UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Juan Carlos Minango Negrete

Low-Complexity Quasi-Optimum Detector Algorithms for Spatial Multiplexing MIMO Systems with Large Number of Antennas

Algoritmos de Detecção Quase-Ótimo de Baixa Complexidade para Sistemas de Multiplexação Espacial com um Grande Número de Antenas

Campinas

2019

UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Juan Carlos Minango Negrete

Low-Complexity Quasi-Optimum Detector Algorithms for Spatial Multiplexing MIMO Systems with Large Number of Antennas

Algoritmos de Detecção Quase-Ótimo de Baixa Complexidade para Sistemas de Multiplexação Espacial com um Grande Número de Antenas

Thesis presented to the School of Electrical and Computer Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering, in the area of Telecommunications and Telematics.

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica, na Área de Telecomunicações e Telemática.

Supervisor: Prof. Dr. Celso de Almeida

Este exemplar corresponde à versão final da tese defendida pelo aluno Juan Carlos Minango Negrete, e orientada pelo Prof. Dr. Celso de Almeida

____

Campinas

2019

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

M66L	Minango Negrete, Juan Carlos, 1987- Low-complexity quasi-optimum detector algorithms for spatial multiplexing MIMO systems with large number of antennas / Juan Carlos Minango Negrete. – Campinas, SP : [s.n.], 2019. Orientador: Celso de Almeida. Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação. 1. Antenas. 2. Diversidade. 3. Probabilidade de erro (Matemática). 4. Telefonia celular - Sistema. I. Almeida, Celso de, 1957-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Algoritmos de detecção quase-ótimo de baixa complexidade para sistemas de multiplexação espacial com grande número de antenas
Palavras-chave em inglês:
Antennas
Diversity
Error probability (Mathematic)
Cell phone - System
Área de concentração: Telecomunicações e Telemática
Titulação: Doutor em Engenharia Elétrica
Banca examinadora:
Celso de Almeida [Orientador]
Reginaldo Palazzo Júnior
Luiz César Martini
Carlos Eduardo Câmara
Henry William Merino Acuna
Data de defesa: 10-07-2019
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-2682-8602
- Currículo Lattes do autor: http://lattes.cnpq.br/3741980811663834

COMISSÃO JULGADORA – TESE DE DOUTORADO

Candidato: Juan Carlos Minango Negrete **RA**:134115

Data da Defesa: 10 de julho de 2019

Título da Tese: "Algoritmos de Detecção Quase-Ótimo de Baixa Complexidade

para Sistemas de Multiplexação Espacial com Grande Número de Antenas".

Prof. Dr. Celso de Almeida (Presidente, FEEC/UNICAMP)
Prof. Dr. Reginaldo Palazzo Júnior (FEEC/UNICAMP)
Prof. Dr. Luiz César Martini (FEEC/UNICAMP)
Prof. Dr. Carlos Eduardo Câmara (UniAnchieta)
Prof. Dr. Henry William Merino Acuna (PUC/PUC Peru)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

Acknowledgements

Firstly, I thank God for the blessings throughout my life. I would also like to thank to my parents Silvia and Marcelo for their love, for their sacrifice, for their advice an example. They help me and support me to continue growing as a person and by always be on my side. To my brother David, for its friendship. I would also like to thank my beloved girlfriend, Carolina. Thank you for all her love and support me for everything, and especially I can't thank you enough for encouraging me throughout this experience.

Furthermore, I would like to express my sincere gratitude to my advisor Prof. Celso de Almeida for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of this research.

Besides my advisor, I would like to thank the rest of my thesis committee, Prof. Dr. Henry Merino, Prof. Dr. Carlos Câmara, Prof. Dr. Reginaldo Palazzo and Prof. Dr. Luiz César Martini, for their encouragement, insightful comments, and contributions aimed at improving the present thesis.

Thanks to Unicamp and my friends from Unicamp, who made this academic stage more bearable.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Finally, thanks SENESCYT-IFTH for the financial support.

Success consists of going from failure to failure without loss of enthusiasm (Winston Churchill).

Abstract

Spatial multiplexing systems employing a large number (tens to hundreds) of transmit and receive antennas have been considered in order to meet the requirement of high data rates and spectral efficiency of the fifth generation (5G) wireless communication standard. However, these systems present some challenges that must be overcome. One of them is the increasing complexity in spatial multiplexing detection, when the number of antennas is large. Thus, this thesis proposes detector algorithms that perform as close as possible to the performance of the optimum maximum likelihood (ML) detector, while yielding the complexity of spatial multiplexing systems with large number of antennas low.

In the first part of this thesis, the case where the number of receive antennas is much larger than the number of transmit antennas is analyzed, where a closed-form expression is derived to determine the performance difference between ML detector and zero-forcing (ZF) detector. Thus, ZF detector achieves near-optimum performance when the number of receive antennas is much larger than the number of transmit antennas due to the asymptotically orthogonal channel matrix property. Thus, two detectors for this case are proposed, which are based on the damped Jacobi (DJ) and Newton-Schultz iterative (NSI) algorithms to avoid the matrix inversion used in the ZF detector and consequently reducing the complexity. Numerical results show that the proposed algorithms achieve the ZF performance, keeping the low-complexity, where NSI outperforms DJ algorithm in performance and complexity. In the second part of this thesis, the case where the number of transmit antennas is equal to the number of receive antennas is considered. In this case, the ZF detector is not more able to achieve near-optimum performance. Thus, a precoding approach is proposed, which is based on channel matrix orthogonalization achieves the optimum performance using the ZF detector. Despite the precoding approach achieves the optimum performance at the expense of power transmission increase, its drawback is that the transmitter needs to know the channel matrix through the feedback link. Finally, a low-complexity detector based on symbol-flipping (SF) algorithm, which is based on local sub-optimum solution searching, is proposed. It achieves a quasi-optimum performance for spatial-multiplexing systems with very large number of transmit and receive antennas by employing simple SF procedures.

Keywords: spatial multiplexing; MIMO; massive MIMO; symmetric MIMO; bit error rate; optimum performance; Rayleigh fading; complexity; detection.

Resumo

Os sistemas de multiplexação espacial que empregam uma grande quantidade de antenas transmissoras e receptoras, na ordem de dezenas a centenas, têm sido considerados para cumprir com os requsitos de altas taxas de dados e eficiência espectral do padrão de comunicação sem fio da quinta geraçõ (5G). No entanto, esses sistemas apresentam alguns desafios que devem ser superados. Um desses desafios é a crescente complexidade na detecção uma vez que o número de antenas é grande. Assim, esta tese propõe algoritmos de detecção cujo desempenho seja o mais próximo possível do desempenho ótimo do detector de máxima verossimilhança (ML), atingindo uma baixa complexidade para os sistemas de multiplexação espacial que empregam um grande número de antenas.

Na primeira parte desta tese, o caso em que o número de antenas de recepção é bem maior que o número de antenas de transmissão é avaliado. Neste quesito, uma expressão fechada é obtida para determinar a diferença de desempenho entre o detector ML e o detector zero-forcing (ZF). Assim, o detector ZF atinge um desempenho quase ótimo quando o número de antenas receptoras é bem maior que o número de antenas transmissoras devido a que a matriz de canal apresenta a propriedade de ortogonalidade assintótica. Baseados nesta premissa, dois detectores para este caso são propostos, os quais são baseados nos algoritmos iterativos de damped Jacobi (DJ) e Newton-Schultz (NSI). Esses algoritmos têm como objetivo evitar a inversão de matrizes usada no detector ZF e consequentemente reduzir a complexidade. Os resultados numéricos mostram que os dois algoritmos propostos atingem o desempenho do detector ZF mantendo uma baixa complexidade, onde o algoritmo NSI supera ao algoritmo DJ em termos de desempenho e de menor complexidade. Na segunda parte desta tese, considera-se o caso em que o número de antenas de transmissão é igual ao número de antenas de recepção. Neste caso, o detector ZF não é mais capaz de atingir um desempenho ótimo. Assim, um novo pré-codificador é proposto, o qual é baseado na ortogonalização da matriz de canal a fim de atingir o desempenho ótimo ao usar o detector ZF. Apesar do que o pré-codificador atinge o desempenho ótimo à custa do aumento da potência de transmissão, sua desvantagem é que o transmissor precisa conhecer a matriz do canal através de um enlace de retroalimentação. Finalmente, um detector de baixa complexidade baseado no algoritmo de inversão de símbolos (SF), cujo principio é a procura de soluções sub-ótimas locais, é proposto. Dito detector atinge um desempenho quase ótimo para sistemas de multiplexação espacial com um grande número de antenas de transmissão e recepção ao empregar procedimentos simples de SF.

Palavras-chaves: multiplexação espacial; MIMO; MIMO massivo; MIMO simétrico; tasa de erro de bit; desempenho ótimo; desvanecimento Rayleigh; complexidade; detecção.

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Connotation
5G	fifth generation
AWGN	additive white Gaussian noise
BER	bit error rate
BM	band matrix
BPSK	binary phase shift keying
BS	base station
CDF	cumulative distribution function
CG	conjugate gradient
CJ	conventional Jacobi
CSCG	circular symmetric complex Gaussian
CSI	channel state information
DD	diagonal dominant
DJ	damped Jacobi
GA	genie-aided
GS	Gauss-Seidel
HKZ	Hermite-Korkin-Zolotarev
HPD	Hermitian positive-definite
ICI	inter-carrier interference
i.i.d.	independent and identically distributed
IoT	internet of things
ISI	inter-symbolic interference
LD	linear detector
LLL	Lenstra-Lenstra-Lovasz
LR	lattice reduction
LS	local search
LTE	long term evolution
MF	matched-filter
MIMO	multiple-input multiple-output

Abbreviation	Connotation
ML	maximum likelihood
MMSE	minimum mean-square error
MSE	mean-square error
NS	Neumann series
NSI	Newton-Schultz iterative
PDF	probability density function
PEP	pairwise error probability
PSK	phase shift keying
QAM	quadrature amplitude modulation
Q.o.S	quality of service
SD	sphere detector
SDD	strictly diagonal dominant
SF	symbol-flipping
SISO	single-input single-output
SM	spatial multiplexing
SNR	signal-to-noise ratio
SVD	singular value decomposition
ZF	zero-forcing

List of Symbols

Symbol	Connotation
A	amplitude
B	system bandwidth
B_c	coherent bandwidth
c	light speed
$\mathbb{C}$	complex elements
$\mathcal{CN}(0,1)$	standard complex normal random variable
$\mathrm{diag}[\cdot]$	diagonal matrix
$\mathrm{d}_{\mathsf{min}}$	minimum Euclidean distance of the constellation
E_b	energy per bit
E_b/N_0	energy per bit to noise power spectral density ratio
E_s	energy per symbol
$\mathbb{E}\{\cdot\}$	mean value
f_c	carrier frequency
$f(n)$	PDF of the random variable x
$\mathbf{I}_m$	$m \times m$ identity matrix
M	modulation order
N_0	noise unilateral power spectral density
N_R	number of receive antennas
N_T	number of transmit antennas
P	probability of error
$\bar{P}$	average transmitted power
$\mathrm{P_r}$	pairwise error probability
$Q(\cdot)$	Q-function
$Q\{\cdot\}$	quantization process
R_b	bit rate
T_b	bit time interval
T_c	coherence time interval
T_s	symbol time interval
$\mathbb{R}$	real elements
γ	overrall SNR
$\delta(\cdot)$	Dirac delta function
$\kappa(\cdot)$	condition number
$\lambda_{\max}(\cdot)$	largest eigenvalue
$\lambda_{\mathsf{min}}(\cdot)$	smallest eigenvalue
λ_w	wavelength

Symbol	**Connotation**
μ_n	mean of the noise components
ξ	spectral efficiency
$\rho(\cdot)$	spectral radius
σ_a	parameter of the Rayleigh PDF
σ_n	variance of the noise components
ω	damped parameter
$\|\mathbf{A}\|_1$	1-norm of matrix $\mathbf{A}$
$\|\mathbf{A}\|_\infty$	infinite-norm of matrix $\mathbf{A}$
$\|\mathbf{A}\|_2$	2-norm of matrix $\mathbf{A}$
$\|\mathbf{A}\|_F$	Frobenius norm of matrix $\mathbf{A}$
$\mathrm{Tr}[\mathbf{A}]$	trace of matrix $\mathbf{A}$
$\Re\{\cdot\}$	real-part operation
$\Im\{\cdot\}$	imaginary-part operation
$(\cdot)^T$	transposition operation
$(\cdot)^*$	conjugation operation
$(\cdot)^H$	Hermitian operation
$(\cdot)^{-1}$	inverse operation
$(\cdot)^\dagger$	Moore-Penrose pseudoinverse
j	imaginary, $\sqrt{-1}$

Contents

1 Introduction

Over the last years, wireless communications traffic has experimented an exponential growth, due to the increasing demand for broadband internet access and multimedia services, such as voice, data and video transmission. In this context, wireless communications are in continuous development motivated by the growing requirements in terms of data rates, reliability and spectral efficiency. Thus, new challenges have arisen for researchers of both academic and industry to pursue solutions, which must be able to deliver higher data rates with lower probability of error.

In wireless communications, two obvious solutions for increasing the date rates are to increase the bandwidth and the transmission power. However, increasing the bandwidth is not so simple, since the frequency spectrum is a limited and a controlled resource. In contrast, increasing transmission power raises system implementation and operating costs, since it can find regulatory barriers that seek to limit system interference and possible health damage.

Nowadays, multiple antennas at the transmitter and receiver is a technique know as multiple-input multiple-output (MIMO) systems (ANDREWS *et al.*, 2007). It is being considered as a solution to achieve higher data rates and reliability without any expansion of the required bandwidth or increase in the system transmission power by the exploitation of both spatial multiplexing and diversity. Thus, the spectral efficiency of MIMO systems increases with the number of transmit antennas. MIMO presents high data rates transmission as possible. On the other hand, multiple receive antennas is the key to combat fading and provide receive diversity gain. Because of these advantages, MIMO systems have been extensively researched. In fact, it can be safely argued that MIMO systems employing 4 to 8 antennas are a fairly mature area nowadays. Thus, the technological issues in such small MIMO systems are well understood and practical implementation of these systems have become common (CHOCKALINGAM; RAJAN, 2014). Indeed, small MIMO systems have already become part of wireless standards like 802.11n/Wi-Fi for local area networks (PEFKIANAKIS *et al.*, 2013), 802.16e/WiMAX for broadband wireless networks (ANDREWS *et al.*, 2007) and long term evolution (LTE) for cellular networks (DAHLMAN *et al.*, 2008).

For the next generation of wireless communications, 5th Generation (5G) is the candidate which promises higher data rates, higher user densities and significant improvement in users' perceived Quality of Service (Q.o.S) (XIANG *et al.*, 2016). In this context, the deployment of MIMO systems with large number of (tens to hundreds) antennas in large-scale MIMO systems, has attracted significant attention of the research community, since this technology allows a much higher spectral efficiency and consequently very high data rates, serving a large number of users at the same time without requiring extra bandwidth resources or additional transmit power (MARZETTA, 2010; RUSEK *et al.*, 2013). This large-scale version of conventional MIMO systems, where a restricted number of antennas is employed, is designed to exploit the advantages of extra degrees of freedom obtained by using more antennas. Thus, large-scale MIMO systems are strong candidates for the 5G wireless communications. standard.

The remainder of this chapter is organized as follows. In Section 1.1, some related works are summarized and the motivations of this thesis are introduced. Finally, the contributions and the outline of this thesis are presented in Section 1.2.

1.1 Related Works and Motivation

The concept of large-scale MIMO systems with tens to hundreds of transmit and receive antennas was first proposed in (RUSEK *et al.*, 2013) in order to achieve high spectral efficiencies. However, several challenges need to be addressed in realizing practical large-scale MIMO systems. One of them is to develop detection algorithms that practically implement low complexity high-performance MIMO systems (MARZETTA *et al.*, 2016).

The MIMO detector jointly detect the symbols that are transmitted simultaneously by the multiple transmit antennas subject to the contamination of fading and random noise. Unfortunately, in MIMO systems the optimum detection problem was proven non-deterministic polynomial-time hard (NP-hard) (YANG; HANZO, 2015),. Thus, the detectors projected for optimum detection have exponential complexity that increases with the number of transmit antennas. This makes these detectors impractical for large-scale MIMO systems.

On the other hand, several works have been published in the literature on MIMO detection (LARSSON, 2009; HUANG *et al.*, 2011; BAI; CHOI, 2014; YANG; HANZO, 2015), which are predominantly focused to conventional small-scale MIMO systems, that is until a maximum of eight antennas. In this context, known small-scale MIMO detectors either perform well but present high complexity, or behave well in complexity but perform poorly in large-scale MIMO systems (CHOCKALINGAM; RAJAN, 2014). Fortunately, since a large number of antennas are deployed in MIMO systems, the channel hardening phenomenon, which is discussed in sub-section 2.5.2, observed in these systems becomes helpful to explore low-complexity high-performance detectors for practical large-scale MIMO systems. Thus, large-scale MIMO detection is one of the main research challenges for 5G wireless communications.

Furthermore, since in 5G it is expected that a large number of users or devices, such as sensors, environmental monitoring devices, smart grid components, smartphones and tablets - which represents the internet of things (IoT) (LIN *et al.*, 2017) - with different configurations and requirements are connected to the network (AGIWAL *et al.*, 2016). Receivers capable of detecting the transmitted information must be designed according to their different data requirements, which has motivated the research of low-complexity detectors for large-scale MIMO systems. Finally, once wireless communication standards do not impose how network providers design their detectors, thus opens up the opportunity for the implementation of different kind of detectors capable of self adjusting according to the goals required at any instant of the transmission.

1.2 Contributions and Thesis Outline

This thesis comprises most of the research results in the field of large-scale MIMO detection algorithms. In general, this thesis proposes the implementation of novel detector algorithms that are suitable for large-scale MIMO systems in terms of performance and complexity. For this, the study of MIMO systems basic concepts, the benefits of scaling-up MIMO systems, the implementation of existing state-of-the-art detectors and a consequent comparative analysis through Monte Carlo simulations are included. The remainder of the thesis is structured as follows:

- **Chapter 2.** This chapter contains the basic concepts for understanding the thesis. An introduction to MIMO systems is made, including a brief review of fundamental concepts and their mathematical representation, Moreover, the benefits of traditional MIMO are described, together with the advantages and challenges of deployment of large-scale MIMO systems. Finally, optimum and sub-optimum MIMO detectors are also presented, along with the distinction between massive and symmetric MIMO systems.

- **Chapter 3.** This chapter contains the contribution published in (MINANGO *et al.*, 2018). At first, based on the channel hardening phenomenon, it is verified that in massive MIMO systems with hundreds of antennas at the base station (BS) and tens users with a single transmitter antenna, linear detectors, such as zero-forcing (ZF), are able to achieve near-optimum performance due to the property of asymptotically orthogonal channel matrix. However, how far away is ZF from the optimum performance in massive MIMO systems?. In order to answer this question, in this chapter closed-form bit error rate (BER) expressions of the ZF and optimum detector are derived. Then, these BER expressions are subsequently used to evaluate the performance difference between ZF and the optimum detector, which is a function of the number of antennas at the BS and the number of users. Finally, numerical results verify the tightness of the expressions obtained.

- **Chapter 4.** This chapter contains the contributions published in (MINANGO; ALMEIDA, 2017c; MINANGO *et al.*, 2017). Once linear ZF detector is able to achieve the near-optimum performance in massive MIMO systems, its practical implementation becomes of interest. Unfortunately, ZF detector involves matrix inversion with high complexity especially when the number of users is large. In order to reduce this complexity, in this chapter, a low-complexity ZF detector based on the Damped Jacobi (DJ) algorithm is proposed. Further, a simple approach to determine the optimum and quasi-optimum damped parameter by exploiting the channel property of asymptotic orthogonality is developed. The analysis shows that the DJ algorithm can reduce the complexity of the ZF detector by about one order of magnitude without performance loss. Finally, it is verified through numerical simulations that the ZF detector based on DJ algorithm achieves the same near-optimum performance of the ZF detector employing matrix inversion, but with a reduced number of iterations, keeping the complexity low.

- **Chapter 5.** This chapter contains the contribution published in (MINANGO; ALMEIDA, 2018b). Although ZF detector based on DJ algorithm achieves a near-optimum performance with low-complexity, its convergence is linear. Thus, in this chapter, a low-complexity ZF detector based on the Newton-Schultz Iterative (NSI) algorithm is proposed, which yields similar performance that ZF detector with exact matrix inversion. Moreover, a relationship between NSI and DJ algorithms is introduced, which shows that NSI outperforms DJ in terms of convergence rate. Besides, in order to further accelerate the convergence rate and reduce the complexity of NSI algorithm, a novel initial matrix inversion solution of NSI algorithm based on Tchebychev polynomial is proposed. Numerical results show that NSI algorithm with the proposed initial matrix inversion solution achieves the near-optimum ZF performance in just two iterations outperforming DJ algorithm in terms of performance and complexity.

- **Chapter 6.** This chapter contains the contribution published in (MINANGO; ALMEIDA, 2018a). This chapter is focused on large-scale symmetric MIMO systems, where ZF detector is unable to achieve near-optimum performance, because the symmetric channel matrix is not asymptotically orthogonal. Thus, ZF performance is clearly inferior to the optimum detector. However, this is no longer true if the symmetric channel matrix was orthogonal. Motivated by this statement, in this chapter, a novel precoding approach for channel matrix orthogonalization of symmetric MIMO systems is proposed. In general, at the transmitter, the precoding matrix that transforms a symmetric channel matrix into an orthogonal matrix is first obtained. After that, at the receiver, ZF detector is applied to the equivalent orthogonal channel matrix. Thus, ZF with precoding presents the same performance that the optimum detector. Numerical results show that the proposed approach achieves the optimum performance with reduced complexity at the expense of a small increase in the total average transmitted power.

- **Chapter 7.** In this chapter a low-complexity detector based on symbol-flipping (SF) algorithm is proposed. This detector algorithm achieves the same optimum performance of the ML detector for symmetric spatial-multiplexing MIMO systems with very large number of transmit and receive antennas by employing simple SF procedures, whose starting-point is the matched filter (MF) vector solution. However, if the number of antennas is on the order of tens, several SF procedures must be employed, whose starting-points are a set generated by the MF vector solution changed by random vectors. The best SF result is chosen as the detected solution. Numerical results indicate that the proposed detector algorithm achieves the optimum performance, presenting a quadratic low-complexity in comparison to the exponential ML detector.

- **Chapter 8.** This chapter presents the final conclusions followed by a discussion of open research directions for future works within the spatial multiplexing MIMO systems and 5G context.

2 Basic Concepts and System Model

2.1 Introduction

In this chapter, the basic concepts and the wireless communication system model are described. Firstly, some channel types, such as additive white Gaussian noise (AWGN) and AWGN with fading, are described. In the following, there is a brief summary of the diversity techniques used to mitigate the fading effects. Later, a description of some digital modulation schemes is presented, which is an important aspect in order to increase the data rate of a wireless communication service.

Furthermore, in this chapter, the system model of a spatial multiplexing multiple-input multiple-output (MIMO) is discussed, where the benefits and issues when the number of antennas is increased are presented. Later, some theoretical background of MIMO channel model is presented. Finally, an introduction to the conventional spatial multiplexing MIMO detector algorithms is discussed.

For better understanding this chapter is organized as follows. Channel model is described in Section 2.2. The diversity techniques are described in Section 2.3. Digital modulation schemes are described in Section 2.4. Finally, spatial multiplexing MIMO system, its system model, MIMO channel and detector algorithms are detailed in Section 2.5.

2.2 Channel Model

The signals or messages transmitted through wireless channels has as main characteristic the low reliability detection inherent to these channels. The signals that propagate in wireless channels suffer additions, distortions and attenuations due to the peculiarities of the environment, such as noise, fading and attenuation with distance. In the following, some of these degradation factors are described.

2.2.1 AWGN Channel

The additive white Gaussian noise (AWGN) is a basic and generally accepted model for thermal noise and also for shot noise in a communication channel. In this thesis, the density spectrum noise is constant and present in the entire frequency range. The noise is added to the received signal and its samples follow a gaussian probability density function (PDF), given by (PAPOULIS; PILLAI, 2002):

$$f(n) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{\frac{-(n-\mu_n)^2}{2\sigma_n^2}}, \tag{2.1}$$

where μ_n and σ_n^2 denote the mean and variance of the random variable n.

2.2.2 Fading Channel

Fading is a physical phenomenon caused by receiving multiple replicas of the same transmitted signal at the destination. Each of these replicas undergoes differences in phase shift, attenuation and time delay during the path between transmitter and receiver over a wireless communication channel. This can result in either constructive or destructive interference, amplifying or attenuating the signal power at the destination.

In the following, the characteristics of the fading in wireless communication channels in relation to their variability in time and frequency domains of the transmitted signal are described.

2.2.2.1 Slow and Fast Fading

An important parameter used to measure the temporal variability of the wireless channel is the coherence time interval, T_c. This parameter is defined as the time interval where received signals have non-zero amplitude correlation. Based on this parameter, the fading can be classified as slow or fast (PROAKIS; MANOLAKIS, 2006).

2.2.2.2 Slow Fading

Slow fading occurs when the coherence time interval of the channel is greater than the symbol time interval, T_s, that is $T_c \gg T_s$. Thus, the amplitude and phase changes imposed by the channel can be considered roughly constant during the symbol duration.

2.2.2.3 Fast Fading

Fast fading occurs when the coherence time interval of the channel is less than the symbol time interval, that is $T_c \ll T_s$. Therefore, the amplitude and phase changes imposed by the channel varies during the symbol duration.

2.2.2.4 Flat and Selective Fading

Another important parameter related to wireless channels is the coherent bandwidth, B_c, which is a statistical measure of the frequency interval in which the fading presents high correlation. Thus, if two signals are separated by frequency interval greater than B_c, they are affected differently by the channel, that is, both signals are affected by statistically independent fadings (RAPPAPORT, 2002). The presence of flat or selective fading can be determined by comparing the channel coherence bandwidth, B_c, and the bandwidth of the transmitted signal, $B = 1/T_s$.

2.2.2.5 Flat Fading

Non-selective or flat fading appears when the coherence bandwidth is larger than the bandwidth of the transmitted signal, that is, $B_c > B$. In this case, the spectral components of the signal are subject to practically the same magnitude of fading. Thus, there is no inter-symbolic interference between consecutively transmitted signals.

2.2.2.6 Selective Fading

The channels with selective fading appear when the coherence bandwidth is smaller than the bandwidth of the transmitted signal, that is, $B_c < B$. In this scenario, different spectral components of the signal are subjected to different magnitudes of fading causing significant distortion in the signal and, therefore, inter-symbolic interference (ISI) and inter-carrier interference (ICI) occur.

Throughout this thesis, slow and flat fading channels are considered, characterized by a Rayleigh fading, whose PDF is represented by:

$$f(a) = \begin{cases} \frac{a}{\sigma_a^2} e^{\frac{-a^2}{2\sigma_a^2}} & a \geq 0 \\ 0 & a < 0, \end{cases} \tag{2.2}$$

where σ_a is the parameter of the Rayleigh PDF.

The Rayleigh random variable a, which represents the fading amplitude, is obtained by two independent and identically distributed (i.i.d) Gaussian random variables with zero mean and variance σ^2 as follows:

$$a = \sqrt{a_{\mathrm{I}}^2 + a_{\mathrm{Q}}^2}, \tag{2.3}$$

where a_{I} and a_{Q} are Gaussian random variables (PAPOULIS; PILLAI, 2002). Furthermore, the mean square value of the fading amplitude is related to the variance of the in-phase and quadrature components as:

$$\mathbb{E}\{a^2\} = 2\sigma^2. \tag{2.4}$$

In this thesis, it is considered that the mean square value of the fading amplitude is unitary. Therefore, the variances of a_{I} and a_{Q} of the Rayleigh parameter is equal to $\sigma^2 = \frac{1}{2}$.

2.3 Diversity

The increasing demand of wireless communications users that seek high data rate transmission with high reliability and hence low error probability has motivated the study of different diversity techniques in order to combat the performance loss over fading channels (PROAKIS; MANOLAKIS, 2006; TSE; VISWANATH, 2005).

Diversity is a common technique for combating the low performance of fading channels. It is based on the fact that different transmitted signals affected by uncorrelated fadings have a small probability of reaching significant performance loss at the same time. Thus, the basic idea is to have different replicas of the transmitted signal, which are affected by uncorrelated fadings. This number of replicas is known as the diversity order. The greater the diversity order, the better the system performance.

Several diversity techniques have been studied in the literature (TSE; VISWANATH, 2005), among the main ones, there are frequency diversity, time diversity and spatial diversity, which are described below.

2.3.1 Frequency Diversity

This technique consists of transmitting the same information-bearing signal on L different carrier frequencies available in the channel. The frequency separation must be enough to achieve uncorrelated fadings. This frequency separation is a function of the coherent bandwidth, B_c. As a disadvantage, frequency diversity requires greater bandwidth and the number of receivers is equal to the number of carriers used.

2.3.2 Time Diversity

In time diversity a signal is transmitted in L different time slots. The time separation between two sucessive instants of time must exceed the coherent time of the channel, which allow uncorrelated fading channels. The disadvantage of this technique is the reduction of the effective transmission rate, since the time slots used in the transmission increases proportionally with diversity.

2.3.3 Spatial Diversity

Spatial diversity technique, also named as antenna diversity technique, is usually achieved by employing multiple L antennas at the transmitter and/or the receiver. These L antennas must have enough distance separation between them in order to be pairwise independent, in such a way that the received replicas of the signal are uncorrelated. Ideally, the transmitted or received antennas should be separated by a distance

$d = \lambda_w/2$ (TSE; VISWANATH, 2005), where λ_w represents the wavelength, which can be computed by $\lambda_w = c/f_c$, where c is the light speed and f_c is the carrier frequency.

Systems that employ multiple transmit and multiple receive antennas are known in the literature as multiple-input-multiple-output (MIMO) systems (ANDREWS *et al.*, 2007; BIGLIERI *et al.*, 2007). Fig. 2.1 shows spatial diversity with a MIMO system employing L transmit and receive antennas.

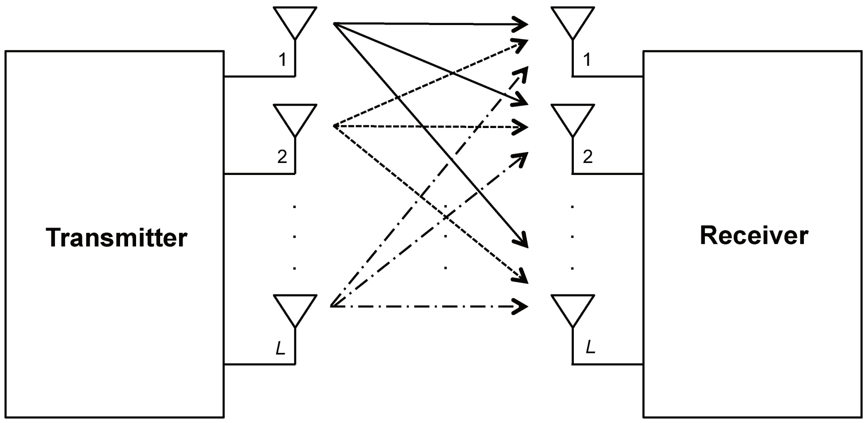


Figure 2.1 – Spatial diversity with a MIMO system employing L transmit and receive antennas.

Both frequency and time diversity technique requires only a single transmit and receive antenna. However, their performance is inferior to spatial diversity technique, which employs multiple antennas. This can be explained by the fact that frequency diversity employs L times larger bandwidth to transmit and time diversity employs L times more time slots to transmit a single information-bearing signal, which also increases the bandwidth L times. Therefore, the efficiency of both diversity technique is reduced by a factor L in comparison to spatial diversity.

2.4 Digital Modulation

The term modulation means moving the spectrum of a baseband signal to a carrier frequency f_c different from zero. Thus, digital modulation appears from the need to transmit a digital signal through a channel that presents a frequency response around f_c (ALMEIDA, 2014). Note that a digital signal can be modulated by changing one of the following parameters of a carrier: amplitude, phase or frequency.

Basically, digital modulation maps one or more bits of information into a set of waveforms that can be transmitted in the channel. The set of waveforms can be graphically represented by a constellation diagram as it is shown in Fig. 2.2. In the following, the digital modulation schemes used in this thesis are described below. After that, some additional concepts referent to modulation are presented.

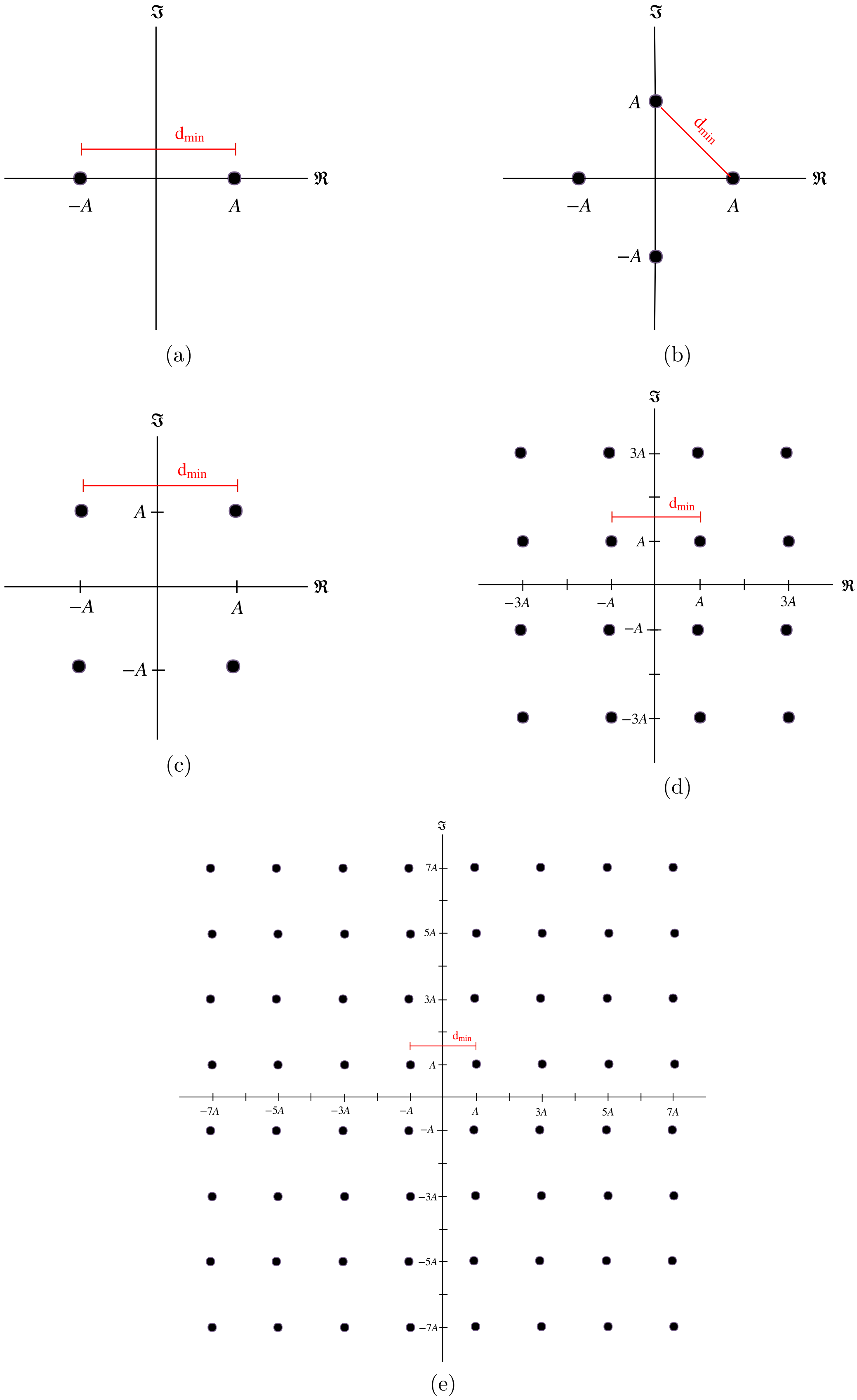


Figure 2.2 – Constellation diagrams for (a) BPSK, (b) QPSK, (c) 4-QAM, (d) 16-QAM and (e) 64-QAM modulations.

2.4.1 BPSK Modulation

The binary phase shift keying (BPSK or 2-PSK) is a two phase modulation scheme in which the phase of a carrier is varied according to the bit to be transmitted. The amplitude and frequency of the carrier are constant. Fig. 2.2a shows the BPSK constellation diagram. The BPSK modulated signal can be written as (ALMEIDA, 2014):

$$s_n(t) = \sum_{i=-\infty}^{\infty} A\cos[2\pi f_c(t - iT_b) + \pi n], \quad \text{for } n = 0, 1, \tag{2.5}$$

where A represents the amplitude, $1/T_b$ the bit rate. Therefore, the energy per bit of the BPSK signal, given by (2.5), is $E_b = A^2T_b/2$.

2.4.2 QPSK Modulation

The quadrature phase-shift keying modulation is without doubt, the most important of all PSK modulations (ALMEIDA, 2014). QPSK modulation is equivalent to two BPSK modulations, one being transmitted in phase and the other in quadrature. Fig. 2.2b shows the QPSK constellation diagram. Thus, the QPSK modulated signal can be denoted by:

$$s_n(t) = \sum_{i=-\infty}^{\infty} A\cos\left[2\pi f_c(t - iT_s) + \frac{\pi n}{2}\right], \quad \text{for } n = 0, 1, 2, 3, \tag{2.6}$$

Note from (2.6) that the carrier can assume 4 different phase values. Therefore, QPSK can modulate two bits per symbol and consequently $T_s = 2T_b$. The energy per bit of the QPSK signal is $E_b = A^2T_b$.

2.4.3 M-QAM Modulation

Quadrature amplitude modulation (QAM) is a digital modulation in which the amplitude and phase of two orthogonal carriers are simultaneously varied. Thus, M-QAM requires less energy per symbol than M-PSK modulations for $M \geq 8$. For M-QAM modulation, there are M symbols and $\log_2 M$ bits, where M is known as the modulation order. An M-QAM modulated signal can be written as (ALMEIDA, 2014):

$$s_n(t) = \sum_{i=-\infty}^{\infty} a_i\cos[2\pi f_c(t - iT_s) + \phi] - \sum_{i=-\infty}^{\infty} b_i\sin[2\pi f_c(t - iT_s) + \phi], \tag{2.7}$$

where a_i and b_i are independent random amplitude variables, that can assume the values $\pm A, \pm 3A, \cdots, \pm(\sqrt{M} - 1)A$. Fig. 2.2c, 2.2d and 2.2e show 4-QAM, 16-QAM and 64-QAM modulation constellation diagrams, respectively. The energy per symbol of M-QAM modulated signal, expressed in (2.7), is given by:

$$E_s = \frac{(M-1)A^2T_s}{3}. \tag{2.8}$$

Furthermore, the relation between symbol energy and bit energy is obtained as:

$$\begin{aligned} E_s &= bE_b \\ &= E_b \log_2 M. \end{aligned} \tag{2.9}$$

Note from Fig. 2.2b and Fig. 2.2c, that 4-QAM can be obtained by a phase shift of $\pi/4$ on QPSK and an amplitude increase of $\sqrt{2}$ times. Thus, their performances are equivalents, as long as their average powers are equal.

2.4.4 Gray Mapping

Gray mapping is a technique that digital modulation schemes often use to minimize the bit error rate (BER). It consists of labeling the modulation symbols, so that the binary representations of adjacent symbols differ by only one bit. Thus, one symbol error corresponds exactly to one bit error.

2.4.5 Euclidean Minimum Distance

The distance between any two symbols of a constellation determines the probability of one of them to be confused with the other. This distance is named Euclidean distance. Moreover, the smaller the Euclidean distance between two symbols, the greater is the susceptibility of error. The minimum Euclidean distance of the constellation ($\mathrm{d}_{\mathrm{min}}$) represents the minimum distance between two symbols of a constellation (ALMEIDA, 2014).

2.4.6 Bit Rate

By considering a digital modulation with M symbols and since $1/T_s$ symbols are transmitted per second, the bit rate of this modulation is given by (ALMEIDA, 2014)

$$R_b = \frac{\log_2 M}{T_s}, \tag{2.10}$$

where its unit is bits/s, or loosely, b/s. Note that the bit rate given by (2.10) can be increased by increasing the modulation order M, or by increasing the symbol rate $1/T_s$. However, the symbol rate is bounded by the bandwidth B of the channel. Thus, the combination of these two parameters limits the available bit rate of a given channel.

2.4.7 Spectral Efficiency

Spectral efficiency refers to the information rate that can be transmitted through a given channel bandwidth (BARRY *et al.*, 2003). Thus, spectral efficiency ξ

is given as the ratio of the information bit rate R_b to the channel bandwidth B, that is:

$$\xi = \frac{R_b}{B}, \tag{2.11}$$

where its unit is bits/sec-Hz, or loosely b/s/Hz. Considering that the minimum Nyquist bandwidth is equal to $B = 1/T_s$ and substituting (2.10) into (2.11), the spectral efficiency can be rewritten as:

$$\xi = \log_2 M. \tag{2.12}$$

2.5 Spatial Multiplexing MIMO System

In ordinary wireless communication systems, the structure of a link is composed of one transmit and one receive antenna. This system, denoted as single-input single-output (SISO), was widely used in the past few decades for both mobile and fixed wireless communications (ANDREWS *et al.*, 2007). For SISO systems it is well known that given a fixed bandwidth, bit rate and consequently spectral efficiency can only be increased by increasing the modulation order. However, this approach could have limitations to support higher data rates, once higher modulation orders are more susceptible to fading effects (ALMEIDA, 2014). Therefore, approaches with increased spectral efficiencies are required.

From a spectral efficiency standpoint, the most interesting type of MIMO systems is spatial multiplexing (SM) (CLERCKX; OESTGES, 2013), which refers to the use of multiple transmit and receive antennas in order to accomplish a multiplexing gain, thus enhancing the bit rate and spectral efficiency with the number of transmit antennas employed without requiring extra bandwidth. In fact, it has been demostrated in (MINANGO, 2014) that to achieve a given spectral efficiency, using a small modulation order and increasing the number of transmit antennas is more efficient than using a single transmit antenna and increasing the modulation order.

2.5.1 System Model

The basic idea behind spatial multiplexing is transmitting different symbols across different transmit antennas at the same time and in the same bandwidth in order to increase the bit rate and consequently the spectral efficiency. At the receiver, the transmitted symbols are detected by employing an interference cancellation-type of algorithm.

In this thesis, a spatial multiplexing MIMO system with N_T transmit and N_R receive antennas, where N_T symbols are simultaneously transmitted over the N_T antennas is considered, as depicted in Fig. 2.3. Thus, the corresponding received signal vector $\mathbf{y} = [y_1, y_2, \cdots, y_{N_R}]^T \in \mathbb{C}^{N_R \times 1}$ at the N_R antennas is given by:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.13}$$

where $\mathbf{x} = [x_1, x_2, \cdots, x_{N_T}]^T \in M^{N_T}$ is the transmitted signal vector, whose elements belong to a given symbol constellation with modulation order M, $\mathbf{n} = [n_1, n_2, \cdots, n_{N_R}]^T \in \mathbb{C}^{N_R \times 1}$ represents the additive complex white Gaussian noise (AWGN) vector, whose entries consists of independent and identically distributed (i.i.d) complex Gaussian random variables with zero mean and variance σ_n^2 and $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ denotes the flat Rayleigh fading MIMO channel matrix, whose entries consists of i.i.d. complex Gaussian random variables with zero mean and unit variance. The MIMO channel matrix in particular is of the form:

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1} & h_{N_R,2} & \cdots & h_{N_R,N_T} \end{bmatrix}, \tag{2.14}$$

where $h_{i,j} \sim \mathcal{CN}(0,1)$ denotes the entry associated with the link between the i-th receive antenna and the j-th transmit antenna.

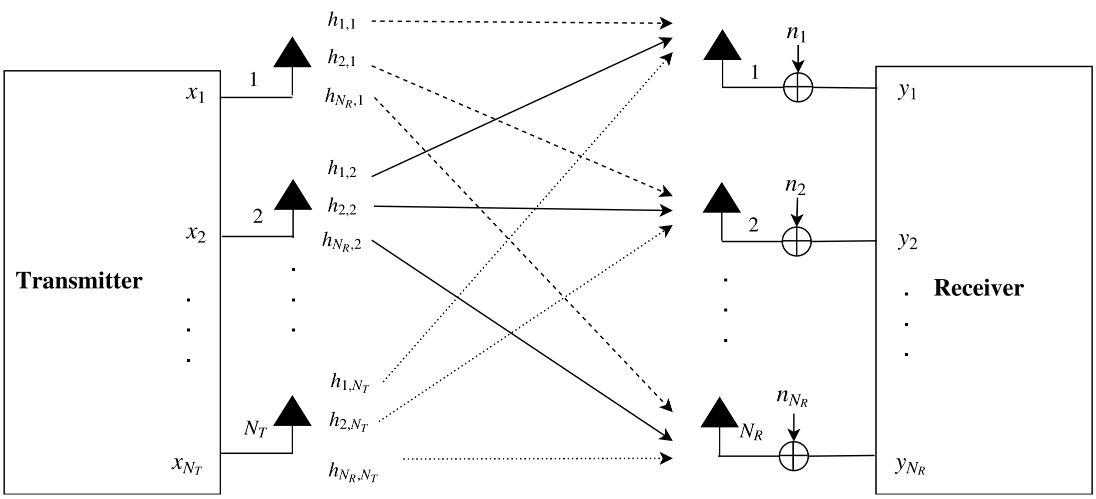


Figure 2.3 – Spatial multiplexing MIMO system.

Since the elements of the transmitted signal vector $\mathbf{x}$ are independent, then, the vector $\mathbf{x}$ has zero-mean and a covariance matrix $\mathbf{R_x} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \sigma_x^2 \mathbf{I}_{N_T}$, where σ_x^2 is the signal power of each transmit antenna. Similarly, the covariance of the independent AWGN vectors is $\mathbf{R_n} = \mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_{N_R}$. Furthermore, the vectors $\mathbf{x}$ and $\mathbf{n}$ are assumed to be independent, that is $\mathbb{E}\{\mathbf{x}\mathbf{n}^H\} = \mathbf{0}$. Based on this, the covariance matrix of $\mathbf{y}$ for a given $\mathbf{H}$ can be found, using the fact that, if both $\mathbf{x}$ and $\mathbf{n}$ have zero mean, then $\mathbf{y}$ has zero mean as well:

$$\begin{aligned}\mathbf{R_y} = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\} &= \mathbb{E}\{(\mathbf{Hx}+\mathbf{n})(\mathbf{Hx}+\mathbf{n})^H\} \\ &= \mathbf{H}\mathbf{R_x}\mathbf{H}^H + \mathbf{R_n} \\ &= \sigma_x^2\mathbf{H}\mathbf{H}^H + \sigma_n^2\mathbf{I}_{N_R} \end{aligned} \tag{2.15}$$

Furthermore, from (2.13), the overall signal-to-noise ratio (SNR) at the receiver is:

$$\gamma = \frac{\mathbb{E}\{\|\mathbf{Hx}\|^2\}}{\mathbb{E}\{\|\mathbf{n}\|^2\}} = \frac{\mathbb{E}\{\sum_{i=1}^{N_R}\sum_{j=1}^{N_T}|h_{i,j}x_j|^2\}}{\mathbb{E}\{\sum_{i=1}^{N_R}|n_i|^2\}} = \frac{N_T N_R \sigma_x^2}{N_R\sigma_n^2} = N_T\frac{\sigma_x^2}{\sigma_n^2}, \tag{2.16}$$

which is in fact the same as N_T times the SNR of a SISO system, since each receive antenna receives the incoming power from N_T transmit antennas, while each one of the N_R receive antennas perceive the same amount of noise power as of a SISO system. This analysis is only valid on the average and when $h_{i,j} \sim \mathcal{CN}(0,1)$, i.e., y_i receives the sum of N_T transmitted symbols weighted by unitary power channel random variables $\mathbb{E}\{|h_{i,j}|^2\} = 1$.

Once the signal power of each transmit antenna and the noise power of each receive antenna can be expressed, respectively, as $\sigma_x^2 = E_s/T_s$ and $\sigma_n^2 = BN_0$, where N_0 is the noise unilateral power spectral density. Then, its ratio is given by:

$$\frac{\sigma_x^2}{\sigma_n^2} = \frac{E_s}{T_s B N_0} = \frac{E_s}{N_0}, \tag{2.17}$$

where $B = 1/T_s$ by assuming the minimum Nyquist bandwidth and E_s/N_0 denotes the average energy per symbol to the noise power spectral density. Since the average energy per bit is $E_b = E_s/\log_2 M$ and by considering (2.17), the overall SNR given by (2.16) can be rewritten as:

$$\gamma = N_T\frac{E_b}{N_0}\log_2 M, \tag{2.18}$$

where E_b/N_0 represents the energy per bit to the noise power spectral density ratio.

2.5.1.1 Spectral Efficiency

In spatial multiplexing MIMO systems, by employing N_T transmit antennas, the overall bit rate compared to SISO systems is thus enhanced by a factor of N_T. Therefore, the bit rate of a MIMO system measured in b/s is given by:

$$R_b = N_T\frac{\log_2 M}{T_s} \quad \text{b/s}. \tag{2.19}$$

where T_s is the symbol time interval. Since spatial multiplexing MIMO systems do not require extra bandwidth to increase the bit rate, thus, their bandwidth is equal to $B = 1/T_s$ Hz and as a consequence the spectral efficiency of a MIMO system measured in b/s/Hz is given by:

$$\xi = \frac{R_b}{B} = N_T\log_2(M). \tag{2.20}$$

This is certainly exciting, as it implies that adding transmit antennas can increase the viability of the high data rates desired in wireless communications.

2.5.2 MIMO Channel

This section presents some relevant theoretical background about MIMO channels in order to address the properties and issues which arise in spatial multiplexing MIMO systems with large number of antennas.

2.5.2.1 Useful Definitions

2.5.2.1.1 Orthogonal Matrix

A $n \times m$ matrix $\mathbf{A}$ is an orthogonal matrix if (GOLUB; LOAN, 1996):

$$\mathbf{A}^H \mathbf{A} = \mathbf{I}_m, \tag{2.21}$$

where $\mathbf{A}^H$ is the conjugate transpose or the Hermitian of $\mathbf{A}$ and $\mathbf{I}_m$ is the $m \times m$ identity matrix. Note that to satisfy (2.21), the column vectors of $\mathbf{A}$ are pairwise orthogonal; likewise for the row vectors. Furthermore, an orthogonal matrix is always invertible, with:

$$\mathbf{A}^{-1} = \mathbf{A}^H. \tag{2.22}$$

Since $\mathbf{A}$ is an orthogonal matrix, so are $\mathbf{A}^H$ and $\mathbf{A}^{-1}$. Moreover, the relation given by (2.22) makes orthogonal matrices particularly easy to compute with, since the Hermitian operation is much simpler than computing an inverse. Moreover, since $\mathbf{A}$ is an orthogonal matrix, its eigenvalues are equal, i.e., $\lambda_1 = \lambda_2 = \cdots = \lambda_m$

2.5.2.1.2 Diagonal Dominant Matrix

An $n \times n$ square matrix $\mathbf{A}$ is diagonal dominant (DD) if for every row or column of $\mathbf{A}$, the absolute value of its diagonal entry is larger than or equal to the sum of the absolute value of all the other (non-diagonal) entries in that row or column. More exactly, matrix $\mathbf{A}$ is DD if:

$$|a_{i,i}| \geq \sum_{j=1, j\neq i}^{n} |a_{i,j}| \quad \text{or} \quad |a_{i,i}| \geq \sum_{j=1, j\neq i}^{n} |a_{j,i}| \tag{2.23}$$

where $a_{i,j}$ denotes the entry in the i-th row and j-th column of matrix $\mathbf{A}$. On the other hand, matrix $\mathbf{A}$ is called strictly diagonal dominant (SDD) if:

$$|a_{i,i}| > \sum_{j=1, j\neq i}^{n} |a_{i,j}| \quad \text{or} \quad |a_{i,i}| > \sum_{j=1, j\neq i}^{n} |a_{j,i}| \tag{2.24}$$

The following metric is considered in order to know the diagonal dominance of matrix $\mathbf{A}$:

$$D = \frac{\sum_{i=1}^{n} |a_{i,i}|}{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} |a_{i,j}|}. \tag{2.25}$$

2.5.2.1.3 Hermitian and Positive Definite Matrix

A $n \times n$ complex square matrix $\mathbf{A}$ is Hermitian, if it is equal to its own conjugate transpose, that is:

$$\mathbf{A} = \mathbf{A}^H. \tag{2.26}$$

On the other hand, a $n \times n$ Hermitian matrix $\mathbf{A}$ is said to be positive-definite if:

$$\mathbf{z}^H \mathbf{A} \mathbf{z} > 0 \tag{2.27}$$

for all nonzero complex vectors $\mathbf{z} \in \mathcal{C}^n$, where $\mathbf{z}^H$ represents the conjugate transpose of the vector $\mathbf{z}$.

2.5.2.1.4 Eigen-decomposition

The eigen-decomposition of a $n \times n$ square matrix $\mathbf{A}$ is given by (GOLUB; LOAN, 1996):

$$\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^H, \tag{2.28}$$

where the matrix $\mathbf{V}$ contains the eigenvectors of $\mathbf{A}$, and $\mathbf{\Lambda} = \text{diag} = [\lambda_1, \lambda_2, \cdots, \lambda_n]$ denotes the diagonal matrix containing the eigenvalues of $\mathbf{A}$. Matrix $\mathbf{V}$ is invertible as long as matrix $\mathbf{A}$ is Hermitian. Furthermore, if matrix $\mathbf{A}$ is Hermitian and positive-definite (HPD), then all its eigenvalues are positive and real.

2.5.2.1.5 Spectral Radius

The spectral radius of a $n \times n$ square matrix $\mathbf{A}$, denoted as $\rho(\mathbf{A})$, is the largest absolute value of its eigenvalues. Thus, let $\lambda_1, \cdots, \lambda_i, \cdots, \lambda_n$ be the eigenvalues of matrix $\mathbf{A}$, its spectral radius is defined as:

$$\begin{aligned} \rho(\mathbf{A}) &= \max_{1 \leq i \leq n} |\lambda_i| \\ &= |\lambda_{\max}(\mathbf{A})|, \end{aligned} \tag{2.29}$$

where $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$.

2.5.2.1.6 Condition Number

Applying the eigen-decomposition given by (2.28) to a $n \times n$ HPD matrix $\mathbf{A}$ and by assuming that the n nonzero eigenvalues are sorted in decreasing order of magnitude, that is $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then the condition number of $\mathbf{A}$ is given by (GOLUB; LOAN, 1996):

$$\kappa(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}, \tag{2.30}$$

where $\lambda_{\min}(\mathbf{A})$ represents the smallest eigenvalue of $\mathbf{A}$. A condition number of one, $\kappa(\mathbf{A}) = 1$, means that matrix $\mathbf{A}$ is orthogonal therefore $\lambda_{\max}(\mathbf{A}) = \lambda_{\min}(\mathbf{A})$, while a large condition number, $\kappa(\mathbf{A}) \gg 1$, implies that matrix $\mathbf{A}$ is highly non-orthogonal or ill-conditioned.

2.5.2.1.7 Vector and Matrix Norm

In this thesis, the Euclidean norm of a vector $\mathbf{z} \in \mathcal{C}^{n\times 1}$ is given by:

$$\|\mathbf{z}\| = \left(\mathbf{z}^H\mathbf{z}\right)^{1/2}. \tag{2.31}$$

On the other hand, given a $n \times m$ complex matrix $\mathbf{A}$, a matrix norm $\|\mathbf{A}\|$ is a non-negative number associated with $\mathbf{A}$. Following, some matrix norms are discussed:

- The maximum absolute column sum norm or 1-norm $\|\mathbf{A}\|_1$ is defined as:

$$\|\mathbf{A}\|_1 = \max_{1\leq j\leq m} \sum_{i=1}^{n} |a_{i,j}|. \tag{2.32}$$

- The maximum absolute row sum norm or infinite-norm $\|\mathbf{A}\|_\infty$ is given by:

$$\|\mathbf{A}\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{m} |a_{i,j}|. \tag{2.33}$$

- Since matrix $\mathbf{A}$ is HPD, its spectral norm or 2-norm $\|\mathbf{A}\|_2$ is the square root of the largest eigenvalue of $\mathbf{A}$, that is:

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A})}. \tag{2.34}$$

- The Frobenius norm $\|\mathbf{A}\|_F$ is given by:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{j=1}^{m}\sum_{i=1}^{n} |a_{i,j}|^2} = \sqrt{\mathrm{Tr}\left[\mathbf{A}^H\mathbf{A}\right]}, \tag{2.35}$$

  where $\mathrm{Tr}[\cdot]$ denotes the trace of a matrix.

2.5.2.2 Channel Hardening

The advantages of spatial multiplexing MIMO systems with large number of antennas are the increase in the spectral efficiency and diversity gain. Furthermore, the large dimensionality results in other advantages that do not appear in spatial multiplexing MIMO systems with low number of antennas as shown below.

Once the channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ given by (2.14) becomes larger, that is both N_T and N_R increases with ratio $\beta = N_T/N_R$, the distribution of its eigenvalues becomes less sensitive to the current distribution of its entries $h_{i,j}$ of the channel matrix $\mathbf{H}$. This is a result of the Marčenko-Pastur law, which is described in the following theorem (TULINO; VERDú, 2004).

Theorem 1 *Since the entries of the channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ are i.i.d. complex Gaussian random variables with zero mean and unit variance, that is $\mathcal{CN}(0,1)$, then the distribution of the eigenvalues of $\mathbf{H}^H\mathbf{H}$, which is a HPD matrix, converges, as $N_T, N_R \to \infty$ to the PDF $f(\lambda)$ given by:*

$$f(\lambda) = \left(1 - \frac{1}{\beta}\right)_+ \delta(\lambda) + \frac{\sqrt{(\lambda - a)_+(b-\lambda)_+}}{2\pi\beta\lambda}, \tag{2.36}$$

where $\delta(\lambda)$ is the Dirac delta function, $(x)_+ \triangleq \max(0,x)$, $a \triangleq (1-\sqrt{\beta})^2$ and $b \triangleq (1+\sqrt{\beta})^2$ for $a \leq \lambda \leq b$ and $0 \leq \beta \leq 1$.

Fig. 2.4 shows the theoretical PDF of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ given by (2.36) and the equivalent simulated PDF for different β. The simulated PDF has been obtained by averaging over 10^5 MIMO channels. From this figure, notice that when β decreases, the range of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ decreases too, which means that the eigenvalues becomes deterministic as the number of receive antennas is larger than the number of transmit antennas, once $\beta = N_T/N_R$. Note that for $\beta = 0$, there is an impulse at $\lambda = 1$. This behavior is known in the literature as channel hardening (TULINO; VERDú, 2004).

Moreover, an interesting feature of the channel hardening is that as the dimensionality of $\mathbf{H}$ increases, the off-diagonal entries of the matrix $\mathbf{H}^H\mathbf{H}$ become increasingly smaller in comparison to the diagonal entries. In particular, two scenarios are distinguished in literature (RUSEK *et al.*, 2013; MARZETTA *et al.*, 2016):

- Larger N_R than N_T, that is $N_R \gg N_T$, which is named as massive MIMO channel.
- Large N_R and N_T with $N_R = N_T$, which is named as symmetric MIMO channel.

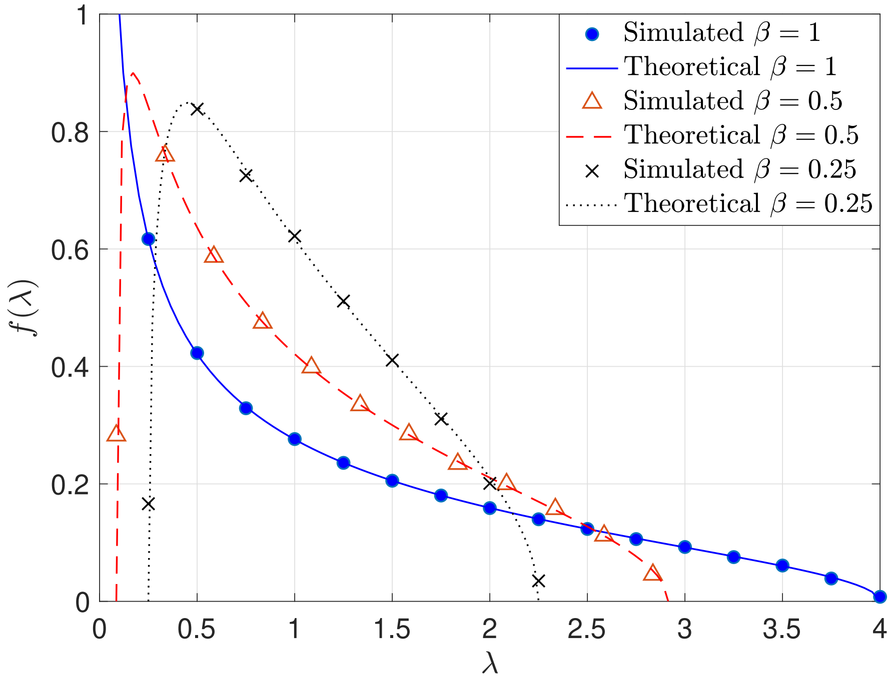


Figure 2.4 – PDF of the eigenvalues of $\mathbf{H}^H\mathbf{H}$ for different values of the ratio β.

2.5.2.3 Massive MIMO Channel

Consider a spatial multiplexing MIMO system, where the number of receive antennas is much larger than the number of transmit antennas, that is $N_R \gg N_T$. In this case, the product $\mathbf{H}^H\mathbf{H}$ convergences asymptotically to $N_R\mathbf{I}_{N_T}$, once N_R tends to infinity for fixed N_T, where each entry of the tall channel matrix $\mathbf{H}$ is an i.i.d. random variable that follows a $\mathcal{CN}(0,1)$ distribution, that is:

$$\lim_{N_R\to\infty}\left(\mathbf{H}^H\mathbf{H}\right) = N_R\mathbf{I}_{N_T}. \tag{2.37}$$

Under the condition (2.37), all eigenvalues of $\mathbf{H}^H\mathbf{H}$ become equal to one, i.e., $\lambda_1 = \lambda_2 = \cdots = \lambda_{N_T} = 1$. Thus, its condition number is equal to one and therefore the matrix $\mathbf{H}$ becomes asymptotically orthogonal. More exactly, the column channel vectors of $\mathbf{H}$ satisfies (RUSEK *et al.*, 2013; NGO *et al.*, 2013):

$$\mathbf{h}_i^H\mathbf{h}_j = \begin{cases} 0, & i,j = 1,\cdots,N_T, \quad i \neq j \\ \|\mathbf{h}_k\|^2 = N_R, & k = 1,\cdots,N_T \end{cases} \tag{2.38}$$

Note that the idealistic conditions given by (2.37) is based on the assumptions that the number of receive antennas N_R tends to infinite. However, since the employment of an infinite number of receive antennas is impractical, it seems natural to know the number of receive antennas required to the channel begins to exhibit this condition.

In practical spatial multiplexing massive MIMO systems, the number of receive antennas N_R is large. However, condition (2.37) shows the asymptotic results when N_R

goes to infinity. Thus, it does not give an account for how close the product $\mathbf{H}^H\mathbf{H}$ is from the identity matrix $N_R\mathbf{I}_{N_T}$, when N_R is finite. The convergence of $\mathbf{H}^H\mathbf{H}$ to $N_R\mathbf{I}_{N_T}$ can be numerically evaluated by considering the condition number and the Frobenius norm of the error matrix $\mathbf{E} = \mathbf{H}^H\mathbf{H} - N_R\mathbf{I}_{N_T}$ metrics, given, respectively, by (2.30) and (2.35). Both metrics are evaluated and averaged over 10^5 channel realizations.

Fig. 2.5 shows the condition number of $\mathbf{H}^H\mathbf{H}$ versus the number of receive antennas N_R, where N_T is fixed at 16 and 64 transmit antennas. It is evident that for both number of transmit antennas of N_T, the condition number tends to one, that is $\kappa\left(\mathbf{H}^H\mathbf{H}\right) \approx 1$, when N_R is very large in the order of ten thousands receive antennas. Furthermore, note that even for values of N_R as large as 100 receive antennas, the condition number for $N_T = 64$ antennas is more than 12 times larger than the corresponding condition number for $N_T = 16$ antennas. Therefore, when the number of transmit antennas increases, it is necessary to employ more receive antennas in order to satisfy the relation $N_R \gg N_T$, which is the principle of spatial multiplexing massive MIMO systems.

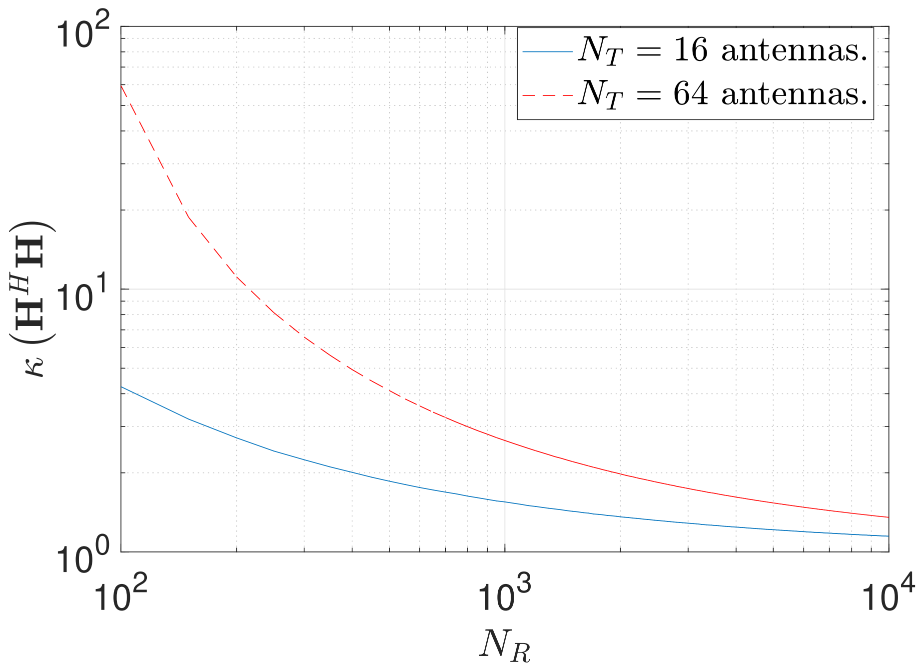


Figure 2.5 – Condition number of $\mathbf{H}^H\mathbf{H}$ against the number of receive antennas N_R with the number of transmit antennas N_T fixed, where $N_R \gg N_T$.

On the other hand, Fig. 2.6 shows the Frobenius norm of the error matrix $\mathbf{E}$ as a function of the number of receive antennas N_R, for N_T =16 and 64 transmit antennas. Matrix $\mathbf{E}$ represents the mean absolute error of $\mathbf{H}^H\mathbf{H}$ from the identity matrix, $N_R\mathbf{I}_{N_T}$. From this figure, notice that this error tends quickly to zero as the number of receive antennas increases, which demonstrate that the tall channel matrix $\mathbf{H}$, for $N_R \gg N_T$, becomes very well conditioned and deterministic.

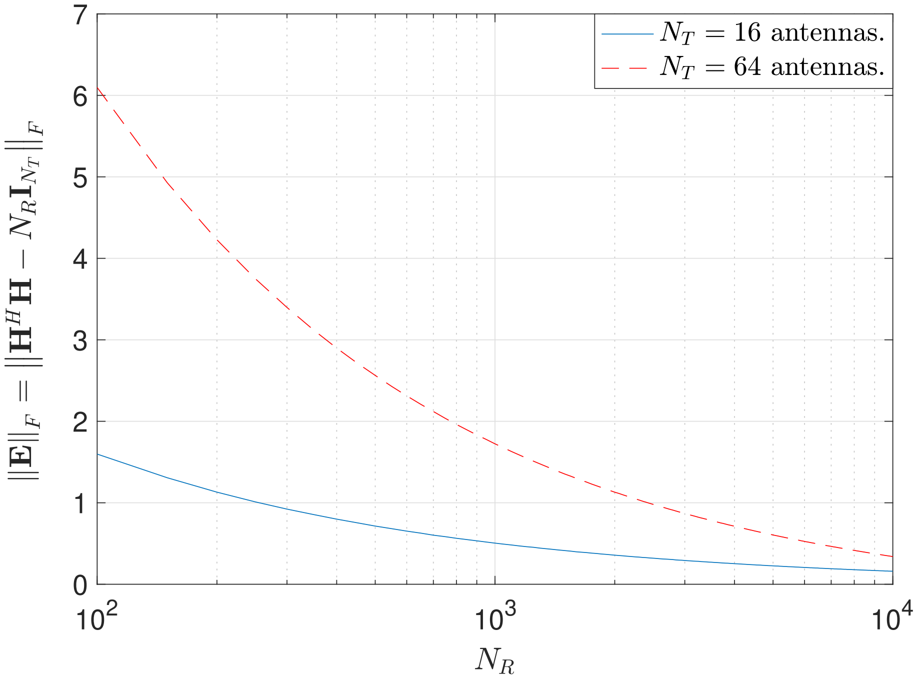


Figure 2.6 – Frobenius norm of the error matrix $\mathbf{E}$ against the number of receive antennas N_R with the number of transmit antennas N_T fixed, where $N_R \gg N_T$.

Finally, for $N_R \gg N_T$ and using that the eigenvalues of $\mathbf{H}^H\mathbf{H}$ follow the PDF given by (2.36) and considering the analysis conducted in (EDELMAN, 1989, Proposition 5.1 and 6.1), the smallest and the largest eigenvalue of $\mathbf{H}^H\mathbf{H}$ can be approximated, respectively, by:

$$\lambda_{\min}\left(\mathbf{H}^H\mathbf{H}\right) \approx N_R\left(1 - \sqrt{\beta}\right)^2, \tag{2.39a}$$

$$\lambda_{\max}\left(\mathbf{H}^H\mathbf{H}\right) \approx N_R\left(1 + \sqrt{\beta}\right)^2, \tag{2.39b}$$

where $\beta = N_T/N_R$. Fig. 2.7 shows a comparison between the approximated and simulated smallest and largest eigenvalues of $\mathbf{H}^H\mathbf{H}$ as a function of the ratio β, where $N_T = 16$ antennas. Note that the approximated smallest and largest eigenvalues of $\mathbf{H}^H\mathbf{H}$ are given, respectively, by (2.39a) and (2.39b). Note that the approximated values of the smallest and largest eigenvalues of $\mathbf{H}^H\mathbf{H}$ are quite close to the simulated results.

2.5.2.4 Symmetric MIMO Channel

In the case that of both N_R and N_T grow to infinity, that is $N_R, N_T \to \infty$, with $N_R = N_T$, the eigenvalues of $\mathbf{H}^H\mathbf{H}$ are not equal anymore as in the massive MIMO scenario and therefore the eigenvalues range does not decrease any more. Thus, in general the condition number is larger than 1, that is $\kappa(\mathbf{H}^H\mathbf{H}) \gg 1$, which implies that the channel matrix $\mathbf{H}$ is highly non-orthogonal or ill-conditioned.

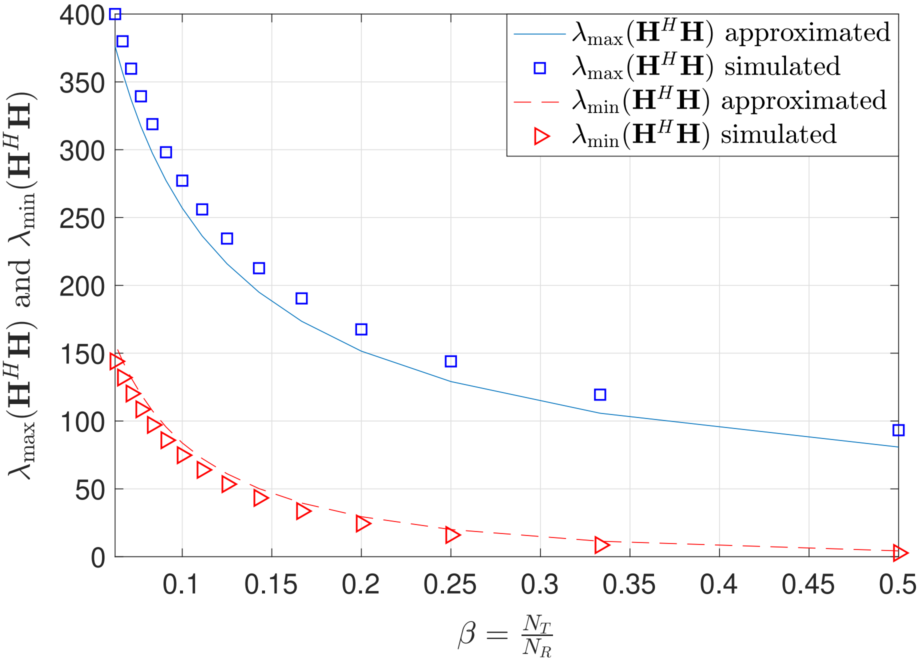


Figure 2.7 – The largest and the smallest eigenvalue of the matrix $\mathbf{H}^H\mathbf{H}$ against the ratio $\beta = N_T/N_R$, where $N_T = 16$ antennas.

To corroborate the last affirmation, Fig. 2.8 shows the condition number of $\mathbf{H}^H\mathbf{H}$ as a function of the number of receive and transmit antennas, where $N_R = N_T$. This simulation result is averaged over 10^5 channel realizations. From this figure, notice that as N_T and N_R increase, the condition number increases too. Therefore, channel matrix $\mathbf{H}$ become even worse ill-conditioned as $N_R, N_T \to \infty$. For example, for a channel matrix $\mathbf{H}$ with $N_R = N_T = 1000$ antennas, its condition number, on average, is $\kappa(\mathbf{H}^H\mathbf{H}) \approx 3500$, therefore $\mathbf{H}$ is highly non-orthogonal.

On the other hand, since the dimensionality of the symmetric MIMO channel matrix $\mathbf{H}$ increases with N_R and N_T, the off-diagonal entries of the product $\mathbf{H}^H\mathbf{H}$ becomes increasingly weaker in comparison to the diagonal entries. This is illustrated in Fig. 2.9, where the average diagonal dominance D metric given by (2.25) for $\mathbf{H}^H\mathbf{H}$ is plotted as a function of the number of receive and transmit antennas with $N_R = N_T$. From this figure, notice that as N_R and N_T increase, $\mathbf{H}^H\mathbf{H}$ is more DD, even for a large number of antennas, i.e., $N_R = N_T > 100$. the matrix $\mathbf{H}^H\mathbf{H}$ can be considered as SDD.

2.5.3 Spatial Multiplexing MIMO Detectors

A predominant aspect of spatial multiplexing MIMO systems is the effectiveness of the detection algorithms (BAI; CHOI, 2014). In comparison to SISO detection in fading channels, detection in spatial multiplexing MIMO systems is more complicated.

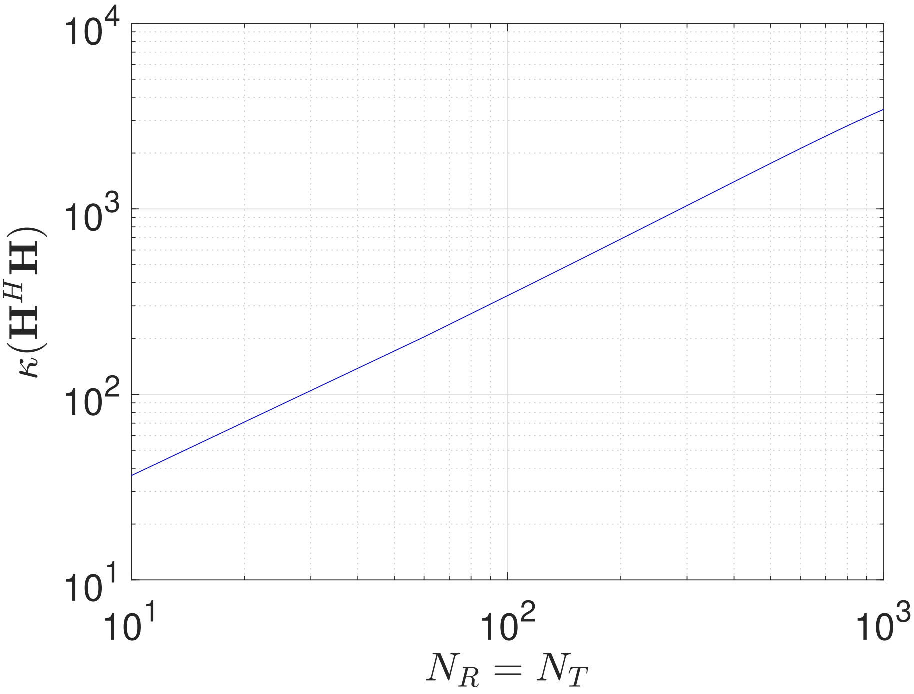


Figure 2.8 – Condition number of $\mathbf{H}^H\mathbf{H}$ against the number of receive and transmit antennas where $N_R = N_T$.

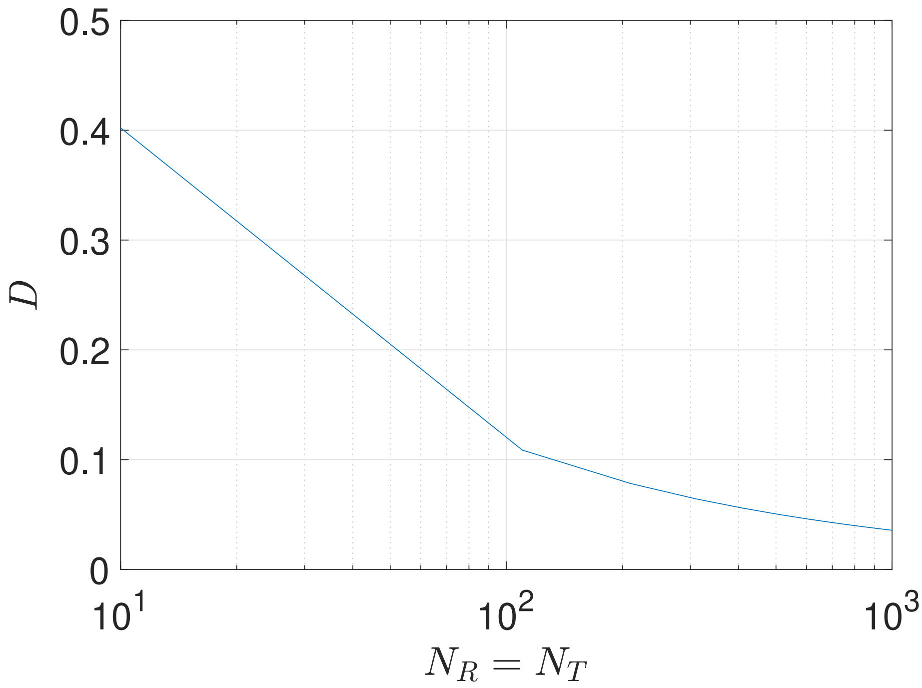


Figure 2.9 – Diagonal dominance metric of $\mathbf{H}^H\mathbf{H}$ as a function of the number of receive and transmit antennas with $N_R = N_T$.

This is because, besides the fading, the N_R receive antennas need to combat against spatial interference produced by simultaneous transmission from the N_T transmit antennas. Thus, efficient detection algorithms in the presence of fading and spatial interference are a demanding task for spatial multiplexing MIMO systems.

In this section, the most important detector algorithms for spatial multiplexing MIMO systems are examined. For the detection process, it is assumed that the channel state information (CSI) at the receiver side was previously estimated.

2.5.3.1 Optimum Detector

Considering the spatial multiplexing MIMO system model given by (2.13), the main purpose of a detector is, from the given received signal vector $\mathbf{y}$ and from the MIMO channel matrix $\mathbf{H}$, to reveal the transmitted signal vector $\mathbf{x}$. Supposing that the transmit vector $\mathbf{x}$ is chosen uniformly from a symbol constellation with modulation order M, the detector that minimizes the probability of error $P(\mathbf{x} \neq \hat{\mathbf{x}} \mid \mathbf{y}, \mathbf{H})$ is the optimum, which is achieved by the maximum likelihood (ML) detector described below.

2.5.3.1.1 Maximum Likelihood Detector

Maximum likelihood (ML) detector achieves the optimum performance by choosing the transmit signal vector $\hat{\mathbf{x}}_{\text{ML}}$ among all possible $\hat{\mathbf{x}} \in M^{N_T}$ transmit signal vectors which is the nearest, in terms of square Euclidean distance, to the received signal vector $\mathbf{y}$ for a given channel matrix $\mathbf{H}$. Mathematically, this is established as (ANDREWS *et al.*, 2007):

$$\hat{\mathbf{x}}_{\text{ML}} = \underset{\hat{\mathbf{x}} \in M^{N_T}}{\arg\min} \left\| \mathbf{y} - \mathbf{H}\hat{\mathbf{x}} \right\|^2 . \tag{2.40}$$

where $\left\| \mathbf{y} - \mathbf{H}\hat{\mathbf{x}} \right\|^2$ denotes the ML metric.

Computing (2.40) through an exhaustive search requires exponential complexity in N_T, that is $O(M^{N_T})$. Even though ML detector attains the optimum performance and a diversity order of N_R (full diversity), its complexity is not feasible even for a moderate number of transmit antennas ($N_T > 5$) and constellation with order 4, i.e., 4-QAM. A more efficient ML implementation can be obtained using the sphere detector (SD) (ANDREWS *et al.*, 2007; BAI; CHOI, 2014).

2.5.3.1.2 Sphere Detector

The sphere detector (SD) is a variant algorithm of the ML detector with lower complexity than the ML detector (MURUGAN *et al.*, 2006; JALDEN; OTTERSTEN, 2005). The idea behind SD is to calculate the same ML metric given by (2.40) by only transmit signal vectors $\hat{\mathbf{x}}$ that are located within a hyper-sphere of radius r centered at the received signal vector $\mathbf{y}$, that is, all the vectors $\hat{\mathbf{x}}$ which satisfy the criterion:

$$\left\| \mathbf{y} - \mathbf{H}\hat{\mathbf{x}} \right\|^2 \leq r^2. \tag{2.41}$$

Thus, SD avoids the complexity of an exhaustive search by only considering the vectors $\hat{\mathbf{x}}$ that satisfy (2.41). Note that the radius r of the hyper-sphere must be chosen carefully. If r is too small, there will be no vectors $\hat{\mathbf{x}}$ inside, and if r is too big, there will be little benefit over an exhaustive search (BARRY *et al.*, 2003). By assuming that at least one possible transmit vector $\hat{\mathbf{x}}$ satifies (2.41), it then follows that this nearest vector $\hat{\mathbf{x}}$ to $\mathbf{y}$ in the smaller SD search set must also be the nearest one in the entire ML search set M^{N_T}. Although the SD has less complexity than the ML detector, its complexity is still exponential but with a smaller exponent for low and medium SNR, that is $O\left(M^{\eta N_T}\right)$, where $0 < \eta < 1$ (HASSIBI; VIKALO, 2005), making it also impractical for large N_T.

On the other hand, knowing ML or SD solution is desired since it serves as a benchmark to assess how various detector algorithms work relative to the optimum performance. When N_T is large, that is from tens of transmit antennas, computing ML or SD solution becomes infeasible due to the exponential complexity. However, the optimum performance can also be achieved by the hypothetical Genie-Aided (GA) detector (BARRY *et al.*, 2003) described below.

2.5.3.1.3 Hypothetical Genie-Aided Detector

From the spatial multiplexing MIMO system model given by (2.13) considers the product:

$$\begin{aligned} \mathbf{r} &= \mathbf{H}^H\mathbf{y} \\ &= \mathbf{H}^H(\mathbf{H}\mathbf{x} + \mathbf{n}) \\ &= \mathbf{G}\mathbf{x} + \mathbf{w}, \end{aligned} \tag{2.42}$$

where $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ is a Gram matrix that can be interpreted as a matrix of cross-correlations, once its $g_{i,j}$ entry is equal to the correlation between i-th and j-th column of the channel matrix $\mathbf{H}$, that is $g_{i,j} = \mathbf{h}_i^H\mathbf{h}_j$. Let $\mathbf{G}^d = \text{diag}[\|\mathbf{h}_1\|^2, \cdots, \|\mathbf{h}_{N_T}\|^2]$ denotes the diagonal matrix of $\mathbf{G}$, then, (2.42) can be rewritten as:

$$\mathbf{r} = \mathbf{G}^d\mathbf{x} + (\mathbf{G} - \mathbf{G}^d)\mathbf{x} + \mathbf{w}, \tag{2.43}$$

where $\mathbf{G}^d\mathbf{x}$ and $(\mathbf{G} - \mathbf{G}^d)\mathbf{x}$ represents the desired and the spatial interference terms, respectively. The basic idea behind the genie-aided (GA) detector is that somehow the spatial interference term $(\mathbf{G} - \mathbf{G}^d)\mathbf{x}$ is known. Thus, since $(\mathbf{G} - \mathbf{G}^d)\mathbf{x}$ is deterministic and known at the receiver side, the best one can do is to subtract it from (2.43), obtaining:

$$\begin{aligned} \mathbf{z} &= \mathbf{r} - (\mathbf{G} - \mathbf{G}^d)\mathbf{x} \\ &= \mathbf{G}^d\mathbf{x} + \mathbf{w}. \end{aligned} \tag{2.44}$$

Once the spatial interference has been eliminated, each entry of the vector $\mathbf{z}$ given by (2.44) can be quantized with respect to the employed symbol constellation with order M, that is:

$$\hat{\mathbf{x}}_{\mathrm{GA}} = Q\left\{\mathbf{z}\right\}, \tag{2.45}$$

where $Q\left\{\cdot\right\}$ and $\hat{\mathbf{x}}_{\mathrm{GA}}$ represent the quantization process and the GA detected vector of the transmitted signal vector $\mathbf{x}$, respectively.

The GA presents a BER bound that is lower of any practical detector and is commonly named as the "perfect-cancellation" bound or "single-user" bound (KRAMER, 2001; BARRY *et al.*, 2003; XIANG *et al.*, 2016).

2.5.3.2 Linear Detectors

From a complexity viewpoint, a low complexity option for spatial multiplexing MIMO are linear detectors (LDs) that are sub-optimum detectors. The LDs attempt to eliminate the spatial interference between the N_T transmitted signals through linear operations performed at the receiver side (MINANGO, 2014).

The idea behind LDs is that the detected transmitted signal vector of $\mathbf{x}$, denoted as $\hat{\mathbf{x}}_{\mathrm{LD}}$, is achieved by multiplying the received signal vector $\mathbf{y}$ given by (2.13) by the equalization matrix $\mathbf{A}$, followed by a component-wise quantization over the employed symbol constellation in the following way:

$$\hat{\mathbf{x}}_{\mathrm{LD}} = Q\left\{\mathbf{A}\mathbf{y}\right\}. \tag{2.46}$$

Althogh LDs possess the advantage of low complexity, their performance is bad if the channel matrix $\mathbf{H}$ is highly non-orthogonal or ill-conditioned, that is LDs do not achieve the full diversity of the optimum detector. However, when the channel matrix $\mathbf{H}$ is orthogonal, these detectors are able to achieve the optimum performance, as is verified in the following chapters. The most common LD algorithms are next.

2.5.3.2.1 Matched-Filter Detector

The matched-filter (MF) detector is the simplest of the LDs, which essentially ignores the spatial interference. Thus, MF detector treats spatial interference produced by the N_T transmit antennas as pure noise by making $\mathbf{A} = \mathbf{H}^H$. Therefore, from (2.13), the MF detection of the transmitted signal vector $\mathbf{x}$ is given by:

$$\hat{\mathbf{x}}_{\mathrm{MF}} = Q\left\{\mathbf{H}^H\mathbf{y}\right\}. \tag{2.47}$$

Note that computing (2.47) requires a complexity of order $N_T N_R$, that is $O(N_T N_R)$, which is quite attractive. However, MF detector only works properly when

$N_R \gg N_T$, that is in massive MIMO channels. On the other hand, MF performance degrades severely when N_T increases, due to high levels of spatial interference.

2.5.3.2.2 Zero-Forcing Detector

The principal aim of zero-forcing (ZF) detector is to eliminate spatial interference completely, regardless of noise enhancement. Thus, the ZF equalization matrix that eliminates the spatial interference between the transmit antennas N_T is given by (MINANGO, 2014):

$$\mathbf{A} = \mathbf{H}^{\dagger} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H, \tag{2.48}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse of the channel matrix $\mathbf{H}$ (GOLUB; LOAN, 1996). In order for the pseudoinverse to exist, the number of transmit antennas N_T must be less than or equal to the number of receive antennas N_R, that is $N_T \leq N_R$. On the other hand, if N_T is larger than N_R, $\mathbf{H}^H\mathbf{H}$ is singular and its inverse does not exit (GOLUB; LOAN, 1996). Thus, from (2.13), the ZF detection of the transmitted signal vector $\mathbf{x}$ is given by:

$$\hat{\mathbf{x}}_{\text{ZF}} = Q\left\{\mathbf{H}^{\dagger}\mathbf{y}\right\}. \tag{2.49}$$

The complexity of ZF detector is in computing $\mathbf{H}^{\dagger}$, which has cubic complexity on the number of transmit antennas, that is $O(cN_T^3)$, where c is a multiplicative factor that depends on the procedure used to obtain $\mathbf{H}^{\dagger}$.

On the other hand, in terms of performance, ZF detector achieves, on average, a diversity order of $N_R - N_T + 1$ for highly non-orthogonal channel matrices (MINANGO, 2014), which is still $N_T - 1$ less than the diversity order of the optimum detector.

2.5.3.2.3 Minimum Mean-Square Error Detector

An alternative to the ZF detector is the minimum mean-square error (MMSE) detector, which attempts to strike a balance between noise enhancement and spatial interference suppression in the receiver. Therefore, MMSE detector aims to find the equalization matrix that minimizes the mean-square error (MSE) between the transmitted signal vector $\mathbf{x}$ and the transformed received signal vector $\tilde{\mathbf{A}}\mathbf{y}$, which is given by the solution of the following minimization problem (ANDREWS *et al.*, 2007):

$$\min_{\tilde{\mathbf{A}}} \; \mathbb{E}\{\|\mathbf{x} - \tilde{\mathbf{A}}\mathbf{y}\|^2\}. \tag{2.50}$$

The solution of (2.50) is given by:

$$\mathbf{A} = (\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}_{N_T})^{-1}\mathbf{H}^H, \tag{2.51}$$

where $\sigma^2 = N_T/\gamma$ and γ is the overall SNR given by (2.16). For high SNR, $\sigma^2 \to 0$, then, (2.50) becomes the ZF equalization matrix given by (2.48). Finally, from (2.13) and (2.50), the MMSE detection of the transmitted signal vector $\mathbf{x}$ is given by:

$$\hat{\mathbf{x}}_{\mathrm{MMSE}} = Q\left\{(\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}_{N_T})^{-1}\mathbf{H}^H\mathbf{y}\right\}. \tag{2.52}$$

Like ZF detector, because of the matrix inversion involved in (2.50), the complexity of the MMSE detector is cubic with the number of transmit antennas. Furthermore, for highly non-orthogonal channel matrix MMSE does not improve the diversity order, which is still $N_R - N_T + 1$ (TSE; VISWANATH, 2005; MINANGO, 2014)

2.5.3.3 Non-linear Lattice Reduction Detector

The non-linear lattice reduction (LR) detector has the potential to achieve the full diversity of the optimum detector by reducing the condition number of the channel matrix $\mathbf{H}$, thus allowing to use LDs and consequently a reduced computational complexity (BAI; CHOI, 2014; WUBBEN *et al.*, 2011).

The basic idea behind LR detector is to turn the channel matrix $\mathbf{H}$ in near-orthogonal as possible. For this purpose, the system model given by (2.13) is transformed into an equivalent system model using the LR technique. The equivalent system model is:

$$\mathbf{y} = \tilde{\mathbf{H}}\mathbf{s} + \mathbf{n}, \tag{2.53}$$

where $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{T}$ and $\mathbf{s} = \mathbf{T}^{-1}\mathbf{x}$ are the near-orthogonal channel matrix and the transformed transmitted signal vector, respectively, where $\mathbf{T}$ is an unimodular matrix which entries belong to the set of complex integers. This can be obtained by appling LR techniques as: Minkowski (AFFLERBACH; GROTHE, 1985), Hermite–Korkin-Zolotarev (HKZ) (ZHANG *et al.*, 2012), Lenstra-Lenstra-Lovasz (LLL) (YANG; KIM, 2013; WEN *et al.*, 2014) and Seysen (SEETHALER *et al.*, 2007), which have polynomial average complexity. In fact, the most used technique in the literature is the LLL (WUBBEN *et al.*, 2011), mainly due to its polynomial complexity $O(N_T^4)$.

Once the equivalent system given by (2.53) has been obtained, ZF or MMSE detectors can be employed in order to perform the quantization on $\mathbf{s}$ instead of $\mathbf{x}$. For LR followed by ZF (LR-ZF), the output signal vector is written as:

$$\hat{\mathbf{s}}_{\mathrm{LR-ZF}} = \tilde{\mathbf{H}}^{\dagger}\mathbf{y} = \mathbf{s} + \tilde{\mathbf{H}}^{\dagger}\mathbf{n}, \tag{2.54}$$

where the multiplication by $\tilde{\mathbf{H}}^{\dagger}$ usually causes less noise amplification due to near-orthogonal columns of the equivalent channel matrix $\tilde{\mathbf{H}}$. Finally, from (2.54) the detection of the transmitted signal vector $\mathbf{x}$ is obtained by applying the following transformation:

$$\hat{\mathbf{x}}_{\mathrm{LR-ZF}} = \mathbf{T}\hat{\mathbf{s}}_{\mathrm{LR-ZF}}. \tag{2.55}$$

Note that a similar procedure can be used for LR followed by MMSE (LR-MMSE). It can be shown that LR-ZF detector has the potential to achieve full diversity at the expense of a SNR gap in comparison to the optimum performance. This is due to the LR technique can reduce the condition number of the channel matrix but never to the unity. That is, the LR technique does not obtain a true orthogonal channel matrix (WUBBEN *et al.*, 2011; YANG; KIM, 2013). On the other hand, the LR-ZF detector has a computational complexity of $O(N_T^4 \log_{10} N_T)$ (YANG; KIM, 2013), which can be infeasible for large spatial multiplexing MIMO systems.

2.5.4 Detection of Large-Scale Spatial Multiplexing MIMO Systems

From the antennas deployment point of view, it can be distinguished two configurations of large-scale spatial multiplexing systems. Fig. 2.10 shows the first configuration which essentially deals with the large-scale MIMO detection problem encountered on an underloaded uplink. This configuration is known in the literature as spatial multiplexing massive MIMO system (RUSEK *et al.*, 2013). In these systems only the base station (BS) receiver is equipped with a large number of antennas, while the total number of transmit antennas is significantly smaller (one antenna per mobile station). Hence, the antenna configuration of spatial multiplexing massive MIMO systems can be characterized as:

$$N_R \gg N_T. \tag{2.56}$$

In spatial multiplexing massive MIMO systems, since the number of receive antennas is significantly higher than the total number of transmit antennas, a high receive diversity order is obtained. On the other hand, in the extreme case, that is when $N_R \to \infty$, the receive diversity gain obtained is so high, that the effects of both multi-antenna spatial interference and noise disappears. In addition, the channel vectors associated with distinct transmit antennas become asymptotically orthogonal, as shown in (2.37). Furthermore, another advantage of spatial multiplexing massive MIMO systems is that very tall, with large N_R, channel matrices $\mathbf{H}$ are very well conditioned, as it was verified in Fig. 2.5. Thus, in spatial multiplexing massive MIMO systems, even the simplest MF detector is capable of achieving a near-optimum performance, as is shown in the next chapter.

In the second configuration of large-scale MIMO systems, a large number of antennas are deployed at receiver, and also at the transmitter (YANG; HANZO, 2015) as is shown in Fig. 2.11. This configuration is characterized by:

$$N_R = N_T, \tag{2.57}$$

Thus, this second configuration is known in the literature as symmetric spatial multiplexing MIMO systems.

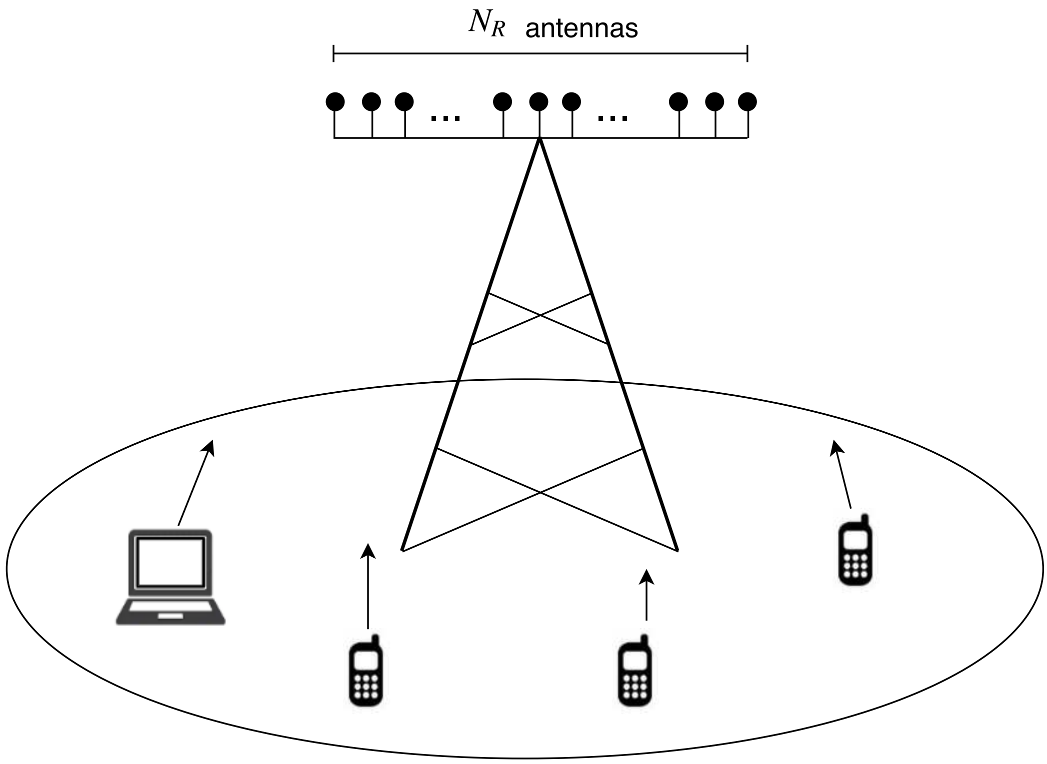


Figure 2.10 – Spatial multiplexing massive MIMO systems with $N_R \gg N_T$, multi-user application.

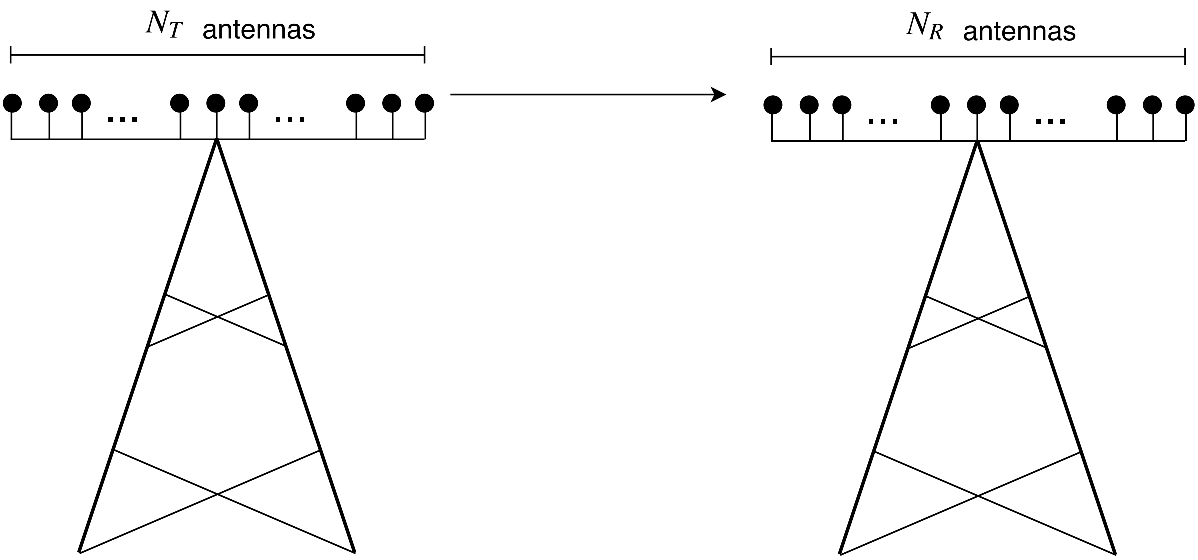


Figure 2.11 – Symmetric spatial multiplexing massive MIMO systems with $N_R = N_T$, poit-to-point application.

For symmetric spatial multiplexing MIMO systems, it has been shown in Fig. 2.8 that as the symmetric channel matrix $\mathbf{H}$ becomes large, in terms of both N_R and N_T, its condition number becomes large too. Thus, $\mathbf{H}$ becomes an ill-conditioned matrix. On the other hand, as the size of $\mathbf{H}$ increases, the diagonal entries of $\mathbf{H}^H\mathbf{H}$ become increasingly larger in magnitude than the off-diagonal entries (see Fig. 2.9). This behavior is associated with the channel-hardening phenomenon, which is considered in Chapter 7.

2.6 Chapter Conclusions

In this chapter, the basic concepts and spatial multiplexing MIMO system model used throughout the thesis were presented. Briefly, the concepts of channel model, diversity and digital modulation were covered.

Furthermore, spatial multiplexing, which is a multiple transmit/receive antenna technique (MIMO), transmits independent symbols by the transmit antennas at the same bandwidth in order to increase the overall bit rate and consequently the spectral efficiency of wireless communication systems. In addition, the background theory of MIMO systems were presented, where two scenarios were considered: massive and symmetric MIMO systems. Finally, optimum, linear and non-linear detectors for spatial multiplexing MIMO detection were reviewed.

3 Performance Difference Between ZF and ML Detector in Massive MIMO Systems

3.1 Introduction

In chapter 2, the benefits of increasing the number of antennas from a spectral efficiency point of view were discussed for a point-to-point spatial multiplexing MIMO system as shown in Fig. 2.3. However, in recent years the focus has deviated to multi-user spatial multiplexing MIMO systems where N_T single-antenna users transmit information simultaneously to a BS with N_R antennas at the same bandwidth achieving multiplexing gain and greater chances of uncorrelated fading. This leads to channel matrices with higher rank, which is much more desirable. Another important advantage of multi-user spatial multiplexing MIMO systems is tha N_T users require a cheap single-antenna transceivers, while the most expensive equipment is located at the BS. Based on these advantages, multi-user spatial multiplexing MIMO systems have been progressively used as a part of wireless communication standards, such as LTE-A (LIM *et al.*, 2013; NAGEL *et al.*, 2016), 802.11 (Wi-Fi) (BEJARANO *et al.*, 2013) and 802.16 (Wi-MAX) (LI *et al.*, 2010).

In 5G cellular networks, higher data rates and user densities are demanded (AGIWAL *et al.*, 2016). But, BS typically employs a small number of antennas in the deployed multi-user spatial multiplexing MIMO systems, i.e., at most 8 antennas for LTE-A (LIM *et al.*, 2013), which limits the data rates and therefore the spectral efficiency. Recently, several works (RUSEK *et al.*, 2013; NGO *et al.*, 2013; NARASIMHAN *et al.*, 2014; MARZETTA *et al.*, 2016) have considered a multi-user spatial multiplexing MIMO systems with tens of users to simultaneously transmit information to a very large number of antennas at BS in order to achieve high spatial multiplexing gains and energy efficiency than conventional multi-user spatial multiplexing MIMO systems. The jargon used in the literature when tens of users transmit information to a large number of BS antennas is known as spatial multiplexing massive MIMO systems(MARZETTA *et al.*, 2016), which is considered as a promising technology for 5G cellular network (VANNITHAMBY; TALWAR, 2017).

On the other hand, an additional advantage of using large number of receive antennas at the BS is that linear detectors (LDs), such as ZF, are able to achieve a near-optimum performance due to property of asymptotically orthogonal channel matrix, that is due to the massive MIMO channel hardening phenomenon described earlier in the sub-section 2.5.2.3. However, how far away is LDs from the optimum ML performance in spatial multiplexing massive MIMO systems? In this chapter, firstly it is verified through Monte Carlo simulation that LDs perform near to the optimum detector. Next, closed-form BER expressions for ZF and ML detectors employing M-QAM modulation are derived. Then, these BER expressions are subsequently used to evaluate the performance difference between both detectors, which is a function of the number of antennas at the BS and the number of users, but it is not a function of the modulation order. Note that this chapter focuses on the BER as a performance measure in contrast to the ergodic capacity-approach commonly used in the literature (MARZETTA *et al.*, 2016). Finally, the derived BER expressions together with the performance difference are validated by means of numerical results. The remainder of this chapter is organized as follows. Section 3.2 reviews the near-optimality of LDs for spatial multiplexing massive MIMO systems. The BER of ML and ZF together with the performance difference expression between ZF and ML detector are obtained in Section 3.3. Numerical results are presented in Section 3.4. Finally, the conclusions are drawn in Section 3.5.

3.2 On the Near-Optimality of Linear Detectors

Consider the uplink of a spatial multiplexing massive MIMO system as shown in Fig. 3.1, which employs N_R antennas at the BS to receive the $N_T \times 1$ signal vector $\mathbf{x}$ transmitted by the N_T single-antenna users simultaneously and at the same bandwidth, where $N_R \gg N_T$, i.e., $N_R = 128$ and $N_T = 16$ was considered in (MINANGO; ALMEIDA, 2017c). Thus, the corresponding $N_R \times 1$ received signal vector $\mathbf{y}$ at the BS antennas can be expressed by (2.13). Note that the elements of $\mathbf{x}$ are symbols of a M-QAM constellation of order M.

The ML detector algorithm given by (2.40), achieves the optimum performance. But, its complexity increases exponentially with the number of users, that is $O(M^{N_T})$, making it impractical. Fortunately, since $N_R \gg N_T$, LDs are able to achieve the near-optimum performance due to the property of asymptotically orthogonal channel matrix, as shown in the following.

LDs multiplies the received signal vector $\mathbf{y}$ given by (2.13) by the equalization matrix $\mathbf{A}$, that is

$$\begin{aligned} \mathbf{r} &= \mathbf{A}\mathbf{y} \\ &= \mathbf{A}\mathbf{H}\mathbf{x} + \mathbf{A}\mathbf{n}, \end{aligned} \tag{3.1}$$

where $\mathbf{r} \in \mathbb{C}^{N_T \times 1}$ is the vector containing information about the N_T transmit symbols and $\mathbf{A}$ for the three conventional detectors MF, ZF and MMSE, described in sub-section

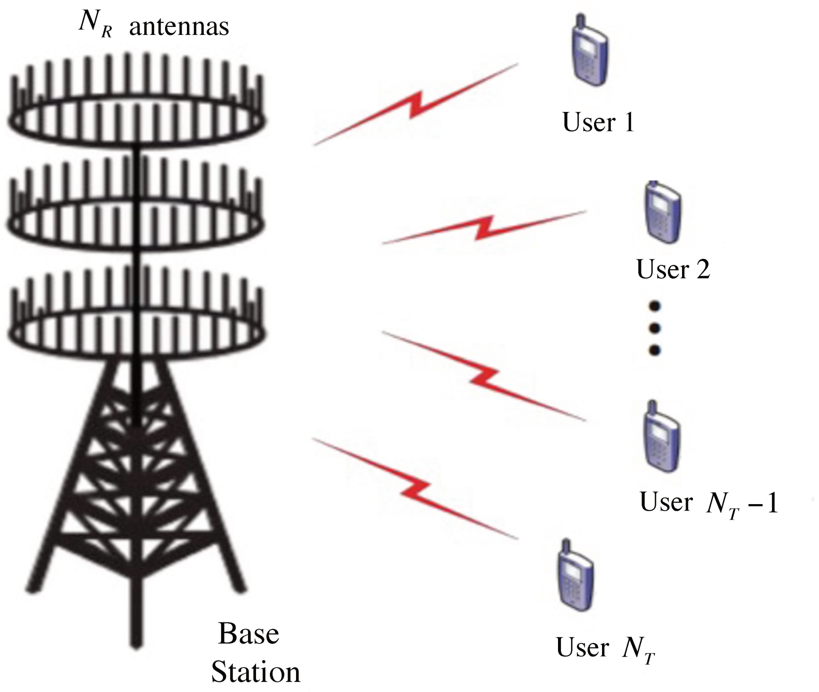


Figure 3.1 – Multi-user spatial multiplexing massive MIMO system representation.

2.5.3.2, is given by:

$$\mathbf{A} = \begin{cases} \mathbf{H}^H, & \text{for} \quad \text{MF}, \\ \mathbf{H}^\dagger = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H, & \text{for} \quad \text{ZF}, \\ \left(\mathbf{H}^H\mathbf{H} + \sigma^2\mathbf{I}_{N_T}\right)^{-1}\mathbf{H}^H, & \text{for} \quad \text{MMSE}. \end{cases} \tag{3.2}$$

Let r_k and x_k the k-th elements of vectors $\mathbf{r}$ and $\mathbf{x}$, respectively. Then:

$$\begin{aligned} r_k &= \mathbf{a}_k\mathbf{H}\mathbf{x} + \mathbf{a}_k\mathbf{n} \\ &= \underbrace{\mathbf{a}_k\mathbf{h}_k x_k}_{\text{desired signal}} + \underbrace{\sum_{i=1, i\neq k}^{N_T} \mathbf{a}_k\mathbf{h}_i x_i}_{\text{multi-user interference}} + \underbrace{\mathbf{a}_k\mathbf{n}}_{\text{enhanced noise}}, \end{aligned} \tag{3.3}$$

where $\mathbf{a}_k$ and $\mathbf{h}_k$ are the k-th row and column vectors of $\mathbf{A}$ and $\mathbf{H}$, respectively. Using the MF detector and considering the massive MIMO channel hardening effects, once it is assumed that $N_R \to \infty$. Then for $k \neq i, k = 1, \cdots, N_T$, it is possible to show that:

$$\frac{1}{N_R}\mathbf{h}_k^H\mathbf{h}_i \to 0, \qquad \text{multi-user interference vanishes;} \tag{3.4}$$

$$\frac{1}{N_R}\mathbf{h}_k^H\mathbf{n} \to 0, \qquad \text{noise vanishes.} \tag{3.5}$$

From (3.3), considering (3.4) and (3.5), the k-th useful signal satisfies:

$$r_k = \frac{1}{N_R}\mathbf{h}_k^H\mathbf{h}_k x_k = \frac{1}{N_R}\left\|\mathbf{h}_k\right\|^2 x_k \propto x_k, \tag{3.6}$$

which indeed represents the symbol transmitted by the k-th user. However, it is evident that the condition (3.6) is rather unrealistic, once N_R is large but finite. Nevertheless (3.6) provides a useful insight of the advantages of increasing the number of antennas at BS.

Since N_R is large but finite, ZF detector is widely considered in spatial multiplexing massive MIMO systems, as it can eliminate the multi-user interference, regardless of noise enhancement (MINANGO, 2014). On the other hand, MMSE detector achieves an optimal balance between multi-user interference suppression and noise enhancement. Thus, from the three usual LDs, MMSE is expected to be the best. Further, from (3.2) note that the filtering matrix $\mathbf{A}$ for both ZF and MMSE detector involves the matrix inverse computation, which yields a non-negligible complexity when a very large number of users are served. This issue will be addressed in the following two chapters.

In the following, numerical results obtained via Monte Carlo simulations are presented in order to demonstrate the near-optimum performance of LDs in spatial multiplexing massive MIMO systems. In the simulations, an uncorrelated flat Rayleigh fading channel matrix is considered. Moreover, during each channel use, all N_T users transmit simultaneously 64-QAM symbols to the BS antennas. The 64-QAM modulation is employed due to its high spectral efficiency of 6 b/s/Hz.

Fig. 3.2 shows the BER simulation results as a function of SNR for spatial multiplexing massive MIMO systems with $N_T = 16$ single-antenna users and a BS with $N_R = 64, 128, 256$ and 512 antennas. The performances of MF, ZF and MMSE detector are compared to the optimum ML detector. Note that the optimum ML performance is achieved by the hypothetical GA detector described earlier in sub-section 2.5.3.1.

From the sub-figures of Fig. 3.2, the performance of LDs are shown versus γ_b for different number of BS antennas. Notice that, the performance penalty of MF detector in comparison to ML detector is reduced significantly by increasing N_R. However, MF performance is still far away from the optimum performance in all considered cases. On the other hand, both ZF and MMSE detector achieves a close performance to ML detector, once $N_R \gg N_T$. Thus, the performance penalty of both ZF and MMSE detector in comparison to ML detector tends to vanish as N_R increases, which verifies the near-optimum performance of ZF and MMSE detectors in spatial multiplexing massive MIMO systems. In summary, MF, ZF and MMSE detectors approach the ML performance as N_R increases, however, both ZF and MMSE performance penalty to ML performance diminishes faster in comparison to the MF performance penalty. Therefore, ZF and MMSE detector for spatial multiplexing massive MIMO systems are considered.

Furthermore, despite that MMSE detector is expected to achieve a better performance than ZF detector, almost from sub-figures of Fig. 3.2, notice that both ZF and

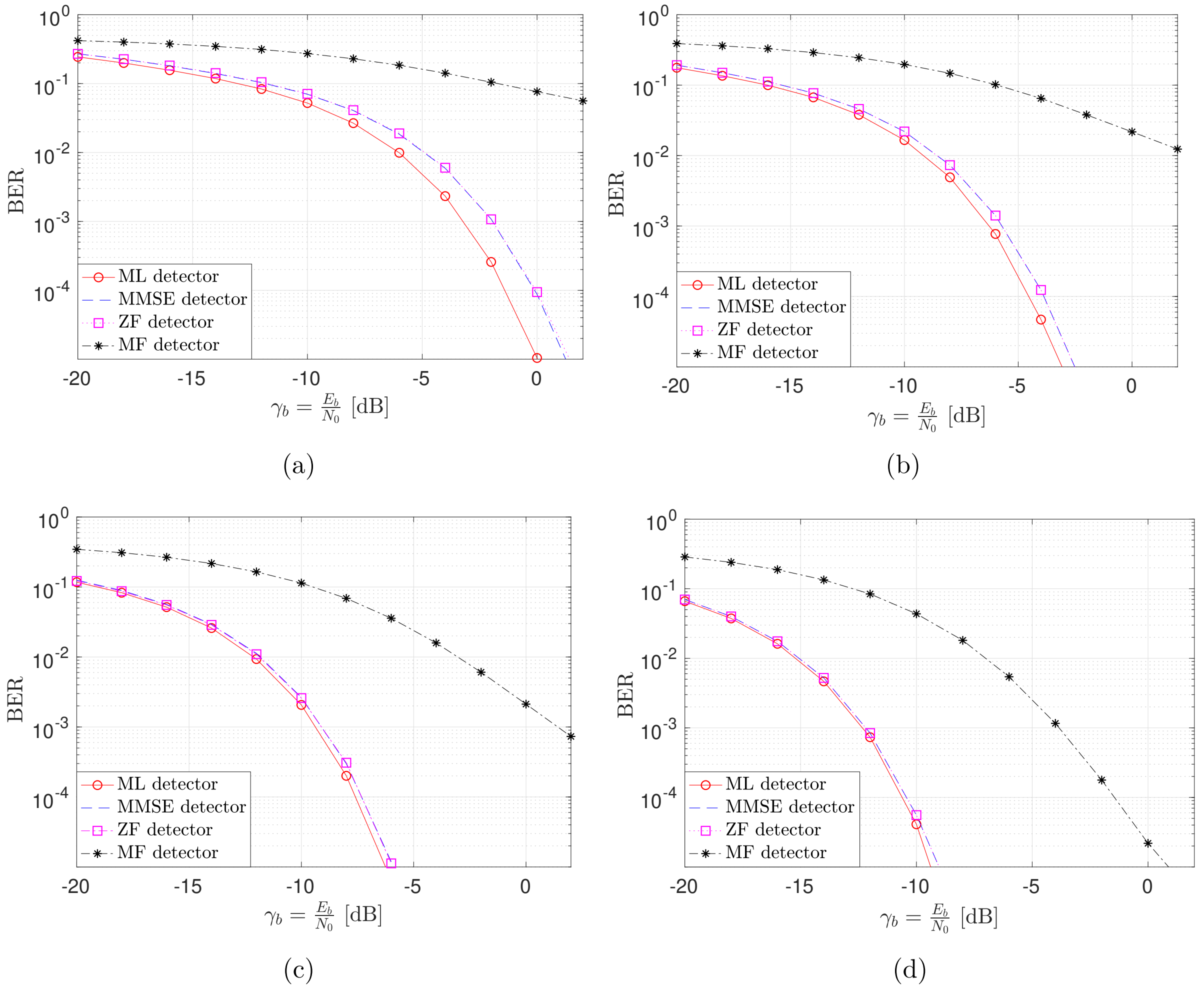


Figure 3.2 – BER as a funtion of γ_b for spatial multiplexing massive MIMO systems with $N_T = 16$ users and (a) $N_R = 64$, (b) $N_R = 128$, (c) $N_R = 256$ and (d) $N_R = 512$ BS antennas.

MMSE detector achieves the same performance for spatial multiplexing massive MIMO systems. Also, both ZF and MMSE detector yields the same complexity $O(N_T^3)$ and diversity order of $N_R - N_T + 1 \approx N_R$, for $N_R \gg N_T$. For these reasons, in this chapter only ZF detector is considered in order to establish the performance difference between near-optimum and optimum detector in spatial multiplexing massive MIMO systems.

3.3 Performance Difference

In this section, closed-form BER expressions based on pairwise error probability (PEP) for ML and ZF detector in spatial multiplexing massive MIMO systems are derived. Then, these BER expressions are used to determine the performance difference between both detectors.

3.3.1 ML Performance Analysis

The vector $\mathbf{x}$ is transmitted and the ML detector makes an erroneous detection $\hat{\mathbf{x}}_{\mathrm{ML}}$ of $\mathbf{x}$ based on the received signal vector $\mathbf{y}$ given by (2.13). Then, the PEP of the ML detector, $\mathrm{P_r}(\mathbf{x} \rightarrow \hat{\mathbf{x}}_{\mathrm{ML}})$, for a given channel matrix $\mathbf{H}$ is defined as the probability that $\mathbf{y}$ is closer to $\mathbf{H}\hat{\mathbf{x}}_{\mathrm{ML}}$ than $\mathbf{H}\hat{\mathbf{x}}$ once $\hat{\mathbf{x}}$ was transmitted. Mathematically, this is established as (BARRY *et al.*, 2003):

$$\begin{aligned}
\mathrm{P_r}(\mathbf{x} \rightarrow \hat{\mathbf{x}}_{\mathrm{ML}}) &= \mathrm{P_r}(\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}_{\mathrm{ML}}\|^2 \leq \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}_{\mathrm{ML}}) + \mathbf{n}\|^2 \leq \|\mathbf{n}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{H}\mathbf{d}_{\mathrm{ML}}\|^2 + 2\Re\{\mathbf{n}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}}\} + \|\mathbf{n}\|^2 \leq \|\mathbf{n}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{H}\mathbf{d}_{\mathrm{ML}}\|^2 \leq -2\Re\{\mathbf{n}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}}\}), \qquad (3.7)
\end{aligned}$$

where $\mathbf{d}_{\mathrm{ML}} = \mathbf{x} - \hat{\mathbf{x}}_{\mathrm{ML}}$ denotes the ML error vector and $\Re\{\cdot\}$ the real-part operation.

Since $\mathbf{n}$ is a zero-mean circular symmetric complex Gaussian (CSCG) random vector with $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2\mathbf{I}_{N_R}$, then the product $\mathbf{n}^H\mathbf{H}\mathbf{d}$ is also a CSCG random vector, that has the following properties:

$$\mathbb{E}\{\mathbf{n}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}}\} = \mathbf{0}, \qquad (3.8a)$$

$$\begin{aligned}
\mathbb{E}\{\mathbf{n}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}}\mathbf{d}_{\mathrm{ML}}^H\mathbf{H}^H\mathbf{n}\} &= \mathbb{E}\{\mathbf{d}_{\mathrm{ML}}^H\mathbf{H}^H\mathbf{n}\mathbf{n}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}}\} \\
&= \mathbf{d}_{\mathrm{ML}}^H\mathbf{H}^H\mathbb{E}\{\mathbf{n}\mathbf{n}^H\}\mathbf{H}\mathbf{d}_{\mathrm{ML}} \\
&= \sigma_n^2\mathbf{d}_{\mathrm{ML}}^H\mathbf{H}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}} \\
&= \sigma_n^2\|\mathbf{H}\mathbf{d}_{\mathrm{ML}}\|^2. \qquad (3.8b)
\end{aligned}$$

Let $\Psi = \Re\{\mathbf{n}^H\mathbf{H}\mathbf{d}_{\mathrm{ML}}\}$. According to the properties given by (3.8), it can be show that $\Psi \sim \mathcal{N}(0, \frac{\sigma_n^2}{2}\|\mathbf{H}\mathbf{d}_{\mathrm{ML}}\|^2)$. Then, from (3.7) the PEP of the ML detector can be obtained as (BARRY *et al.*, 2003):

$$\begin{aligned}
\mathrm{P_r}(\mathbf{x} \rightarrow \hat{\mathbf{x}}_{\mathrm{ML}}) &= \mathrm{P_r}(\|\mathbf{H}\mathbf{d}_{\mathrm{ML}}\|^2 \leq -2\Psi) \\
&= Q\left(\sqrt{\frac{\|\mathbf{H}\mathbf{d}_{\mathrm{ML}}\|^2}{2\sigma_n^2}}\right), \qquad (3.9)
\end{aligned}$$

where

$$Q\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy \qquad (3.10)$$

is the Q-function (PAPOULIS; PILLAI, 2002).

Considering that the ML detector makes a single error detection for the i-th user, which indeed represents the most probable error detection (BIGLIERI *et al.*, 2007), the PEP given by (3.9) can be written as:

$$\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\mathrm{ML}}) = Q\left(\sqrt{\frac{\|\mathbf{h}_i\|^2 \mathrm{d}_{\mathrm{min}}^2}{2\sigma_n^2}}\right), \tag{3.11}$$

where $\mathbf{h}_i$ is the i-th column vector of $\mathbf{H}$ and $\mathrm{d}_{\mathrm{min}}$ denotes the minimum distance of the constellation belonging to the i-th user.

Since the entries of $\mathbf{H}$ are i.i.d. random variables with distribution $\mathcal{CN}(0, 1)$ and assuming that $N_R \to \infty$, it can be shown that (NGO *et al.*, 2013):

$$\lim_{N_R \to \infty} \frac{1}{N_R} \|\mathbf{h}_i\|^2 = 1. \tag{3.12}$$

Finally, replacing (3.12) into (3.11), the PEP of the ML detector for spatial multiplexing massive MIMO systems is given by:

$$\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\mathrm{ML}}) = Q\left(\sqrt{\frac{N_R \mathrm{d}_{\mathrm{min}}^2}{2\sigma_n^2}}\right). \tag{3.13}$$

For M-QAM constellation, $\mathrm{d}_{\mathrm{min}}$ and σ_n^2 are given, respectively, by (ALMEIDA, 2014):

$$\mathrm{d}_{\mathrm{min}}^2 = \frac{6E_b}{(M-1)\,T_b}, \tag{3.14}$$

and

$$\sigma_n^2 = \frac{E_b}{\gamma_b T_b \log_2 M}, \tag{3.15}$$

where E_b and $1/T_b$ represent the energy per bit and the bit rate, respectively, $\gamma_b = E_b/N_0$ is the SNR and N_0 is the noise unilateral power spectral density. Substituting (3.14) and (3.15) into (3.13), the PEP of the optimum ML detector for spatial multiplexing massive MIMO systems employing M-QAM is given by:

$$\mathrm{P_r}\left(\mathbf{x} \to \hat{\mathbf{x}}_{\mathrm{ML}}\right) = Q\left(\sqrt{\frac{3N_R \gamma_{b,\mathrm{ML}} \log_2 M}{M-1}}\right) \tag{3.16}$$

3.3.2 ZF Performance Analysis

Employing ZF detector to the received signal vector $\mathbf{y}$ given by (2.13), the ZF equalized received vector is given by:

$$\mathbf{H}^\dagger \mathbf{y} = \mathbf{x} + \mathbf{H}^\dagger \mathbf{n}, \tag{3.17}$$

which shows that the multi-user interference is completely eliminated from $\mathbf{H}^\dagger \mathbf{y}$. The metric used with ZF detector is then $\|\mathbf{H}^\dagger \mathbf{y} - \mathbf{x}\|^2$.

Assuming again that the vector $\mathbf{x}$ is transmitted, while $\hat{\mathbf{x}}_{\text{ZF}}$ is erroneously detected, from (2.13) and (3.17), the PEP for the ZF detector is given by:

$$\begin{aligned}
\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\text{ZF}}) &= \mathrm{P_r}(\|\mathbf{H}^{\dagger}\mathbf{y} - \hat{\mathbf{x}}_{\text{ZF}}\|^2 \leq \|\mathbf{H}^{\dagger}\mathbf{y} - \mathbf{x}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{x} + \mathbf{H}^{\dagger}\mathbf{n} - \hat{\mathbf{x}}_{\text{ZF}}\|^2 \leq \|\mathbf{H}^{\dagger}\mathbf{n}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{d}_{\text{ZF}} + \mathbf{H}^{\dagger}\mathbf{n}\|^2 \leq \|\mathbf{H}^{\dagger}\mathbf{n}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{d}_{\text{ZF}}\|^2 + 2\Re\{\mathbf{n}^H(\mathbf{H}^{\dagger})^H\mathbf{d}_{\text{ZF}}\} + \|\mathbf{H}^{\dagger}\mathbf{n}\|^2 \leq \|\mathbf{H}^{\dagger}\mathbf{n}\|^2) \\
&= \mathrm{P_r}(\|\mathbf{d}_{\text{ZF}}\|^2 \leq -2\Re\{\mathbf{n}^H(\mathbf{H}^{\dagger})^H\mathbf{d}_{\text{ZF}}\}),
\end{aligned} \tag{3.18}$$

where $\mathbf{d}_{\text{ZF}} = \mathbf{x} - \hat{\mathbf{x}}_{\text{ZF}}$ is the ZF error vector.

Since $\mathbf{n}$ is a zero-mean CSCG random vector, the following properties can be written:

$$\mathbb{E}\{\mathbf{n}^H(\mathbf{H}^{\dagger})^H\mathbf{d}_{\text{ZF}}\} = \mathbf{0}, \tag{3.19a}$$

$$\begin{aligned}
\mathbb{E}\{\mathbf{n}^H(\mathbf{H}^{\dagger})^H\mathbf{d}_{\text{ZF}}\mathbf{d}_{\text{ZF}}^H\mathbf{H}^{\dagger}\mathbf{n}\} &= \mathbb{E}\left\{\mathbf{d}_{\text{ZF}}^H\mathbf{H}^{\dagger}\mathbf{n}\mathbf{n}^H(\mathbf{H}^{\dagger})^H\mathbf{d}_{\text{ZF}}\right\} \\
&= \mathbf{d}_{\text{ZF}}^H\mathbf{H}^{\dagger}\mathbb{E}\left\{\mathbf{n}\mathbf{n}^H\right\}(\mathbf{H}^{\dagger})^H\mathbf{d}_{\text{ZF}} \\
&= \sigma_n^2\mathbf{d}_{\text{ZF}}^H(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{d}_{\text{ZF}} \\
&= \sigma_n^2\mathrm{Tr}[(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{d}_{\text{ZF}}\mathbf{d}_{\text{ZF}}^H],
\end{aligned} \tag{3.19b}$$

where the following relation is used (GOLUB; LOAN, 1996):

$$\mathbf{H}^{\dagger}(\mathbf{H}^{\dagger})^H = (\mathbf{H}^H\mathbf{H})^{-1}. \tag{3.20}$$

Thus, from (3.18) and based on the properties given by (3.19), the PEP for the ZF detector is given by:

$$\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\text{ZF}}) = Q\left(\sqrt{\frac{\mathrm{Tr}[(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{d}_{\text{ZF}}\mathbf{d}_{\text{ZF}}^H]}{2\sigma_n^2}}\right). \tag{3.21}$$

In spatial multiplexing massive MIMO systems, when both the number of BS antennas N_R and the number of users N_T are large, but with a fixed ratio $\alpha = N_R/N_T$, the matrices $(\mathbf{H}^H\mathbf{H})^{-1}$ and $\mathbf{d}_{\text{ZF}}\mathbf{d}_{\text{ZF}}^H$ are asymptotically independent (TULINO; VERDú, 2004; RUSEK *et al.*, 2013), and hence the following properties can be written:

$$\mathbb{E}\left\{\mathrm{Tr}\left[(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{d}\mathbf{d}^H\right]\right\} = N_T\mathbb{E}\left\{\mathrm{Tr}\left[(\mathbf{H}^H\mathbf{H})^{-1}\right]\right\}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{d}\mathbf{d}^H\right]\right\}, \tag{3.22a}$$

$$\mathbb{E}\left\{\mathrm{Tr}\left[(\mathbf{H}^H\mathbf{H})^{-1}\right]\right\} = \frac{N_R - N_T}{N_T}. \tag{3.22b}$$

Considering (3.22) into (3.21), then the PEP of the ZF detector for spatial multiplexing massive MIMO systems is given by:

$$\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\text{ZF}}) = Q\left(\sqrt{\frac{(N_R - N_T)\,\|\mathbf{d}_{\text{ZF}}\|^2}{2\sigma_n^2}}\right). \tag{3.23}$$

Finally, considering the most probable event in which ZF detector makes a single error detection for the i-th user, then (3.23) can be expressed as:

$$\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\mathrm{ZF}}) \simeq Q\left(\sqrt{\frac{(N_R - N_T)\mathrm{d}_{\min}^2}{2\sigma_n^2}}\right). \tag{3.24}$$

On the other hand, considering (3.14) and (3.15) into (3.24), the PEP of the ZF detector employing M-QAM is given by:

$$\mathrm{P_r}(\mathbf{x} \to \hat{\mathbf{x}}_{\mathrm{ZF}}) = Q\left(\sqrt{\frac{3\,(N_R - N_T)\,\gamma_{b,\mathrm{ZF}}\log_2 M}{M-1}}\right) \tag{3.25}$$

3.3.3 Performance Gap

In order to obtain the performance gap, in terms of SNR, between detectors ZF and ML for spatial multiplexing massive MIMO systems, the PEPs expressions given by (3.16) and (3.25) are equaled, then:

$$\frac{\gamma_{b,\mathrm{ZF}}}{\gamma_{b,\mathrm{ML}}} = \frac{N_R}{N_R - N_T} = \frac{\alpha}{\alpha - 1}, \qquad \text{where } \alpha = N_R/N_T. \tag{3.26}$$

Note that the performance gap or SNR penalty between detectors ZF and ML for spatial multiplexing massive MIMO systems given by (3.26) depends only on the relation between the number of BS antennas N_R and the number of users N_T.

3.4 Numerical Results

Fig. 3.3 shows the BER versus γ_b for ML and ZF detectors. In the simulation, different $\alpha = N_R/N_T$ ($\alpha = 4, 8, 16$) were considered for 64-QAM modulation. Note again that the simulation performance of the optimum ML detector is achieved by the GA detector.

From Fig. 3.3, note that as α increases, the performance gap between ML and ZF detector decreases. Thus, for a BER $= 10^{-6}$ with $\alpha = 4$, $\alpha = 8$ and $\alpha = 16$, the performance gap between both detectors is around 1.25 dB, 0.6 dB and 0.3 dB, respectively. Therefore, the penalty between ML and ZF detector is a shift of the BER curve, which varies according with the relation between the number of BS antennas N_R and the number of users N_T. On the other hand, note that the BER expressions of ML and ZF detector, given respectively by (3.16) and (3.25), fit perfectly the simulation results.

Table 3.1 shows the performance gap, in terms of γ_b, obtained theoretically in (3.26) and by simulation in Fig. 3.3. Notice that both theoretical and simulated performance gap fit perfectly for different α. Thus, through the performance gap obtained

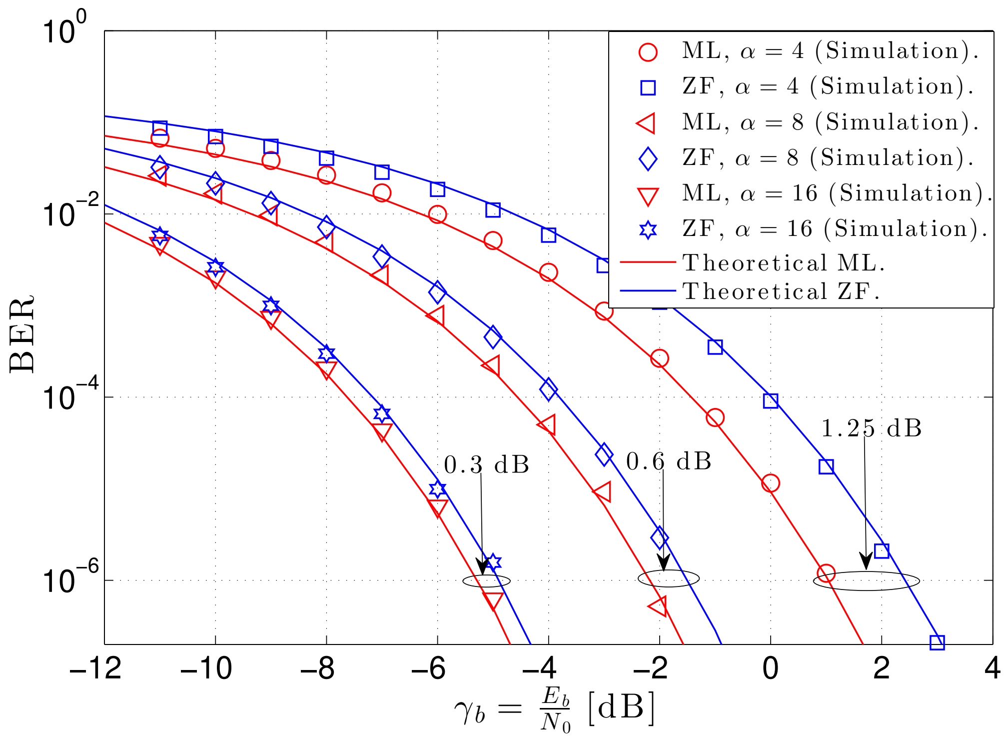


Figure 3.3 – Performance gap comparison between ML and ZF detector as a function of γ_b for spatial multiplexing massive MIMO systems with different $\alpha = N_R/N_T$ employing 64-QAM.

Table 3.1 – Performance gap comparison.

α	Performance gap simulation from Fig. 3.3	Performance gap theoretical from (3.26)
4	1.25 dB	1.24 dB
8	0.60 dB	0.58 dB
16	0.30 dB	0.28 dB

theoretically it is possible to know the performance gap for the design of spatial multiplexing massive MIMO systems employing ZF detector. Thus, for example, it is possible to know a priori that a spatial multiplexing massive MIMO system with $N_R = 128$ antennas and performance difference of 2 dB, then from (3.26) the maximum number of users that achieves this requirement is $N_T = 47$. Fig. 3.4 shows the theoretical and numerical results of an $N_R \times N_T = 128 \times 47$ spatial multiplexing massive MIMO system employing 64-QAM modulation. It is evident that for a BER $= 10^{-6}$, the performance difference between ZF and ML detector is around 2 dB, which validates the requirement previously established for the spatial multiplexing massive MIMO design.

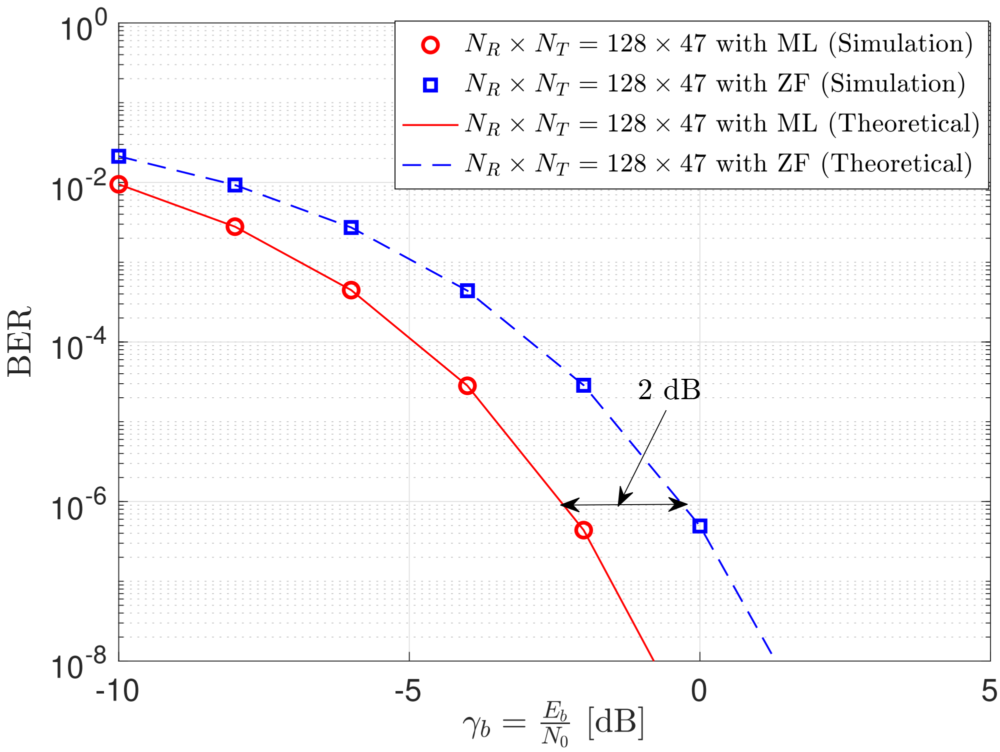


Figure 3.4 – BER against γ_b for $N_R \times N_T = 128 \times 47$ spatial multiplexing massive MIMO system employing 64-QAM.

3.5 Chapter Conclusions

In this chapter, the advantages of increasing the number of antennas at the BS have been identified. In particular, it has been shown that under the assumption that $N_R \gg N_T$ linear detectors, such as ZF and MMSE, were shown to perform close to optimum. Thus, spatial multiplexing massive MIMO systems offer the opportunity of increasing the spectral efficiency by employing low complexity linear detectors.

Moreover, closed-form BER expressions for the optimum ML and the near-optimum ZF detector for spatial multiplexing massive MIMO systems have been derived and used to obtain the performance difference between both detectors. This performance difference depends only on the ratio between the number of base station antennas N_R and the number of users N_T. Numerical results fit perfectly to the BER theoretical equation and to the performance difference equation. The expressions obtained in this chapter are fundamental and tight to know the difference in performance between ML and ZF detectors in the design of spatial multiplexing massive MIMO systems.

4 Low-Complexity ZF Detector for Massive MIMO Systems Based on Damped Jacobi Algorithm

4.1 Introduction

As it was pointed out in Chapter 3, due to the large number of BS antennas, ZF detector is able to achieve near-optimum performance. Unfortunately, this detector involves complicated exact channel matrix inversion, especially when the dimension of the spatial multiplexing massive MIMO system is large (NARASIMHAN *et al.*, 2014).

Recently, several approximated ZF detector algorithms were proposed in order to avoid channel matrix inversion (GAO *et al.*, 2014; YIN *et al.*, 2014a; DAI *et al.*, 2015). Richardson and conjugate gradient (CG) algorithms were proposed in (GAO *et al.*, 2014) and (YIN *et al.*, 2014a), respectively, to achieve near-optimum performance, but both requires large numbers of iterations to converge and CG involves also many divisions. In order to reduce the numbers of iterations, the approach based on Gauss-Seidel (GS) algorithm was presented, respectively, in (DAI *et al.*, 2015) and (MINANGO; ALMEIDA, 2017a), where only a small number of iterations is needed to achieve the near-optimum performance. However, the GS inner iterations make it not suitable for parallel implementation. On the other hand, for a large number of BS antennas, a ZF detector based on the popular Jacobi (CJ) algorithm was proposed in (KONG; PARK, 2016), but its convergence rate is low. Besides, the authors of (KONG; PARK, 2016) have adopted the real-valued spatial multiplexing massive MIMO system model to propose a new initial solution for the CJ algorithm in order to accelerate the convergence rate. However, it is well known that the the real-valued spatial multiplexing massive MIMO system model representation doubles its dimensionality and consequently the computational complexity.

Thus, in this chapter, a low-complexity ZF detector based on the Damped Jacobi (DJ) algorithm is proposed in order to reduce the computational complexity of ZF detector employing exact matrix inversion for spatial multiplexing massive MIMO systems. In order to ensure the fast convergence rate of the DJ algorithm, a simple way to determine the quasi-optimum damping parameter by exploiting the channel matrix property of asymptotic orthogonality is developed. This quasi-optimum damping parameter depends only on the dimension of the spatial multiplexing massive MIMO system, that is, the number of BS antennas and the number of users. Numerical results show that the proposed algorithm outperforms the CJ algorithm, and it can also achieve the near-optimum performance with about one order of magnitude less in complexity than the ZF detector employing exact matrix inversion.

The remainder of this chapter is organized as follows. Section 4.2 describes briefly the system model together with the CJ algorithm. Section 4.3 details the proposed algorithm. Numerical results are presented in Section 4.4. Finally, the conclusions are drawn in Section 4.5.

4.2 System Model

In this chapter, the uplink of a spatial multiplexing massive MIMO system equipped with N_T single-antenna users and N_R antennas at BS where $N_R \gg N_T$ is considered. The N_T users map their own information bits to the symbols of a modulation of order M. The $N_T \times 1$ vector $\mathbf{x}$, which contains the modulated symbols, is transmitted over the $N_R \times N_T$ massive MIMO channel matrix $\mathbf{H}$ and the $N_R \times 1$ signal vector $\mathbf{y}$ received at the BS antennas is given by (2.13).

Since $N_R \gg N_T$, the column channel vectors of $\mathbf{H}$ are asymptotically orthogonal, i.e., $\mathbf{H}$ is a near-orthogonal or well-conditioned channel matrix (TULINO; VERDú, 2004), and the ZF detector is able to achieve near-optimum performance.

4.2.1 ZF Detector

Appling a ZF detector, described in sub-section 2.5.3.2, to the received signal vector $\mathbf{y}$, the detected signal vector of $\mathbf{x}$ before the quantization process is given by:

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{ZF}} = \mathbf{H}^{\dagger}\mathbf{y} &= \left(\mathbf{H}^{H}\mathbf{H}\right)^{-1}\mathbf{H}^{H}\mathbf{y} \\ &= \mathbf{G}^{-1}\tilde{\mathbf{y}},\end{aligned} \tag{4.1}$$

where

$$\tilde{\mathbf{y}} = \mathbf{H}^{H}\mathbf{y} \tag{4.2}$$

represents the matched-filter output of $\mathbf{y}$, and

$$\mathbf{G} = \mathbf{H}^{H}\mathbf{H} \tag{4.3}$$

is the $N_T \times N_T$ Gram matrix, which is strictly diagonal dominant (SDD), that is satisfy (2.24), and it is Hermitian positive definite (HPD), that is satisfies (2.26) and (2.27), respectively.

From (4.1), note that ZF detector involves the exact matrix inversion $\mathbf{G}^{-1}$ with large size which has complexity $O\left(cN_T^3\right)$, where c is a multiplicative factor that depends on the algorithm used to obtain $\mathbf{G}^{-1}$. Thus, the computational complexity to obtain the exact matrix inversion of $\mathbf{G}$ is high, since N_T is usually large in spatial multiplexing massive MIMO systems.

4.2.2 Conventional Jacobi Algorithm

Based on the property that the Gram matrix $\mathbf{G}$ is SDD, in (KONG; PARK, 2016) the conventional Jacobi (CJ) algorithm was employed to detect the transmitted signal vector $\mathbf{x}$ in an iterative way without the need to compute $\mathbf{G}^{-1}$, as follows:

$$\begin{aligned}\hat{\mathbf{x}}_{t+1}^{\mathrm{CJ}} &= \mathbf{D}^{-1}\left(\tilde{\mathbf{y}} - \mathbf{R}\hat{\mathbf{x}}_t^{\mathrm{CJ}}\right) \\ &= \mathbf{D}^{-1}\tilde{\mathbf{y}} - \mathbf{D}^{-1}\mathbf{R}\hat{\mathbf{x}}_t^{\mathrm{CJ}}, \;\; t = 0, 1, 2, \ldots \end{aligned} \tag{4.4}$$

where $\mathbf{D}$ and $\mathbf{R}$ are matrices composed, respectively, by the diagonal and the off-diagonal elements of $\mathbf{G}$, that is $\mathbf{G} = \mathbf{D} + \mathbf{R}$, and $\hat{\mathbf{x}}_{t+1}^{\mathrm{CJ}}$ is the solution of the t-th iteration. Note that from (4.1), $\mathbf{G}\hat{\mathbf{x}}_{\mathrm{ZF}} = \tilde{\mathbf{y}}$. Thus, substituting that $\mathbf{G} = \mathbf{D} + \mathbf{R}$, it can get to (4.4) easily.

4.2.2.0.1 Convergence Rate of CJ Algorithm

From (4.4), the iteration matrix of the CJ algorithm is defined by:

$$\begin{aligned}\mathbf{B}^{\mathrm{CJ}} &= -\mathbf{D}^{-1}\mathbf{R} \\ &= -\mathbf{D}^{-1}\left(\mathbf{G} - \mathbf{D}\right) \\ &= \mathbf{I}_{N_T} - \mathbf{D}^{-1}\mathbf{G}. \end{aligned} \tag{4.5}$$

Thus, the necessary and sufficient condition for the convergence of (4.4) is that the spectral radius of the iteration matrix, given by (4.5), should satisfy (OLSHEVSKY; TYRTYSHNIKOV, 2010) that:

$$\rho\left(\mathbf{B}^{\mathrm{CJ}}\right) = \left|\lambda_{\max}\left(\mathbf{B}^{\mathrm{CJ}}\right)\right| < 1. \tag{4.6}$$

where $\lambda_{\max}\left(\mathbf{B}^{\mathrm{CJ}}\right)$ is the greatest eigenvalue of $\mathbf{B}^{\mathrm{CJ}}$.

On the other hand, since $N_R \gg N_T$, the smallest and largest eigenvalue of the Gram matrix $\mathbf{G}$ can be approximated, respectively, by (see Theorem 1):

$$\lambda_{\min}\left(\mathbf{G}\right) \approx N_R\left(1 - \sqrt{\frac{N_T}{N_R}}\right)^2, \tag{4.7a}$$

$$\lambda_{\max}\left(\mathbf{G}\right) \approx N_R\left(1 + \sqrt{\frac{N_T}{N_R}}\right)^2 \tag{4.7b}$$

and also, due to the massive MIMO channel hardening phenomenon, it is possible to approximate the diagonal matrix $\mathbf{D}$ by:

$$\mathbf{D} = N_R\mathbf{I}_{N_T}. \tag{4.8}$$

Thus, from (4.5), considering (4.7b) and (4.8), the largest eigenvalues of the iteration matrix $\mathbf{B}^{\mathrm{CJ}}$ is given by:

$$\begin{aligned}\lambda_{\max}\left(\mathbf{B}^{\mathrm{CJ}}\right) &= \lambda_{\max}\left(\mathbf{I}_{N_T}\right) - \lambda_{\max}\left(\mathbf{D}^{-1}\mathbf{G}\right)\\ &= 1 - \frac{1}{N_R}\lambda_{\max}\left(\mathbf{G}\right)\\ &\approx 1 - \left(1 + \sqrt{\frac{N_T}{N_R}}\right)^2. \end{aligned}\tag{4.9}$$

Finally, substituting (4.9) into (4.6), the spectral radius of the CJ iteration matrix can be approximated by:

$$\rho\left(\mathbf{B}^{\mathrm{CJ}}\right) \approx \left|1 - \left(1 + \sqrt{\frac{N_T}{N_R}}\right)^2\right| < 1, \tag{4.10}$$

which satisfy its condition once $\frac{N_T}{N_R} \ll 1$.

4.3 Proposed Damped Jacobi Algorithm

Substituting $\mathbf{R} = \mathbf{G} - \mathbf{D}$ in (4.4) and the CJ iterative algorithm can be written as:

$$\hat{\mathbf{x}}_{t+1}^{\mathrm{CJ}} = \hat{\mathbf{x}}_t^{\mathrm{CJ}} - \mathbf{D}^{-1}\mathbf{r}_t, \;\; t = 0, 1, 2, \ldots \tag{4.11}$$

where $\mathbf{D}^{-1}\mathbf{r}_t$ is a correction factor and

$$\mathbf{r}_t = \mathbf{G}\hat{\mathbf{x}}_t^{\mathrm{CJ}} - \tilde{\mathbf{y}} \tag{4.12}$$

denotes the residual vector at the t-th iteration. Then, a damped parameter $\omega \in \mathbb{R}$ is introduced in (4.11) in order to accelerate the convergence rate, i.e., to reduce the number of iterations. This algorithm is now referred as damped Jacobi (DJ) algorithm. Thus, from (4.11) the detection of $\mathbf{x}$ using the DJ iterative algorithm is defined as:

$$\hat{\mathbf{x}}_{t+1}^{\mathrm{DJ}} = \hat{\mathbf{x}}_t^{\mathrm{DJ}} - \omega\mathbf{D}^{-1}\mathbf{r}_t, \;\; t = 0, 1, 2, \ldots \tag{4.13}$$

From (4.13), it can be shown that:

$$\begin{aligned}\hat{\mathbf{x}}_{t+1}^{\mathrm{DJ}} &= \hat{\mathbf{x}}_t^{\mathrm{DJ}} - \omega\mathbf{D}^{-1}\left(\mathbf{G}\hat{\mathbf{x}}_t^{\mathrm{DJ}} - \tilde{\mathbf{y}}\right)\\ &= \hat{\mathbf{x}}_t^{\mathrm{DJ}} - \omega\mathbf{D}^{-1}\left(\left[\mathbf{D} + \mathbf{R}\right]\hat{\mathbf{x}}_t^{\mathrm{DJ}} - \tilde{\mathbf{y}}\right)\\ &= (1-\omega)\hat{\mathbf{x}}_t^{\mathrm{DJ}} + \omega\mathbf{D}^{-1}\left(\tilde{\mathbf{y}} - \mathbf{R}\hat{\mathbf{x}}_t^{\mathrm{DJ}}\right)\\ &= (1-\omega)\hat{\mathbf{x}}_t^{\mathrm{DJ}} + \omega\left(\mathbf{D}^{-1}\tilde{\mathbf{y}} - \mathbf{D}^{-1}\mathbf{R}\hat{\mathbf{x}}_t^{\mathrm{DJ}}\right)\\ &= (1-\omega)\hat{\mathbf{x}}_t^{\mathrm{DJ}} + \omega\hat{\mathbf{x}}_{t+1}^{\mathrm{CJ}}, \end{aligned}\tag{4.14}$$

where the new DJ solution $\hat{\mathbf{x}}_{t+1}^{\mathrm{DJ}}$ results from the weighted mean of the old DJ solution $\hat{\mathbf{x}}_{t}^{\mathrm{DJ}}$ and one step ahead of the CJ solution $\hat{\mathbf{x}}_{t+1}^{\mathrm{CJ}}$, given by (4.4). Therefore, the DJ algorithm accelerates the convergence rate of the CJ algorithm. Note that if ω equals to one, then $\hat{\mathbf{x}}_{t+1}^{\mathrm{DJ}} = \hat{\mathbf{x}}_{t+1}^{\mathrm{CJ}}$, therefore the DJ algorithm can be considered as a generalization of the CJ algorithm. Thus, an important issue of DJ algorithm is the choice of ω, which influences on the convergence rate (MINANGO; ALMEIDA, 2017c; MINANGO *et al.*, 2017).

4.3.1 Convergence Rate of DJ Algorithm

From (4.13), the iteration matrix of DJ algorithm is given by:

$$\mathbf{B}^{\mathrm{DJ}} = \mathbf{I}_{N_T} - \omega \mathbf{D}^{-1}\mathbf{G}. \tag{4.15}$$

Then, the convergence of the DJ algorithm is guaranteed if the spectral radius of its iteration matrix $\mathbf{B}^{\mathrm{DJ}}$ given by (4.15) is less than 1, that is:

$$\rho\left(\mathbf{B}^{\mathrm{DJ}}\right) = \left|\lambda_{\max}\left(\mathbf{B}^{\mathrm{DJ}}\right)\right| < 1. \tag{4.16}$$

Again, since $N_R \gg N_T$, and considering (4.7b) and (4.8), the largest eigenvalue of the iteration matrix $\mathbf{B}^{\mathrm{DJ}}$ given by (4.15) can be expressed as:

$$\begin{aligned}\lambda_{\max}\left(\mathbf{B}^{\mathrm{DJ}}\right) &= \lambda_{\max}\left(\mathbf{I}_{N_T}\right) + \omega\lambda_{\max}\left(\mathbf{D}^{-1}\mathbf{G}\right) \\ &\approx 1 - \omega\left(1 + \sqrt{\frac{N_T}{N_R}}\right)^2.\end{aligned} \tag{4.17}$$

Replacing (4.17) into (4.16), the spectral radius of $\mathbf{B}^{\mathrm{DJ}}$ is given by:

$$\rho\left(\mathbf{B}^{\mathrm{DJ}}\right) \approx \left|1 - \omega\left(1 + \sqrt{\frac{N_T}{N_R}}\right)^2\right|. \tag{4.18}$$

4.3.1.1 Damped Parameter

From (4.18), the correct choice of the damped parameter ω plays an important role in the convergence rate of the DJ algorithm for any initial solution $\hat{\mathbf{x}}_0^{\mathrm{DJ}}$.

Theorem 2 *For the convergence of the DJ algorithm and given any initial solution $\hat{\mathbf{x}}_0^{\mathrm{DJ}}$, the damped parameter ω must satisfy that:*

$$0 < \omega < \frac{2}{\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})}, \tag{4.19}$$

where $\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})$ is the largest eigenvalue of the matrix $\mathbf{D}^{-1}\mathbf{G}$.

Proof 1 *Consider the smallest and the largest eigenvalues of the matrix* $\mathbf{D}^{-1}\mathbf{G}$*, denoted as* $\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})$ *and* $\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})$*, respectively, which are real and positive (OLSHEVSKY; TYRTYSHNIKOV, 2010). Then, from (4.15) the eigenvalues of* $\mathbf{B}^{\mathrm{DJ}} = \mathbf{I}_{N_T} - \omega\mathbf{D}^{-1}\mathbf{G}$ *are in the range:*

$$1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G}) \leq \lambda_i(\mathbf{B}^{\mathrm{DJ}}) \leq 1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G}), \tag{4.20}$$

where $\lambda_i(\mathbf{B}^{\mathrm{DJ}})$ *denotes the i-th eigenvalue of the iteration matrix* $\mathbf{B}^{\mathrm{DJ}}$*.*

From (4.16) and (4.20), the following conditions must be satisfied for the convergence of the DJ algorithm (MINANGO et al.*, 2017):*

$$1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G}) < 1, \tag{4.21a}$$

$$1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G}) > -1. \tag{4.21b}$$

The first condition (4.21a) implies that $\omega > 0$*, while the second (4.21b) requires that* $\omega < \frac{2}{\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})}$*. In other words, the DJ algorithm converges for any initial solution* $\hat{\mathbf{x}}_{\mathrm{DJ}}^{(0)}$*, since* $\omega \in \mathbb{R}$ *satisfy:*

$$0 < \omega < \frac{2}{\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})}, \tag{4.22}$$

which completes the proof of Theorem 1.

From Theorem 1, the damped parameter ω can be chosen in the range given by (4.19) for helping in the convergence of the DJ algorithm. However, it is desirable to choose the optimum value of ω, that minimizes $\rho\left(\mathbf{B}^{\mathrm{DJ}}\right)$. Thus, considering the conditions given by (4.21), the spectral radius $\rho\left(\mathbf{B}^{\mathrm{DJ}}\right)$ given by (4.16) can be rewritten as:

$$\rho\left(\mathbf{B}^{\mathrm{DJ}}\right) = \max\left\{\left|1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})\right|, \left|1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})\right|\right\}. \tag{4.23}$$

Fig. 4.1 shows the spectral radius $\rho\left(\mathbf{B}^{\mathrm{DJ}}\right)$ as a function of ω for $0 < \omega < \frac{2}{\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})}$. From this figure, for low values of ω, the quantity $|1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})|$ dominates, i.e., $|1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})| > |1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})|$, whereas for high values of ω, the quantity $|1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})|$ dominates, i.e., $|1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})| > |1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})|$. Thus, the optimum ω is reached at the point where the curve $|1 - \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})|$ crosses the curve $|1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})|$. This occurs, when:

$$-1 + \omega\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G}) = 1 - \omega\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G}). \tag{4.24}$$

Solving (4.24), the optimum damped parameter is given by:

$$\omega_{\mathrm{opt}} = \frac{2}{\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G}) + \lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})}. \tag{4.25}$$

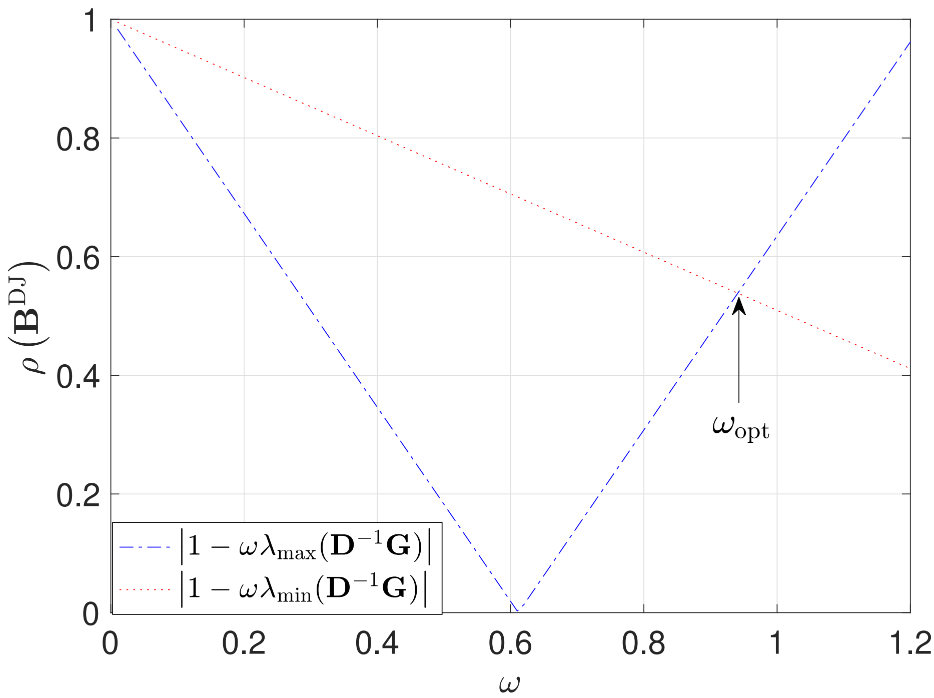


Figure 4.1 – The spectral radius $\rho\left(\mathbf{B}^{\mathrm{DJ}}\right)$ as a function of the damped parameter ω.

Note clearly that to compute (4.25), it is necessary to know a priori $\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})$ and $\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})$, which is hard to obtain in practice. Moreover, if the channel matrix $\mathbf{H}$ changes rapidly in fast time-varying channels, $\mathbf{G}$ changes rapidly too, consequently ω_{opt} given by (4.25) needs to be recalculated constantly. Thus, using (4.25) to determine the optimum damping parameter is not the best way in practical spatial multiplexing massive MIMO systems. However, by exploiting the channel property of asymptotic orthogonality (TULINO; VERDú, 2004), a quasi-optimum damped parameter $\hat{\omega}_{\mathrm{opt}}$ can be obtained in a much simpler way.

Theorem 3 *For spatial multiplexing massive MIMO systems, the quasi-optimum damped parameter of the DJ algorithm can be obtained by:*

$$\hat{\omega}_{\mathrm{opt}} = \frac{N_R}{N_R + N_T}, \tag{4.26}$$

which depends only on the number of BS antennas N_R and the number of users N_T.

Proof 2 *Considering (4.7a), (4.7b) and (4.8), the smallest and the largest eigenvalues of the matrix $\mathbf{D}^{-1}\mathbf{G}$ can be approximated, respectively, by:*

$$\lambda_{\min}\left(\mathbf{D}^{-1}\mathbf{G}\right) = \frac{1}{N_R}\lambda_{\min}\left(\mathbf{G}\right) \approx \left(1-\sqrt{\frac{N_T}{N_R}}\right)^2, \tag{4.27a}$$

$$\lambda_{\max}\left(\mathbf{D}^{-1}\mathbf{G}\right) = \frac{1}{N_R}\lambda_{\max}\left(\mathbf{G}\right) \approx \left(1+\sqrt{\frac{N_T}{N_R}}\right)^2. \tag{4.27b}$$

Finally, replacing (4.27) into (4.25), the quasi-optimum damped parameter is given by:

$$\hat{\omega}_{\text{opt}} \approx \frac{N_R}{N_R + N_T}, \tag{4.28}$$

which completes the proof of Theorem 2.

Theorem 2 indicates that the quasi-optimum damped parameter $\hat{\omega}_{\text{opt}}$ depends only on the number of receive antennas N_R at BS and the number of users N_T, which are deterministic and known. Thus, $\hat{\omega}_{\text{opt}}$ does not need to be recomputed when $\mathbf{H}$ and consequently $\mathbf{G}$ vary. Furthermore, $\hat{\omega}_{\text{opt}}$ does not need to calculate $\lambda_{\min}(\mathbf{D}^{-1}\mathbf{G})$ and $\lambda_{\max}(\mathbf{D}^{-1}\mathbf{G})$ given by (4.25), whereby $\hat{\omega}_{\text{opt}}$ is ideal for practical low-complexity spatial multiplexing massive MIMO detectors based on DJ algorithm.

Fig. 4.2 shows the difference between the optimum damped parameter given by (4.25) and the quasi-optimum damped parameter given by (4.28) versus the number of BS antennas N_R, while N_T is fixed as $N_T = 16$ users. It can be seen that $\hat{\omega}_{\text{opt}}$ is quite close to ω_{opt}, especially when N_R is large.

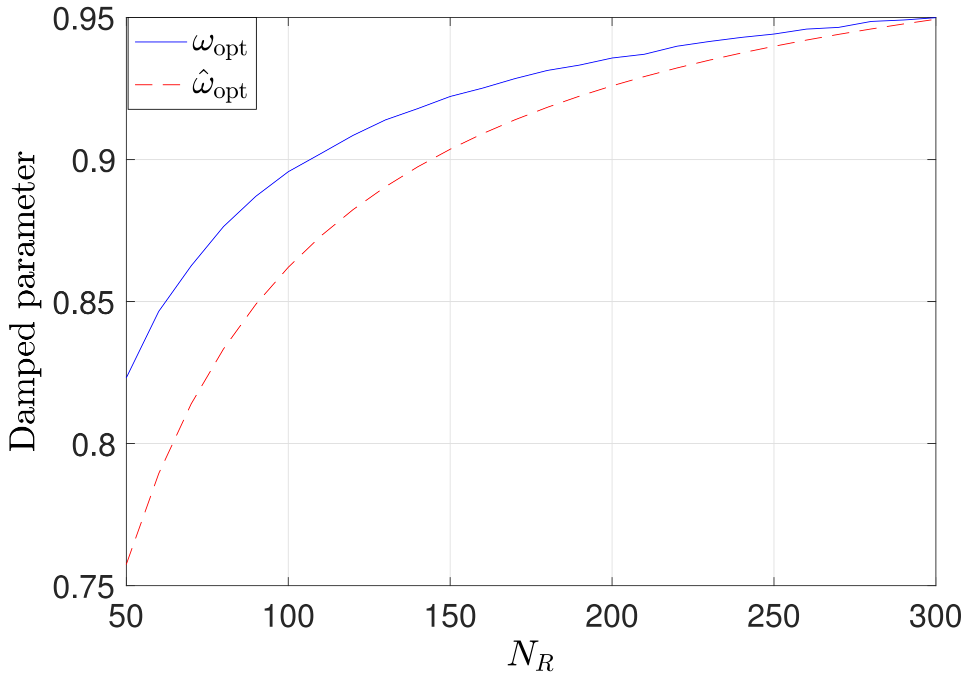


Figure 4.2 – Comparison between the optimum, ω_{opt}, and the quasi-optimum damped parameter, $\hat{\omega}_{\text{opt}}$, as a function of the number of BS antennas N_R for $N_T = 16$ users.

Finally, substituting (4.28) into (4.18), the spectral radius of the DJ iteration matrix $\mathbf{B}^{\mathrm{DJ}}$ can be approximated by:

$$\rho\left(\mathbf{B}^{\mathrm{DJ}}\right) \approx \left|1 - \frac{N_R}{N_R + N_T}\left(1 + \sqrt{\frac{N_T}{N_R}}\right)^2\right| < 1, \tag{4.29}$$

which is satisfied once $N_R \gg N_T$.

Fig. 4.3 shows both the theoretical and approximate spectral radius for both CJ and DJ algorithm as a funtion of the number of BS antennas N_R, where the number of users is fixed, $N_T = 16$. The theoretical spectral radius for CJ and DJ algorithm were computed by using, respectively, (4.6) and (4.16) over 10^5 channel realizations, while the approximate spectral radius for CJ and DJ algorithm are given, respectively, by (4.10) and (4.29).

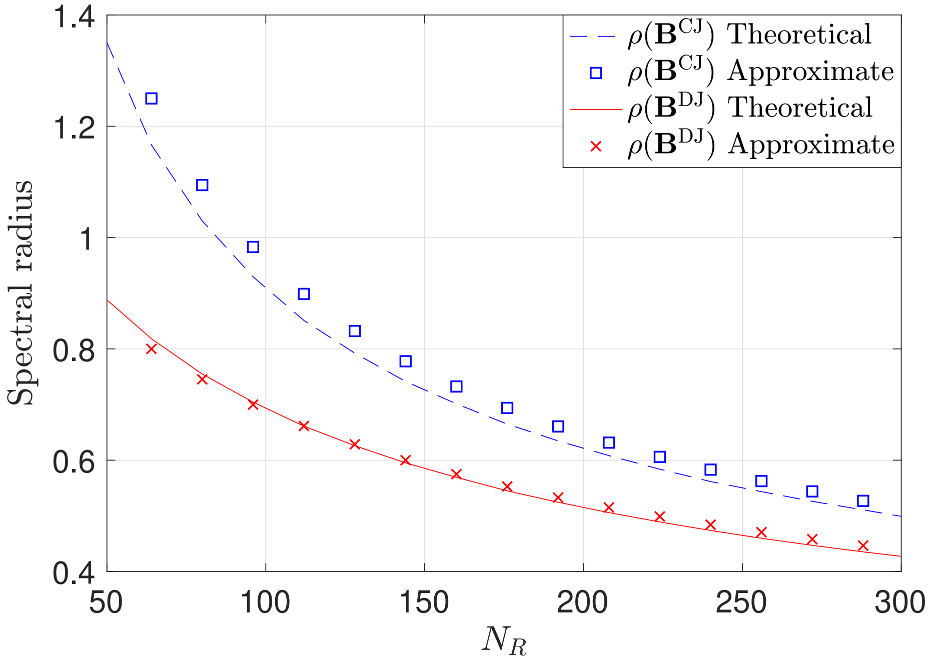


Figure 4.3 – Spectral radius of CJ and DJ algorithm as a funtion of the number of BS antennas N_R for $N_T = 16$ users.

From this figure, note that the gap between the theoretical and approximate values of the spectral radius for both CJ and DJ algorithm is negligible, which means that the approximate spectral radius is tight to the theoretical spectral radius of both algorithms. Furthermore, notice that DJ algorithm enjoys an obviously faster convergence rate than CJ algorithm, which is equivalent to a reduced number of iterations in order to achieve near-optimum performance.

In general, practical spatial multiplexing massive MIMO systems consider that BS can be equipped with $N_R = 64$, $N_R = 128$ or $N_R = 256$ receive antennas, which represents realistic scenarios (MINANGO; ALMEIDA, 2017c). Thus, considering Fig. 4.3

and these three scenarios of BS antennas, notice that when N_R is equal to 128 or 256 antennas, the spectral radius of both CJ and DJ algorithms is lesser than one consequently both algorithms converge. However, for $N_R = 64$ antennas the spectral radius of CJ algorithm is higher than one, and thus CJ algorithm does not converge while DJ algorithm is able to converge once its spectral radius is still lesser than one. This observation will be more evident in the numerical results section.

4.3.2 Initial Solution

In order to start the iterative process of the DJ method, given by (4.13), as the Gram matrix $\mathbf{G}$ is SDD, the initial solution can be written as:

$$\hat{\mathbf{x}}_0^{\mathrm{DJ}} = \hat{\omega}_{\mathrm{opt}} \mathbf{D}^{-1} \tilde{\mathbf{y}}, \tag{4.30}$$

where the computational complexity to invert the diagonal matrix $\mathbf{D}$ is low.

4.4 Numerical Results

In this section, numerical results of the BER against the SNR are provided in order to show the efficiency of DJ algorithm in comparison to CJ algorithm (KONG; PARK, 2016). The BER of ZF detector employing exact matrix inversion is also included as a benchmark. Three typical spatial multiplexing massive MIMO systems are considered with $N_R \times N_T = 64 \times 16$, $N_R \times N_T = 128 \times 16$ and $N_R \times N_T = 256 \times 16$ on an uncorrelated flat Rayleigh fading channel and 64-QAM modulation.

Fig. 4.4 shows the BER of DJ and CJ algorithms for a $N_R \times N_T = 64 \times 16$ spatial multiplexing massive MIMO system, where t denotes the number of iterations. Notice that the BER of DJ algorithm improves with the number of iterations. Thus, for $t = 14$ iterations the performance difference between DJ algorithm and ZF detector employing exact matrix inversion is within 0.1 dB, which results in an excellent performance with reduced complexity. In contrast and despite the increasing number of iterations, CJ algorithm does not converge, producing obvious BER floor, which corroborate the simulation results of the spectral radius presented in Fig. 4.3.

The BER comparisons between DJ and CJ algorithm for $N_R \times N_T = 128 \times 16$ and $N_R \times N_T = 256 \times 16$ are shown, respectively, in Fig. 4.5 and Fig. 4.6. From these figures, notice that the BER of both DJ and CJ algorithm improves with the number of iterations. However, DJ algorithm outperforms CJ algorithm for equal number of iterations. Thus, only $t = 8$ and $t = 5$ iterations are necessary to DJ algorithm achieves the ZF near-optimum performance for $N_R \times N_T = 128 \times 16$ and $N_R \times N_T = 256 \times 16$, respectively. Therefore, DJ algorithm outperforms notably CJ algorithm.

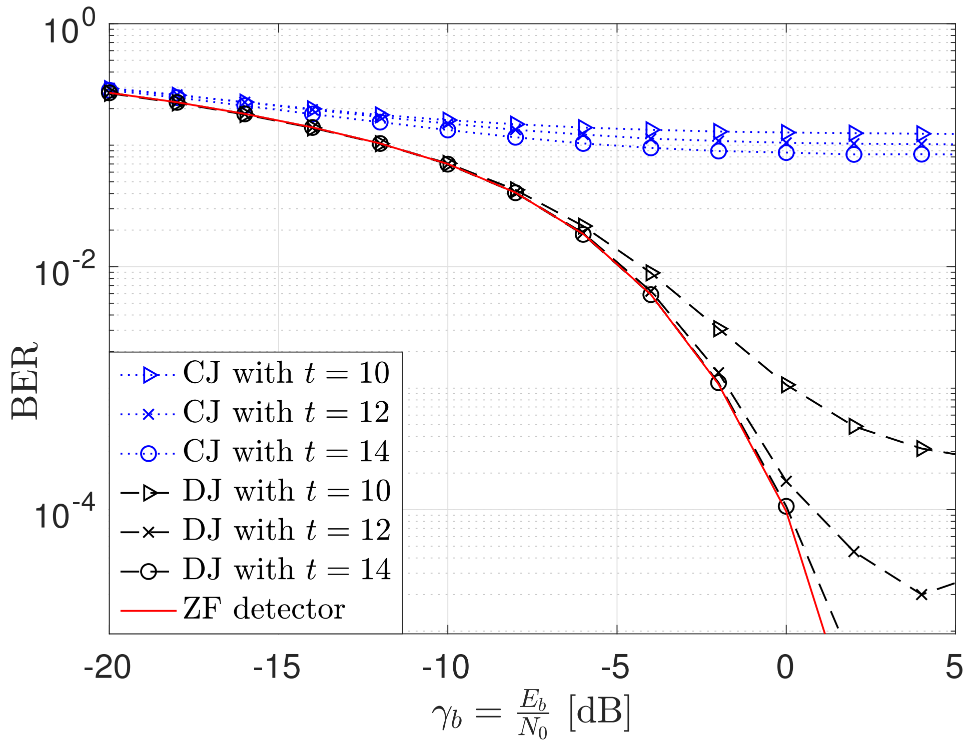


Figure 4.4 – BER versus γ_b of DJ and CJ algorithms for a $N_R \times N_T = 64 \times 16$ spatial multiplexing massive MIMO system employing 64-QAM.

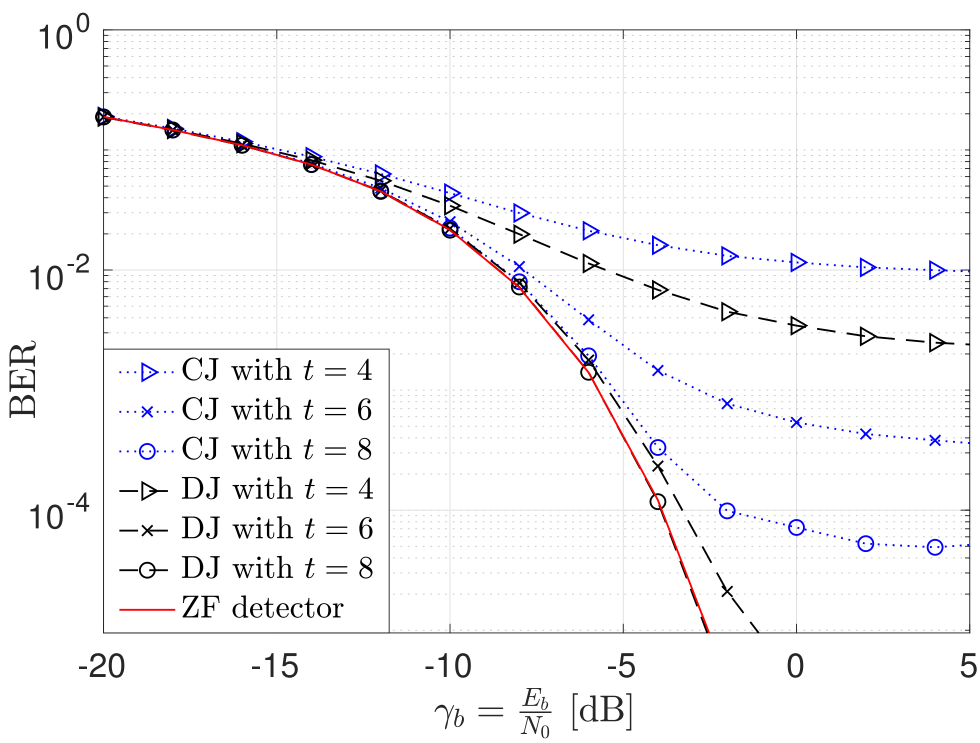


Figure 4.5 – BER versus γ_b of DJ and CJ algorithms for a $N_R \times N_T = 128 \times 16$ spatial multiplexing massive MIMO system employing 64-QAM.

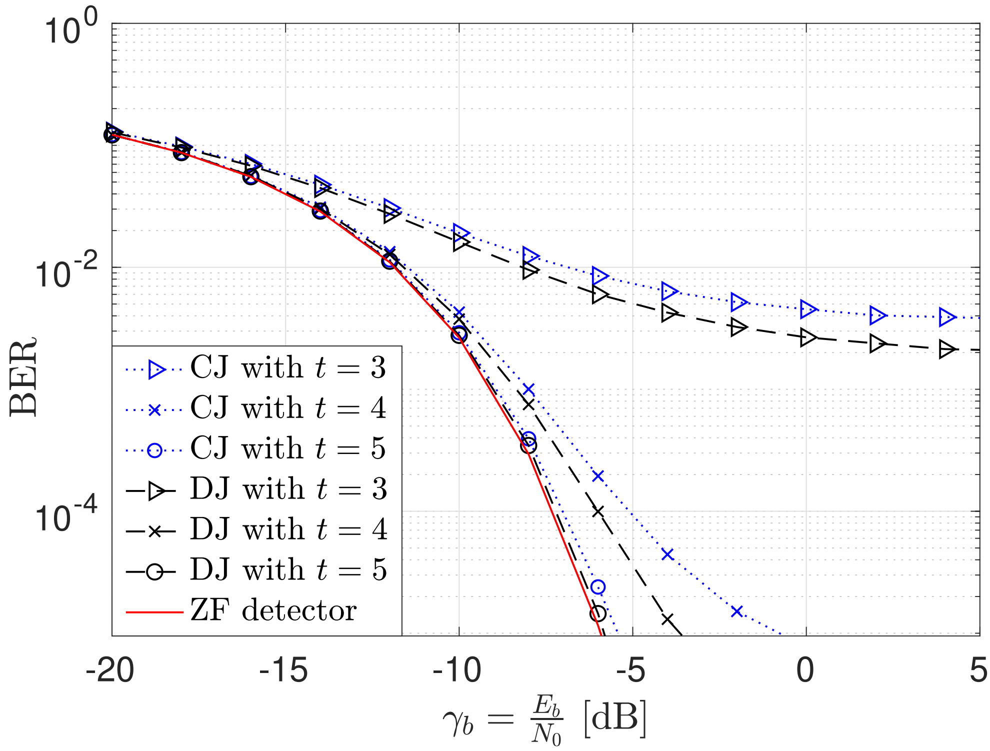


Figure 4.6 – BER versus γ_b of DJ and CJ algorithms for a $N_R \times N_T = 256 \times 16$ spatial multiplexing massive MIMO system employing 64-QAM.

Comparing Fig. 4.4, Fig. 4.5 and Fig. 4.6, in summary notice that the BER of CJ algorithm becomes worse with a decreasing number of BS antennas N_R, to the point where the algorithm does not converge, presenting a BER floor. In contrast, DJ algorithm is able to achieve the near-optimum performance as the number of iteration increases for all spatial multiplexing massive MIMO systems considered. This indicates that the convergence rate of DJ algorithm is more robust with respect to the number of BS antennas. Furthermore, given the same number of iterations, a faster convergence rate is reached by DJ algorithm in comparison to CJ algorithm as the number of BS antennas N_R increases. Note that a faster convergence rate means a smaller number of iterations required to achieve a certain accuracy and therefore less computational complexity.

4.4.1 Complexity Analysis

For the sake of analyzing the computational complexity of the proposed DJ algorithm, the term "*flop*" is used to describe multiply-add operations.

The total computational complexity is split into two parts. In the first part, the initialization step computes $\hat{\mathbf{x}}_0^{\text{DJ}}$ given by (4.30), which has $2N_T$ flops, including the inverse of $\mathbf{D}$ and the matrix-vector multiplication $\mathbf{D}^{-1}\tilde{\mathbf{y}}$. The second part comes from the t iterations given by (4.13), which involves matrix-vector multiplication and addition.

Thus, the number of operations for each iteration is $2N_T^2 - N_T$. Finally, the total number of flops is given by the sum of flops of the first and second part, given as:

$$C_{\mathrm{DJ}} = 2N_T + t\left(2N_T^2 - N_T\right). \tag{4.31}$$

Fig. 4.7 shows the computational complexity as a function of the number of users N_T for a spatial multiplexing massive MIMO system employing DJ detector algorithm with different number of iterations t. Note that the computational complexity is given in term of the number of flops. Furthermore, ZF detector with exact matrix inversion, which is obtained through singular value decomposition (SVD) (GOLUB; LOAN, 1996), is also included as benchmark. Remark that using SVD for obtaining a matrix inversion could be considered "exact" due to its numerical stability. Thus, the number of flops for computing ZF detector is given by (HIGHAM, 2008):

$$C_{\mathrm{ZF}} = 11N_T^3 + 2N_T^2. \tag{4.32}$$

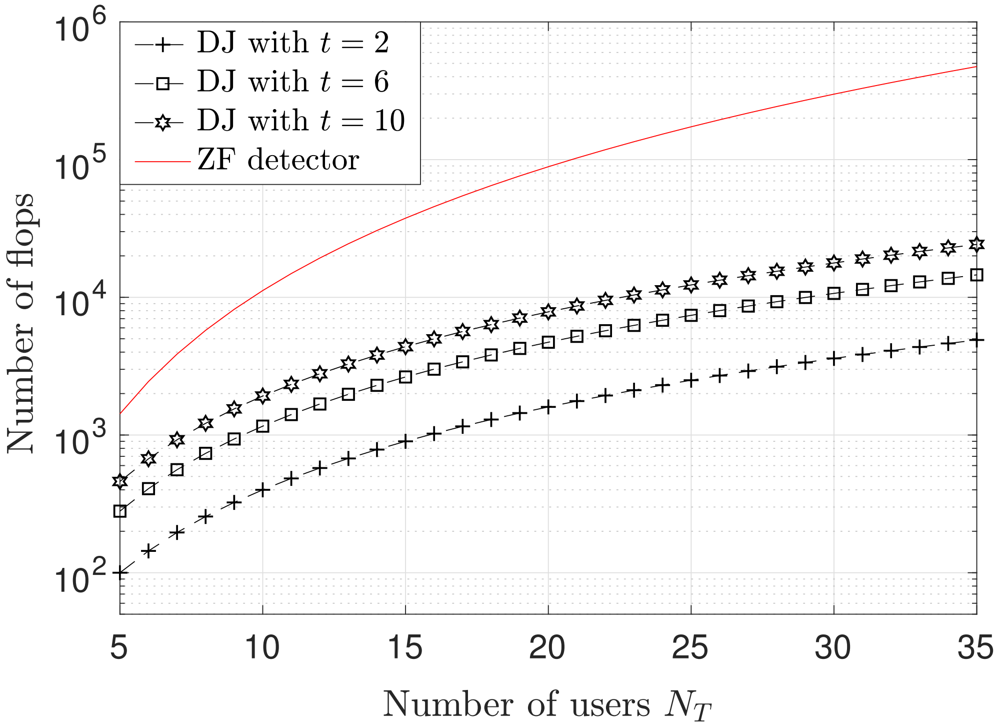


Figure 4.7 – Complexity as a function of the number of users N_T for a spatial multiplexing massive MIMO system employing DJ detector algorithm.

From Fig. 4.7, notice that the complexity of ZF detector is drastically reduced by employing the proposed DJ algorithm. Therefore, DJ algorithm can reduce the complexity from $O\left(N_T^3\right)$ to $O\left(N_T^2\right)$.

4.5 Chapter Conclusion

In this chapter, a low-complexity detector based on the Damped Jacobi (DJ) algorithm was proposed in order to achieve the near-optimum performance of ZF detector that employs exact matrix inversion in spatial multiplexing massive MIMO systems. Moreover, by exploiting the channel matrix property of asymptotic orthogonality, a simple way to compute the quasi-optimum damped parameter of DJ algorithm was developed, which depends only on the dimensions of the spatial multiplexing massive MIMO systems, that is, on the number of received antennas at the base station and the number of users. This parameter is essential to accelerate the convergence rate of DJ algorithm in comparison to the known Jacobi (CJ) algorithm. Finally, it was verified through numerical results that DJ outperforms CJ algorithm achieving the near-optimum performance of ZF detector with a small number of iterations that reduces the complexity from $O(N_T^3)$ to $O(N_T^2)$.

5 Low-Complexity ZF Detector Based on Newton-Schultz Iterative Algorithm for Massive MIMO Systems

5.1 Introduction

In spatial multiplexing massive MIMO systems, since the number of antennas at the base station (BS) is much larger than the number of users, as was pointed out in Chapter 3, ZF detector can achieve near-optimum performance due to the asymptotic orthogonal channel property (RUSEK *et al.*, 2013). However, it faces exact matrix inversion challenging problems, whose complexity grows with the cubic of the number of users.

Matrix inversion methods can be divided into two categories: direct and iterative. Direct methods basically compute the solution in a finite number of operations. Iterative methods, in turn, do not find an exact inversion solution in finite time but they converge to the inversion solution asymptotically for a prescribed tolerance (HIGHAM, 2008). Direct methods, such as Gaussian elimination, Cholesky decomposition and SVD (GOLUB; LOAN, 1996; HIGHAM, 2008) may require a not acceptable time to compute the matrix inversion when the size of the matrices is high, becoming unfeasible for practical spatial multiplexing massive MIMO systems. On the contrary, iterative methods such as Neumann series (MINANGO; ALMEIDA, 2017b), Damped Jacobi (MINANGO *et al.*, 2017), Gauss-Seidel (MINANGO; ALMEIDA, 2017a) and Newton-Schultz (BARRETT *et al.*, 1994) are preferred in problems of medium/large matrices size, due to their smaller storage requirements and computational time efficiency, becoming ideal in software/hardware implementation of spatial multiplexing massive MIMO systems.

In order to reduce the complexity of matrix inversion, (MINANGO; ALMEIDA, 2017b) and (YIN *et al.*, 2014b) have presented an approximate matrix inversion algorithm based on the polynomial expansion of Neumann series. However, this approximate approach suffers from significant performance loss when massive MIMO scales up (MINANGO; ALMEIDA, 2017b). On the other hand, although Damped Jacobi (DJ) algorithm for signal detection in spatial multiplexing massive MIMO systems described in Chapter 4 achieves a near-optimum performance with low-complexity, however, it converges linearly. Very recently, Newton-Schultz Iterative (NSI) algorithm has been employed to find the approximate matrix inversion for sphere decoding of MIMO systems (WANG; LEIB, 2013) and for massive MIMO signal detection (TANG *et al.*, 2016), respectively. Unfortunately, the number of NSI iterations are strongly dependent on the initial matrix inversion solution. Although, there are many initial matrix inversion solutions and in general, such initial solutions take a high number of iterations.

The convergence of NSI algorithm is fully related to the choice of the initial matrix inversion solution. In this chapter, it is established a relationship between DJ and NSI algorithm, where the solution after t iterations employing NSI algorithm can be seen as the solution after $2^t - 1$ iterations employing DJ algorithm. Thus, based on this relationship and due to the low latency and good numerical stability (BEN-ISRAEL, 1965) of NSI algorithm, an improvement of this algorithm for spatial multiplexing massive MIMO signal detection is proposed. The improvement is obtained by a novel initial matrix inversion solution based on Tchebychev polynomial in order to substantially accelerate the convergence rate and reduce the number of iterations of NSI algorithm. It means that when the number of iterations is limited, NSI algorithm with the proposed initial matrix inversion solution presents a faster convergence rate and consequently a low-complexity approximate matrix inversion useful for the signal detection of spatial multiplexing massive MIMO. Numerical results show that NSI algorithm can achieve the near-optimum ZF performance in just two iterations. Additionally, the band matrix (BM) concept is employed in NSI algorithm in order to further reduce the computational complexity.

The remainder of this chapter is organized as follows. Section 5.2 briefly describes the spatial multiplexing massive MIMO system model. Section 5.3 details the NSI algorithm employed to obtain the approximated matrix inversion and discusses the relationship between DJ and NSI algorithm. The proposed improved initial matrix inversion is described in Section 5.4. Then, Section 5.5 shows the performance results through simulation. Finally, conclusions are drawn in Section 5.6.

5.2 System Model

The uplink of a spatial multiplexing massive MIMO system employing N_R BS receiving antennas and N_T single-antenna transmitting users, where $N_R \gg N_T$.

Let $\mathbf{x}$ denotes the $N_T \times 1$ transmitted signal vector containing the transmitted symbols from all N_T users, $\mathbf{H}$ denotes the $N_R \times N_T$ flat Rayleigh fading channel matrix and $\mathbf{n}$ is the $N_R \times 1$ AWGN vector. Then, the $N_R \times 1$ received signal vector $\mathbf{y}$ at the BS antennas can be presented (2.13).

5.2.1 ZF Detector

For $N_R \gg N_T$, ZF detector presents as shown in Chapter 3 close to optimum performance due to the asymptotic orthogonality channel property. Then, a matrix inversion-dependent ZF detector is considered for analysis purposes in this chapter. Thus, the ZF detection of the transmitted signal vector $\mathbf{x}$ before the quantization process is given by (4.1), which was described in sub-section 4.2.1.

From (4.1), it is again emphasized that ZF detector involves the exact matrix inversion of the $N_T \times N_T$ Gram matrix $\mathbf{G}$ given by (4.3), whose computational complexity is $O(cN_T^3)$, where c denotes a multiplicative factor that depends on the algorithm used to compute $\mathbf{G}^{-1}$.

5.3 Newton-Schultz Iterative Algorithm for Computing Matrix Inversion

An iterative algorithm for computing $\mathbf{G}^{-1}$ is a procedure for generating a sequence $\{\mathbf{Z}_t : t = 0, 1, \dots\}$, where t is the number of iterations, that converges to $\mathbf{G}^{-1}$. This procedure specifies how to select the initial matrix inversion solution $\mathbf{Z}_0$, how to proceed from $\mathbf{Z}_t$ to $\mathbf{Z}_{t+1}$ for each t, and when to stop after having obtained a reasonable approximation of $\mathbf{G}^{-1}$.

In this chapter, the NSI algorithm is employed in order to compute an approximate matrix inversion. As it involves only matrix multiplications-additions, it can be implemented very efficiently on high-performance computer or hardware (BEN-ISRAEL; COHEN, 1966; TANG *et al.*, 2016). Thus, the inverse of $\mathbf{G}$ can be obtained by the NSI algorithm using the following procedure:

$$\begin{aligned}\mathbf{Z}_{t+1} &= \mathbf{Z}_t + \mathbf{Z}_t\left(\mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_t\right)\\ &= \mathbf{Z}_t\left(2\mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_t\right),\end{aligned} \tag{5.1}$$

where $\mathbf{Z}_t$ is the approximate matrix inversion solution of $\mathbf{G}$ after t iterations.

The initial matrix inversion solution $\mathbf{Z}_0$ should be chosen properly, as it determines the number of iterations required for the iterative algorithm to converge. Thus, let

$$\mathbf{E}_0 = \mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_0 \tag{5.2}$$

be the initial error matrix, for the convergence of (5.1). Then, $\mathbf{E}_0$ must satisfy the following condition:

$$\rho\left(\mathbf{E}_0\right) = \left|\lambda_{\max}\left(\mathbf{E}_0\right)\right| < 1, \tag{5.3}$$

where $\rho\left(\mathbf{E}_0\right)$ and $\lambda_{\max}\left(\mathbf{E}_0\right)$ denote, respectively, the spectral radius and the largest eigenvalue of matrix $\mathbf{E}_0$.

In the literature, there are some known initialization matrix inversion solution which guarantee convergence. In (BEN-ISRAEL; COHEN, 1966), Theorem 1 shows that the initial matrix inversion solution should be given by:

$$\mathbf{Z}_0 = \eta \mathbf{I}_{N_T}, \tag{5.4}$$

where η is positive and sufficiently small, satisfying that:

$$0 < \eta < \frac{2}{\lambda_{\max}\left(\mathbf{G}\right)}, \tag{5.5}$$

where $\lambda_{\max}\left(\mathbf{G}\right)$ is the largest eigenvalue of matrix $\mathbf{G}$.

Finally, the ZF solution of (4.1) at the $(t+1)$-th iteration using the NSI algorithm given by (5.1) is computed by the matrix-vector product given by:

$$\hat{\mathbf{x}}_{t+1}^{\mathrm{NSI}} = \mathbf{Z}_{t+1}\tilde{\mathbf{y}}. \tag{5.6}$$

5.3.1 Convergence Rate of the NSI Algorithm

In the following, it is shown that NSI algorithm converges to $\mathbf{G}^{-1}$ in a quadratic form. Let,

$$\mathbf{E}_t = \mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_t, \tag{5.7}$$

be the error matrix at the t-th iteration. Then, substituting $t+1$ by t into (5.1) and then in (5.7), $\mathbf{E}_t$ can be rewritten as:

$$\begin{aligned}
\mathbf{E}_t &= \mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_{t-1}\left(\mathbf{I}_{N_T} + \mathbf{E}_{t-1}\right) \\
&= \mathbf{E}_{t-1} - \mathbf{G}\mathbf{Z}_{t-1}\mathbf{E}_{t-1} \\
&= \left(\mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_{t-1}\right)\mathbf{E}_{t-1} \\
&= \mathbf{E}_{t-1}^2.
\end{aligned} \tag{5.8}$$

where it is used that $\mathbf{I}_{N_T} - \mathbf{G}\mathbf{Z}_{t-1} = \mathbf{E}_{t-1}$. Thus, by induction, it can be shown that:

$$\mathbf{E}_t = \left(\mathbf{E}_0\right)^{2^t}, \tag{5.9}$$

which proves that NSI algorithm has quadratic convergence order.

From (BEN-ISRAEL, 1965), it is shown that:

$$\lim_{t\to\infty}\mathbf{E}_t = \lim_{t\to\infty}\left(\mathbf{E}_0\right)^{2^t} = \mathbf{0}. \tag{5.10}$$

Then, from the definition of $\mathbf{E}_t$ given by (5.7)

$$\mathbf{Z}_t = \mathbf{G}^{-1}\left(\mathbf{I}_{N_T} - \mathbf{E}_t\right). \tag{5.11}$$

Therefore:

$$\begin{aligned}\lim_{t\to\infty}\mathbf{Z}_t &= \lim_{t\to\infty}\mathbf{G}^{-1}\left(\mathbf{I}_{N_T}-\mathbf{E}_t\right)\\ &= \mathbf{G}^{-1}.\end{aligned} \tag{5.12}$$

5.3.2 Initial Matrix Inversion Solution $\mathbf{Z}_0$

In (BEN-ISRAEL; COHEN, 1966), Theorem 5 shows that the optimum choice of η in (5.4) is given by:

$$\eta = \frac{2}{\lambda_{\max}\left(\mathbf{G}\right)+\lambda_{\min}\left(\mathbf{G}\right)}, \tag{5.13}$$

where $\lambda_{\min}\left(\mathbf{G}\right)$ is the smallest eigenvalue of $\mathbf{G}$. This optimum η has been employed in (WANG; LEIB, 2013).

In practice, it is difficult to calculate $\lambda_{\max}\left(\mathbf{G}\right)$ and $\lambda_{\min}\left(\mathbf{G}\right)$. Thus, a sub-optimum alternative for η employs (HAGHANI; SOLEYMANI, 2014):

$$\eta \cong \frac{1}{\left\|\mathbf{G}\right\|_1\left\|\mathbf{G}\right\|_\infty}, \tag{5.14}$$

where

$$\left\|\mathbf{G}\right\|_1 = \max_j\left(\sum_{i=1}^{N_T}\left|g_{ij}\right|\right), \tag{5.15}$$

and

$$\left\|\mathbf{G}\right\|_\infty = \max_i\left(\sum_{j=1}^{N_T}\left|g_{ij}\right|\right). \tag{5.16}$$

On the other hand, since the $N_T \times N_T$ Gram matrix $\mathbf{G}$ given by (4.3) is SDD, it can be approximated by $\mathbf{G}\approx\mathbf{D}$ (RUSEK *et al.*, 2013), where $\mathbf{D}$ is the diagonal matrix of $\mathbf{G}$. Thus, another choice of η which does not require an estimate of $\lambda_{\max}\left(\mathbf{G}\right)$ and $\lambda_{\min}\left(\mathbf{G}\right)$ is available in (TANG *et al.*, 2016), that is given by:

$$\eta = \mathbf{D}^{-1} \approx \frac{1}{N_T}. \tag{5.17}$$

However, as both (5.14) and (5.17) are sub-optimal solutions of η, the number of iterations necessary to converge increases, as well as, the complexity.

In Section 5.4, a new method for finding an improved initialization matrix inversion solution $\mathbf{Z}_0$ is proposed, which reduces the number of iterations and the complexity ensuring convergence.

5.3.3 Stopping Criteria

The stopping criterion is one of the most important factors, which could affect the computational time of NSI algorithm in practical implementations.

A termination criterion is to stop the algorithm after a fixed number of iterations, until the required tolerance in relation to the exact matrix inversion solution is obtained. An alternative is the a posteriori stopping criteria, which use already computed iterations to decide when to stop. The main drawback is that the exact matrix inversion solution is unknown. In this paper, based on (5.7), the error matrix Frobenius norm termination is used as stopping criterion, given by:

$$\|\mathbf{E}_t\|_F \leq \epsilon, \tag{5.18}$$

where ϵ is the prescribed tolerance.

5.3.4 Comparison between DJ and NSI algorithm

In this sub-section, it is shown that the ZF solution of (4.1) after t iterations employing NSI algorithm given by (5.6) is the ZF solution after $2^t - 1$ iterations employing DJ algorithm given by (4.13). For this, firstly a relationship between DJ and Neumann series (NS) algorithm is established. Then, based on this relationship, DJ and NSI algorithms are compared.

5.3.4.1 Relationship between DJ and NS algorithm

5.3.4.1.1 NS algorithm

NS algorithm is employed to carry out the approximate matrix inversion through polynomial expansion. Thus, according to NS algorithm the matrix inversion of $\mathbf{G}$ can be written by a t-order matrix polynomial as

$$\begin{aligned}\mathbf{G}^{-1} &\approx \sum_{n=1}^{t} \left(\mathbf{I}_{N_T} - \mathbf{D}^{-1}\mathbf{G}\right)^{n-1} \mathbf{D}^{-1} \\ &\approx \sum_{n=1}^{t} (-1)^{n-1} \left(\mathbf{D}^{-1}\mathbf{R}\right)^{n-1} \mathbf{D}^{-1},\end{aligned} \tag{5.19}$$

where $\mathbf{R}$ is the off-diagonal matrix of $\mathbf{G}$.

The convergence of (5.19) is guaranteed if the spectral radius of $(\mathbf{I}_{N_T} - \mathbf{D}^{-1}\mathbf{G})$ is less than one, that is:

$$\rho\left(\mathbf{I}_{N_T} - \mathbf{D}^{-1}\mathbf{G}\right) < 1. \tag{5.20}$$

If condition (5.20) is satisfied, the approximation approaches equality as $t \to \infty$ (MINANGO; ALMEIDA, 2017b). However, note that for $t = 3$, the complexity of NS algorithm is $O(N_T^3)$, which shows that only a marginal complexity reduction can be obtained.

5.3.4.1.2 Revisited DJ Algorithm

Rewriting the DJ iterative algorithm given by (4.13)

$$\begin{aligned}\hat{\mathbf{x}}_{t+1}^{\mathrm{DJ}} &= \hat{\mathbf{x}}_t^{\mathrm{DJ}} - \omega\mathbf{D}^{-1}\left(\mathbf{G}\hat{\mathbf{x}}_t^{\mathrm{DJ}} - \tilde{\mathbf{y}}\right)\\ &= \left(\mathbf{I}_{N_T} - \omega\mathbf{D}^{-1}\mathbf{G}\right)\hat{\mathbf{x}}_t + \omega\mathbf{D}^{-1}\tilde{\mathbf{y}}\\ &= \mathbf{B}^{\mathrm{DJ}}\hat{\mathbf{x}}_t + \omega\mathbf{D}^{-1}\tilde{\mathbf{y}},\ \ t = 0,1,2,\ldots \end{aligned} \tag{5.21}$$

Expanding it for $t = 0,1,2$ in terms of the initial solution $\hat{\mathbf{x}}_0^{\mathrm{DJ}} = \omega\mathbf{D}^{-1}\tilde{\mathbf{y}}$, that is:

$$\begin{aligned}\hat{\mathbf{x}}_1^{\mathrm{DJ}} &= \mathbf{B}^{\mathrm{DJ}}\hat{\mathbf{x}}_0 + \omega\mathbf{D}^{-1}\tilde{\mathbf{y}} = \left[\mathbf{B}^{\mathrm{DJ}}\mathbf{D}^{-1} + \mathbf{D}^{-1}\right]\omega\tilde{\mathbf{y}},\\ \hat{\mathbf{x}}_2^{\mathrm{DJ}} &= \mathbf{B}^{\mathrm{DJ}}\hat{\mathbf{x}}_1 + \omega\mathbf{D}^{-1}\tilde{\mathbf{y}} = \left[(\mathbf{B}^{\mathrm{DJ}})^2\mathbf{D}^{-1} + \mathbf{B}^{\mathrm{DJ}}\mathbf{D}^{-1} + \mathbf{D}^{-1}\right]\omega\tilde{\mathbf{y}},\\ \hat{\mathbf{x}}_3^{\mathrm{DJ}} &= \mathbf{B}^{\mathrm{DJ}}\hat{\mathbf{x}}_2 + \omega\mathbf{D}^{-1}\tilde{\mathbf{y}} = \left[(\mathbf{B}^{\mathrm{DJ}})^3\mathbf{D}^{-1} + (\mathbf{B}^{\mathrm{DJ}})^2\mathbf{D}^{-1} + \mathbf{B}^{\mathrm{DJ}}\mathbf{D}^{-1} + \mathbf{D}^{-1}\right]\omega\tilde{\mathbf{y}}.\end{aligned}$$

It is easy to shown that the general form of this expansion that represents the $(t+1)$-th iteration of DJ algorithm is given by:

$$\begin{aligned}\hat{\mathbf{x}}_{t+1}^{\mathrm{DJ}} &= \left[\sum_{n=1}^{t+1}\left(\mathbf{B}^{\mathrm{DJ}}\right)^{n-1}\mathbf{D}^{-1}\right]\omega\tilde{\mathbf{y}}\\ &= \left[\sum_{n=1}^{t+1}\left(\mathbf{I}_{N_T} - \omega\mathbf{D}^{-1}\mathbf{G}\right)^{n-1}\mathbf{D}^{-1}\right]\omega\tilde{\mathbf{y}}\\ &= \left[\sum_{n=1}^{t+1}(-1)^{n-1}\left(\omega\mathbf{D}^{-1}\mathbf{R}\right)^{n-1}\mathbf{D}^{-1}\right]\omega\tilde{\mathbf{y}}\end{aligned} \tag{5.22}$$

Comparing (5.19) and (5.22), observe that the ZF solution after t iterations of the DJ algorithm is equivalent to t terms of the polynomial expansion of the NS algorithm, which means that both algorithms have the same precision.

5.3.4.2 Relationship between DJ and NSI Algorithm

Consider the ZF solution employing the NSI algorithm given by (5.6). Expanding $\mathbf{Z}_{t+1}$ for $t = 0,1,2$ in terms of the initial matrix inversion solution $\mathbf{Z}_0$, given by (5.4) with $\eta = \mathbf{D}^{-1}$, observe that:

$$\begin{aligned}\hat{\mathbf{x}}_1^{\mathrm{NSI}} &= \mathbf{Z}_1\tilde{\mathbf{y}} = [2\mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{G}\mathbf{D}^{-1}]\tilde{\mathbf{y}},\\ \hat{\mathbf{x}}_2^{\mathrm{NSI}} &= \mathbf{Z}_2\tilde{\mathbf{y}} = [4\mathbf{D}^{-1} - 6(\mathbf{D}^{-1}\mathbf{G})\mathbf{D}^{-1} + 4(\mathbf{D}^{-1}\mathbf{G})^2\mathbf{D}^{-1} - (\mathbf{D}^{-1}\mathbf{G})^3\mathbf{D}^{-1}]\tilde{\mathbf{y}},\\ \hat{\mathbf{x}}_3^{\mathrm{NSI}} &= \mathbf{Z}_3\tilde{\mathbf{y}} = [8\mathbf{D}^{-1} - 28(\mathbf{D}^{-1}\mathbf{G})\mathbf{D}^{-1} + 56(\mathbf{D}^{-1}\mathbf{G})^2\mathbf{D}^{-1} - 70(\mathbf{D}^{-1}\mathbf{G})^3\mathbf{D}^{-1} + 56(\mathbf{D}^{-1}\mathbf{G})^4\mathbf{D}^{-1}\\ &\qquad - 28(\mathbf{D}^{-1}\mathbf{G})^6\mathbf{D}^{-1} + 8(\mathbf{D}^{-1}\mathbf{G})^6\mathbf{D}^{-1} - (\mathbf{D}^{-1}\mathbf{G})^7\mathbf{D}^{-1}]\tilde{\mathbf{y}}.\end{aligned}$$

Thus, the general form of the expansion of the ZF solution employing the NSI algorithm is given by:

$$\hat{\mathbf{x}}_{t+1}^{\mathrm{NSI}} = \left[\sum_{n=1}^{2^t} \binom{2^t}{n} (-1)^{n-1} \left(\mathbf{D}^{-1}\mathbf{G}\right)^{n-1} \mathbf{D}^{-1} \right] \tilde{\mathbf{y}}, \tag{5.23}$$

where the number of terms to be added grows exponentially with 2^t. Therefore, NSI algorithm converges quadratically.

Comparing (5.22) and (5.23), notice that the ZF solution employing NSI algorithm after t iterations is the ZF solution employing DJ algorithm after $2^t - 1$ iterations. Thus, the following relationship between DJ and NSI algorithm can be established:

$$\hat{\mathbf{x}}_{t}^{\mathrm{NSI}} = \hat{\mathbf{x}}_{\sum_{n=0}^{t-1} 2^n}^{\mathrm{DJ}}. \tag{5.24}$$

From (5.24) observe that the NSI algorithm converges much faster than the DJ algorithm for the same spatial multiplexing massive MIMO system. In order to confirm this, the normalized mean-square error (MSE) is considered as a metric to evaluate the convergence rate of the ZF solution employing DJ and NSI algorithms. Thus, for the exact ZF solution $\hat{\mathbf{x}}_{\mathrm{ZF}}$, given by (4.1), and the ZF solution employing DJ and NSI algorithms given, respectively, by (5.22) and (5.23), their normalized MSE are given, respectively, by:

$$\mathrm{MSE}^{\mathrm{DJ}} = \frac{\mathbb{E}\left\{\left\|\hat{\mathbf{x}}_{\mathrm{ZF}} - \hat{\mathbf{x}}_t^{\mathrm{DJ}}\right\|\right\}}{N_T^2} \quad \text{and} \quad \mathrm{MSE}^{\mathrm{NSI}} = \frac{\mathbb{E}\left\{\left\|\hat{\mathbf{x}}_{\mathrm{ZF}} - \hat{\mathbf{x}}_t^{\mathrm{NSI}}\right\|\right\}}{N_T^2}. \tag{5.25}$$

Fig. 5.1 shows the normalized MSE of DJ and NSI algorithm versus the number of iterations for a $N_R \times N_T = 128 \times 16$ spatial multiplexing massive MIMO system employing 64-QAM. From this figure, note that the normalized MSE of NSI algorithm decreases faster in comparison to the the normalized MSE of DJ algorithm. Thus, after 5 iterations, NSI algorithm has achieved an accurate ZF solution with a normalized MSE less than 10^{-15}, while DJ algorithm requires about 31 iterations in order to achieve the same normalized MSE, which verifies the relationship between DJ and NSI algorithm given by (5.24).

5.4 An Improved Initial Matrix Inversion Solution for Newton-Schultz Iterative Algorithm

Be the eigen-decomposition of matrix $\mathbf{G}$ as (HIGHAM, 2008):

$$\mathbf{G} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H, \tag{5.26}$$

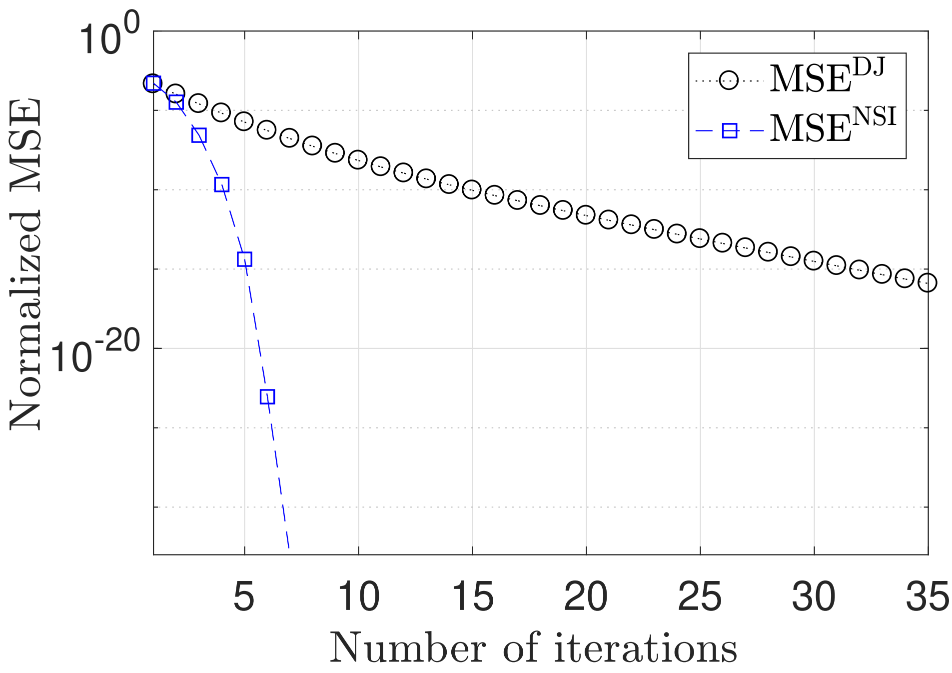


Figure 5.1 – Normalized MSE of DJ and NSI versus algorithm versus the number of iterations for a $N_R \times N_T = 128 \times 16$ spatial multiplexing massive MIMO system employing 64-QAM.

where $\mathbf{V}$ is a orthogonal matrix of size $N_T \times N_T$, whose i-th column is the eigenvector $\mathbf{v}_i$ of $\mathbf{G}$ and

$$\boldsymbol{\Lambda} = \text{diag}\left[\lambda_{\max}\left(\mathbf{G}\right) \geq \cdots \geq \lambda_i\left(\mathbf{G}\right) \geq \cdots \geq \lambda_{\min}\left(\mathbf{G}\right)\right] \tag{5.27}$$

is a diagonal matrix, whose elements are the corresponding eigenvalues. The initial matrix inversion solution is proposed as:

$$\mathbf{Z}_0 = \zeta \mathbf{I}_{N_T} + \phi \mathbf{G}. \tag{5.28}$$

From (5.28), the matrix product $\mathbf{G}\mathbf{Z}_0$ can be written as:

$$\mathbf{G}\mathbf{Z}_0 = \mathbf{G}\left(\zeta \mathbf{I}_{N_T} + \phi \mathbf{G}\right). \tag{5.29}$$

Doing the eigen-decomposition of $\mathbf{G}$ in (5.29), using (5.26), (5.29) can be rewritten as:

$$\begin{aligned} \mathbf{G}\mathbf{Z}_0 &= \left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H\right)\left(\zeta \mathbf{I}_{N_T} + \phi \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H\right) \\ &= \mathbf{V}\underbrace{\left(\zeta\boldsymbol{\Lambda} + \phi\boldsymbol{\Lambda}^2\right)}_{\mathbf{P}}\mathbf{V}^H \end{aligned} \tag{5.30}$$

As $\boldsymbol{\Lambda}$ is a diagonal matrix, $\mathbf{P}$ is a diagonal matrix too. Thus, the i-th diagonal element of $\mathbf{P}$ is given by:

$$p_i = \zeta \lambda_i\left(\mathbf{G}\right) + \phi \lambda_i^2\left(\mathbf{G}\right), \quad i = 1, \ldots, N_T, \tag{5.31}$$

where $\lambda_i(\mathbf{G})$ is the i-th diagonal element of $\mathbf{\Lambda}$.

Therefore, from (5.30) and (5.31), ζ and ϕ can be chosen such that the spectral radius of the initial error matrix:

$$\begin{aligned}\rho\left(\mathbf{I}_{N_T}-\mathbf{G}\mathbf{Z}_0\right)&=\rho\left(\mathbf{I}_{N_T}-\mathbf{P}\right)\\&=\max_{1\leq i\leq N_T}\left|1-p_i\right|,\end{aligned}\tag{5.32}$$

is minimized. In (FABER *et al.*, 2010), $1-p_i$ is a scaled Tchebyshev polynomial on the interval:

$$\left[\lambda_{\min}\left(\mathbf{G}\right),\ \lambda_{\max}\left(\mathbf{G}\right)\right]\tag{5.33}$$

that is represented by:

$$1-p_{\mathrm{i}}=\mathrm{T}_2\left[\lambda_i\left(\mathbf{G}\right)\right]\triangleq\frac{2\left[\lambda_i\left(\mathbf{G}\right)-\lambda_{\mathrm{mid}}\left(\mathbf{G}\right)\right]^2-\delta^2}{2\lambda_{\mathrm{mid}}^2\left(\mathbf{G}\right)-\delta^2},\tag{5.34}$$

where

$$\lambda_{\mathrm{mid}}\left(\mathbf{G}\right)=\frac{\lambda_{\max}\left(\mathbf{G}\right)+\lambda_{\min}\left(\mathbf{G}\right)}{2},\tag{5.35}$$

and

$$\delta=\lambda_{\max}\left(\mathbf{G}\right)-\lambda_{\mathrm{mid}}\left(\mathbf{G}\right).\tag{5.36}$$

Therefore, using (5.34) - (5.36) into (5.32), it can be shown that:

$$\begin{aligned}\rho\left(\mathbf{I}_{N_T}-\mathbf{G}\mathbf{Z}_0\right)&=\max_{1\leq i\leq N_T}\left|\mathrm{T}_2\left[\lambda_i\left(\mathbf{G}\right)\right]\right|\\&=\left|\mathrm{T}_2\left[\lambda_{\max}\left(\mathbf{G}\right)\right]\right|\\&=\frac{\delta^2}{2\lambda_{\mathrm{mid}}^2\left(\mathbf{G}\right)-\delta^2}.\end{aligned}\tag{5.37}$$

To find ζ, ϕ and hence $\mathbf{Z}_0$, the following equations should be used:

$$\begin{aligned}p_i=\zeta\lambda_i+\phi\lambda_i^2&=1-\mathrm{T}_2\left[\lambda_i\left(\mathbf{G}\right)\right]\\&=\frac{2\lambda_{\mathrm{mid}}^2\left(\mathbf{G}\right)-\delta^2-2\left[\lambda_i\left(\mathbf{G}\right)-\lambda_{\mathrm{mid}}\left(\mathbf{G}\right)\right]^2+\delta^2}{2\lambda_{\mathrm{mid}}^2-\delta^2}\\&=\frac{-2\lambda_i^2\left(\mathbf{G}\right)+4\lambda_i\left(\mathbf{G}\right)\lambda_{\mathrm{mid}}\left(\mathbf{G}\right)}{2\lambda_{\mathrm{mid}}^2\left(\mathbf{G}\right)-\delta^2}.\end{aligned}\tag{5.38}$$

Hence:

$$\zeta=\frac{4\lambda_{\mathrm{mid}}\left(\mathbf{G}\right)}{2\lambda_{\mathrm{mid}}^2\left(\mathbf{G}\right)-\delta^2};\quad\text{and}\quad\phi=-\frac{2}{2\lambda_{\mathrm{mid}}^2\left(\mathbf{G}\right)-\delta^2}.\tag{5.39}$$

From (5.28) and the previous developments, notice that the proposed initialization matrix inversion solution $\mathbf{Z}_0$ depends on the knowledge of $\lambda_{\min}(\mathbf{G})$ and $\lambda_{\max}(\mathbf{G})$, which is difficult to obtain in practice. However, in massive MIMO systems, as N_R is very large in comparison to N_T, the largest and the smallest eigenvalues of $\mathbf{G}$ can be approximated, respectively, by (4.7a) and (4.7b).

Thus, based on (4.7), (5.35) and (5.36), the following equation can be rewritten as:

$$\lambda_{\text{mid}}\left(\mathbf{G}\right) \approx N_R\left(1+\frac{1}{\beta}\right), \tag{5.40}$$

$$\delta \approx 2N_R\sqrt{\frac{1}{\beta}}. \tag{5.41}$$

where $\beta = \frac{N_T}{N_R}$.

Finally, the practical $\hat{\zeta}$ and $\hat{\phi}$ are obtained, respectively, by:

$$\hat{\zeta} = \frac{2\beta\left(1+\beta\right)}{N_R\left(1+\beta^2\right)} \quad \text{and} \quad \hat{\phi} = -\frac{\beta^2}{N_R^2\left(1+\beta^2\right)}, \tag{5.42}$$

where $\hat{\zeta}$ and $\hat{\phi}$ depends only on the number of transmit, N_T, and receive, N_R, antennas. Thus, once the massive MIMO system configuration is fixed, that is N_T and N_R are established, the computation of the proposed initial matrix inversion solution $\mathbf{Z}_0$ is easy.

Fig. 5.2 shows the theoretical and practical $\rho\left(\mathbf{Z}_0\right)$ obtained by employing, respectively, (5.39) and (5.42) versus β. Notice that the gap between theoretical and practical values is negligible, especially when $N_R \gg N_T$, which reflects the good approximation between (5.39) and (5.42) to compute $\mathbf{Z}_0$.

5.5 Numerical Results

In this section, firstly, the numerical convergence rate and the stopping criteria of NSI algorithm for distinct initialization matrices inversion solution $\mathbf{Z}_0$ are analyzed. Subsequently, the effects of the number of iterations on the NSI performance using the proposed $\mathbf{Z}_0$ are investigated. Then, the complexity of NSI algorithm is determined. Finally, a method to reduce further the complexity is presented. The 64-QAM modulation is employed in all simulations.

5.5.1 Numerical Convergence Rate and Stopping Criteria

Fig. 5.3 presents $\|\mathbf{E}_t\|_F$ given by (5.18) as a function of the number of iterations for four distinct initial solution for $N_R = 128$ and $N_T = 16$.

Observe in Fig. 5.3 that the proposed $\mathbf{Z}_0$ (5.28) has a very fast convergence rate. Only two NSI iterations are necessary to obtain that $\|\mathbf{E}_t\|_F \leq 10^{-3}$, which is an excellent prescribed tolerance ϵ to stop the iterative algorithm and find a good approximation of

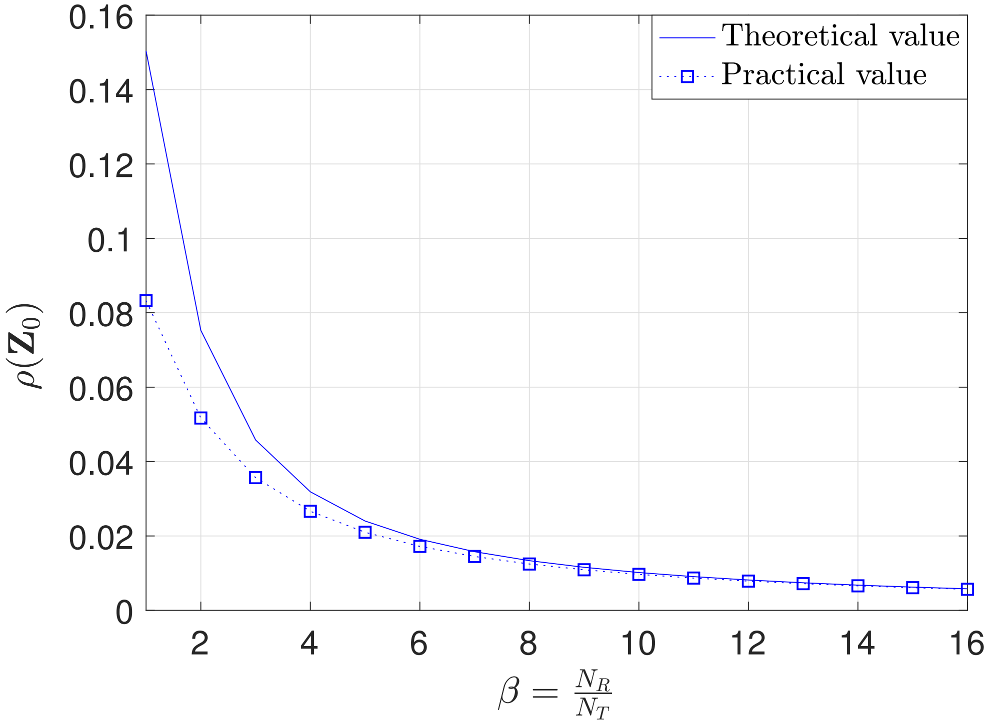


Figure 5.2 – Comparison between the theoretical and practical $\rho(\mathbf{X}_0)$ against $\beta = \frac{N_R}{N_T}$ where $N_T = 16$.

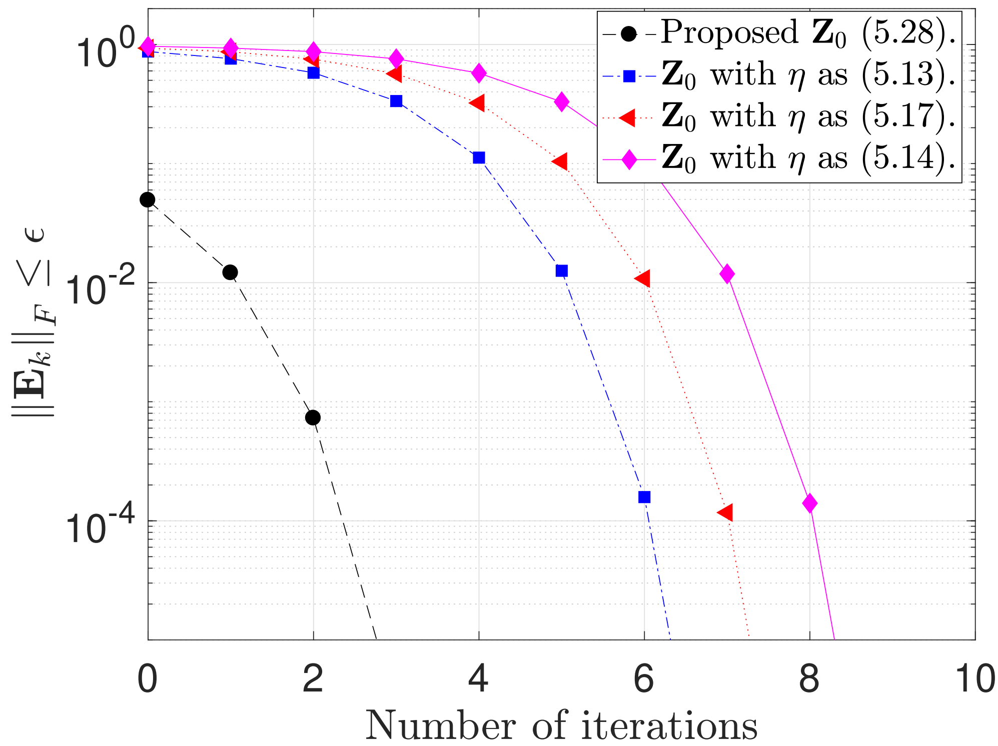


Figure 5.3 – Error matrix Frobenius norm $\|\mathbf{E}_t\|_F$ as a function of the number of iterations for a 16×16 matrix $\mathbf{G}$.

the inverse of $\mathbf{G}$. On the other hand, using the initial solution given by (5.4) with the optimum η given by (5.13) and the sub-optimums η given by (5.17) and (5.14) more than 5, 6 and 7 iterations are necessary, respectively, in order to reach that $\|\mathbf{E}_t\|_F \leq 10^{-3}$. Hence, the proposed initial solution, given by (5.28), clearly outperforms the solutions proposed in the literature.

5.5.2 BER Performance

In order to evaluate the performance of NSI algorithm with the proposed $\mathbf{Z}_0$ given by (5.28), the BER simulation results includes also DJ algorithm. Furthermore, the BER of ZF detector with exact matrix inversion using SVD decomposition is also included as benchmark. In this section, two spatial multiplexing massive MIMO systems with $N_R \times N_T = 128 \times 16$ and $N_R \times N_T = 128 \times 32$, respectively, are considered. Note that $N_R = 128$ received antennas represents a realistic scenario nowadays (MARZETTA *et al.*, 2016). Perfect knowledge of the Rayleigh channel matrix $\mathbf{H}$ is assumed at the receiver.

Fig. 5.4 shows the BER versus γ_b for NSI algorithm with the proposed initial solution $\mathbf{Z}_0$ and DJ algorithm. The spatial multiplexing massive MIMO system is $N_R \times N_T = 128 \times 16$, where t denotes the number of iterations. Notice that the BER performance of both algorithms improves with the number of iterations. However, NSI algorithm outperforms DJ algorithm. Observe that NSI algorithm with the proposed $\mathbf{Z}_0$ can achieve ZF performance in just 2 iterations, which evidences the faster convergence rate of NSI algorithm.

Fig. 5.5 shows the BER versus γ_b for NSI algorithm with the proposed $\mathbf{Z}_0$ and DJ algorithm for a massive MIMO system with $N_R \times N_T = 128 \times 32$. By comparing Fig. 5.4 and Fig. 5.5, note that increasing N_T, the BER performance of DJ algorithm becomes worse. DJ algorithm needs around 8 iterations to approach near-optimum ZF performance for $N_R \times N_T = 128 \times 16$ (see Fig. 4.5 in chapter 4), while for $N_R \times N_T = 128 \times 32$ around 12 iterations are needed. In contrast, NSI algorithm with the proposed $\mathbf{Z}_0$ requires just 2 iterations in both spatial multiplexing massive MIMO systems. This indicates that, as it has addressed in Section 5.3, a suitable initialization solution produces faster convergence rate, that requires a smaller number of iterations to achieve a certain accuracy.

Finally, Fig. 5.6 shows the BER versus N_T for ZF detector with exact matrix inversion and NSI algorithm with the proposed $\mathbf{Z}_0$ after 2 iterations, employing 64-QAM, for $\gamma_b = -15$ dB and $N_R = 128$ antennas. Also, the GA detector is employed as a benchmark. Note that NSI performance fits perfectly the ZF performance for $N_T \leq 60$. For $N_T > 60$, the NSI performance is pretty close to ZF performance.

Furthermore, from Fig. 5.6 notice that the performance gap between GA de-

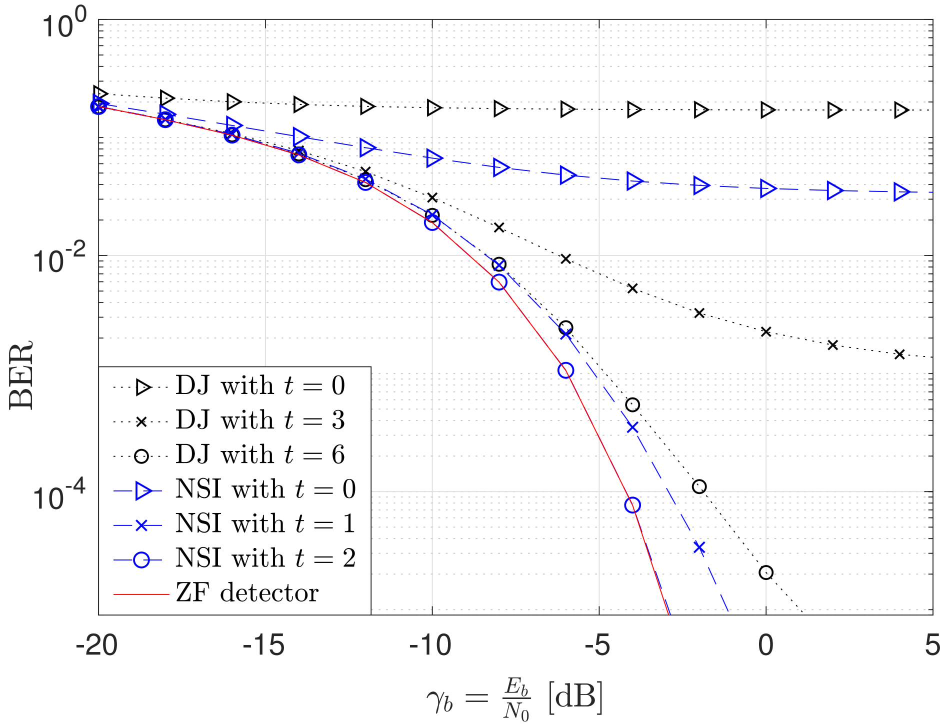


Figure 5.4 – BER versus γ_b of NSI and DJ algorithm for a $N_R \times N_T = 128 \times 16$ spatial multiplexing massive MIMO system employing 64-QAM.

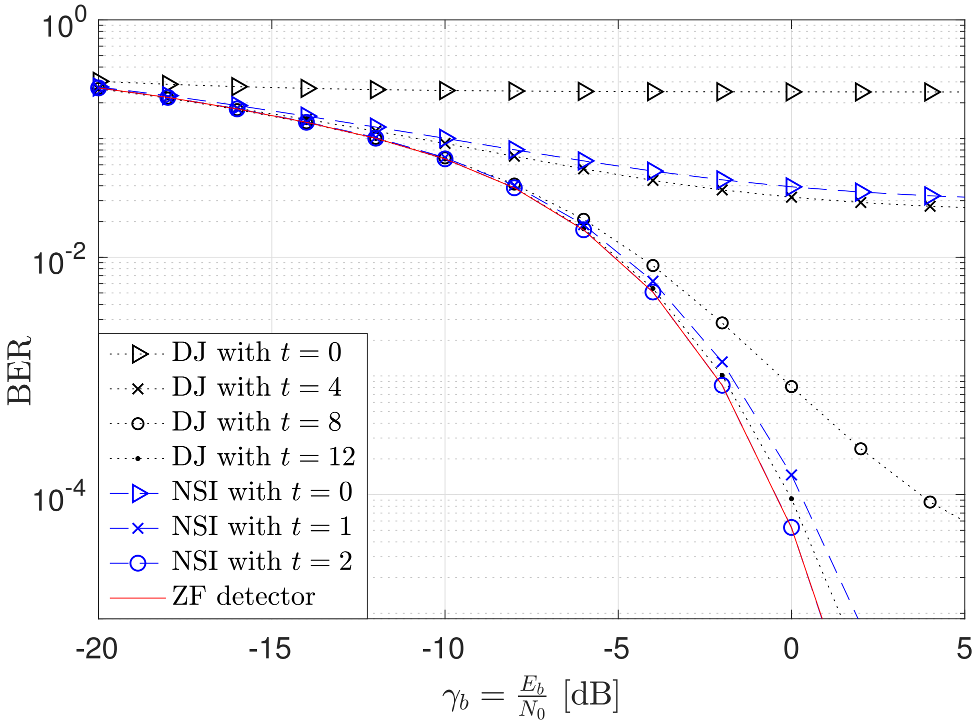


Figure 5.5 – BER versus γ_b of NSI and DJ algorithm for a $N_R \times N_T = 128 \times 32$ spatial multiplexing massive MIMO system employing 64-QAM.

tector and NSI algorithm increases as N_T increases. Thus, for relatively low N_T, NSI algorithm achieves near-optimum performance (valid for spatial multiplexing massive MIMO systems), while for $N_T = N_R$, NSI algorithm does not achieves the near-optimal performance (symmetric spatial multiplexing MIMO systems). However, in this situation, NSI algorithm can be employed as an initial vector solution for algorithms in symmetric spatial multiplexing MIMO systems (QIN *et al.*, 2015; GYAMFI *et al.*, 2015).

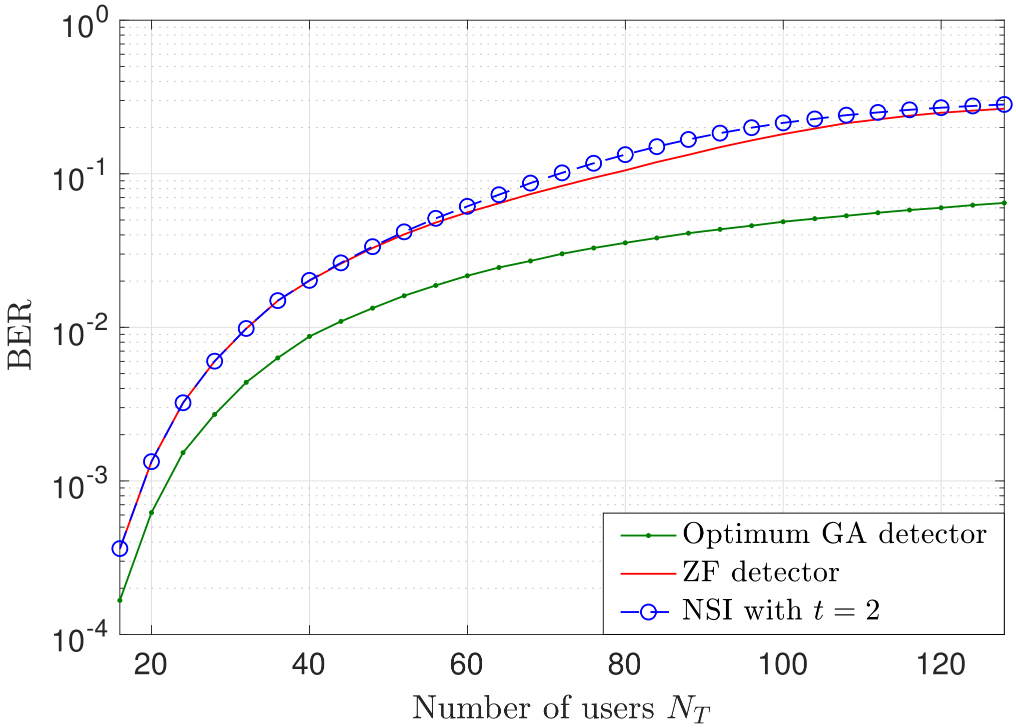


Figure 5.6 – BER as a function of the number of users N_T, employing 64-QAM, for a $\gamma_b = 15\,\text{dB}$ and $N_R = 128$ antennas.

5.5.3 Complexity Analysis

The term "*flop*" is used again to denote multiply-add pair operations in order to analyze the computational complexity of NSI algorithm. From (5.1), each NSI iteration t essentially amounts to two matrix multiplications. Thus, by exploiting the symmetry of $\mathbf{GZ}_t$, NSI algorithm has a complexity of $\frac{2}{3}N_T^3 - \frac{1}{3}N_T^2$ flops by iteration. However, once NSI algorithm with the proposed $\mathbf{Z}_0$ in 2 iterations is able to achieve near-optimum performance for spatial multiplexing massive MIMO systems, the number of flops is given by:

$$C_{\text{NSI}} = \frac{4}{3}N_T^3 - \frac{2}{3}N_T^2. \tag{5.43}$$

Although NSI algorithm presents cubic complexity, that is $O\left(\frac{4}{3}N_T^3\right)$, it has high accuracy, strong numerical stability and contains only matrix additions and multiplications, which is strongly preferred in software/hardware implementation and it is also less complex than the exact matrix inversion obtained through SVD decomposition given by (4.32).

Fig. 5.7 shows the complexity of DJ and NSI algorithms given respectively, by (4.31) and (5.43) as a function of the number of users N_T for different number of iterations t. The ZF detector with exact matrix inversion given by (4.32) is also included as benchmark. In order to ensure near-optimum performance, a large number of iterations ($t \geq 8$) is required by DJ algorithm, as shown in Fig. 5.4 and Fig. 5.5. In contrast, NSI algorithm with the proposed $\mathbf{Z}_0$ requires only 2 iterations. However, from Fig. 5.7 observe that both algorithms have similar computational complexity in terms of the number of flops. In the following, a complexity reduction for NSI algorithm is presented.

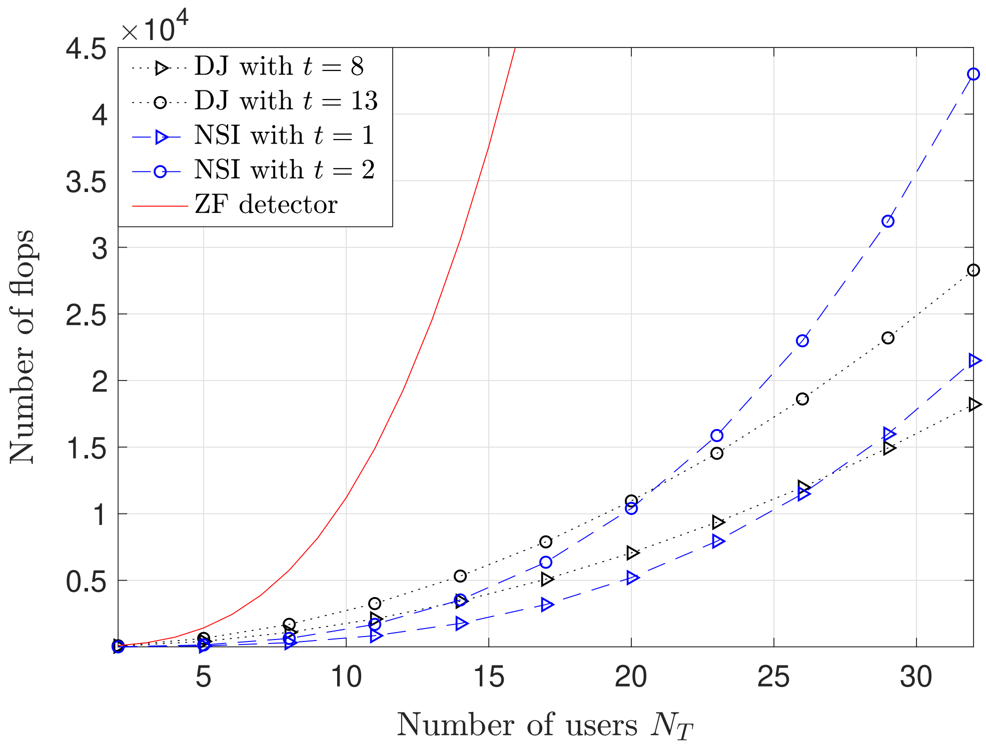


Figure 5.7 – Complexity comparison as a function of the number of users N_T for NSI, DJ and ZF with exact matrix inversion.

5.5.4 Complexity Reduction of NSI Algorithm

In order to reduce the computational complexity of NSI algorithm, the band matrix (BM) concept applied to NSI algorithm is proposed. A BM is a matrix with non-zero entries confined only to a band around the diagonal, that comprises the diagonal and secondary diagonals.

Consider the diagonal dominance of the initial solution $\mathbf{Z}_0$ (RUSEK *et al.*, 2013), where the elements outside the band matrix are set to zero. Thus, the first iteration can be written as:

$$\widehat{\mathbf{Z}}_1 = \overline{\mathbf{Z}}_0 \left(2\mathbf{I}_{N_T} - \tilde{\mathbf{G}}^{(\upsilon)}\overline{\mathbf{Z}}_0\right) \tag{5.44}$$

where $\overline{\mathbf{Z}}_0 = \operatorname{diag}[\mathbf{Z}_0]$ and $\tilde{\mathbf{G}}^{(\upsilon)}$ is the BM of $\mathbf{G}$ given by:

$$\begin{cases} \tilde{g}_{ij}^{(\upsilon)} = g_{ij} & |j-i| \leq \upsilon, \\ \tilde{g}_{ij}^{(\upsilon)} = 0 & |j-i| > \upsilon. \end{cases} \tag{5.45}$$

where υ represents the bandwidth, that means that only υ adjacent diagonals in the band matrix consist of non-zero entries.

Employing $\widehat{\mathbf{Z}}_1$, given by (5.44), as the input for the second iteration,

$$\widehat{\mathbf{Z}}_2 = \widehat{\mathbf{Z}}_1 \left(2\mathbf{I}_{N_T} - \tilde{\mathbf{G}}^{(\upsilon)}\widehat{\mathbf{Z}}_1\right). \tag{5.46}$$

Thus, the number of flops in the first and second iterations are $N_T(2\upsilon+1) - 2\upsilon(\upsilon+1)$ and $N_T^2(2\upsilon+1) - N_T(2\upsilon^2+2\upsilon-1)$, respectively. Then, the total number of flops is given by:

$$C_{\text{NSI,BM}} = N_T^2(2\upsilon+1) - 2N_T\left(\upsilon^2-1\right) - 2\upsilon(\upsilon+1). \tag{5.47}$$

From (5.47), observe that the complexity of NSI algorithm is now quadratic. Thus, through the BM concept, the complexity has been reduced from $O\left(\frac{4}{3}N_T^3\right)$ to $O\left(N_T^2(2\upsilon+1)\right)$.

The bandwidth parameter υ influences on the accuracy and complexity of the NSI algorithm. As υ increases, the number of flops also increases monotonically. Fig. 5.8 shows the effects of υ in the BER for a $N_R \times N_T = 128 \times 32$ massive MIMO system. Notice that as υ increases, the BER improves, but the computational complexity increases. However, for a relatively small $\upsilon = 4$, similar BER performance is achieved in relation to the near-optimum ZF performance.

Despite that there is a compromise between performance and complexity when employing NSI algorithm with BM, for relatively small υ the computational complexity is inferior to NSI algorithm without BM. Thus, Fig. 5.9 shows the complexity comparison as a function of the number of users N_T for NSI algorithm with and without the application of BM, where 2 iterations are considered for both NSI algorithms. In the case of NSI algorithm with BM, a $\upsilon = 4$ is considered. Furthermore, both the complexity of ZF detector with exact matrix inversion and DJ algorithm with $t = 8$ iterations are included as benchmark. From this figure, observe that the complexity reduction employing NSI algorithm with BM is effective for all N_T, which means that the application of BM in the NSI algorithm is effective for large N_T.

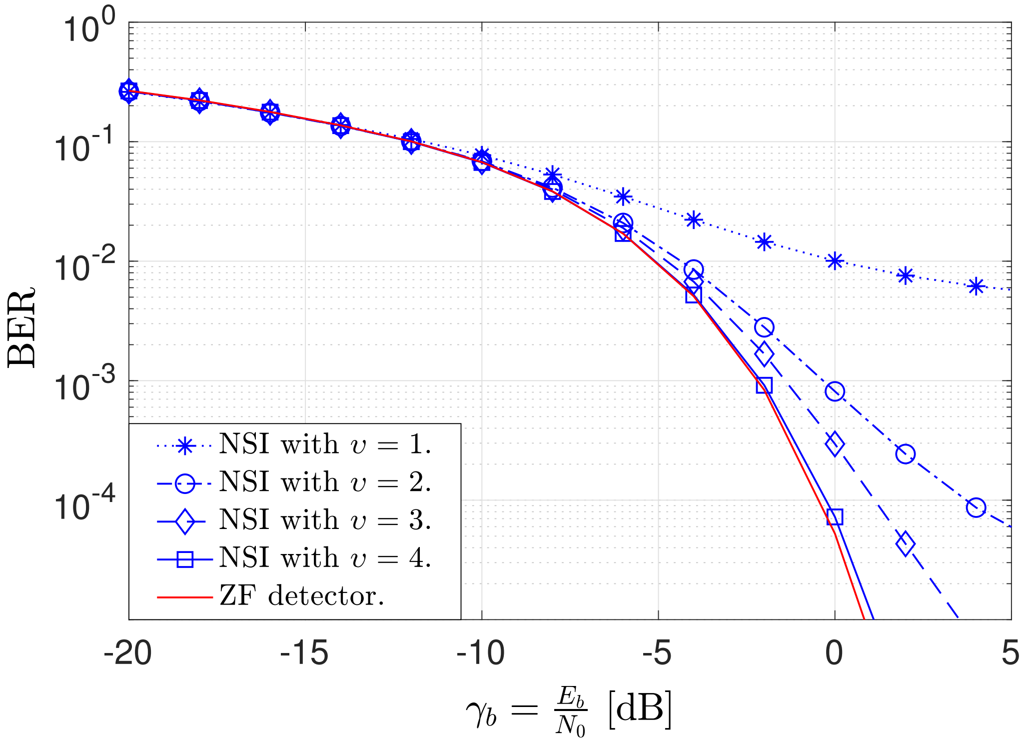


Figure 5.8 – BER versus γ_b for different υ of a $N_R \times N_T = 128 \times 32$ spatial multiplexing massive MIMO system employing 64-QAM.

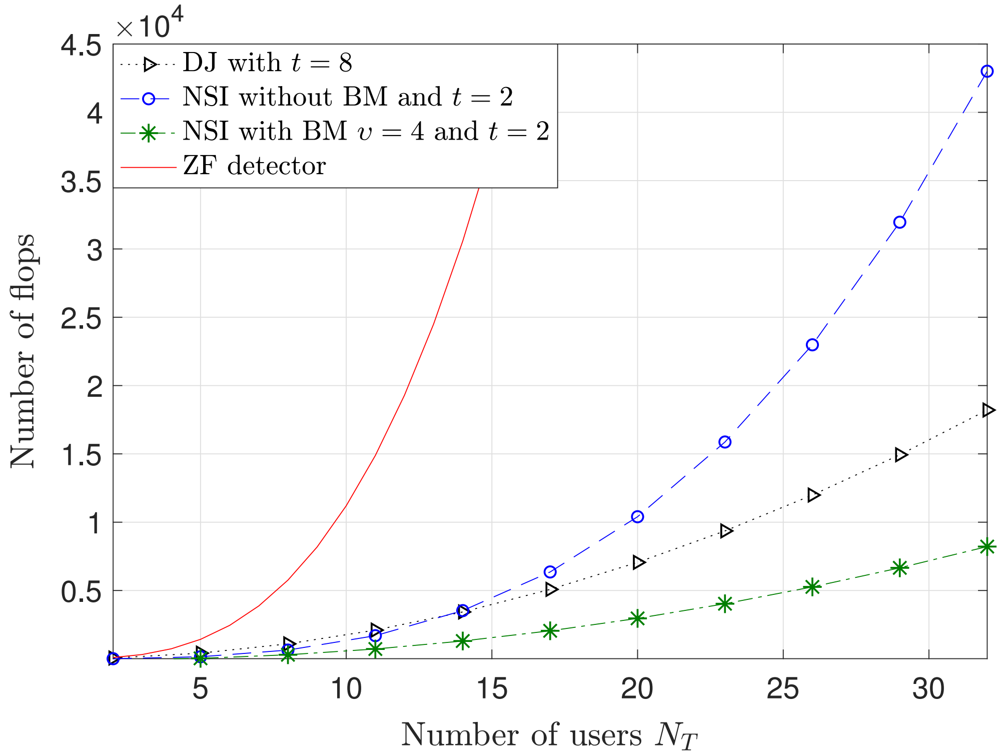


Figure 5.9 – Complexity as a function of the number of users N_T for NSI with and without BM, DJ and ZF with exact matrix inversion.

5.6 Chapter Conclusions

In this chapter, the NSI algorithm for the detection of spatial multiplexing massive MIMO systems has been proposed, avoiding the complex exact matrix inversion of the ZF detector. Furthermore, in order to speed up the convergence of NSI algorithm, a novel initial matrix inversion solution based on Tchebychev polynomial has been proposed. The numerical results reveal that NSI algorithm with the proposed initial matrix inversion solution outperforms Damped Jacobi (DJ) algorithm. It achieves the near-optimum ZF performance in just two iterations, keeping the complexity low. Finally, in order to reduce additionally the complexity, the band matrix concept has been employed, which reduces the complexity from $O\left(\frac{4}{3}N_T^3\right)$ to $O\left(N_T^2\left(2v+1\right)\right)$, which is effective when the number of users N_T is large.

6 Achieving Optimum Performance with ZF Detector by Using a Novel Precoding Approach in Symmetric Spatial Multiplexing MIMO Systems

6.1 Introduction

Despite ZF detector is attractive in spatial multiplexing massive MIMO systems due to their small complexity in comparison to the optimum ML detector. When the number of users N_T is equal to the number of BS antennas N_R, a high spectral efficiency is achieved in contrast to massive MIMO systems where $N_T \ll N_R$. However, the performance of ZF detector is severely degraded when $N_R = N_T$. In this case, the condition number of $\mathbf{H}^H\mathbf{H}$ is high (see Fig. 2.8), and consequently, the channel matrix $\mathbf{H}$ is highly non-orthogonal. Furthermore, the degraded performance of ZF detector when $N_R = N_T$ can also be explained by its diversity order, which is given, on average, by $N_R - N_T + 1$ where $N_R \geq N_T$ (MINANGO, 2014). Thus, when $N_R = N_T$, the ZF detector attains a mean diversity order of 1, which is far away from the full diversity order N_R of the optimum ML detector. Therefore, the key challenge in practical realizing of systems with $N_R = N_T$, which are known as symmetric spatial multiplexing MIMO, is the detection complexity at the receiver if full diversity order is desirable.

In order to guarantee full diversity and optimum performance, ML detector should be employed, whose complexity is $O(M^{N_T})$, where M is the modulation order. Thus making it impractical to perform detection in symmetric spatial multiplexing MIMO systems with large N_T (ANDREWS *et al.*, 2007). Though ZF detector achieves a poor performance compared to the optimum detector in symmetric spatial multiplexing MIMO systems. This is no longer true if the symmetric channel matrix $\mathbf{H}$ is orthogonal (MAURER *et al.*, 2007; ROGER *et al.*, 2008; WUBBEN *et al.*, 2011). Thus, several studies have focused on the implementation of ZF detector preceded by lattice reduction (LR) (WUBBEN *et al.*, 2011; JING *et al.*, 2016; LIU *et al.*, 2017; IZADINASAB *et al.*, 2017), whose purpose is to reduce the condition number. For a condition number of one, symmetric channel matrices are orthogonal and therefore the ZF detector has optimum performance, i.e., equivalent to ML detector (WUBBEN *et al.*, 2011).

Although the ZF detector with lattice reduction (ZF-LR) has the potential to achieve full diversity, there is a SNR gap from the ML performance. This is due to LR can significantly reduce the condition number of the symmetric channel matrix but never to the unity. Furthermore, as the symmetric channel matrix has large dimension, the probability that the channel matrix has large condition number is high (WUBBEN *et al.*, 2011). Consequently the SNR gap is also large. To summarize this point, the major problem with the ZF detector is its poor performance for symmetric channel matrices with large condition number or ill-conditioned symmetric channel matrices.

Thus, in this chapter, in order to corroborate the above affirmations, firstly the effects of the symmetric channel matrix condition number on the ZF detector performance are studied. Specifically, simulations show that the poor ZF performance is heavily linked to symmetric channel matrices with large condition number. Motivated by this insight, a novel precoding approach for symmetric channel matrix orthogonalization is proposed. In general form, the proposed approach first find out the precoder matrix at the transmitter side, which transforms the symmetric channel matrix with any condition number into an orthogonal matrix. At the receiver side, the received signal can then be applied to a ZF detector achieving optimum performance. Simulation results show the validity of the proposed approach, where the price to be paid to achieve the optimum performance is a small increase in the transmitted average power.

Furthermore, the proposed precoding approach for symmetric spatial multiplexing MIMO systems consider perfect channel state information (CSI) available at both transmitter and receiver side. The CSI could be sent by the receiver to the transmitter, when there are separate frequency bands for uplink and downlink transmission, frequency division duplex (FDD), or the transmitter could estimate the channel, if it is reciprocal, like in a time division duplexing (TDD) system, by receiving pilot signals from the receiver (ZHOU *et al.*, 2011; THOMAS; VOOK, 2014).

The remainder of this chapter is organized as follows. Section 6.2 presents the system model. In Section 6.3, the effects of the symmetric channel matrix condition number on the ZF performance are discussed. Section 6.4 details the proposed precoding approach, while in Section 6.5, the simulation results are presented. Finally, the conclusions are drawn in Section 6.6.

6.2 System Model

A symmetric spatial multiplexing MIMO system with N_T transmit and N_R receive antennas, where $N_R = N_T$ is considered in this chapter. Thus, N_T symbols are simultaneously transmitted over N_T antennas and the corresponding received signal vector $\mathbf{y}$ at the N_R antennas is given by (2.13).

Furthermore, the total average transmit power $\bar{P}$ is equal to the sum of the average powers of all transmit antennas N_T, that is:

$$\bar{P} = \sum_{i=1}^{N_T} \bar{P}_o = N_T \bar{P}_o, \tag{6.1}$$

where $\bar{P}_o$ is the transmit average power by each antenna.

Besides, both ML and ZF detectors, which were described in chapter 2, are considered in this chapter.

6.3 Effects of the Symmetric Channel Matrix Condition Number on the ZF Detector Performance

Applying the ZF equalization matrix given by (2.48) on the received signal vector $\mathbf{y}$, given by (2.13), the ZF equalized vector $\mathbf{r}$, before the quantization process, can be written as:

$$\mathbf{r} = \mathbf{H}^{\dagger}\mathbf{y} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{y} = \mathbf{x} + \mathbf{n}', \tag{6.2}$$

which is equal to the transmitted signal vector $\mathbf{x}$ corrupted by the modified noise $\mathbf{n}' = \mathbf{H}^{\dagger}\mathbf{n}$, which has correlation matrix given by (CHOCKALINGAM; RAJAN, 2014):

$$\mathbf{R}_{\mathbf{n}'} = \sigma_n^2 \left(\mathbf{H}^H\mathbf{H}\right)^{-1}. \tag{6.3}$$

From (6.2), notice that the interference caused by the symmetric channel matrix $\mathbf{H}$ has been completely removed (forced to zero). However, the variance of $\mathbf{n}'$ is equal or greater than the original variance of $\mathbf{n}$, i.e., there is an increase in noise power (ROGER *et al.*, 2008; CHOCKALINGAM; RAJAN, 2014).

As ZF equalization does not produce loss of information, the ML detector given by (2.40) can use the ZF equalized vector $\mathbf{r}$. Thus, the ML detector based on $\mathbf{r}$ becomes:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\hat{\mathbf{x}} \in M^{N_T}} \left\{(\mathbf{r} - \hat{\mathbf{x}})^H \mathbf{H}^H\mathbf{H}\,(\mathbf{r} - \hat{\mathbf{x}})\right\}. \tag{6.4}$$

which is proved in appendix B.

Equation (6.4) may be interpreted as “ML detection after ZF equalization”, different of “direct ML detection” given by (2.40). Note, however, that (2.40) and (6.4) are strictly equivalent.

On the other hand, a channel matrix $\mathbf{H}$ is orthogonal, if the following condition is satisfied (GOLUB; LOAN, 1996):

$$\mathbf{H}^H\mathbf{H} = \mathbf{I}_{N_T}. \tag{6.5}$$

Hence, for an orthogonal symmetric channel matrix $\mathbf{H}$, the noise correlation matrix given by (6.3) becomes:

$$\mathbf{R}_{\mathbf{n}'} = \sigma_n^2, \tag{6.6}$$

that is the noise power is not increased.

As a consequence, the components of the noise vector $\mathbf{n}' = \mathbf{H}^{\dagger}\mathbf{n} = \mathbf{n}$ are uncorrelated, and the ML detector in (6.4) can be simplified to:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\hat{\mathbf{x}} \in M^{N_T}}{\arg\min} \left\| \mathbf{r} - \hat{\mathbf{x}} \right\|^2 . \tag{6.7}$$

From (6.7), since each component of $\hat{\mathbf{x}}$ affects only the corresponding component of $\mathbf{r} = \mathbf{x} + \mathbf{n}'$, and since $\mathbf{n}'$ is uncorrelated, the ML detector given by (6.7) can be decoupled into a set of symbol-by-symbol optimizations as:

$$\hat{\mathbf{x}}_{\mathrm{ML}}[i] = \underset{\hat{\mathbf{x}}[i] \in M}{\arg\min} \left\| \mathbf{r}[i] - \hat{\mathbf{x}}[i] \right\|^2 , \quad i = 1, \cdots, N_T, \tag{6.8}$$

which can be solved using a component-wise quantization with respect to the symbol constellation of order M. Thus, (6.8) can be rewritten as:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = Q\left\{\mathbf{r}\right\}, \tag{6.9}$$

which is equivalent to the linear ZF detector solution given by (2.49).

In general, the above analysis shows that when the symmetric channel matrix $\mathbf{H}$ is orthogonal, ZF detector is equivalent to ML detector, achieving optimum performance and full diversity. In order to corroborate this statement, it is shown through simulation that the bad performance of ZF detector in comparison to ML detector is mainly caused by the occurrence of ill-conditioned symmetric channel matrices.

Fig. 6.1 shows the BER performance of the ML and ZF detectors as a function of the condition number of $\mathbf{H}$, given by $\kappa(\mathbf{H}^H\mathbf{H}) = \frac{\lambda_{\max}(\mathbf{H}^H\mathbf{H})}{\lambda_{\min}(\mathbf{H}^H\mathbf{H})} \geq 1$, where $\lambda_{\max}(\mathbf{H}^H\mathbf{H})$ and $\lambda_{\min}(\mathbf{H}^H\mathbf{H})$ denote the largest and the smallest eigenvalues of $\mathbf{H}^H\mathbf{H}$, respectively. In this simulation, a symmetric spatial multiplexing MIMO system with $N_R = N_T = 5$, 4-QAM modulation, and a γ_b of -9 dB is considered. Observe that for $\kappa(\mathbf{H}^H\mathbf{H}) < 3$, ZF detector has approximately the same performance of ML detector. However, as $\kappa(\mathbf{H}^H\mathbf{H}) > 3$, the performance of both detectors separates, emphasizing that ML detector is obviously the one that has better performance.

From Fig. 6.1, notice that the bad performance of ZF detector depends on the probability of channel matrices with large $\kappa(\mathbf{H}^H\mathbf{H})$. The simulated probability density function (PDF) of $\kappa(\mathbf{H}^H\mathbf{H})$ is presented in Fig. 6.2a, whereas Fig. 6.2b shows the cumulative distribution function (CDF) of $\kappa(\mathbf{H}^H\mathbf{H})$. From Fig. 6.2b, observe that the

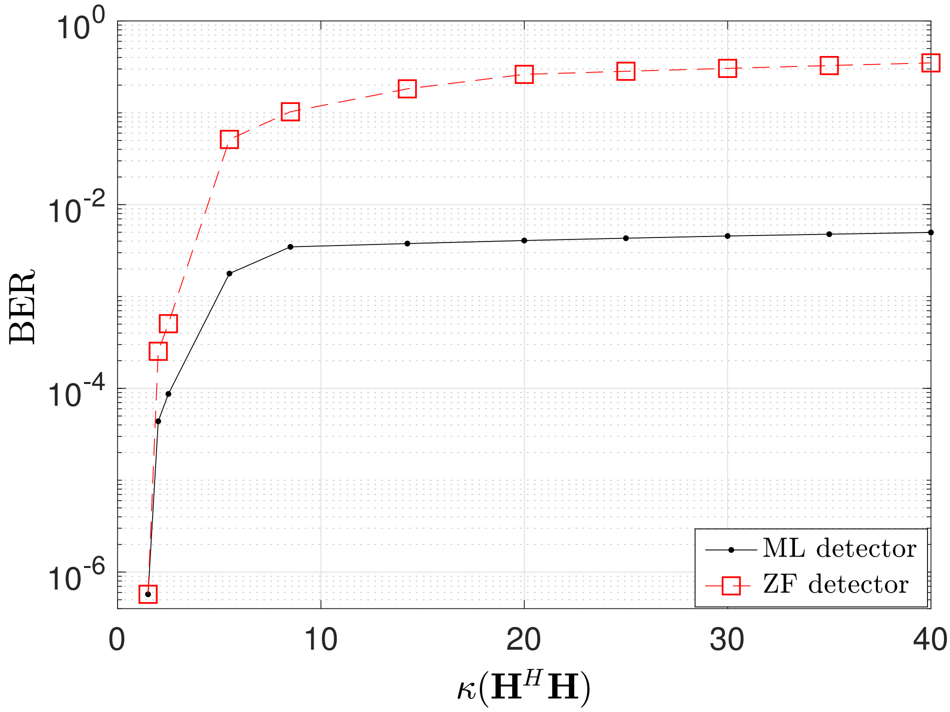


Figure 6.1 – BER as a function of the condition number $\kappa(\mathbf{H}^H\mathbf{H})$ for random symmetric channel matrices $\mathbf{H}$ for a spatial multiplexing MIMO system with $N_R = N_T = 5$, 4-QAM modulation, and $\gamma_b = -9$ dB.

probability that $\kappa(\mathbf{H}^H\mathbf{H})$ exceeds 10, 15, and 20 is 50%, 25%, and 15%, respectively. This suggests that ill-conditioned channel matrices occur frequently enough to cause a significant degradation on the BER performance of ZF detectors. Furthermore, increasing both N_T and N_R, the PDF and CDF are basically spreaded (TULINO; VERDú, 2004). Thus, by increasing N_T and N_R with $N_R = N_T$, channel matrices become even worse ill-conditioned.

6.4 Proposed Precoding Approach

As already stated in the previous section, in symmetric spatial multiplexing systems, linear ZF detector is optimum if the channel matrix $\mathbf{H}$ is orthogonal, but has poor performance otherwise. It is therefore natural to apply the ZF detector not directly to $\mathbf{H}$, but rather to a transformed symmetric spatial multiplexing MIMO system with channel matrix $\breve{\mathbf{H}}$, which is orthogonal. Thus, according to Fig. 6.3 the system model (2.13) can be rewritten as:

$$\begin{aligned} \mathbf{y} &= \breve{\mathbf{H}}\mathbf{x} + \mathbf{n}, \\ &= \mathbf{H}\mathbf{V}\mathbf{x} + \mathbf{n}, \end{aligned} \tag{6.10}$$

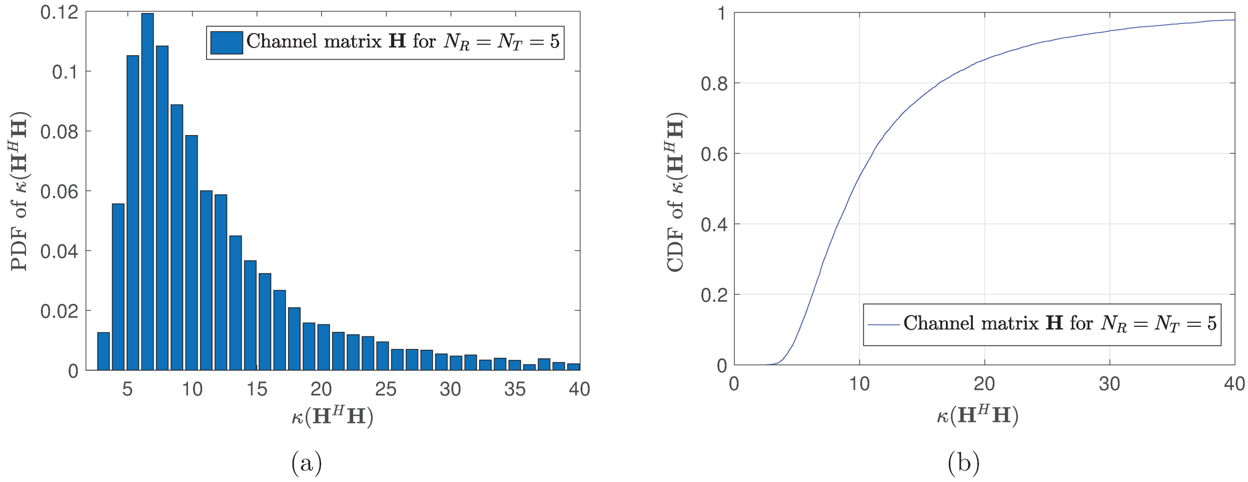


Figure 6.2 – (a) Probability density function of $\kappa(\mathbf{H}^H\mathbf{H})$. (b) Cumulative distribution function of $\kappa(\mathbf{H}^H\mathbf{H})$ for $N_R = N_T = 5$.

where $\mathbf{V} \in \mathbb{C}^{N_T \times N_T}$ represents the precoder matrix and $\mathbf{Vx}$ is the new transmitted symbol. In order to ZF detector to be optimum, it is necessary that $\mathbf{HV}$ be an orthogonal matrix, that is:

$$(\mathbf{HV})^H (\mathbf{HV}) = \mathbf{I}_{N_T}. \tag{6.11}$$

Thus, equalizing (6.10) with the pseudo-inverse $(\mathbf{HV})^\dagger$ leads to:

$$\check{\mathbf{r}} = (\mathbf{HV})^\dagger \mathbf{y} = \mathbf{x} + (\mathbf{HV})^\dagger \mathbf{n}. \tag{6.12}$$

Since $\mathbf{HV}$ is orthogonal, the noise power is not increased. Furthermore the pseudo-inverse of $\mathbf{HV}$ can be computed easily once $(\mathbf{HV})^\dagger = (\mathbf{HV})^H$. Finally, the equalized vector $\check{\mathbf{r}}$ given by (6.12) is quantized as:

$$\hat{\mathbf{x}} = Q\{\check{\mathbf{r}}\}, \tag{6.13}$$

which achieves the optimum performance.

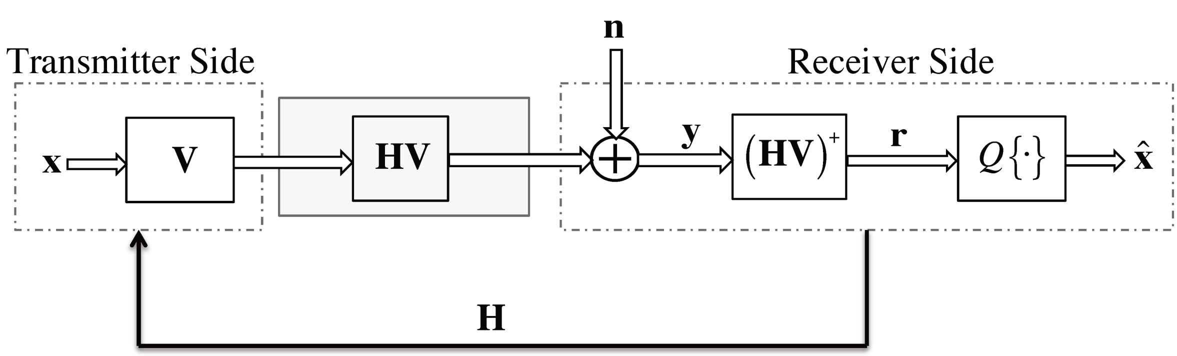


Figure 6.3 – Block diagram of the precoding approach.

6.4.1 Determination of the Precoder Matrix

The main idea of the proposed approach is to find out the precoder matrix $\mathbf{V}$ that transforms the symmetric channel matrix $\mathbf{H}$ into an orthogonal channel matrix $\mathbf{HV}$.

The symmetric channel matrix $\mathbf{H}$ is assumed to be perfectly known at the transmitter side through the feedback link.

Solving (6.11), it can be shown that:

$$\mathbf{V}^H\mathbf{V} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}, \tag{6.14}$$

where it was used that $(\mathbf{HV})^H = \mathbf{V}^H\mathbf{H}^H$. From (6.14) and supposing that $\mathbf{V}$ is a symmetric matrix, i.e., $\mathbf{V} = \mathbf{V}^H$, the precoder matrix is given by:

$$\mathbf{V} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1/2}. \tag{6.15}$$

Note that computing $\mathbf{V}$ from (6.15) has cubic complexity with respect to the number of transmit antennas, $O\left(N_T^3\right)$, due to the inverse of $\mathbf{H}^H\mathbf{H}$.

6.4.2 Total Transmitted Power

The total average transmit power using the precoder matrix $\mathbf{V}$ is given by:

$$\bar{P} = \left\|\mathbf{V}\right\|_F^2 \sum_{i=1}^{N_T} \bar{P}_o = \left\|\mathbf{V}\right\|_F^2 N_T \bar{P}_o. \tag{6.16}$$

By comparing the total average transmit power, $\bar{P}$, of the symmetric spatial multiplexing MIMO system given by (6.1) with the total average transmitted power employing the precoding matrix given by (6.16), it is concluded that the precoding approach increases the total average transmit power by $\left\|\mathbf{V}\right\|_F^2$. This means, that the proposed precoding approach increments the transmit power in order to get the optimum performance.

However, since the wireless channels experienced by each transmit antenna are different. Consequently. It is desirable to allocate different power (FANG *et al.*, 2011) to the transmit antennas based on their respective channel quality in order to achieve optimum performance.

6.5 Numerical Results

In this section, the simulation results of the BER of the proposed precoding approach for symmetric spatial multiplexing MIMO systems is presented.

Fig. 6.4 shows the BER performance against γ_b of the precoding approach in comparison to ML and ZF detectors for a symmetric MIMO system with $N_R = N_T = 20$

antennas employing 64-QAM modulation. The optimum ML performance is achieved by the hypothetical GA detector, which was described earlier in sub-section 2.5.3.1. From this figure, observe that the precoding approach achieves the optimum performance with the full diversity order of $N_R = 20$, that is, the same performance of the ML detector.

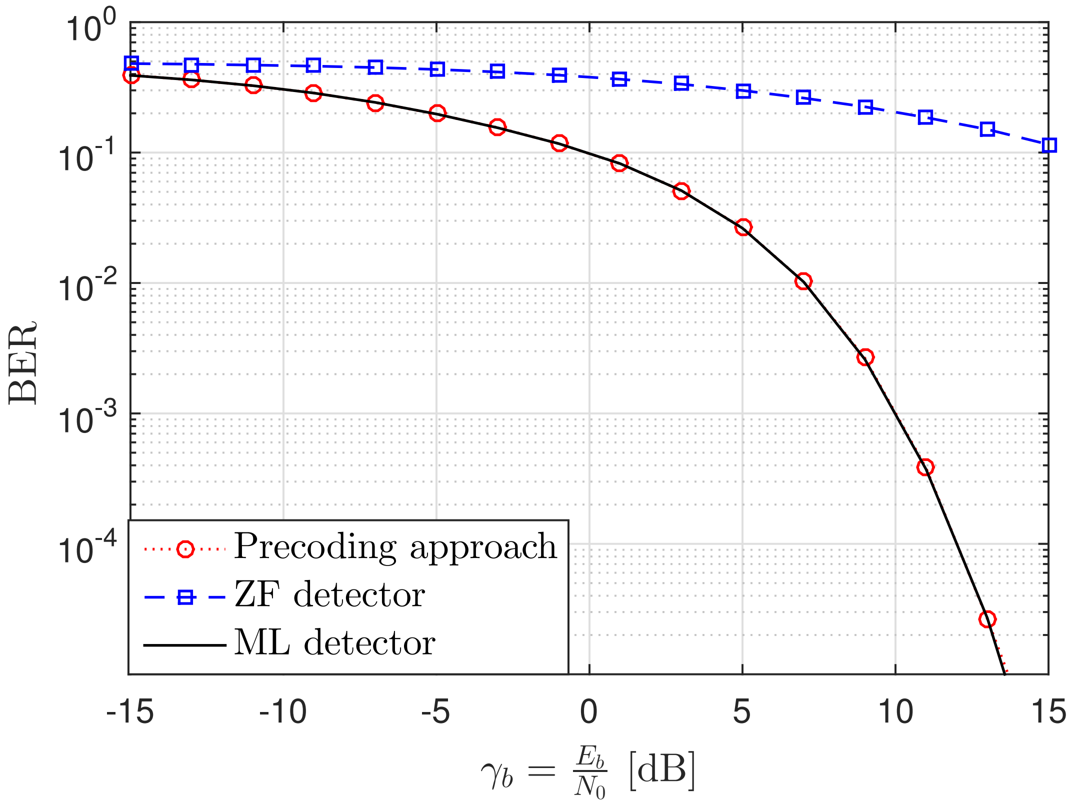


Figure 6.4 – BER as a function of γ_b for a symmetric spatial multiplexing MIMO with $N_R = N_T = 20$ antennas and 64-QAM employing the precoding approach, ML and ZF detector.

Fig. 6.5 and Fig. 6.6 present the BER as a function of γ_b for the symmetric spatial multiplexing MIMO systems with $N_R = N_T = 50$ and $N_R = N_T = 100$ antennas for 16-QAM and 4-QAM modulations, respectively. In both symmetric MIMO systems, the precoding approach, ML and ZF detector are employed. In both figures, similar conclusions to Fig. 6.4 are obtained, where the precoding approach outperforms ZF detector and achieves the optimum performance with full diversity order of $N_R = 50$ and $N_R = 100$, respectively.

Notice from Fig. 6.4, Fig. 6.5 and Fig. 6.6 that the precoding approach achieves the optimum performance independently of the modulation order and the number of transmit N_T and receive N_R antennas. As stated earlier, the price to be paid is a small increase in the average transmit power due to the precoder matrix $\mathbf{V}$. Thus, more transmit power in order to modify the ill-conditioned channel matrix into an orthogonal matrix. Fig. 6.7 shows that the total average transmit power of the precoding approach is about

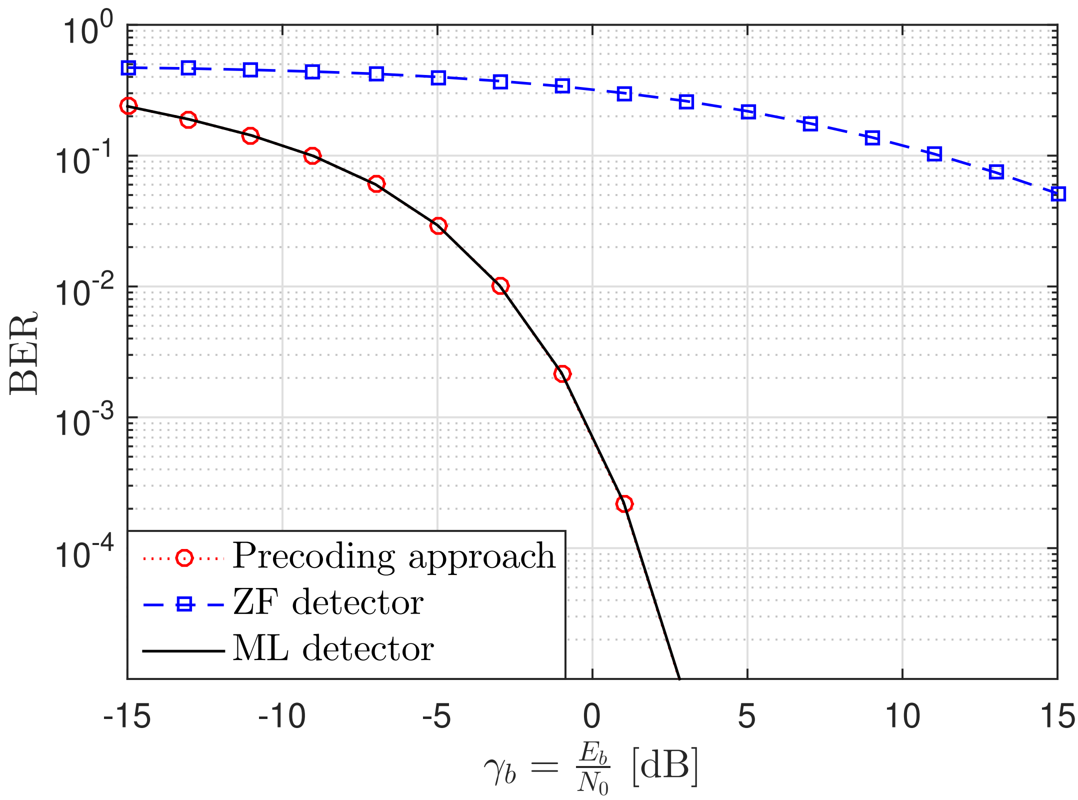


Figure 6.5 – BER as a function of the γ_b for a symmetric spatial multiplexing MIMO with $N_R = N_T = 50$ antennas and 16-QAM employing the precoding approach, ML and ZF detector.

1.5 times (approximately 2 dB) the total average transmit power of a symmetric spatial multiplexing MIMO system without precoding. Note that the total average transmit power increases linearly with the number of transmit antennas N_T.

6.6 Chapter Conclusions

In this chapter, a novel precoding approach was proposed, which makes ZF detector to have the same optimum performance of ML detector in symmetric spatial multiplexing MIMO systems. At first, it was shown that if a symmetric channel matrix is orthogonal, ML and ZF detectors have same performance. Otherwise, ZF detector has poor performance that is justified heavily by the occurrence of ill-conditioned symmetric channel matrices. Based on the above, the precoding approach orthogonalizes the symmetric equivalent channel matrix in order to apply, at the receiver, the ZF detection that achieves the same performance of ML detector, but with much lower complexity. As the equivalent channel matrix is orthogonal, its inverse is easily calculate by taking its Hermitian $(\mathbf{HV})^{-1} = (\mathbf{HV})^{H}$, whose complexity is $O(N_T^2)$. The drawback is a small increase in the total average transmit power. Simulation results have showed that the

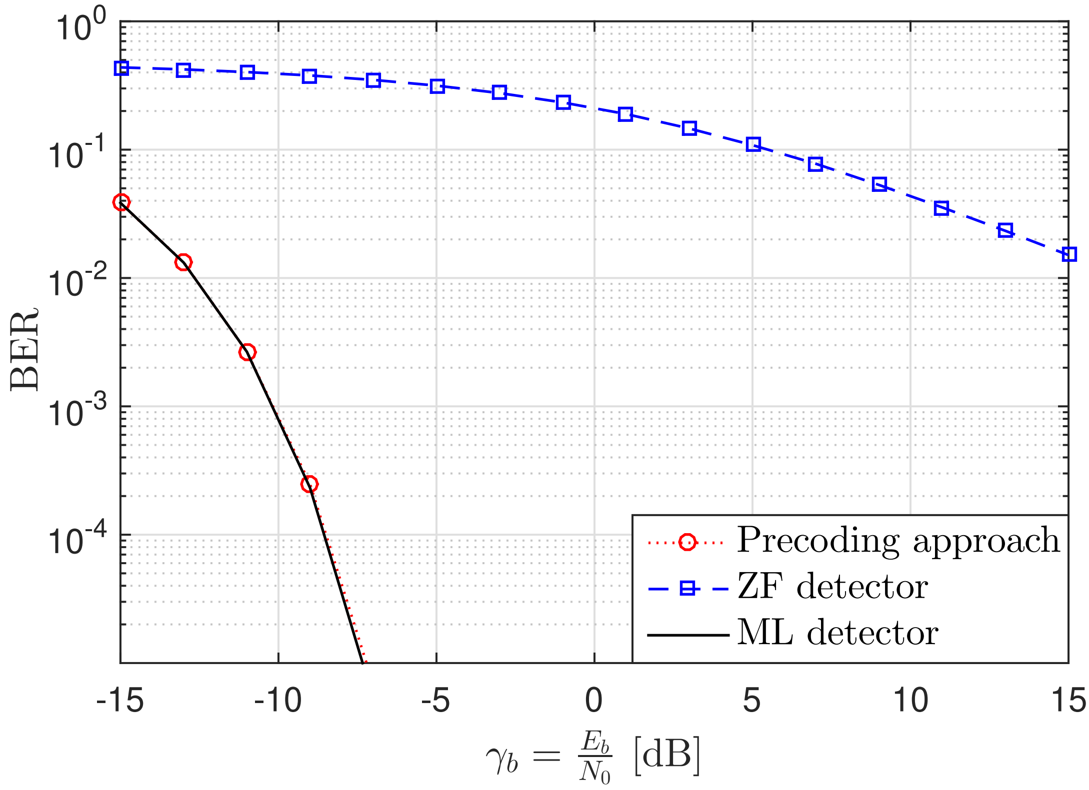


Figure 6.6 – BER as a function of the γ_b for a symmetric spatial multiplexing MIMO with $N_T = N_R = 100$ antennas and 4-QAM employing the precoding approach, ML and ZF detector.

precoding approach achieves optimum performance. Remark that the precoding approach allow low-complexity implementation of symmetric MIMO systems with very high spectral efficiency.

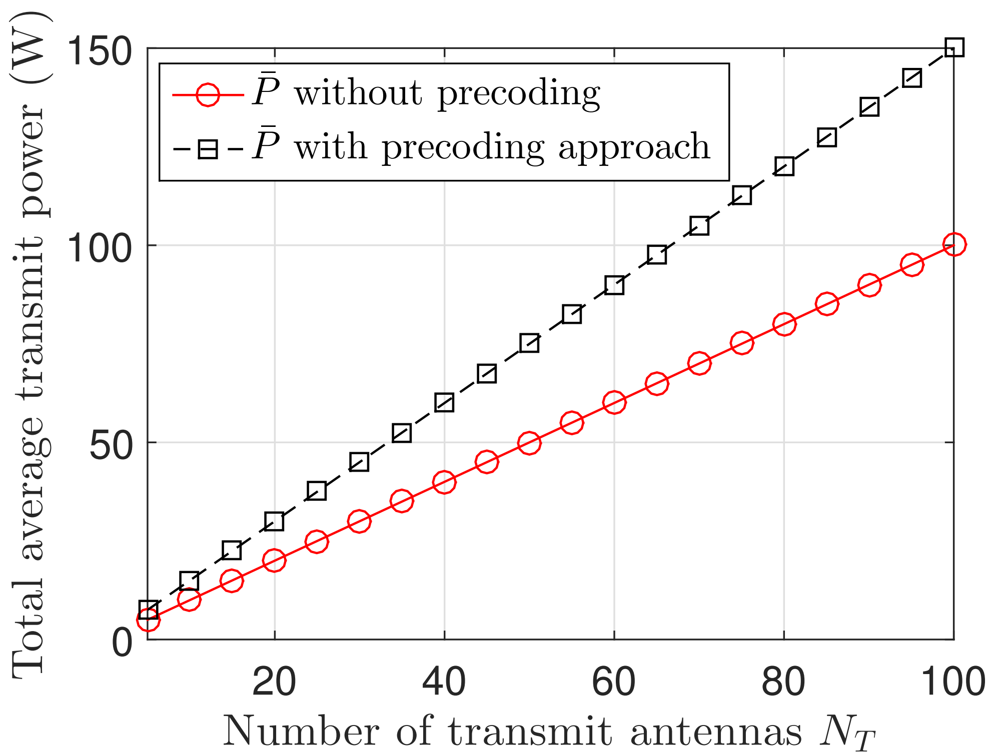


Figure 6.7 – Total average transmit power as a function of the number of transmit antennas N_T of symmetric spatial multiplexing MIMO system with and without precoding.

7 Low-Complexity Symbol-Flipping Algorithm Detector for Symmetric Spatial Multiplexing MIMO Systems with Large Number of Antennas

7.1 Introduction

In Chapter 6 the optimum performance was achieved with a ZF detector by using a novel precoding approach in symmetric spatial multiplexing MIMO systems. The principal goal of the precoding approach is to transforms the channel matrix into an orthogonal matrix, such that a ZF detector can achieve optimum performance. Thus, during each transmission, the transmitter needs to know a priori the channel matrix in order to compute the precoder matrix for the channel orthogonalization. In general, a transmitter has no knowledge of the current channel matrix. Therefore, current channel matrix must be estimated at the receiver side and then, fedback to the transmitter side. The main drawback of this process is that additional resource is required for transmitting the feedback information, where the amount of feedback information increases with the number of transmit and receive antennas (THOMAS; VOOK, 2014). Therefore, the overhead problem together with the feedback delay become critical when symmetric MIMO systems with large number of antennas are employed. Therefore, alternative detection approaches with low-complexity and near-optimum performance should be developed.

On the other hand, ML detector has high complexity because it searches for the global optimum solution over all possible transmitted signal vectors, which depends on the modulation order and on the number of transmit antennas employed in the MIMO system. In this chapter, by focusing in the spatial search, the probability of finding the global optimum solution from local sub-optimum solutions using symbol-flipping (SF) procedures is verified to increase with the number of transmit and receive antennas. As a consequence, the SF procedure present very close performance to the optimum detector in spatial multiplexing MIMO systems with hundreds or more transmit and receive antennas. Specifically, the SF procedure, which is a heuristic local search (LS) algorithm for solving computationally hard optimization problems (BLUM; ROLI, 2003; AARTS; LENSTRA, 1997), allows to move among different local sub-optimum solutions in the spatial search of limited possible transmitted signal vectors until a solution deemed optimum is found.

However, for symmetric spatial multiplexing MIMO systems with tens to hundreds of antennas, SF procedure does not achieve the optimum performance, once generally it is trapped in local sub-optimum solutions, which can be far away from the global optimum solution. In this situation, some SF procedures with different initial solution vectors used as starting-point is considered. The set of initial solution vectors allows to escape from local sub-optimum solutions reaching to the solution deemed as optimum and consequently approaching to the ML performance.

In summary, the main contribution of this chapter is the proposition of a low-complexity detector algorithm based on SF procedure for symmetric spatial multiplexing MIMO systems in order to achieve very close performance to ML detector. The algorithm can be easily extended for non-symmetric spatial multiplexing MIMO systems.

The remainder of this chapter is organized as follows. Section 7.2 presents the system model. In Section 7.3, the proposed detector algorithm is described. The performance simulation together with the complexity results are presented in Section 7.4. Finally, the conclusions are drawn in Section 7.5.

7.2 System Model

In this chapter, similar to Chapter 6, a symmetric spatial multiplexing MIMO system with N_T transmit and N_R receive antennas is considered, where N_T independent symbols are simultaneously transmitted over N_T antennas. Thus, the corresponding $N_R \times 1$ complex received signal vector $\mathbf{y}$ at the N_R antennas is given by (2.13)

7.3 Proposed Detector Algorithm

7.3.1 Symbol Flipping Procedure

Symbol flipping (SF) procedure is based on an iterative exploration of local sub-optimum solutions in order to get to the global optimum solution. The SF procedure that is applied to an initial solution is defined by the set of symbols of a constellation of order M.

7.3.1.1 Initial solution vector

The SF procedure starts from an initial solution vector that iteratively moves to neighbor sub-optimum solutions. Thus, let k be the index of $\hat{\mathbf{x}}^{(k)}$ that denotes the iteration number. The initial solution vector $\hat{\mathbf{x}}^{(0)}$ that starts the SF procedure can be obtained by:

$$\hat{\mathbf{x}}^{(0)} = Q\left\{\mathbf{B}\mathbf{y}\right\}, \tag{7.1}$$

where $\mathbf{B}$ is the initial solution matrix filter, which can be a matched-filter (MF), a zero-forcing (ZF) filter or an identity matrix for a random initial solution vector. The operation $Q\left\{\mathbf{B}\mathbf{y}\right\}$ denotes the quantization of the vector $\mathbf{B}\mathbf{y}$ to the set of symbols of a constellation of order M.

For MF, $\mathbf{B} = \mathbf{H}^H$ and hence $\hat{\mathbf{x}}^{(0)} = Q\left\{\mathbf{H}^H\mathbf{y}\right\}$, which is used as the initial solution vector of the SF procedure. The reason for this choice is that unlike MF detector which minimizes noise, ZF detector mitigates the inter-antenna interference to zero but fails to treat noise. In fact, this mitigation comes at the cost of noise enhancement, which in the low SNR regime leads to a performance inferior in comparison to MF detector. For a numerical demonstration of this fact, Fig. 7.1 shows the BER as a function of the number of transmit and receive antennas $N_T = N_R$, for $\gamma_b = -15$ dB, and as a function of γ_b, for $N_T = N_R = 100$ antennas, respectively, for MF and ZF detector employing 4-QAM modulation. In both figures, ML detector is also shown as an ideal objective. Furthermore, since simulating the performance of ML detector for large number of antennas involves prohibitively high complexity, the optimum ML performance is achieved by the GA detector introduced in sub section 2.5.3.1.3.

From Fig. 7.1a notice that MF detector outperforms ZF detector for any number of transmit antennas. Furthermore, from Fig. 7.1b, MF detector presents better performance for low values of γ_b. Thereby, in these cases, the best strategy is to use MF.

Another reason to choose MF detector is that computing $\mathbf{B}$ for MF involves only the conjugate transposition of $\mathbf{H}$, whose complexity is quadratic, while computing $\mathbf{B}$ for ZF involves the matrix inversion of $\mathbf{H}$, whose complexity is cubic. Therefore, $\hat{\mathbf{x}}^{(0)}$ obtained with the MF detector is a low-complexity initial solution vector that can be used in the SF procedure.

7.3.1.2 Symbol Flipping

From (2.40), let

$$\Phi[\hat{\mathbf{x}}] = \hat{\mathbf{x}}^H\mathbf{G}\hat{\mathbf{x}} - 2\Re\{\hat{\mathbf{x}}^H\tilde{\mathbf{y}}\} \tag{7.2}$$

be the ML cost function of vector $\hat{\mathbf{x}}$, where $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ represents the Gram matrix and $\tilde{\mathbf{y}} = \mathbf{H}^H\mathbf{y}$. Then, once $\hat{\mathbf{x}}^{(0)}$ is obtained, the SF procedure is applied from the first to the last element of vector $\hat{\mathbf{x}}^{(0)}$, which consists of flipping $\hat{\mathbf{x}}^{(0)}$ one symbol position at a time. The SF procedure discards $\hat{\mathbf{x}}^{(0)}$ when the ML cost of $\hat{\mathbf{x}}^{(1)}$, which differs from $\hat{\mathbf{x}}^{(0)}$ in one symbol position, is lesser than the ML cost of the current solution $\hat{\mathbf{x}}^{(0)}$. In the following, the SF procedure is applied from the first to the last position of $\hat{\mathbf{x}}^{(1)}$ and so on.

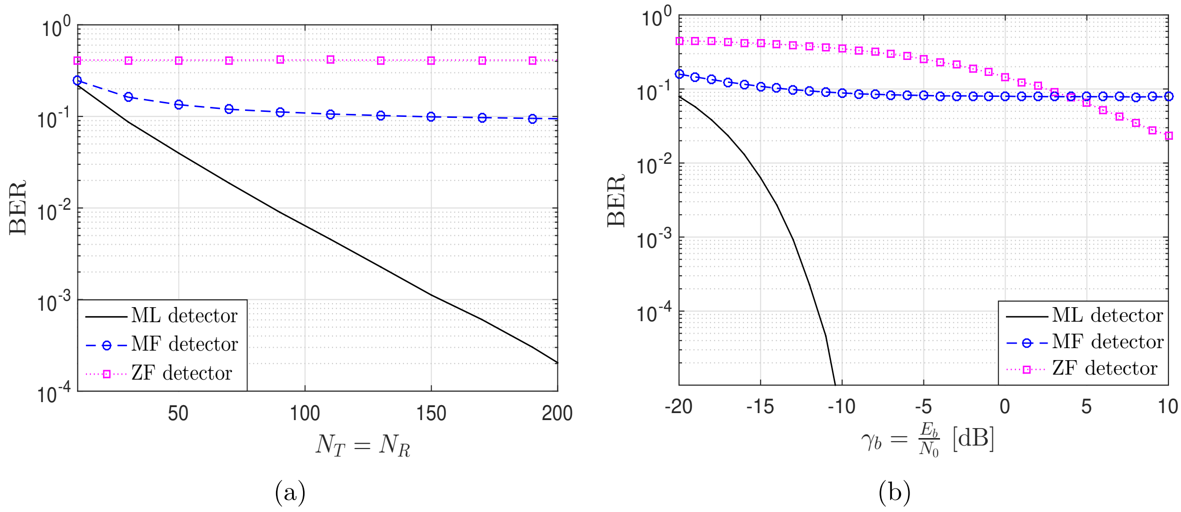


Figure 7.1 – (a) BER as a function of $N_T = N_R$ for a fixed $\gamma_b = -15$ dB employing MF and ZF detector. (b) BER as a function of γ_b for a SM-MIMO system with $N_T = N_R = 100$ antennas employing MF and ZF detector. ML detector perform is included as a target.

The SF procedure stops when no solution is achieved, i.e., there is no new solution, that differs from the current solution by one symbol position, presents lower ML cost. Note that when no local additional sub-optimum solution is achieved, one could pass to a second stage where the SF are performed by flipping two consecutive symbol positions of the current solution, then three symbol positions and so forth. However, in order to maintain the complexity low, in this chapter it is considered one SF at a time of the current solution. The SF procedure is formally described in the following in more detail.

Let $\hat{\mathbf{x}}^{(k)}$ be the solution of the k-th iteration whose elements are $\hat{x}_i^{(k)}$, for $i = 1, \cdots, N_T$. In the $(k+1)$-th iteration, the solution $\hat{\mathbf{x}}^{(k+1)}$ can be obtained from $\hat{\mathbf{x}}^{(k)}$ by flipping the i-th element, that is:

$$\hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{x}}^{(k)} + \delta_i^{(k)} \mathbf{e}_i, \tag{7.3}$$

where $\mathbf{e}_i$ represents a column vector whose i-th entry is equal to $+1$ and all other elements are zero. The possible values of $\delta_i^{(k)} = \hat{x}_i^{(k+1)} - \hat{x}_i^{(k)}$ depends on the constellation of order M employed and the current value of $\hat{x}_i^{(k)}$. For example, the symbols of constellations of the 4-QAM modulation are given by $\{1+j, 1-j, -1-j, -1+j\}$. If $\hat{x}_i^{(k)} = 1+j$, the possible values of $\delta_i^{(k)}$ are $\mathbb{P} = \{-2, -2j, -2-2j\}$.

The SF is successful, if the ML cost function given by (7.2) decreases, that is if:

$$\Delta^{(k+1)} = \Phi[\hat{\mathbf{x}}^{(k+1)}] - \Phi[\hat{\mathbf{x}}^{(k)}] < 0. \tag{7.4}$$

On the other hand, if $\Delta^{(k+1)} \geq 0$, the SF procedure is terminated in the last vector position. Then $\hat{\mathbf{x}}^{(k)}$ is declared the detected signal vector.

From (7.2) and (7.3) into (7.4), the decrement in $\Delta^{(k+1)}$ of the ML cost function can be written as:

$$\begin{aligned}
\Delta^{(k+1)} &= [(\hat{\mathbf{x}}^{(k)} + \delta_i^{(k)}\mathbf{e}_i)^H \mathbf{G}(\hat{\mathbf{x}}^{(k)} + \delta_i^{(k)}\mathbf{e}_i) - 2\Re\{(\hat{\mathbf{x}}^{(k)} + \delta_i^{(k)}\mathbf{e}_i)^H \tilde{\mathbf{y}}\}] - [\hat{\mathbf{x}}^{(k)^H}\mathbf{G}\hat{\mathbf{x}}^{(k)} - 2\Re\{\hat{\mathbf{x}}^{(k)^H}\tilde{\mathbf{y}}\}] \\
&= \left|\delta_i^{(k)}\right|^2 \mathbf{e}_i^H \mathbf{G}\mathbf{e}_i + \delta_i^{(k)}\hat{\mathbf{x}}^{(k)^H}\mathbf{G}\mathbf{e}_i + \delta_i^{(k)^*}\mathbf{e}_i^H\mathbf{G}\hat{\mathbf{x}}^{(k)} - 2\Re\left\{\delta_i^{(k)}\mathbf{e}_i^H\tilde{\mathbf{y}}\right\} \\
&= \left|\delta_i^{(k)}\right|^2 (\mathbf{G})_{i,i} + \delta_i^{(k)}\hat{\mathbf{x}}^{(k)^H}\mathbf{g}_i + \delta_i^{(k)^*}\mathbf{g}_i^H\hat{\mathbf{x}}^{(k)} - 2\Re\left\{\delta_i^{(k)^*}\tilde{y}_i\right\} \\
&= \left|\delta_i^{(k)}\right|^2 (\mathbf{G})_{i,i} + \delta_i^{(k)}\hat{\mathbf{x}}^{(k)^H}\mathbf{g}_i + \delta_i^{(k)^*}(\hat{\mathbf{x}}^{(k)^H}\mathbf{g}_i)^H - 2\Re\left\{\delta_i^{(k)^*}\tilde{y}_i\right\},
\end{aligned} \tag{7.5}$$

where $(\mathbf{G})_{i,i}$ denotes the (i,i)-th entry of $\mathbf{G}$, $\mathbf{g}_i$ is the i-th column vector of $\mathbf{G}$, and $\tilde{y}_i$ is the i-th element of vector $\tilde{\mathbf{y}}$.

From (7.5), note that $\Delta^{(k+1)}$ depends on the current solution $\hat{\mathbf{x}}^{(k)}$ and on the value of $\delta_i^{(k)}$, which takes only integer values according to the constellation employed. Then, the minimum of $\Delta^{(k+1)}$ can be evaluated as:

$$\Delta^{(k+1)} = \min_{\delta \in \mathbb{P}} \left\{ \left|\delta^{(k)}\right|^2 (\mathbf{G})_{i,i} + \delta^{(k)}\hat{\mathbf{x}}^{(k)^H}\mathbf{g}_i + \delta_i^{(k)^*}(\hat{\mathbf{x}}^{(k)^H}\mathbf{g}_i)^H - 2\Re\left\{\delta^{(k)^*}\tilde{y}_i\right\} \right\}, \tag{7.6}$$

where $\mathbb{P}$ is the set of all possible $\delta_i^{(k)}$ values. Finally, from (7.6), the k-th SF is successful if $\Delta^{(k+1)} < 0$.

In summary, the SF procedure starts from the MF initial solution $\hat{\mathbf{x}}^{(0)}$ and performs SF from the 1st to N_T-th symbol position of $\hat{\mathbf{x}}^{(0)}$. Based on the symbol in the i-th position of $\hat{\mathbf{x}}^{(0)}$ and on the constellation, the minimum value of $\delta_i^{(0)}$ given by (7.6) is determined. If $\Delta^{(1)} < 0$, a new solution $\hat{\mathbf{x}}^{(1)}$ by updating the i-th position of $\hat{\mathbf{x}}^{(0)}$ as in (7.3) is employed. This new solution replaces the early one and the SF procedure is applied from the 1st to N_T-th symbol position of the new solution $\hat{\mathbf{x}}^{(1)}$. When no better solution is found, the SF procedure is finished. The SF procedure is presented in Algorithm 7.1.

Fig. 7.2 shows the BER versus the number of transmit and receive antennas, $N_T = N_R$, for a MIMO system employing the SF procedure for 4-QAM and γ_b of -15 dB. From this figure, note that for a low number of transmit and receive antennas, there is a degradation on the BER for the SF procedure in comparison to the optimum performance represented by the ML detector.

The above behavior can be explained according to the random matrix theory (BAI; CHOI, 2014; EDELMAN, 1989), where it is known that when N_T and N_R are high, the matrix $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ becomes a diagonal matrix $N_R\mathbf{I}_{N_T}$ for an uncorrelated Rayleigh fading channel, which is known in the literature as channel hardening (RUSEK *et al.*, 2013).

Pseudocódigo 7.1 SF procedure

Input:
$\tilde{\mathbf{y}} = \mathbf{H}^H \mathbf{y}$;
$\hat{\mathbf{x}}^{(0)} = Q\{\tilde{\mathbf{y}}\}$;
$\mathbf{G} = \mathbf{H}^H \mathbf{H}$;
$k = 0$;
$i = 1$;
SF iterative process:
while $i \le N_T$ **do**
  Based on $\hat{x}_i^{(k)}$ generate $\mathbb{P} = \{\delta_{i,j}^{(k)}\}, j = 1, \cdots, M-1$;
  $\Delta^{(k+1)} = \min_{\delta \in \mathbb{P}} \left\{ \left|\delta^{(k)}\right|^2 (\mathbf{G})_{i,i} + \delta^{(k)} \hat{\mathbf{x}}^{(k)^H} \mathbf{g}_i + \delta_i^{(k)^*} (\hat{\mathbf{x}}^{(k)^H} \mathbf{g}_i)^H - 2\Re\left\{\delta^{(k)^*} \tilde{y}_i\right\} \right\}$;
  if $\Delta^{(k+1)} < 0$ **then**
    $\hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{x}}^{(k)} + \delta_i^{(k)} \mathbf{e}_i$;
    $i = 0$;
    $k = k + 1$;
  end if
  $i = i + 1$;
end while
Output:
  Detected vector $\hat{\mathbf{x}}^{(k)}$.

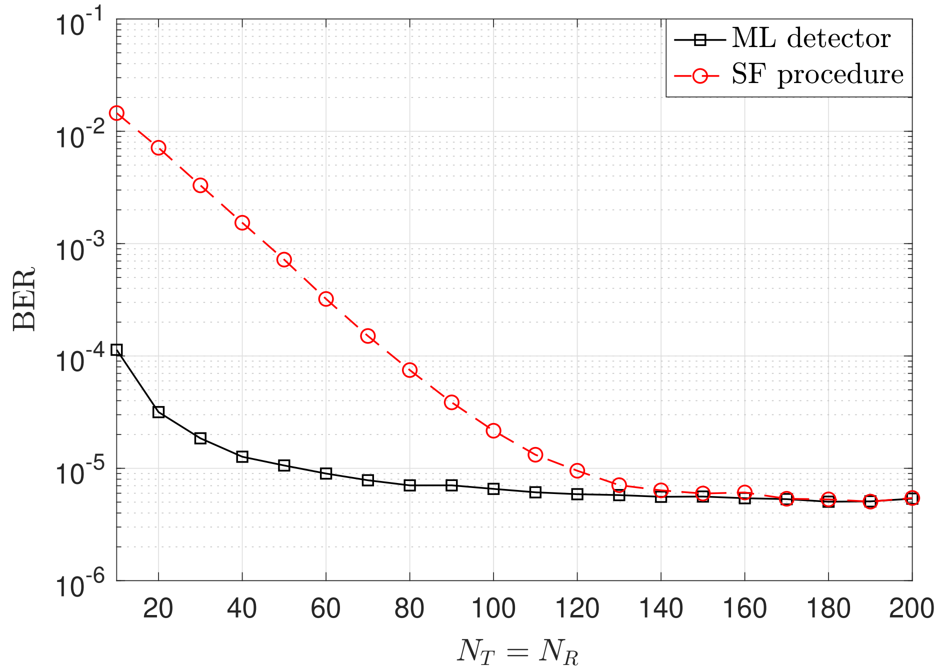
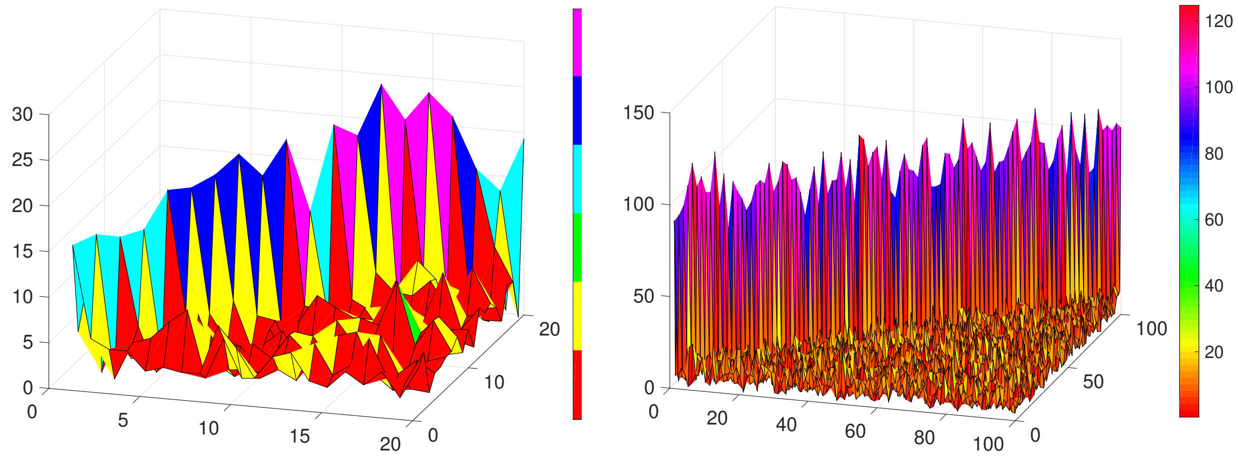


Figure 7.2 – BER as a function of the number of transmit and receive antennas ($N_T = N_R$) for $\gamma_b = -15$ dB.

Fig. 7.3 shows the channel hardening behavior of spatial multiplexing MIMO systems for $N_T = N_R = 20$ and $N_T = N_R = 100$ antennas. Notice that as the dimension of $\mathbf{H}$ increases, the modulus of the diagonal elements of $\mathbf{G}$ becomes larger than the



(a) (b)

Figure 7.3 – Channel hardening behavior of $\mathbf{G} = \mathbf{H}^H\mathbf{H}$ for spatial multiplexing MIMO systems for (a) $N_T = N_R = 20$, (b) $N_T = N_R = 100$ antennas.

modulus of the off-diagonal elements. Thus, local sub-optimum solutions converge easily to the global optimum solution for $N_T > 130$ antennas. Nevertheless, for a low number of antennas, the SF procedure gets trapped stuck at local sub-optimum solutions, as shown in Fig. 7.2. In such a situation, an action should take place in order to escape from a local sub-optimum solution and to reach to the global optimum solution.

Based on the above, in the following subsection, a simple but very effective method to improve further the performance of SF procedure for a low number of transmit and receive antennas is proposed.

7.3.2 Initial Solution Vectors

In order to the SF procedure to escape from a local sub-optimum solution, new initial solution vectors are necessary. These initial solution vectors are used as a starting-point for the SF procedure to increase the probability of achieving the global optimum solution.

A set of L initial solution vectors is constructed, where the first initial solution vector is the MF $\hat{\mathbf{x}}^{(0)}$ solution. Thus, $L-1$ initial solution vectors should be generated, which are different from $\hat{\mathbf{x}}^{(0)}$ in c symbol positions. For $\hat{\mathbf{x}}^{(0)}$, c random positions are replaced with symbols from the chosen constellation. Then, the SF procedure for each one of the L initial solution vectors in order to obtain L local sub-optimum solutions is performed, i.e., $\hat{\mathbf{x}}_0^{(k_0)}, \hat{\mathbf{x}}_1^{(k_1)}, \cdots, \hat{\mathbf{x}}_{L-1}^{(k_{L-1})}$, where k_p denotes the number of steps to SF convergence. Finally, the local sub-optimum solution that presents lesser ML cost is chosen as the final solution, deemed as the global optimum.

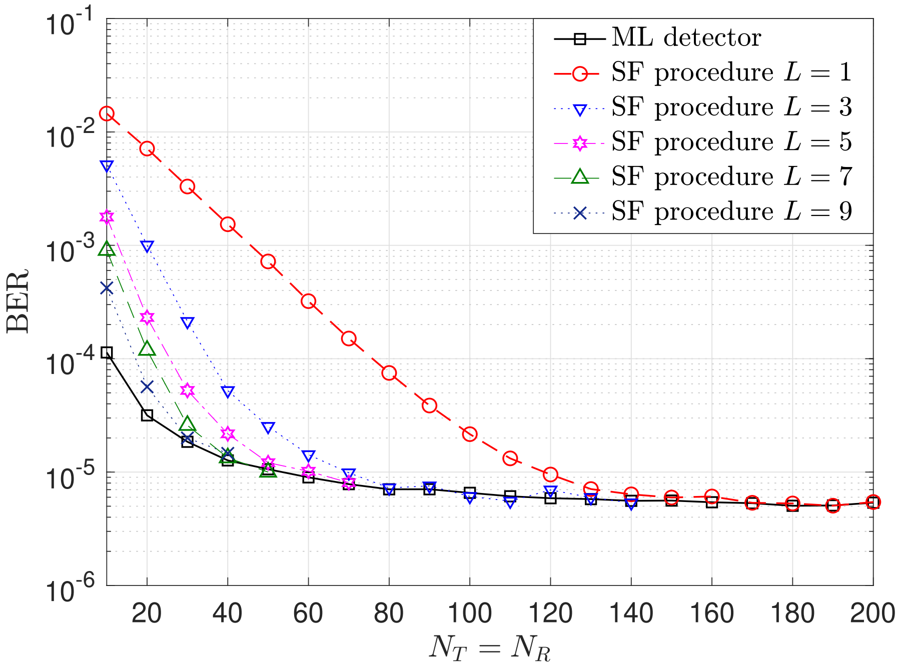


Figure 7.4 – BER as a function of the number of transmit and receive antennas with $N_T = N_R$ for $\gamma_b = -15$ dB.

Fig. 7.4 shows the BER versus the number of transmit and receive antennas for a MIMO system employing different number of initial solution vectors L, for 4-QAM, $\gamma_b = -15$ dB and $c = \lfloor \frac{N_T}{2} \rceil$. Observe that as the number of initial solution vectors L increases, better performance is obtained especially for low number of antennas. Thus, for example, for $N_T = N_R = 40$ antennas, $L = 7$ initial solution vectors are necessary to achieve the optimum performance, which demonstrates the effectiveness of this approach.

In order to obtain the $L - 1$ initial solutions, the number of symbols changed from $\hat{\mathbf{x}}^{(0)}$ in Fig. 7.4 was given by $c = \lfloor \frac{N_T}{2} \rceil$. As the maximum number of possible symbol changes is $c_{\max} = N_T$, an interesting task is to determine the optimum number of symbols to be changed from $\hat{\mathbf{x}}^{(0)}$. This is answered in the next figure.

Fig. 7.5 shows the BER as a function of γ_b for a MIMO system with $N_T = N_R = 20$ antennas and 4-QAM employing the SF procedure for L initial solution vectors with c random positions changed from $\hat{\mathbf{x}}^{(0)}$. In Fig. 7.5a for $L = 1$, obviously, increasing on c does not improve SF performance. However, as L and c increase, better performance is achieved by the SF procedure (see Fig. 7.5b, Fig. 7.5c and Fig. 7.5d). Furthermore, note from Fig. 7.5d that for $L = 10$, to change $c = 5$ random positions from $\hat{\mathbf{x}}^{(0)}$ is enough to approximate the ML curve for a BER of 10^{-5}. Changing $c = 10$ positions maintain the performance of $c = 5$. Increasing c from $\lfloor \frac{N_T}{2} \rceil$, it is expected the performance gets worsen again.

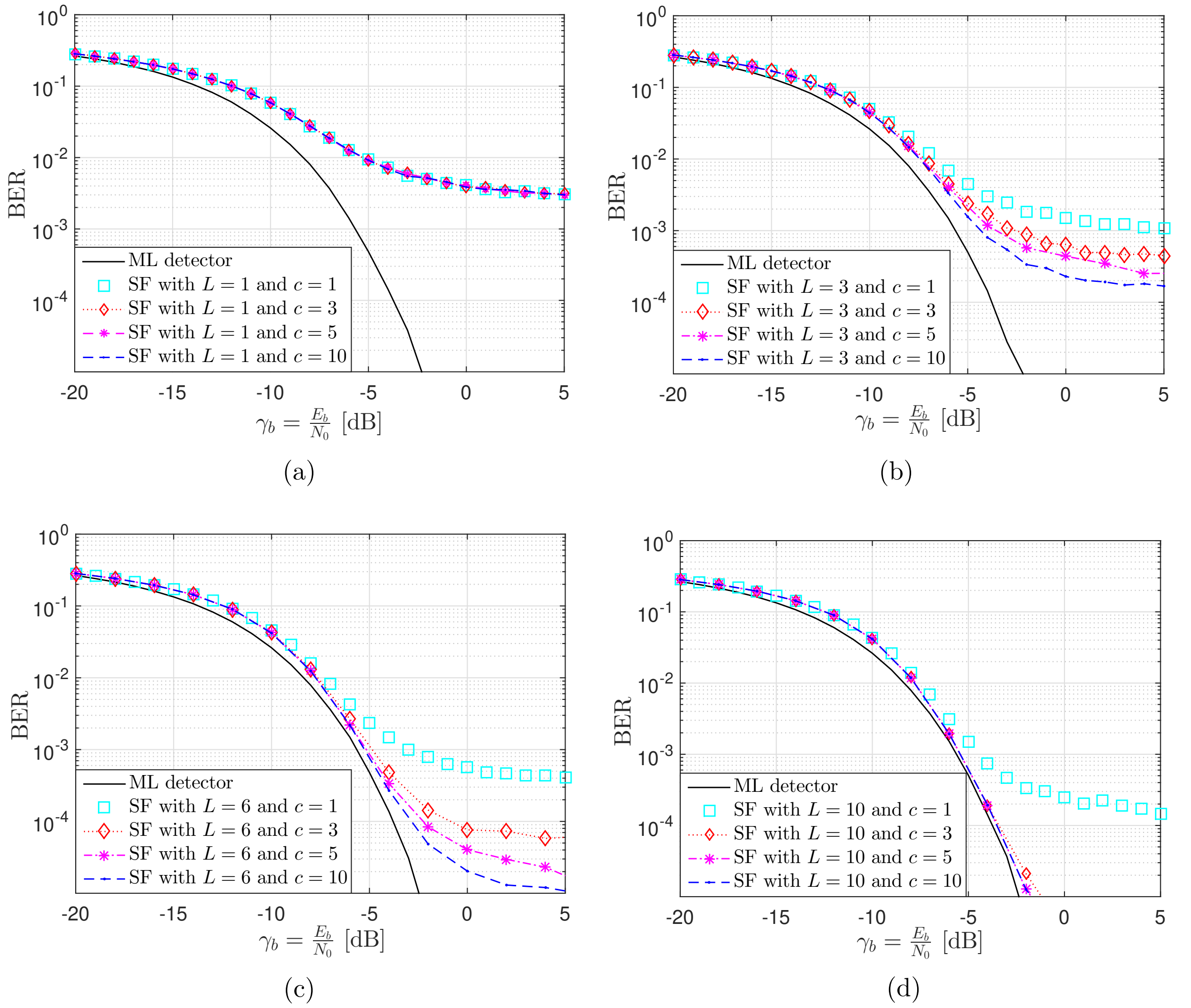


Figure 7.5 – BER as a function of γ_b for a spatial multiplexing MIMO system with $N_T = N_R = 20$ antennas and 4-QAM employing the SF procedure for different c for (a) $L = 1$, (b) $L = 3$, (c) $L = 6$ and (d) $L = 10$ initial solution vectors

7.4 Numerical Results

In this section, the performance and complexity of the SF algorithm for MIMO systems are presented.

7.4.1 BER Performance

Fig. 7.6 and Fig. 7.7 show the BER of SF algorithm for $N_T = N_R = 50$ and $N_T = N_R = 100$ antennas and for 16-QAM, respectively. From these figures, notice that since N_T is large, for $L = 5$ with $c = 3$ the BER is near ML until 10^{-5}, respectively.

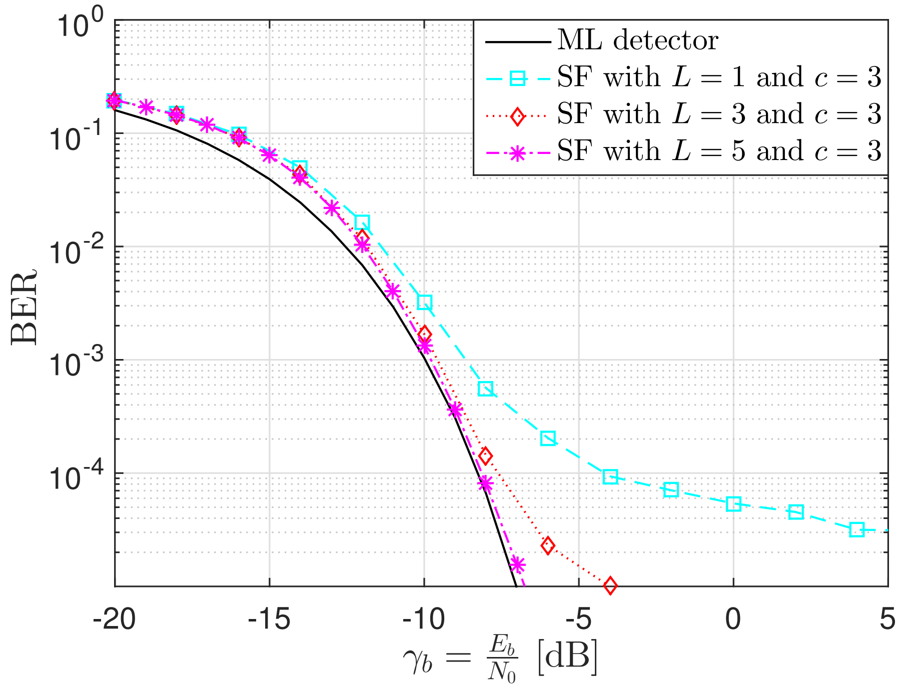


Figure 7.6 – BER as a function of γ_b for a spatial multiplexing MIMO system for $N_T = N_R = 50$ antennas and 16-QAM employing SF algorithm.

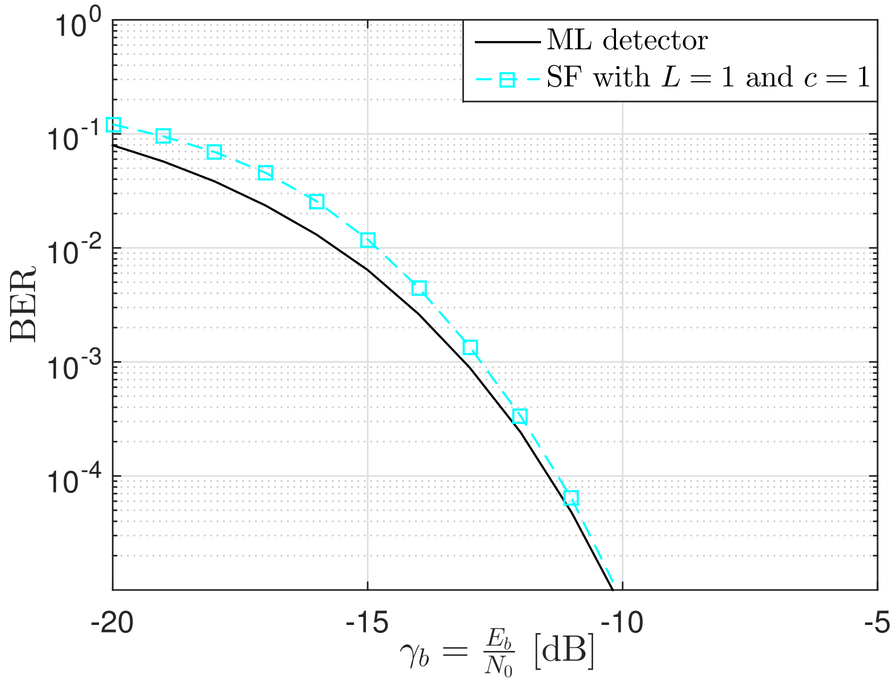


Figure 7.7 – BER as a function of γ_b for a spatial multiplexing MIMO system with $N_T = N_R = 100$ antennas and 16-QAM employing SF algorithm.

7.4.2 Computational Complexity

Basically, the complexity of SF procedure depends on the number of initial solution vectors L and the number of antennas N_T.

Thus, the total complexity of the SF procedure is split in two parts. In the first part, the initialization step computes the MF initial vector solution $\hat{\mathbf{x}}^{(0)}$, which consists of a matrix-vector multiplication and a component-wise quantization, whose complexity is equal to $2N_T^2 - N_T$ flops. The complexity of generating $L-1$ initial vector solutions which differ in c random position from $\hat{\mathbf{x}}^{(0)}$, is $N_T(L-1)$ flops. Therefore, the computational complexity of the first part is equal to:

$$C_I = 2N_T^2 + N_T L - 2N_T. \tag{7.7}$$

The second part of the SF iterative process (lines 8 to 18 of algorithm 7.1) is due to each of the L initial solution vectors. Once the SF iterative process involves the computation of $\Delta^{(k+1)}$ (line 10) for the i-th element of the current solution $\hat{\mathbf{x}}^{(k)}$, whose complexity is $(M-1)(2N_T+10)$ flops. Then, if the condition $\Delta^{(k+1)} < 0$ (line 11) is satisfied, 2 flops are required for updating $\hat{\mathbf{x}}^{(k+1)}$ (see line 12). Otherwise, if condition $\Delta^{(k+1)} < 0$ is not satisfied, the iterative process involves just the computation of $\Delta^{(k+1)}$. Therefore, considering L initial solution vectors, the computational complexity of the second part of the procedure is equal to:

$$C_{\text{II}} = LK(2N_T M + 10M - 2N_T - 8) \tag{7.8}$$

flops, where K is the average number of successful iterations, that is when $\Delta^{(k+1)} < 0$ is satisfied.

Fig. 7.8a shows the average number of successful iterations K against γ_b for MIMO systems with $N_T = N_R = 20, 50, 100$ antennas and 4-QAM. Note, as expected, K increases with N_T. On the other hand, Fig. 7.8b presents the relation between K and N_T for $\gamma_b = -10$ dB, 4-QAM and c optimum that has been obtained by simulation. From this figure, observe that K has a linear increase with N_T, that is

$$K = 0.59N_T - 1.2. \tag{7.9}$$

Finally, adding the first and second part given by (7.7) and (7.8), respectively, the total number of flops by employing SF algorithm is given by:

$$\begin{aligned} C_{\text{T}} &= 2N_T^2 + N_T L - 2N_T + 2N_T LKM + 10LKM - 2N_T LK - 8LK \\ &\approx N_T^2(2 + 1.8LM - 1.18L), \end{aligned} \tag{7.10}$$

where (7.9) was used in (7.8) plus (7.7).

7.5 Chapter Conclusions

In this chapter, a low-complexity SF detector for spatial multiplexing MIMO systems with a large number of transmit and receive antennas was presented. The pro-

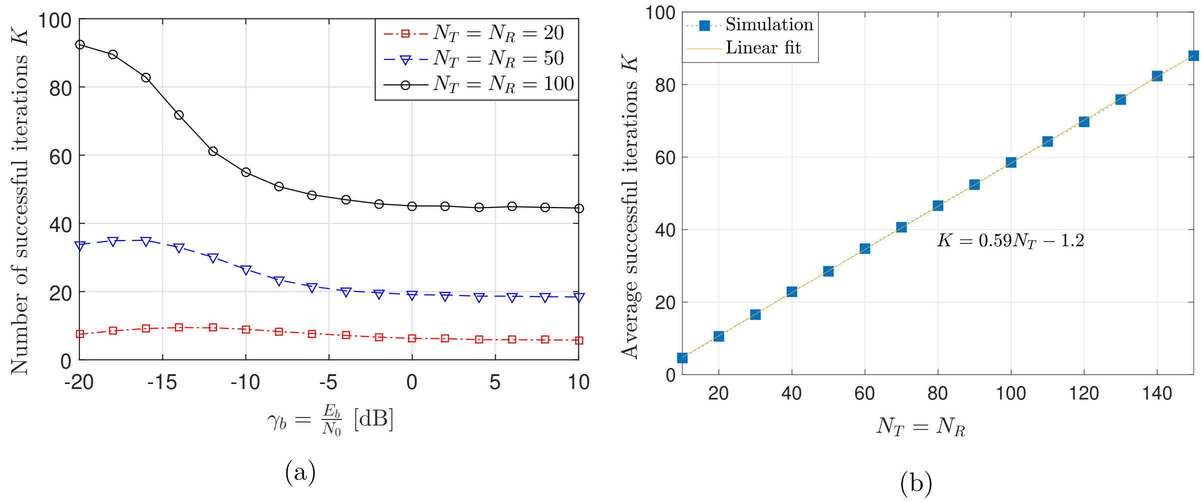


Figure 7.8 – Average number of successful iterations K against (a) γ_b for MIMO systems with $N_T = N_R = 20, 50$ and 100 antennas, (b) $N_T = N_R$ for $\gamma_b = -10$ dB and c optimum.

posed SF detector was shown to have excellent attributes in terms of both low complexity as well as quasi-optimum performance, achieving high spectral efficiency of the order of tens to hundreds of bps/Hz. In general, it achieves optimum performance by employing one simple SF procedure, when the number of antennas is of the order of hundreds. On the contrary, when the number of antennas is on the order of tens to achieve the optimum performance several SF procedures must be employed, presenting a quadratic complexity with N_T. Finally, numerical results have confirmed that the proposed detector algorithm achieves optimum performance.

8 Conclusions and Future Works

This chapter presents the concluding remarks of the thesis and future research works.

8.1 Conclusions

This thesis was focused on aspects of design and implementation of low-complexity detector algorithms for spatial multiplexing MIMO systems with large number of transmit and receive antennas. These systems constitute one of the most promising techniques to achieve higher spectral efficiency in 5G wireless communications standard. Thus the proposal of novel detector algorithms with high-performance and low-complexity are necessary.

The contributions of this thesis are two-fold. Firstly, when the number of receive antennas is much larger than the number of transmit antennas, that is named spatial multiplexing massive MIMO systems, linear ZF detector was proven to achieve near-optimum performance and full diversity order due to the property of asymptotically orthogonal channel matrix. Thus, a closed-form BER expression were derived in order to determine the performance difference between ZF and the optimum ML detector. In this context, ZF detector has attracted interest for its employ in spatial multiplexing massive MIMO systems.

However, once ZF detector involves the computation of an inverse matrix, its practical implementation might be cumbersome. Thus, by exploiting the property of asymptotically orthogonal channel matrix of spatial multiplexing massive MIMO systems, two low-complexity ZF detectors based on Damped Jacobi (DJ) and Newton-Schultz iterative (NSI) algorithm were proposed in order to avoid the computation of an inverse matrix. The performance of both algorithms were verified resulting in acceptable BER performance, especially when the number of iterations increases. Furthermore, both algorithms are able to reduce the number of required computations from $O(N_T^3)$ to $O(N_T^2)$. Beside that, NSI algorithm outperforms significantly DJ algorithm in terms of performance and complexity, which makes it ideal for spatial multiplexing massive MIMO detection.

Secondly, contrary to the previous case, when the number of transmit and receive antennas is large and equal ($N_T = N_R$), that is named symmetric spatial multiplexing MIMO systems, ZF detector achieves a poor performance in comparison to the optimum ML performance, attaining a diversity order of one, which is undesirable. Since

ML detector is impractical for large number of transmit antennas, due to its exponential complexity, more suitable algorithms that yield low complexity and perform as close to optimum, as possible, are necessary. In this context, a novel precoding approach and a symbol-flipping (SF) detector algorithm were proposed and evaluated for symmetric spatial multiplexing MIMO detection. The proposed precoding approach transforms a symmetric channel matrix into an orthogonal matrix in order to apply ZF detector at the receiver side achieving optimum performance at the expense of a small increase in the total average transmitted power. Despite achieving optimum performance through precoding approach, its disadvantage is that the transmitter needs to know perfectly the channel matrix, which limits its practical implementation. On the other hand, the proposed SF detector algorithm, which is based on local sub-optimum searching, is able to achieve the optimum ML performance when a large number, in the order of hundreds, of transmit and receive antennas is available. However, if the number of antennas is in the order of tens, SF detector algorithm must employ different random starting-point in order to achieve the optimum performance.

As a general conclusion, it was shown in this thesis that problems arising in large-scale spatial multiplexing MIMO detection can be efficiently overcome.

8.2 Future Works

In this section, some proposals related to the contributions and results of this thesis can be studied in the future, such as:

- The detection algorithms proposed in this thesis considers single-cell scenarios. However, some additional challenges could appear in multi-cell scenarios due to the inter-cell interference. In this context, the proposed detection algorithms should be generalized to multi-cell scenarios.

- In this thesis, error correcting codes were not considered. For this reason, in order to improve the performance of large-scale spatial multiplexing MIMO systems, it is suggested as future research to consider error correcting codes, such as convolutional codes, turbo codes and low-density parity-check (LDPC) codes. Thus, developing iterative detection and decoding algorithms can be an interesting challenge.

- Throughout this thesis, the Rayleigh fading channels have been considered. In this context, other fading channel models can be considered. For example, Nakagami-m, Rician fading channels or a combination of channel models with and without line-of-sight.

- Perfect channel state information (CSI) at receiver was assumed in this thesis. However, in practice, the CSI is not perfect. Thus, it is proposed as a future research to consider imperfect CSI in a more real large-scale spatial multiplexing MIMO system model. Furthermore, it would be interesting to study and implement channel estimation algorithms for large-scale spatial multiplexing MIMO systems.

- Hardware implementation of the algorithms proposed in this thesis is an interesting challenge for future works.

- Employing large-scale spatial multiplexing MIMO systems in cooperative networks.

Bibliography

AARTS, E.; LENSTRA, J. K. (Ed.). *Local Search in Combinatorial Optimization*. 1st. ed. New York, NY, USA: John Wiley & Sons, Inc., 1997. ISBN 0471948225. Citado na página 109.

AFFLERBACH, L.; GROTHE, H. Calculation of Minkowski-Reduced Lattice Bases. *Computing*, v. 35, n. 3, p. 269–276, Sep 1985. ISSN 1436-5057. Disponível em: <https://doi.org/10.1007/BF02240194>. Citado na página 50.

AGIWAL, M.; ROY, A.; SAXENA, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Communications Surveys Tutorials*, v. 18, n. 3, p. 1617–1655, thirdquarter 2016. ISSN 1553-877X. Citado 2 vezes nas páginas 23 and 54.

ALMEIDA, C. de. *Livro não publicado*. Campinas, SP, Brazil: Unicamp, 2014. Citado 5 vezes nas páginas 30, 32, 33, 34, and 60.

ANDREWS, J. G.; GHOSH, A.; MUHAMED, R. *Fundamentals of WiMAX: Understanding Broadband Wireless Networking (Prentice Hall Communications Engineering and Emerging Technologies Series)*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2007. ISBN 0132225522. Citado 6 vezes nas páginas 21, 30, 34, 46, 49, and 98.

BAI, L.; CHOI, J. *Low Complexity MIMO Detection*. [S.l.]: Springer Publishing Company, Incorporated, 2014. ISBN 1489998241, 9781489998248. Citado 5 vezes nas páginas 22, 44, 46, 50, and 113.

BARRETT, R.; BERRY, M.; CHAN, T. F.; DEMMEL, J.; DONATO, J.; DONGARRA, J.; EIJKHOUT, V.; POZO, R.; ROMINE, C.; VORST, H. Van der. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. 2nd ed. ed. [S.l.]: Society for Industrial and Applied Mathematics (SIAM), 1994. Citado na página 79.

BARRY, J. R.; MESSERSCHMITT, D. G.; LEE, E. A. *Digital Communication: Third Edition*. Norwell, MA, USA: Kluwer Academic Publishers, 2003. ISBN 0792375483. Citado 4 vezes nas páginas 33, 47, 48, and 59.

BEJARANO, O.; KNIGHTLY, E. W.; PARK, M. IEEE 802.11ac: From Channelization to Multi-User MIMO. *IEEE Communications Magazine*, v. 51, n. 10, p. 84–90, October 2013. ISSN 0163-6804. Citado na página 54.

BEN-ISRAEL, A. An Iterative Method for Computing the Generalized Inverse of an Arbitrary Matrix. *Mathematics of Computation*, v. 19, p. 452–455, 1965. Citado 2 vezes nas páginas 80 and 82.

BEN-ISRAEL, A.; COHEN, D. On Iterative Computation of Generalized Inverses and Associated Projections. *J. Soc. Indust. Appl. Math. Ser. Numer. Anal. Math. of Comp*, v. 3, p. 410–419, 1966. Citado 3 vezes nas páginas 81, 82, and 83.

BIGLIERI, E.; CALDERBANK, R.; CONSTANTINIDES, A.; GOLDSMITH, A.; PAULRAJ, A.; POOR, H. V. *MIMO Wireless Communications*. New York, NY, USA:

Cambridge University Press, 2007. ISBN 0521873282. Citado 2 vezes nas páginas 30 and 59.

BLUM, C.; ROLI, A. Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. *ACM Computing Surveys*, ACM, New York, NY, USA, v. 35, n. 3, p. 268–308, Sep 2003. Citado na página 109.

CHOCKALINGAM, A.; RAJAN, B. S. *Large MIMO Systems*. New York, NY, USA: Cambridge University Press, 2014. ISBN 1107026652, 9781107026650. Citado 3 vezes nas páginas 21, 22, and 100.

CLERCKX, B.; OESTGES, C. Chapter 2 - From Multi-Dimensional Propagation to Multi-Link MIMO Channels. In: CLERCKX, B.; OESTGES, C. (Ed.). *MIMO Wireless Networks (Second Edition)*. Second edition. Oxford: Academic Press, 2013. p. 29 – 58. ISBN 978-0-12-385055-3. Disponível em: <http://www.sciencedirect.com/science/article/pii/B978012385055300002X>. Citado na página 34.

DAHLMAN, E.; PARKVALL, S.; SKOLD, J.; BEMING, P. *3G Evolution, Second Edition: HSPA and LTE for Mobile Broadband*. 2. ed. Orlando, FL, USA: Academic Press, Inc., 2008. ISBN 0123745381, 9780123745385. Citado na página 21.

DAI, L.; GAO, X.; SU, X.; HAN, S.; I, C. L.; WANG, Z. Low-Complexity Soft-Output Signal Detection Based on Gauss-Seidel Method for Uplink Multiuser Large-Scale MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 64, n. 10, p. 4839–4845, Oct 2015. ISSN 0018-9545. Citado na página 65.

EDELMAN, A. *Eigenvalues and Condition Number of Random Matrices*. Tese (Doutorado) — Massachusetts Institute of Technology, Department of Mathematics, 5 1989. Citado 2 vezes nas páginas 43 and 113.

FABER, V.; LIESEN, J.; TICHý, P. On Chebyshev Polynomials of Matrices. *SIAM Journal on Matrix Analysis and Applications*, v. 31, n. 4, p. 2205–2221, 2010. Disponível em: <http://dx.doi.org/10.1137/090779486>. Citado na página 88.

FANG, W.; SUN, H.; YANG, L. Power Allocation for Maximizing Sum Capacity of Multiuser MIMO Downlink with Transmit Precoding Based on SLNR. In: *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*. [S.l.: s.n.], 2011. p. 1–5. ISSN 1550-2252. Citado na página 104.

GAO, X.; DAI, L.; MA, Y.; WANG, Z. Low-Complexity Near-Optimal Signal Detection for Uplink Large-Scale MIMO Systems. *Electronics Letters*, v. 50, n. 18, p. 1326–1328, August 2014. ISSN 0013-5194. Citado na página 65.

GOLUB, G. H.; LOAN, C. F. V. *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996. ISBN 0-8018-5414-8. Citado 8 vezes nas páginas 37, 38, 39, 49, 61, 77, 79, and 100.

GYAMFI, K. S.; BAEK, J.; LEE, K. Lattice Reduction Aided Tabu Search with Channel-Dependent Stopping Criterion for MIMO Systems. *Electronic Letters*, v. 51, n. 24, p. 2062–2064, 2015. ISSN 0013-5194. Citado na página 93.

HAGHANI, F. K.; SOLEYMANI, F. An Improved Schulz-Type Iterative Method for Matrix Inversion with Application. *Transactions of the Institute of Measurement and Control*, v. 36, p. 983–991, 2014. Citado na página 83.

HASSIBI, B.; VIKALO, H. On the Sphere-Decoding Algorithm i. Expected Complexity. *IEEE Transactions on Signal Processing*, v. 53, n. 8, p. 2806–2818, Aug 2005. ISSN 1053-587X. Citado na página 47.

HIGHAM, N. J. *Functions of Matrices: Theory and Computation*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2008. xx+425 p. ISBN 978-0-898716-46-7. Citado 3 vezes nas páginas 77, 79, and 86.

HUANG, H.; PAPADIAS, C. B.; VENKATESAN, S. *MIMO Communication for Cellular Networks*. [S.l.]: Springer Publishing Company, Incorporated, 2011. ISBN 0387775218, 9780387775210. Citado na página 22.

IZADINASAB, M. K.; DAMEN, M. O.; NAJAFI, H. Efficient Lattice-Reduction-Aided Conditional Detection for MIMO Systems. In: *2017 15th Canadian Workshop on Information Theory (CWIT)*. [S.l.: s.n.], 2017. p. 1–5. Citado na página 98.

JALDEN, J.; OTTERSTEN, B. On the Complexity of Sphere Decoding in Digital Communications. *IEEE Transactions on Signal Processing*, v. 53, n. 4, p. 1474–1484, April 2005. ISSN 1053-587X. Citado na página 46.

JING, C.; WANG, X.; CHEN, B.; MA, Y.; WEI, J. A Low-Complexity MIMO Detector Based on Fast Dual-Lattice Reduction Algorithm. In: *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*. [S.l.: s.n.], 2016. p. 1–6. Citado na página 98.

KONG, B. Y.; PARK, I. C. Low-Complexity Symbol Detection for Massive MIMO Uplink Based on Jacobi Method. In: *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. [S.l.: s.n.], 2016. p. 1–5. Citado 3 vezes nas páginas 65, 67, and 74.

KRAMER, G. Genie-Aided Outer Bounds on the Capacity of Interference Channels. In: *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No.01CH37252)*. [S.l.: s.n.], 2001. p. 103–. Citado na página 48.

LARSSON, E. G. MIMO Detection Methods: How They Work [lecture notes]. *IEEE Signal Processing Magazine*, v. 26, n. 3, p. 91–95, May 2009. ISSN 1053-5888. Citado na página 22.

LI, Q.; LI, G.; LEE, W.; LEE, M.; MAZZARESE, D.; CLERCKX, B.; LI, Z. MIMO Techniques in WiMAX and LTE: A Feature Overview. *IEEE Communications Magazine*, v. 48, n. 5, p. 86–92, May 2010. ISSN 0163-6804. Citado na página 54.

LIM, C.; YOO, T.; CLERCKX, B.; LEE, B.; SHIM, B. Recent Trend of Multiuser MIMO in LTE-Advanced. *IEEE Communications Magazine*, v. 51, n. 3, p. 127–135, March 2013. ISSN 0163-6804. Citado na página 54.

LIN, J.; YU, W.; ZHANG, N.; YANG, X.; ZHANG, H.; ZHAO, W. A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications. *IEEE Internet of Things Journal*, v. 4, n. 5, p. 1125–1142, Oct 2017. ISSN 2327-4662. Citado na página 23.

LIU, J.; XING, S.; SHEN, L. Lattice-Reduction-Aided Breadth-First Tree Searching Algorithm for MIMO Detection. *IEEE Communications Letters*, v. 21, n. 4, p. 845–848, April 2017. ISSN 1089-7798. Citado na página 98.

MARZETTA, T. L. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, v. 9, n. 11, p. 3590–3600, November 2010. ISSN 1536-1276. Citado na página 22.

MARZETTA, T. L.; LARSSON, E. G.; YANG, H.; NGO, H. Q. *Fundamentals of Massive MIMO*. [S.l.]: Cambridge University Press, 2016. Citado 5 vezes nas páginas 22, 40, 54, 55, and 91.

MAURER, J.; MATZ, G.; SEETHALER, D. Low-Complexity and Full-Diversity MIMO Detection Based on Condition Number Thresholding. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2007. v. 3, p. III–61–III–64. ISSN 1520-6149. Citado na página 98.

MINANGO, J. *Avaliação de Desempenho da Técnica de Multiplex Espacial na Presença de Interferência de Co-Canal*. Dissertação (Mestrado) — University of Campinas, Brazil, 2014. Citado 6 vezes nas páginas 34, 48, 49, 50, 57, and 98.

MINANGO, J.; ALMEIDA, C. de. Low-Complexity MMSE Detector Based on Refinement Gauss-Seidel Method for Massive MIMO Systems. In: *2017 IEEE 9th Latin-American Conference on Communications (LATINCOM)*. [S.l.: s.n.], 2017. p. 1–5. Citado 2 vezes nas páginas 65 and 79.

MINANGO, J.; ALMEIDA, C. de. Low-Complexity MMSE Detector Based on the First-Order Neumann Series Expansion for Massive MIMO Systems. In: *2017 IEEE 9th Latin-American Conference on Communications (LATINCOM)*. [S.l.: s.n.], 2017. p. 1–5. Citado 2 vezes nas páginas 79 and 84.

MINANGO, J.; ALMEIDA, C. de. Optimum and Quasi-Optimum Relaxation Parameters for Low-Complexity Massive MIMO Detector Based on Richardson Method. *Electronics Letters*, v. 53, n. 16, p. 1114–1115, 2017. ISSN 0013-5194. Citado 4 vezes nas páginas 24, 55, 69, and 73.

MINANGO, J.; ALMEIDA, C. de. Achieving Optimum Performance with ZF Detector by Using a Novel Precoding Approach in Spatial Multiplexing MIMO Systems. *Physical Communication*, v. 31, p. 62 – 68, 2018. ISSN 1874-4907. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1874490718302489>. Citado na página 24.

MINANGO, J.; ALMEIDA, C. de. Low Complexity Zero Forcing Detector Based on Newton-Schultz Iterative Algorithm for Massive MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 67, n. 12, p. 11759–11766, Dec 2018. ISSN 0018-9545. Citado na página 24.

MINANGO, J.; ALMEIDA, C. de; ALTAMIRANO, C. D. Low-Complexity MMSE Detector for Massive MIMO Systems Based on Damped Jacobi Method. In: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. [S.l.: s.n.], 2017. p. 1–5. ISSN 2166-9589. Citado 4 vezes nas páginas 24, 69, 70, and 79.

MINANGO, J.; ALTAMIRANO, C. D.; ALMEIDA, C. de. Performance Difference Between Zero Forcing and Maximum Likelihood Detectors in Massive MIMO Systems. *Electronics Letters*, v. 54, n. 25, p. 1464–1466, 2018. ISSN 0013-5194. Citado na página 23.

MURUGAN, A. D.; GAMAL, H. E.; DAMEN, M. O.; CAIRE, G. A Unified Framework for Tree Search Decoding: Rediscovering the Sequential Decoder. *IEEE Transactions on Information Theory*, v. 52, n. 3, p. 933–953, March 2006. ISSN 0018-9448. Citado na página 46.

NAGEL, L.; PRATSCHNER, S.; SCHWARZ, S.; RUPP, M. Efficient Multi-User MIMO Transmissions in the LTE-A Uplink. In: *2016 1st International Workshop on Link- and System Level Simulations (IWSLS)*. [S.l.: s.n.], 2016. p. 1–6. Citado na página 54.

NARASIMHAN, T. L.; RAVITEJA, P.; CHOCKALINGAM, A. Large-Scale Multiuser SM-MIMO Versus Massive MIMO. In: *2014 Information Theory and Applications Workshop (ITA)*. [S.l.: s.n.], 2014. p. 1–9. Citado 2 vezes nas páginas 54 and 65.

NGO, H. Q.; LARSSON, E. G.; MARZETTA, T. L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Transactions on Communications*, v. 61, n. 4, p. 1436–1449, April 2013. ISSN 0090-6778. Citado 3 vezes nas páginas 41, 54, and 60.

OLSHEVSKY, V.; TYRTYSHNIKOV, E. *Matrix Methods: Theory, Algorithms and Applications*. [S.l.]: World Scientific, 2010. Citado 2 vezes nas páginas 67 and 70.

PAPOULIS, A.; PILLAI, S. U. *Probability, Random Variables, and Stochastic Processes*. Fourth. Boston: McGraw Hill, 2002. ISBN 0071122567 9780071122566 0073660116 9780073660110 0071226613 9780071226615. Disponível em: <http://www.worldcat.org/search?qt=worldcat_org_all&q=0071226613>. Citado 3 vezes nas páginas 27, 28, and 59.

PEFKIANAKIS, I.; LEE, S.; LU, S. Towards MIMO-Aware 802.11n Rate Adaptation. *IEEE/ACM Transactions on Networking*, v. 21, n. 3, p. 692–705, June 2013. ISSN 1063-6692. Citado na página 21.

PROAKIS, J. G.; MANOLAKIS, D. K. *Digital Signal Processing (4th Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 0131873741. Citado 2 vezes nas páginas 27 and 29.

QIN, Z.; XU, J.; TAO, X.; ZHOU, X. Improved Depth-First-Search Sphere Decoding Based on LAS for MIMO-OFDM Systems. In: *IEEE Vehicular Technology Conference (VTC)*. [S.l.: s.n.], 2015. p. 1–5. Citado na página 93.

RAPPAPORT, T. *Wireless Communications: Principles and Practice*. 2nd. ed. [S.l.]: Prentice Hall, 2002. (Prentice Hall communications engineering and emerging technologies series). ISBN 0130422320. Citado na página 28.

ROGER, S.; GONZALEZ, A.; ALMENAR, V.; VIDAL, A. Combined K-Best Sphere Decoder Based on the Channel Matrix Condition Number. In: *2008 3rd International Symposium on Communications, Control and Signal Processing*. [S.l.: s.n.], 2008. p. 1058–1061. Citado 2 vezes nas páginas 98 and 100.

RUSEK, F.; PERSSON, D.; LAU, B. K.; LARSSON, E. G.; MARZETTA, T. L.; EDFORS, O.; TUFVESSON, F. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Processing Magazine*, v. 30, n. 1, p. 40–60, Jan 2013. ISSN 1053-5888. Citado 10 vezes nas páginas 22, 40, 41, 51, 54, 61, 79, 83, 95, and 113.

SEETHALER, D.; MATZ, G.; HLAWATSCH, F. Low-Complexity MIMO Data Detection Using Seysen's Lattice Reduction Algorithm. In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. [S.l.: s.n.], 2007. v. 3, p. III–53–III–56. ISSN 1520-6149. Citado na página 50.

TANG, C.; LIU, C.; YUAN, L.; XING, Z. High Precision Low Complexity Matrix Inversion Based on Newton Iteration for Data Detection in the Massive MIMO. *IEEE Communication Letters*, v. 20, n. 3, p. 490–493, Mar. 2016. ISSN 1089-7798. Citado 3 vezes nas páginas 79, 81, and 83.

THOMAS, T. A.; VOOK, F. W. Method for Obtaining Full Channel State Information for RF Beamforming. In: *2014 IEEE Global Communications Conference*. [S.l.: s.n.], 2014. p. 3496–3500. ISSN 1930-529X. Citado 2 vezes nas páginas 99 and 109.

TSE, D.; VISWANATH, P. *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005. ISBN 0-5218-4527-0. Citado 3 vezes nas páginas 29, 30, and 50.

TULINO, A. M.; VERDú, S. Random Matrix Theory and Wireless Communications. *Foundations and Trends® in Communications and Information Theory*, v. 1, n. 1, p. 1–182, 2004. ISSN 1567-2190. Disponível em: <http://dx.doi.org/10.1561/0100000001>. Citado 5 vezes nas páginas 40, 61, 66, 71, and 102.

VANNITHAMBY, R.; TALWAR, S. Massive MIMO Communications. In: _____. *Towards 5G: Applications, Requirements and Candidate Technologies*. Wiley, 2017. ISBN 9781118979846. Disponível em: <https://ieeexplore.ieee.org/document/8042298>. Citado na página 54.

WANG, Y.; LEIB, H. Sphere Decoding for MIMO Systems with Newton Iterative Matrix Inversion. *IEEE Communication Letters*, v. 17, n. 2, p. 389–392, Feb. 2013. ISSN 1089-7798. Citado 2 vezes nas páginas 79 and 83.

WEN, Q.; ZHOU, Q.; MA, X. An Enhanced Fixed-Complexity LLL Algorithm for MIMO Detection. In: *2014 IEEE Global Communications Conference*. [S.l.: s.n.], 2014. p. 3231–3236. ISSN 1930-529X. Citado na página 50.

WUBBEN, D.; SEETHALER, D.; JALDEN, J.; MATZ, G. Lattice Reduction. *IEEE Signal Processing Magazine*, v. 28, n. 3, p. 70–91, May 2011. ISSN 1053-5888. Citado 4 vezes nas páginas 50, 51, 98, and 99.

XIANG, W.; ZHENG, K.; SHEN, X. S. *5G Mobile Communications*. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2016. ISBN 3319342061, 9783319342061. Citado 2 vezes nas páginas 21 and 48.

YANG, S.; HANZO, L. Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs. *IEEE Communications Surveys Tutorials*, v. 17, n. 4, p. 1941–1988, Fourthquarter 2015. ISSN 1553-877X. Citado 2 vezes nas páginas 22 and 51.

YANG, Y.; KIM, J. Fixed-Complexity LLL-Based Signal Detection for MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 62, n. 3, p. 1415–1419, March 2013. ISSN 0018-9545. Citado 2 vezes nas páginas 50 and 51.

YIN, B.; WU, M.; CAVALLARO, J. R.; STUDER, C. Conjugate Gradient-Based Soft-Output Detection and Precoding in Massive MIMO Systems. In: *2014 IEEE Global Communications Conference*. [S.l.: s.n.], 2014. p. 3696–3701. ISSN 1930-529X. Citado na página 65.

YIN, B.; WU, M.; WANG, G.; DICK, C.; CAVALLARO, J. R.; STUDER, C. A 3.8Gb/s Large-Scale MIMO Detector for 3GPP LTE-Advanced. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2014. p. 3879–3883. ISSN 1520-6149. Citado na página 79.

ZHANG, W.; QIAO, S.; WEI, Y. HKZ and Minkowski Reduction Algorithms for Lattice-Reduction-Aided MIMO Detection. *IEEE Transactions on Signal Processing*, v. 60, n. 11, p. 5963–5976, Nov 2012. ISSN 1053-587X. Citado na página 50.

ZHOU, B.; JIANG, L.; ZHAO, S.; HE, C. BER Analysis of TDD Downlink Multiuser MIMO Systems with Imperfect Channel State Information. *EURASIP Journal on Advances in Signal Processing*, v. 2011, n. 1, p. 104, Nov 2011. ISSN 1687-6180. Disponível em: <https://doi.org/10.1186/1687-6180-2011-104>. Citado na página 99.

APPENDIX A – List of Publications

Journal Articles

- J. Minango, and C. de Almeida, “Low-Complexity Symbol-Flipping Detector for Symmetric Spatial Multiplexing MIMO Systems with Large Number of Antennas”, *submitted to a scientific journal*, 2019.
- J. Minango, D. Altamirano, and C. de Almeida, “Performance Difference Between Zero-Forcing and Maximum Likelihood Detectors in Massive MIMO Systems”, *IEEE Electronics Lett.*, 54(25): 1464-1466, Dec. 2018.
- J. Minango, and C. de Almeida, “Low Complexity Zero Forcing Detector Based on Newton-Schultz Iterative Algorithm for Massive MIMO Systems”, *IEEE Trans. Veh. Technol.*, 67(12): 11759-11766, Dec. 2018.
- J. Minango, and C. de Almeida, “Achieving Optimum Performance with ZF Detector by Using a Novel Precoding Approach in Spatial Multiplexing MIMO Systems”, *Elsevier Physical Commun.*, 31: 62-68, Dec. 2018.
- J. Minango, and C. de Almeida, “Optimum and Quasi-Optimum Relaxation Parameters for Low-Complexity Massive MIMO Detector Based on Richardson Method”, *IEEE Electronics Lett.*, 53(16): 1114-1115, Aug. 2017.

Conference Articles

- J. Minango, and C. de Almeida, “A Low-Complexity Linear Precoding Algorithm Based on Jacobi Method for Massive MIMO Systems” in *IEEE 87th Veh. Technol. Conf.*, Porto, Portugal, Jun. 2018.
- J. Minango, and C. de Almeida, “Low-Complexity MMSE Detector Based on Refinement Gauss-Seidel Method for Massive MIMO Systems” in *IEEE 9th Latin-American Conf. Commun.*, Guatemala City; Guatemala, Nov. 2017.
- J. Minango, and C. de Almeida, “Low-Complexity MMSE Detector Based on the First-Order Neumann Series Expansion for Massive MIMO Systems” in *IEEE 9th Latin-American Conf. Commun.*, Guatemala City; Guatemala, Nov. 2017.
- J. Minango, D. Altamirano, and C. de Almeida, “Low-complexity MMSE detector for massive MIMO systems based on damped Jacobi Method” in *IEEE 28th Int.*

Symposium on Personal, Indoor and Mobile Radio Commun., Montreal, Canada, Oct. 2017.

APPENDIX B – Derivation of Equation (6.4)

Here, it is proved that:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\hat{\mathbf{x}} \in M^{N_T}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2 \triangleq \arg\min_{\hat{\mathbf{x}} \in M^{N_T}} \left\{ (\mathbf{r} - \hat{\mathbf{x}})^H \mathbf{H}^H \mathbf{H} (\mathbf{r} - \hat{\mathbf{x}}) \right\}, \tag{B.1}$$

where $\mathbf{r} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$ is the unconstrained ZF solution.

From (B.1), the Euclidean distance $\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2$ of the ML detector in (2.40) can be expanded as:

$$\begin{aligned}
\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2 &= \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}} - \mathbf{H}\mathbf{r} + \mathbf{H}\mathbf{r}\|^2 \\
&= (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}} - \mathbf{H}\mathbf{r} + \mathbf{H}\mathbf{r})^H (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}} - \mathbf{H}\mathbf{r} + \mathbf{H}\mathbf{r}) \\
&= \{(\mathbf{y} - \mathbf{H}\mathbf{r})^H + (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}})^H\}\{(\mathbf{y} - \mathbf{H}\mathbf{r}) + (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}})\} \\
&= (\mathbf{y} - \mathbf{H}\mathbf{r})^H (\mathbf{y} - \mathbf{H}\mathbf{r}) + (\mathbf{y} - \mathbf{H}\mathbf{r})^H (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}}) \\
&\quad + (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}})^H (\mathbf{y} - \mathbf{H}\mathbf{r}) + (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}})^H (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}}) .
\end{aligned} \tag{B.2}$$

Since $\mathbf{r}$ is the ZF solution, then:

$$(\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}})^H (\mathbf{y} - \mathbf{H}\mathbf{r}) = (\mathbf{y} - \mathbf{H}\mathbf{r})^H (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}}) = 0. \tag{B.3}$$

Thus (B.2) reduces to:

$$\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2 = (\mathbf{y} - \mathbf{H}\mathbf{r})^H (\mathbf{y} - \mathbf{H}\mathbf{r}) + (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}})^H (\mathbf{H}\mathbf{r} - \mathbf{H}\hat{\mathbf{x}}) . \tag{B.4}$$

Substituting $\mathbf{r}$ with $(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$ into (B.4), it is obtained that:

$$\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2 = \{\mathbf{y} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}\}^H\{\mathbf{y} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}\} + (\mathbf{r} - \hat{\mathbf{x}})^H \mathbf{H}^H \mathbf{H} (\mathbf{r} - \hat{\mathbf{x}}) . \tag{B.5}$$

Since

$$\mathbf{y} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y} = \{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\}\mathbf{y}, \tag{B.6}$$

the first term of (B.5) becomes:

$$
\begin{aligned}
&\{\mathbf{y} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}\}^H\{\mathbf{y} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}\} \\
&\quad = \mathbf{y}^H\{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\}^H\{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\}\mathbf{y} \\
&\quad = \mathbf{y}^H\{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-H}\mathbf{H}^H\}\{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\}\mathbf{y} \\
&\quad = \mathbf{y}^H\{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-H}\mathbf{H}^H + \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-H}\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\}\mathbf{y} \\
&\quad = \mathbf{y}^H\{\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H + \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\}\mathbf{y} \\
&\quad = \mathbf{y}^H\{\mathbf{I}_{N_R} - 2\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H + (\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H)^2\}\mathbf{y} \\
&\quad = \mathbf{y}^H\{(\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H)^2\}\mathbf{y},
\end{aligned}
\tag{B.7}
$$

where it is used that $(\mathbf{H}^H\mathbf{H})^{-1} = (\mathbf{H}^H\mathbf{H})^{-H}$ once $\mathbf{H}^H\mathbf{H}$ is a SPD matrix. Then, substituting (B.7) into (B.5), it can be show that:

$$
\|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2 = \mathbf{y}^H\{(\mathbf{I}_{N_R} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H)^2\}\mathbf{y} + (\mathbf{r} - \hat{\mathbf{x}})^H \mathbf{H}^H\mathbf{H}\,(\mathbf{r} - \tilde{\mathbf{x}})\,. \tag{B.8}
$$

Note that the first term on the right side of (B.8) is constant with respect to $\hat{\mathbf{x}}$. Hence, this term can be ignored on the minimization. Thus, the equivalence in (B.1) immediately follows:

$$
\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\hat{\mathbf{x}} \in M^{N_T}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|^2 \triangleq \arg\min_{\hat{\mathbf{x}} \in M^{N_T}} \left\{(\mathbf{r} - \hat{\mathbf{x}})^H \mathbf{H}^H\mathbf{H}\,(\mathbf{r} - \hat{\mathbf{x}})\right\}, \tag{B.9}
$$

which completes the proof.