathbf{u}}^T \overline{\mathbf{G}} \overline{\mathbf{u}} + 2\overline{\mathbf{g}}^T \overline{\mathbf{u}}) \tag{6.-7}$$

where

$$\overline{\mathbf{G}} = \sum_n \overline{\boldsymbol{y}}^i(n) \overline{\boldsymbol{y}}^i(n)^T \quad \text{and} \quad \overline{\mathbf{g}} = \sum_n \overline{\alpha}^i(n) \overline{\boldsymbol{y}}^i(n)$$

$$\begin{aligned}
\overline{\mathbf{u}} &= [\cos(2\theta), \sin(2\theta)\cos(\alpha), \sin(2\theta)\sin(\alpha)]^T, \\
\overline{\boldsymbol{y}}^i(n) &= \begin{bmatrix} \left(|Z^i(n)|^2 - |Z^L(n)|^2\right)/2 \\ -\Re e(Z^i(n)Z^L(n)^*) \\ -\Im m(Z^i(n)Z^L(n)^*) \end{bmatrix}, \\
\overline{\alpha}^i(n) &= \frac{|Z^i(n)|^2 + |Z^L(n)|^2}{2} - r,
\end{aligned}$$

that can be solved in the same way as equation (6.5).

6.6 Simulations

In this section we present simulation results for both least squares fitting and Constant Modulus approach introduced respectively in 6.4.1,6.4.2 and 6.5 in order to asses the performance of the proposed algorithms. The algorithm performance is measured in terms of distance between estimated and real channel, as in [38], by :

$$\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{h}}) &\triangleq \min_{\alpha,k\geq 0} \left\| \alpha\hat{\boldsymbol{h}} - \begin{bmatrix} \mathbf{0}_{qk,1} \\ \boldsymbol{h} \\ \mathbf{0}_{q(M'-M-k),1} \end{bmatrix} \right\|^2 \\
&= \min_{k\geq 0} \left\| (\mathbf{I} - \hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^{\#}) \begin{bmatrix} \mathbf{0}_{qk,1} \\ \boldsymbol{h} \\ \mathbf{0}_{q(M'-M-k),1} \end{bmatrix} \right\|^2
\end{aligned}$$

where $\hat{\boldsymbol{h}}^{\#} = \hat{\boldsymbol{h}}^T/\|\hat{\boldsymbol{h}}\|^2$. Statistics are evaluated over 100 Monte-Carlo runs.

6.6.1 Simulation results for Least Squares Fitting Approach

We consider a two-output one-input system. The input sequence is an i.i.d. zero mean unit variance BPSK process. The multi path channel impulse response for the first output is generated from a raised-cosine spectrum pulse. Then it is attenuated, delayed and sampled. Afterwards, it is windowed such that the polynomial degree is M. The other channel transfer functions are generated by assuming a plane propagation model for each path (see [7] for more details). Here we choose two different sets of channel realizations. In the first realization, the channel is generated assuming that there are three paths. The main path is not attenuated, whereas two other paths are attenuated about 5 dB each. The

delay corresponding to the main path is zero and for the two other paths delays are integer multiples of path numbers. The resulted channel coefficients are :

$$\begin{aligned} h_1(z) &= -0.6046 + (0.0312)z^{-1} + (0.3089)z^{-2} \\ h_2(z) &= -0.5978 + (0.0887)z^{-1} + (0.2521)z^{-2} \end{aligned}$$

For the second channel realization, the attenuation and delays of paths are chosen in a way to obtain a channel impulse response where the energy is concentrated in the middle of channel. The channel coefficients are :

$$\begin{aligned} h_1(z) &= 0.0021 + (0.0786)z^{-1} - (0.6953)z^{-2} \\ &\quad +(0.5689)z^{-3} - (0.0009)z^{-4} + (0.0051)z^{-5} \\ h_2(z) &= 0.0067 - (0.0168)z^{-1} + (0.2558)z^{-2} \\ &\quad -(0.8727)z^{-3} + (0.0223)z^{-4} - (0.0116)z^{-5} \end{aligned}$$

For non-linear optimization algorithm, the equalizer delay is chosen to be equal to $M'-1$. The number of considered iterations are 10. Optimal step size is determined as explained in section 6.4. Simulation results are summarized in following tables for several values of L and M'.

Table 6.1 and 6.2 present the performance of both algorithms over the first channel realization derived at SNR=10 dB and $T = 500$. It is shown that, the bilinear optimization algorithm achieves better performance than the non-linear optimization algorithm in only two iterations due to the better choice of initial point, for a low value of SNR. For the second channel realization and low SNR

	MSE (dB) Non-linear	
M'	$L = M' - M + 1$	$L = M' - M + 2$
3	-29.6	-28
4	-22.94	-22.3
5	-22.1	-21.8
6	-21.2	-21.03

Table 6.1: Channel estimates MSE using 'Non-linear' Algorithm with SNR=10dB.

(10 dB), both algorithms have almost the same performance. Tables 6.3 and 6.4 provide simulation results for SNR=40 dB over second channel realization. Bilinear optimization algorithm shows better performance in this case. But is computationally more expensive than the 'non-linear' optimization algorithm.

	MSE (dB) Bilinear	
M'	$L = M' - M + 1$	$L = M' - M + 2$
3	-31.7	-31
4	-27.5	-31
5	-29	-27
6	-22.5	-26

Table 6.2: Channel estimates MSE using 'Bilinear' Algorithm with SNR=10dB.

	MSE (dB) Non-linear	
M'	$L = M' - M + 1$	$L = M' - M + 2$
6	-22.7	-21.3
7	-20.0	-20.2

Table 6.3: Channel estimates MSE using 'Non-linear' Algorithm with SNR=40dB.

6.6.2 Simulation results for Constant Modulus Approach

We have considered a one-input two-output system. The input sequence is an i.i.d. zero mean unit variance BPSK process. We consider a channel of degree 2. The channel coefficients are :

$$\begin{aligned} h_1(z) &= -0.2931 - (0.0151)z^{-1} - (0.1497)z^{-2} \\ h_2(z) &= 0.5029 + (0.7448)z^{-1} + (0.1505)z^{-2} \end{aligned}$$

Figure 6.1 presents the MSE of the estimated channel for different values of SNR with a sample size $T = 1000$. For this simulation L is fixed to $M' - M + 1$. The performance of the algorithm when an overestimation of channel order occurs remain acceptable. For example, for a SNR = 25 dB, $M' = M + 2$ and $L = 3$, the distance between estimated and real channel is -23 dB.

Table 6.5 provides simulation results for several values of M' and L. The SNR is fixed to 30 dB. It is shown that even for large values of M' and L we obtain relatively good performance.

Table 6.6 demonstrates the performance of proposed algorithm for a SNR=30 dB. Different values of T and M' have been considered. The value of L is equal to $M' - M + 1$. The performance of the algorithm depends on the sample size.

	MSE (dB) Bilinear	
M'	$L = M' - M + 1$	$L = M' - M + 2$
6	-28.5	-31
7	-25.4	-24.5

Table 6.4: Channel estimates MSE using 'Bilinear' Algorithm with SNR=40dB.

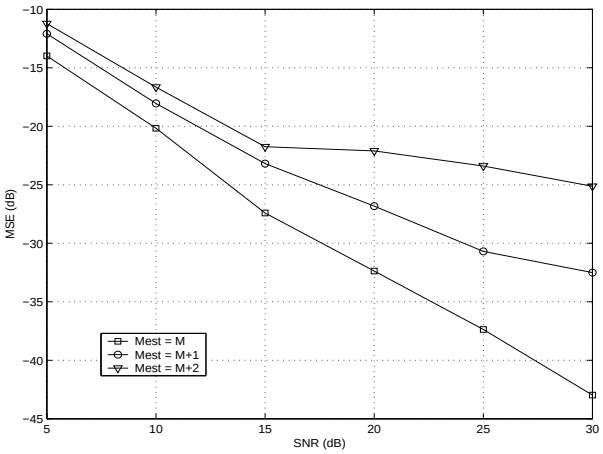

Figure 6.1: MSE versus SNR

A sample size $T = 500$ seems to be sufficient to achieve acceptable MSE even for large values of the estimated order M'.

6.7 Conclusions

In this chapter we have proposed several subspace identification algorithms robust to channel order overestimation in a SIMO context. When an overestimation of channel order occurs, the subspace criteria can be used to obtain a certain structural property related to estimated channel. We exploit this structural property in conjunction with Least Squares Fitting criteria and Constant Modulus criteria respectively to obtain robust algorithms.

Least Squares Fitting Approach In this contribution, no additional assumption is made over input data. Based on a Least squares Fitting Approach we propose two identification algorithms called Non-linear optimization and Bilinear optimization algorithms. Non-linear optimization algorithm uses an opti-

	MSE (dB)		
M'	$L=M'-M+1$	$L=M'-M+2$	$L=M'-M+3$
3	-32.15	-20.37	-26.45
4	-24.10	-23.22	-21.35
5	-19.77	-17.84	-18.18
6	-16.17	-18.76	-18.35

Table 6.5: Channel estimates MSE using CMA approach

	MSE (dB)		
T	$M'=M+1$	$M'=M+2$	$M'=M+3$
100	-15.23	-12.39	-7.65
250	-22.38	-18.96	-15.78
500	-28.42	-20.49	-18.27

Table 6.6: Channel estimates MSE using CMA approach

mal step size gradient algorithm, whereas Bilinear optimization uses a Newton algorithm.

We also propose an efficient way of initialization for both algorithms that guarantee the convergence over a few iterations. It is shown by simulation that both algorithms achieve good performance in low/ moderate signal to noise ratio.

Constant Modulus Approach This algorithm is based on the minimization of a CM cost function in conjunction with the subspace based criterion. It is shown by simulation that when constant modulus source signal is used, this algorithm achieves good performance in low/moderate signal to noise ratio.

However, it is shown in simulations that the Least Squares Fitting approach provides slightly better results than that of CM approach in a few iterations.

Chapter 7

Concluding remarks and Perspectives

In this document several aspects concerning Multiple-antenna wireless systems over fading channels have been considered. Main contributions have been divided on two different parts : *transmitter side* and *receiver side*.

7.1 Transmitter Side

At the first part of this report we have focused on Algebraic Space-Time codes that exploit maximum diversity gain and achieve a good coding gain in a multiple-antenna environment [1].

In this part we have proposed a new approach to combine Algebraic Space-Time codes with CDMA system. This approach consists of first using Algebraic Space-Time codes to encode information data, then spread the encoded data using appropriate spreading sequence and send the spread information over several transmit antennas. At the receiver side, there are one or several receive antennas and decoding procedure consists of decorrelating all users first, then applying Maximum Likelihood Sphere Decoder (SD) to each user separately.

We have also proposed a technique for blind channel estimation based on the orthogonality between noise and signal subspaces.

We have shown via several simulations the significant gain that is obtained using Algebraic Space-Time codes over several transmit antennas versus a tradi-

[1] We do not use the inherent time and/or space diversity available in a multiple antenna system which needs certain assumptions over propagation channel and can be used for a limited number of transmit antennas.

tional system that use only several transmit antennas without algebraic Space-Time codes. We have also shown that when the number of transmit antennas increase, an improvement of performance is also achieved since the diversity gain increases with the number of transmit antennas. Moreover, the spectral efficiency of the system increases with no band-width increase due to the structure of Algebraic Space-Time codes. We have shown also the improvements achieved when several receiver antennas are used at the receiver. The proposed scheme has been also adapted to the case of multi-user and multi-path. It is shown that in the case of multi-user the Sphere Decoding of each user separately yields near optimum performance. It is shown that, in the multi-path case we obtain an additional gain due to the delay diversity added to the system.

It is shown that the performance loss in the system when we use blind channel identification might be quite significant and can be lead to a diversity loss. Therefore, some more reliable identification techniques have to be considered in future.

Moreover, several aspects concerning Space-Time codes might be investigated in detail. For example, we can combine the Diagonal Algebraic Space-Time codes with Turbo codes to further improve the performance of the system.

7.2 Receiver Side

In the second part of this report we focused on the main drawbacks of subspace based identification techniques : *Computational complexity* and *lack of robustness to the order overestimation* of propagation channel.

7.2.1 Minimum Noise Subspace-like methods

MNS-like methods are promising because they are computationally more efficient than standard subspace methods. In these methods the large matrix decomposition is avoided and the noise vectors are computed in a parallel scheme as the least eigenvectors of covariance matrix computed from a special set of outputs.

We have proposed several extensions of MNS method called Symmetric Minimum Noise Subspace (SMNS) and Orthogonal Minimum Noise Subspace (OMNS). We introduced SMNS and OMNS algorithms and computed the exact computational cost of each one. We have also compared the performance of the proposed methods.

We have shown that SMNS has the advantage of being more robust and more performant because it uses all of the system outputs in a symmetric manner. In the case of OMNS we have shown that the orthogonality of the noise vectors can improve the performance of the system specially when there are a large number of sensors.

We have also proposed an efficient way of computing the least eigen vectors of the covariance matrix using a RLS-type technique followed by a power method. We have shown in simulations that this method has a high convergence rate and outperforms the other proposed methods.

MNS-like algorithms are promising and can be extended to other applications like as DOA estimation and array calibration. Actually a paper is under preparation in order to unify and complete previousely obtained results.

It is important to compute the asymptotic performance of MNS algorithm in the MIMO case in order to obtain more insight on the behavior of these algorithms and the effect of each parameter on the overall performance.

We have seen that under an additional assumption we can go 'beyond the minimum'. This must be studied later.

7.2.2 Robust subspace method

In this contribution we have exploited a structural property of the estimated channel derived from subspace-based criterion in conjunction with least squares criterion and constant modulus criterion to obtain robust algorithms.

It is shown by simulation that the approach using the Least Squares Fitting approach achieve better performance especially when we use the efficient initialization procedure proposed in this chapter. Moreover, no additional assumption over input data is made when we use this algorithm.

Bibliography

[1] C-implementation: algorithm of fincke and pohst for computing short vectors in lattices. Technical report, http://www.matha.mathematik.uni-dortmund.de/ fv/.

[2] UMTS overview. Technical report, http://www.umtsworld.com/.

[3] Space-Time block coded transmit antenna diversity for WCDMA. Technical report, proposed TDOC No. 662/98 to ETSI SMG2 UMTS standards, Dec. 1998.

[4] Downlink improvement through space-time spreading. Technical report, proposed TDOC $C30 - 19990817 - 014$ to 3GPP2 IS200 standards, aug. 1999.

[5] K. Abed-Meraim, S. Attallah, A. Chkeif, and Y. Hua. Orthogonal Oja algorithm. *IEEE Signal processing letters*, 7(5), May 2000.

[6] K. Abed-Meraim, J. Cardoso, A. Gorokhov, P. Loubaton, and E. Moulines. On subspace methods for blind identification of single-input multi-output FIR systems. *IEEE Trans. on Signal processing*, 45(1):42–55, January 1997.

[7] K. Abed-Meraim and Y. Hua. Blind identification of Multi-Input Multiple-Output system using Minimum Noise Subspace. *IEEE Trans. on Signal processing*, 45(1):254–258, January 1997.

[8] K. Abed-Meraim, Y. Hua, and A. Belouchrani. Superfast noise subspace computation: Application to array calibration and DOA estimation. *Proc. of* 15^th *annual Benjamin Franklin Symposium*, May 1997.

[9] K. Abed-Meraim, P. Loubaton, and E. Moulines. A Subspace algorithm for certain blind identification problems. *IEEE Trans. on Information Theory*, 43:499–511, March 1997.

[10] S. M. Alamouti. A simple transmit diversity technique for wireless communications,. *IEEE J. Selec. Areas on Commun.*, 16:1451–1458, Oct. 1998.

[11] J.-C. Belfiore, X. Giraud, and J. Rodriguez. Optimal linear labeling for the minimization of both source and channel distortion. In *ISIT'2000*, Sorrento, Italy, June 2000.

[12] D. P. Berteskas. *Constraint Optimization and Lagrange Multiplier Methods*. New York Academic Press, 1982.

[13] H. Bolcskei and A. Paulraj. Performance of Space-Time codes in the presence of spatial fading correlation. In *Asilomar Conf.*, Pacific Grove, CA, Oct. 2000.

[14] M. Borgmann and M. B. Pelaez. *Performance of non-orhtogonal space-time codes under transmitter-side spatial fading correlation*. PhD thesis, Stanford University, Dec. 2000.

[15] K. Boulle and J.-C. Belfiore. Modulation scheme designed for the Rayleigh Fading channel. In *CISS'1992*, Princeton,NJ, March 1992.

[16] J. Boutros and E. Viterbo. Signal space diversity : A power and bandwidth efficient diversity technique for the Rayleigh fading channel. *IEEE Trans. Information Theory*, 44:1453–1467, July 1998.

[17] J. Boutros, E. Viterbo, C. Rastello, and J. C. Belfiore. Good lattice constellations for both Rayleigh Fading and Gaussian channels. *IEEE Trans. Information Theory*, 42(2):502–518, March 1996.

[18] L. Brunel. Optimum and sub-optimum multiuser detection based on sphere decoding for multi-carrier code division multiple access systems. In *ICC 2002*, 2002.

[19] A. M. Chan and I. Lee. A new reduced-complexity sphere decoder for multiple antenna systems. *IEEE Int. Conf. on Communications*, 1:460–464, 2002.

[20] A. Chkeif, K. Abed-Meraim, G. Kawas-Kaleh, and Y. Hua. Spatio-Temporal blind adaptive multiuser detection. *IEEE Trans. Communications*, 48:729–732, May 2000.

[21] J. H. Conway and N. J. Sloane. *Sphere Packings, Lattices and groups*. New York: Springer-Verlag, 2nd edition edition, 1993.

[22] M. O. Damen, , A. Safavi, and K. Abed-Meraim. On cdma with Space-Time codes over multipath fading channels. *IEEE Trans. on Wireless Communications*, 2(1):11–19, January 2003.

[23] M. O. Damen. *Joint coding/decoding in a multiple access system, Application to mobile communications*. PhD thesis, ENST, Paris, 1998.

[24] M. O. Damen, A. Abdi, and M. Kaveh. On the effect of the correlated fading on several space-time coding and detection schemes. In *VTC fall 2001*, 2001.

[25] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore. Generalized Sphere Decoder for asymmetrical space-time communication architecture. *Electronics Letters*, 36(2):166–167, January 2000.

[26] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore. Transmit diversity using rotated constellations with Hadamard transform. *in Proc. of ASSPCC'2000, Alberta, Canada*, Oct. 2000.

[27] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore. Diagonal algebraic space-time block codes. *IEEE Trans. on Information Theory*, 48(3):628–636, March 2002.

[28] M. O. Damen, K. Abed-Meraim, and M. S. Lemdani. Further results on the sphere decoder algorithm. In *ISIT'2001*, Washington, USA, June 2001.

[29] M. O. Damen, A. Chkeif, and J.-C. Belfiore. Lattice codes decoder for Space-Time codes. *IEEE Communications Letters*, 4:161–163, May 2000.

[30] D. Divsalar and M. K. Simon. The design of the trellis coded MPSK for Fading channels: Performance criteria. *IEEE Trans. on Communicatins*, 36(9):1004–1012, September 1998.

[31] P. Fan and M. Darnell. *Sequence Design for Communication Applications*. Taunton, Somerset, England: Research Studies Press Ltd., 1996.

[32] I. Fijalkow, A. Touzni, and J. R. Treichler. Fractionally spaced equalization using CMA: robustness to channel noise and lack of disparity. *IEEE Trans. on Signal processing*, 45:56–67, 1997.

[33] U. Finke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Matematics of computations*, 44(170):463–471, April 1985.

[34] G. Forney. Minimal bases of rational vector spaces, with applications to multivariable linear systems. *SIAM J. Contr.*, 13:493–520, March 1975.

[35] G. J. Foschchini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky. Samplified processing for high spectral efficiency wireless communication emploing multi-element arrays. *IEEE J. on Selected Areas in Communications*, pages 1841–1852, Nov 1999.

[36] S. Galliou and J.-C. Belfiore. A new family of full rate, fully diverse Space -Time codes based on Galois theory. In *ISIT'2002*, Lausanne, Switzerland, July 2002.

[37] H. Gazzah and K. Abed-Meraim. Blind equalization with controlled delay robust to order overestimation. In *ISSPA 2001*, Aug. 2001.

[38] H. Gazzah, P. Regalia, J. Delmas, and K. Abed-Meriam. A blind multichannel identification algorithm robust to order overestimation. *IEEE Trans. on Signal processing*, 50(6):1449–1458, June 2002.

[39] H. Gazzah, P.-A. Regalia, and J.-P. Delmas. A blind identification algorithm robust to order over estimation. In *ICASSP 2000*, June 2000.

[40] D. Gesbert and P. Duhamel. Robust blind joint data/channel estimation based on bilinear optimization. *Proc.* 8^{th} *IEEE Int. Workshop on SSAP*, pages 168–171, June 1996.

[41] D. Gesbert, M. Shafi, D. S. Shui, P. Smith, and A. Naguib. From theory to pactice: An overview of MIMO Space-Time coded wireless system. *Journal of Selected Areas in Communications*, 21(3):281–303, April 2003.

[42] X. Giraud and J.-C. Belfiore. Constellations matched to Rayleigh Fading channel. *IEEE Trans. Information Theory*, 42(1):106–115, January 1996.

[43] X. Giraud, E. Boutillon, and J.-C. Belfiore. Algebraic tools to build modulation schemes for fading channels. *IEEE Trans. Information Theory*, 43(3):938–952, May 1997.

[44] S. Glisic and B. Vucetic. *Spread Spectrum CDMA systems for wireless communications*. Artech House Inc., 1997.

[45] D. Godard. Self-recovering equalization and carrier tracking in two dimensional data communication systems. *IEEE Comm.*, 28:1867–1875, 1980.

[46] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolonian-sky. Detection algorithm and initial laboratory results using the v-blast space-time communication architecture. *Electronics Letters*, 35(1):14–15, January 1999.

[47] G. H. Golub and C. F. V. Loan. *Matrix Computations*. John Hopkins University Press, 3-rd edition, 1996.

[48] A. Gorokhov, M. Kristensson, and B. Ottersten. Robust blind second order deconvolution of multiple FIR channels. In *GLOBECOM 98*, Nov. 1998.

[49] A. Gorokhov, M. Kristensson, and B. Ottersten. Robust blind second-order deconvolution. *IEEE Signal processing letters*, 6(1):13–17, January 1999.

[50] A. Gorokhov and P. Loubaton. Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources. *IEEE Trans. Circuits Syst.*, 44:813–820, Sept. 1997.

[51] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo. Signal design for transmitter diversity wireless communication systems over Rayleigh Fading channels. In *Proc. IEEE VTC'96*, 1996.

[52] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Trans. onInf. Theory*, 49(4):951–963, April 2003.

[53] B. Hassibi and H. Vikalo. On the expected complexity of Sphere Decoding. In *The 35th Asilomar Conf.*, Nov 2001.

[54] B. Hochwald, T. Marzetta, and C. Papadias. A novel Space-Time spreading scheme for wireless CDMA systems. In *in Proc. of 37 Allerton Conf.*, Monticello, IL, Sept. 1999.

[55] B. M. Hochwald and S. Brink. Achieving Near-Capacity on a Multiple-Antenna channel. *submitted to IEEE Trans. on Communications*, 2001.

[56] B. M. Hochwald, T. Marzetta, and C. Papadias. A transmitter diversity scheme for wideband CDMA systems based on Space-Time spreading. *IEEE J. of Selected Areas in Communications*, 19:48 –60, Jan. 2001.

[57] Y. Hua, K. Abed-Meraim, and M. Wax. Blind system identification using minimum noise subspace. *IEEE Trans. on Signal processing*, 45(3):770–773, March 1997.

[58] W. C. Jakes. *Microwave mobile communications.* New York IEEE Press, 1974.

[59] C. R. Johnson, P. Schniter Jr., T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas. Blind equalization using the constant modulus criterion: A review. *Proc. IEEE*, 86:1927–1950, Nov. 1998.

[60] T. Kailath. *Linear Systems.* Prentice-Hall, 3-rd edition, 1980.

[61] A. K. Lenstra, H. W. Lenstra, and L. Lovasz. Factoring polynomials with rational coefficients. *Mat. Ann.*, 261:515–534, Oct. 1982.

[62] Y. Li and Z. Ding. Blind channel identification based on second order cyclostationary statistics. *Proc. ICASSP*, 4:81–84, 1993.

[63] A. P. Liavas, P. A. Regalia, and J.-P. Delmas. Blind channel approximation : Effective channel order determination. *IEEE Trans. on Signal processing*, 47(12):3336–3344, December 1999.

[64] H. Liu and G. Xu. A subspace method for signature waveform estimation in synchronous CDMA systems. *IEEE Trans. on Communications*, pages 1346–1354, Oct. 1996.

[65] J. Manton. A new algorithm for computing the extreme eigenvectors of a complex hermitien matrix. In *Proc. of IEEE workshop on Statistical Signal Processing*, August 2001.

[66] S. Marcos. Calibration of a distorted towed array using a propagation operator. *Journal of Acoustical Society of America*, Dec. 1991.

[67] S. Marcos, A. Marsal, and M. Benidir. The propagator method for source bearing estimation. *Signal Processing*, 42:121–138, March 1995.

[68] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue. Subspace methods for the blind identification of the multichannel FIR filters. *IEEE Trans. on Signal processing*, 43:516–525, Feb. 1995.

[69] A. Narula, M. D. Trott, and G. W. Wornell. Performance limits of coded diversity methods for transmitter antenna arrays. *IEEE Trans. On Inf. Theory*, 45(7):2418–2433, Nov. 1999.

[70] L. Perros-Meilhac, E. Moulines, K. Abed-Meraim, P. Chevalier, and P. Duhamel. Blind identification of multipath channels : A parametric subspace approach. *IEEE Trans. on Signal processing*, 49(7):1468–1479, July 2001.

[71] E. Pite and P. Duhamel. Bilinear methods for blind channel equalization: (no) local minimum issue. *in Proc. of ICASSP'98*, 4:2113 –2116, 1998.

[72] M. Pohst. On the computation of lattice vectors of minimal length, succesive minima and reduced bases with applications. *ACM SIGSAM Bull*, 15:37–44, 1981.

[73] B. Porat. *Digital Processing of Random Signals*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[74] J. G. Proakis. *Digital Communications*. Mc. Grow-Hill series in Electrical and Computer Engineering, 4-th edition, 2000.

[75] J. Radon. *lineare scharen orthogonaler matrizen*, pages 1–14. abhandlungen aus dem mathematischen seminar der hamburgishen universitat, 1922.

[76] G. C. Rayleigh and J. M. Coiffi. Spatio-Temporal coding for wireless communications. *IEEE Trans. on Communications*, 46:357–366, March 1998.

[77] G. Rekaya and J. C. Belfiore. Complexity of ML lattice decoders for the decoding of linear full rate Space-Time Codes. *submitted IEEE Trans. on Wireless Communications*, Dec. 2002.

[78] K. Rohani and J. Jalloul. Diversity for Direct Spread CDMA. Technical report, ETSI SMG2 Wideband CDMA Concept Group, Sept. 1997.

[79] A. Safavi and K. Abed-Meraim. Blind channel identification robust to order overestimation. *to be submitted to IEEE trans. on Signal procesing.*

[80] A. Safavi and K. Abed-Meraim. Orthogonal Minimum Noise Subspace for Multi-Input Multi-Output system Identification. In *Proc. of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, Oct. 2001.

[81] A. Safavi and K. Abed-Meraim. Blind channel identification robust to order overestimation: a Constant Modulus Approach. In *Proc. of the IEEE International Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, 2003.

[82] A. Safavi and K. Abed-Meraim. Symmetric Minimum Noise Subspace for Multi-Input Multi-Output system Identification. In *submitted to GRETSI*, France, Sept. 2003.

[83] A. Safavi, K. Abed-Meraim, and Ph. Ciblat. Blind channel identification robust to order overestimation. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, California, U.S.A., Nov. 2002.

[84] A. Safavi, K. Abed-Meraim, and Y. Hua. Minimun Noise Subspace array processing. *to be submitted to IEEE trans. on Signal procesing.*

[85] N. Seshadri and J. H. Winters. Two signaling schemes for improving the error performance of frequency-division-duples (FDD) transmission systems using transmitter antenna diversity. *Int. J. Wireless Inform. Networks*, 1(1), 1994.

[86] D. S. Shiu, G. J. Foschini, M. J.'Gans, and J. M. Kahn. Fading correlation and its effects on the capacity of multielement antenna systems. *IEEE Trans. on Communications*, 48(3):502–513, march 2000.

[87] V. Tarokh, H. Jafarkhani, and A. R. Calderbanck. The application of orthogonal designs to wireless communication. In *Information Theory Workshop*, Jun 1998.

[88] V. Tarokh, H. Jafarkhani, and A. R. Calderbanck. Space-Time block codes from orthogonal designs. *IEEE Trans. Information Theory*, 45:1456–1466, July 1999.

[89] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank. Space-Time codes for high data rate wireless communications: performance criteria in the presence of channel estimation errors, mobility, and multiple paths. *IEEE Trans. on Communications*, pages 199–207, Feb. 1999.

[90] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-Time codes for high data rate wireless communications: performance criterion and code construction. *IEEE Trans. Information Theory*, 44:744–765, March 1998.

[91] E. Telatar. Capacity of multi-antanna Gaussian Channels. Technical report, ATT Bell Laboraratories, Murray Hill, NJ, 1995.

[92] L. Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelihood methods. *Proc. IEEE*, 86:1951–1967, Oct. 1998.

[93] L. Tong and S. Perreau. Multichannel blind identification: from ubspace to maximum likelihood methods. *Processing of the IEEE*, 86(10):1951–1968, Oct. 1998.

[94] L. Tong, G. Xu, and T. Kailath. A new approach to blind identification and equalization of multi-path channels. 25^{th} *Asilomar Conf.*, pages 856–860, 1991.

[95] J. R. Treichler and B. G. Agee. A new approach to multi-path correction of constant modulus signal. *IEEE Trans. on Signal processing*, 31:459–72, 1983.

[96] J. R. Treichler, L. Tong, I. Fijalkow, C. R. Johnson, Jr., and C. U. Berg. On the current shape of FSE-CMA behavior theory. In *First IEEE signal processing workshop on Signal processing advances in wireless communications*, 1997.

[97] A. Viterbi. *CDMA- Principals of Spread Spectrum Communication.* Addison-Wesley, 1995.

[98] E. Viterbo and E. Biglieri. A universal lattice code decoder. In *GRETSI*, Juan-les-Pins, France, Sept. 1993.

[99] E. Viterbo and J. Boutros. A universal lattice code decoder for Fading channel. *IEEE Trans. Information Theory*, 45(5):1639–1642, July 1999.

[100] E. Walter and L. Pronzato. *Identification de modèles paramétriques à partir de données experimentales*. Masson, 1994.

[101] X. Wang and H. V. Poor. Blind multiuser detection: A subspace approach. *IEEE Trans. Information Theory*, 44:677–690, March 1998.

[102] M. Wax and T. Kailath. Detection of signal by information theoretic criteria. *IEEE Trans. on Acoustics, Speech and Signal processing*, ASSP-33(2):387–392, April 1985.

[103] V. Weerackody. Diversity for the direct-sequence spread spectrum system using multiple transmit antennas. *ICC'93*, 3:23–26, May 1993.

[104] J. H. Winters. On the capacity of radiocommunication systems with diversity in a Rayleigh Fading environement. *IEEE J. Selec. Areas on Commun.*, SAC-5(5), June 1987.

[105] A. Wittneben. A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation. *ICC 93*, 3:1630–1634, 1993.

Appendix A

Publications

Contributions of this thesis gave rise to several publications in :

Journal Publications

- M. O. Damen, A. Safavi and K. Abed-Meraim. "On CDMA with Space-Time Codes Over Multi path Fading Channels", *IEEE Transactions on Wireless Communication*, Jan. 2003, pp. 11-19

- A. Safavi, K. Abed-Meraim and Y. Hua. "Array Processing using Minimum Noise Subspace", To be submitted to *IEEE transactions on Signal Processing.*

- A. Safavi and K. Abed-Meraim. "Blind Channel Identification Robust to Order Overestimation", To be submitted to *IEEE Signal Processing letters.*

Conferences and Workshop

- M. O. Damen, K. Abed-Meraim and A. Safavi, "On CDMA with Space-Time Codes Over Multi path Fading Channels" *The 7th IEEE Singapore International Conference on Communications Systems (ICCS)*, Nov. 2000

- A. Safavi and K. Abed-Meraim, " Orthogonal Minimum Noise Subspace for Multiple-Input Multiple-Output System Identification" *The 11th IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 285 -288, Singapore 2001.

- A. Safavi, K. Abed-Meraim and Ph. Ciblat, "Blind Channel Identification Robust to Order Overestimation", *The 36th Asilomar Conference on Signals, Systems and Computers*, vol.2, pp. 1248 -1251, Nov. 2002, USA.

- A. Safavi and K. Abed-Meraim, "Blind Channel Identification Robust to Order Overestimation: A Constant Modulus Approach", *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol.4, pp. 313-316, April 2003, Hong Kong.

- A. Safavi and K. Abed-Meraim, " Symmetric Minimum Noise Subspace for Multi-Input Multi-Output System Identification", accepted for publication in *19eme colloque GRETSI sur le traitement du signal et des Images.*, Sept. 2003.