On Lattice Sequential Decoding for Large MIMO Systems

Thesis by

Konpal Ali

In Partial Fulfillment of the Requirements

For the Degree of

Masters of Science

King Abdullah University of Science and Technology, Thuwal,

Kingdom of Saudi Arabia

April, 2014

The thesis of Konpal Ali is approved by the examination committee

Committee Chairperson: Mohamed-Slim Alouini

Committee Member: Tareq Al-Naffouri

Committee Member: Ahmed Sultan Saleem

Committee Member: George Turkiyyah

ABSTRACT

On Lattice Sequential Decoding for Large MIMO Systems

Konpal Ali

Due to their ability to provide high data rates, Multiple-Input Multiple-Output (MIMO) wireless communication systems have become increasingly popular. Decoding of these systems with acceptable error performance is computationally very demanding.

In the case of large overdetermined MIMO systems, we employ the Sequential Decoder using the Fano Algorithm. A parameter called the bias is varied to attain different performance-complexity trade-offs. Low values of the bias result in excellent performance but at the expense of high complexity and vice versa for higher bias values. We attempt to bound the error by bounding the bias, using the minimum distance of a lattice. Also, a particular trend is observed with increasing SNR: a region of low complexity and high error, followed by a region of high complexity and error falling, and finally a region of low complexity and low error. For lower bias values, the stages of the trend are incurred at lower SNR than for higher bias values. This has the important implication that a low enough bias value, at low to moderate SNR, can result in low error and low complexity even for large MIMO systems. Our work is compared against Lattice Reduction (LR) aided Linear Decoders (LDs). Another impressive observation for low bias values that satisfy the error bound is that the Sequential Decoder's error is seen to fall with increasing system size, while it grows

for the LR-aided LDs.

For the case of large underdetermined MIMO systems, Sequential Decoding with two preprocessing schemes is proposed 1) Minimum Mean Square Error Generalized Decision Feedback Equalization (MMSE-GDFE) preprocessing 2) MMSE-GDFE preprocessing, followed by Lattice Reduction and Greedy Ordering. Our work is compared against previous work which employs Sphere Decoding preprocessed using MMSE-GDFE, Lattice Reduction and Greedy Ordering. For the case of large systems, this results in high complexity and difficulty in choosing the sphere radius. Our schemes, particularly 2), perform better in terms of complexity and are able to achieve almost the same error curves, depending on the bias used.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my adviser, Dr. Mohamed-Slim Alouini, for his guidance and mentoring. Also, I want to thank Dr. Walid Abediseid, for his advice and assistance during the course of my work. I want to acknowledge my thesis committee members: Dr. Ahmed Sultan Salem, for his devotion and insight on any queries I had; Dr. Tareq Al-Naffouri, for his very valuable feedback and meticulousness in going through my thesis; and Dr. George Turkiyyah, for his interest in my work and his constant encouragement. I truly appreciate all of your help and guidance.

I want to express my gratitude to my loving family. My father, for always encouraging me to pursue my dreams and doing everything possible to help me achieve my goals. My mother, for her constant support, advice and for listening to me whine. My brother Owais, for being the person I can always rely on in every situation. My brother Taha, for never failing to cheer me.

Finally, I want to thank my friends for making these last two years so amazing, for their disparate advice on disparate situations, and for being so awesome. In particular I want to thank (in alphabetical order): Amber, Annie, Basmah, Farhan, Hanan, Huma, Idris, Iqra, Islam, Rishabh, and Rkia, for always being there.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Symbol	Meaning
AWGN	Additive White Gaussian Noise
CLLL	Complex Lenstra Lenstra Lovász
CLPS	Closest Lattice Point Search
CSI	Channel State Information
DLSD	Double Layer Sphere Decoder
D-SLB	Dual Shortest Longest Basis
D-SLV	Dual Shortest Longest Vector
ELR	Element-based Lattice Reduction
GSD	Generalized Sphere Decoder
i.i.d.	independent identically distributed
ILS	Integer Least Squares
LD	Linear Decoder
LLL	Lenstra Lenstra Lovász
LOS	Line Of Sight
LR	Lattice Reduction
MAP	Maximum A Posteriori
MD	Minimum Distance
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MMSE-DFE	Minimum Mean Square Error Decision Feedback Equalization
MMSE-GDFE	Minimum Mean Square Error Generalized Decision Feedback Equalization
PEP	Pair-wise Error Probability
PVC	Prevoting Cancellation
PVS	Postvoting Vector Selection
SISO	Single Input Single Output
ZF	Zero Forcing
ZF-DFE	Zero Forcing Decision Feedback Equalization

LIST OF SYMBOLS

Symbol	Meaning
$\lambda_{min}(\mathbf{H})$	minimum eigenvalue of matrix $\mathbf{H}$
$d^2_{min}(\mathbf{H})$	squared minimum Euclidean distance of constellation obtained from channel $\mathbf{H}$
M	number of transmit antennas
N	number of receive antennas
m	number of real unknowns, $m = 2M$
n	number of real equations, $n = 2N$
$\mathcal{Q}(.)$	quantization operation, translates any point to the nearest QAM constellation point
^{T}	transpose operation
^{H}	Hermitian operation
*	conjugate operation
†	Moore-Penrose pseudoinverse
$\Re[\mathbf{x}]$	real components of vector $\mathbf{x}$
$\Im[\mathbf{x}]$	imaginary components of vector $\mathbf{x}$
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{1}_{N\times 1}$	vector of dimension $N \times 1$ consisting of all ones
$\mathbb{Z}$	integer set
$\mathbb{C}$	complex field
$\mathbb{R}$	real numbers
$\mathbb{Z}[j]$	Gaussian integer ring with elements of the form $\mathbb{Z} + j\mathbb{Z}$

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background

The two goals of any communication system are achieving reliability and capacity, in other words, achieving high performance and high data rates, respectively. Multiple Input Multiple Output (MIMO) systems employ multiple antennas at the transmitter and receiver, which allow them to exploit the spatial dimension in order to achieve higher data rates and performance.

With the increasing demand of data rate in recent years, researchers have shown great interest in MIMO Systems. Unfortunately, a trade-off exists between achieving reliability and capacity. MIMO systems for instance, when used for improving data rates by sending different data streams on each transmit antenna - as opposed to sending the same data stream on multiple transmit antennas to achieve better performance - require very complex reliable detection methods, especially as the system size increases. The Maximum Likelihood (ML) decoder in particular, although optimal and therefore most reliable, suffers from exponential complexity in terms of the number of transmit antennas (M) and constellation size ($\mathcal{M}$). On the other hand, Linear Decoders (LDs) such as Zero-Forcing (ZF) and the Minimum Mean Square Error (MMSE) have only polynomial complexity and are thus widely adopted in a number of systems. In MIMO systems, however, these decoders result in very poor

performance compared to the ML decoder due to their sensitivity to ill-conditioned channel matrices.

Analysis in [1] shows that for MIMO V-BLAST systems with M transmit antennas and N receive antennas, conventional LDs such as ZF and MMSE, can only collect a diversity of $N - M + 1$, though they enjoy very low computational complexity. The ML decoder, on the other hand, collects receive-diversity. However, Lattice Reduction (LR) techniques used to aid LDs, achieve receive-diversity at the expense of a small increase in complexity. It is important to note that a gap does exist between the performance curves of these LR-aided LDs and the ML detector due to the suboptimality of the LDs.

In our work with overdetermined systems, i.e. systems where $N \geq M$, we employ the Sequential Decoder with the Fano Algorithm to not only achieve receive-diversity, but to decrease the gap that exists between the ML detectors performance and that of LR-aided LDs. By varying a parameter called the bias in the Sequential Decoder, we are able to attain a very good performance-complexity trade-off.

The case of underdetermined systems is more difficult to decode, as the MIMO system here has $N < M$ and is therefore equivalent to a linear system with a larger number of unknowns, M, than the number of equations, N. ML decoding can of course still be applied to achieve good performance, but since the complexity is exponential, like the overdetermined case, we seek other solutions for the decoding of this problem. A number of optimal and suboptimal strategies have been proposed for solving these problems. Chapter 4 covers some important past work for the decoding of underdetermined systems, and Chapter 5 presents our contribution on decoding strategies. We have again employed the Sequential Decoder using the Fano Algorithm for the detection. Since the problem is underdetermined, the rank of the channel matrix is less than the number of unknowns, and hence we require some preprocessing before the actual decoding can be done. We employ Minimum Mean Square Er-

ror - Generalized Decision Feedback Equalization (MMSE-GDFE) preprocessing to convert our system into a square system, with M equations and M unknowns.The effective channel matrix is then full rank, and we can proceed with the Sequential Decoder.

1.2 Fading Channels and Techniques to Achieve High Diversity

Fading Channels:

Fading channels refer literally to the fading of the signal power as a signal traverses through a wireless channel. The attenuation of signal strength is accounted for by a number of factors. As a signal radiates outward from its source and propagates towards the receiver, its strength falls off with the distance squared, assuming a straight path between the transmitter and receiver without obstacles. This is one very important component of fading known as 'Path Loss'. In reality since no path without obstacles exists between the transmitter and receiver, the signal suffers from an additional loss called 'Shadowing'. Shadowing occurs due to the absorption, reflection, scattering and diffraction of the signal by the obstacles in the path. The attenuation caused by Shadowing and Path Loss occur over relatively larger distances.

Often there is more than one path between the transmitter and receiver, and multiple copies of the signals are obtained at the receiver, each having traversed a different path and encountered different obstacles. These various paths are called multipaths and the associated signals are called multipath components (MPCs). Since the different multipath components have phase differences, having traversed different paths, they add both constructively and destructively at the receiver. This kind of addition results in a very fast variation in the amplitude of the signal component and is known as 'Fast Fading' or 'Fading due to Multipath Components'. Another way of

looking at fast fading is that it is the fading that occurs when the coherence time of the channel is much smaller than the symbol period. This results in rapid changing of the channel impulse response during a symbol interval. Multipath variations, unlike Path Loss and Shadowing, occur over very small distances (of the order of the signal wavelength).

Another type of fading known as 'Flat Fading' occurs when the coherence bandwidth of the channel is much larger than the bandwidth of the signal being transmitted. As a result the channel response is composed of a constant gain and a linear phase. Flat fading results in a decreased SNR at the receiver. Two of the most commonly used fading models are the Rayleigh Fading model and the Rician Fading model. Rayleigh fading assumes there is no Line of Sight (LOS) component in the different multipath components at the receiver and thus represents the worst case fading scenario. Its power is exponentially distributed and phase is uniform. The phase and the amplitude in Rayleigh fading are independent of one another. Rayleigh fading is the most commonly used model as most practical scenarios (especially urban) do not have an LOS component. Rician fading on the other hand does include the LOS component(s) in addition to the other MPC's.

Techniques to Achieve High Diversity Order:

Diversity techniques are used to improve performance in a communication system that undergoes severe degradation due to fading. It involves the usage of multiple signal paths, each of which fades independently, to ensure reliable communication as long as one of the paths is strong enough. In other words, the different techniques of diversity aim to reduce the average error probability of the system by increasing the number of independent channels between the transmitter and receiver over which the information is transmitted. Another useful interpretation of diversity is that it is the power with which the SNR falls in the average error probability expression of the system.

A number of techniques can be used to achieve a high diversity order. The fundamentals behind achieving high diversity order can be explained by the following:

Time diversity involves achieving diversity by averaging the fading channel over time. The basic idea behind achieving time diversity is to transmit information multiple times, with the transmission times separated by at least the coherence time of the channel. This way each copy of the information will experience an independent realization of the channel. The price paid for this reduction in error probability is the delay associated with decoding and the reduced data rate. The data rate, though, can be improved by employing different coding techniques.

Frequency Diversity involves transmitting the information over multiple frequency bands that are separated by at least the coherence bandwidth of the channel. Like the case of time diversity, this separation of the frequency bands results in each transmission encountering an independent channel realization. The price paid for the reduction of error here is the expense of more bandwidth.

Spatial Diversity, aka Antenna Diversity, involves transmitting the information using multiple antennas at the transmitter and/or receiver. The antennas need to be placed sufficiently apart in order to attain independent channels. The price paid is in the form of the hardware complexity associated with the use of the multiple antennas. Receiver diversity involves employing multiple antennas at the receiver only, while a single antenna is used at the transmitter. The essence of diversity lies in the receiving of higher total SNR because of the multiple signal copies received. The diversity gain is of the order of the number of receiving antennas. Transmit diversity on the other hand requires a single antenna at the receiver but multiple antennas at the transmitter. When CSI (Channel Side Information) is known at the transmitter, the essence of diversity is in allocating a higher transmit-power to the antennas which correspond to better channels, and thus SNR is maximized this way at the receiver. On the other hand when CSI is unknown at the transmitter, combinations of space

and time diversity are required at the receiver. The Alamouti scheme is one such example of space and time diversity when the CSI is unknown at the transmitter.

1.3 MIMO, Multiuser MIMO, and Massive MIMO

MIMO (Multiple Input Multiple Output) systems are systems that take advantage of the dimension of space in their transmission by employing multiple antennas at the transmitter and receiver. The spatial dimension is exploited to achieve reliability and capacity. Unfortunately, a trade-off exists between achieving reliability and capacity, and we can maximize either one while minimizing the other or have some combination of the two.

Using MIMO to enhance reliability is known as Spatial Diversity. Reliability refers to the reduced probability of error that comes with employing a MIMO system. Redundant streams of data are transmitted and received in parallel along independent spatial paths. Since the probability of all the channels (the independent spatial paths) having a degraded performance is much lower than that of one channel being degraded, the probability of all the duplicates of the message being detected incorrectly is much lower than that of one message being detected incorrectly. Hence the system becomes more reliable. When MIMO is used to enhance reliability completely, the error probability may fall up to SNR^{-NM} compared to the error probability being inversely proportional to SNR in the case of Single Input Single Output (SISO) transmission systems.

Capacity on the other hand, is related to the improvement in data rate associated with employing MIMO. It involves sending different data streams over each of the independent channels at the same time and over the same frequencies, without distorting the information being transmitted. This is done to enhance throughput of the system. This technique is also referred to as Spatial Multiplexing. The spatial

degree of freedom attained by the use of MIMO changes the relationship between power and Shannon Capacity, and adding antennas has the potential to scale the capacity linearly.

MIMO technology was introduced so that multiple transmit and receive antennas could be used to simultaneously transmit multiple data streams. This results in enhanced data rates and reliability by making use of the additional degrees of freedom the propagation channel can provide by the additional antennas. The price paid for the enhanced performance is the increase in hardware complexity, energy consumption and signal processing required at both ends of the transmission scheme.

Multiuser MIMO (MU-MIMO) is MIMO used to communicate with different terminals simultaneously. It results in improvement of data rate, enhancement of reliability, improvement in energy efficiency and reduced interference, though all of these benefits cannot be achieved simultaneously.

Massive MIMO aims to achieve all the benefits of conventional MIMO but on a much larger scale, by scaling up MIMO by orders of magnitude. A very large number of antennas simultaneously serve a much smaller number of terminals, thus providing excess degrees of freedom. The benefits of massive MIMO are achieved by spatial multiplexing which depends on the base station having good knowledge of the channel on both the uplink and downlink. Hence, the number of terminals that can be served simultaneously is not limited by the number of antennas but by the inability to acquire channel state information for an unlimited number of terminals. An extensive survey on Massive MIMO systems was done in [2].

Massive MIMO results in enhanced capacity and significant improvement in the radiated energy efficiency. The enhanced capacity results from the spatial multiplexing used. It is limited, though, by the channel information available. The dramatic increase in energy efficiency occurs because the very large number of antennas available can be used to focus the energy into small regions of space with extreme sharpness,

as required.

It also has the additional advantage of requiring inexpensive, low-power components compared to traditional MIMO. While traditional MIMO requires fewer but a lot more expensive high power amplifiers to feed a few antennas; massive MIMO uses hundreds of extremely low cost amplifiers for its many antennas. The difference in expenditure is significant. This works because although massive MIMO reduces the accuracy and linearity of each individual amplifier, it relies on the Law of Large Numbers to average out the noise, fading and hardware imperfections when the signals from the many antennas are combined. Additionally, bulky items such as large, expensive coaxial cables are no longer required for massive MIMO. Moreover, since massive MIMO has an excess of unused degrees of freedom (= the number of antennas − the number of terminals), these can be made use of in hardware-friendly signal shaping.

By spreading information over the multiple antennas, the system is made more robust to jamming as well. This is not possible in a traditional MIMO system as there are no extra degrees of freedom provided by multiple antennas.

It should be mentioned, that although Massive MIMO renders many of the problems faced by conventional MIMO insignificant, it brings up some completely new issues that need to be dealt with.

1.4 System Model

Our system is one that employs multiple antennas at both the transmitter and receiver. Let M denote the number of transmit antennas and N the number of receive antennas. For the overdetermined case, $N \geq M$ and for the underdetermined case, $N < M$. The vectors $\mathbf{s}_c = [s_{c1},\ s_{c2}, \ldots, s_{cM}]^T$, $\mathbf{x}_c = [x_{c1},\ x_{c2}, \ldots, x_{cN}]^T$, and $\mathbf{w}_c = [w_{c1},\ w_{c2}, \ldots, w_{cN}]^T$ denote the transmit, receive, and noise vectors respectively.

The $N \times M$ matrix $\mathbf{H}_{c_1}$ represents the channel matrix. The noise is assumed to be Additive White Gaussian Noise (AWGN) and the channel to undergo Rayleigh fading; each component of the noise vector and the channel matrix is independent identically distributed (i.i.d.) and has a complex Gaussian distribution with zero mean and unit variance. The channel is also assumed to be stationary throughout a transmission block and to vary independently from block to block. The Channel Side Information (CSI) is assumed to be available at the receiver but not at the transmitter. The following equation describes the channel model:

$$\begin{aligned} \mathbf{x}_c &= \sqrt{\frac{D\ \ SNR}{M}}\mathbf{H}_{c_1}\mathbf{s}_c + \mathbf{w}_c && (1.1) \\ &= \mathbf{H}_c\mathbf{s}_c + \mathbf{w}_c, && (1.2) \end{aligned}$$

where SNR is the signal to noise ratio, D is a normalizing factor equal to $12/(\mathcal{M}-1)$, and $\mathbf{H}_c = \sqrt{D\ \ SNR/M}\mathbf{H}_{c_1}$. We have chosen the entries of $\mathbf{s}_c$ to be complex, independent and drawn from the $\mathcal{M}$-QAM constellation $\mathcal{S}$, where $\mathcal{M} = 2^{2k}$. The real and imaginary parts of $\mathbf{s}_c$ are therefore drawn from the $\sqrt{\mathcal{M}}$-PAM constellation (or equivalently the 2^k-PAM constellation), denoted by the set $\chi_k = \{\pm 1, \pm 3, \ldots, \pm(2^k - 1)\}$.

The vectors $\mathbf{s}$, $\mathbf{x}$ and $\mathbf{w}$ are the vectors $\mathbf{s}_c$, $\mathbf{x}_c$ and $\mathbf{w}_c$ respectively, but with their real components stacked on top of their imaginary components i.e.

$$\mathbf{s} = \begin{bmatrix} \Re[\mathbf{s}_c] \\ \Im[\mathbf{s}_c] \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \Re[\mathbf{x}_c] \\ \Im[\mathbf{x}_c] \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \Re[\mathbf{w}_c] \\ \Im[\mathbf{w}_c] \end{bmatrix}.$$

The counterpart of the matrix $\mathbf{H}_c$ is the matrix $\mathbf{H}$, where

$$\mathbf{H} = \begin{bmatrix} \Re[\mathbf{H}_c] & -\Im[\mathbf{H}_c] \\ \Im[\mathbf{H}_c] & \Re[\mathbf{H}_c] \end{bmatrix},$$

so that,

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w}. \tag{1.3}$$

Note: We will denote the number of real unknowns by m, where $m = 2M$, and the number of real equations by n, where $n = 2N$. Therefore $\mathbf{s}$ is of dimensions $m \times 1$, $\mathbf{x}$ and $\mathbf{w}$ of $n \times 1$ and $\mathbf{H}$ of $n \times m$.

Notation: Throughout this thesis, vectors will be denoted by lower bold faced letters and matrices by upper bold faced letters. The superscript ^{T} denotes the transpose, * denotes the conjugate, ^{H} denotes the Hermitian, and † denotes the Moore-Penrose pseudoinverse. Real and imaginary parts are denoted by $\Re[\ .\ ]$ and $\Im[\ .\ ]$ respectively. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix and $\mathbf{1}_{N\times 1}$ the vector of dimension $N \times 1$ consisting of all ones. $\mathbb{Z}$ denotes the the integer set, $\mathbb{C}$, the complex field, $\mathbb{R}$ for real numbers, and $\mathbb{Z}[j]$ for the Gaussian integer ring with elements of the form $\mathbb{Z} + j\mathbb{Z}$.

Chapter 2

Decoding Techniques

2.1 Well known Decoding Techniques

2.1.1 Optimal Decoding

Assume an AWGN channel, i.e. a channel that only suffers from noise but no channel distortion, so that the transmission system is,

$$\mathbf{x}_c = \mathbf{s}_c + \mathbf{w}_c, \tag{2.1}$$

where $\mathbf{w}_c$ is $\mathcal{N}(0, N_0/2)$.

Since we have available at the receiver the received signal $\mathbf{x}_c$, and $\mathbf{w}_c$ is known to have a zero-mean Gaussian distribution; logically speaking, a decoding strategy that searches for the message that is most probable, given the received signal $\mathbf{x}_c$, is the optimal strategy. Equivalently this can be written as,

$$\hat{\mathbf{s}}_c = \max_{\mathbf{s}_c \in \mathcal{S}^N} P(\mathbf{s}_c|\mathbf{x}_c). \tag{2.2}$$

This is known as the maximum a posteriori (MAP) detector, and is the optimal decoding strategy. If we do consider a non-identity channel matrix, the transmitted

point can be thought of as the lattice point $\mathbf{H}_c\mathbf{s}_c$. Since the channel matrix is assumed to be known at the receiver, and is independent of the transmit vector $\mathbf{s}_c$, the MAP rule can be written as,

$$\hat{\mathbf{s}}_c = \max_{\mathbf{s}_c \in \mathcal{S}^N} P(\mathbf{H}_c\mathbf{s}_c|\mathbf{x}_c). \tag{2.3}$$

Note that the maximization is still over the different possible $\mathbf{s}_c$, and independent of $\mathbf{H}_c$, so the MAP rule is still equivalent to the case of the MAP rule for an AWGN channel. It is important to realize that the decision being made is affected by the Gaussian noise $\mathbf{w}_c$, and not the channel matrix $\mathbf{H}_c$. This lies in the fact that $\mathbf{H}_c$ is known at the receiver, but the noise $\mathbf{w}_c$ is not.

From Bayes rule, the conditional probability can be re-written as

$$P(\mathbf{s}_c|\mathbf{x}_c) = \frac{P(\mathbf{x}_c|\mathbf{s}_c)P(\mathbf{s}_c)}{P(\mathbf{x}_c)}. \tag{2.4}$$

Since the denominator is independent of $\mathbf{s}_c$, the maximization for optimum detection can be done over the numerator only, i.e.

$$\hat{\mathbf{s}}_c = \max_{\mathbf{s}_c \in \mathcal{S}^N} P(\mathbf{x}_c|\mathbf{s}_c)P(\mathbf{s}_c). \tag{2.5}$$

If the probability of transmission of each message $\mathbf{s}_c$, is the same, then $P(\mathbf{s}_c)$ is a constant and independent of the particular transmitted message. For transmission schemes with equi-probable transmit messages, we can therefore ignore the last term in the maximization, and use the Maximum Likelihood (ML) rule for optimal detection instead, described as follows:

$$\hat{\mathbf{s}}_c = \max_{\mathbf{s}_c \in \mathcal{S}^N} P(\mathbf{x}_c|\mathbf{s}_c). \tag{2.6}$$

Since the additive noise in the transmission schemes being considered are Gaussian, the ML rule can further be simplified to a Minimum Distance (MD) rule. The $P(\mathbf{x}_c|\mathbf{s}_c)$ is a Gaussian distribution centered at the transmit symbol $\mathbf{s}_c$, with variance equal to that of the noise component $\mathbf{w}_c$'s.

$$P(\mathbf{x}_c|\mathbf{s}_c) = \frac{1}{(\pi N_0)^{N/2}} \exp\left(- \frac{||\mathbf{x}_c - \mathbf{s}_c||^2}{N_0} \right). \tag{2.7}$$

Since the only term dependent on the particular transmit vector $\mathbf{s}_c$ is $||\mathbf{x}_c - \mathbf{s}_c||^2$, maximizing $P(\mathbf{x}_c|\mathbf{s}_c)$ is equivalent to minimizing the squared distance $||\mathbf{x}_c - \mathbf{s}_c||^2$. For channels with AWGN, and equi-probable transmit messages, we can therefore use the Minimum Distance rule given by,

$$\hat{\mathbf{s}}_c = \min_{\mathbf{s}_c \in \mathcal{S}^N} ||\mathbf{x}_c - \mathbf{s}_c||^2, \tag{2.8}$$

for optimal decoding.

In our work, we will be dealing with systems that transmit all possible messages with equal probability. The terms 'optimal decoding' and 'ML decoding' will therefore be used interchangeably.

Since we deal with transmit vectors containing integers only, the MD rule in our work, can equivalently be thought of as a box constrained Integer Least Squares (ILS) problem.

2.1.2 Linear Decoders

Linear Decoders are a class of suboptimal decoders that have low computational complexity. They perform fairly well for SISO systems, but in the case of MIMO systems their performance may not be acceptable. We will briefly discuss the working of the following LDs:

Zero Forcing: The Zero Forcer aims to find a matrix $\mathbf{W}_c$, such that $\mathbf{W}_c\mathbf{H}_c = \mathbf{I}$.

The Zero Forcing linear detector that meets this constraint is the Moore-Penrose pseudoinverse given by, $\mathbf{H}_c^{\dagger} = (\mathbf{H}_c^H \mathbf{H}_c)^{-1}\mathbf{H}_c^H$. The corresponding estimate of the transmitted vector is found by,

$$\hat{\mathbf{s}}_{c_{ZF}} = \mathcal{Q}(\mathbf{H}_c^{\dagger}\mathbf{x}_c).$$

The quantization operation is required as the transmitted message belonged to an $\mathcal{M}$-QAM constellation, and the vector obtained by multiplying the received message $\mathbf{x}_c$ with $\mathbf{H}_c^{\dagger}$, does not in general belong to the $\mathcal{M}$-QAM constellation. Even if we were not considering boundary control, the quantization operation would be replaced by a rounding operation, to limit the values of $\mathbf{H}_c^{\dagger}\mathbf{x}_c$ to integers.

Since the off-diagonal elements of the matrix $(\mathbf{H}_c^H \mathbf{H}_c)$ are not zero, the ZF technique tries to null the interference caused on the term being detected by the other transmit components. This often results in noise amplification and hence accounts for the poor performance of ZF detection in the case of large MIMO systems.

Minimum Mean Square Error: As the name suggests, this technique aims to find a matrix $\mathbf{W}_c$ that minimizes the criteria $E\big[(\mathbf{W}_c\mathbf{x}_c - \mathbf{s}_c)^H(\mathbf{W}_c\mathbf{x}_c - \mathbf{s}_c)\big]$. Solving this minimization results in the matrix $\mathbf{W}_c = (\mathbf{H}_c^H\mathbf{H}_c + \mathbf{I})^{-1}\mathbf{H}_c^H$. The MMSE estimate of the transmit vector is obtained as follows,

$$\hat{\mathbf{s}}_{c_{MMSE}} = \mathcal{Q}(\mathbf{W}_c\mathbf{x}_c).$$

The quantization operation is applied for the same reasons as in the ZF case.
Equivalently, this can be solved by using the Moore-Penrose pseudoinverse of the extended channel matrix, $\bar{\mathbf{H}}_c \triangleq \begin{bmatrix} \mathbf{H}_c \\ \mathbf{I}_M \end{bmatrix}$ and multiplying it with the extended receive

vector $\bar{\mathbf{x}}_c \triangleq \begin{bmatrix} \mathbf{x}_c \\ \mathbf{0}_{M\times 1} \end{bmatrix}$ i.e. by applying the ZF technique to the extended channel system $\bar{\mathbf{x}}_c = \bar{\mathbf{H}}_c \mathbf{s}_c + \bar{\mathbf{n}}_c$.

2.2 Lattices

A complex lattice, λ, is a set of N dimensional vectors or points in the complex Euclidean space $\mathbb{C}^N$. This can be represented by, $\{\sum_{i=1}^{M} a_i \mathbf{b}_i : a_i \in \mathbb{Z}[j], \mathbf{b}_i \in \mathbb{C}^N\}$, where a_i are Gaussian integers, and $\mathbf{b}_i$ are N dimensional complex vectors. In the real case, a_i are integers, $\mathbf{b}_i$ are real vectors, and the lattice therefore belongs to the real Euclidean space.

Each lattice λ, is therefore generated by a linear combination of a set of linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M \in \lambda$ with Gaussian integer coefficients. The set of vectors $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M\}$ is called a basis of λ, and the matrix $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M]$, which has the basis vectors as its columns and is therefore of dimensions $N \times M$, is called the generator matrix of λ.

The basis of a lattice is not unique, and in fact an infinite number of bases exist for a particular lattice. A new basis can be obtained by multiplying the generator matrix with an $M \times M$ unimodular matrix. This will be discussed in the next section.

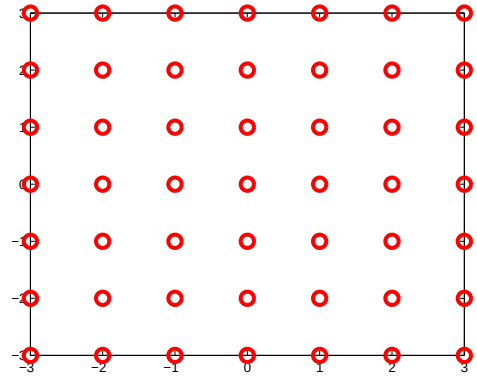
Since a_i are Gaussian integers, lattice decoding schemes result in an output that is an estimate of the transmitted Gaussian integer vector. In other words, lattice decoding schemes work with transmit vectors belonging to the infinite Gaussian integer space, and so do not have boundary control. Since our transmission schemes generally employ QAM constellations, and the transmit message therefore belongs to a finite set of odd Gaussian integers, a simple transformation is required before decoding so that the transmitted message is a Gaussian integer vector containing odd and even numbers. Also, the possible transmit vectors are limited to to a finite set; the out-

put of the lattice decoder, however, is not. Lattice Decoding therefore needs to be followed by a transformation back to the odd integer space and then a quantization step.

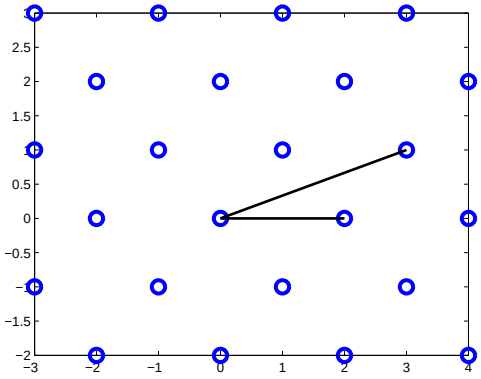
2.3 Lattice Reduction Techniques

A vector of information symbols transmitted undergoes distortion by its surroundings; this is modeled as the multiplication of the transmit vector with a channel matrix. It also experiences the addition of noise, which is modeled by the addition of an AWGN vector. If at the receiver we consider noiseless transmission and thus ignore the AWGN component, the received vector, $\mathbf{H}_c\mathbf{s}_c$, can be viewed as a lattice point and the corresponding received constellation as the lattice, with the columns of the channel matrix as its basis vectors. In other words, the noiseless received vector can be interpreted as a point in an M dimensional lattice, $\lambda(\mathbf{H}_c)$. The received constellation is thus not uniformly distributed, and may in fact have a very uneven distribution depending on the channel matrix $\mathbf{H}_c$. This makes decoding of this irregular lattice a very difficult problem, as the minimum distance required for erroneous detection, d_{min}, may be much smaller than in the case of the original problem, hence increasing the probability of erroneous detection.

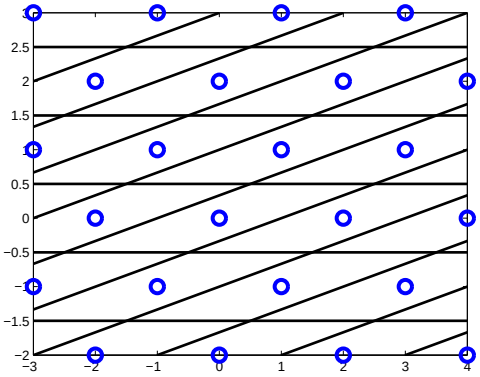
A very good example to illustrate this is presented in [3]; by considering a 2×2 real channel matrix, $H = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$. Figure 2.1a shows part of a large integer constellation that is transmitted to obtain the noiseless received constellation in Figure 2.1b. The column vectors of the channel matrix are shown in the figure. Figures 2.1c and 2.1d show the new decision boundaries for the ZF detector and the optimal ML detection respectively. As can be seen from the figures, multiplication of the integer constellation with the channel matrix results in smaller d_{min} particularly for the suboptimal detection schemes; this explains the drop in performance for these



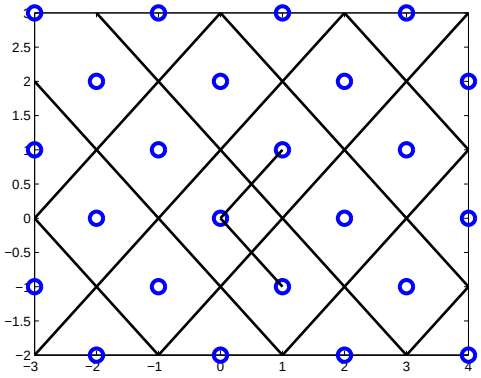
(a) Part of transmit integer constellation.



(b) Noiseless received constellation.



(c) Recieved constellation with Zero Forcing boundaries.



(d) Recieved constellation with optimal Maximum Likelihood decoding boundaries.

Figure 2.1: Transmit and noiseless-receive constellations with decision boundaries.

schemes.

If the bases of the lattice are more orthogonal, the noiseless receive constellation has a more regular shape, and the boundary regions of the suboptimal detectors are a lot closer to that of optimal detection. This results in better performance of the suboptimal detectors. However, in the case of MIMO systems, particularly with increasing system size, the columns of the channel matrix $\mathbf{H}_c$, are not in general uncorrelated and may, in fact, have very high correlation. This results in a very skewed lattice at the receiver. Detection using low-complexity schemes thus results in very poor performance, as the d_{min} may be very small.

For any lattice λ, there are an infinite number of possible bases, some of which

are more orthogonal than others. Since changing the lattice basis does not change the lattice, the original detection problem being solved remains unaltered. This leads to the notion of performing lattice reduction, to transform the given problem into a problem with a more orthogonal basis and then applying low-complexity detection schemes to the altered basis. The solution obtained in this reduced basis can then be transformed back to the original basis. This way performing low-complexity detection in a 'better' bases results in significant performance improvements. In fact, linear detectors when applied to reduced-lattice bases result in achieving receive-diversity.

In particular we aim to find reduced lattice bases that have more orthogonal and shorter basis in order to get better decision boundaries for low-complexity suboptimal detection schemes. This is because matrices that are more orthogonal have 'better' properties, such as ease of inversion, than matrices that are singular or ill-conditioned. The condition number of a matrix can be used as a measure of the orthogonality of its columns, with a perfectly orthogonal matrix having a condition number of 1 and more singular matrices having higher condition numbers. Performing lattice reduction on matrices not only results in the deviation of the distribution of the condition number to fall, but in fact even the mean of the condition number decreases significantly and falls a lot closer to 1.

For a channel matrix $\mathbf{H}_c$, a change in lattice basis results in a new generator matrix, $\widetilde{\mathbf{H}}_c$, such that both matrices generate the same lattice and have different basis. The following relation holds between the two generator matrices:

$$\widetilde{\mathbf{H}}_c = \mathbf{H}_c\mathbf{T}, \tag{2.9}$$

where $\mathbf{T}$, of dimensions $M \times M$, is a unimodular matrix and so, by definition, for the complex channel matrix $\mathbf{H}_c$, all elements of matrices $\mathbf{T}$ and $\mathbf{T}^{-1}$ contain Gaussian integers only and the determinant of $\mathbf{T}$ is ± 1 or $\pm j$ (for real $\mathbf{H}_c$, all elements of

matrices $\mathbf{T}$ and $\mathbf{T}^{-1}$ contain integers only and the determinant of $\mathbf{T}$ is ± 1).
We can now translate our original transmission system to

$$\begin{aligned} \mathbf{x}_c &= (\mathbf{H}_c\mathbf{T})(\mathbf{T}^{-1}\mathbf{s}_c) + \mathbf{w}_c \quad (2.10) \\ &= \widetilde{\mathbf{H}}_c\widetilde{\mathbf{s}}_c + \mathbf{w}_c, \quad (2.11) \end{aligned}$$

where $\widetilde{\mathbf{s}}_c \triangleq \mathbf{T}^{-1}\mathbf{s}_c$.

Linear detection schemes, such as ZF and MMSE, are applied to the modified system to obtain $\widehat{\widetilde{\mathbf{s}}}_c$, which is an estimate of $\widetilde{\mathbf{s}}_c$. The estimate $\widehat{\widetilde{\mathbf{s}}}_c$ is then translated back by multiplying with $\mathbf{T}$ followed by quantization to obtain $\widehat{\mathbf{s}}_c$ belonging to the $\mathcal{M}$-QAM constellation, which is an estimate of $\mathbf{s}_c$.

$$\widehat{\mathbf{s}}_c = \mathcal{Q}(\mathbf{T}\widehat{\widetilde{\mathbf{s}}}_c). \quad (2.12)$$

As mentioned in Chapter 1, MIMO V-BLAST systems using conventional linear decoders such as ZF and MMSE, are only able to collect a diversity of $N - M + 1$, though they enjoy very low computational complexity. The exponential complexity ML decoder, on the other hand, collects receive-diversity. However, lattice reduction techniques used to aid LDs, such as those in [1], [4], [3], [5], [6], [7], [8], are able to achieve receive-diversity at the expense of a small increase in complexity. It is important to note that a gap does exist between the performance curves of these LR-aided LDs and the ML detector due to the suboptimality of the LDs; namely the imperfect orthogonalization of the channel matrix by lattice reduction and the imperfections introduced by the quantization operation. Other work such as [9] and [10] employ lattice reduction techniques as well, to achieve ML detection diversity.

Finding the optimal basis of a lattice is a very computationally demanding task and in fact, lattice reduction for high-dimensions is known to be an NP-hard problem. For smaller dimensional systems, optimal techniques can be used; for instance for the

case of two dimensional systems, the Gaussian Reduction Algorithm (GRA) is optimal as it generates the shortest basis vectors in a lattice. However, for larger system sizes, reduction techniques that result in lower complexity at the cost of suboptimal basis reduction are required. In fact, a lot of techniques exist in the literature that result in the suboptimal reduction of the basis to more orthogonal basis, with varying performance (in the sense of the amount of lattice reduction achieved) and complexity trade-offs.

One of the most popular reduction algorithms is the Lenstra Lenstra Lovász (LLL) algorithm in [11]. It has a polynomial complexity on average and results in suboptimal reduction. The algorithm works by guaranteeing to find the shortest basis vector up to an exponential factor by upper bounding the orthogonality defect. There exist other variations of this algorithm such as some Complex-LLL algorithms, the Dual-LLL algorithm etc.

Another popular alternative to the LLL algorithm is Seysen's reduction algorithm. Seysen's Algorithm (SA) simultaneously reduces the primal and dual basis by minimizing a metric called the Seysen's measure iteratively until a local minimum is found. Seysen's measure quantifies the orthogonality of the basis in both the primal and dual spaces; hence the minimization results in lattice reduction.

SA performs better lattice reduction than the LLL algorithm and so better error rates are achieved when decoding systems that have been reduced using SA, than systems that have been reduced using LLL. The difference is more obvious for larger MIMO systems. This performance improvement of the SA compared to the LLL, however, occurs at the higher computational cost of the lattice reduction using the SA, associated with the high complexity of each basis update that occurs. Both Seysen's measure and the orthogonality measure for the LLL achieve their minima when the reduced basis is orthogonal.

Other popular techniques for lattice reduction would include Minkowski reduction,

the Hermite criterion, the Korkine-Zolotareff (KZ) reduction, but we will not discuss these.

2.3.1 Element-based Lattice Reduction

In general, for large MIMO systems, the LLL algorithm does not result in good enough performance and the SA is too computationally expensive. Techniques are required that have lower complexity, whilst maintaining or enhancing performance. Element-based Lattice Reduction (ELR) algorithms are proposed to tackle this by having goals different from those of other reduction techniques.

It has been shown in [4] that the Pair-wise Error Probability (PEP) of incorrect detection of the i^{th} transmitted symbol, x_{c_i}, increases with the corresponding diagonal element $C_{i,i}$ of matrix $\mathbf{C}$, where $\mathbf{C} = ((\mathbf{H}_c^H)\mathbf{H}_c)^{-1}$. Similarly in the reduced basis the PEP of incorrect detection of the i^{th} transmitted symbol of $\widetilde{x}_{c_i}$ increases with the i^{th} diagonal element, $\widetilde{C}_{i,i}$, of matrix $\widetilde{\mathbf{C}}$ where

$$\begin{aligned} \widetilde{\mathbf{C}} &= ((\widetilde{\mathbf{H}}_c^H)\widetilde{\mathbf{H}}_c)^{-1} \\ &= \mathbf{T}^{-1}\mathbf{C}(\mathbf{T}^{-1})^H. \end{aligned}$$

Thus to reduce the PEP, the diagonal elements of $\mathbf{C}$ must be minimized. In particular, increasing SNR results in the largest diagonal element of $\mathbf{C}$ dominating the PEP; minimizing this is therefore of utmost importance. Two optimization problems can be formulated from the above analysis:

- The Dual-Shortest Longest Vector (D-SLV) reduction, which minimizes the largest diagonal element of $\widetilde{\mathbf{C}}$ by finding an appropriate unimodular matrix
- The Dual-Shortest Longest Basis (D-SLB) reduction, which also by finding an appropriate unimodular matrix, minimizes each diagonal element of $\widetilde{\mathbf{C}}$ in descending

order of value

Finding the optimum reduced-basis using the D-SLV and D-SLB is computationally very demanding. So instead ELR algorithms are proposed in [4] which iteratively calculate suboptimal solutions for the D-SLV and D-SLB reductions.

2.3.2 CLLL Reduction

The LLL algorithm described, works with real matrices and hence the dimensionality of the channel matrices is doubled as computations are carried out on $\mathbf{H}$ instead of $\mathbf{H}_c$. The Complex-LLL (CLLL) algorithm in [1], on the other hand, performs complex operations and can reduce complex matrices. As a result, the computational complexity of the CLLL algorithm is half of the LLL. Impressively, this reduction in complexity occurs without any performance loss.

2.4 Sphere Decoding

The optimal decoding strategy, the ML detection, involves minimizing the noise in the received signal. In other words the decoder's output $\hat{\mathbf{s}}$, is chosen corresponding to that value of $\mathbf{s}$ that results in the noiseless receive lattice point, $\lambda = \mathbf{Hs}$, closest to the received point $\mathbf{x}$ i.e.

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}\in\mathcal{S}^n} ||\mathbf{x} - \mathbf{Hs}||^2. \tag{2.13}$$

From [12], this minimization is equivalent to the search for the closest lattice point $\mathbf{Hs}$, to the received vector $\mathbf{x}$, and can therefore be thought of as a search for the point closest to the center, in an n dimensional sphere containing all possible lattice points $\mathbf{Hs}$, centered at $\mathbf{x}$. The search for the closest point to the center can be reduced from searching over all possible noiseless received vectors $\mathbf{Hs}$, to a smaller sphere that contains a subset of $\mathbf{Hs}$. This smaller sphere will include the lattice point closest

to $\mathbf{x}$, and the search space will be reduced, hence preserving the optimality of the decoding procedure but reducing the complexity of the search at the same time.

As we are to employ lattice decoding, boundary control needs to be given up by taking into consideration the entire n-dimensional integer space. This is done by replacing $\mathcal{S}^n$ with the infinite n-dimensional integer set, $\mathbb{Z}^n$.

Since we are to search over a subset of the entire space, by searching for the closest received lattice point inside the sphere centered at the received vector, we assume a sphere with a squared radius of C_0 that is large enough to include the closest received lattice point. The minimization problem can then be written as

$$||\mathbf{x} - \mathbf{Hs}||^2 \leq C_0. \tag{2.14}$$

2.4.1 Pohst Enumeration

Since the channel matrix, $\mathbf{H}$, can be equivalently written as $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q}$ is an $n \times m$ unitary matrix with orthonormal columns, and $\mathbf{R}$ an $m \times m$ upper triangular matrix, the minimization can be rewritten as

$$\begin{aligned} ||\mathbf{Q}^H(\mathbf{x} - \mathbf{Hs})||^2 &\leq C_1 \\ ||\mathbf{Q}^H\mathbf{x} - \mathbf{Q}^H(\mathbf{QR})\mathbf{s})||^2 &\leq C_1 \\ ||\mathbf{xD} - \mathbf{Rs}||^2 &\leq C_1, \end{aligned}$$

to obtain a new upper triangular system, where $\mathbf{xD} \triangleq \mathbf{Q}^H\mathbf{x}$ and C_1 is the radius of the new sphere formed by the altered system. The following inequalities hold true for $i = 1, \ldots, m$

$$\sum_{j=i}^{m} ||xD_j - \sum_{l=j}^{m} r_{j,l}s_l||^2 \leq C_1 \qquad 1 \leq i \leq m. \tag{2.15}$$

These inequalities can be used for backward substitution starting with $i = m$, the last level, and decoding each level of the possible allowed codewords within the sphere of radius $\sqrt{C_1}$. The inequality at each level i will give us an interval of admissible possibilities of s_i that lie within the sphere for the given set of symbols previously decoded $s_{i+1}, s_{i+2}, \ldots, s_m$. Let $\mathbf{s}_i^m = [s_i \ s_{i+1} \ldots s_m]$ denote the last $m - i + 1$ components of $\mathbf{s}$. Then for s_i, the corresponding admissible interval is given as a function of $\mathbf{s}_{i+1}^m$ by the range of integers $\mathcal{I}_i(\mathbf{s}_{i+1}^m)$, where

$$\mathcal{I}_i(\mathbf{s}_{i+1}^m) = [\mathcal{A}_i(\mathbf{s}_{i+1}^m), \ \ \mathcal{B}_i(\mathbf{s}_{i+1}^m)]$$

and

$$\mathcal{A}_i(\mathbf{s}_{i+1}^m) = \left\lceil \frac{1}{r_{i,i}} \left(xD_i - \sum_{j=i+1}^{m} r_{i,j} s_j - \sqrt{C_1 - \sum_{j=i+1}^{m} |xD_j - \sum_{l=j}^{m} r_{j,l} s_l|^2} \ \right) \right\rceil \quad (2.16)$$

$$\mathcal{B}_i(\mathbf{s}_{i+1}^m) = \left\lfloor \frac{1}{r_{i,i}} \left(xD_i - \sum_{j=i+1}^{m} r_{i,j} s_j + \sqrt{C_1 - \sum_{j=i+1}^{m} |xD_j - \sum_{l=j}^{m} r_{j,l} s_l|^2} \ \right) \right\rfloor. \quad (2.17)$$

When the lattice point lies outside of the sphere, $\sum_{j=i+1}^{m} \left| xD_i - \sum_{l=j}^{m} r_{j,l} s_l \right|^2 > C_1$. If the lattice point corresponding to the components $\mathbf{s}_{i+1}^m$ decoded already, exceeds the sphere of radius $\sqrt{C_1}$, then no values of possible s_i exist that will result in an m dimensional lattice point $\mathbf{Rs}$ that exists inside the sphere (equivalently no n dimensional lattice point $\mathbf{Hs}$ will exist inside the sphere of radius $\sqrt{C_0}$); hence $\mathcal{I}_i(\mathbf{s}_{i+1}^m)$ is empty.

Additionally, when $\mathcal{A}_i(\mathbf{s}_{i+1}^m) > \mathcal{B}_i(\mathbf{s}_{i+1}^m)$, there is no value of s_i that satisfies the inequalities in Eqn. (2.15) and hence no point $\mathbf{Hs}$, that has its last $m-i$ components equal to $\mathbf{s}_{i+1}^m$, exists inside the sphere.

This way, using Sphere Decoding at each level i, starting from $i = m$, we are able to span the admissible interval $\mathcal{I}_i(\mathbf{s}_{i+1}^m)$ and climb up to the level $i = 1$. At level $i = 1$,

a non-empty $\mathcal{I}_1(\mathbf{s}_2^m)$ corresponds to the vectors $\mathbf{s} = [s_1 \;\; \mathbf{s}_2^{mT}]^T$, where $s_1 \in \mathcal{I}_1(\mathbf{s}_2^m)$, that are the lattice points **Hs** that lie inside the sphere of radius $\sqrt{C_0}$ centered at **xD**.

2.4.2 Schnorr-Euchner Enumeration

While the Pohst Enumeration involves natural spanning of the possibilities of x_i by considering the values in the order $\mathcal{A}_i(\mathbf{s}_{i+1}^m), \mathcal{A}_i(\mathbf{s}_{i+1}^m) + 1, \ldots, \mathcal{B}_i(\mathbf{s}_{i+1}^m)$, the Schnorr-Euchner Enumeration involves starting in the middle of the interval $\mathcal{I}_i(\mathbf{s}_{i+1}^m)$, and zigzagging to the ends of the interval. The starting point of the Schnorr-Euchner enumeration for any x_i is,

$$\mathcal{S}_i(\mathbf{s}_{i+1}^m) = \left\lceil \frac{1}{r_{i,i}} \Big(xD_i - \sum_{j=i+1}^{m} r_{i,j} s_j \Big) \right\rfloor. \tag{2.18}$$

Hence the ordered sequence of x_i produced at any level is as follows,

$$x_i \in \{\mathcal{S}_i(\mathbf{s}_{i+1}^m),\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) + 1,\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) - 1,\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) + 2,\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) - 2,\ \ldots\} \bigcap \mathcal{I}$$

if

$$xD_i - \sum_{j=i+1}^{m} r_{i,j} s_j - r_{i,i} \mathcal{S}_i(\mathbf{s}_{i+1}^m) \geq 0,$$

and the ordered sequence

$$x_i \in \{\mathcal{S}_i(\mathbf{s}_{i+1}^m),\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) - 1,\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) + 1,\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) - 2,\ \mathcal{S}_i(\mathbf{s}_{i+1}^m) + 2,\ \ldots\} \bigcap \mathcal{I}(\mathbf{s}_{i+1}^m)$$

if

$$xD_i - \sum_{j=i+1}^{m} r_{i,j}s_j - r_{i,i}\mathcal{S}_i(\mathbf{s}_{i+1}^m) < 0.$$

It is interesting to note that unlike the Pohst Enumeration, the Schnorr-Euchner Enumeration allows one to set the sphere of the radius to infinity. This way the entire space is considered and the event of a sphere being declared empty never occurs. The complexity is, of course, high in this case as the algorithm spans a large number of points, due to the larger space being considered, before finding the ML solution. With $C_0 = \infty$, the first point found is the Babai point or the Zero-Forcing Decision-Feedback-Equalization (ZF-DFE) point. As it corresponds to the first value x_i found for each level i in the entire integer space, this point is also equivalently given by

$$\begin{aligned} x_i^{zf-dfe} &= \mathcal{S}_i(x_{i+1}^{zf-dfe}, \ldots, x_m^{zf-dfe}) && (2.19) \\ &= \left\lceil \frac{1}{r_{i,i}} \left(xD_i - \sum_{j=i+1}^{m} r_{i,j} x_j^{zf-dfe} \right) \right\rfloor, && (2.20) \end{aligned}$$

starting from $i = m$ and working back to $i = 1$.

2.4.3 On the Radii of Sphere Decoders

The radius of a Sphere Decoder is a very important parameter as it directly plays a role in the complexity of the decoding operation by varying the size of the space being considered. Choosing a large radius ensures finding the closest lattice point at the expense of high complexity. Smaller radii on the other hand may result in wasteful decoding if the sphere is too small, does not include any receive lattice point and is therefore declared empty; decoding in these cases will have to be repeated with a

larger radius. There exist techniques for choosing an initial radius, however, these are not robust methods in general and therefore neither guarantee a non-empty sphere, nor low complexity.

Another important factor to realize is that since there is no hard and fast rule for choosing the radius, the complexity of decoding will vary accordingly, and so in a sense complexity comparison with other decoding schemes may not exactly be fair. Furthermore the sphere size required varies with not only the system size, but also the SNR being used for transmission as this affects the spacing of the received constellation. In our simulations we have chosen radii by trial and error for each system size and changed the radii for different SNR ranges. This has been done to save simulation time, as a result the complexity of Sphere Decoding for our simulations is rather on the lower end. This ought to be taken into account when comparing the results for schemes that employ Sphere Decoding versus those that employ other decoding techniques.

2.5 Sequential Decoding

In our work we employ the suboptimal Sequential Decoder using the Fano Algorithm. Due to its suboptimality, the decoder has a much lower computational complexity than the ML decoder but an error performance that, although acceptable, is worse than that of the ML. From [14], the working of a Sequential decoder can be divided into two stages, namely the Preprocessing Stage followed by a Tree Search Stage. The purpose of the Preprocessing Stage is to tame the channel, make it sparser, and to put the problem in the form of a tree structure. Taming the channel involves a QR decomposition of the channel matrix so that the detection can be done recursively due to the upper triangular structure of the modified problem. MMSE-Decision Feedback Equalizer (MMSE-DFE) may also be applied to achieve better results at

the decision point. To induce sparsity in the modified problem, a lattice reduction may be applied to obtain an upper triangular structure more sparse than the original one. Permutation of the columns of the upper triangular matrix also results in increased sparsity. To ensure the problem has a tree structure it must be in an upper triangular form. The Tree Search Stage involves finding the best path in the tree of possible codewords.

Since we are trying to avoid high complexity, we have not focused much on the Preprocessing Stage and simply apply a QR decomposition to the channel matrix to put our problem in the form of an upper triangular structure so that the problem has a tree structure and detection can be done recursively. For the ease of analysis we will denote the upper triangular system as follows:

$$\underbrace{\begin{bmatrix} z_m \\ \vdots \\ z_1 \end{bmatrix}}_{\mathbf{z}} = \underbrace{\begin{bmatrix} r_{m,m} & \cdots & \cdots & r_{m,1} \\ 0 & r_{m-1,m-1} & \cdots & r_{m-1,1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{1,1} \end{bmatrix}}_{\mathbf{R}} \underbrace{\begin{bmatrix} s_m \\ \vdots \\ s_1 \end{bmatrix}}_{\mathbf{s}} + \underbrace{\begin{bmatrix} n_m \\ \vdots \\ n_1 \end{bmatrix}}_{\mathbf{n}},$$

where $\mathbf{z}$ represents the modified receive vector, $\mathbf{R}$ the modified channel, $\mathbf{s}$ the vector of symbols to be decoded, and $\mathbf{n}$ the modified noise vector.

This is followed by a Fano Search Stage which is a type of iterative search. The Fano Algorithm, unlike other search stage algorithms (e.g. the Stack Algorithm), requires almost negligible memory due its iterative nature. The price paid for the low memory requirements is the need to revisit nodes that the other algorithms do not.

In particular, the Fano Algorithm is a Best-First Search algorithm which means that at any level in the search stage, the algorithm chooses the best possible child node and then checks the validity of the newly formed path. In general, the tree structure of a code is used to decode the received sequence by making tentative hypotheses on successive branches of the tree. These hypotheses may be changed when subsequent ones indicate an error in the previous hypotheses.

Nodes are represented by $\mathbf{s}^k$, where $1 \leq k \leq m$ and $\mathbf{s}^k = [s_1\ s_2\ \ldots s_k]$, and the bias is denoted by b. Sequential decoding involves deciding a codeword that minimizes a certain function. The function is a bit metric called the path value and is denoted by $f(\mathbf{s}^k)$. The path value used in the Fano Algorithm of sequential decoding is

$$f(\mathbf{s}^k) = \sum_{j=1}^{k} w_j(\mathbf{s}^j) - bk,$$

where

$$w_i(\mathbf{s}^i) = \left|\left| z_i - \sum_{j=1}^{i} r_{i,j} s_j \right|\right|^2.$$

Note that $w_i(\mathbf{s}^i)$ is the component of the norm of the noise term at level i, and that at each level k, the Sphere Decoder searches for nodes which satisfy $\sum_{j=1}^{k} w_j(\mathbf{s}^j) < C_1$. It has been proved that at any decoding stage, extending the path with the smallest Fano metric minimizes the probability that the extending path does not belong to the optimal path, thereby justifying the use of the path value. Making such a 'locally' optimal decision at every decoding stage, however, does not guarantee finding the 'global' optimal path (i.e. the ML detectors result), and hence the error performance of the Sequential Decoder using the Fano metric is inferior to ML decoding. The dynamic threshold, T, is another metric and is constrained to change in increments of a fixed number Δ, called the step size. The changes in the value of T are determined by the algorithm which tightens and loosens the bound T as required. Since the Fano Algorithm is a Best-First Search, only the child node with the best Fano Metric is

considered in an iteration. If this metric is less than T, the child node is valid. For a valid child node the algorithm is terminated if the child node is a leaf node, otherwise tightening of the dynamic threshold T is done. If the child node isn't valid, the threshold will be increased if it is too small and if it isn't, the algorithm will move back a node and look for the next best child node.

The Sequential Decoder thus hypothesizes in such a way that the path value, $f(\mathbf{s}^k)$, is always less than the dynamic threshold T. If f is greater than T, and T is not too small, the decoder is on the wrong path and searching for a different path in the tree needs to be done.

The Fano Algorithm allows two types of movements from one node to another:

1. Forward − the decoder goes one branch to the right in the received value tree from the previously hypothesized node

2. Backward − moving one branch to the left in the received value tree when an incorrect hypothesis has been made, so that the next best child node can be found.

A record of the previous, current, and successor nodes and path metrics is kept (i.e. $\mathbf{s}^{k-1}$, $f(\mathbf{s}^{k-1})$, $\mathbf{s}^k$, $f(\mathbf{s}^k)$, $\mathbf{s}^{k+1}$ and $f(\mathbf{s}^{k+1})$) and the threshold T at every node. Initially (i.e. at $k = 0$), $\mathbf{s}^k$ is the origin and $T = 0$.

Chapter 3

Decoding of Large Overdetermined Systems using Lattice Decoding Techniques

3.1 Introduction

In this chapter we will discuss the case of overdetermined systems, i.e. systems with the number of receive antennas greater than or equal to the number of transmit antennas. This results in the mathematical model of the system having the number of equations larger than or equal to the number of unknowns. The channel is assumed to be known at the receiver end of the transmission system. Since the channel is random, decoding of such a problem, especially for larger system sizes can be quite tedious. We present in this chapter techniques, for decoding the transmitted symbol for large system sizes, i.e. as the number of antennas grows at both the transmit and receive side. The focus on large systems is due to the increasing demand of data rates, which can be fulfilled by the use of larger numbers of antennas.

As described before, the improvement in performance achieved by using LR-aided LDs is quite significant compared to that of un-aided LDs, particularly for MIMO

systems with increasing size. The price paid is the increase in complexity that comes with the lattice reduction process being employed, but the benefits of these schemes outweigh their cost. In essence, LR-aided LDs are able to achieve receive diversity at modest complexity. It is important to note, however, that despite being able to achieve receive diversity, there exists a gap between the performance of LR-aided LDs and the optimal ML detection. In our work with overdetermined systems, we employ the Sequential Decoder with the Fano Algorithm to not only achieve receive-diversity, but to decrease the gap that exists between the ML detector's performance and that of LR-aided LDs. By varying a parameter called the bias in the Sequential Decoder, we are able to attain a very good performance-complexity trade-off.

Our work with overdetermined systems will be compared against the Element-based Lattice Reduction (ELR) techniques used in [4] and the Complex Lenstra-Lenstra-Lovász (CLLL) reduction technique employed in [1] to decode MIMO systems. The bias is varied according to the transmit-to-receive-antenna ratio in order to upper-bound the error probability according to [15]. Increasing the bias reduces complexity but results in higher error rates and vice versa for decreasing the bias. This happens because as the bias is increased, the decoder approaches the ZF decoder, while decreasing the bias results in it approaching the optimal decoder. It is shown that the Sequential Decoder, like the CLLL and ELR, attains receive-diversity and, in fact, performs better than them.

3.2 Framework

Since the $m \times 1$ vector $\mathbf{s}$ contains elements drawn from $\mathcal{M}$-QAM constellation i.e. $\{\pm 1, \pm 3, \ldots, \pm(\sqrt{\mathcal{M}} - 1)\}$, $\mathbf{s}$ can be rewritten as $\acute{\mathbf{s}}$, where $\mathbf{s} = 2\acute{\mathbf{s}} - \mathbf{1}_{m\times 1}$, so that $\acute{\mathbf{s}}$ contains elements from $\mathbb{Z}^m$. This translation is required as the Sequential Decoder, being a lattice decoder, works with integer vectors. The original real system can thus

be expressed as

$$\begin{aligned} \mathbf{x} &= \mathbf{H}(2\acute{\mathbf{s}} - \mathbf{1}_{m\times 1}) + \mathbf{w} && (3.1)\\ &= 2\mathbf{H}\acute{\mathbf{s}} - \mathbf{H}\,\mathbf{1}_{m\times 1} + \mathbf{w}. && (3.2) \end{aligned}$$

The system can then be translated to

$$\begin{aligned} \acute{\mathbf{x}} &= \mathbf{x} + \mathbf{H}\,\mathbf{1}_{m\times 1} && (3.3)\\ &= 2\mathbf{H}\acute{\mathbf{s}} + \mathbf{w} && (3.4)\\ &= \acute{\mathbf{H}}\acute{\mathbf{s}} + \mathbf{w}. && (3.5) \end{aligned}$$

where $\acute{\mathbf{H}} = 2\mathbf{H}$. The set $\Lambda = \{\acute{\mathbf{H}}\acute{\mathbf{s}} : \acute{\mathbf{s}} \in \mathbb{Z}^m\}$ is an m dimensional lattice in $\mathbb{R}^n$. Applying a QR-decomposition to $\acute{\mathbf{H}}$, we obtain $\acute{\mathbf{Q}}$ and $\acute{\mathbf{R}}$. This is used to calculate

$$\begin{aligned} \mathbf{x}_D &= \acute{\mathbf{Q}}^H\acute{\mathbf{x}} && (3.6)\\ &= \acute{\mathbf{Q}}^H\acute{\mathbf{Q}}\acute{\mathbf{R}}\acute{\mathbf{s}} + \acute{\mathbf{Q}}^H\mathbf{w} && (3.7)\\ &= \acute{\mathbf{R}}\acute{\mathbf{s}} + \mathbf{w}_1, && (3.8) \end{aligned}$$

where $\mathbf{w}_1 = \acute{\mathbf{Q}}^H\mathbf{w}$. As the columns of $\acute{\mathbf{Q}}$ are unit vectors, the distribution of the elements of $\mathbf{w}_1$ is the same as that of $\mathbf{w}$. $\mathbf{x}_D$ is input to the Fano Algorithm in [14] along with $\acute{\mathbf{R}}$, the step size which is set to be equal to 1 for our work, and the bias. The bias is the parameter which is varied to obtain different performance-complexity trade-offs. The output of the Fano Algorithm is the $m \times 1$ vector $\widehat{\acute{\mathbf{s}}}$, consisting of integers. It is then translated back to obtain the $m \times 1$ vector $\widehat{\mathbf{s}}_1$ by the following

$$\widehat{\mathbf{s}}_1 = 2\widehat{\acute{\mathbf{s}}} - \mathbf{1}_{m\times 1}.$$

The vector $\widehat{\mathbf{s}}_1$ consists of the real and imaginary parts stacked on top of one another, these are then used to obtain the complex vector $\widehat{\mathbf{s}}_{c1}$ of dimensions $M \times 1$. Since the Fano Algorithm, being a lattice decoder and therefore not having any boundary control, outputs a vector containing integers from the infinite ring, whereas the original transmitted vector contains elements belonging to the $\mathcal{M}$-QAM constellation only, a quantization step is required to ensure the decoded symbols belong to the $\mathcal{M}$-QAM constellation. Hence $\widehat{\mathbf{s}}_c$, which is an estimate of $\mathbf{s}_c$, is obtained by $\widehat{\mathbf{s}}_c = \mathcal{Q}(\widehat{\mathbf{s}}_{c1})$.

In our work we also apply the Sequential Decoder to the extended MMSE system. In this case, the receive vector $\mathbf{x}$ and the channel matrix $\mathbf{H}$ used above, are simply replaced by their extended MMSE system counterparts $\bar{\mathbf{x}}$ and $\bar{\mathbf{H}}$, where $\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{0}_{m\times 1} \end{bmatrix}$ and $\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{I}_m \end{bmatrix}$. This way, we proceed with the preprocessing transformations in Eqns (3.1) to (3.8), as in the case of the original real system, and then apply the Sequential Decoder.

3.3 Minimum Eigenvalue of Channel Matrices

Consider a real matrix, $\mathbf{A}$ of dimensions $r \times t$, where each element of the matrix is i.i.d and has a Gaussian distribution with zero mean and unit variance. Let $r \geq t$, so that the ratio of the transmit to receive antennas, denoted by y, always satisfies $y \in (0, 1]$. Then from the Marchenko-Pastur Law, as the number of antennas is increased asymptotically for a fixed antenna ratio y, the ratio of the minimum eigenvalue of the corresponding matrix $\mathbf{A}^T\mathbf{A}$ to the number of transmit antennas, λ_{min}/t, approaches $(1 - \sqrt{y})^2$, i.e.

$$\lim_{t\to\infty} \frac{\lambda_{min}(\mathbf{A}^T\mathbf{A})}{t} \to (1 - \sqrt{y})^2 \qquad y \in (0, 1]$$

.

For a complex matrix, $\mathbf{A}_c$ of dimensions $N \times M$, with i.i.d. elements that have a Gaussian distribution with zero mean and unit variance, when $N \geq M$, the ratio of the transmit to receive antennas, $y = M/N$, still lies in the range $(0, 1]$. From [16], as the number of antennas is increased asymptotically for this matrix, keeping the antenna ratio y fixed, the ratio of the minimum eigenvalue of the matrix $\mathbf{A}_c^T \mathbf{A}_c$ to the number of transmit antennas, λ_{min}/M, approaches $2(1 - \sqrt{y})^2$, i.e.

$$\lim_{M \to \infty} \frac{\lambda_{min}(\mathbf{A}_c^H \mathbf{A}_c)}{M} \to 2(1 - \sqrt{y})^2 \qquad y \in (0, 1].$$

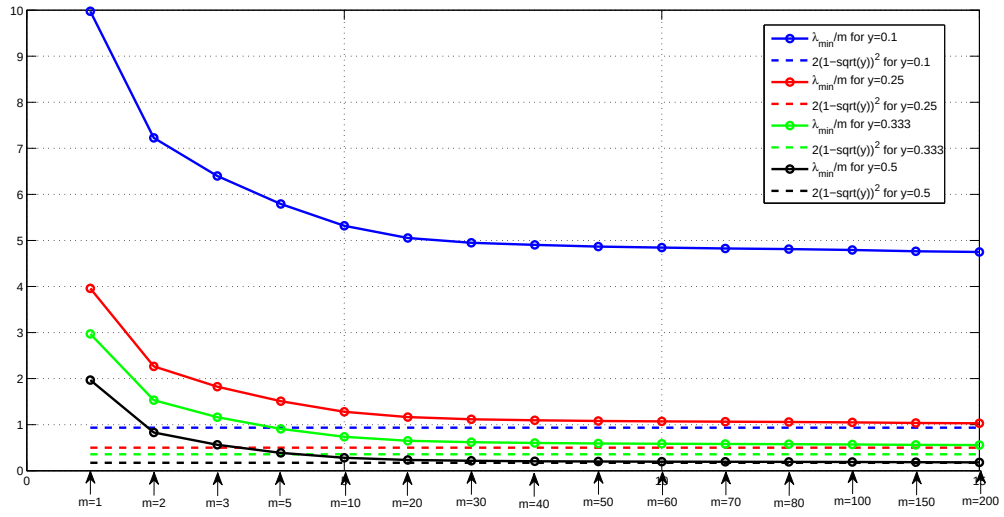


Figure 3.1: The ratio of the minimum eigenvalue to the number of transmit antennas, λ_{min}/m, for various transmit to receive antenna ratios, y, with increasing antennas, and the asymptotic value of the ratio, $2(1 - \sqrt{y})^2$ for each corresponding y.

A plot of the ratio of the minimum eigenvalue to the number of transmit antennas, as the number of antennas is increased for various transmit to receive antenna ratios is shown in Figure 3.1. The value of λ_{min}/m can be seen to approach the theoretical asymptote, particularly in the cases of larger y values, as the number of antennas is increased.

3.4 Minimum Distance of a Lattice

In this section we will discuss how the minimum Euclidean distance of a lattice generated by the $n \times m$ real channel matrix, $\mathbf{H}$, is used to bound error probability by bounding the bias. Assuming the all-zero lattice point of dimensions $m \times 1$ is transmitted, the error probability of the Sequential Decoder can be upper bounded as a function of the bias as shown in Eqn. (9) of [15]

$$P_e(b) \leq Pr\left(\bigcup_{\substack{\mathbf{s}\in\mathbb{Z}^m \\ \mathbf{s}\neq\mathbf{0}}} \left\{ 2\mathbf{s}^T\mathbf{w} > ||\mathbf{s}||^2\Big(1 - \frac{bm}{d^2_{min}(\mathbf{H})}\Big) \right\} \right).$$

From this we can see that we require the term $bm/d^2_{min}(\mathbf{H})$ to be less than 1. Hence, the bias, b, must be lower than $d^2_{min}(\mathbf{H})/m$, where $d^2_{min}(\mathbf{H})$ is defined as

$$d^2_{min}(\mathbf{H}) = \min_{\substack{\mathbf{s}\in\mathbb{Z}^m \\ \mathbf{s}\neq\mathbf{0}}} ||\mathbf{Hs}||^2.$$

Thus we are interested in finding the minimum distance of the noiseless received vector, obtained after multiplication of an integer transmit vector $\mathbf{s}$ with the channel matrix $\mathbf{H}$. As $\mathbf{s}$ can be any point on the m-dimensional integer space and the integer space has infinite points, using an exhaustive search such as ML decoding to find the minimum distance is not possible. Hence for the calculation of $d_{min}(\mathbf{H})$ we opt for a technique that has finite computational complexity, namely the Sphere Decoder presented in [12]. Note, as $\mathbf{H}$ is random, an average minimum distance can be found by averaging over a large number of realizations of the channel matrix. This average minimum distance is used for calculating the bias in our work.

For a particular ratio of transmit to receive antennas, $y = m/n$, increasing the number of transmit antennas and hence the number of receive antennas as well, an asymptotic bound on the minimum and maximum eigenvalues of channel matrices was found in [16]. The ratio of the minimum eigenvalue to the number of transmit

antennas, λ_{min}/m, of an $n \times m$ channel matrix approaches $2(1-\sqrt{y})^2$ as the number of antennas is increased for a fixed transmit to receive antenna ratio y. Additionally $d^2_{min}(\mathbf{H})$ is lower bounded by λ_{min} as:

$$\begin{aligned} d^2_{min}(\mathbf{H}) &= \min_{\substack{\mathbf{s}\in\mathbb{Z}^m \\ \mathbf{s}\neq\mathbf{0}}} ||\mathbf{Hs}||^2 \\ &\geq \min_{\substack{\mathbf{s}\in\mathbb{Z}^m \\ \mathbf{s}\neq\mathbf{0}}} \mathbf{s}\mathbf{H}^T\mathbf{Hs} \\ &\geq \lambda_{min}(\mathbf{H}^T\mathbf{H}) \min_{\substack{\mathbf{s}\in\mathbb{Z}^m \\ \mathbf{s}\neq\mathbf{0}}} ||\mathbf{s}||^2 \\ &= \lambda_{min}(\mathbf{H}^T\mathbf{H}). \end{aligned}$$

Note:

$$\min_{\substack{\mathbf{s}\in\mathbb{Z}^m \\ \mathbf{s}\neq\mathbf{0}}} ||\mathbf{s}||^2 = 1,$$

as the smallest norm corresponds to the vector with all entries equal to 0, except one entry which is equal to 1.

Through simulations using the Sphere Decoder we were able to calculate and plot $d^2_{min}(\mathbf{H})/m$ for different antenna ratios y with increasing antennas. We require this as we want to investigate the effect of using bias values in the range of the lowest $d^2_{min}(\mathbf{H})/m$ for a particular y.

Figure 3.2 is a plot of the average $d^2_{min}(\mathbf{H})/m$ of transmission systems for various transmit to receive antenna ratios, y, and with increasing number of antennas for each value of y. For larger system sizes, where calculating $d^2_{min}(\mathbf{H})/m$ may be difficult as using Sphere Decoding to find the minimum distance of large systems is extremely time consuming, the asymptotic value of λ_{min}/m may be a useful alternative. This is because as shown above, the asymptotic value of the ratio of the minimum eigenvalue to the number of transmit antennas, as the number of antennas grows, $2(1-\sqrt{y})^2$, lower bounds the average $d^2_{min}(\mathbf{H})/m$. This can thus be used for estimating a bias value that is small enough to satisfy the error probability bound. Of course larger

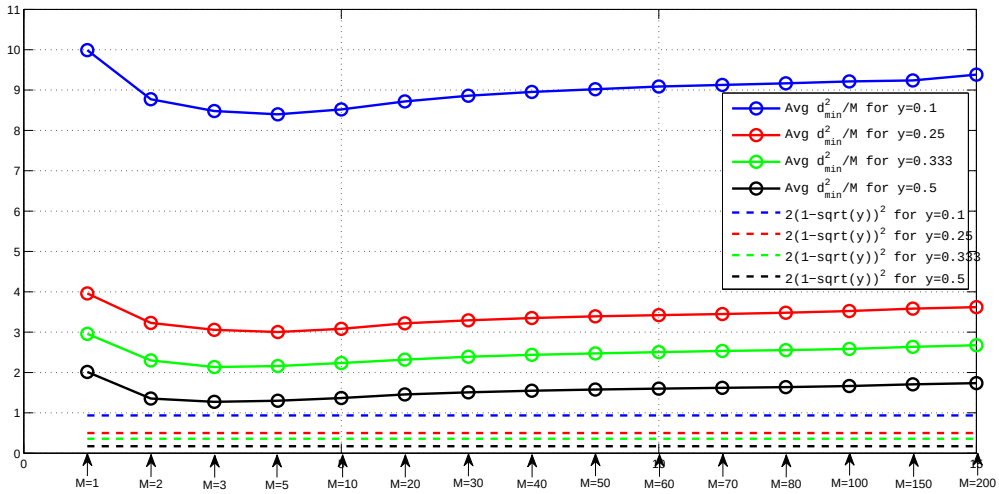


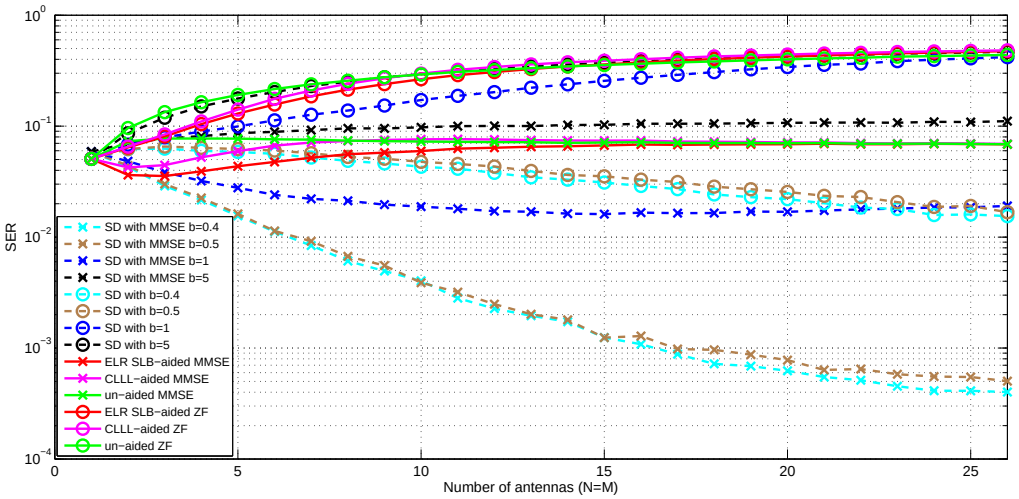
Figure 3.2: Asymptotic values of the ratio of the minimum eigenvalue to the number of transmit antennas, for various transmit to receive antenna ratios, y, and the average d^2_{min}/M for systems with the corresponding y values and increasing number of antennas.

values of the bias exist, that are still less than $d^2_{min}(\mathbf{H})/m$, and therefore satisfy the error bound at lower complexities, but in the case of larger system sizes evaluating these bias values may be difficult.

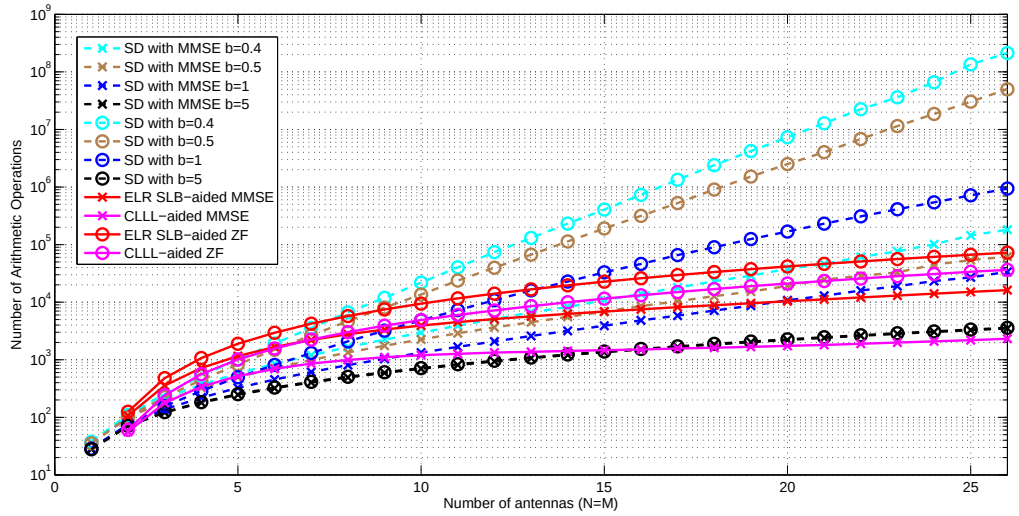
3.5 Numerical Results

In this section we compare the performance and complexity of Sequential Decoders with and without an MMSE extended system, using different bias values, against LR-aided ZF and MMSE LDs. The LR techniques employed for comparison are the dual ELR SLB and the CLLL. The bias values used for the Sequential Decoder are 0.4, 0.5, 1 and 5 so that the performance-complexity trade-off can be observed. All the systems being analyzed contain equal numbers of transmit and receive antennas. The bias values of 0.4 and 0.5 were chosen as the values are in the range lower than $d^2_{min}(\mathbf{H})/m$ for $y = 1$.

Figure 3.3a is a plot of the symbol error rate as the number of antennas is



(a) Performance of different detectors for increasing number of antennas.



(b) Complexity of different detectors for increasing number of antennas.

Figure 3.3: Performance and complexity of different detectors for MIMO Systems employing 4-QAM, SNR=3dB and equal number of transmit and receive antennas.

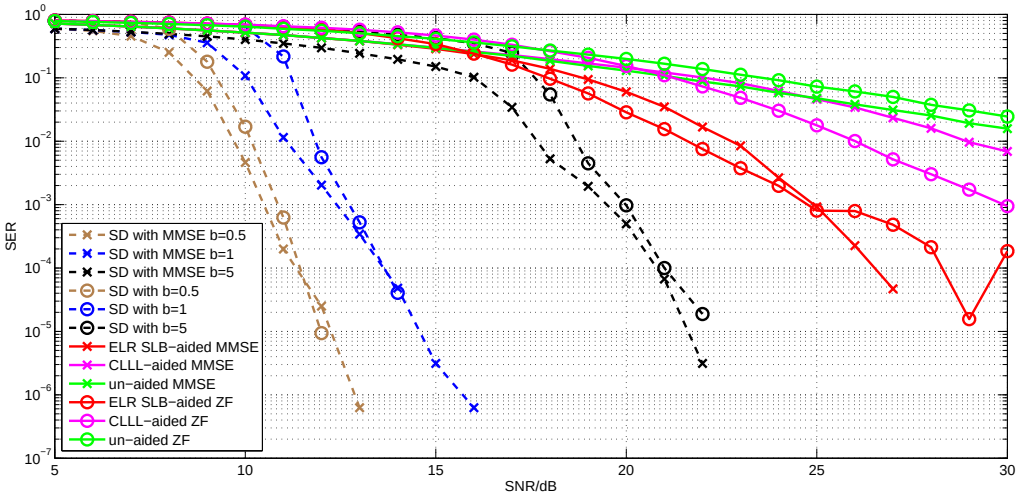
increased. A 4-QAM constellation is used for transmission and an SNR of 3dB. As expected, the un-aided ZF decoder has the worst performance. The CLLL-aided ZF shows some improvement and the ELR SLB-aided ZF has the best performance among the LR aided ZF decoders. The Sequential Decoder with a bias of 5 performs not much better than the unaided ZF, but at a bias of 1 there is a marked improvement in performance, and bias values of 0.4 and 0.5 show particularly good performance. In the cases of bias values equal to 0.4 and 0.5, it is also very interesting to note that unlike the other decoders, error is decreasing with increasing number of antennas. The LR aided MMSE and unaided MMSE showed marked improvement but similar performance trends to their ZF counterparts. The extended MMSE system was also applied to the Sequential Decoders and performance improvement was again noted from the corresponding un-extended case; the performance with bias values equal to 0.4 and 0.5 is exceptional.

Figure 3.3b shows the corresponding complexities of the Sequential Decoders and the complexity of the lattice reduction for each technique, applied to both the channel matrix (ZF case) and the extended channel matrix (MMSE case). As expected, the complexity of the MMSE systems is lower than their ZF counterparts and the complexities grow with increasing antennas. Among the LR schemes, ELR SLB reduction has the highest complexity, followed by CLLL. The rate at which complexity for the Sequential Decoders increases is larger than that of the lattice reductions particularly for the smaller bias values. The complexity of the Sequential Decoder with the extended MMSE system shows considerable improvement and is even lower than the SLB-aided MMSE for smaller systems, but when the number of antennas exceeds around 12 to 20, the complexity of the Sequential Decoder with lower bias values becomes larger. The price paid for the performance gain of the Sequential Decoder is its high complexity for large systems. For small systems, such as of five antennas, it is interesting to note that there is a performance gain without higher

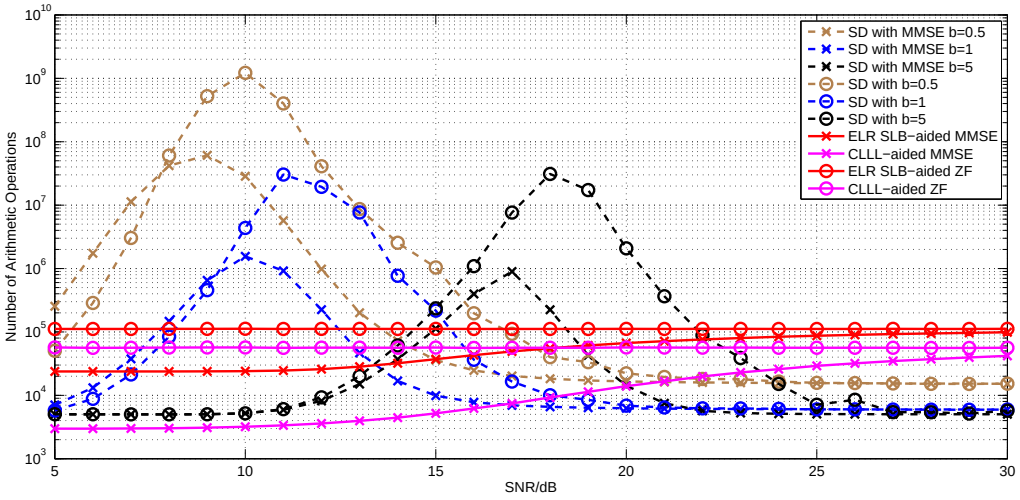
complexity.

Figure 3.4a is the plot of the symbol error versus SNR for a 32×32 system employing a 16-QAM constellation. As can be seen from the figure, the performance of the Sequential Decoder is superior and its error curves fall a lot more sharply and at considerably lower SNR values. The error curves corresponding to lower bias values fall at lower SNR and the error curves fall almost parallel to one another. Additionally, the Sequential Decoder when applied to an extended MMSE system results in the error curves falling at even lower SNR values; the gap can be seen between them and their corresponding un-extended system. The ELR SLB-aided ZF and MMSE have the next best performance after the Sequential Decoders, though the ELR SLB-aided ZF has better performance at moderate SNR values and at higher SNR the ELR SLB-aided MMSE outperforms it. The CLLL-aided MMSE and ZF perform next best. At low to moderate SNR, the CLLL-aided MMSE outperforms the CLLL-aided ZF, but at an SNR of around 20dB, the CLLL-aided ZF's error falls and it outperforms its MMSE counterpart. The worst performance as expected is by the unaided ZF, followed by the unaided MMSE.

Figure 3.4b is a plot of the corresponding complexity of the Sequential Decoders with and without the extended MMSE system, and the Lattice Reduction techniques applied to the channel matrix and the extended MMSE channel matrix. The complexity for Lattice Reduction of the un-extended ZF systems is constant and independent of SNR; the ELR SLB having the highest complexity, followed by the CLLL. The corresponding MMSE systems have lower complexity at low SNR values, but as SNR increases the complexity of the MMSE system's Lattice Reduction reaches that of the ZF system's. This can be explained by the fact that in the case of the MMSE, the channel matrix is replaced by the extended channel matrix, which is better conditioned. Since both the LR algorithms perform reduction until the matrix is 'good enough' in terms of orthogonality of the columns, the extended channel matrix in



(a) Performance vs. SNR for different detectors.



(b) Complexity vs. SNR for different detectors.

Figure 3.4: Performance and complexity of different detectors for a 32×32 MIMO System employing 16-QAM.

a lot of instances does not require reduction as it is already good enough. When the unit variance channel matrix is multiplied by a larger factor, the need to reduce the channel matrix and make it closer to what is defined as 'minimal orthogonality' is increased. Increasing SNR therefore results in more channel matrices undergoing reduction. This increases until all the channel matrices need to be reduced and hence we have constant complexity for the higher SNR values. This can also be seen by varying a parameter of the CLLL reduction algorithm, δ, which is a measure of how much the channel matrix should be reduced to make it close to orthogonal. Increasing δ results in more reduced matrices, higher computational complexity and better performance. δ can take any value between 0.5 and 1, and was set to 0.5 for all of our simulations.

The Sequential Decoders in Figure 3.4b have a particular complexity trend: a region of low complexity (comparable to the LR techniques) at low SNR, a region of high complexity in a particular SNR range, and low complexity (comparable to the LR techniques) at higher SNR. The range of high complexity shifts towards higher SNR values as the bias is increased. It should be noted that for each bias value the range of SNR that has high complexity corresponds to the SNR range in the error performance curves where the error probability first starts to drop steeply. After this high-complexity range, at higher SNR values the complexity of the Sequential Decoders falls to values comparable to and even lower than the LR techniques. These trends can be explained as follows; initially at low SNR as the noise is more significant relative to the signal part of the received message, the decoder finds incorrect branches on the tree and makes mostly erroneous decisions without much effort; this corresponds to high error probability and low complexity. As the SNR is increased, the decoder is able to differentiate between the noise and signal components, so the complexity increases and error decreases as more work is done to decode correctly by finding the right branch. At higher SNR values, it is easier for the decoder to

differentiate between the error and signal components, and correct detection is done without too much work by the decoder; hence low error at low complexity. It should also be noted that the Sequential Decoders applied to the extended MMSE systems, although still have higher complexity than the Lattice Reduction techniques, show a significant complexity improvement from their un-extended counterparts, have a narrower range of high-complexity and this range occurs a little before the that of the un-extended Sequential Decoder's high-complexity range, corresponding to the fact that their error curves drop before those of the un-extended.

Chapter 4

Past work on Underdetermined Systems

4.1 Background

In the case of the most frequently used communication system, one end is a base station and the other a mobile. The base station can be large in size and therefore accommodate a large number of antennas; the mobile on the other hand is limited in size and therefore number of antennas. Hence the downlink always has a larger number of receive antennas than transmit antennas, resulting in a tall channel matrix and therefore an overdetermined system. The uplink, however, has a smaller number of receive antennas than the transmit antennas; the channel matrix is therefore fat, and the system is underdetermined.

The rank of a matrix, $\mathbf{H}$, satisfies $rank(\mathbf{H}) \leq min\{m, n\}$. For overdetermined transmission systems the rank of the channel matrix is assumed equal to m and therefore the number of unknowns, and so the Sphere Decoding algorithm can be applied directly. However in the case of underdetermined systems or for MIMO channels that are highly correlated, the rank of the channel matrix is assumed to be equal to n and is therefore less than the number of unknowns m.

Also, the matrix $\mathbf{H}^H\mathbf{H}$ is full rank and therefore positive definite for an overde-

termined system. For the case of the underdetermined system, $\mathbf{H}^H\mathbf{H}$ is positive semidefinite, its Cholesky factor $\mathbf{R}$ is not full rank, and therefore the last $M - N$ rows of $\mathbf{R}$ are equal to zero.

In an overdetermined system, the channel matrix $\mathbf{H}$ generates a lattice over $\mathbb{R}^m$, and the problem can be solved using a Sphere Decoder. However, in the underdetermined case, since the number of receive antennas is smaller than the number of transmit antennas, $\mathbf{H}$ generates only a projection over $\mathbb{R}^n$ of a lattice in $\mathbb{R}^m$.

Although the general Closest Lattice Point Search (CLPS) problem is known to be NP-hard, the average complexity of the Sphere Decoder is polynomial in the number of unknowns for overdetermined, Rayleigh fading channels that suffer additive white Gaussian noise. However, for the case of underdetermined systems, the polynomial complexity of the Sphere Decoder does not hold and an exponential term in the difference between the number of unknowns and equations exists.

A number of algorithms and techniques have been proposed that aim to decode underdetermined systems efficiently. Various techniques that result in optimal decoding of the underdetermined system with reduced computational complexity are presented in Section 4.2. Section 4.3 captures some interesting techniques proposed for the suboptimal decoding of underdetermined systems; these techniques obviously have the advantage of decreased complexity compared to the optimal ones at the price of performance degradation. Our work on underdetermined MIMO systems also involves a suboptimal approach and will be presented in Chapter 5.

4.2 Optimal Detection Schemes for the Underdetermined Transmission Case

4.2.1 Generalized Sphere Decoding Algorithm

The problem of decoding an asymmetric MIMO channel, and in particular the uplink, which is an underdetermined MIMO system, is solved using a Generalized Sphere Decoding algorithm in [17] and [18]. Since a Sphere Decoding algorithm, which is a type of lattice decoder, is to be employed, we deal with the real channel model. The algorithm starts by finding the ZF point $\boldsymbol{\rho} = \mathbf{H}^{\dagger}\mathbf{x}$, this way the sphere centered at the received signal is transformed into an ellipsoid centered at the origin. Cholesky factorization is used to convert the problem of minimizing the Euclidean distance, into an upper triangular form, followed by the Sphere Decoder. Since the system is underdetermined, the upper triangular matrix has the last $m - n$ rows equal to zero, hence the last $m - n$ components of $\mathbf{s}$ need to be chosen exhaustively. This leads to the algorithm having exponential complexity in the difference between the number of unknowns and equation. The inequalities imposed due to the requirement that $\mathbf{s}$ lie within the sphere of radius C_0, and the upper triangular nature of the metric being minimized, results in the requirement that the values of allowed s_i, for $i \leq n$, lie in a certain range. Fortunately the upper and lower bounds of this range depend on the choice of s_i for $n < i \leq m$. This way choices of s_i for $n < i \leq m$ that make the range of s_i for $i \leq n$ empty can be discarded, hence making the algorithm more efficient. Also, note that once a point $\mathbf{s}$ is found inside the sphere, the radius is updated if the distance of this point to the center is less than the current radius. The new bounds of the inequalities to be tested are updated accordingly.

4.2.2 Improved Generalized Sphere Decoding Algorithm

An improved version of the Generalized Sphere Decoder (GSD) is proposed in [19]. This algorithm is significantly faster than the one proposed in [17] and a step closer towards achieving non-exponential complexity for the decoding of underdetermined systems. The GSD in [17] does not use the knowledge of the failure of a certain hypothesized s_i for $n < i \leq m$, to predict the failure of other hypotheses of s_i for $n < i \leq m$; the improved version of the GSD, however, is motivated by this key observation which makes it a lot more efficient. Like the case of the GSD, only those s_i where $n < i \leq m$ are acceptable that make an inequality based on the sphere radius true. However, we partition and order the possible combinations of s_i for $n < i \leq m$ in such a way that if certain combinations fail, then a number of other combinations will also fail, hence these can be discarded without actually being tested. This way we are able to attain the optimal decision without necessarily having to employ an exhaustive search over all possible combinations of the $m - n$ unknowns. This results in a significant improvement in the computational cost of the search process.

4.2.3 Recursive Improved Partitioning Approach

The algorithm proposed in [20] is also an optimal GSD, and has a recursive implementation. It involves partitioning of sets as in the case of [19], but unlike [19], where the depth of the algorithm must be chosen, this approach starts from the deepest level. It results in complexity significantly lower than that of the algorithms proposed in [17] and [19].

4.2.4 Double-Layer Sphere Decoder

An optimal approach for decoding underdetermined MIMO systems is presented in [21] that employs a Double-Layer Sphere Decoder (DLSD) to improve the speed from

the algorithms presented in [17] and [19]. Cholesky factorization is used to transform the transmission system into an equivalent upper-trapezoidal system. Let $\mathbf{F}$ represent the $M \times M$ upper-triangular matrix with the last $M - N$ rows containing all zeros, ($\mathbf{F}$ is therefore equivalently an upper-trapezoidal matrix). The minimization can therefore be written as:

$$||\mathbf{F}(\boldsymbol{\rho} - \mathbf{s})||^2 = \sum_{i=1}^{n} \left|\left| \sum_{j=i}^{m} F_{ij}(\rho_j - s_j) \right|\right|^2 \tag{4.1}$$

$$= \sum_{i=1}^{n-1} \left|\left| \sum_{j=i}^{m} F_{ij}(\rho_j - s_j) \right|\right|^2 + \left|\left| \sum_{j=n}^{m} F_{nj}(\rho_j - s_j) \right|\right|^2. \tag{4.2}$$

As conventional sphere decoding requires this metric which is to be minimized to be less than a certain sphere radius C, we can equivalently use the usual sphere decoding inequalities to write the following

$$\left|\left| \sum_{j=n}^{m} F_{nj}(\rho_j - s_j) \right|\right|^2 \leq ||\mathbf{F}(\boldsymbol{\rho} - \mathbf{s})||^2 \leq C \tag{4.3}$$

and therefore

$$\sum_{j=n}^{m} F_{nj}\rho_j - \sqrt{C} \leq \sum_{j=n}^{m} F_{nj}s_j \leq \sum_{j=n}^{m} F_{nj}\rho_j + \sqrt{C}. \tag{4.4}$$

After applying some transformations to this inequality, the details of which we will not discuss here, the algorithm is able to get a set of inequalities that allow a slightly modified Sphere Decoder to be applied. This way we are able to decode to the 'outer layer' in order to obtain the last $m - n + 1$ values of the transmitted vector $\mathbf{s}$. Once these unknowns are calculated, their values are subtracted from the upper-trapezoidal problem, and used to transform it to an overdetermined problem. A regular Sphere

Decoder is then applied to decode the 'inner layer', namely the first $n-1$ values of the transmit vector $\mathbf{s}$.

4.2.5 A Regularization Approach - Transforms Underdetermined System into an Overdetermined System

An optimal algorithm for the decoding of underdetermined MIMO systems that employ a constant modulus constellation is proposed in [22]. It is also extended to the case of non-constant modulus QAM constellations at the price of higher complexity. Particularly for the case of constant modulus constellation symbols, this algorithm results in much lower complexity than the complexity of the GSD and the improved GSD in [17] and [19] respectively. In fact the reduction in complexity of the proposed algorithm compared to that of the complexity in [17] and [19] grows as the difference between the number of transmitters and receivers grows.

The algorithm modifies the ML metric using the fact that a constant modulus modulation is being employed, in order to increase the rank of the effective 'channel matrix' after the modification to be equal to the number of unknowns M. This way the regular sphere decoder can be applied to the modified system. The reduction in complexity occurs as the new channel matrix is full rank, and therefore the upper triangular matrix obtained from Cholesky factorization does not have the last $M-N$ rows equal to zero, and hence the corresponding symbols of $\mathbf{s}_c$ do not need to be decoded separately (using an exhaustive search in the case of [18] or on average a non-exhaustive search in the case of [19]).

The modification is done as follows, since the ML rule requires minimization of the magnitude of the noise term over all possibilities of $\mathbf{s}_c$, adding an additional term independent of $\mathbf{s}_c$ does not affect the minimization rule. As a constant modulus constellation is being employed, $\alpha \mathbf{s}_c^H \mathbf{s}_c = \alpha M$, which is a constant and independent

of the particular $\mathbf{s}_c$ being considered. The minimization is therefore equivalently,

$$\hat{\mathbf{s}}_c = \arg \min_{\mathbf{s}_c \in \mathcal{M}^M} ||\mathbf{x}_c - \mathbf{H}_c \mathbf{s}_c||^2 + \alpha \mathbf{s}_c^H \mathbf{s}_c \quad (4.5)$$

$$= \arg \min_{\mathbf{s}_c \in \mathcal{M}^M} \mathbf{x}_c^H \mathbf{x}_c - \mathbf{x}_c^H \mathbf{H}_c \mathbf{s}_c - \mathbf{s}_c^H \mathbf{H}_c^H \mathbf{x}_c + \mathbf{s}_c^H (\mathbf{H}_c^H \mathbf{H}_c + \alpha \mathbf{I}_M) \mathbf{s}_c. \quad (4.6)$$

Denote by $\mathbf{G}_c$, the positive definite matrix $\mathbf{H}_c^H \mathbf{H}_c + \alpha \mathbf{I}_M$. $\mathbf{G}_c$ is Cholesky factorized into $\mathbf{G}_c = \mathbf{D}_c^H \mathbf{D}_c$, where $\mathbf{D}_c$ is an upper triangular matrix. Defining $\mathbf{b} = \mathbf{G}_c^{-1} \mathbf{H}_c^H \mathbf{x}_c$, and adding and subtracting $\mathbf{b}^H \mathbf{D}_c^H \mathbf{D}_c \mathbf{b}$ from the last equation, the optimization problem can be restated as,

$$\hat{\mathbf{s}}_c = \arg \min_{\mathbf{s}_c \in \mathcal{M}^M} ||\mathbf{D}_c (\mathbf{b} - \mathbf{s}_c)||^2. \quad (4.7)$$

The new $M \times M$ upper triangular channel matrix $\mathbf{D}_c$, has rank equal to the number of unknowns M and therefore has all non-zero diagonal elements. Regular sphere decoding can be applied to this system to obtain all M unknowns without requiring a separate search over any components of the unknown transmit vector $\mathbf{s}_c$. Note since the proposed algorithm works independent of the rank of the channel matrix $\mathbf{H}_c$, it can be applied to both under and overdetermined transmission systems.

For the case of non-constant QAM modulations, the constellation can be decomposed into a 4-QAM constellation of a larger dimension and a correspondingly modified channel matrix. This decomposition results in a larger transmit vector and a channel matrix with increased columns. The algorithm above is then applied to the new system the way it would be to any constant modulus constellation system. Of course an increase in complexity is incurred in these cases due to the increase in system size, but as the difference between the number of transmit and receive antennas grows, even these non-constant modulus QAM modulations outperform the GSD and improved GSD in [17] and [19].

To transform a square $\mathcal{M}$-QAM constellation system into an equivalent system that employs a 4-QAM constellation, the following transformations are applied. Let $\mathcal{M} = 2^{2k}$, then an $\mathcal{M}$-QAM constellation can be represented as a weighted sum of k 4-QAM constellations. For an element z belonging to the $\mathcal{M}$-QAM constellation, $z_i \in$ 4-QAM for $0 \leq i < k$ and,

$$z = \sum_{i=0}^{k-1} 2^i z_i. \tag{4.8}$$

For the case of a 16-QAM system for example,

$$\mathbf{s}_c = \mathbf{s}_{c1} + 2\mathbf{s}_{c2},$$

and the transmission system can therefore be represented as,

$$\mathbf{x}_c = \begin{bmatrix} \mathbf{H}_c & 2\mathbf{H}_c \end{bmatrix} \begin{bmatrix} \mathbf{s}_{c1} \\ \\ \mathbf{s}_{c2} \end{bmatrix} + \mathbf{w}_c, \tag{4.9}$$

which is an equivalent system of dimensions $N \times 2M$ in this case, and $N \times kM$ in general.

4.2.6 Improved Regularization Approach - Transforms Underdetermined System into an Overdetermined System

The algorithm presented in [23] is a modified version of the algorithm in [22] and transforms the underdetermined Integer Least Squares (ILS) problem into an equivalent overdetermined integer least squares problem. This is achieved by doing a regularization using part of the transmit vector, $\mathbf{s}$. QR factorization of the real channel matrix is used to translate the transmission system into an equivalent upper trapezoidal system. The upper trapezoidal matrix is partitioned into an $n \times n$ upper triangular

matrix, $\mathbf{R}_1$ and another matrix, $\mathbf{R}_2$ of dimensions $n \times (m-n)$. The channel matrix $\mathbf{H}$ is partitioned in such a way that $\mathbf{H}_1$ contains the first $n-1$ columns of $\mathbf{H}$, and $\mathbf{H}_2$ contains the last $m-n+1$ columns. The $n-1$ unknowns corresponding to $\mathbf{H}_1$ are called $\mathbf{s}_1$, and the $l = m-n+1$ unknowns corresponding to $\mathbf{H}_2$ are called $\mathbf{s}_2$. For $\mathbf{s}_c$ belonging to QAM constellations larger than 4-QAM, and the corresponding $\mathbf{s}$ belonging to constellations larger than 2-PAM, vector $\mathbf{s}_2$ is translated into the equivalent vector $\bar{\mathbf{s}}_2$, that contains all elements belonging to a 2-PAM constellation. $\bar{\mathbf{s}}_2$ is therefore k times the size of vector $\mathbf{s}_2$, where the 2^{2k}-QAM constellation is being employed. The matrix $\mathbf{H}_2$ corresponding to $\mathbf{s}_2$ is modified accordingly as well to produce matrix $\bar{\mathbf{H}}_2$ of dimensions $n \times kl$. Since $\bar{\mathbf{s}}_2$ has elements belonging to a constant modulus constellation, the $||\bar{\mathbf{s}}_2||^2$ is a constant, and the original ILS problem, or equivalently the ML detection rule, can be re-written as:

$$\min_{\mathbf{s}_1 \in \chi_k^{n-1}, \bar{\mathbf{s}}_2 \in \chi_1^{kl}} \left|\left| \mathbf{x} - [\mathbf{H}_1 \;\; \bar{\mathbf{H}}_2] \begin{bmatrix} \mathbf{s}_1 \\ \bar{\mathbf{s}}_2 \end{bmatrix} \right|\right|^2 + \alpha^2 ||\bar{\mathbf{s}}_2||^2, \tag{4.10}$$

where α is a regularization parameter. Defining the following,

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_1 & \bar{\mathbf{H}}_2 \\ \mathbf{0} & \alpha \mathbf{I} \end{bmatrix} \in \mathbb{R}^{(n+kl)\times(m+(k-1)l)},$$

and

$$\bar{\mathbf{s}} = \begin{bmatrix} \mathbf{s}_1 \\ \bar{\mathbf{s}}_2 \end{bmatrix} \in \mathbb{R}^{m+(k-1)l}, \quad \bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n+kl} \tag{4.11}$$

$$\bar{\chi} = \left\{ \begin{bmatrix} \mathbf{s}_1 \\ \bar{\mathbf{s}}_2 \end{bmatrix} : \mathbf{s}_1 \in \chi_k^{n-1}, \bar{\mathbf{s}}_2 \in \chi_1^{kl} \right\} \tag{4.12}$$

allows us to rewrite the minimization problem as,

$$\min_{\bar{\mathbf{s}} \in \bar{\chi}} ||\bar{\mathbf{x}} - \bar{\mathbf{H}}\bar{\mathbf{s}}||_2^2. \tag{4.13}$$

As the number of equations is not less than the number of unknowns anymore, this is a box constrained overdetermined ILS problem and can thus be solved using a regular Sphere Decoder.

Compared to [22], this modified algorithm has a lower computational complexity. In the case of $k = 1$, this stems from the fact that this modified approach adds a regularization term based on the last $m - n + 1$ components of $\mathbf{s}$ whereas the original algorithm in [22] added the regularization term based on the whole transmit vector $\mathbf{s}$; this results in a larger number of rows of the effective channel matrix $\bar{\mathbf{H}}$ for the original algorithm than the proposed algorithm, while the number of columns is the same in both, implying a larger system size for the original algorithm and therefore higher computational complexity. For the case of $k \geq 2$, reduction in computational complexity occurs as the modified algorithm results in a smaller number of columns for the effective channel matrix $\bar{\mathbf{H}}$ than the original algorithm. This occurs as partitioning of elements s_i transmitted from a non-constant modulus constellation is done for the vector $\bar{\mathbf{s}}_2$ only, while in the original algorithm it is done for the entire transmit vector $\mathbf{s}$. Hence the system size is smaller for the proposed algorithm, resulting in significant complexity reductions.

4.2.7 Tree Search Decoding

An efficient Tree Search Decoder (TSD) is proposed in [24] to optimally decode underdetermined MIMO systems. This is done by integrating the two search processes presented in the Double Layer Sphere Decoding algorithm presented in [21] into one process. This results in a depth first tree search algorithm. A novel Column Re-

ordering (CR) strategy is also proposed to increase the efficiency of the search. The TSD-CR algorithm presented in this paper does not handle the cases of constant modulation and non-constant modulus QAM constellations separately, hence does not enlarge the dimensions of the ILS problem and is therefore more efficient.

Using QR decomposition the ILS problem is transformed to an equivalent upper trapezoidal problem, and using the concept of sphere decoding into a set of inequalities. The inequalities are divided into a set that is used to decode s_i for $i = m : -1 : n$ and another for $i = n - 1 : -1 : 1$. For decoding s_i for $i = m : -1 : n$, a method similar to the decoding of the last $m - n + 1$ elements of the transmit vector $\mathbf{s}$ in [21], using an outer-layer Sphere Decoder by applying linear transformations, is used. Modifying the inequalities for $i = n - 1 : -1 : 1$, a range of allowable values of s_i for $i = n - 1 : -1 : 1$ is obtained. Once the last $m - n + 1$ components of $\mathbf{s}$ are obtained, it can be checked if the range of the allowable s_i for $i = n - 1$ is non-empty, in which case the estimate of s_{n-1} is obtained, and we then proceed to level $n - 2$. However, if the allowable range for $i = n - 1$ is empty, we go back to level $i = n$ to find suitable transmit elements that result in a non-empty range.

It is important to mention that in the outer-layer sphere decoding applied in this algorithm, non-constant modulus QAM constellations were not transformed into equivalent but larger 4-QAM systems. This is because the non-constant modulus was already incorporated in the linear transformations applied in this case, unlike those applied in [21]. Hence the system size was not increased and so the search process was more efficient.

A novel column reordering strategy is also proposed in this paper but we will not be discussing it. Another optimal Tree-search ML detection for underdetermined systems is presented in [25]. It has certain benefits as the same algorithm works with both overdetermined and underdetermined systems, and it requires no preprocessing. It results in complexity lower than that of the Generalized Sphere Decoder. However,

this scheme works with M-PSK constellations only, and so we will not be discussing it.

4.3 Suboptimal Detection Schemes for the Underdetermined transmission case

Despite the attractiveness that comes with an algorithm that achieves optimal performance, it is crucial to realize that in the underdetermined case, decoding is in fact a very difficult problem - especially when the system size increases and when the difference between the number of transmit and receive antennas grows; which is the case in real transmission systems that have a large base station and a small mobile device. Small degradation in performance in the case of some suboptimal schemes is not a very large price to pay for the reductions in computational complexity they offer.

4.3.1 Prevoting Cancellation-based detection and Optimal Postvoting Vector Selection

A Prevoting Cancellation (PVC) based detection approach is employed for the decoding of underdetermined MIMO systems in [26] and [27]. This suboptimal scheme results in very good and even near-ML performance depending on whether or not optimal column rearrangement is employed. The basic idea of the prevoting cancellation-based detection is to divide the columns of a fat channel matrix, $\mathbf{H}_c$, into a square matrix, $\mathbf{H}_{cq}$, and another matrix, $\mathbf{H}_{cp}$. The elements of $\mathbf{s}$ corresponding to the columns of $\mathbf{H}_{cq}$ are denoted by $\mathbf{s}_{cq}$, and those corresponding to the columns of $\mathbf{H}_{cp}$, $\mathbf{s}_{cp}$. The difference between the number of transmit and receive antennas is denoted by R, i.e. $R = M - N$. Hence, the original complex transmission system can be denoted as

follows:

$$\mathbf{x}_c = \mathbf{H}_{cq}\mathbf{s}_{cq} + \mathbf{H}_{cp}\mathbf{s}_{cp} + \mathbf{w}_c. \tag{4.14}$$

The $R \times 1$ vector $\mathbf{s}_{cp}$ is called the prevoting vector and the $N \times 1$ vector $\mathbf{s}_{cq}$, the postvoting vector, for reasons which will become clear shortly. An exhaustive search is done over part of the transmitted message by considering all $\mathcal{M}^R$ possibilities of the R elements in the prevoting vector $\mathbf{s}_{cp}$. The k^{th} possibility of $\mathbf{s}_{cp}$ is denoted by $\mathbf{s}_{cp}^{(k)}$, where $k = 1, 2, \ldots, \mathcal{M}^R$. By subtracting the contribution of the prevoting vector from the received vector, $\mathbf{x}_c$, a square system can be formed as follows:

$$\begin{aligned} \mathbf{y}_c^{(k)} &= \mathbf{x}_c - \mathbf{H}_{cp}\mathbf{s}_{cp}^{(k)} && (4.15) \\ &= \mathbf{H}_{cq}\mathbf{s}_{cq} + \mathbf{w}_c. && (4.16) \end{aligned}$$

Simpler techniques such as LD or LR aided detection, are then employed for the detection of the remaining symbols of $\mathbf{s}_c$, namely the elements of the postvoting vector $\mathbf{s}_{cq}$, from the square subsystem. It should be noted that each possibility of the prevoting vector, i.e. each $\mathbf{s}_{cp}^{(k)}$ for $k = 1, 2, \ldots, \mathcal{M}^R$, results in a different subsystem and therefore a different estimate of its corresponding postvoting vector $\mathbf{s}_{cq}$. The final decision of which prevoting and postvoting vector pair to choose from the $\mathcal{M}^R$ possibilities of $\hat{\mathbf{s}}$, where $\hat{\mathbf{s}}^{(k)} = [\mathbf{s}_{cp}^{(k)T} \quad \hat{\mathbf{s}}_{cq}^{(k)T}]^T$ for $1 \leq k \leq \mathcal{M}^R$, as the estimate of the transmitted vector $\mathbf{s}_c$, is done by choosing the pair which results in the smallest Euclidean distance metric, in other words the pair that minimizes the noise in the received symbol. The scheme results in very good performance, but at the cost of high complexity, particularly as the difference between the number of transmit and receive antennas grows.

Without optimal-Postvoting Vector Selection (PVS), i.e. when optimal column rearrangement is not considered, the complexity of PVC-MIMO detection is simply

equal to $C_{tot} = \mathcal{M}^R C_{sub}$. C_{sub} includes the complexity of the interference removal (the removal of the contribution of $\mathbf{H}_{cp}\mathbf{s}_{cp}$), the detection of the square subsystem, and the noise calculation; C_{sub} therefore needs to be multiplied by the factor $\mathcal{M}^R$ as there are $\mathcal{M}^R$ possibilities of $\mathbf{s}_{cp}$.

If optimal column re-arrangement is employed, the criteria used to decide which columns to select, comprises of the following. Since choosing the optimal prevoting vector is equivalent to selecting the optimal postvoting vector, we opt for optimal Postvoting Vector Selection. The optimal postvoting vector set is denoted by $\mathcal{Q}$, where $\mathcal{Q} \subset \{1, \ldots, M\}$ The selection criteria are as follows:
for unaided LDs:

$$\mathcal{Q} = \arg\max_{\mathcal{Q}} \lambda_{min}(\mathbf{H}_{cq}^H \mathbf{H}_{cq})$$

for LR-aided MMSE :

$$\mathcal{Q} = \arg\max_{\mathcal{Q}} \lambda_{min}(\mathbf{G}_{cq}^H \mathbf{G}_{cq}),$$

where $\mathbf{G}_{cq}$ is the lattice-reduced basis from $\mathbf{H}_{cq}$, and $\lambda_{min}(\mathbf{A})$ denotes the minimum eigenvalue of the matrix $\mathbf{A}$.

The complexity analysis of the prevoting cancellation with optimal PVS can be broken down as follows: $C_{tot} = \frac{(\prod_{i=0}^{N-1}(M-i))C_{sel}}{W} + \mathcal{M}^R C_{sub}$, where C_{sel} is the complexity of the lattice reduction process and the column rearrangement required for the selection of the optimal $\mathbf{H}_{cq}$. W denotes the number of transmissions over which the channel remains unchanged, and therefore the column rearrangement and lattice reduction do not need to be repeated for W transmissions. The second term in C_{tot} is the same as that for PVC without optimal PVS.

As can be seen from the above analysis, the complexity of decoding becomes very high, particularly for the cases of large values of R, and when large $\mathcal{M}$-QAM

constellations are used. Furthermore, for the case of slow-fading channels, where W is large, complexity will be low; the complexity will, however, grow very quickly for fast-fading channels with lower values of W.

The error rate for the case of the PVC without optimal PVS is much higher and the complexity a lot less than the case of the PVC with optimal PVS. Also, a significant gap exists between the error curves of the PVC without optimal PVS and the error curves for ML decoding. With Optimal PVS, PVC error curves are very close to the ML error curves and in a lot of cases the difference between the two is almost negligible.

PVC with PVS using Sequential Decoding: We have also applied PVC with optimal PVS using the Sequential Decoder instead of an LR-aided/un-aided LD, to see performance improvements. These were negligible, as the technique is suboptimal and the inferior latter decoding technique already approached very near-optimal performance. The complexity however was much higher, and so the technique did not improve the performance-complexity trade-off enough to be analyzed more.

4.3.2 MMSE-GDFE Preprocessing followed by Lattice Reduction, Greedy Ordering and Sphere Decoding

It has been shown in [12] that MMSE-GDFE followed by Greedy Ordering greatly reduces the complexity of the Sphere Decoder. The algorithm presented in [28] and [29] can basically be broken down into the following steps: 1) MMSE-GDFE preprocessing 2) lattice reduction 3) column reordering using Greedy Ordering 4) lattice decoding using the Schnorr-Euchner enumeration with a finite radius (Sphere Decoding) 5) estimation of the original transmitted vector by reordering the output of the Sphere Decoder, translating back to the original lattice, and quantization.

Applying MMSE-GDFE front-end filtering to an underdetermined channel matrix, $\mathbf{H}_c$ results in obtaining a full rank linear system, with the number of equations

equal to the number of unknowns M. The system obtained is also upper triangular. It is important to note, however, that MMSE-GDFE preprocessing introduces a suboptimality in the decoding process. The details of this suboptimality are described in Section 4.4.1. Since we will be using lattice decoding techniques, the transmitted message to be decoded must be an integer vector, and so we will apply the MMSE-GDFE preprocessing to the real transmit system. Decompose the extended channel matrix $\bar{\mathbf{H}}$ into an $(n+m) \times m$ matrix $\bar{\mathbf{Q}}$ and an $m \times m$ upper triangular matrix $\mathbf{R}_1$ using QR-decomposition. Obtain the matrix $\mathbf{Q}_1$ by taking the first n columns of $\bar{\mathbf{Q}}$. MMSE-GDFE preprocessing is done as follows to obtain a new upper triangular transmission system:

$$\mathbf{xD} = \mathbf{Q}_1^H \mathbf{x}, \tag{4.17}$$

where $\mathbf{xD}$ is an $m \times 1$ vector. The equivalent new upper triangular transmission system after the preprocessing is therefore,

$$\mathbf{xD} = \mathbf{R}_1 \mathbf{s} + \mathbf{nD}, \tag{4.18}$$

where $\mathbf{nD}$ is the $m \times 1$ non-Gaussian noise vector obtained because of the MMSE-GDFE front-end filtering. It is the non-Gaussian distribution of the noise that introduces suboptimality in the decoding process.

Since $\mathbf{s}$ contains elements belonging to the $\sqrt{\mathcal{M}}$-PAM constellation, such that $s_i \in \{\pm 1, \pm 3, \ldots, \pm(\sqrt{\mathcal{M}} - 1)\}$, the translation $\mathbf{s} = 2\acute{\mathbf{s}} - \mathbf{1}_{m \times 1}$ results in a $m \times 1$ vector $\acute{\mathbf{s}}$ containing elements from the integer set.

The real system in the equation above can therefore be re-written as

$$\begin{aligned} \mathbf{xD} &= \mathbf{R}_1(2\acute{\mathbf{s}} - \mathbf{1}_{m \times 1}) + \mathbf{nD} && (4.19) \\ &= 2\mathbf{R}_1 \acute{\mathbf{s}} - \mathbf{R}_1 \mathbf{1}_{m \times 1} + \mathbf{nD}. && (4.20) \end{aligned}$$

This system can be translated into,

$$\begin{aligned} \mathbf{xD}_1 &= \mathbf{xD} + \mathbf{R}_1\mathbf{1}_{m\times 1} && (4.21)\\ &= 2\mathbf{R}_1\acute{\mathbf{s}} + \mathbf{nD} && (4.22)\\ &= \mathbf{R}_{1d}\acute{\mathbf{s}} + \mathbf{nD}, && (4.23) \end{aligned}$$

where $\mathbf{R}_{1d} = 2\mathbf{R}_1$.

Since the system is now in a form to which lattice decoding can be applied, we perform lattice reduction on the effective channel matrix, $\mathbf{R}_{1d}$, to obtain a more orthogonal basis that will aid in the decoding process. Using the LLL reduction technique, we obtain the corresponding reduced basis $\mathbf{H}_2$, where $\mathbf{H}_2 = \mathbf{R}_{1d}\mathbf{T}_d$. The altered system then becomes,

$$\begin{aligned} \mathbf{xD}_1 &= (\mathbf{R}_{1d}\mathbf{T}_d)(\mathbf{T}_d^{-1}\acute{\mathbf{s}}) + \mathbf{nD} && (4.24)\\ &= \mathbf{H}_2\mathbf{z} + \mathbf{nD}, && (4.25) \end{aligned}$$

where $\mathbf{z} = (\mathbf{T}_d)^{-1}\acute{\mathbf{s}}$. Since the system is real and $\mathbf{T}_d$ is a unimodular matrix, by definition $\mathbf{T}_d$ and its inverse contain integers only. Hence the altered system has the unknown vector $\mathbf{z}$, containing integers and so lattice decoding can still be performed on this altered system.

Finally, we use Greedy Ordering to permute the system, in order to obtain better performance. The purpose of Greedy Ordering is to obtain a permutation of the channel matrix that results in a QR-decomposition with the property that the smallest diagonal element of the upper triangular matrix is maximized. This alteration of the channel matrix is beneficial as the system will undergo Sphere Decoding and having large diagonal entries of the upper triangular matrix will result in decreasing the range of the intervals of the Pohst Enumeration, thereby decreasing complexity and

therefore enhancing efficiency of the Sphere Decoder. The permuted channel matrix is denoted by $\mathbf{H}_{2g}$, such that $\mathbf{H}_{2g} = \mathbf{H}_2\mathbf{P}_g$, where $\mathbf{P}_g$ is the permutation matrix. The system is therefore modified as follows:

$$\begin{aligned} \mathbf{xD}_1 &= (\mathbf{H}_2\mathbf{P}_g)(\mathbf{P}_g^{-1}\mathbf{z}) + \mathbf{nD} && (4.26) \\ &= \mathbf{H}_{2g}\mathbf{z}_1 + \mathbf{nD}, && (4.27) \end{aligned}$$

where $\mathbf{z}_1 = \mathbf{P}_g^{-1}\mathbf{z}$.

A QR-decomposition is applied to the new effective channel matrix $\mathbf{H}_{2g}$, to obtain the factors $\mathbf{Q}_g$ and $\mathbf{R}_g$. The system is finally translated into an upper triangular form by the following multiplication

$$\begin{aligned} \mathbf{xD}_2 &= \mathbf{Q}_g^H\mathbf{xD}_1 && (4.28) \\ &= \mathbf{R}_g\mathbf{z}_1 + \mathbf{nD}_1, && (4.29) \end{aligned}$$

where $\mathbf{nD}_1 = \mathbf{Q}_g^H\mathbf{nD}$. The final effective upper triangular matrix $\mathbf{R}_g$, the effective receive-vector $\mathbf{xD}_2$, and the radius are input to the Sphere Decoder. The output of the Sphere Decoder is the $m \times 1$ vector $\hat{\mathbf{z}}_1$ which is an estimate of the vector $\mathbf{z}_1$. The estimate of the vector $\mathbf{z}$ is obtained my permuting the elements of $\hat{\mathbf{z}}_1$ back to their original order. This is done by

$$\hat{\mathbf{z}} = \mathbf{P}_g\hat{\mathbf{z}}_1. \tag{4.30}$$

Since we performed the decoding in a reduced lattice basis, the decoded output needs to be transferred back to the original basis by the following,

$$\hat{\hat{\mathbf{s}}} = \mathbf{T}_d\hat{\mathbf{z}}, \tag{4.31}$$

where $\hat{\acute{\mathbf{s}}}$ is the estimate of the $m \times 1$ vector $\acute{\mathbf{s}}$ consisting of integer elements. The estimated vector is then translated back so that its elements belong to the set of odd numbers using

$$\hat{\mathbf{s}_1} = 2\hat{\acute{\mathbf{s}}} - \mathbf{1}_{m\times 1}. \tag{4.32}$$

Since the original transmitted real vector, $\mathbf{s}$, contained elements belonging to the $\sqrt{\mathcal{M}}$-PAM constellation strictly, we require a quantization that limits the elements of the estimate to within the boundaries of the PAM constellation. The estimate $\hat{\mathbf{s}} = \mathcal{Q}(\hat{\mathbf{s}_1})$, therefore contains the decoded estimate of the real transmit vector $\mathbf{s}$.

4.4 MMSE-GDFE Preprocessing

Applying the MMSE-GDFE filter involves first defining the extended channel matrix $\bar{\mathbf{H}}$ such that,

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{I}_m \end{bmatrix},$$

$\bar{\mathbf{H}}$ of dimensions $(n+m) \times m$ undergoes QR decomposition to give $\bar{\mathbf{H}} = \bar{\mathbf{Q}}\mathbf{R}_1$, where $\bar{\mathbf{Q}}$ of dimensions $(n+m) \times m$ has orthonormal columns, and $\mathbf{R}_1$ of dimensions $m \times m$ is an upper triangular matrix; each diagonal element of $\mathbf{R}_1$ is positive. The submatrix $\mathbf{Q}_1$, comprising of the first n rows of $\bar{\mathbf{Q}}$ satisfies $\mathbf{H} = \mathbf{Q}_1\mathbf{R}_1$. However, it is important to note that since the columns of $\mathbf{Q}_1$ are in general non-orthogonal, the relation $\mathbf{Q}_1^H\mathbf{H} = \mathbf{R}_1$, cannot be claimed.

The MMSE-GDFE forward and backward filters are thus given by $\mathbf{Q}_1$ and $\mathbf{R}_1$,

respectively. Since $\bar{\mathbf{Q}}$ is orthonormal, we have

$$\begin{aligned}
\bar{\mathbf{H}}^H\bar{\mathbf{H}} &= (\bar{\mathbf{Q}}\mathbf{R}_1)^H(\bar{\mathbf{Q}}\mathbf{R}_1) && (4.33)\\
&= \mathbf{R}_1^H\bar{\mathbf{Q}}^H\bar{\mathbf{Q}}\mathbf{R}_1 && (4.34)\\
&= \mathbf{R}_1^H\mathbf{I}_m\mathbf{R}_1 && (4.35)\\
&= \mathbf{R}_1^H\mathbf{R}_1. && (4.36)
\end{aligned}$$

Also, by the definition of $\bar{\mathbf{H}}$, we have

$$\bar{\mathbf{H}}^H\bar{\mathbf{H}} = \mathbf{H}^H\mathbf{H} + \mathbf{I}_m. \tag{4.37}$$

Then $\mathbf{R}_1^H\mathbf{R}_1 = \mathbf{H}^H\mathbf{H} + \mathbf{I}_m$, and hence the backward filter $\mathbf{R}_1$ is always invertible for finite SNR values.

Applying the forward filter to the received signal we obtain the following

$$\begin{aligned}
\mathbf{xD} &= \mathbf{Q}_1^H\mathbf{x} && (4.38)\\
&= \mathbf{Q}_1^H(\mathbf{Hs} + \mathbf{w}) && (4.39)\\
&= \mathbf{Q}_1^H\mathbf{Hs} + \mathbf{Q}_1^H\mathbf{w} && (4.40)\\
&= \mathbf{Q}_1^H\mathbf{Hs} + \mathbf{Q}_1^H\mathbf{w} + \mathbf{R}_1\mathbf{s} - \mathbf{R}_1\mathbf{s} && (4.41)\\
&= \mathbf{R}_1\mathbf{s} + \underbrace{(\mathbf{Q}_1^H\mathbf{H} - \mathbf{R}_1)\mathbf{s} + \mathbf{Q}_1^H\mathbf{w}}_{\mathbf{nD}}. && (4.42)
\end{aligned}$$

The effective noise term, $\mathbf{nD}$, includes a Gaussian component, $\mathbf{Q}_1^H\mathbf{w}$, and a non-Gaussian component, $(\mathbf{Q}_1^H\mathbf{H} - \mathbf{R}_1)\mathbf{s}$; $\mathbf{nD}$, therefore, does not have a Gaussian distribution. Note that the non-Gaussian component of $\mathbf{nD}$ is a self-interference term, as it is a function of the transmitted message, $\mathbf{s}$.

The non-Gaussian distribution of the effective noise vector has the important implication that the minimum Euclidean distance is now in general not the optimal

metric for decoding. Since we will proceed to use decoding strategies that aim to minimize the minimum Euclidean distance, MMSE-GDFE preprocessing of these decoding strategies will introduce a suboptimality.

4.4.1 On the Suboptimality (and Optimality) of MMSE-GDFE Preprocessing

MMSE-GDFE preprocessing is known to compromise the optimality of the Minimum Distance (MD) decoding scheme as it results in the introduction of the self-noise term. This stems from the fact that preprocessing involves applying a QR-decomposition on the extended real channel matrix $\bar{\mathbf{H}}$, rather than the actual real channel matrix, $\mathbf{H}$, and then extracting only the first n columns of the matrix $\bar{\mathbf{Q}}$, which has orthonormal columns, to get the matrix $\mathbf{Q}_1$, which does not have orthogonal columns in general. When $\mathbf{Q}_1$ is used to translate the system into the upper triangular system, the distribution of the noise term in the original system is not preserved and hence $\mathbf{nD}$ is non-Gaussian in general. This results in the MD decoder not being equivalent to the ML decoder and therefore suboptimality is introduced in the decoding process.

In [30] it is proved that any constant modulus QAM constellation, or equivalently any constant modulus PAM constellation for the corresponding real system, results in MD decoding equivalent to ML decoding even with MMSE-GDFE preprocessing. In other words, MMSE-GDFE preprocessing does not introduce a suboptimality in the MD decoding of systems that employ 4-QAM or BPSK modulation schemes. Hence, Sphere Decoding of MMSE-GDFE preprocessed 4-QAM or BPSK systems is optimal. Simply put, although the matrix $\mathbf{Q}_1$ still has in general non-orthogonal columns in the case of 4-QAM or BPSK modulations (as the channel is independent of the modulation being employed), and the preprocessed noise, $\mathbf{nD}$, therefore non-Gaussian, the minimum distance rule for optimal decoding is still preserved when constant modulus constellations are used. For non-constant modulus constellations,

however, it cannot be guaranteed; hence MD decoding in these cases is not equivalent to ML decoding.

To emphasize on the suboptimality introduced by MMSE-GDFE preprocessing, let us analyze the minimum distance decoding rule when applied to the original system and compare it with the analysis when it is applied to the system after MMSE-GDFE preprocessing.

For a system $\mathbf{x} = \mathbf{Hs} + \mathbf{w}$, the MD solution (which is the ML solution) is $\hat{\mathbf{s}}_{ML} = \arg\min_{s \in \mathcal{M}^m} ||\mathbf{x} - \mathbf{Hs}||^2$, and so

$$||\mathbf{x} - \mathbf{H}\hat{\mathbf{s}}_{ML}||^2 \leq ||\mathbf{x} - \mathbf{Hs}||^2 \qquad \forall \mathbf{s} \in \mathcal{M}^m,$$

which can equivalently be written as

$$\mathbf{x}^H\mathbf{x} + \hat{\mathbf{s}}_{ML}^H\mathbf{H}^H\mathbf{H}\hat{\mathbf{s}}_{ML} - 2\mathbf{x}^H\mathbf{H}\hat{\mathbf{s}}_{ML} \leq \mathbf{x}^H\mathbf{x} + \mathbf{s}^H\mathbf{H}^H\mathbf{Hs} - 2\mathbf{x}^H\mathbf{Hs},$$

and therefore

$$\hat{\mathbf{s}}_{ML}^H\mathbf{H}^H\mathbf{H}\hat{\mathbf{s}}_{ML} - \mathbf{s}^H\mathbf{H}^H\mathbf{Hs} + 2\mathbf{x}^H\mathbf{H}(\mathbf{s} - \hat{\mathbf{s}}_{ML}) \leq 0.$$

When the MMSE-GDFE preprocessed system $\mathbf{xD} = \mathbf{R}_1\mathbf{s} + \mathbf{nD}$ is used to decode $\mathbf{s}$ with MD decoding, the metric being minimized is $||\mathbf{xD} - \mathbf{R}_1\mathbf{s}||^2$. Denote the decoded message by $\hat{\mathbf{s}}$. Then,

$$||\mathbf{xD} - \mathbf{R}_1\hat{\mathbf{s}}||^2 \leq ||\mathbf{xD} - \mathbf{R}_1\mathbf{s}||^2 \qquad \forall \mathbf{s} \in \chi_k^m,$$

which is equivalent to

$$\hat{\mathbf{s}}^H\mathbf{R}_1^H\mathbf{R}_1\hat{\mathbf{s}} - \mathbf{s}^H\mathbf{R}_1^H\mathbf{R}_1\mathbf{s} + 2\mathbf{xD}^H\mathbf{R}_1(\mathbf{s} - \hat{\mathbf{s}}) \leq 0.$$

Since $\mathbf{x}\mathbf{D}^H = \mathbf{x}^H\mathbf{Q}_1$, the above equation can be rewritten as,

$$\hat{\mathbf{s}}^H\mathbf{R}_1^H\mathbf{R}_1\hat{\mathbf{s}} - \mathbf{s}^H\mathbf{R}_1^H\mathbf{R}_1\mathbf{s} + 2\mathbf{x}^H\mathbf{H}(\mathbf{s} - \hat{\mathbf{s}}) \leq 0.$$

As stated earlier, $\mathbf{R}_1^H\mathbf{R}_1 = \mathbf{H}^H\mathbf{H} + \mathbf{I}_m$, and the detection rule is therefore,

$$\hat{\mathbf{s}}^H\mathbf{H}^H\mathbf{H}\hat{\mathbf{s}} - \mathbf{s}^H\mathbf{H}^H\mathbf{H}\mathbf{s} + 2\mathbf{x}^H\mathbf{H}(\mathbf{s} - \hat{\mathbf{s}}) + \underbrace{\hat{\mathbf{s}}^H\hat{\mathbf{s}} - \mathbf{s}^H\mathbf{s}}_{suboptimality} \leq 0.$$

The rule can now be observed as the ML decoding rule with an additional term which introduces the suboptimality. It is important to note that if the term that introduces suboptimality is non-negative, then the detection rule will still be optimal. For a general constellation, it cannot be guaranteed that this term is non-negative, as the term may be negative for some combinations of $\mathbf{s}$ and $\hat{\mathbf{s}}$ and non-negative for others.

In the case of a 4-QAM constellation (or any other constant-modulus constellation), however, the term $\hat{\mathbf{s}}^H\hat{\mathbf{s}} - \mathbf{s}^H\mathbf{s}$ causing the suboptimality is always zero, and the minimum distance rule for the MMSE-GDFE preprocessed system is therefore equivalent to the minimum distance rule for the original system with Gaussian noise. The MD rule for MMSE-GDFE preprocessed constant-modulus constellation systems is therefore equivalent to the ML rule, and no suboptimality is introduced by the preprocessing.

Chapter 5

Decoding of Large Underdetermined Systems using Lattice Decoding Techniques

5.1 Introduction

We applied a lattice decoding technique for the decoding of underdetermined MIMO systems in Chapter 4, when the Sphere Decoder preceded by MMSE-GDFE preprocessing, Lattice Reduction, and Greedy Ordering was used. Since MMSE-GDFE preprocessing is, in general, a suboptimal technique and as we already established that optimality is not the most important criteria when decoding a system, especially one as difficult as a large underdetermined MIMO system; following up the generally suboptimal preprocessing with a suboptimal decoding technique should therefore not be concerning. In our work with underdetermined systems, we opt for the suboptimal Sequential Decoder as this allows us the flexibility to trade performance for complexity and vice versa.

5.2 Framework

5.2.1 MMSE-GDFE Preprocessing Followed by Sequential Decoding

In our work with underdetermined systems, we apply the MMSE-GDFE filter to the real channel matrix to get not only a better channel, but to effectively obtain an overdetermined system to which lattice decoding can be applied. In essence, we increase the rank of the channel matrix from n to m, by applying this preprocessing.

As described before, a QR decomposition is applied to the extended real channel matrix and the first n rows of the extended matrix $\bar{\mathbf{Q}}$ are taken, and used to obtain the $m \times m$ upper triangular system $\mathbf{xD} = \mathbf{R}_1\mathbf{s} + \mathbf{nD}$ of rank m, as shown in Eqns(4.17) and (4.18).

As the elements of the complex transmitted message, $\mathbf{s}_c$, are drawn from the $\mathcal{M}$-QAM constellation, the real counterpart of the transmit message belong to χ_k^m and its elements therefore to $\{\pm 1, \pm 3, \ldots, \pm(\sqrt{\mathcal{M}} - 1)\}$. Since lattice decoding searches over the set of real integers, the translation $\mathbf{s} = 2\acute{\mathbf{s}} - \mathbf{1}$, is applied to $\mathbf{s}$, to obtain $\acute{\mathbf{s}}$ belonging to $\mathbb{Z}^m$. The system is therefore modified as shown in Eqns (4.19) to (4.22), to obtain the $m \times m$ system, $\mathbf{xD}_1 = \mathbf{R}_{1d}\acute{\mathbf{s}} + \mathbf{nD}$ of Eqn (4.23).

Since the system is in an upper triangular form, we input the matrix $\mathbf{R}_{1d}$, the vector $\mathbf{xD}_1$ and the bias, to the Sequential Decoder to obtain the estimate $\hat{\acute{\mathbf{s}}}$ of the altered transmit vector $\acute{\mathbf{s}}$. Since $\hat{\acute{\mathbf{s}}}$ belongs to $\mathbb{Z}^m$, this is translated to obtain $\hat{\mathbf{s}}_f = 2\hat{\acute{\mathbf{s}}} - \mathbf{1}$.

The vector $\hat{\mathbf{s}}_f$, belongs to the set of infinite odd numbers; a quantization operation is therefore required to limit the estimated vector, so that each entry belongs to the $\sqrt{\mathcal{M}}$-PAM constellation and the vector to χ_k^m. This is done as follows,

$$\hat{\mathbf{s}} = \mathcal{Q}(\hat{\mathbf{s}}_f), \tag{5.1}$$

and $\hat{\mathbf{s}}$ is the final estimate of the scheme.

5.2.2 MMSE-GDFE Preprocessing, Lattice Reduction and Greedy Ordering Followed by Sequential Decoding

In this scheme, like the scheme proposed in [28], preprocessing is done by applying MMSE-GDFE to the real channel matrix, followed by Lattice Reduction and finally Greedy Ordering. The obtained system is then decoded using Sequential Decoding. This scheme is used to observe the effect of Lattice Reduction and Greedy Ordering in the preprocessing, before Sequential Decoding is applied to an underdetermined system that has been made full-rank using MMSE-GDFE.

The real transmission system $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$ therefore undergoes the transformations in Eqns (4.17 - 4.29), to obtain the $m \times m$ system $\mathbf{x}\mathbf{D}_2 = \mathbf{R}_g\mathbf{z}_1 + \mathbf{n}\mathbf{D}_1$. The vector $\mathbf{x}\mathbf{D}_2$, the matrix $\mathbf{R}_g$ and the bias are input to the Sequential Decoder, and $\hat{\mathbf{z}}_1$, the estimate of $\mathbf{z}_1$ is obtained. The elements of $\hat{\mathbf{z}}_1$ are then permuted back to their original order using Eqn (4.30), after which Eqn (4.31) is used to transform the estimates back to their original lattice basis, and Eqn (4.32) transforms the obtained integer vector into an odd-integer vector set. As the original transmitted message belonged to the $\mathcal{M}$-QAM constellation and lattice decoding does not have boundary control, a quantization operation is applied to limit the estimated vector to the $\mathcal{M}$-QAM constellation.

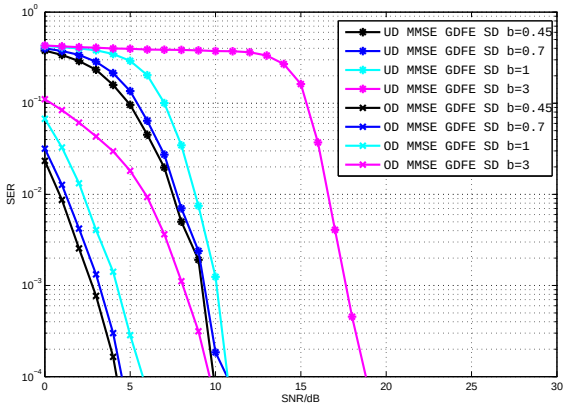
5.3 Results and Analysis

Figure 5.1a and Figure 5.1b are the performance vs. SNR plots, for both the 20×20 overdetermined and 10×20 underdetermined transmission systems, employing 4-QAM and 16-QAM constellations respectively. The decoding technique used is the MMSE-GDFE preprocessed Sequential Decoding mentioned in Section 5.2.1

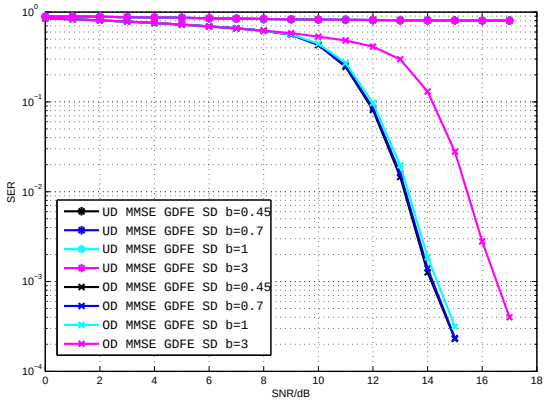
with different bias values. As can be seen from the plots, the error for the overdetermined systems falls much faster than for the underdetermined case, which is expected as overdetermined systems have much better channels even after both systems have undergone preprocessing. In the case of the 4-QAM system, a large gap exists between the overdetermined and underdetermined case, and the gap increases as the constellation size grows; in the case of 16-QAM the gap is too large for us to see the underdetermined system's error curves drop in the simulated SNR range. Also, the gap between the overdetermined and underdetermined error curves increases as bias is increased.

The corresponding complexity graphs are plotted in Figure 5.1c and Figure 5.1d. From Figure 5.1c we see that the for the underdetermined case, the complexity trend is similar to that for Sequential Decoding without MMSE-GDFE preprocessing applied to overdetermined systems and mentioned in the results in Chapter 3 - a region of low complexity and high error, followed by a region where complexity peaks and the error curve starts to fall, and finally a region of low complexity and low error. Since the error curves for MMSE-GDFE preprocessed Sequential Decoder applied to the overdetermined case fall at very low SNR, we are not able to observe the first of the aforementioned three parts in the complexity graph, but only see a region of high complexity when the error curve is falling, followed by a region of low complexity at low error when SNR is high. The three regions are more visible for the overdetermined case in Figure 5.1d, because the error curves for the 16-QAM case fall at SNR a little higher than in the 4-QAM case, so the complexity curves are shifted accordingly as well. In the 16-QAM case the third region of the complexity graph is not visible for the underdetermined case as the error curves never fell in the SNR region being observed.

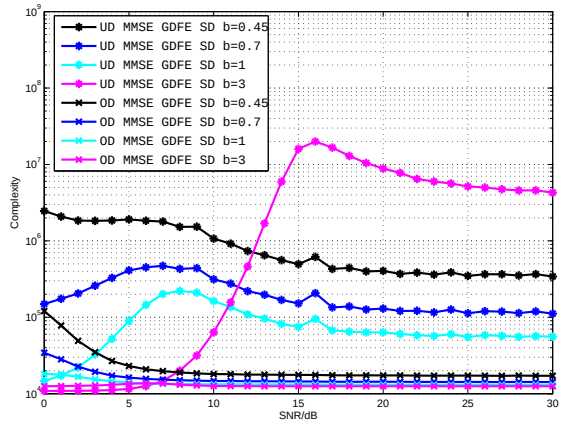
Notice, for both the overdetermined and undetermined cases, like the Sequential Decoding in Chapter 3, lower bias values correspond to error curves falling faster



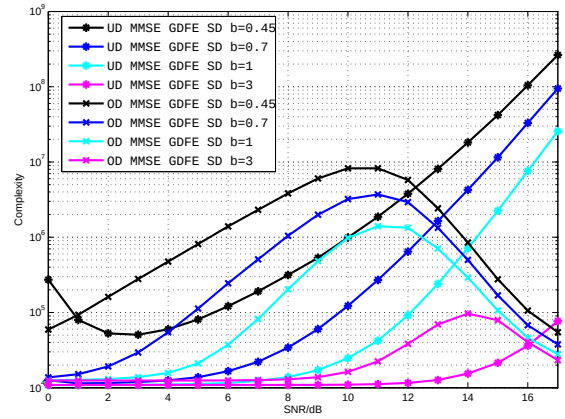
(a) Performance with increasing SNR for 4-QAM.



(b) Performance with increasing SNR for 16-QAM.



(c) Complexity with increasing SNR for 4-QAM.

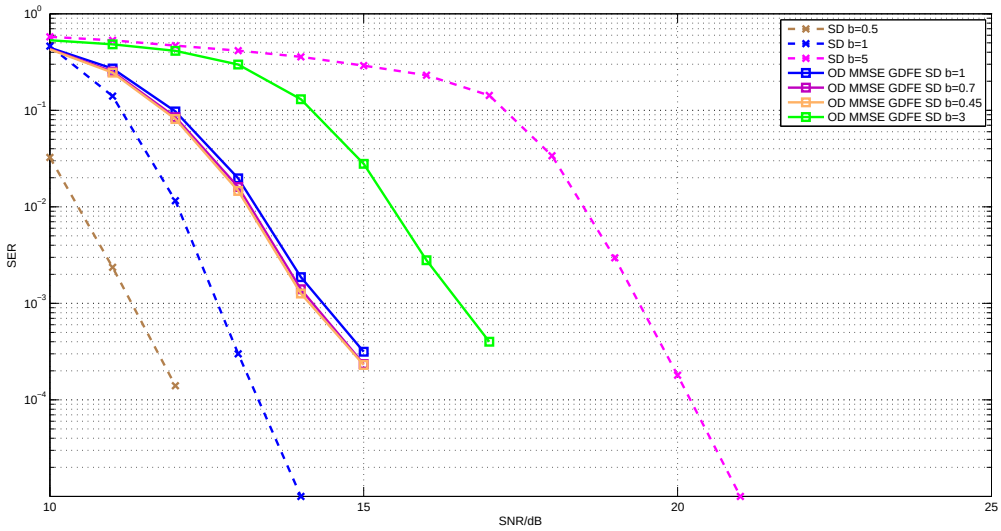


(d) Complexity with increasing SNR for 16-QAM.

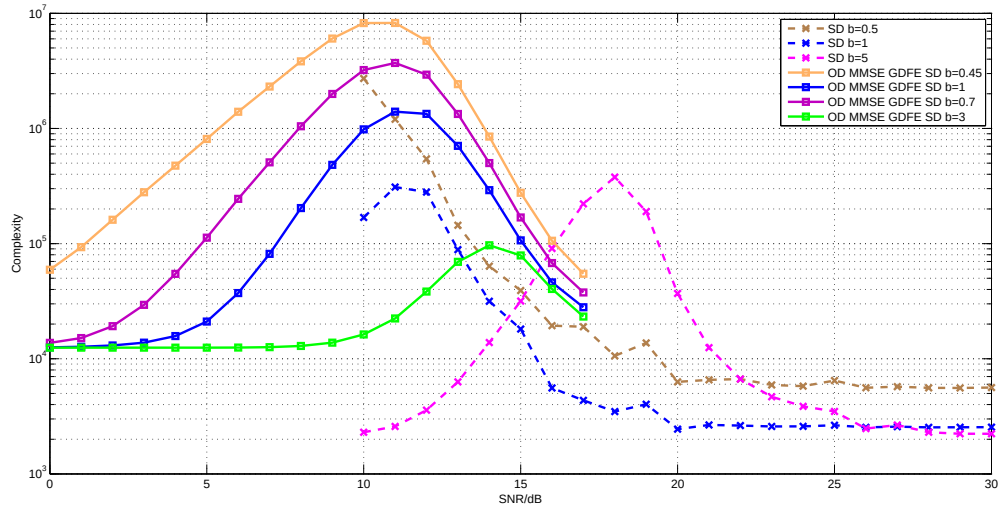
Figure 5.1: MMSE-GDFE preprocessed Sequenital Decoding for 20×20 and 10×20 Systems

and therefore the complexity peak occurring faster. However, in Chapter 3, lower bias values also corresponded to narrower high complexity regions, this does not seem to be the case in MMSE-GDFE preprocessed Sequential decoders. In fact, with decreasing bias, the complexity peaks tend to widen.

Figure 5.2a shows the error curves for a 20×20 system employing 16-QAM. This is used to show the effect of MMSE-GDFE on Sequential Decoding by applying it to the overdetermined case. The suboptimality introduced by MMSE-GDFE can



(a) Performance with increasing SNR.



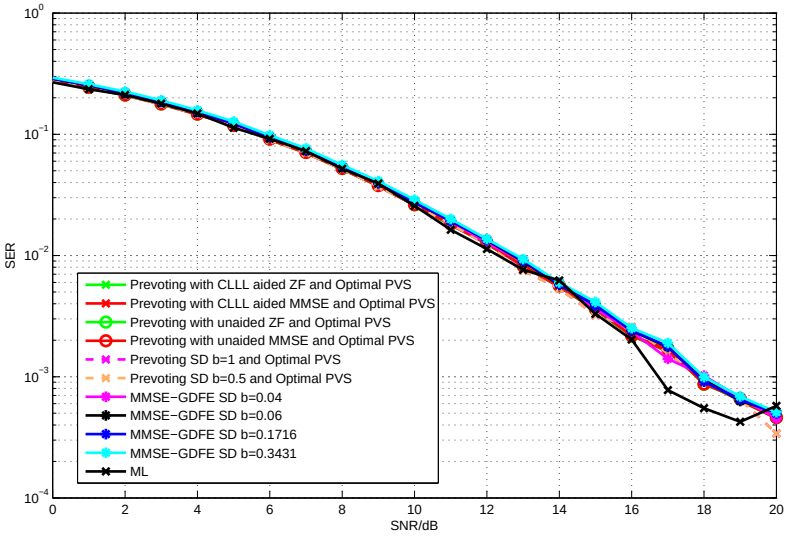
(b) Complexity with increasing SNR.

Figure 5.2: Performance and complexity of the Sequential Decoder, with and without MMSE-GDFE preprocessing, for a 20×20 MIMO System employing 16-QAM.

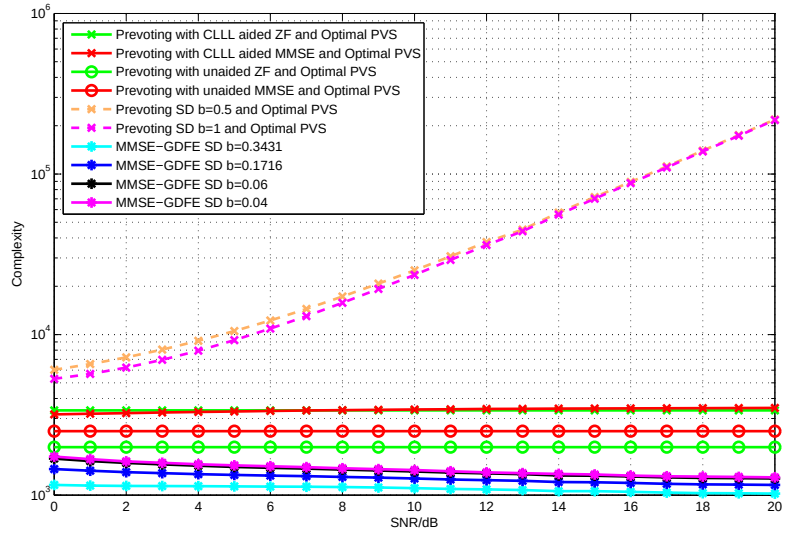
be seen, as the error curves of the Sequential Decoding preprocessed with MMSE-GDFE are shifted to the right from the error curves of Sequential Decoding that is not preprocessed. Additionally, from the complexity plot in Figure 5.2b, we see that MMSE-GDFE preprocessed Sequential Decoders have higher complexity than the unprocessed for the same bias value. The widening of the high-complexity range, as bias value is decreased, can be seen in the plot as well. The two decoding strategies were simulated in different SNR ranges to avoid excess high-complexity computations, and the error curves are plotted to show the portions of significance, the complexity is however plotted throughout the SNR range to show the different trends.

Figure 5.3a is a plot of the error performance with increasing SNR for an under-determined system with 2 receive and 4 transmit antennas. A 4-QAM constellation was employed for transmission. From the plot, all the schemes being used have almost equivalent performance and this is extremely close to that of the ML detector's performance. It should be noted that as the ML detection is computationally very expensive, the simulations for it were averaged for a lower number of iterations and thus this curve is rougher than the others; as it is only being used as a benchmark, our work remains unaffected. From the figure we therefore see that there is negligible performance difference in the case of small systems at least, between the Prevoting with Optimal PVS that employs the Sequential Decoder, LR-aided LDs and even un-aided LDs. Also, the MMSE-GDFE preprocessed Sequential Decoder scheme performs similarly.

Figure 5.3b plots the complexity of these schemes. Among the schemes that employ Prevoting with Optimal PVS, as expected, the complexity in the case of the unaided LDs is the lowest, followed by that of LR-aided LDs, and the highest is that of Sequential Decoders. It should be noted that the Prevoting with Optimal PVS that employs Sequential Decoding has a significantly higher complexity, and this increases with SNR, unlike the other Prevoting with Optimal PVS schemes. Since, Prevoting



(a) Performance of different detectors with increasing SNR.



(b) Complexity of different detectors with increasing SNR.

Figure 5.3: Performance and complexity of different detectors for a 2×4 MIMO System employing 4-QAM.

with Optimal PVS that uses LR-aided LDs in general reaches performance very close to that of the ML, it is unwise to use a scheme that does not have a significant performance gain to provide for the significant complexity increase it causes. We will thus not discuss Prevoting with Optimal PVS using Sequential Decoding further.

As can be seen from the complexity plot, the MMSE-GDFE aided Sequential Decoders outperform the schemes that employ Prevoting with Optimal PVS in terms of complexity. This can be explained by the exhaustive search part of the Prevoting Scheme which results in complexity exponential in the difference between the unknowns and knowns, as well as the expensive Optimal PVS calculation. Due to the nature of the Prevoting based cancellation technique, we did not apply it to larger systems and opted instead for MMSE-GDFE preprocessed Sequential Decoding.

It ought to be mentioned here that the complexity analysis done for Figures 5.1, 5.2, 5.3, as well as the simulations in Chapter 3, measures the number of flops counted as operations are done. Each real addition is counted as one flop, each real multiplication as one flop, each complex addition as two flops, each complex multiplication as six flops and the multiplication of a real number with a complex number as two flops. Rounding operations and hard-limiting are considered negligible. For Figures 5.4, 5.5, and 5.6, as a lot of matrix operations were involved, particularly with the preprocessing techniques used, it was considered wiser to measure complexity using Matlab's inbuilt tic-toc function which measures CPU time. Due to this, the 'value' of complexity may be very different in the graphs that measure the number of flops and those that use the tic-toc function, but since we are interested in the trends of the complexity, our work remains unaffected.

Figure 5.4a is a plot of the error with increasing SNR for a 2×10 transmission system employing 4-QAM. The size of the system was chosen to model a realistic scenario with a mobile that has 2 antennas and a larger base station with 10. The figure shows plots of the following: 1) the scheme employed in [28], namely a preprocessing that

involves applying MMSE-GDFE, Lattice Reduction and Greedy Ordering, followed by Sphere Decoding, 2) the preprocessing of [28] followed by the search for the Babai Point, 3) the MMSE-GDFE preprocessed Sequential Decoder scheme in Section 5.2.1 and 4) the scheme in Section 5.2.2 that involves the preprocessing from [28] followed by Sequential Decoding. A range of bias values is used for 3) and 4) to observe the performance complexity trade-off.

As MMSE-GDFE when used with a 4-QAM constellation, does not introduce suboptimality in the minimum distance decoder, and as Sphere Decoding is an optimal decoding technique, for this particular simulation, 1) is an optimal detector. As can be seen from the plot, 1) therefore has the lowest error curves. The error curves of the SD schemes in 3) and 4) when used with small enough bias values result in performance extremely close to that of the optimal detector. In the cases of the bias values of 0.3431 and 0.1716 shown in the figure, the difference is almost negligible. For larger bias values, however, the performance of the Sequential Decoding schemes in 3) and 4) have a non-negligible gap from the optimal performance as can be seen in the cases of bias values equal to 0.5 and 1, and the gap increases as the bias value becomes larger. It should be noted that for any bias value, the performance of the scheme in 4) is better than that of the scheme in 3), particularly for lower SNR values. Schemes 1), 3) and 4) achieve receive diversity, but scheme 2) does not and a large gap is seen to exist between the performance of 2) and the other schemes.

Figure 5.4b compares the complexities of the above mentioned schemes. Since finding the Babai Point has low complexity, the complexity of 2) is dominated by that of the preprocessing scheme, and is not very high. The complexity of 1) is dominated by the Sphere Decoding process. In our simulations we used as small a radius as possible corresponding to each SNR, that would not result in an empty sphere. The radii for each SNR were found using trial and error. In practical situations this may not be possible and a larger radius would have to be used to ensure a non-empty

sphere at the expense of higher complexity. The complexity of scheme 2) in this plot should therefore be taken as a sort of lower bound of complexity for this scheme. The few peaks in the complexity curve correspond to the use of an updated larger radius as the SNR is increased.

The complexity of the Sequential Decoders in 3) increases with SNR, corresponding to the drop in error. This becomes obvious in the case of a bias value of 1, where an increase in complexity can be seen when the error falls. In general, the complexity of scheme 4), impressively, is lower than that of the scheme in 3). Also, it does not increase with SNR for any bias value. This implies the superiority of the scheme in 4) over that in 3) as it results in lower error at lower complexity. The gap in the complexity curves increases for lower bias values, but decreases for higher bias values. For the high bias value of 1, in the low SNR region, the gap decreases so much that the complexity of scheme 3) is actually lower than that of 4). This, however, only happens because scheme 3) is in its first phase and has low complexity and low error. Scheme 4) too, is in the first phase, but has higher complexity due its more expensive preprocessing stage. With increasing SNR, the complexity of scheme 3) becomes larger than that of 4).

The performance-complexity trade-off advantage could be explained by the superiority of the preprocessing employed in scheme 4). It may result in the matrix being input to the Sequential Decoder to be tamed well enough that the complexity of the Sequential Decoding operation is decreased dramatically. This is in contrast to 3) where the preprocessing is less expensive than the preprocessing of 4), but the complexity of the Sequential Decoder much larger. It should be noted that for scheme 4) using small bias values such as 0.3431 and 0.1716, we are able to attain near-optimal performance at complexity much lower than that of the optimal scheme in 1).

Figure 5.5a is a plot of the performance for a system with a fixed number of receive antennas, $N = 18$, and increasing transmit antennas, M. All values of M being

considered are greater than N, and hence the system is always underdetermined. A 4-QAM constellation is employed for each system and the SNR is fixed to 3dB. The decoders employed are the two Sequential Decoding schemes in 5.2.1 and 5.2.2. The error curves increase with increasing transmit antennas. One may be tempted to think that in the overdetermined case, increasing the system size resulted in error falling when Sequential Decoders with bias values small enough were employed. However, in this case, increasing the system size makes decoding harder. This stems from the fact that the number of receivers is fixed and we are dealing with the underdetermined case, as the number of transmit antennas increases, the difference between the number of unknowns and equations increases; thus we deal with systems more difficult to decode as M is increased.

In general, the error of the scheme proposed in 5.2.2 is lower than that of the scheme in Section 5.2.1, but this difference decreases with increasing M, and decreases with decreasing bias as well. We have not simulated results with the Sphere Decoder employing the preprocessing from [28], as computing a radius small enough for each system size is a very tedious task, and employing a large radius would result in extremely high complexity for these large systems.

The corresponding complexity graphs are plotted in Figure 5.5b. As can be seen from the curves, the complexity increases with increasing system size, as it should. The scheme in Section 5.2.1 has higher complexity than the Scheme in Section 5.2.2 for the lower bias values particularly. The complexity curve for the higher bias value of 1, is much lower than the other complexity curves, but this is due to the fact that at the low SNR being employed, the decoder with a bias value of 1 is in the stage when SNR is low and the complexity is therefore also low and error is high. At a higher SNR the complexity differences between the two schemes would be more pronounced.

The complexity of the Lattice Reduction and Greedy Ordering operations involved in the scheme of Section 5.2.2 are also plotted. It is interesting to note that

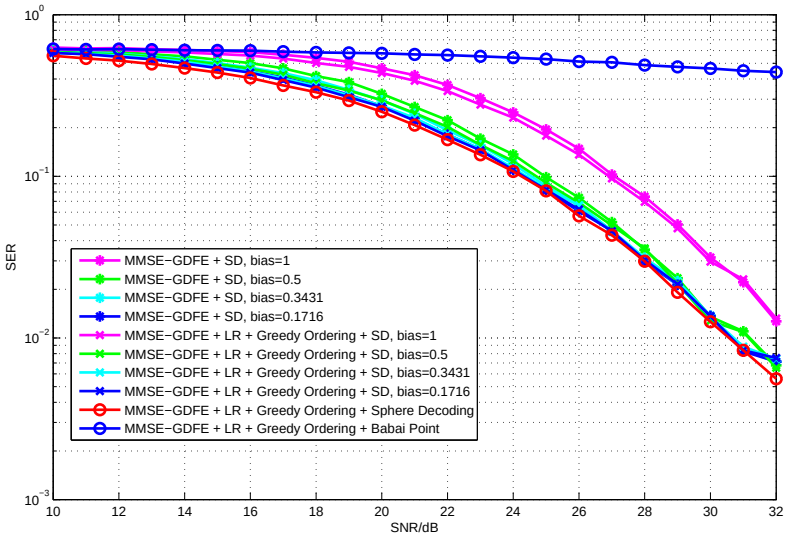
the preprocessing, and in particular the Greedy Ordering employed, dominates the complexity for smaller systems especially. The preprocessing tames the channel being input to the Sequential Decoder so well that the complexity of the Sequential Decoder plays a much smaller part in the total complexity of the scheme. For larger bias values such as that of b=1, we see that there is an almost negligible gap between the complexity curves for Greedy Ordering and the entire scheme. As bias is reduced, the complexity of the Sequential Decoding operation increases, and the gap between the curves increases. The difference is still, however, not very drastic. Increasing the number of transmit antennas makes the system more difficult to decode, and the gap between the Greedy Ordering complexity curve and the total system complexity curve increases.

Figure 5.6a is the error plot against the increasing number of receive antennas. MIMO systems using 4-QAM and an SNR of 3dB are employed. The number of transmit antennas is fixed to $M = 24$, and N is increased. The decoding schemes in Section 5.2.1 and 5.2.2 are employed with different bias values. From the plot, the error curves fall for all decoding schemes as N increases. The number of transmit antennas is constant and the size of the system being input to the Sequential Decoder is therefore fixed to $m \times m$ in each case regardless of the value of N. Increasing the number of receive antennas results in 'better' channels being considered in terms of information at the receiver side. Due to this, the error curves fall with increasing N, and this can be seen from the plot for all decoding schemes being employed. It is important to note that this is unlike the case in Chapter 3 with overdetermined systems where error fell with increasing antennas, because in that case the system size was also increasing; the system size over here is, however, constant.

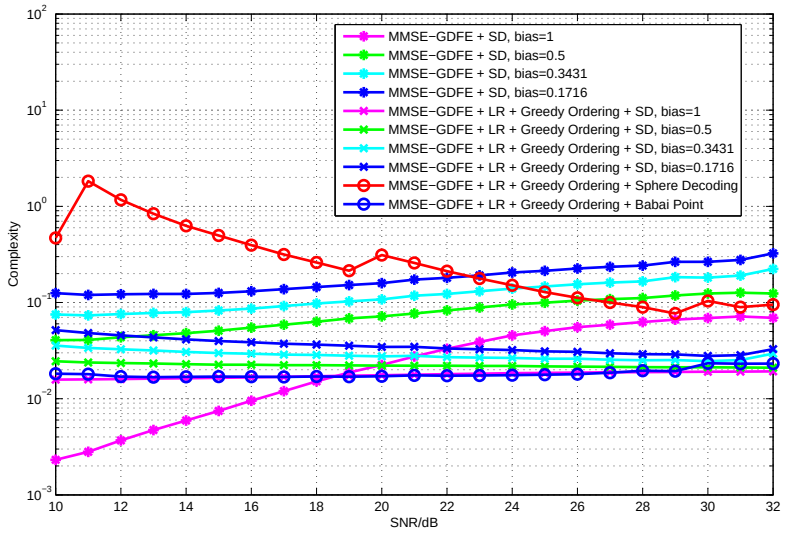
As before, the performance of the scheme in 5.2.2 is better than that in 5.2.1 and the gap between the two error curves increases as N is increased. The gap between the scheme in 5.2.1 and 5.2.2 decreases as the bias is decreased, as the performance

of both becomes better due to the improvement in the actual decoding process.

Figure 5.6b contains the corresponding complexity curves. The complexity curves for the scheme in 5.2.1 are seen to fall with increasing N, while the curves for the scheme in 5.2.2 fall initially and then become constant. Since the number of receive antennas N is changing, the complexity corresponding to each N belongs to a different system which maybe in a different region of the three phased complexity curve. The scheme in 5.2.1 for instance, with lower bias values at smaller values of N has falling error and high complexity, implying it is in the second phase; the bias of 1 however has high error and low complexity implying it is in the first phase. At higher values of N, the lower bias values move to the third phase with low complexity and low error. The bias value of 1 on the other hand, first goes through the second phase and then when N has increased enough, reaches the third stage. The scheme in 5.2.2 with lower bias values is in the second stage through the low values of N as the complexity is rather high and error starts falling. At higher values of N it reaches the third stage as the complexity curve falls very close to the Greedy Ordering complexity curve implying the complexity of the Sequential Decoding has reduced considerably and the complexity of the Greedy Ordering technique dominates the complexity of the process.

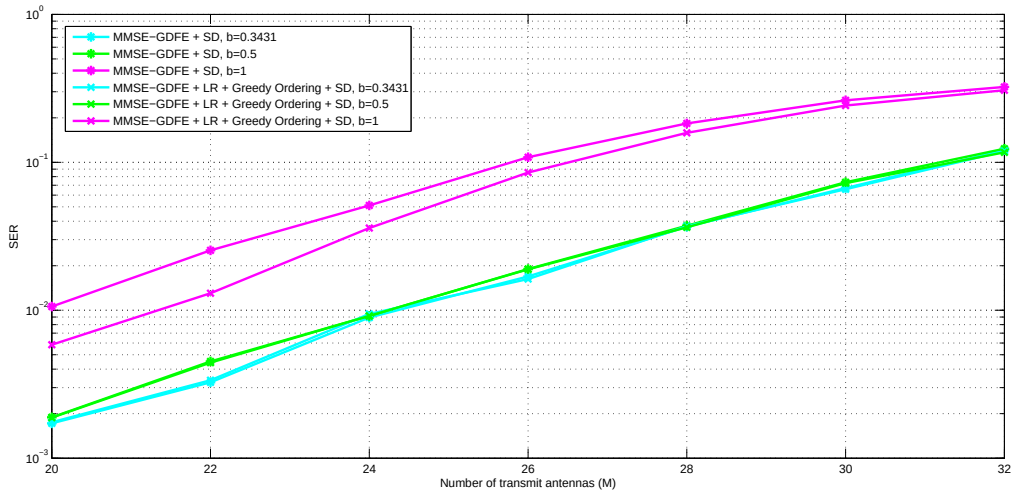


(a) Performance of different detectors with increasing SNR.

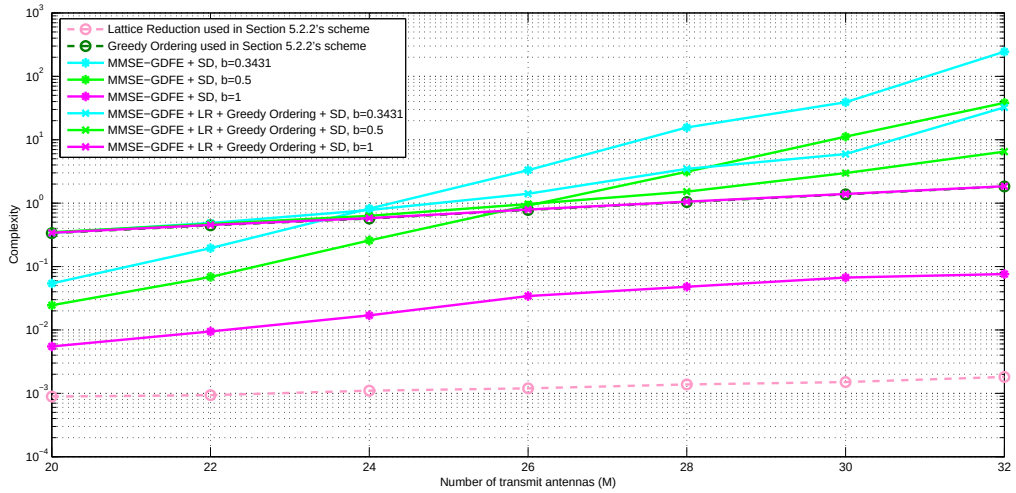


(b) Complexity of different detectors with increasing SNR.

Figure 5.4: Performance and complexity of different detectors for a 2×10 MIMO System employing 4-QAM.

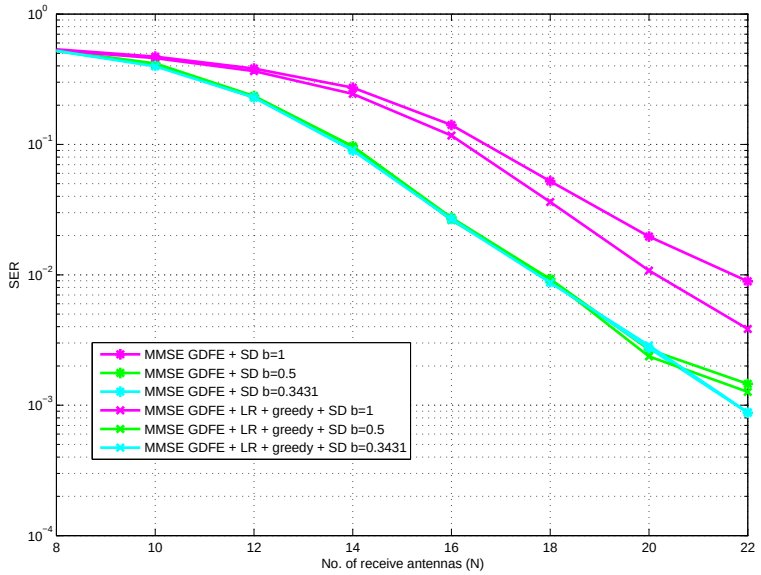


(a) Performance of different detectors with increasing transmit antennas.

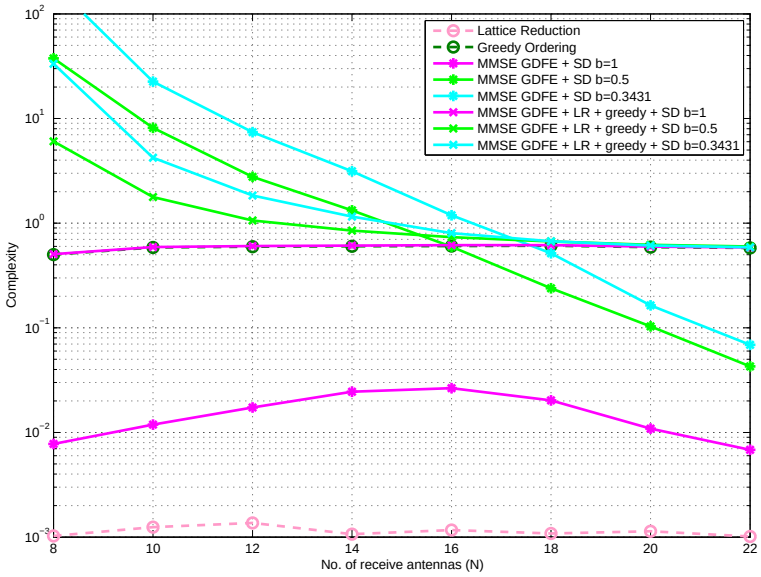


(b) Complexity of different detectors with increasing transmit antennas.

Figure 5.5: Performance and complexity of different detectors for MIMO Systems with 18 receive antennas employing 4-QAM and SNR=3dB.



(a) Performance of different detectors with increasing receive antennas.



(b) Complexity of different detectors with increasing receive antennas.

Figure 5.6: Performance and complexity of different detectors for MIMO Systems with 24 transmit antennas employing 4-QAM and SNR=3dB.

Chapter 6

Concluding Remarks

6.1 Summary

To meet the growing demand of high data rates, large MIMO systems need to be employed in transmission schemes. In our work we have presented the application of Sequential Decoding to large overdetermined and underdetermined MIMO systems. We have also analyzed the use of popular techniques, such as various lattice reduction aided linear decoding techniques for the decoding of large overdetermined systems, and different techniques employing the Sphere Decoder for the case of the underdetermined systems.

Since Sequential Decoders are able to trade performance for complexity gain and vice versa by altering the bias, we are able to achieve various performance-complexity trade-offs. The error probability of the Sequential Decoder can be bounded by bounding the bias, using the minimum distance of a lattice. The minimum distance in turn, can be bounded by the minimum eigenvalue of the matrix $\mathbf{H}^H\mathbf{H}$. Since for a fixed transmit-to-receive antenna ratio, the ratio of the minimum eigenvalue to the number of transmit antennas achieves a constant as the number of antennas grows; for large MIMO systems, we can bound the error by setting the bias to be less than the asymptote of $\lambda_{min}(\mathbf{H}^H\mathbf{H})/m$. This way we are able to achieve the error bound using the Sequential Decoder.

In the case of the overdetermined systems, Sequential Decoding is compared with ELR-aided linear decoders and CLLL-aided linear decoders. The results show that the Sequential Decoder outperforms LR-aided LDs in terms of performance and the gap between the error curves of the Sequential Decoder and LR-aided LDs is quite significant, particularly for large MIMO systems, even with large bias values. Interestingly, as the number of antennas is increased, the error of the Sequential Decoders with the low bias values chosen according to the minimum distance of the lattice criterion, is seen to fall. This is very impressive as none of the other decoding schemes show such behavior. Additionally, a particular trend in the performance of the Sequential Decoder is seen as SNR increases: a region of low error and low complexity at low SNR, a region of high complexity and falling error at moderate SNR, a region of low complexity (as low or lower than LR-aided LDs) and low error at higher SNR. The three phases occur at lower SNR for lower bias values, and the peaks in the second phase are narrower with decreasing bias. From this important observation it is extrapolated that there is room for improvement of the overall system performance, as the bias and SNR for a particular number of transmit and receive antennas can be chosen so that significant performance improvements are achieved in the same or even lower complexity range. In particular, choosing a low bias value would allow one to achieve the third phase and enjoy low error and low complexity in the low to moderate SNR range, even for large MIMO systems.

In the case of the underdetermined systems, we have presented two schemes: 1) MMSE-GDFE preprocessed Sequential Decoding, 2) MMSE-GDFE preprocessing, Lattice Reduction and Greedy Ordering followed by Sequential Decoding. The MMSE-GDFE preprocessing step is required as underdetermined MIMO systems are not full rank, and the preprocessing increases the rank to be equal to the number of unknowns. It should be noted, however, that MMSE-GDFE preprocessing introduces a suboptimality in the decoding process, except when the constellation being

employed has a constant modulus such as the 4-QAM cosntellation.

We have compared our work with the Sphere Decoder that employs MMSE-GDFE preprocessing, Lattice Reduction and Greedy Ordering in [28]. The second scheme proposed, results in some very good performance-complexity trade-offs. The first scheme also works fairly well, but the second is more robust. The scheme in [28] results in near-optimal performance, but the complexity associated is difficult to analyze as the Sphere Decoder requires choosing a radius that directly affects complexity. The radius must be large enough, so that the sphere is not empty, and small enough to avoid high complexity. In our simulations, through trial and error, small radii were chosen to achieve low complexity. In practical situations, choosing a radius like this may not be possible, and a larger radius may have to be used, dramatically increasing the complexity. Our schemes achieved error very close to that of the scheme in [28] when small enough bias values were used, and in the case of the first scheme, the complexity was comparable to that of the Sphere Decoding scheme's. In the case of the second scheme, the complexity was lower than that of the Sphere Decoding scheme's. It must be reiterated that in a practical scenario, [28] would have much higher complexity and so our schemes would have a more significant complexity advantage.

It should be mentioned that the superiority of the second scheme stems from its superior preprocessing. In fact when the system is good enough (the difference $m-n$ is not too large), the complexity of the second scheme is dominated by the complexity of the Greedy Ordering technique, as it tames the channel so that the decoding effort required is much less. This difference is also observed by noting that the complexity of lower bias values, that put in greater effort in Sequential Decoding, exceeds the Greedy Ordering complexity if the system becomes 'bad'; high bias values, however, are dominated by the complexity of Greedy Ordering for longer than the lower bias cases.

Our work with underdetermined systems also shows that with increasing systems size, unlike the overdetermined case, we are not able to achieve decreasing error at any bias. This can be explained by the fact that in the underdetermined case, we fixed the number of receivers (the equations) and increased the number of transmitters (the unknowns), which made the problem not only larger, but also more difficult to solve as the difference $m - n$ grew. When the number of receivers was increased, however, the error fell, but this was because the number of transmitters was fixed, and hence the system size was constant. The only change was that the difference between the number of unknowns and knowns was decreased by increasing N, hence the drop in error.

6.2 Future Research Work

Future work on the underdetermined case would involve quantifying the benefits of the additional preprocessing proposed in Section 5.2.2. Techniques other than Greedy Ordering, that have lower complexity, could be considered. Also, the effects of the lattice reduction being employed could be analyzed further, and other lattice reduction techniques that have lower complexity but reduce the lattice less, could be employed to see the effect on the error performance of the system.

REFERENCES

[1] X. Ma and W. Zhang, "Performance analysis for MIMO systems with lattice-reduction aided linear equalization," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 309–318, February 2008.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.

[3] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 1, Nov 2002, pp. 424–428 vol.1.

[4] Q. Zhou and X. Ma, "Element-Based Lattice Reduction Algorithms for Large MIMO Detection," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 274–286, February 2013.

[5] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. Information Theory Workshop, 2003*, March 2003, pp. 345–348.

[6] Y. H. Gan, C. Ling, and W.-H. Mow, "Complex Lattice Reduction Algorithm for Low-Complexity Full-Diversity MIMO Detection," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2701–2710, July 2009.

[7] M. Taherzadeh, A. Mobasher, and A. Khandani, "LLL Reduction Achieves the Receive Diversity in MIMO Decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4801–4805, Dec 2007.

[8] C.-E. Chen and W.-H. Sheen, "A New Lattice Reduction Algorithm for LR-Aided MIMO Linear Detection," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp.

2417–2422, August 2011.

[9] K. Singhal, T. Datta, and A. Chockalingam, "Lattice reduction aided detection in large-MIMO systems," in *Signal Processing Advances in Wireless Communications (SPAWC), 2013 IEEE 14th Workshop on*, June 2013, pp. 594–598.

[10] Q. Zhou and X. Ma, "Joint transceiver designs using lattice reduction algorithms," in *Signal and Information Processing (ChinaSIP), 2013 IEEE China Summit International Conference on*, July 2013, pp. 584–588.

[11] A. Lenstra, H. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Annalen*, vol. 261, no. 4, pp. 515–534, 1982.

[12] M. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49., no. 10., pp. 2389–2402, Oct. 2003.

[13] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul 1999.

[14] A. Murugan, H. El-Gamal, M.-O. Damen, and G. Caire, "A unified framework for tree search decoding: rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933–953, March 2006.

[15] W. Abediseid and M. Alouini, "On Lattice Sequential Decoding for the Unconstrained AWGN Channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2446–2456, June 2013.

[16] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, M.I.T., 1989.

[17] M.-O. Damen, K. Abed-Meraim, and J. C. Belfiore, "A generalized lattice decoder for asymmetrical space-time communication architecture," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, vol. 5, 2000, pp. 2581–2584 vol.5.

[18] M. Damen, M. O. Damen, J.-C. Belfiore, K. Abed-Meraim, and J. claude Bel Ore, "A Generalized Sphere Decoder for Asymmetrical Space-Time Communication," *Electron. Lett*, vol. 36, pp. 16–6, 2000.

[19] P. Dayal and M. K. Varanasi, "A fast generalized sphere decoder for optimum decoding of under-determined MIMO systems," in *in Proc. Allerton Conf. Communication, Control, and Computing*, 2003, pp. 1216–1225.

[20] X.-W. Chang and X. Yang, "A New Fast Generalized Sphere Decoding Algorithm for Under-Determined MIMO Systems," in *23rd Biennial Symposium on Communications*, 2006, pp. 18–21.

[21] Z. Yang, C. Liu, and J. He, "A new approach for fast generalized sphere decoding in MIMO systems," *Signal Processing Letters, IEEE*, vol. 12, no. 1, pp. 41–44, Jan 2005.

[22] T. Cui and C. Tellambura, "An efficient generalized sphere decoder for rank-deficient MIMO systems," *Communications Letters, IEEE*, vol. 9, no. 5, pp. 423–425, May 2005.

[23] X. Chang and X. Yang, "An Efficient Regularization Approach for Underdetermined MIMO System Decoding."

[24] X.-W. Chang and X. Yang, "An Efficient Tree Search Decoder with Column Reordering for Underdetermined MIMO Systems," in *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, Nov 2007, pp. 4375–4379.

[25] G. Romano, D. Ciuonzo, P. Salvo Rossi, and F. Palmieri, "Tree-search ML detection for underdetermined MIMO systems with M-PSK constellations," in *Wireless Communication Systems (ISWCS), 2010 7th International Symposium on*, Sept 2010, pp. 102–106.

[26] L. Bai, C. Chen, and J. Choi, "Prevoting Cancellation-Based Detection for Underdetermined MIMO Systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, pp. 1–11, 2010.

[27] ——, "Lattice Reduction Aided Detection for Underdetermined MIMO Systems: A Pre-Voting Cancellation Approach," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, May 2010, pp. 1–5.

[28] M. O. Damen, H. El Gamal, and G. Caire, "MMSE-GDFE lattice decoding for under-determined linear channels," in *CISS 2004, 38th annual Conference on Information Sciences and Systems, March 17-19, 2004, Princeton University, USA*, Princeton University, UNITED STATES, 05 2004. [Online]. Available: https://www.eurecom.fr/publication/1453

[29] M.-O. Damen, H. El-Gamal, and G. Caire, "MMSE-GDFE lattice decoding for solving under-determined linear systems with integer unknowns," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, June 2004, pp. 539–.

[30] S. Hwang and P. Schniter, "On the optimality of MMSE-GDFE pre-processed sphere decoding," in *in Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, 2005.

A Papers Submitted and Under Preparation

• Konpal Shaukat Ali, Walid Abediseid, and Mohamed-Slim Alouini, "Sequential Decoders for Large MIMO Systems", *In the 3rd International Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks, PHYSComNet 2014*, in conjunction with the 12th International IEEE Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOpt 2014.

• "Sequential Decoding for Large Asymmetric MIMO Systems" (under preparation).