Aalto University
School of Science
Degree Programme in Mathematics and Operations Research

Niko Väisänen

Beamforming Techniques for Optimizing Channel Capacity in Wireless Communications

Master's Thesis
Espoo, January 29, 2018

Supervisors: Assoc. Professor Camilla Hollanti
Advisor: D.Sc. (Tech.) Jussi Salmi

Aalto University
School of Science
Degree Programme in Mathematics and Operations Research

ABSTRACT OF
MASTER'S THESIS

Author: Niko Väisänen

Title:
Beamforming Techniques for Optimizing Channel Capacity in Wireless Communications

Date: January 29, 2018 **Pages:** vii + 96

Major: Mathematics **Code:** SCI3054

Supervisors: Assoc. Professor Camilla Hollanti

Advisor: D.Sc. (Tech.) Jussi Salmi

Many modern devices rely on wireless Internet connection. This has increased the demand for high performance wireless networks. However, the traditional single-transmitting single-receiving antenna systems cannot provide the desired data rates and latencies in the future.

The capacity of wireless communication systems can be increased by using more than one transmitting and receiving antennas. These multiple-input multiple-output (MIMO) systems allow the use of advanced signal precoding methods including beamforming. This method enables transmission of several concurrent data streams by controlling the electromagnetic wave interference.

In this thesis, we review various beamforming techniques for single user and multi-user MIMO systems. In the single user case, the well-known eigen-beamforming method is derived, which achieves the optimal channel capacity. In the multi-user MIMO context, we focus on zero-forcing techniques that are based on a zero inter-user interference constraint. Finally, we define a new geometry-based MIMO channel model that is used for beamforming simulations. Results indicate that using beamforming can yield substantial increase in performance compared to the traditional communication systems.

Keywords: beamforming, block diagonalization, channel capacity, channel inversion, channel model, eigen-beamforming, MIMO, multi-user MIMO, Shannon's theorem, single user MIMO, water-filling principle, zero-forcing beamforming

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Matematiikan ja systeemianalyysin koulutusohjelma

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Niko Väisänen

Työn nimi:
Langattoman viestinnän kanavakapasiteetin optimointi keilanmuodostustekniikalla

Päiväys:	29. tammikuuta 2018	**Sivumäärä:**	vii + 96
Pääaine:	Matematiikka	**Koodi:**	SCI3054
Valvojat:	Professori Camilla Hollanti		
Ohjaaja:	Tekniikan tohtori Jussi Salmi		

Monet nykyajan laitteet vaativat toimiakseen langattoman internet-yhteyden. Lisääntyvä älylaitteiden määrä on kasvattanut tarvetta korkean suorituskyvyn langattomille tietoverkoille. Perinteiset yhden lähetys- ja vastaanottoantennin viestintäsysteemit eivät kuitenkaan pysty tulevaisuudessa tuottamaan tarvittavia datansiirtonopeuksia ja viiveaikoja.

Langattomassa viestinnässä kapasiteettia voidaan kasvattaa käyttämällä useita lähetys- ja vastaanottoantenneja. Nämä niin kutsutut MIMO-systeemit mahdollistavat edistyneiden koodaustekniikoiden käyttämisen. Yksi näistä tekniikoista on keilanmuodostus, jonka avulla voidaan lähettää useita datapaketteja samanaikaisesti. Menetelmä perustuu sähkömagneettisten aaltojen superpositioperiaatteen hyödyntämiseen.

Tässä diplomityössä tarkastellaan useita keilanmuodostustekniikoita sekä yhden että monen käyttäjän MIMO-systeemeissä. Yhden käyttäjän tapauksessa johdetaan tunnettu menetelmä, joka saavuttaa optimaalisen kanavakapasiteetin. Usean käyttäjän MIMO-systeemien yhteydessä taas keskitytään tekniikoihin, joissa käyttäjien väliset häiriöt minimoidaan. Lopuksi määritellään uusi geometriaperusteinen kanavamalli, jota käytetään keilanmuodostussimulaatioissa. Tulokset osoittavat, että keilanmuodostusta käyttämällä voidaan parantaa merkittävästi langattomien viestintäsysteemien suorituskykyä.

Asiasanat: blokkidiagonalisointi-menetelmä, kanavakapasiteetti, kanavankääntömenetelmä, kanavamalli, keilanmuodostus, MIMO, monen käyttäjän MIMO, nollahäiriökeilanmuodostus, ominaisvektorikeilanmuodostus, Shannonin teoreema, vesitäyttömenetelmä

Kieli: Englanti

Acknowledgements

I would like to express my gratitude to my supervisor, Professor Camilla Hollanti, for her support over all these years in Aalto. Your inspirational attitude towards algebra and mathematics in general has had a major impact on my selected study path to which I am very pleased about.

I am also very thankful to my advisor, D.Sc. Jussi Salmi, for all the simulation ideas and fruitful conversations that greatly furthered my understanding of radio technology and mobile networks beyond the theoretic scope I was mainly focused on in this thesis.

Thanks also belongs to my squad group leader Janne Aho, former line manager Jyri Suvanen and my tribe leader Matti Lehtimäki for firstly giving this great opportunity to work at Nokia Networks, but also for allowing me to concentrate on this thesis and my studies at work hours. Special thanks also to all the team members for being such great colleagues—it is a pleasure to work with you.

Finally, I want to thank my family—Ria, Minttu, and Kuisma—for making me smile always when I come home. Working hard and thinking math all day long is sometimes tough, but your support has been a huge source of motivation for me during this work.

Thank you!

Espoo, January 29, 2018

Niko Väisänen

List of Acronyms

5G	5th generation (mobile networks)
AWGN	additive white Gaussian noise
CDF	cumulative density function
CSCG	circularly symmetric complex Gaussian
CSI	channel state information
LOS	line-of-sight
LTE	Long-Term Evolution (mobile networks)
MIMO	multiple-input multiple-output
MU-MIMO	multi-user MIMO
MSE	mean squared error
PAM	pulse-amplitude modulation
PDF	probability density function
PMF	probability mass function
SINR	signal-to-interference-plus-noise ratio
SISO	single-input single-output
SNR	signal-to-noise ratio
SU-MIMO	single-user MIMO
SVD	singular value decomposition

Contents

List of Acronyms **v**

1 Introduction **1**

2 Linear Algebra Preliminaries **3**
2.1 Basic Definitions 3
2.2 Matrix Decompositions 5
2.3 Other Important Results 10

3 Channel Capacity **12**
3.1 Capacity via Sphere Packing 12
3.2 Shannon's Theorem 16

4 Single-User MIMO **23**
4.1 Single-User Channel Capacity 24
4.2 Eigen-Beamforming 26
4.2.1 Single Spatial Stream 27
4.2.2 Multiple Spatial Streams 30
4.2.3 Practical View 33
4.3 Water-Filling Principle 35

5 Multi-User MIMO **39**
5.1 Multi-User Channel Capacity 40
5.2 Eigen-Beamforming Approach 41
5.3 Zero-Forcing Beamforming 42
5.3.1 Channel Inversion 44
5.3.2 Regularized Channel Inversion 45
5.3.3 Vector Perturbation 47
5.3.4 Block Diagonalization 51
5.3.4.1 Maximum Throughput 51
5.3.4.2 Power Control 55

5.3.4.3 Partial Channel Knowledge 56
5.4 Advanced Beamforming Solutions 58

6 Beamforming Simulations **61**
6.1 Channel Model 61
6.1.1 Geometry-Based SISO Channel Model 62
6.1.2 MIMO Extension 63
6.2 Eigen-Beamforming Results 65
6.2.1 Benefit of Large Antenna Array 66
6.2.2 Eigen-Beamforming Visualization and Channel Aging . 69
6.3 Zero-Forcing Beamforming Results 71
6.3.1 Channel Inversion Visualization and Channel Aging . . 71
6.3.2 Effect of Extra Degrees of Freedom 75

7 Conclusion **77**

A Additional Simulation Results **85**
A.1 Antenna Array Radiation Pattern 85
A.2 Zero-Forcing Beamforming 87

B Additional Proofs **91**
B.1 Proof of Equation (3.3) 91
B.2 Proof of Equation (4.3) 91
B.3 Proof of Lemma 4.4 93
B.4 Proof of Proposition 5.2 94

Chapter 1

Introduction

Wireless communication is in a major role in our everyday life. Modern technologies such as mobile devices, smart televisions, and even cars all rely on Internet connections that are usually established wirelessly. The demand for high performance networks is ever increasing due to the growth in the number of endpoint devices.

In a traditional wireless communication system there exists one transmitter and one receiver that both have only one available antenna for the communication. This system is called *single-input single-output* (SISO). Shannon-Hartley Theorem[1] tells that the data rate in a SISO system is limited by three factors: the bandwidth of the system, the signal power at the receiver (or the signal level), and the noise power.

Assuming that the noise power has been made as small as possible, then to increase system capacity either the bandwidth or the received power needs to be increased. However, for example in mobile networks, the network providers have only fixed amount of bandwidth available and they cannot increase it on their own. Therefore, the only option for improving the system capacity is to increase the signal level. The easiest way to achieve this is just to add more input power at the transmitter. However, this increases the power consumption and creates interference to neighboring cells, so this option is not desirable. Therefore, increasing capacity in SISO systems is not possible in practice.

To overcome this issue, the transmitter can employ multiple transmitting antennas instead of just one. This allows the transmitter to send multiple simultaneous data streams, thus enabling higher data rates. However, in order to efficiently utilize this, there must be multiple receiving antennas. This creates a system that is called *multiple-input multiple-output* (MIMO).

[1]Discussion on this theorem can be found in Chapters 3 and 4.

The requirement for multiple receiving antennas can be achieved in two ways. The first option is that the receiver has multiple receiving antennas. Since there is only one receiver called a user, this system is referred to as *single user* MIMO (SU-MIMO). The second option is that there exists multiple users that all have one or more receiving antennas. This option is called *multi-user* MIMO (MU-MIMO).

Using a MIMO system does not immediately enhance performance. This is due to the superposition principle of electromagnetic waves that tells that the received signal is the sum of the signals received from different sources. Therefore, using a MIMO system improperly may lead to destructive interference at the receiver which decreases the system performance. This issue is solved by beamforming.

Beamforming is a precoding technique that can be used in MIMO wireless communication systems to increase system performance by allowing multiple simultaneous data transmissions. Beamforming controls the relative phases of the transmitted or received signals to create constructive and destructive interference at the desired directions. In the single user transmit beamforming, this means that the transmitted signal is controlled so that a constructive interference is obtained at the intended receiver. On the other hand, in the multi-user case the objective is to create destructive interference at known undesired receivers.

In this thesis, we review beamforming techniques that can be used to increase system performance in wireless MIMO communications. The thesis is divided into three parts. In the first part, the preliminary theory is given. This includes various common results in linear algebra that are presented and proved in Chapter 2. Then, in Chapter 3 the channel capacity of a system is defined, and the description for Shannon's Theorem is given. In the second part, different beamforming methods are discussed. First, SU-MIMO beamforming is considered in Chapter 4. In this chapter we will find that a closed-form solution for the maximum capacity can be obtained leading to the optimal beamforming method for single user cases. Then, in Chapter 5 MU-MIMO beamforming is studied. Since the optimal solutions have issues in practical applications, we concentrate on reviewing sub-optimal methods obtained via beamforming. Finally, in the third part, we define a new geometry-based MIMO channel model that is used for beamforming simulations. Description of this model and the obtained results are presented in Chapter 6.

Chapter 2

Linear Algebra Preliminaries

In this chapter, we derive all the necessary linear algebra tools that are needed for channel capacity and beamforming discussions. Most of these results are well-known, but they are listed here for completeness. All the proofs presented here are writer's own contribution, however, similar proofs can be found from many linear algebra text books, e.g. [1], [2].

This chapter is divided into three sections. First, in Section 2.1, the basic definitions and notations that are used throughout the thesis are presented. Then, in Section 2.2, we describe matrix decompositions that will be used in beamforming method derivations. Key results in this section are the spectral theorem, the eigenvalue decomposition, and the singular value decomposition (SVD). Finally, in Section 2.3, we present other important linear algebra results that are needed later on in the thesis.

2.1 Basic Definitions

Definition 2.1. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a matrix, $\mathbf{v} \in \mathbb{C}^{n\times 1}$ be a non-zero vector, and $\lambda \in \mathbb{C}$ be a scalar satisfying $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Then, we call $\mathbf{v}$ an *eigenvector* and λ the corresponding *eigenvalue* of $\mathbf{A}$. If $\|\mathbf{v}\| = 1$, then we call $(\lambda, \mathbf{v})$ an *eigenpair* of $\mathbf{A}$.

Definition 2.2. A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is *normal* if $\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger$ where $\mathbf{A}^\dagger$ is the Hermitian transpose of $\mathbf{A}$. If $\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger = \mathbf{I}$, then $\mathbf{A}$ is said to be *unitary*. If $\mathbf{A} = \mathbf{A}^\dagger$, then $\mathbf{A}$ is *Hermitian*.

Proposition 2.1. *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Then, every eigenvalue of $\mathbf{A}$ is real. Moreover, $\mathbf{x}^\dagger\mathbf{A}\mathbf{x} \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{C}^{n\times 1}$.*

Proof. Let $\mathbf{x} \in \mathbb{C}^{n\times 1}$ and suppose $\mathbf{x}^\dagger\mathbf{A}\mathbf{x} = a \in \mathbb{C}$. Then,

$$a = \mathbf{x}^\dagger\mathbf{A}\mathbf{x} = \mathbf{x}^\dagger\mathbf{A}^\dagger\mathbf{x} = \left(\mathbf{x}^\dagger\mathbf{A}\mathbf{x}\right)^\dagger = a^*$$

where $(\cdot)^*$ denotes the complex conjugate. Since $a = a^*$, we may conclude that $a \in \mathbb{R}$. If we let $\mathbf{x}$ to be a unit eigenvector of $\mathbf{A}$, then a becomes an eigenvalue of $\mathbf{A}$. Therefore, all the eigenvalues are real. □

Definition 2.3. A Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is *positive definite* if for any non-zero vector $\mathbf{x} \in \mathbb{C}^{n\times 1}$ it holds that $\mathbf{x}^\dagger \mathbf{A}\mathbf{x} > 0$. If instead $\mathbf{x}^\dagger \mathbf{A}\mathbf{x} \geq 0$, then $\mathbf{A}$ is *positive semi-definite*. Analogously can be defined *negative definite* and *negative semi-definite* matrices.

Proposition 2.2. *Let $\mathbf{A} \in \mathbb{C}^{m\times n}$. Then, the matrices $\mathbf{A}^\dagger\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\dagger$ are positive semi-definite.*

Proof. Let $\mathbf{x} \in \mathbb{C}^{n\times 1}$. A direct computation gives

$$\mathbf{x}^\dagger \mathbf{A}^\dagger \mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^\dagger (\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \geq 0\,.$$

Now, this result can be applied to $\mathbf{A}^\dagger$ and the statement follows. □

Definition 2.4. Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{B} \in \mathbb{C}^{M\times N}$. Then the *direct sum* of matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{0}_{m\times N} \\ \mathbf{0}_{M\times n} & \mathbf{B} \end{bmatrix},$$

i.e., the $(m + M) \times (n + N)$ block diagonal matrix defined by $\mathbf{A}$ and $\mathbf{B}$. Here $\mathbf{0}_{\ell\times r}$ denotes the $\ell \times r$ zero matrix. In general, let $\mathbf{A}_i \in \mathbb{C}^{m_i\times n_i}$ for $i = 1, 2, \ldots, k$. We define the direct sum of matrices $\mathbf{A}_i$ to be

$$\bigoplus_{i=1}^{k} \mathbf{A}_i = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0}_{m_1\times n_2} & \cdots & \mathbf{0}_{m_1\times n_k} \\ \mathbf{0}_{m_2\times n_1} & \mathbf{A}_2 & \cdots & \mathbf{0}_{m_2\times n_k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{m_k\times n_1} & \mathbf{0}_{m_k\times n_2} & \cdots & \mathbf{A}_k \end{bmatrix}.$$

Notice that the direct sum of matrices is not commutative.

Definition 2.5. Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ be a matrix. The *null space* (or *kernel*) of $\mathbf{A}$ is the linear subspace $\ker(\mathbf{A})$ of vectors $\mathbf{x} \in \mathbb{C}^{n\times 1}$ satisfying

$$\mathbf{A}\mathbf{x} = \mathbf{0}\,.$$

The *nullity* of $\mathbf{A}$, denoted as $\mathrm{nul}(\mathbf{A})$, is the dimension of the null space as a vector space.

2.2 Matrix Decompositions

Theorem 2.3. (Schur decomposition) *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *be any* $n \times n$ *complex matrix. Then, there exists a decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^\dagger$$

where $\mathbf{U}$ *is a unitary matrix and* $\mathbf{T}$ *is an upper triangular matrix.*

Proof. We use induction on the size n of $\mathbf{A}$. If $n = 1$, then the statement holds trivially. Suppose that the statement holds for all $(n-1) \times (n-1)$ matrices. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$, and let $(\lambda, \mathbf{v})$ be an eigenpair of $\mathbf{A}$ (which existence is guaranteed by the fundamental theorem of algebra[1]). Further, let $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ be such that $\{\mathbf{v}, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}\}$ forms an orthogonal basis. Let $\mathbf{U} = \begin{bmatrix}\mathbf{v} & \mathbf{u}_1 & \ldots & \mathbf{u}_{n-1}\end{bmatrix}^\dagger$. Then, we may notice that $(\mathbf{U}\mathbf{A}\mathbf{U}^\dagger)_{11} = \lambda$ and $(\mathbf{U}\mathbf{A}\mathbf{U}^\dagger)_{i1} = 0$ for all $i > 1$. Therefore,

$$\mathbf{U}\mathbf{A}\mathbf{U}^\dagger = \begin{bmatrix}\lambda & \mathbf{a}^\dagger \\ \mathbf{0} & \mathbf{A}'\end{bmatrix}$$

for some vector $\mathbf{a} \in \mathbb{C}^{(n-1)\times 1}$ and a matrix $\mathbf{A}' \in \mathbb{C}^{(n-1)\times(n-1)}$. By the induction hypothesis, $\mathbf{A}'$ has a decomposition $\mathbf{A}' = \mathbf{V}\mathbf{T}\mathbf{V}^\dagger$ where $\mathbf{V} \in \mathbb{C}^{(n-1)\times(n-1)}$ is unitary and $\mathbf{T} \in \mathbb{C}^{(n-1)\times(n-1)}$ is upper triangular. Let $\mathbf{W} = \big(1 \oplus \mathbf{V}^\dagger\big)\mathbf{U}$ which is unitary by construction. Now, we have

$$\begin{aligned}\mathbf{W}\mathbf{A}\mathbf{W}^\dagger &= \begin{bmatrix}1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^\dagger\end{bmatrix}\begin{bmatrix}\lambda & \mathbf{a}^\dagger \\ \mathbf{0} & \mathbf{A}'\end{bmatrix}\begin{bmatrix}1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}\end{bmatrix} \\ &= \begin{bmatrix}\lambda & \mathbf{a}^\dagger\mathbf{V} \\ \mathbf{0} & \mathbf{V}^\dagger\mathbf{A}'\mathbf{V}\end{bmatrix} = \begin{bmatrix}\lambda & \mathbf{a}^\dagger\mathbf{V} \\ \mathbf{0} & \mathbf{T}\end{bmatrix}\end{aligned}$$

which is an upper triangular matrix. Therefore, by induction the statement holds. □

Definition 2.6. A matrix $\mathbf{A}$ is *similar* to a matrix $\mathbf{B}$ if there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$. $\mathbf{A}$ is called *diagonalizable* if $\mathbf{B}$ is diagonal. Further, if the matrix $\mathbf{P}$ is a unitary matrix, then $\mathbf{A}$ is called *unitarily diagonalizable.*

Proposition 2.4. *Suppose* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *is similar to a diagonal matrix* $\mathbf{\Gamma}$. *Then,* $\mathbf{\Gamma}$ *is uniquely determined up to permutation.*

[1]Fundamental theorem of algebra [3]: Every non-constant polynomial with complex coefficients has at least one root.

Proof. Suppose $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Gamma} = \bigoplus_i \gamma_i$ and $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{\Gamma}' = \bigoplus_i \gamma'_i$. This implies $\mathbf{P}\mathbf{\Gamma}\mathbf{P}^{-1} = \mathbf{Q}\mathbf{\Gamma}'\mathbf{Q}^{-1}$. Let $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{P}$. Then, we have

$$\mathbf{B}\mathbf{\Gamma} = \mathbf{\Gamma}'\mathbf{B}\,.$$

for some invertible matrix $\mathbf{B} = (b_{ij})$. This is true if and only if for all pairs (i, j) holds that either $b_{ij} = 0$ or $\gamma_j = \gamma'_i$. Let $a \in \mathbb{C}$, and let $\mathcal{D}_a$ and $\mathcal{D}'_a$ be the maximal sets such that $\gamma_j = a$ for all $j \in \mathcal{D}_a$ and $\gamma'_i = a$ for all $i \in \mathcal{D}'_a$. To show that $\mathbf{\Gamma}$ and $\mathbf{\Gamma}'$ are equal up to permutation, it suffices to show that $|\mathcal{D}_a| = |\mathcal{D}'_a|$ for all $a \in \mathbb{C}$.

Consider a submatrix $\mathbf{B}'$ of $\mathbf{B}$ defined by the index sets $\mathcal{D}_a$ for the columns and $\mathcal{D}'_a$ for the rows. Since $\mathcal{D}_a$ and $\mathcal{D}'_a$ are maximal, we must have $b_{ij} = 0$ for all $i \in \mathcal{D}'_a$, $j \notin \mathcal{D}_a$ and for all $j \in \mathcal{D}_a$, $i \notin \mathcal{D}'_a$. Let $\mathbf{P}$ and $\mathbf{Q}$ be such permutation matrices that $\mathbf{P}\mathbf{B}\mathbf{Q}$ has $\mathbf{B}'$ in its upper left corner. Since $\mathbf{B}$ is invertible and $\mathbf{P}, \mathbf{Q}$ are unitary as permutation matrices, we have

$$\mathbf{I} = (\mathbf{P}\mathbf{B}\mathbf{Q})(\mathbf{P}\mathbf{B}\mathbf{Q})^{-1} = \begin{bmatrix} \mathbf{B}' & \mathbf{0} \\ \mathbf{0} & \mathbf{B}'' \end{bmatrix} \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{B}'\mathbf{C} & \mathbf{B}'\mathbf{D} \\ \mathbf{B}''\mathbf{E} & \mathbf{B}''\mathbf{F} \end{bmatrix}$$

for some matrices $\mathbf{B}'', \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}$. This equation is true only if $\mathbf{B}'\mathbf{C} = \mathbf{I}$. Similarly can be showed that there exists $\mathbf{C}'$ such that $\mathbf{C}'\mathbf{B}' = \mathbf{I}$. Therefore, $\mathbf{B}'$ has both left and right inverses, so it must be a square matrix. Since the number of columns and rows were defined by $\mathcal{D}_a$ and $\mathcal{D}'_a$ respectively, we have $|\mathcal{D}_a| = |\mathcal{D}'_a|$, which implies the result. □

Lemma 2.5. *An upper (lower) triangular matrix is normal if and only if it is diagonal.*

Proof. If a matrix is diagonal, then it clearly is normal. Now, let $\mathbf{T} = (t_{ij}) \in \mathbb{C}^{n\times n}$ be a normal upper triangular matrix. We use induction on n. If $n = 1$, the statement holds trivially. Now, suppose that the statement holds for all values less than n. Since $\mathbf{T}$ is upper triangular, we may write it in the form

$$\mathbf{T} = \begin{bmatrix} t_{11} & \mathbf{t} \\ \mathbf{0} & \mathbf{T}' \end{bmatrix}$$

for some upper triangular matrix $\mathbf{T}' \in \mathbb{C}^{(n-1)\times(n-1)}$ and a vector $\mathbf{t} \in \mathbb{C}^{(n-1)\times 1}$. Now,

$$\mathbf{T}\mathbf{T}^\dagger = \begin{bmatrix} t_{11} & \mathbf{t}^\dagger \\ \mathbf{0} & \mathbf{T}' \end{bmatrix} \begin{bmatrix} t_{11}^* & \mathbf{0} \\ \mathbf{t} & (\mathbf{T}')^\dagger \end{bmatrix} = \begin{bmatrix} |t_{11}|^2 + \|\mathbf{t}\|^2 & (\mathbf{T}'\mathbf{t})^\dagger \\ \mathbf{T}'\mathbf{t} & \mathbf{T}'(\mathbf{T}')^\dagger \end{bmatrix}$$
$$\mathbf{T}^\dagger\mathbf{T} = \begin{bmatrix} t_{11}^* & \mathbf{0} \\ \mathbf{t} & (\mathbf{T}')^\dagger \end{bmatrix} \begin{bmatrix} t_{11} & \mathbf{t}^\dagger \\ \mathbf{0} & \mathbf{T}' \end{bmatrix} = \begin{bmatrix} |t_{11}|^2 & t_{11}^*\mathbf{t}^\dagger \\ t_{11}\mathbf{t} & \mathbf{t}\mathbf{t}^\dagger + (\mathbf{T}')^\dagger\mathbf{T}' \end{bmatrix}.$$

Since $\mathbf{T}$ is normal, $|t_{11}|^2 + \|\mathbf{t}\|^2 = |t_{11}|^2$ which implies $\mathbf{t} = \mathbf{0}$. This, on the other hand, means $\mathbf{T}'(\mathbf{T}')^\dagger = (\mathbf{T}')^\dagger\mathbf{T}'$, which by the induction hypothesis implies that $\mathbf{T}'$ is a diagonal matrix. Therefore, also $\mathbf{T}$ is a diagonal matrix, and the statement holds by induction. Since $\mathbf{T}^\dagger$ is lower triangular, the lower case is also proven.

□

Theorem 2.6. (Spectral Theorem for normal matrices) *A matrix is unitarily diagonalizable if and only if it is a normal matrix.*

Proof. Suppose $\mathbf{A}$ is unitarily diagonalizable as $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ for some unitary matrix $\mathbf{U}$ and diagonal matrix $\mathbf{\Lambda}$. Since diagonal matrices commute,

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^\dagger\mathbf{U}^\dagger = \mathbf{U}\mathbf{\Lambda}^\dagger\mathbf{\Lambda}\mathbf{U}^\dagger = \mathbf{A}^\dagger\mathbf{A}\,,$$

so $\mathbf{A}$ is normal.

To prove the opposite direction, let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a normal matrix. Then, by the Schur decomposition, $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^\dagger$ for a unitary matrix $\mathbf{U}$ and an upper diagonal matrix $\mathbf{T}$. Since $\mathbf{A}$ is normal, we have

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger\mathbf{A} \iff \mathbf{U}\mathbf{T}\mathbf{T}^\dagger\mathbf{U}^\dagger = \mathbf{U}\mathbf{T}^\dagger\mathbf{T}\mathbf{U}^\dagger \iff \mathbf{T}\mathbf{T}^\dagger = \mathbf{T}^\dagger\mathbf{T}$$

which by Lemma 2.5 implies $\mathbf{T}$ is a diagonal matrix. Therefore, $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^\dagger$ is a unitary diagonalization of $\mathbf{A}$. □

Corollary 2.7. *Eigenvectors of a normal matrix form an orthogonal basis.*

Proof. Let $\mathbf{A}$ be a normal matrix. Then, by the spectral theorem, $\mathbf{A}$ is unitarily diagonalizable, i.e. $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ for some unitary matrix $\mathbf{U}$. This implies $\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$ which means by the definition of eigenpairs that each column of $\mathbf{U}$ and the corresponding diagonal value of $\mathbf{\Lambda}$ form an eigenpair. Since $\mathbf{U}$ is unitary, it follows that the eigenvectors of $\mathbf{A}$ form an orthogonal basis for the entire space. □

Proposition 2.8. *A matrix is positive definite if and only if its eigenvalues are all strictly positive. Similar results hold for positive semi-definite, negative definite and negative semi-definite matrices.*

Proof. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and let $\mathbf{x} \in \mathbb{C}^{n\times 1}$ be any non-zero vector. Since every Hermitian matrix is normal, by Corollary 2.7, the eigenvectors of $\mathbf{A}$ form a basis for the entire space, so $\mathbf{x}$ can be written as a linear combination of eigenvectors of $\mathbf{A}$. Let $\mathbf{x} = \sum_{i\in\mathcal{D}} c_i\mathbf{v}_i$ where

$c_i \in \mathbb{C}\backslash\{0\}$, $\mathbf{v}_i$ is an eigenvector of $\mathbf{A}$, and $\mathcal{D}$ is some non-empty index set $\mathcal{D} \subseteq \{1, \ldots, n\}$. Since the eigenvectors of $\mathbf{A}$ are orthogonal, we have

$$\begin{aligned}
\mathbf{x}^\dagger \mathbf{A} \mathbf{x} &= \left(\sum_{i \in \mathcal{D}} c_i^* \mathbf{v}_i^\dagger \right) \mathbf{A} \left(\sum_{i \in \mathcal{D}} c_i \mathbf{v}_i \right) \\
&= \sum_{i \in \mathcal{D},\ j \in \mathcal{D}} c_i^* c_j \mathbf{v}_i^\dagger \mathbf{A} \mathbf{v}_j \\
&= \sum_{i \in \mathcal{D},\ j \in \mathcal{D}} c_i^* c_j \lambda_j \mathbf{v}_i^\dagger \mathbf{v}_j \\
&= \sum_{j \in \mathcal{D}} |c_j|^2 \lambda_j
\end{aligned}$$

where λ_j is the eigenvalue corresponding to the eigenvector $\mathbf{v}_j$. Since $|c_i|^2 > 0$, we can see that $\mathbf{x}^\dagger \mathbf{A} \mathbf{x}$ is positive (non-negative, negative, non-positive) for all $\mathbf{x}$ if and only if every eigenvalue of $\mathbf{A}$ is positive (non-negative, negative, non-positive). □

Theorem 2.9. (Eigenvalue decomposition) *Suppose $\mathbf{A} \in \mathbb{C}^{n \times n}$ has n linearly independent eigenvectors. Then, there exists a decomposition*

$$\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1}$$

where the columns of $\mathbf{Q}$ are the eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda}$ is a diagonal matrix which diagonal entries are the eigenvalues of $\mathbf{A}$.

Proof. Let $\mathbf{Q} \in \mathbb{C}^{n \times n}$ be a matrix which columns are linearly independent eigenvectors of $\mathbf{A}$. Let λ_i be the eigenvalue corresponding to the ith column of $\mathbf{Q}$, and let $\mathbf{\Lambda} = \bigoplus_{i=1}^{n} \lambda_i$. Then, by the definition of eigenpairs

$$\mathbf{A} \mathbf{Q} = \mathbf{Q} \mathbf{\Lambda} \,.$$

Since the columns of $\mathbf{Q}$ are linearly independent, it is an invertible matrix. Therefore, the previous equation implies

$$\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1}$$

which is the desired result. □

Corollary 2.10. *A matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is similar to a diagonal matrix if and only if $\mathbf{A}$ has n linearly independent eigenvectors.*

Proof. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. If $\mathbf{A}$ has n linearly independent eigenvectors, then it has an eigenvalue decomposition, which shows it is similar to a diagonal matrix. On the other hand, if $\mathbf{A}$ is similar to a diagonal matrix, then $\mathbf{A} = \mathbf{P}^{-1}\mathbf{\Lambda}\mathbf{P}$ for some invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{\Lambda}$. This equation implies $\mathbf{A}\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{\Lambda}$, so by the definition of eigenpairs, the columns of $\mathbf{P}^{-1}$ are eigenvectors of $\mathbf{A}$ and the corresponding diagonal elements of $\mathbf{\Lambda}$ are the eigenvalues of $\mathbf{A}$. Moreover, since $\mathbf{P}$ is invertible, also $\mathbf{P}^{-1}$ is, which implies the independence of columns, i.e., the eigenvectors. □

Theorem 2.11. (Singular value decomposition) *Let $\mathbf{A} \in \mathbb{C}^{m\times n}$. Then, there exists a decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$$

where $\mathbf{U} \in \mathbb{C}^{m\times m}$ and $\mathbf{V} \in \mathbb{C}^{n\times n}$ are unitary matrices, and $\mathbf{\Sigma} \in \mathbb{R}_+^{m\times n}$ is a diagonal matrix with non-negative real entries. Moreover, if the diagonal entries of $\mathbf{\Sigma}$ are arranged in a decreasing order, then $\mathbf{\Sigma}$ is uniquely determined. This matrix decomposition is called the singular value decomposition (SVD). Columns of $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors of $\mathbf{A}$ respectively, and the diagonal elements of $\mathbf{\Sigma}$ are the singular values of $\mathbf{A}$.

Proof. Consider the Hermitian matrix $\mathbf{A}^\dagger\mathbf{A}$. Let $(\lambda, \mathbf{v})$ be an eigenpair of $\mathbf{A}^\dagger\mathbf{A}$ such that $\lambda \neq 0$ (if no such exists, then the statement becomes trivial). By Proposition 2.2, $\mathbf{A}^\dagger\mathbf{A}$ is positive semi-definite, so by Proposition 2.8, $\lambda > 0$. Notice that

$$\|\mathbf{A}\mathbf{v}\| = \sqrt{(\mathbf{A}\mathbf{v})^\dagger(\mathbf{A}\mathbf{v})} = \sqrt{\mathbf{v}^\dagger(\mathbf{A}^\dagger\mathbf{A}\mathbf{v})} = \sqrt{\lambda\mathbf{v}^\dagger\mathbf{v}} = \lambda^{1/2}\,.$$

Define a unit vector

$$\mathbf{u} = \frac{\mathbf{A}\mathbf{v}}{\lambda^{1/2}}\,.$$

Now,

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{u} = \mathbf{A}\frac{\mathbf{A}^\dagger\mathbf{A}\mathbf{v}}{\lambda^{1/2}} = \lambda\frac{\mathbf{A}\mathbf{v}}{\lambda^{1/2}} = \lambda\mathbf{u}\,,$$

so $(\lambda, \mathbf{u})$ is an eigenpair of $\mathbf{A}\mathbf{A}^\dagger$. By Corollary 2.7, there exists unitary matrices $\mathbf{V}$ and $\mathbf{U}$ which columns are the eigenvectors of $\mathbf{A}^\dagger\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\dagger$ respectively. Let $\mathbf{V} = [\mathbf{V}_1\ \mathbf{V}_2]$ and $\mathbf{U} = [\mathbf{U}_1\ \mathbf{U}_2]$ where the subscript 2 denotes the eigenvectors that span the null space of the corresponding matrix. Let the columns of $\mathbf{U}_1$ be constructed as described above. Since the strictly positive eigenvalues of $\mathbf{A}^\dagger\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\dagger$ coincide, we may define a matrix $\mathbf{\Lambda}^{1/2} = \bigoplus_{i=1}^r \lambda_i^{1/2}$ where λ_i corresponding eigenvalue of ith columns of both $\mathbf{V}_1$ and $\mathbf{U}_1$. Finally, define a matrix $\mathbf{\Sigma} \in \mathbb{R}_+^{m\times n}$ by $\mathbf{\Sigma} = \mathbf{\Lambda}^{1/2} \oplus \mathbf{0}_{(m-r)\times(n-r)}$. Now, by the construction of $\mathbf{U}_1$, we have

$$\mathbf{A}\mathbf{V} = \mathbf{A}[\mathbf{V}_1\ \mathbf{V}_2] = [\mathbf{A}\mathbf{V}_1\ \mathbf{A}\mathbf{V}_2] = [\mathbf{U}_1\mathbf{\Lambda}^{1/2}\ \mathbf{0}] = \mathbf{U}\mathbf{\Sigma}\,.$$

Since $\mathbf{V}$ is unitary, we have

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$$

as desired. Moreover, if we arrange the diagonal values of $\mathbf{\Sigma}$ in decreasing order, then by Proposition 2.4, $\mathbf{\Sigma}$ is uniquely determined. □

Corollary 2.12. *Left and right singular vectors of a matrix* $\mathbf{A}$ *are the eigenvectors of* $\mathbf{A}\mathbf{A}^\dagger$ *and* $\mathbf{A}^\dagger\mathbf{A}$ *respectively.*

Proof. See proof of Theorem 2.11. □

2.3 Other Important Results

Proposition 2.13. *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *be a matrix and let* $(\lambda, \mathbf{v})$ *be its eigenpair. Then,* $(\lambda + \alpha, \mathbf{v})$ *is an eigenpair of* $\mathbf{A} + \alpha\mathbf{I}$ *where* $\mathbf{I}$ *is the identity matrix.*

Proof. Suppose $(\lambda, \mathbf{v})$ is an eigenpair of a matrix $\mathbf{A}$. Then, by a straight computation we have

$$(\mathbf{A} + \alpha\mathbf{I})\mathbf{v} = \mathbf{A}\mathbf{v} + \alpha\mathbf{v} = (\lambda + \alpha)\mathbf{v}$$

which proves the claim. □

Proposition 2.14. *Let* $\mathbf{A}$ *be a normal matrix and let* $(\lambda, \mathbf{v})$ *be its eigenpair with* $\lambda \neq 0$*. Then, the matrix* $\mathbf{A} - \lambda\mathbf{v}\mathbf{v}^\dagger$ *has the same eigenvalues than* $\mathbf{A}$ *except that* λ *is replaced by* 0*.*

Proof. Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ be a normal matrix and let $(\lambda, \mathbf{v})$ be an eigenpair of $\mathbf{A}$. Since $\|\mathbf{v}\| = 1$, also $\|\mathbf{v}\|^2 = \mathbf{v}^\dagger\mathbf{v} = 1$. Then,

$$(A - \lambda\mathbf{v}\mathbf{v}^\dagger)\mathbf{v} = \mathbf{A}\mathbf{v} - \lambda\mathbf{v}\mathbf{v}^\dagger\mathbf{v} = \lambda\mathbf{v} - \lambda\mathbf{v} = 0\mathbf{v}\,,$$

so $\mathbf{v}$ is an eigenvector of $\mathbf{A} - \lambda\mathbf{v}\mathbf{v}^\dagger$ with corresponding eigenvalue 0. Now, suppose $(\mu, \mathbf{u})$ is an eigenpair of $\mathbf{A}$ with $\mathbf{u} \neq \mathbf{v}$. Since $\mathbf{A}$ is normal, by Corollary 2.7 its eigenvectors are orthogonal, so

$$(\mathbf{A} - \lambda\mathbf{v}\mathbf{v}^\dagger)\mathbf{u} = \mathbf{A}\mathbf{u} - \lambda\mathbf{v}\mathbf{v}^\dagger\mathbf{u} = \mathbf{A}\mathbf{u} = \mu\mathbf{u}\,.$$

Therefore, $(\mu, \mathbf{u})$ is an eigenpair of $\mathbf{A} - \lambda\mathbf{v}\mathbf{v}^\dagger$. □

Definition 2.7. The *column rank* (*row rank*) of a matrix is the number of linearly independent columns (rows). The *rank* of a matrix is defined to be its column rank[2].

[2]Column rank and row rank of a matrix are always equal [1, p. 72].

Theorem 2.15. (Rank-Nullity Theorem) *Let* $\mathbf{A} \in \mathbb{C}^{m\times n}$*. Then, the rank of* $\mathbf{A}$ *and the dimension of the null space of* $\mathbf{A}$ *add up to the number of columns in* $\mathbf{A}$*, that is,*

$$\operatorname{rank}(\mathbf{A}) + \operatorname{nul}(\mathbf{A}) = n\,.$$

Proof. See [1, pp. 71–72]. □

Theorem 2.16. (Sylvester's Determinant Theorem) *Let* $\mathbf{A} \in \mathbb{C}^{m\times n}$ *and* $\mathbf{B} \in \mathbb{C}^{n\times m}$*. Then,*

$$\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA})$$

where $\mathbf{I}_k$ *is the* $k \times k$ *identity matrix.*

Proof. Proof can be found in [2, pp. 416–417]. □

Theorem 2.17. (Hadamard's inequality) *Let* $\mathbf{A} = (a_{ij}) \in \mathbb{C}^{n\times n}$ *be a Hermitian positive semi-definite matrix. Then,*

$$\det(A) \leq \prod_{i=1}^{n} a_{ii}$$

where equality holds if and only if $\mathbf{A}$ *is diagonal.*

Proof. See [4]. □

Chapter 3

Channel Capacity

In order to measure the performance of a communication system, the concept of channel capacity is necessary. Channel capacity is a measure of maximal transmission rate such that arbitrarily small decoding error rate can be achieved. On the contrary, transmitting at higher rate will always lead to decoding errors. The fundamental result on channel capacity was given by Shannon [5] in 1948.

This chapter consists of two sections. First, Section 3.1 is based on sphere packing arguments on channel capacity used in [6]. This geometric approach serves as an introduction to channel capacity and gives an intuitive basis for Shannon-Hartley theorem which will be presented later in Chapter 4. Section 3.2 is devoted to Shannon's theorem [5] which is the fundamental theorem in information theory. As a background for this theorem, we give definitions for entropy of a random variable describing an information source, and for mutual information of two random sources. The most crucial definitions and results are all formulated so that they support capacity discussions for MIMO channels later in Chapters 4 and 5.

3.1 Capacity via Sphere Packing

In this section, capacity of an additive white Gaussian noise (AWGN) channel is derived via heuristic sphere packing arguments. The approach to find an upper bound for capacity is based on the presentation of [6]. A lower bound is found by a similar approach.

The AWGN channel is the basic model to describe information flow from a transmitter to a receiver. This channel model is the basis of all other models used in this thesis. In the most basic form, AWGN channel is given

by

$$y = x + n \tag{3.1}$$

where $x \in \mathbb{R}$ is a transmitted symbol, $n \sim \mathcal{N}(0, \sigma^2)$ is a real white Gaussian noise, and $y \in \mathbb{R}$ is a received signal. Model (3.1) is referred to as *real* AWGN channel, since the transmitted and received symbols lie on the real axis. From channel capacity point of view it is important to notice that in this basic case, the transmitted symbols lie on a one-dimensional line.

Suppose that the transmitter chooses transmitted symbols from a fixed signal constellation, and the receiver uses maximum likelihood estimation [7, pp. 173–196] for decoding the received signals. A decoding error then occurs if the received signal is closer to some other constellation point than the original symbol. In practice, the constellation points cannot be placed arbitrarily far away from each others since the transmitter has only limited transmitting power. Therefore, with this model the decoding error rate cannot be arbitrarily small.

One simple idea to increase communication reliability is to use repetition code of length N, which means that each symbol is repeated N times before moving on to the next symbol. For example, if a binary input is assumed, then the receiver can find the correct output even if at most $(N-1)/2$ of the received N symbols are erroneous[1]. Therefore, if the decoding error rate is less than $N/2$, then by using repetition code of length N, all the errors can be corrected. However, increasing the code length N decreases the data rate, and thus, in practice this coding method is rather inefficient.

While using the length N repetition code is not the best option for fast and reliable communication, its geometric interpretation is important. We can interpret the transmitted symbols as N-dimensional vectors of the form $\begin{bmatrix} a & a & \dots & a \end{bmatrix}^T$ where a is the symbol that is desired to be transmitted. Therefore, we can consider that the signal constellation is the set of these repetition vectors in the N-dimensional space. Thus, the AWGN channel can be written in the form

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{3.2}$$

where $\mathbf{x} = (x_i) \in \mathbb{R}^N$ is a vector representing one constellation point, $\mathbf{n} = (n_i)$ is the noise whose entries are i.i.d. real Gaussian random variables with $n_i \sim \mathcal{N}(0, \sigma^2)$, and $\mathbf{y} = (y_i)$ is the received signal. Assuming the receiver uses maximum likelihood estimation for decoding in this N-dimensional space, then a decoding error occurs only if the constellation points are too close together compared to the magnitude of noise. In the

[1]Here we assumed maximum likelihood detection for each symbol and majority decision for determining the final output.

case of repetition code, the possible values of $\mathbf{x}$ lie all on the same line. Therefore, due to power limitations only a few constellation points can be used to maintain a large enough space between these points to avoid decoding errors. This inefficient use of the N-dimensional space causes repetition code to have poor performance.

Let us now consider what happens if the constellation points are more efficiently packed in the N-dimensional space. Suppose that the expected power input per channel use is $\mathbb{E}[x_i^2] = P$, and the noise variance is $\mathbb{E}[n_i^2] = \sigma^2$. Then, by Jensen's inequality[2] and linearity of expectation, the expected norm of the received vector follows

$$\mathbb{E}[\|\mathbf{y}\|] \leq \sqrt{N(P+\sigma^2)}\,. \tag{3.3}$$

The exact details can be found in Appendix B.1. Therefore, by the strong law of large numbers[3], in the long term the received symbols $\mathbf{y}$ lie almost surely in an N-sphere of radius $\sqrt{N(P+\sigma^2)}$. This allows us to concentrate only on the interior of this sphere and neglect everything that happens at the exterior.

Since a decoding error occurs only if the received symbol is closer to a different constellation point than that was sent, the constellation points need to be placed sparsely enough. To find the minimum required separation between constellation points, notice that the expected deviation from the transmitted symbol follows

$$\mathbb{E}[\|\mathbf{y}-\mathbf{x}\|] = \mathbb{E}[\|\mathbf{n}\|] \leq \sqrt{N\sigma^2} \tag{3.4}$$

by the same reasoning that is used to prove equation (3.3). Again, by the strong law of large numbers, equation (3.4) implies that the received symbols lie almost surely within an N-sphere of radius $\sqrt{N\sigma^2}$ centered at the corresponding transmitted symbols. Therefore, to achieve arbitrarily small decoding error rate and to satisfy the power constraint, we need to have non-overlapping N-spheres of radius $\sqrt{N\sigma^2}$ that lie within an N-sphere of radius $\sqrt{N(P+\sigma^2)}$.

The amount of information carried by one symbol is logarithmically dependent on the number of constellation points. Therefore, to find an estimate of the maximum possible data rate, we need an estimate of how many N-spheres of radius $\sqrt{N\sigma^2}$ fit into an N-sphere of radius $\sqrt{N(P+\sigma^2)}$. An

[2]Jensen's inequality [8], [9]: $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ if φ is convex. Inequality is reversed if φ is concave.

[3]Strong law of large numbers [10, pp. 237–241]: Suppose a random variable X has finite first moment, that is, $\mathbb{E}[|X|] < \infty$. Then, the sample average of a random variable converges almost surely to the expected value.

upper bound of this value is given by the ratio of the volumes[4]:

$$\frac{\frac{\pi^{N/2}}{\Gamma(N/2+1)}\sqrt{N(P+\sigma^2)}^N}{\frac{\pi^{N/2}}{\Gamma(N/2+1)}\sqrt{N\sigma^2}^N} = \left(1+\frac{P}{\sigma^2}\right)^{N/2}. \tag{3.5}$$

This value can be turned into data rate $C(N)$ by converting it to bits and taking into account the number of channel uses:

$$C(N) \leq \frac{\log_2\left(1+\frac{P}{\sigma^2}\right)^{N/2}}{N} = \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right). \tag{3.6}$$

We will later see in Section 4.1 that this data rate upper bound given by (3.6) is actually the maximum channel capacity of the AWGN channel. While this upper bound was obtained by using non-exact estimates, it is rather straightforward to use these same arguments to get a lower bound to the maximum data rate. In equation (3.4), we used the ratio of volumes to estimate the number of smaller N-spheres that fit into a larger N-sphere. However, this is an optimistic approximation, since in practice there will be gaps between the spheres.

To get a good estimate for the lower bound of the maximum data rate, we use the following theorem conjectured by Minkowski [12] and first proved by Hlawka [13]:

Theorem 3.1. (Minkowski-Hlawka Theorem) *Let Δ be the density of the best packing of N-spheres. Then, Δ satisfies*

$$\Delta \geq \frac{\zeta(N)}{2^N}$$

where $\zeta(N) = \sum_{k=1}^{\infty} k^{-N}$ is the Riemann zeta function.

Proof. Various proofs can be found in the literature, e.g., [13], [14]. □

Since $\zeta(N) \geq 1$ for all positive N, Theorem 3.1 implies a weaker result that the density is at least 2^{-N}. Therefore, a lower bound for the data rate $C(N)$ is given by

$$C(N) \geq \frac{1}{N}\log_2\left(2^{-N}\left(1+\frac{P}{\sigma^2}\right)^{N/2}\right) = \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right) - 1. \tag{3.7}$$

[4]Volume of an N-sphere [11] is given by $\frac{\pi^{N/2}}{\Gamma(N/2+1)}r^N$ where r is the radius of the sphere and $\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}\mathbf{d}x$ is the Gamma function.

Notice that this lower bound is close to the data rate estimate of (3.6). Thus, it seems reasonable that the maximum data rate of an AWGN channel with arbitrarily small error rate, i.e., the channel capacity, is indeed

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right) \tag{3.8}$$

for codes in $\mathbb{R}^N$. If complex constellation points are used instead, then one complex channel use corresponds to two real channel uses. Hence, for complex constellations, equation (3.8) can be written as

$$C' = \log_2\left(1 + \frac{P}{\sigma^2}\right). \tag{3.9}$$

Equations (3.8) and (3.9) suggest that strictly positive data rate can be achieved while having arbitrarily small error rate.

Based on the heuristic arguments used in this section, it seems rather straightforward to implement a communication system that reaches channel capacity in AWGN SISO channels. Moreover, channel capacity seems not to be affected by the dimensionality N, i.e., the block length of the code. However, the arguments used here were heavily relying on the strong law of large numbers which does not state anything about the rate of convergence. Therefore, it may take extremely long time before the average error rate evens out. In the real world applications this behavior is not acceptable, however, and more predictable performance properties are desired.

Many error-correction codes have been designed to detect and correct decoding errors. The state-of-the-art methods include turbo codes used in the Long-Term Evolution (LTE) standard [15], low-density parity-check codes used in Wi-Fi [16], and polar codes [17]. In the future 5th generation (5G) mobile networks, both low-density parity-check codes and polar codes will be used [18]. Using these codes, it is possible to achieve data rates close to channel capacity [19], [20], [21].

3.2 Shannon's Theorem

The foundations of information theory were given by Shannon in [5]. In his paper, Shannon modeled the information flow from a transmitter to a receiver as a Markov process. He considered the information carried by a random variable as its *entropy* which measures how uncertain the outcome of a discrete random event is given its probability mass function (PMF). Using the concept of entropy, Shannon was able to prove his fundamental

theorems of noiseless and noisy channels which are here formulated as a one theorem referred to as *Shannon's Theorem*. This theorem states that there exists a limit in the data rate which can be achieved with arbitrarily small error rate, but cannot be exceeded.

This section considers the key concepts related to Shannon's theorem. We begin with a brief discussion on the background of the entropy in the discrete case, which is based on the approach used by Shannon in [5]. After this, we switch to the continuous case, and then further extend it to the vector notation in order to support discussion on the MIMO channel capacity in the subsequent chapters.

In [5], Shannon looked for a measure of uncertainty of a discrete random variable X with PMF p. He argued that this uncertainty (or entropy), if it can be defined, reflects the information content of a random variable that models an information source. Shannon required that the measure of the entropy H of a random variable X must satisfy the following three properties:

(i) H is continuous in p,

(ii) if X has uniform distribution, then H is a monotonic increasing function on the number of possible states of X, and

(iii) if the event of deciding the value of X can be broken down to two consecutive random events, then the entropy of the original event is a weighted sum of the entropies of the two events.

Shannon proved that the only function that satisfy all these properties is of the form

$$H(X) = -K\mathbb{E}[\log p(X)] \tag{3.10}$$

where K is a positive constant and the logarithm can be taken in any base.

Since the constant K is just a scaling factor, a natural choice for it is $K = 1$. On the other hand, in the information theoretic context the most natural choice for the base of the logarithm is 2 that gives the entropy in bits. This allows us to write the definition of entropy of a discrete random variable as follows.

Definition 3.1. *Entropy* (or *discrete entropy*) H of a discrete random variable X with PMF p is

$$H(X) = -\mathbb{E}[\log_2 p(X)]$$

and the unit of measure is a bit.

Definition 3.1 has multiple natural consequences which should be intuitively true due to the interpretation of entropy as uncertainty. For example, entropy is always non-negative and an event with only one possible outcome has entropy equal to zero. It is also true that uniform random events have the most entropy, since

$$\begin{aligned} h(X) &= -\mathbb{E}[\log_2 p(X)] = \sum_{x\in\mathcal{S}} p(x)\log_2\frac{1}{p(x)} \\ &\leq \log_2\left(\sum_{x\in\mathcal{S}} p(x)\frac{1}{p(x)}\right) = \log_2 n \end{aligned} \tag{3.11}$$

where $\mathcal{S}$ is the set of all possible states of X, n is the size of $\mathcal{S}$, and the equality is attained if and only if X is uniformly distributed. In (3.11), the inequality is obtained by using Jensen's inequality on the function $f(z) = z\log_2(1/z)$ that is concave on the interval $[0, 1]$.

Shannon studied in his paper [5] also continuous channels which can be considered generalizations of discrete channels. Here we concentrate on the continuous channels, since they seem to be more widely used in the modern literature, e.g., [6], [22], [23]. Using random variables, it is possible to obtain the channel capacity of a SISO channel. However, in this thesis, we are interested in the channel capacity of MIMO systems. Therefore, we need to extend the study from random variables to random vectors which can be used to model multiple parallel data transmissions.

We begin the discussion by defining the (differential) entropy of a continuous real random variable.

Definition 3.2. *(Differential entropy of a real random variable)*
Let $X : \Omega \to \mathbb{R}$ be a continuous real random variable, and let p_X be its probability density function (PDF). Then, the differential entropy of X is given by

$$h(X) = -\mathbb{E}[\log_2 p_X(X)]\,.$$

Notice that the definition of differential entropy seems very similar to that of discrete entropy. Unfortunately, differential entropy does not have all the same properties as discrete entropy. For example, differential entropy can be negative since PDFs can be greater than 1. Therefore, differential entropy cannot be seen as a direct generalization of discrete entropy despite its similar definition. An exact generalization of entropy to the case of continuous random variables[5] is derived in [24]. However, for our purposes the differential entropy is sufficient, since we will only use it to define a quantity called

[5]Entropy of a continuous random variable [24]: Let x_i for $i = 1, \ldots, n$ be discrete

mutual information later on in this section. While differential entropy may not be the most exact generalization of discrete entropy, mutual information shares the same intuition in both discrete and continuous cases.

Let us now generalize the definition of differential entropy to real random vectors. To make notation nicer, we will first introduce a shorthand for a specific type of n-fold integrals. Suppose $f : \mathbb{R}^n \to \mathbb{R}_+$ is a PDF and $g : \mathbb{R}_+ \to \mathbb{R}$ is some function. Then, we let

$$\langle g(f(\mathbf{x}))\rangle = \int_{\mathbb{R}^n} f(\mathbf{x})g(f(\mathbf{x}))\mathbf{d}\mathbf{x}\,. \tag{3.12}$$

Using this notation, we may write the definition of the differential entropy of a real random vector. We will also define *joint* and *conditional* entropies which are later needed for the definition of mutual information.

Definition 3.3. *(Differential entropy of a real random vector)*
Let $\mathbf{x} : \Omega \to \mathbb{R}^{n\times 1}$ be a real random vector, and let $p_{\mathbf{x}}$ be the PDF of $\mathbf{x}$. Then, the differential entropy of $\mathbf{x}$ is given by

$$h(\mathbf{x}) = -\langle \log_2 p_{\mathbf{x}}(\mathbf{x})\rangle\,.$$

Further, let $p_{\mathbf{x},\mathbf{y}}$ be the joint PDF of $\mathbf{x}$ and $\mathbf{y}$. Then, the joint differential entropy of $\mathbf{x}$ and $\mathbf{y}$ is given by

$$h(\mathbf{x},\mathbf{y}) = -\langle \log_2 p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})\rangle\,.$$

Finally, the conditional differential entropy of $\mathbf{x}$ given $\mathbf{y}$ can be written as

$$h(\mathbf{x}|\mathbf{y}) = h(\mathbf{x},\mathbf{y}) - h(\mathbf{y})\,.$$

Differential entropy for real valued random vectors is a natural generalization for the one-dimensional definition. To see this, we can find a close connection between these cases if the components of the random vector are independent. Suppose that $\mathbf{x} : \Omega \to \mathbb{R}^n$ is a random vector whose components x_i are independent random variables for all $i = 1,\ldots,n$. Then the

---

points. Then, let

$$\lim_{n\to\infty}\frac{1}{n}|\{x_i \mid x_i \in (a,b)\}| = \int_a^b m(x)\mathbf{d}x\,.$$

Suppose p_X is the PDF of a real random variable X. Then, the entropy of X is given by

$$-\int_{\mathbb{R}} p_X(x)\log_2\frac{p_X(x)}{m(x)}\mathbf{d}x\,.$$

differential entropy of $\mathbf{x}$ satisfies

$$h(\mathbf{x}) = -\langle \log_2 p_{\mathbf{x}}(\mathbf{x}) \rangle = -\left\langle \log_2 \left(\prod_{i=1}^{n} p_{x_i}(x_i) \right) \right\rangle$$
$$= -\sum_{i=1}^{n} \langle \log_2 p_{x_i}(x_i) \rangle = \sum_{i=1}^{n} h(x_i) \,.$$

Intuitively this means that uncertainty increases additively if more independent random events are added. For example, one coin toss has one bit of uncertainty, whereas n coin tosses have n bits of uncertainty.

Definition 3.3 would allow us to define the mutual information of two real random vectors. However, in wireless communications, transmitted and received signals are described by complex valued objects. Therefore, the definition of differential entropy needs to be further generalized for complex random vectors. Definition 3.3 cannot be used with complex vectors as it is, since the logarithm would cause difficulties. To prevent this issue, we may convert complex vectors to their *real representation*, and then use the definition of real valued vectors to this new random vector. This approach is used, for example, in [25].

Definition 3.4. *(Real representation of a complex vector)*
Let $\mathbf{x} \in \mathbb{C}^{n \times 1}$ be a complex vector. Then, the real representation of $\mathbf{x}$ is the vector

$$\mathbf{x}^{(r)} = \left[\Re\{\mathbf{x}^T\} \; \Im\{\mathbf{x}^T\} \right]^T \in \mathbb{R}^{2n \times 1}$$

where $\Re\{\cdot\}$ is the real part and $\Im\{\cdot\}$ is the imaginary part, taken componentwise.

Definition 3.5. *(Differential entropy of a complex random vector)*
Let $\mathbf{x} : \Omega \to \mathbb{C}^{n \times 1}$ be a complex valued random vector. Then, the differential entropy of $\mathbf{x}$ is given by its real representation $\mathbf{x}^{(r)}$, that is

$$h(\mathbf{x}) = h\big(\mathbf{x}^{(r)}\big) \,.$$

The joint and conditional entropies are defined analogously:

$$h(\mathbf{x}, \mathbf{y}) = h(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}) \,,$$
$$h(\mathbf{x}|\mathbf{y}) = h(\mathbf{x}^{(r)}|\mathbf{y}^{(r)}) \,.$$

Finally, we are ready to define the mutual information of two complex valued random vectors. This definition is then used to describe the key theorem that allows us to derive the desired channel capacity formulas in Chapters 4 and 5.

Definition 3.6. *(Mutual information of two random vectors)*
The mutual information of two random vectors $\mathbf{y}$ and $\mathbf{x}$, denoted by $I(\mathbf{y};\mathbf{x})$, is given by

$$\begin{aligned} I(\mathbf{y};\mathbf{x}) &= h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \\ &= h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y}) \\ &= I(\mathbf{x};\mathbf{y})\,. \end{aligned}$$

In the discrete case, this definition is the same except differential entropies are interchanged with discrete entropies.

Intuitively, mutual information describes how much the uncertainty of one source is reduced if the other source is known. In other words, mutual information is a measure of how much information is gained from one source if the other source is known. For example, if random vector sources $\mathbf{x}$ and $\mathbf{y}$ are independent, then knowing $\mathbf{x}$ does not yield any knowledge on $\mathbf{y}$ and vice versa. Therefore, in this case the mutual information is equal to zero.

Shannon [5] found that the mutual information describes the data rate from the transmitter to the receiver. On one hand, the entropy of the received vector tells the amount of information the receiver receives, and on the other hand, the conditional entropy tells how much information the transmitter needs to provide so that the receiver can decode its received information correctly. Therefore, the channel capacity can be seen as the mutual information of the transmitted and received signal. This observation led to the fundamental theorems for noiseless and noisy channels [5], which are stated here as a single general theorem referred to as *Shannon's Theorem*.

Theorem 3.2. (Shannon's Theorem) *Let a random vector $\mathbf{x}$ be an input of a memoryless channel and let a random vector $\mathbf{y}$ be its output. Then, the channel capacity C of the channel is given by*

$$C = \sup_{p_{\mathbf{x}}} I(\mathbf{y};\mathbf{x})$$

where $p_{\mathbf{x}}$ is the PDF of $\mathbf{x}$. More precisely, for every $\varepsilon > 0$ and $R < C$, there exists a block code of length N such that the data rate is at least R and the block error probability is at most ε. Moreover, a data rate strictly greater than C cannot be achieved.

By a block code we here mean that the coding method is applied to each length N message block separately, and the block error probability is the probability that a received block is erroneous even if the required decoding methods are used.

Conclusion

In this chapter, we built up intuition behind channel capacity by studying the sphere packing problem. Then, we defined the information content of a random variable to be its entropy. The definition of entropy was then extended to continuous random vectors. Using this extended definition, we were able to define the mutual information of two random vectors. Finally, this enabled us to describe Shannon's Theorem 3.2 that gives a limit for the maximum channel capacity in communication over a noisy channel.

Chapter 4

Single-User MIMO

In this chapter we study a communication system that consists of a base station and a user that both have multiple transmitting and/or receiving antennas. Since both the transmitter and the receiver have multiple antennas, this is a MIMO system. More precisely, there is only one user, so this system is called single-user MIMO (SU-MIMO). In this thesis, mostly downlink[1] transmission is considered. Therefore, from now on we assume that the base station is the transmitter unless stated otherwise.

Compared to SISO systems, SU-MIMO systems have more degrees of freedom due to the MIMO nature. Since there exists multiple inputs and outputs, SU-MIMO systems support multiple simultaneous spatial streams, which can have significant gain in the system throughput [23]. By a spatial stream we here mean independently encoded data that is simultaneously transmitted to a receiver. However, in order to have the maximum benefit from MIMO, a coding technique must be chosen carefully. If a poor coding technique is used, then different spatial streams interfere each other, which increases the chance of a receiver decoding error. In this chapter, beamforming techniques for SU-MIMO are considered.

This chapter consists of three sections. First, in Section 4.1, formula for SU-MIMO capacity is derived from Shannon's theorem 3.2. As a corollary to this formula, Shannon-Hartley theorem is presented. Then, in Section 4.2, we study the optimal beamforming method for SU-MIMO. This section is subdivided into two parts. The first of them considers a special case in which the number of spatial streams is limited to one. The second part then studies the general beamforming solution for SU-MIMO allowing multiple simultaneous spatial streams. Finally, Section 4.3 derives a water-filling principle that solves the power allocation problem emerged from the gen-

[1]Downlink: The base station is the transmitter and a user is the receiver.

eral SU-MIMO beamforming solution. This principle will be later used in Chapter 5 in the multi-user beamforming context.

4.1 Single-User Channel Capacity

In this section we use Shannon's theorem to derive[2] a formula for SU-MIMO channel capacity. This result is then used in Section 4.2 to find an optimal beamforming method for SU-MIMO. In the end of this section, a connection is made to the AWGN channel capacity result obtained in Section 3.1.

Let us now study implications of Shannon's theorem on SU-MIMO channels. A general SU-MIMO channel can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{4.1}$$

where $\mathbf{x}$ is the transmitted complex vector, $\mathbf{H}$ is the channel matrix, $\mathbf{n}$ is the random noise, and $\mathbf{y}$ is the received complex vector [23]. Complex values are used since they capture both the amplitude and the phase of radio waves that are here referred to as *signals*. The channel matrix $\mathbf{H} = (h_{ij})$ describes the channel response for each pair of antennas. More precisely, the element h_{ij} tells how a unit signal transmitted from the jth transmitting antenna is received at the ith receiving antenna. On the other hand, vectors $\mathbf{x}$ and $\mathbf{y}$ determine the signals transmitted and received at each antenna respectively, so that each element of these vectors corresponds to one transmitting or receiving antenna. Furthermore, the vector $\mathbf{n}$ describes the noise that affects the received signals.

Suppose that the base station has M transmitting antennas and the user has K receiving antennas, then the channel matrix $\mathbf{H} \in \mathbb{C}^{K\times M}$ is a $K \times M$ complex matrix. The vectors $\mathbf{x}$, $\mathbf{n}$, and $\mathbf{y}$ can be seen as random variables. We model the noise as circularly symmetric complex Gaussian (CSCG) random vector with covariance $\mathbb{E}\left[\mathbf{n}\mathbf{n}^\dagger\right] = \sigma^2\mathbf{I}$, which we write as $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. This is a standard assumption in literature, see e.g. [6], [22], [23], and it is motivated by the central limit theorem[3].

By Shannon's theorem, the channel capacity is given by

$$C = \sup I(\mathbf{y}; \mathbf{x})\,. \tag{4.2}$$

Using Definition 3.6, we can write the mutual information of $\mathbf{y}$ and $\mathbf{x}$ in the

[2]Similar derivation for these results can be found in [22].

[3]Central limit theorem [10, pp. 258–265]: Normalized sum of independent random variables with finite expected value converges to Gaussian distribution.

form

$$\begin{aligned} I(\mathbf{y};\mathbf{x}) &= h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \\ &= h(\mathbf{y}) - h(\mathbf{n})\,. \end{aligned} \tag{4.3}$$

More details regarding the latter equation can be found in Appendix B.2. The following lemma gives us tools to find the maximum value of the mutual information.

Lemma 4.1. *Let* $\mathbf{z}$ *be a complex random vector with covariance* $\mathbf{R}_z = \mathbb{E}\big[\mathbf{z}\mathbf{z}^\dagger\big]$. *Then, the differential entropy of* $\mathbf{z}$ *satisfies*

$$h(\mathbf{z}) \leq \log_2 \det(\pi e \mathbf{R}_z)$$

where equality holds if and only if $\mathbf{z}$ *is CSCG.*

Proof. See [22] for a proof. □

Using Lemma 4.1, we may find a strict upper bound for the mutual information of $\mathbf{y}$ and $\mathbf{x}$. Suppose that the random vector $\mathbf{x}$ has covariance $\mathbb{E}\big[\mathbf{x}\mathbf{x}^\dagger\big] = \mathbf{R}_x$. Then, by the linearity of expectation, $\mathbf{y}$ has covariance

$$\mathbb{E}\big[\mathbf{y}\mathbf{y}^\dagger\big] = \mathbb{E}\big[(\mathbf{H}\mathbf{x}+\mathbf{n})(\mathbf{H}\mathbf{x}+\mathbf{n})^\dagger\big] = \sigma^2\mathbf{I} + \mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\,. \tag{4.4}$$

By a straightforward computation, $\mathbf{y}$ is CSCG if and only if $\mathbf{x}$ is CSCG. Therefore, by Lemma 4.1, the mutual information satisfies

$$\begin{aligned} I(\mathbf{y};\mathbf{x}) &\leq \log_2 \det\big(\pi e\big(\sigma^2\mathbf{I} + \mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\big)\big) - \log_2 \det\big(\pi e \sigma^2 \mathbf{I}\big) \\ &= \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right) \end{aligned} \tag{4.5}$$

where equality holds if and only if $\mathbf{x}$ is CSCG. By Shannon's theorem, channel capacity is given as the supremum of (4.5) over the covariance $\mathbf{R}_x$. We summarize this result into the following theorem.

Theorem 4.2. *Suppose that a SU-MIMO channel is given by*

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ *is the transmitted complex signal with covariance* $\mathbb{E}\big[\mathbf{x}\mathbf{x}^\dagger\big] = \mathbf{R}_x$, $\mathbf{H}$ *is the channel matrix,* $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ *is the random noise, and* $\mathbf{y}$ *is the received complex signal. Then, the channel capacity of this system is given by*

$$C = \sup_{\mathbf{R}_x} \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right)$$

which can be achieved only if $\mathbf{x}$ *is CSCG.*

In Theorem 4.2 capacity can be viewed as the spectral efficiency since the bandwidth of the system is not considered. If we would like to emphasize the effect of bandwidth, we could replace C with C/B where B is the system bandwidth. However, in this thesis we choose to identify channel capacity and spectral efficiency. Therefore, the unit of channel capacity C is bit/s/Hz.

An important consequence of Theorem 4.2 is the following result originally due to work of Hartley [26] and Shannon [5].

Corollary 4.3. (Shannon-Hartley Theorem) *Channel capacity of an AWGN SISO channel is given by*

$$C = \log_2\left(1 + \frac{P}{\sigma^2}\right)$$

where P is the average power of the signal, and σ^2 is the variance of the noise.

Proof. AWGN SISO channel with average power constraint P corresponds to case $\mathbf{H} = 1$ and $\mathbf{R}_x = P$ in Theorem 4.2. □

The term P/σ^2 is known as signal-to-noise ratio (SNR). Notice that Shannon-Hartley theorem gives the same channel capacity for AWGN SISO channels as was suggested by equation (3.9).

4.2 Eigen-Beamforming

In the previous section, capacity of a SU-MIMO channel was derived. In this section, we will use Theorem 4.2 to find an optimal beamforming technique for SU-MIMO, which achieves the Shannon capacity[4]. This method requires that both the base station and the user have perfect channel state information (CSI), i.e., the channel matrix $\mathbf{H}$ is known to both of them. Later on, we will see that this requirement can be relaxed to the knowledge of equivalent channel, which means that both parties know the channel that includes the beamforming coefficients of the other party.

This section consists of two parts. In the first part, we consider a special case where the transmitter transmits only one spatial stream at a time. Then, in the second part, the general SU-MIMO case with multiple simultaneous spatial streams is studied. It is found that the optimal beamforming method for the general case can be viewed as iteratively applying the single spatial stream solution. The only problem that will remain is the power allocation which is addressed in Section 4.3.

[4]Shannon capacity: Capacity given by Shannon's theorem.

4.2.1 Single Spatial Stream

Let us consider a SU-MIMO channel where the base station has M transmitting antennas and the user has K receiving antennas. The goal is to find a coding scheme that maximizes the system throughput. Before studying the general case, we suppose that only one spatial stream can be transmitted at a time, which implies that no inter-stream interference[5] is present. We assume perfect CSI both for the transmitter and for the receiver.

Suppose the transmitter transmits a complex data symbol $s \in \mathbb{C}$. This symbol is mapped to the transmitting antennas by applying a precoding $\mathbf{w} \in \mathbb{C}^{M\times 1}$ to it. The transmitted signal can be written as

$$\mathbf{x} = \mathbf{w}s \in \mathbb{C}^{M\times 1}\,. \tag{4.6}$$

Intuitively, precoding means that the transmitted symbol is sent from antennas at different phases and amplitudes. By the superposition principle of electromagnetic waves, this generates a constructive signal interference in some directions and destructive interference at other directions causing a beam-like radiation pattern. Example visualization of this radiation pattern can be found from Appendix A.1. Due to the geometric beam interpretation, this precoding technique is called *beamforming*. The precoding vectors are called *beamforming coefficients*[6].

Using equation (4.6), the channel model for this single spatial stream SU-MIMO system can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{w}s + \mathbf{n} \tag{4.7}$$

where $\mathbf{H}$ is the channel matrix, s is the complex data symbol, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is the random noise, and $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is the received signal. To decode this signal, the receiver employs decoding $\mathbf{m} \in \mathbb{C}^{K\times 1}$ such that the decoded symbol is given as an inner product of the received signal and the decoding vector:

$$\hat{s} = \mathbf{m}^\dagger\mathbf{y} = \mathbf{m}^\dagger\mathbf{H}\mathbf{w}s + \mathbf{m}^\dagger\mathbf{n}\,. \tag{4.8}$$

In order to keep signal and noise levels untouched, we require that $\mathbf{m}$ is a unit vector. Since matrix multiplication is a linear transform and $\mathbf{m}$ is a unit vector, the noise term follows $\mathbf{m}^\dagger\mathbf{n} \sim \mathcal{CN}(0, \sigma^2)$. Further, we may assume

[5]Inter-stream interference: In this thesis, inter-stream interference refers to the interference that may occur if simultaneously received spatial streams intended to the same user interfere each others.

[6]Later on, also precoding matrices as well as decoding vectors and decoding matrices are called beamforming coefficients.

that also the precoding vector $\mathbf{w}$ is a unit vector, since the transmit power can be included in the transmitted symbol s.

Notice that $\mathbf{m}^\dagger\mathbf{H}\mathbf{w}$ is constant if we fix the beamforming and decoding coefficients. Now, instead of considering the channel from the transmitter antenna array to the receiver antenna array, we may consider the channel that includes beamforming and decoding processes. This interpretation allows us to write the channel model (4.8) in the form

$$\hat{s} = s' + n \tag{4.9}$$

where $s' = \mathbf{m}^\dagger\mathbf{H}\mathbf{w}s$ is the input, and $n = \mathbf{m}^\dagger\mathbf{n}$ the noise. Therefore, this is an AWGN SISO channel, so the capacity is given by Shannon-Hartley theorem 4.3. Assuming the average transmit power is P_t, the channel capacity is given by

$$C = \sup_{\mathbf{w},\mathbf{m}} \log_2\left(1 + \frac{P_t}{\sigma^2}|\mathbf{m}^\dagger\mathbf{H}\mathbf{w}|^2\right). \tag{4.10}$$

Therefore, the optimal precoding and decoding vectors are such that they maximize $|\mathbf{m}^\dagger\mathbf{H}\mathbf{w}|^2$. This means that the optimal capacity in this case is achieved by maximizing signal power at the receiver.

Let the singular value decomposition (SVD) of the channel matrix $\mathbf{H}$ be

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger \tag{4.11}$$

where $\mathbf{U} \in \mathbb{C}^{K\times K}$ and $V \in \mathbb{C}^{M\times M}$ are unitary matrices and $\mathbf{\Sigma} \in \mathbb{R}_+^{K\times M}$ is a diagonal matrix with non-negative real diagonal entries in decreasing order. Let us denote $\mathbf{m}^\dagger\mathbf{U} = \mathbf{u}^\dagger$ and $\mathbf{V}^\dagger\mathbf{w} = \mathbf{v}$. Since $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, they preserve vector (Euclidean) norms. Hence, maximization of $|\mathbf{m}^\dagger\mathbf{H}\mathbf{w}|^2$ is reduced to maximization of

$$|\mathbf{u}^\dagger\mathbf{\Sigma}\mathbf{v}|^2 \tag{4.12}$$

where $\mathbf{u}$ and $\mathbf{v}$ are unit vectors. If $K \neq M$, then only $\min\{K, M\}$ first dimensions have a contribution in (4.12). Therefore, without loss of generality, we may assume that $K = M$.

Let $\mathbf{\Lambda}^2 = \mathbf{\Sigma}$ where $\mathbf{\Lambda} = \bigoplus_{i=1}^M \lambda_i^{1/2}$ is a diagonal matrix with non-negative real entries. Then, by Cauchy-Schwarz inequality[7]

$$\begin{aligned} |\mathbf{u}^\dagger\mathbf{\Sigma}\mathbf{v}|^2 = |(\mathbf{\Lambda}\mathbf{u})^\dagger(\mathbf{\Lambda}\mathbf{v})|^2 &\leq \|\mathbf{\Lambda}\mathbf{u}\|^2\|\mathbf{\Lambda}\mathbf{v}\|^2 \\ &= \sum_{i=1}^K \lambda_i|u_i|^2 \sum_{i=1}^M \lambda_i|v_i|^2 . \end{aligned} \tag{4.13}$$

[7]Cauchy-Schwarz inequality [9]: $|\mathbf{x}^\dagger\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$ with equality if and only if $\mathbf{x} = c\mathbf{y}$ for some $c \in \mathbb{C}$.

It turns out that the last expression in (4.13) is maximized at the same time when the equality holds in the Cauchy-Schwarz approximation. To demonstrate this, we need the following lemma.

Lemma 4.4. *Let $a_i, x_i \in \mathbb{R}$ for all $i = 1, \ldots, n$ satisfying $a_1 \geq \cdots \geq a_n \geq 0$ and $\sum_{i=1}^{n} x_i^2 = 1$. Then,*

$$\max_{x_1,\ldots,x_n} \sum_{i=1}^{n} a_i x_i^2 = a_1$$

and the maximum is achieved with $x_1 = 1$, $x_2 = \cdots = x_n = 0$.

Proof. See Appendix B.3. □

The equality in the Cauchy-Schwarz inequality holds when the two vectors are parallel. On the other hand, by Lemma 4.4, the last expression in (4.13) is maximized when $\mathbf{\Lambda u}$ and $\mathbf{\Lambda v}$ are both parallel to the first coordinate axis. Now, we may apply Lemma 4.4 to equation (4.13) to find

$$|\mathbf{u}^\dagger \mathbf{\Sigma v}|^2 \leq \sum_{i=1}^{K} \lambda_i |u_i|^2 \sum_{i=1}^{M} \lambda_i |v_i|^2 \leq \lambda_1^2 \,. \tag{4.14}$$

Equality in (4.14) holds if and only if $\mathbf{u}$ and $\mathbf{v}$ are both parallel to the first coordinate axis. Since $\mathbf{u} = \mathbf{m}^\dagger \mathbf{U}$ and $\mathbf{v} = \mathbf{V}^\dagger \mathbf{w}$, and $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, the optimal $\mathbf{m}$ and $\mathbf{w}$ are the first columns of $\mathbf{U}$ and $\mathbf{V}$ respectively. We formalize this result in the following proposition.

Proposition 4.5. *Let $\mathbf{H} \in \mathbb{C}^{K \times M}$ be the channel matrix of a SU-MIMO channel known to both transmitter and receiver, and let $\mathbf{H} = \mathbf{U\Sigma V}^\dagger$ be the SVD of $\mathbf{H}$. Then, the capacity while transmitting a single spatial stream is maximized when the transmitter and the receiver choose their precoding and decoding vectors to be the first column of $\mathbf{U}$ and $\mathbf{V}$ respectively.*

Intuitively, Proposition 4.5 means that if only one spatial stream is used, then all the transmitting power should be concentrated on the direction[8] of the most dominant right singular vector of the channel matrix $\mathbf{H}$. Receiver, on the other hand, receives the signal best from the direction[9] of the most

[8]Remark that the direction given by beamforming coefficients does not mean the geometric real world direction the signal is transmitted to. After all, the beamforming coefficients lie on a high-dimensional space whereas the transmitted signals are in the 3-dimensional space. Correspondence between these two spaces is determined by the geometry of the antenna array.

[9]See footnote 8.

dominant left singular vector of $\mathbf{H}$. Therefore, capacity is maximized by using channels as an advantage rather than trying to cancel its effects.

In the next part of this section, we will study the general SU-MIMO capacity. An optimal solution will be derived, and connection to this single spatial stream case is made. This leads to an iterative interpretation of the general solution in which each iteration is a single spatial stream optimization.

4.2.2 Multiple Spatial Streams

Previously, we studied SU-MIMO beamforming with an additional constraint that only one spatial stream can be transmitted at a time. Now, we will consider the general case where multiple simultaneous spatial streams are allowed. This will lead to the optimal general SU-MIMO beamforming solution. Various similar derivations as presented here for this result can be found in literature, e.g., [22] and [23].

Suppose the base station employs M transmitting antennas and the receiver has K receiving antennas. Then, the general SU-MIMO channel model is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{4.15}$$

where $\mathbf{H} \in \mathbb{C}^{K\times M}$ is the channel matrix, $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the transmitted signal, $\mathbf{n} \in \mathbb{C}^{K\times 1}$ is the CSCG noise with covariance $\sigma^2\mathbf{I}$, and $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is the received signal. By Theorem 4.2, the capacity of this channel is

$$C = \sup_{\mathbf{R}_x} \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right) \tag{4.16}$$

where $\mathbf{R}_x = \mathbb{E}\left[\mathbf{x}\mathbf{x}^\dagger\right]$ is the covariance matrix of the transmitted signals. Therefore, finding the capacity C amounts to maximizing the determinant $\det\left(\mathbf{I} + (1/\sigma^2)\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right)$.

As a covariance matrix, $\mathbf{R}_x$ is Hermitian and therefore normal. By the spectral theorem it has a decomposition

$$\mathbf{R}_x = \mathbf{W}\mathbf{\Gamma}\mathbf{W}^\dagger \tag{4.17}$$

where $\mathbf{W}$ is a unitary matrix and $\mathbf{\Gamma}$ is a diagonal matrix with non-negative real entries. Let the SVD of $\mathbf{H}$ be

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger \tag{4.18}$$

where $\mathbf{U} \in \mathbb{C}^{K\times K}$ and $\mathbf{V} \in \mathbb{C}^{M\times M}$ are unitary matrices, and $\mathbf{\Sigma} \in \mathbb{R}_+^{K\times M}$ is a diagonal matrix. Using equations (4.17) and (4.18), we have

$$\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right) = \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger\mathbf{W}\mathbf{\Gamma}\mathbf{W}^\dagger\mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^\dagger\right). \tag{4.19}$$

Let us denote $\mathbf{V}^\dagger\mathbf{W} = \mathbf{B}$. Then, using Sylvester's determinant theorem 2.16, equation (4.19) simplifies to

$$\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right) = \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{B}\mathbf{\Gamma}\mathbf{B}^\dagger\right). \tag{4.20}$$

By Hadamard's inequality 2.17, to maximize this expression, the matrix inside the determinant should be diagonal. This is true if and only if $\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{B}\mathbf{\Gamma}\mathbf{B}^\dagger$ is diagonal. On the other hand, if this is the case, then by Proposition 2.4, $\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{B}\mathbf{\Gamma}\mathbf{B}^\dagger$ is equal to $\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}$ up to permutation of $\mathbf{\Gamma}$. Since this permutation can be included into $\mathbf{\Gamma}$, we have

$$\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger\right) \leq \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\right) \tag{4.21}$$

where equality holds if $\mathbf{B} = \mathbf{I}$, i.e. $\mathbf{V}^\dagger\mathbf{W} = \mathbf{I}$. Therefore, the general SU-MIMO capacity can be re-written as

$$C = \sup_{\mathbf{\Gamma}} \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\right). \tag{4.22}$$

The optimal selection of $\mathbf{\Gamma}$ is discussed in the next section.

Now that we have found the general SU-MIMO capacity, it is easy to design a beamforming method that achieves this capacity. Capacity given by (4.22) is satisfied if $\mathbf{V}^\dagger\mathbf{W} = \mathbf{I}$. Notice that this is true if and only if the beamforming coefficients $\mathbf{W}$ are chosen to be the right singular vectors of the channel matrix $\mathbf{H}$, that is, $\mathbf{W} = \mathbf{V}$. Now, let $\mathbf{x} = \mathbf{V}\mathbf{s}$ where $\mathbf{s} \in \mathbb{C}^{M\times 1}$ is the vector containing the transmitted data symbols. Then, by the linearity of expectation, $\mathbf{x}$ has covariance

$$\mathbf{R}_x = \mathbf{V}\mathbb{E}\left[\mathbf{s}\mathbf{s}^\dagger\right]\mathbf{V}^\dagger. \tag{4.23}$$

Assuming the transmitted symbols are independent of each other, the covariance $\mathbf{R}_s$ of $\mathbf{s}$ may be chosen as

$$\mathbf{R}_s = \mathbf{\Gamma} \tag{4.24}$$

where $\mathbf{\Gamma}$ is a diagonal matrix. Since the diagonal entries of $\mathbf{\Gamma}$ can be controlled by altering the input power to each data stream, $\mathbf{\Gamma}$ is called the power allocation matrix. Now, $\mathbf{x}$ has the desired covariance to attain capacity given by (4.22). This result is summarized in the following theorem.

Theorem 4.6. (Eigen-beamforming) *Suppose a transmitter and a receiver communicate over a SU-MIMO channel defined by the channel matrix* $\mathbf{H}$*. Then, the optimal channel capacity up to power allocation is attained by transmitter choosing the beamforming coefficients to be the right singular vectors of the channel matrix* $\mathbf{H}$*.*

Theorem 4.6 shows that the optimal channel capacity is obtained by choosing the transmit beamforming coefficients to be the right singular vectors of the channel matrix $\mathbf{H}$. Notice that there are no requirements for the receiver when maximizing the capacity. The reason for this is that the information has already transfered when the receiver receives the signal. Therefore, it is up to the receiver how it extracts the information content, but the amount of received data cannot be affected at this point.

The result given by Theorem 4.6 is analogous to the single spatial stream result given in Proposition 4.5 which tells that using the most dominant right singular vector for transmit beamforming coefficients is the optimal solution. Even stronger connection between the single and multiple spatial stream results may be found by considering this general solution as an iterative process.

Applying Sylvester's determinant theorem to SU-MIMO capacity given by Theorem 4.2 yields

$$C = \sup_{\mathbf{R}_x} \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}^\dagger\mathbf{H}\mathbf{R}_x\right). \tag{4.25}$$

Instead of saying that the beamforming coefficients should be the right singular vectors of $\mathbf{H}$, we may equivalently consider them as the eigenvectors of $\mathbf{H}^\dagger\mathbf{H}$ (see Corollary 2.12). Then, maximizing the channel capacity for a single spatial stream, the most dominant eigenvector is selected. By Hadamard's inequality, to maximize the general SU-MIMO capacity, there should not be inter-stream interference.

Suppose the most dominant eigenvector of $\mathbf{H}^\dagger\mathbf{H}$ is $\mathbf{v}$ with corresponding eigenvalue λ. Then, consider the matrix $\mathbf{H}' = \mathbf{H}^\dagger\mathbf{H} - \lambda\mathbf{v}\mathbf{v}^\dagger$ which has the same eigenpairs as $\mathbf{H}^\dagger\mathbf{H}$, but the eigenvalue corresponding to $\mathbf{v}$ is changed to 0. Now, we may apply the single spatial stream selection to $\mathbf{H}'$ and continue the process iteratively. Ultimately, the resulting beamforming coefficients will be the eigenvectors of $\mathbf{H}^\dagger\mathbf{H}$ which are the same as the right singular vectors of $\mathbf{H}$. This shows that the general SU-MIMO solution can be regarded as iteratively applying the single spatial stream solution while requiring there is no inter-stream interference.

Notice that the receiver can use this same method to derive beamforming coefficients to extract the information in the optimal way. The single spatial stream solution suggest that the receiver should use the the most dominant left singular vector. Following the previous arguments, we may see that in the multiple spatial stream case this means that as many left singular vectors as there are spatial streams should be used. If the transmitter uses right singular vectors and the receiver uses left singular vectors of a channel

matrix $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$, then the channel model is written as

$$\hat{\mathbf{s}} = \mathbf{U}^\dagger\mathbf{H}\mathbf{V}\mathbf{s} + \mathbf{U}^\dagger\mathbf{n} = \mathbf{\Sigma}\mathbf{s} + \mathbf{U}^\dagger\mathbf{n} \tag{4.26}$$

where the elements of $\hat{\mathbf{s}}$ are the received estimates for the transmitted symbols $\mathbf{s}$, and $\mathbf{n}$ is the noise. Since $\mathbf{\Sigma}$ is a diagonal matrix, the receiver can decode each symbol independently. Moreover, the matrix $\mathbf{U}^\dagger$ is unitary, so the noise properties are not changed.

4.2.3 Practical View

In this subsection, SU-MIMO beamforming is considered from a practical point of view. For deriving the optimal beamforming method, it was easiest to work with SVD due to its generality. However, SVD provides unnecessary information since, for example, the transmitter only needs to find the right singular vectors[10] of the channel matrix $\mathbf{H}$. Therefore, in practice the desired coefficients are determined from the eigenvalue decomposition of $\mathbf{H}^\dagger\mathbf{H}$ (or from $\mathbf{H}\mathbf{H}^\dagger$ depending on the direction at which the channel is viewed). This matrix is sometimes referred to as the *channel covariance matrix*. Since the eigenvalue decomposition has all the necessary information, it provides a more efficient method to compute the beamforming coefficients. Due to the eigenvector interpretation of the beamforming coefficients, this beamforming method is called *eigen-beamforming*. For now on this term is used to describe the SU-MIMO beamforming method.

Let us now study some aspects of eigen-beamforming that were neglected while deriving the method. First, we assumed perfect CSI, but did not motivate this. Perfect CSI is a rather optimistic assumption due to delays between channel estimation and applying the beamforming coefficients. However, with static users the channel matrix $\mathbf{H}$ can be estimated accurately assuming stationary environment. Details about channel estimation in orthogonal frequency division multiplexing systems, used in current mobile networks, can be found in [27]. Therefore, we can conclude that perfect CSI is a sound assumption with static users. Moreover, in mobile networks the channel estimation is done very frequently (around every one millisecond in 5G [18]), so even moving users with relatively low velocity can be viewed to be stationary within this time frame.

Now that we know the channel matrix can be obtained in practice, let us consider what happens if there are some random errors in the channel

[10] In section 4.3, we will find that also the singular values of the channel matrix $\mathbf{H}$ are needed for the optimal solution

estimation. Suppose that the correct channel matrix is $\mathbf{H}$, but the estimated channel matrix is

$$\mathbf{H}' = \mathbf{H} + \mathbf{H}_e \tag{4.27}$$

where $\mathbf{H}_e$ an error matrix. Let us consider the matrix $\mathbf{H}_e$ as a random matrix, and assume its element are zero mean. Moreover, let us assume that the error matrix satisfies $\mathbb{E}\left[\mathbf{H}_e^\dagger \mathbf{H}_e\right] = \sigma_e^2 \mathbf{I}$. This assumption is motivated by the fact that each antenna has its own signal pipeline. Supposing the errors occur at these pipelines, we may see that the errors are related to individual antennas rather than pairs of antennas.

The expected channel covariance matrix can be written as

$$\mathbb{E}\Big[(\mathbf{H}')^\dagger \mathbf{H}'\Big] = \mathbf{H}^\dagger \mathbf{H} + \sigma_e^2 \mathbf{I}\,. \tag{4.28}$$

Now, if $(\lambda^2, \mathbf{v})$ is an eigenpair of $(\mathbf{H}')^\dagger \mathbf{H}'$, then by Proposition 2.13, $(\lambda^2 + \sigma^2, \mathbf{v})$ is an eigenpair of $\mathbb{E}[(\mathbf{H}')^\dagger \mathbf{H}']$. We notice that the expected eigenvectors are the same as the correct ones, but the corresponding eigenvalues are slightly perturbed. Since the error variance σ_e^2 is presumably rather small, we can see that channel estimation errors should not have major impact on the system performance.

Previous discussion should not be confused with channel aging which means that the estimated channel matrix is outdated once it is used. In general, in channel aging the error matrix $\mathbf{H}_e$ does not necessarily have diagonal covariance which means that the beamforming coefficients will be different for the aged matrix and for the correct one. Intuitively, these dependencies in the error matrix arise from the geometry of the channel—usually most of the environment remains unchanged, but some local change, e.g. movement of the user or bypassing car, causes the channel matrix to change. Channel aging will be discussed in Chapter 6.

Finally, let us study how eigen-beamforming affects the channel estimation. We assume perfect channel reciprocity [23, pp. 153–154]. In channel estimation, information about the channel matrix $\mathbf{H}$ can be measured from carefully chosen reference symbols [28]. If these symbols are beamformed, then instead of obtaining the channel matrix $\mathbf{H}$, the receiver obtains an estimate for the equivalent channel that is for the matrix $\mathbf{H}' = \mathbf{H}\mathbf{W}$ (or $\mathbf{H}' = \mathbf{H}^\dagger \mathbf{W}$ depending on the direction we are considering the channel) where $\mathbf{W}$ are the beamforming coefficients. Assuming $\mathbf{W}$ is unitary, as in eigen-beamforming, we may notice that the desired channel covariance matrix $\mathbf{H}\mathbf{H}^\dagger$ (or $\mathbf{H}^\dagger \mathbf{H}$) is obtained as

$$\mathbf{H}'(\mathbf{H}')^\dagger = \mathbf{H}\mathbf{W}\mathbf{W}^\dagger \mathbf{H}^\dagger = \mathbf{H}\mathbf{H}^\dagger\,. \tag{4.29}$$

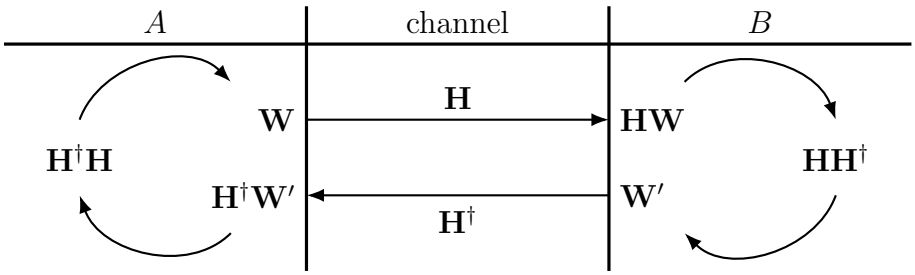


Figure 4.1: Flow for computing beamforming coefficients $\mathbf{W}$ and $\mathbf{W}'$ based on unitarily beamformed reference symbols.

Therefore, the eigen-beamforming coefficients can be computed regardless of the beamforming coefficients $\mathbf{W}$.

Figure 4.1 visualizes the flow of computing beamforming coefficients based on unitarily beamformed reference symbols. In this figure, two parties A and B communicate over a channel that is assumed to be constant and to be defined by a matrix $\mathbf{H}$. Parties A and B can be regarded as a base station and a user respectively. The cycle can start with either A or B choosing unitary beamforming coefficients which can be as simple as the identity matrix. Then, the other party estimates the equivalent channel matrix, which can be then used to compute the covariance of the actual channel matrix. This covariance matrix is then used to calculate the eigen-beamforming coefficients, which are then applied in reply. Both parties are able to use this same exact procedure to find their beamforming coefficients.

4.3 Water-Filling Principle

In Section 4.2.2, we found that the channel capacity C is given by

$$C = \sup_{\mathbf{\Gamma}} \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\right) \tag{4.30}$$

where σ^2 is the noise power, $\mathbf{\Sigma}$ is the singular value matrix of the channel, and $\mathbf{\Gamma}$ is the power allocation matrix. Therefore, in order to find the channel capacity, we are left with optimizing the diagonal matrix $\mathbf{\Gamma}$. This will lead to well-known *water-filling* solution (e.g. [29]).

Let $\mathbf{\Sigma}^\dagger\mathbf{\Sigma} = \bigoplus_{i=1}^{M} \lambda_i^2$ and $\mathbf{\Gamma} = \bigoplus_{i=1}^{M} P_i$. Since $\mathbf{I}$, $\mathbf{\Sigma}$, and $\mathbf{\Gamma}$ are all diagonal,

we have

$$\log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{\Gamma}\right) = \log_2 \prod_{i=1}^{M}\left(1 + \frac{\lambda_i^2 P_i}{\sigma^2}\right) = \sum_{i=1}^{M}\log_2\left(1 + \frac{\lambda_i^2 P_i}{\sigma^2}\right). \tag{4.31}$$

Now, the optimization problem can be written formally as

$$\begin{aligned} \underset{P_1,\ldots,P_M}{\text{maximize}} \quad & \sum_{i=1}^{M}\log_2\left(1 + \frac{\lambda_i^2 P_i}{\sigma^2}\right) \\ \text{subject to} \quad & \sum_{i=1}^{M} P_i \le P \\ & P_i \ge 0 \quad \forall i = 1,\ldots,M\,. \end{aligned} \tag{4.32}$$

The constraints here require that the system total input power may not exceed the given value P, and naturally, all the powers need to be non-negative.

We can solve this problem by using Lagrange multipliers. Since a change of logarithm base is just a scalar multiplication, we may change the base 2 to the natural base e without affecting the optimal power allocations. Let us write the Lagrange function in the form

$$\mathcal{L}(P_1,\ldots,P_M,\mu) = \sum_{i=1}^{M}\ln\left(1 + \frac{\lambda_i^2 P_i}{\sigma^2}\right) - \mu\left(\sum_{i=1}^{M} P_i - P'\right) \tag{4.33}$$

where $0 \le P' \le P$. Now, the partial derivatives are given by

$$\frac{\partial\mathcal{L}}{\partial P_i} = \frac{\lambda_i^2}{\sigma^2 + \lambda_i^2 P_i} - \mu \tag{4.34}$$

$$\frac{\partial\mathcal{L}}{\partial\mu} = P' - \sum_{i=1}^{M} P_i\,. \tag{4.35}$$

Critical points are those where (4.34) and (4.35) are zero. Now, the condition $\partial\mathcal{L}/\partial P_i = 0$ leads to

$$P_i = \xi - \frac{\sigma^2}{\lambda_i^2} \tag{4.36}$$

where $\xi = 1/\mu$. Notice that if $P_i = 0$ for some i, the corresponding term vanishes from the objective function. Therefore, since P_i is required to be non-negative for each i, we may take

$$P_i = \left(\xi - \frac{\sigma^2}{\lambda_i^2}\right)^+ \tag{4.37}$$

where $(x)^+ = \max\{0, x\}$. On the other hand, ξ can be chosen such that the condition $\partial\mathcal{L}/\partial\mu = 0$ is met. From equation (4.37), we can see that each P_i grows with ξ, and on the other hand, the objective function grows with P_i. Therefore, the maximum channel capacity is achieved when we choose ξ as large as possible, that is, to satisfy

$$\sum_{i=1}^{M}\left(\xi - \frac{\sigma^2}{\lambda_i^2}\right)^+ = P\,. \tag{4.38}$$

Choosing ξ as in (4.38)[11] and allocating power as in (4.37) is called the *water-filling principle* [29]. We summarize the general SU-MIMO results in the following proposition.

Proposition 4.7. *In SU-MIMO communications, the optimal transmitting beamforming coefficients are given by the right singular vectors of the channel matrix. Power allocation to the ith stream with the corresponding singular value λ_i is given by*

$$P_i = \left(\xi - \frac{\sigma^2}{\lambda_i^2}\right)^+$$

where ξ satisfies

$$\sum_{i=1}^{M}\left(\xi - \frac{\sigma^2}{\lambda_i^2}\right)^+ = P\,.$$

In this case, the maximum theoretical channel capacity is given by

$$C = \sum_{i=1}^{M}\left(\log_2\left(\frac{\xi\lambda_i^2}{\sigma^2}\right)\right)^+.$$

[11] Notice that the left-hand-side defines a continuous function as it is a sum of composites of continuous functions. Therefore, by Bolzano's theorem [30], a solution to this equation exists.

Conclusion

In this chapter, we discussed SU-MIMO beamforming. We began with deriving the SU-MIMO capacity from Shannon's theorem. Then, we maximized this capacity, and found that the optimal beamforming coefficients for SU-MIMO are given by the right singular vectors of the channel matrix. Since the singular vectors can be interpreted as eigenvectors of the channel covariance matrix, this beamforming method is called eigen-beamforming. We further optimized precoding by deriving the water-filling power allocation method. Practical aspects of eigen-beamforming were also discussed.

Chapter 5

Multi-User MIMO

In the previous chapter, we discussed SU-MIMO communication systems where eigen-beamforming was used to achieve the maximum channel capacity. We saw that in this single user case it is possible to analytically derive the optimal solution including beamforming coefficients and power allocation to each spatial stream. In this chapter, we will study a more general setup—the multi-user MIMO (MU-MIMO). As the name suggests, in this case there are multiple users instead of just one. Furthermore, these users are assumed to be non-cooperative since otherwise they could be treated as a single decentralized user.

The key difference between the single user and the multi-user cases is that in MU-MIMO the transmitter simultaneously transmits spatial streams that are not all intended to the same user. This means that for each user there is a desired and an undesired signal transmitted from the base station. The undesired signals, i.e., *inter-user interference*, are treated as noise. Due to this interference, in general, capacity for each user cannot be maximized independently.

Dirty paper coding [31] achieves the optimal channel capacity in channels subject to interference. Instead of trying to cancel the interference, transmitter adapts to it, and as a result, the achieved performance is the same as if there was no interference. While this method gives theoretically the best possible performance, it is hard to implement in practice.

Another method to achieve optimal channel capacity is interference alignment [32]. This technique maximizes interference overlapping, and thus, minimizes the signal space that is occupied by interference. In the other words, the interference-free signal space is maximized. However, there exists multiple practical issues that makes implementation of interference alignment difficult in large networks [33].

In this thesis, we concentrate on the sub-optimal methods based on beam-

forming. Similarly to Chapter 4, we start by using Shannon's theorem to derive the channel capacity for MU-MIMO in Section 5.1. Then, in Sections 5.2 and 5.3, multiple beamforming methods for MU-MIMO are discussed. In Section 5.2 a greedy approach using eigen-beamforming is considered. We will see that this approach may generate too much interference causing poor system performance. However, this method turns out to be a solid choice if user scheduling is used. In Section 5.3, we set an additional constraint of zero inter-user interference[1], and study beamforming methods that satisfy this requirement. This study is divided into two parts. In the first part, it is assumed that each user has only one receiving antenna. Then, in the second part, the general case where each user may have one or more receiving antennas is considered. Finally, in Section 5.4, we discuss further beamforming solutions for MU-MIMO.

5.1 Multi-User Channel Capacity

So far, we have implicitly assumed that there is no inter-user interference in the transmission. This is the case for example, if all the data is intended to one user and eigen-beamforming is used. In general, however, the data can be intended to multiple independent users and inter-user interference is likely to occur. This interference is perceived as noise at the receiver, thus decreasing the channel capacity.

The channel model for a user experiencing interference is of the form

$$\mathbf{y} = \mathbf{H}(\mathbf{x} + \tilde{\mathbf{x}}) + \mathbf{n} = \mathbf{H}\mathbf{x} + \mathbf{H}\tilde{\mathbf{x}} + \mathbf{n} \tag{5.1}$$

where $\mathbf{x}$ is the intended signal, $\tilde{\mathbf{x}}$ is the unintended signal (i.e., $\mathbf{H}\tilde{\mathbf{x}}$ is the interference), and $\mathbf{n}$ is the random noise. Since the interference-plus-noise part is independent from the intended signal $\mathbf{x}$, the mutual information of $\mathbf{x}$ and $\mathbf{y}$ is then

$$I(\mathbf{y}; \mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) = h(\mathbf{y}) - h(\mathbf{H}\tilde{\mathbf{x}} + \mathbf{n}) \tag{5.2}$$

which follows from the same arguments that was used with equation (4.3). Let us consider all $\mathbf{x}$, $\tilde{\mathbf{x}}$, and $\mathbf{n}$ as CSCG random vectors, and denote $\tilde{\mathbf{n}} = \mathbf{H}\tilde{\mathbf{x}} + \mathbf{n}$. Suppose that the covariance of $\tilde{\mathbf{n}}$ is non-singular. Then, by Lemma

[1]This constraint is also called the zero-forcing condition.

4.1, we have

$$I(\mathbf{y};\mathbf{x}) = \log_2 \det\left(\frac{\mathbb{E}\left[\mathbf{y}\mathbf{y}^\dagger\right]}{\mathbb{E}[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^\dagger]}\right) = \log_2 \det\left(\mathbf{I} + \frac{\mathbf{H}\mathbb{E}\left[\mathbf{x}\mathbf{x}^\dagger\right]\mathbf{H}^\dagger}{\mathbb{E}[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^\dagger]}\right) \tag{5.3}$$

where the division means left or right multiplication with inverse which are the same in this case due to Sylvester's determinant theorem 2.16. Let the covariances be $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\dagger\right] = \mathbf{R}_x$, $\mathbb{E}\left[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^\dagger\right] = \mathbf{R}_{\tilde{x}}$, and $\mathbb{E}\left[\mathbf{n}\mathbf{n}^\dagger\right] = \sigma^2\mathbf{I}$. By the linearity of expectation, (5.3) reduces to

$$I(\mathbf{y};\mathbf{x}) = \log_2 \det\left(\mathbf{I} + \frac{\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger}{\sigma^2\mathbf{I} + \mathbf{H}\mathbf{R}_{\tilde{x}}\mathbf{H}^\dagger}\right) \tag{5.4}$$

which describes the data rate to one user in MU-MIMO system. This generalizes to MU-MIMO capacity by observing that the total system data rate is the sum of data rates to all users. Therefore, we have the following theorem.

Theorem 5.1. *Capacity of a MU-MIMO channel with N users is given by*

$$C = \sup_{\mathbf{R}_{x_1},\ldots,\mathbf{R}_{x_N}} \sum_{i=1}^{N} \log_2 \det\left(\mathbf{I} + \frac{\mathbf{H}_i\mathbf{R}_{x_i}\mathbf{H}_i^\dagger}{\sigma^2\mathbf{I} + \mathbf{H}_i\mathbf{R}_{\tilde{x}_i}\mathbf{H}_i^\dagger}\right)$$

where $\mathbf{H}_i$ is the channel matrix from the base station to the ith user, and $\mathbf{R}_{x_i}$ and $\mathbf{R}_{\tilde{x}_i}$ are the covariances of the intended and unintended signals to the ith user respectively.

5.2 Eigen-Beamforming Approach

In Chapter 4, we learned that the optimal SU-MIMO channel capacity is achieved by eigen-beamforming. In this section, we study how this optimal single-user beamforming method is suited for multi-user applications. This leads to a greedy solution whose performance highly depends on a user-scheduler. Due to the limited scope of this work, only a short discussion on possible solutions with eigen-beamforming is given.

It is straightforward to extend eigen-beamforming to MU-MIMO: consider each user's channel independently and compute eigen-beamforming coefficients using the SU-MIMO solution. The problem with this approach is the correlation between the beamforming coefficients of different users. These

correlations can be caused, for example, by geometrically similar user locations. As a result, users experience interference from other users, since this aspect was neglected while computing the beamforming coefficients.

To overcome the issue with inter-user interference, a user-scheduler needs to be used. The function of this scheduler is to decide for which users the base station can transmit simultaneous signals. In order to take the maximum benefit from this solution, we need to assume there exists a large number of users, which is the case, e.g., in mobile networks.

For convenience, let us say, that beamforming coefficients for one single spatial stream is called an *eigen-beam*. Then, in the ideal case, the scheduler is able to find a set of users such that from their strong eigen-beams, it is possible to construct an orthogonal basis for $\mathbb{C}^M$ where M is the number of base station transmitting antennas[2]. In other words, the scheduler is ideally able to find such users that eigen-beamforming can be used without caring about inter-user interference. If such a set of users can be found, then the channel capacity will be asymptotically the same as that of dirty paper coding. However, in general, this is not the case and the scheduling problem is much more difficult[3].

In [34] and [35], similar iterative scheduling algorithms were proposed. The idea is to successfully minimize the correlation between the current and the previously selected eigen-beams. In [34], it was suggested that this approach is a good candidate for near optimal performance. Due to simplicity of eigen-beamforming compared to dirty paper coding, this approach is rather attractive choice for real world applications.

5.3 Zero-Forcing Beamforming

The sum capacity rate in MU-MIMO promised by the dirty paper coding is hard to achieve in practice [36]. While eigen-beamforming leads to an optimal coding strategy in SU-MIMO, a closed-form optimal beamforming technique in MU-MIMO cannot be found. However, beamforming may lead to the optimal solution under additional constraints. This is the case, for example, if zero inter-user interference is required. This additional requirement leads to a beamforming solution called *zero-forcing beamforming* which proves to be a versatile coding technique.

In this section, various zero-forcing variants are studied. We begin with the case where the base station employs multiple transmitting antennas and

[2]The number of transmitting antennas is the maximum number of different spatial streams.

[3]In [34], it is claimed that the scheduling problem is NP-complete.

there exists multiple independent users each with only one receiving antenna. Section 5.3.1 considers the most basic zero-forcing technique, the *channel inversion* [37]. This technique cancels the channel effect, and therefore, also the inter-user interference is canceled. While this method is simple, it may lead to poor system performance in specific channel conditions. This issue is addressed in Sections 5.3.2 and 5.3.3 in which two different approaches are described that improve channel inversion performance.

In Section 5.3.2, the zero inter-user interference constraint is slightly relaxed. This results in the *regularized inversion* [38] method which removes most of the interference while avoiding the possible problems with poorly conditioned channel matrices. Regularized inversion proves to be a great improvement over the channel inversion in the low SNR regime, however, the high SNR asymptotic performance is the same as that of the plain channel inversion.

In Section 5.3.3, instead of trying to modify inversion process, the transmitted data is modified. The method obtained is called *vector perturbation* [39], which forces the transmitted data not to align with the badly behaving space respect to the channel matrix. However, not an arbitrary modification to the transmitted data can be made. Modification is chosen to be an additive Gaussian integer vector which effect can be removed at the receiver via a modulo operator. The vector perturbation method achieves linear growth in the channel capacity with respect to the number of users even in the high SNR regime, which neither channel inversion nor regularized inversion are capable of achieving.

Section 5.3.4 considers the most general form of MU-MIMO. Instead of each user having only one receiving antenna, any number of receiving antennas at each user are allowed in certain limits that are to be made precise later. While the aforementioned techniques could be used in this case by considering each receiving antenna as an independent receiver, it would be waste of decoding resources. Instead of diagonalizing the effective channel, transmitter may block diagonalize it. This leads to the *block diagonalization* [36] method which decomposes the MU-MIMO channel into independent parallel SU-MIMO channels. This allows eigen-beamforming to be used for each concurrent channel which gives the optimal maximum throughput solution for MU-MIMO if zero inter-user interference is required. However, it turns out this is not the only application of block diagonalization. Using water-filling method reversely results into a solution for *power control* problem. Finally, block diagonalization is extended to the case of partial channel knowledge.

5.3.1 Channel Inversion

We will start the study of various zero-forcing beamforming techniques with the simplest one—the channel inversion [37]. We will study both the downlink and the uplink cases. In downlink, the base station is the transmitter and in uplink the base station is the receiver. We begin with the downlink case.

Suppose that the base station has M transmitting antennas and there exists N non-cooperating users with one receiving antenna each. Assume that the number of users does not exceed the number of transmitting antennas, i.e., $N \leq M$. The channel model for MU-MIMO can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{5.5}$$

where $\mathbf{H} \in \mathbb{C}^{N\times M}$ is the channel matrix, $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the transmitted signal, $\mathbf{n} \in \mathbb{C}^{N\times 1}$ is the CSCG noise, and $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is the received signal at each receiving antenna. Notice that the ith row of the channel matrix is the channel from the base station to the ith user.

Let us assume that the channel matrix has full row rank. Suppose that the transmitter knows the channel matrix $\mathbf{H}$, and it sends a data vector $\mathbf{s} \in \mathbb{C}^{N\times 1}$ where the ith entry corresponds to the data intended to the ith user. To remove the inter-user interference, the transmitter may precode the data vector $\mathbf{s}$ by

$$\mathbf{x} = \mathbf{H}^{+}\mathbf{s} = \mathbf{H}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger})^{-1}\mathbf{s} \tag{5.6}$$

where $\mathbf{H}^{+}$ is the right Moore-Penrose pseudo-inverse [40] of $\mathbf{H}$. Since we assumed that $N \leq M$ and $\mathbf{H}$ has full row rank, the inverse of $\mathbf{H}\mathbf{H}^{\dagger}$ exists. Using this precoding we may notice that the channel equations transforms into

$$\mathbf{y} = \mathbf{s} + \mathbf{n} \tag{5.7}$$

so the data intended to each user can be decoded independently at each receiver assuming that the noise level is small enough. Since the beamforming coefficient are chosen to be the pseudo-inverse of the channel matrix $\mathbf{H}$, this precoding method is called *channel inversion*.

We may notice that a similar beamforming technique can be applied if the base station is in a receiving role instead of being the transmitter. Suppose that the base station has M receiving antennas, and N non-cooperative users transmit data simultaneously to it. It is assumed that $N \leq M$. The channel model is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{5.8}$$

where $\mathbf{H} \in \mathbb{C}^{M\times N}$ is the channel matrix, $\mathbf{s} \in \mathbb{C}^{N\times 1}$ is the data transmitted by the users, $\mathbf{n} \in \mathbb{C}^{M\times 1}$ is the CSCG noise, and $\mathbf{y} \in \mathbb{C}^{M\times 1}$ is the received

data at the base station. Assuming that the channel matrix is of full rank and that the base station knows $\mathbf{H}$, the receiver may apply beamforming by

$$\hat{\mathbf{s}} = \mathbf{H}^{-}\mathbf{y} = (\mathbf{H}^{\dagger}\mathbf{H})^{-1}\mathbf{H}^{\dagger}\mathbf{y} \tag{5.9}$$

where $\mathbf{H}^{-}$ is the left Moore-Penrose pseudo-inverse. Since M has full rank and $N \leq M$, the inverse of $\mathbf{H}^{\dagger}\mathbf{H}$ exists. This uplink beamforming allows us to write the channel model in the form

$$\hat{\mathbf{s}} = \mathbf{s} + \mathbf{H}^{-}\mathbf{n}\,. \tag{5.10}$$

If each component of $\mathbf{H}^{-}\mathbf{n}$ is small enough, maximum likelihood decoder is able to find the data transmitted by the users.

If the channel state information is known at the base station, we can see that it is easy to apply channel inversion to allow multiple simultaneous data streams—at least in theory. In practice, however, poorly conditioned channel matrices may make it difficult to apply channel inversion. In the downlink direction, the base station has some maximum power limitation. If the channel matrix $\mathbf{H}$ has very small singular values, then the inverse covariance matrix $(\mathbf{H}\mathbf{H}^{\dagger})^{-1}$ has very large eigenvalues. This may cause that the antenna data vector $\mathbf{x} = \mathbf{H}^{+}\mathbf{s}$ has too large norm so that the base station power amplifiers cannot provide sufficient input power. In this case, the inversion will be imperfect which may cause poor system performance. On the other hand, looking at equation (5.10), we can see that in the uplink direction, applying channel inversion changes the properties of the noise. This means that if $\mathbf{H}$ has small singular values, then $\mathbf{H}^{-}$ has large singular values, which amplify the noise power. Therefore, even a low noise level may be amplified so that the transmitted data cannot be decoded. In Sections 5.3.2 and 5.3.3 we will discuss methods to overcome the issues with poorly conditioned channel matrices.

5.3.2 Regularized Channel Inversion

Previously, we stated that the channel inversion method is prone to poorly conditioned channel matrices. However, we did not clarify when this undesired behavior occurs. Now, we will show that if the total number of receiving antennas, i.e. the number of users, is the same as the number of transmitting antennas, the expected power needed for the perfect channel inversion is infinite. This motivates the main subject of this section, *regularized inversion*, which overcomes the issue by altering the eigenvalues of the channel covariance matrix before the inversion. On the downside, zero inter-user interference constraint is not strictly satisfied anymore.

Suppose there exists the same number of transmitting and receiving antennas, i.e., $M = N$. Then, the channel inversion described in the previous section becomes

$$\mathbf{x} = \mathbf{H}^{-1}\mathbf{s}\,. \tag{5.11}$$

If $\mathbf{H}$ is a singular matrix, then $\mathbf{x}$ cannot be computed or if $\mathbf{H}$ is near-singular, then the most dominant singular values of $\mathbf{H}^{-1}$ are extremely large causing $\mathbf{x}$ to have too large norm. Since the transmitter power amplifiers can provide only some finite input power, these near-singular channel realizations cause channel inversion to have poor performance in practice.

Let us study this more precisely. Let $\gamma = \|\mathbf{x}\|^2$. It was shown in [41] that γ satisfies PDF[4]

$$f(\gamma) = \frac{N\gamma^{N-1}}{(1+\gamma)^{N+1}}\,. \tag{5.12}$$

Using equation (5.12), we can find the expected value of γ.

$$\begin{aligned}
\mathbb{E}[\gamma] &= \int_0^\infty \gamma f(\gamma)\mathbf{d}\gamma = \int_0^\infty \frac{N\gamma^N}{(1+\gamma)^{N+1}}\mathbf{d}\gamma \\
&= N\int_0^\infty \left(\frac{\gamma}{1+\gamma}\right)^N \frac{1}{1+\gamma}\mathbf{d}\gamma \\
&\geq \frac{N}{2^N}\int_1^\infty \frac{\mathbf{d}\gamma}{1+\gamma} = \frac{N}{2^N}\left(\lim_{\gamma\to\infty}\ln(1+\gamma) - \ln 2\right) \\
&= \infty
\end{aligned}$$

Hence, the expected norm of the transmitted vector is infinite. This causes the asymptotic channel capacity to be constant as the number of users grow (i.e. $N \to \infty$) [38]. Therefore, modifications to the channel inversion scheme must be found in order to support the case $N = M$.

To enhance the performance of zero-forcing beamforming in the case of equal number of transmitting and receiving antennas, we must ensure that the channel matrix $\mathbf{H}$ is not near-singular, that is, the smallest singular value of the channel matrix must be controlled. We know that the singular values of $\mathbf{H}$ are the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$. Since $\mathbf{H}$ was assumed to be a square matrix, its inverse is the same as its pseudo-inverse. Therefore, we may write $\mathbf{H}^{-1} = \mathbf{H}^\dagger\big(\mathbf{H}\mathbf{H}^\dagger\big)^{-1}$. Using this form, we may see that problems arise if the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$ are too small. We may use proposition 2.13 to *regularize*[5]

[4]Assuming that the element of the channel matrix $\mathbf{H}$ can be treated as zero mean unit variance complex random variables.

[5]Regularized inversion given by (5.13) is a special case of Tikhonov regularization [42], [43].

the inversion of $\mathbf{H}$ as

$$\mathbf{H}_\alpha^+ = \mathbf{H}^\dagger(\mathbf{H}\mathbf{H}^\dagger + \alpha\mathbf{I})^{-1} \tag{5.13}$$

where $\alpha \geq 0$ is chosen large enough. This regularization method for channel inversion was suggested in [38]. Now, instead of using the plain channel inversion method, we use

$$\mathbf{x} = \mathbf{H}_\alpha^+\mathbf{s}\,. \tag{5.14}$$

By choosing α appropriately, we can make the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger + \alpha\mathbf{I}$ as large as is necessary, so the norm of $\mathbf{x}$ can be controlled. However, increasing α results into increased inter-user interference, which decreases the system performance. Therefore, the value of α must be chosen carefully.

Using a large-N approximation, it was shown in [38] that the optimal choice of α is

$$\alpha_{\text{opt}} = N\sigma^2 \tag{5.15}$$

independently of the singular values of $\mathbf{H}$. With this value of α, it was shown in the same paper that the channel capacity has an approximate form

$$C \approx N\log_2(1 + \text{SINR}) \tag{5.16}$$

where SINR stands for signal-to-interference-plus-noise ratio. For fixed SINR, we may notice that the capacity grows linearly with the number of users N. This is a major improvement over the plain channel inversion which has a constant asymptotic performance in the same case [38]. However, if $\sigma^2 \to 0$, i.e. $\alpha \to 0$, the regularized inversion approaches the plain channel inversion. Therefore, regularization provides improvement only in low SNR regime, but the performance with high SNR is the same as previously.

5.3.3 Vector Perturbation

In the previous section, we discussed that the regularized inversion method provides a great improvement over the plain channel inversion technique in the low SNR regime. However, in the high SNR scenarios it approaches the channel inversion method, and therefore, it is unable to provide any major improvements in these cases. In this section, we study the method of *vector perturbation*[6] [39], which is able to deliver linear growth in channel capacity with the number of users for the whole SNR range. The key idea is that the transmitted data is altered so that it is not much affected by the large singular values of $\mathbf{H}^{-1}$.

We have discussed that the small singular values of the channel matrix $\mathbf{H}$ cause the transmitted vector $\mathbf{x} = \mathbf{H}^{-1}\mathbf{s}$ to have infinite expected norm in

[6]Vector perturbation is closely related to Tomlinson-Harashima precoding [44], [45].

the case where the number of the transmitting antennas is the same as the number of users. Regularized inversion solves this problem by controlling the inversion by $\mathbf{x} = \mathbf{H}^\dagger(\mathbf{H}\mathbf{H}^\dagger + \alpha\mathbf{I})^{-1}\mathbf{s}$ where $\alpha \geq 0$ is chosen such that the norm of $\mathbf{x}$ is low enough. Another way to solve the problem with inversion is to alter the transmitted data so that it is not much affected by the small singular values. This can be achieved by introducing a perturbation vector that tries to align the most dominant components of the data vector $\mathbf{s}$ along the least dominant singular vectors of $\mathbf{H}^{-1}$. However, the transmitter cannot perturb the data arbitrarily, since otherwise, the receivers could not decode the received signals. Therefore, the goal is to form an altered data vector $\tilde{\mathbf{s}}$ such that $\mathbf{x} = \mathbf{H}^{-1}\tilde{\mathbf{s}}$ has much smaller norm than $\mathbf{H}^{-1}\mathbf{x}$, but $\tilde{\mathbf{s}}$ can be still decoded independently at each receiver.

Instead of an arbitrary complex perturbation, the modified data is set to be of the form

$$\tilde{\mathbf{s}} = \mathbf{s} + \beta\boldsymbol{\ell} \tag{5.17}$$

where $\beta > 0$ is a positive real number and $\boldsymbol{\ell} \in \mathbb{Z}[\mathrm{i}]^{N\times 1}$ is an N-dimensional vector over Gaussian integers[7]. Perturbation defined by (5.17) allows decoding via a modulo function

$$f_\beta(x) = x - \left\lfloor \frac{x + \beta/2}{\beta} \right\rfloor \beta \tag{5.18}$$

where $\lfloor\cdot\rfloor$ is the floor function[8]. To see how this function can be used for decoding, we assume a noiseless channel given by

$$\mathbf{y} = \mathbf{H}\mathbf{x}\,. \tag{5.19}$$

Using the perturbation model (5.17), we can write the previous equation in the form

$$\mathbf{y} = \mathbf{H}\big(\mathbf{H}^{-1}\tilde{\mathbf{s}}\big) = \mathbf{s} + \beta\boldsymbol{\ell}\,. \tag{5.20}$$

The kth user then receives the signal

$$y_k = s_k + \beta\ell_k \tag{5.21}$$

where $s_k \in \mathbb{C}$ and $\ell_k \in \mathbb{Z}[\mathrm{i}]$. Now, both the real and the imaginary part of y_k are of the form $a + \beta b$, where $a \in \mathbb{R}$ and $b \in \mathbb{Z}$. If the modulo function is

[7]Gaussian integers [3, p. 229]: A subset of complex numbers that are defined as $\mathbb{Z}[\mathrm{i}] = \{a + b\mathrm{i} \mid a, b \in \mathbb{Z}\}$.

[8]Floor of real number x, written as $\lfloor x\rfloor$, is the largest integer less than or equal to x.

used for the real and the imaginary part separately, we have

$$\begin{aligned} f_\beta(a+\beta b) &= a+\beta b - \left\lfloor \frac{a+\beta b+\beta/2}{\beta} \right\rfloor \beta \\ &= a + \left(b - \left\lfloor \frac{a}{\beta} + b + \frac{1}{2} \right\rfloor \right) \beta \\ &= a \qquad \text{if and only if } \beta > 2a\,. \end{aligned}$$

Therefore, if the value of β is chosen appropriately and each receiver knows this value, receivers can remove the perturbation effect from the received signal.

Let us now consider the general channel model with noise,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}\,. \tag{5.22}$$

The kth user receives the signal

$$y_k = s_k + \beta \ell_k + n_k \tag{5.23}$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$. Looking at equation (5.23), we may observe that the real and the imaginary parts are now of the form $a + \beta b + n$ for some $a \in \mathbb{R}$, $b \in \mathbb{Z}$, and $n \sim \mathcal{N}(0, \sigma^2)$. Let us now calculate the expected mean squared error (MSE) of the decoded signal. Using the linearity of expectation, we have

$$\mathbb{E}\big[|a - f_\beta(a+\beta b+n)|^2\big] = \sigma^2 + \beta^2 \mathbb{E}\left[\left\lfloor \frac{a+n}{\beta} + \frac{1}{2} \right\rfloor^2 \right] - 2\beta \mathbb{E}\left[\left\lfloor \frac{a+n}{\beta} + \frac{1}{2} \right\rfloor n \right]. \tag{5.24}$$

The following result can be used to show that the asymptotic behavior of this MSE is fully determined by the noise variance.

Proposition 5.2. *Let $a \in \mathbb{R}$ and $n \sim \mathcal{N}(0, \sigma^2)$. Then,*

$$\lim_{\beta \to \infty} \beta^k \mathbb{E}\left[\left\lfloor \frac{a+n}{\beta} + \frac{1}{2} \right\rfloor^\ell n^m \right] = 0$$

for all $k, m \in \mathbb{Z}_+$ and $\ell \in \mathbb{Z}_{\geq 1}$.

Proof. See Appendix B.4. □

By Proposition 5.2 the expected mean square error at limit $\beta \to \infty$ is given by

$$\lim_{\beta\to\infty} \mathbb{E}\big[|a - f_\beta(a + \beta b + n)|^2\big] = \sigma^2 \,. \tag{5.25}$$

Therefore, if large enough β is used at transmission, the decoding error will be at the same level as if the vector perturbation method is not used. However, if β is too large, then the selection of $\boldsymbol{\ell}$ in (5.17) becomes a challenge, which will be discussed next.

The perturbation vector $\boldsymbol{\ell}$ should be chosen such that the modified data vector $\tilde{\mathbf{s}} = \mathbf{s} + \beta\boldsymbol{\ell}$ is not much affected by the small singular values of $\mathbf{H}$. One possible way to achieve this is to choose $\boldsymbol{\ell}$ such that the norm of $\mathbf{x} = \mathbf{H}^{-1}\tilde{\mathbf{s}}$ is minimized [39]. Notice that

$$\|\mathbf{x}\|^2 = \mathbf{x}^\dagger\mathbf{x} = (\mathbf{H}^{-1}\tilde{\mathbf{s}})^\dagger\mathbf{H}^{-1}\tilde{\mathbf{s}} = \tilde{\mathbf{s}}^\dagger(\mathbf{H}\mathbf{H}^\dagger)^{-1}\tilde{\mathbf{s}} \,. \tag{5.26}$$

Therefore, the selection of $\boldsymbol{\ell}$ can be written as an integer lattice optimization problem[9]

$$\boldsymbol{\ell} = \boldsymbol{\ell}(\beta) = \arg\min_{\boldsymbol{\ell}'} (\mathbf{s} + \beta\boldsymbol{\ell}')^\dagger(\mathbf{H}\mathbf{H}^\dagger)^{-1}(\mathbf{s} + \beta\boldsymbol{\ell}') \,. \tag{5.27}$$

Now, if $\beta \to \infty$ we may notice that

$$\lim_{\beta\to\infty} \boldsymbol{\ell}(\beta) = \mathbf{0} \tag{5.28}$$

which basically tells us that β cannot be chosen arbitrarily large. This is an unfortunate results since the decoding MSE is minimized with large β.

In [39], it was suggested that β should be chosen as

$$\beta = 2(|c|_{\max} + \Delta/2) \tag{5.29}$$

where $|c|_{\max}$ is the maximum magnitude of a constellation point and Δ is the separation between the points. This selection is motivated by the fact that $|f_\beta(a + \beta b + n) - a| \leq \Delta/2$ if and only of $\beta > 2(a + n)$ where $a + \beta b + n$ is either the real or the imaginary part of the received signal. If the noise term $n > \Delta/2$, then the signal cannot be decoded anyway. Therefore, selecting β as in (5.29) gives the lowest value for β which enables decoding as if the channel was a plain AWGN.

In [39], it is claimed that the performance of the vector perturbation method is close to Shannon capacity, however, no rigorous proof for this is given. Nevertheless, simulations in this same paper show significant performance gains over the plain channel inversion and over the regularized inversion. The main drawback of the vector perturbation method is the complexity which is caused by solving the integer lattice optimization problem while choosing the perturbation vector $\boldsymbol{\ell}$.

[9] Integer lattice optimization problems can be solved, for example, with the sphere decoder [46].

5.3.4 Block Diagonalization

So far, we have discussed on beamforming solutions in SU- and MU-MIMO scenarios. However, in the multi-user case, we have only considered situations where each user has only one receiving antenna. We will refer this already considered multi-user channel as a *special* MU-MIMO channel to avoid confusion. Now, we will study the *general* MU-MIMO channel in which the base station has M transmitting antennas and there exists N independent users each having K_i receiving antennas for $i = 1, \ldots, N$.

MU-MIMO channel can be viewed as a composition of multiple SU-MIMO channels with an addition of inter-user interference. We have learned that the SU-MIMO capacity is optimized with eigen-beamforming, and zero-forcing methods can be employed in the special MU-MIMO channel. However, these techniques are not suitable for the general MU-MIMO channel.

As discussed in Section 5.2, eigen-beamforming could be used for each user independently. However, this would in general create too much inter-user interference preventing major performance gains. On the other hand, we could interpret each receiving antenna as an individual and use previously discussed zero-forcing methods on this special MU-MIMO channel. However, this would not take advantage on the MIMO processing capabilities of each receiver. Since the general MU-MIMO has properties both from the SU-MIMO and from the special MU-MIMO, one possible solution to achieve good system performance is to use a hybrid method consisting of both eigen-beamforming and zero-forcing elements.

In this section, we concentrate on the method introduced in [36] called *block diagonalization*. This section is divided into three parts. In the first part, we will study the basics of the block diagonalization, and how it is used to achieve maximum throughput while obeying zero inter-user interference constraint. Then, in the second part, the power control problem is considered, that is, how to minimize power consumption while maintaining high enough data rate at each user. Finally, in the third part, the block diagonalization method is generalized into the partial CSI case. In this case, the channel matrix is not fully know, but instead some partial knowledge of it is available.

5.3.4.1 Maximum Throughput

Let us denote the channel matrix and the precoding matrix of the ith receiver by $\mathbf{H}_i \in \mathbb{C}^{K_i \times M}$ and $\mathbf{W}_i \in \mathbb{C}^{M \times S_i}$ respectively where S_i, with $S_i \leq K_i$, is the number of transmitted spatial streams to the ith user. The channel model for the ith receiver can be written as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{n}_i = \mathbf{H}_i \mathbf{W}_i \mathbf{s}_i + \mathbf{n}_i \tag{5.30}$$

where $\mathbf{s}_i$ is the data vector, $\mathbf{x}_i$ is the transmitted signal, $\mathbf{n}_i$ is the noise, and $\mathbf{y}_i$ is the received signal. The goal is to find beamforming coefficients $\mathbf{W}_i$ for $i = 1, \ldots, N$ that maximize the system capacity.

We require that the zero inter-user interference constraint must be satisfied. This requirement can be written formally as

$$\mathbf{H}_j \mathbf{W}_i = \mathbf{0} \tag{5.31}$$

for all $j \neq i$. For convenience, let us denote

$$\tilde{\mathbf{H}}_i = \begin{bmatrix} \mathbf{H}_1^\top & \ldots & \mathbf{H}_{i-1}^\top & \mathbf{H}_{i+1}^\top & \ldots & \mathbf{H}_N^\top \end{bmatrix}^\top . \tag{5.32}$$

Intuitively, $\tilde{\mathbf{H}}_i$ denotes the channel in which the signal intended to the ith user is detected as interference. Using this notation, the zero-forcing condition says that $\mathbf{W}_i$ must lie on the null space of $\tilde{\mathbf{H}}_i$. Under this condition, the maximum number of parallel streams to the ith user is given by

$$\max S_i = \dim(\ker(\tilde{\mathbf{H}}_i)) = \mathrm{nul}(\tilde{\mathbf{H}}_i) \tag{5.33}$$

where $\ker(\tilde{\mathbf{H}}_i) = \{\mathbf{x} \in \mathbb{C}^{M \times 1} \mid \tilde{\mathbf{H}}_i \mathbf{x} = \mathbf{0}\}$ is the kernel of $\tilde{\mathbf{H}}_i$ and $\mathrm{nul}(\tilde{\mathbf{H}}_i)$ is the nullity of $\tilde{\mathbf{H}}_i$. If $\tilde{\mathbf{H}}_i$ has full rank, then by the rank-nullity theorem $\mathrm{nul}(\tilde{\mathbf{H}}_i) = 0$, so the zero inter-user interference constraint cannot be achieved in this case with user i. Therefore, the transmitter can transmit data simultaneously to every user if

$$\max\{\mathrm{rank}(\tilde{\mathbf{H}}_1), \ldots, \mathrm{rank}(\tilde{\mathbf{H}}_N)\} < M \tag{5.34}$$

where M is the number of transmitting antennas.

Let us now write the full system channel equation as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{5.35}$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1^\top & \mathbf{H}_2^\top & \ldots & \mathbf{H}_N^\top \end{bmatrix} \in \mathbb{C}^{\{K_1, \ldots, K_N\} \times M}$ is the full system channel matrix, $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the vector of transmitted signals, $\mathbf{n} \in \mathbb{C}^{\{K_1, \ldots, K_N\} \times 1}$ is the noise, and $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^\top & \mathbf{y}_2^\top & \ldots & \mathbf{y}_N^\top \end{bmatrix}^\top \in \mathbb{C}^{\{K_1, \ldots, K_N\} \times 1}$ is the vector of received signals at every user. Here we emphasized the block structure by writing $\{a_1, \ldots, a_n\}$ in the dimensions meaning that there exists blocks with dimensions $a_1, \ldots, a_n$ and the full matrix dimension is given by $a = \sum_{i=1}^n a_i$. The transmitted signals can be written as

$$\mathbf{x} = \mathbf{W}\mathbf{s}\,, \tag{5.36}$$

where $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \ldots & \mathbf{W}_N \end{bmatrix} \in \mathbb{C}^{M \times \{S_1, \ldots, S_N\}}$ is the precoding matrix and $\mathbf{s} \in \mathbb{C}^{\{S_1, \ldots, S_N\} \times 1}$ is the data vector combining the data intended to each user. Notice the user specific structure of the transmitted signals.

Suppose that the zero-forcing condition (5.34) holds, and the SVD of $\tilde{\mathbf{H}}_i$ is given by

$$\tilde{\mathbf{H}}_i = \tilde{\mathbf{U}}_i \begin{bmatrix} \tilde{\mathbf{\Sigma}}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_i^{(1)} & \tilde{\mathbf{V}}_i^{(0)} \end{bmatrix}^\dagger \tag{5.37}$$

for all i where $\tilde{\mathbf{\Sigma}}_i$ holds all the strictly positive singular values, and the columns of $\tilde{\mathbf{V}}_i^{(1)}$ are the corresponding right singular vectors. Now, $\tilde{\mathbf{V}}_i^{(0)}$ forms an orthogonal basis of the null space of $\tilde{\mathbf{H}}_i$, so we may choose $\mathbf{W}$ as

$$\mathbf{W} = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)} & \tilde{\mathbf{V}}_2^{(0)} & \dots & \tilde{\mathbf{V}}_N^{(0)} \end{bmatrix} \mathbf{W}' \tag{5.38}$$

where $\mathbf{W}'$ is a pre-coding matrix used before zero-forcing. This allows us to write the channel model (5.35) as

$$\mathbf{y} = \left(\bigoplus_{i=1}^{N} \mathbf{H}_i \tilde{\mathbf{V}}_i^{(0)} \right) \mathbf{W}'\mathbf{s} + \mathbf{n}\,. \tag{5.39}$$

Since the first right-hand-side term is block diagonal, we know that there is no inter-user interference. Therefore, the precoding problem is now reduced to find beamforming coefficients $\mathbf{W}_i'$ for each $i = 1, \dots, N$ which maximize the capacities for SU-MIMO channels $\mathbf{y}_i = (\mathbf{H}_i \tilde{\mathbf{V}}_i^{(0)}) \mathbf{W}_i' \mathbf{s}_i + \mathbf{n}_i$ for all $i = 1, \dots, N$.

In Sections 4.2 and 4.3, we found that the optimal SU-MIMO beamforming solution is given by eigen-beamforming and the water-filling principle. Let

$$\mathbf{H}_i \tilde{\mathbf{V}}_i^{(0)} = \mathbf{U}_i \begin{bmatrix} \mathbf{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_i^{(1)} & \mathbf{V}_i^{(0)} \end{bmatrix}^\dagger \tag{5.40}$$

where $\mathbf{\Sigma}_i$ holds the strictly positive singular values and the columns of $\mathbf{V}_i^{(1)}$ are the corresponding right singular vectors. Following the eigen-beamforming method, we choose

$$\mathbf{W}_i' = \mathbf{V}_i^{(1)} \mathbf{\Gamma}_i \tag{5.41}$$

where $\mathbf{\Gamma}_i$ is a power allocation matrix. Therefore, the full precoding matrix can be finally written in the form

$$\mathbf{W} = \begin{bmatrix} \tilde{\mathbf{V}}_1^{(0)} \mathbf{V}_1^{(1)} \mathbf{\Gamma}_1 & \tilde{\mathbf{V}}_2^{(0)} \mathbf{V}_2^{(1)} \mathbf{\Gamma}_2 & \dots & \tilde{\mathbf{V}}_N^{(0)} \mathbf{V}_N^{(1)} \mathbf{\Gamma}_N \end{bmatrix}. \tag{5.42}$$

Optimal power allocation matrices $\mathbf{\Gamma}_i$ can be found by using the water-filling principle on the matrix $\mathbf{\Sigma} = \bigoplus_{i=1}^{N} \mathbf{\Sigma}_i$ with a full system power constraint. If the transmitter uses the precoding matrix $\mathbf{W}$ given by (5.42), then the received signals are given by

$$\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{s} + \mathbf{n} = \left(\bigoplus_{i=1}^{N} \mathbf{H}_i \tilde{\mathbf{V}}_i^{(0)} \mathbf{V}_i^{(1)} \mathbf{\Gamma}_i \right) \mathbf{s} + \mathbf{n}\,. \tag{5.43}$$

Therefore, the receivers may decode the received signals as if the channel was a plain SU-MIMO channel.

Precoding defined by the equation (5.42) is called the *block diagonalization* method since it first block diagonalizes the channel matrix $\mathbf{H}$. Notice that if $K_i = 1$ for all $i = 1, \ldots, N$, then this method diagonalizes the channel matrix $\mathbf{H}$. However, the produced diagonal matrix is not necessarily identity, so the method is not exactly the same as channel inversion even if $K_i = 1$ for all i. Moreover, techniques based on channel inversion require $\sum_{i=1}^{N} K_i \leq M$, but block diagonalization is possible always if condition (5.34) holds. Therefore, block diagonalization is always possible if channel inversion is, but it may be possible even if channel inversion is not.

We finish this section by proving that block diagonalization gives the optimal precoding method subject to the zero-forcing condition. In Section 5.1, we found that the channel capacity in MU-MIMO can be written as

$$C = \sup_{\mathbf{R}_{x_1}, \ldots, \mathbf{R}_{x_N}} \sum_{i=1}^{N} \log_2 \det \left(\mathbf{I} + \frac{\mathbf{H}_i \mathbf{R}_{x_i} \mathbf{H}_i^\dagger}{\sigma^2 \mathbf{I} + \mathbf{H}_i \mathbf{R}_{\tilde{x}_i} \mathbf{H}_i^\dagger} \right) \tag{5.44}$$

where $\mathbf{R}_{x_i}$ and $\mathbf{R}_{\tilde{x}_i}$ are the covariances of the intended and unintended signals to user i respectively. Earlier in this section, we found that the zero-forcing condition holds if and only if the beamforming coefficients for the ith user lie on the null space of the matrix $\tilde{\mathbf{H}}_i$. This means that the interference term $\mathbf{H}_i \mathbf{R}_{\tilde{x}_i} \mathbf{H}_i^\dagger$ is zero for all i, and the covariance of the desired signal can be written as

$$\mathbf{R}_{x_i} = \mathbf{W}_i \mathbf{R}'_{x_i} \mathbf{W}_i^\dagger \tag{5.45}$$

where $\mathbf{W}_i$ are the beamforming coefficients and $\mathbf{R}'_{x_i}$ is a normal matrix. The channel capacity is therefore given by

$$C = \sup_{\substack{\mathbf{R}'_{x_i}, \mathbf{W}_i \in \mathrm{nul}(\tilde{\mathbf{H}}_i) \\ i=1,\ldots,N}} \sum_{i=1}^{N} \log_2 \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{H}_i \mathbf{W}_i \mathbf{R}'_{x_i} \mathbf{W}_i^\dagger \mathbf{H}_i^\dagger \right). \tag{5.46}$$

Notice that each term in the summation is independent from the other terms, so the supremum can be taken from each term separately.

Using the same notation as was used earlier, let $\tilde{\mathbf{V}}_i^{(0)}$ be the orthogonal basis of the null space of $\tilde{\mathbf{H}}_i$ obtained via SVD. Since $\mathbf{W}_i$ must lie on this null space, it can be written as

$$\mathbf{W}_i = \tilde{\mathbf{V}}_i^{(0)} \mathbf{T}_i \tag{5.47}$$

for some matrix $\mathbf{T}_i$ denoting a change of basis. This allows us to write the capacity C_i for the ith user as

$$C_i = \sup_{\mathbf{T}_i, \mathbf{R}'_{x_i}} \log_2 \det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_i\tilde{\mathbf{V}}_i^{(0)}\mathbf{T}_i\mathbf{R}'_{x_i}\mathbf{T}_i^\dagger\left(\tilde{\mathbf{V}}_i^{(0)}\right)^\dagger\mathbf{H}_i^\dagger\right). \tag{5.48}$$

Now, we may notice that this capacity corresponds to a SU-MIMO channel with channel matrix $\mathbf{H}_i\tilde{\mathbf{V}}_i^{(0)}$ and transmitting covariance matrix $\mathbf{R}_{x_i}^{(0)} = \mathbf{T}_i\mathbf{R}_{x_i}\mathbf{T}_i^\dagger$. However, we already learned in Sections 4.2 and 4.3 that the optimal solution for this is given by eigen-beamforming. Moreover, this is exactly the solution block diagonalization uses, so we may conclude that the optimal precoding method subject to zero-forcing condition is given by block diagonalization.

Proposition 5.3. *The channel capacity of the general MU-MIMO subject to zero inter-user interference constraint is maximized by the block diagonalization method.*

5.3.4.2 Power Control

Previously we have discussed on maximizing the system throughput given a maximum power constraint. However, this may lead to a solution that is not fair among the receivers. Since the water-filling principle prefers high SNR streams, it may leave some streams with no allocated power. In general, this is not a desirable feature, e.g. in mobile networks. This motivates a power control problem which is to minimize power consumption while maintaining good enough quality of service at each receiver. In [36], it was suggested that block diagonalization could be used to solve the power control problem in MU-MIMO. In the following, we will study the suggested method in detail.

Suppose the minimum data rate requirements for each user is given by R_i for all $i = 1, \ldots, N$. Then, the power control problem can be formulated as

$$\begin{aligned} \underset{P_1,\ldots,P_N}{\text{minimize}} \quad & \sum_{i=1}^{N} P_i \\ \text{subject to} \quad & r_i \geq R_i \quad \forall i = 1, \ldots, N \\ & P_i \geq 0 \quad \forall i = 1, \ldots, N \end{aligned} \tag{5.49}$$

where P_i is the total power input for the ith user and r_i is the actual data rate to the ith user. In general, this problem is hard to solve. However, if we insist zero inter-user interference, then the optimal solution is given by block diagonalization with slightly modified power allocation algorithm.

While using the block diagonalization method, each SU-MIMO channel is independent, so we can minimize the full system power consumption by minimizing each user's power consumption independently. This gives a rise to the single-user power control problem

$$\begin{aligned} \underset{P_{i,1};\ldots;P_{i,K_i}}{\text{minimize}} \quad & \sum_{k=1}^{K_i} P_{i,k} \\ \text{subject to} \quad & \sum_{k=1}^{K_i} \log_2\left(1+\frac{\lambda_{i,k}^2 P_{i,k}}{\sigma^2}\right) \geq R_i \\ & P_{i,k} \geq 0 \quad \forall k = 1,\ldots,K_i \end{aligned} \tag{5.50}$$

where $P_{i,k}$ and $\lambda_{i,k}$ are the allocated power and the singular value corresponding to the kth stream of the ith user respectively, σ^2 is the noise variance, and R_i is the required data rate of the ith user. The actual capacity expression was obtained similarly as in Sections 4.2 and 4.3. We may notice the duality with the optimization problem (4.32) which led to the water-filling solution. Using Lagrange multipliers to (5.50) yields to the same water-filling power allocation

$$P_{i,k} = \left(\xi_i - \frac{\sigma^2}{\lambda_{i,k}^2}\right)^+ \tag{5.51}$$

as previously. We can solve ξ_i so that the data rate constraint is satisfied:

$$\begin{aligned} R_i = \sum_{k=1}^{K_i}\left(\log_2\left(\frac{\xi_i\lambda_{i,k}^2}{\sigma^2}\right)\right)^+ &= \sum_{k=1}^{S_i}\log_2\left(\frac{\xi_i\lambda_{i,k}^2}{\sigma^2}\right) \\ \iff \qquad \xi_i &= \sigma^2\left(\frac{2_i^R}{\prod_{k=1}^{S_i}\lambda_{i,k}^2}\right)^{1/S_i} \end{aligned} \tag{5.52}$$

where S_i is the number of streams with strictly positive power allocation. Since there are only finitely many values for S_i, namely, $S_i \in \{1,\ldots,K_i\}$, the optimal value for S_i respect to the power consumption can be easily found.

5.3.4.3 Partial Channel Knowledge

In the previous sections, we have learned that block diagonalization can be used for the MU-MIMO maximum throughput and power control applications. So far, we have assumed perfect CSI at the transmitter. Now, we relax

on this assumption and study the use of block diagonalization with imperfect channel knowledge. The method considered here was suggested in [36].

In practice, the transmitter cannot always attain perfect channel state information. In wireless communications, for example, the transmitter may gain knowledge on the directions at which the receivers are located. This will lead to an imperfect CSI which can be used for block diagonalization of the unknown channel matrix.

Let us consider a SU-MIMO channel. Suppose that there exists L distinct multi-paths[10] from the transmitter to the receiver. If we fix a reference phase for the transmitter and the receiver antenna arrays, then the phase at one single transmitting or receiving antenna is given by the direction of the signal. A detailed discussion of this can be found from Appendix A.1. Using this knowledge, the channel feedback of one pair of antennas, i.e. one element of the channel matrix $\mathbf{H} = (h_{ij})$, can be written in the form

$$h_{ij} = \sum_{k=1}^{L} \alpha_{ikj}\gamma_{ikj}\beta_{ikj} \tag{5.53}$$

where α_{ikj} is the relative phase of the received signal of at the ith receiving antenna from the jth transmitting antenna and kth multi-path, γ_{ikj} is the path loss of that path, and β_{ikj} is the relative phase at the transmitter. If we assume that the transmitter and the receiver are far enough from each other, then the phase at the receiver does not depend on the transmitting antenna, so $\alpha_{ikj} = \alpha_{ik}$, and the transmitted phase does not depend on the receiving antenna, so $\beta_{ikj} = \beta_{kj}$. Moreover, the path loss is approximately the same for each pair of antennas for one single multi-path, so we may say that $\gamma_{ikj} = \gamma_k$. This allows us to write the channel matrix as

$$\mathbf{H} = \mathbf{A}' \operatorname{diag}(\gamma_1, \ldots, \gamma_L)\mathbf{B} = \mathbf{AB} \tag{5.54}$$

where $\mathbf{A}' = (\alpha_{ij})$ and $\mathbf{B} = (\beta_{ij})$ hold the information on the receiver and transmitter steering vectors respectively, and $\operatorname{diag}(\gamma_1, \ldots, \gamma_L)$ tells the path loss of each multi-path signal. This model is rather idealistic since it neglects all the random factors that affect the channel matrix. However, for the continuation, we only need the knowledge on $\mathbf{B}$. Therefore, it is sufficient to assume that the random information is present at $\mathbf{A}$.

Suppose that the transmitter has M transmitting antennas and there exists N independent users with K_i receiving antennas for all $i = 1, \ldots, N$. We assume that the channel of the ith user is given by

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i + \mathbf{n}_i = \mathbf{A}_i\mathbf{B}_i\mathbf{x}_i + \mathbf{n}_i \tag{5.55}$$

[10] A multi-path means a signal propagation paths from the transmitter to the receiver.

where $\mathbf{A}_i \in \mathbb{C}^{K_i \times L}$ and $\mathbf{B}_i \in \mathbb{C}^{L \times M}$ are unknown and known to the transmitter respectively. We will now continue by using a similar approach that was used in Section 5.3.4.1. Suppose that the transmitted signal $\mathbf{x}_i$ is precoded as $\mathbf{x}_i = \mathbf{W}_i \mathbf{s}_i$. We impose the zero-forcing condition $\mathbf{H}_j \mathbf{W}_i = \mathbf{0}$ to hold for all $j \neq i$. Notice that this condition holds if we choose $\mathbf{W}_i$ such that $\mathbf{B}_j \mathbf{W}_i = \mathbf{0}$ for all $j \neq i$. Therefore, we may use the block diagonalization method for the matrix $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1^\top & \mathbf{B}_2^\top & \dots & \mathbf{B}_N^\top \end{bmatrix}^\top$.

Define $\tilde{\mathbf{B}}_i = \begin{bmatrix} \mathbf{B}_1^\top & \dots & \mathbf{B}_{i-1}^\top & \mathbf{B}_{i+1}^\top & \dots & \mathbf{B}_N^\top \end{bmatrix}^\top$ and let the singular value decomposition of $\tilde{\mathbf{B}}_i$ be written as

$$\tilde{\mathbf{B}}_i = \tilde{\mathbf{U}}_{B_i} \begin{bmatrix} \tilde{\mathbf{\Sigma}}_{B_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_{B_i}^{(1)} & \tilde{\mathbf{V}}_{B_i}^{(0)} \end{bmatrix}^\dagger \tag{5.56}$$

where $\tilde{\mathbf{V}}_{B_i}^{(0)}$ is the orthogonal basis of the null space of $\tilde{\mathbf{B}}_i$. We choose $\mathbf{W}_i = \tilde{\mathbf{V}}_{B_i}^{(0)} \mathbf{W}_i'$, which yields

$$\mathbf{H}_j \mathbf{W}_i = \mathbf{A}_j \mathbf{B}_j \tilde{\mathbf{V}}_{B_i}^{(0)} \mathbf{W}_i' = \mathbf{0} \tag{5.57}$$

for all $j \neq i$. Equation (5.57) implies that the zero-forcing condition is satisfied even if $\mathbf{A}$ was assumed to be unknown to the transmitter. It was shown in [36] that at high SNR, $\mathbf{A}$ does not affect the system capacity, and capacity-wise the optimal choice for $\mathbf{W}_i'$ is the right singular vectors of $\mathbf{B}_i \tilde{\mathbf{V}}_{B_i}^{(0)}$ corresponding to the strictly positive singular values. In the same paper it is stated that at high SNR the optimal power allocation is to distribute the power evenly among each spatial stream.

5.4 Advanced Beamforming Solutions

We conclude this chapter with a short literature review of further results in the MU-MIMO beamforming. We have presented here four beamforming methods that are based on the zero inter-user interference constraint. These methods have proved to be a good platform for deriving more advanced beamforming solutions that approach the sum capacity promised by dirty paper coding. Much emphasis has been given to user selection algorithms, but also new beamforming methods have been developed. We will first discuss these new methods and then continue with the solutions involving user selection.

In [47], a beamforming method based on maximizing signal-to-leakage ratios was introduced. This method allows some inter-user interference, and the optimization problem is solved via the generalized eigenvalue problem. Vector perturbation has been the subject of multiple recent studies including

[48], [49], and [50]. In [48], the Wiener filter[11] vector perturbation method is introduced. This method is proved to be optimal in the MSE sense, and capacity-wise the performance is similar to the method discussed in Section 5.3.3. Complexity issues of the vector perturbation method are addressed in [49] where the max-SINR vector perturbation method is developed. The method is based on maximizing a SINR approximation derived in the same paper. This approach proves to have better performance compared to the regular and the Wiener filter vector perturbation methods with lower complexity. Finally, extension to the general MU-MIMO is given in [50] where block diagonalization and vector perturbation are combined.

Assuming high number of users, a user selection (or scheduling) algorithm can be used. In [52], semi-orthogonal user selection algorithm is introduced. It is proved that channel inversion combined with this user selection achieves the optimal asymptotic channel capacity. This algorithm is improved by volume-based user selection [53], which reduces the computational complexity and increases the capacity compared to the semi-orthogonal user selection algorithm.

The optimal asymptotic channel capacity can also be achieved by using capacity-based user selection algorithms [54], [55], and [56]. The idea of the zero-forcing with user selection algorithm [54] is to maximize the channel capacity subject to the water-filling power allocation. Since the optimal solution to this problem requires an exhaustive search, a greedy iterative algorithm is used instead. Sequential water-filling algorithm [55] improves the performance of the previous method by removing the users with no allocated power. Further improvement is given by the greedy user selection with swap algorithm [56] that defines an additional swap operation that helps to prevent from the cases where the selection algorithm gets stuck on a poor locally optimum user selection. In [56], it is claimed that this algorithm achieves 99.3% of the optimum capacity with only little added complexity.

Conclusion

In this chapter, we first stated that dirty paper coding and interference alignment achieve the optimal channel capacity in channels subject to interference. However, these methods are not practical, so sub-optimal solutions obtained

[11] Wiener filter [51], [7, pp. 233-256]: A filter designed to minimize MSE between an observed signal and a target signal.

via beamforming were discussed. We derived the general formula for the channel capacity using Shannon's theorem.

MU-MIMO beamforming discussion was started by considering eigenbeamforming in the multi-user context. However, we found out that inter-user interference may decrease the system performance, and recognized the importance of a scheduler. We then restricted the study to beamforming methods that achieve the zero inter-user interference constraint. For single antenna users, the plain channel inversion was shown to satisfy this constraint, however, its performance turned out to be rather poor in certain scenarios. This method was then optimized with inversion regularization and vector perturbation methods.

Finally, we considered the block diagonalization method that generalizes channel inversion to the case where users may have multiple receiving antennas. This method was shown to be the optimal precoding method subject to the zero inter-user interference constraint. In addition to the maximum throughput use-case, we discussed the power control problem and partial channel state information beamforming while using block diagonalization. We concluded the chapter with a literature review of more advanced MU-MIMO beamforming solutions.

Chapter 6

Beamforming Simulations

In this chapter, we will study beamforming by using simulations. The initial goal was to visualize beamforming by using a geometry-based channel model. However, this non-standard model proved to be very versatile which led to various other results in addition to those desired visualizations. A few of these side product results are presented in this chapter, although, not all of them could be fit here.

This chapter is divided into three sections. First, in Section 6.1, the geometry-based channel model used in the simulations is described. Then, in Sections 6.2 and 6.3 the simulation results are presented. In these sections eigen-beamforming and zero-forcing beamforming are considered respectively. In addition to beamforming visualizations, we will study the benefits of large antenna arrays, how channel aging may decrease the performance in the real world applications, and the effects of extra degrees of freedom while using channel inversion.

6.1 Channel Model

In this section, the channel model used in the simulations is discussed. Since the initial target of the simulations was to visualize beamforming techniques, the channel model selection was in a critical part of achieving successful results. The main limitation in choosing the model was to balance the model complexity and the desired simulation accuracy.

In order to visualize the results, a channel model related to some physical environment was selected. *Ray tracing* model—described for example in [57]—seems to be a perfect fit for this purpose, however, its complexity is too high from the perspective of this thesis. Instead, a *geometry-based* channel model is used. This model is motivated by the double-directional

radio channel model described in [58]. In [59], a SISO model based on the double-directional model was used. Due to its simplicity and geometric interpretation, this model was chosen to be the basis of the model used in the simulations. However, this model cannot be used as it is, since it is intended for SISO channels, whereas we need a MIMO channel model.

6.1.1 Geometry-Based SISO Channel Model

The geometry-based SISO channel model is based on the idea of multi-path radio propagation. Suppose we have a base station and a user that transmit signals to each other. Then, suppose that they are in an environment of objects that can scatter radio signals to all possible directions. Assuming that the transmitting antenna is perfectly isotropic, the received signal is the superposition of all the signals arriving from different propagation paths, i.e. multi-paths. Let us assume that at most one scattering is possible for each of these paths. Then, the propagation paths from the transmitter to the receiver are given by the paths from the transmitter via scatterer objects to the receiver. The line-of-sight (LOS) signal can be included if desired. Example visualization of those propagation paths is given in Figure 6.1a.

Now, that we know how the propagation paths are defined, let us consider the computation of the channel response. For this we need to define what are the phase and the amplitude changes for each multi-path. We assume that there is no phase shift during scattering, and therefore, the total phase shift during one propagation path is fully determined by the propagation path length and the wavelength of the signal. For a zero-phase transmitted signal, the phase shift $\Delta\varphi$ is given by

$$\Delta\varphi = \exp\left(-\frac{d}{\lambda} \cdot 2\pi\mathrm{i}\right) \tag{6.1}$$

where d is the propagation path length, λ is the signal wavelength, and i is the imaginary unit. Amplitude changes are modeled via log-distance path loss model. This is motivated by the free space propagation path loss model [60] which states that the path loss L_P in the free space propagation is given by

$$L_P = 20\log_{10}(d) + 20\log_{10}(f) - 27.55 \tag{6.2}$$

where the path loss is measured in decibels (dB), d is the propagation path length in meters (m), and f is the radio frequency in megahertz (MHz).

The channel response h_k for a zero-phase unit-signal arriving from the kth multi-path is then given by

$$h_k = 10^{-L_k/20}\exp\left(-\frac{d_k}{\lambda} \cdot 2\pi\mathrm{i}\right) \tag{6.3}$$

where d_k and L_k are the length and the path loss of the kth propagation path respectively. Since the receiver observes the superposition of all the multi-path components, the total channel response h, is then written as

$$h = \sum_k 10^{-L_k/20} \exp\left(-\frac{d_k}{\lambda} \cdot 2\pi \mathrm{i}\right) \tag{6.4}$$

where the sum is taken over all the propagation paths.

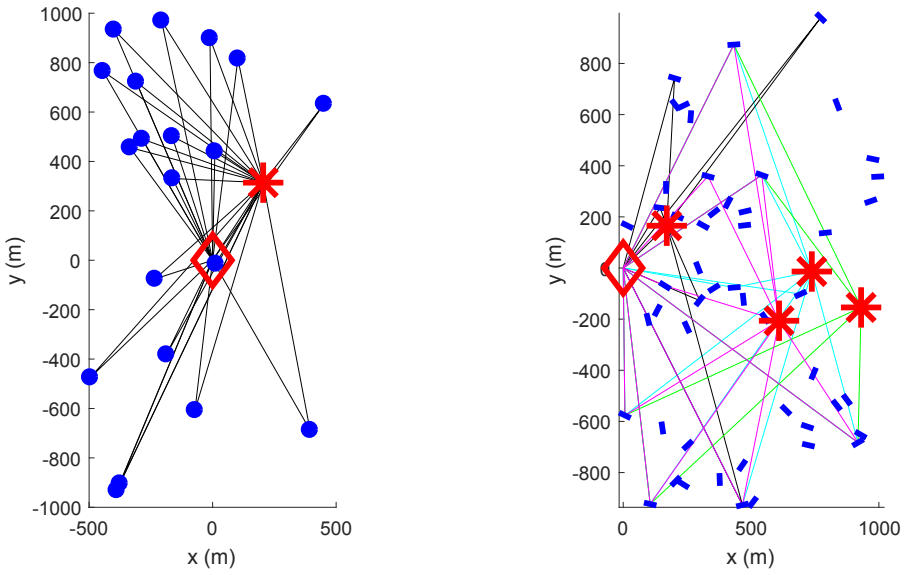
The geometry-based SISO channel model can be viewed as extremely low complexity approximation of the ray tracing channel model. It may not be the most accurate representation of the real world radio channels, however, in [59] it was shown that fading follows widely accepted Rician and Rayleigh fading models depending on the number of scatterers, and moreover, the Doppler effect is modeled implicitly for moving users. Due to these results, we claim that this model is sufficiently accurate for the scope of this thesis. More importantly, the geometry-based channel model gives an excellent platform for visualizing radio propagation.

6.1.2 MIMO Extension

In the previous section, we defined a geometry-based SISO channel model that allows lightweight visualization of radio signal propagation. In this section, we will describe, how this SISO model can be extended to support MIMO channels. As a result, we obtain a geometry-based MIMO channel model that is used for the simulations.

The idea of the geometry-based SISO channel model is very simple—predefine propagation paths, and compute the channel response based on the path distances. This same idea can be used in MIMO context, however, propagation paths cannot be defined via single point on a map. The reason for this is that one of the key channel properties, that is, the relative phase differences at different antennas, is lost. This can be noticed by considering a situation in which the base station has multiple transmitting antennas and a user has only one receiving antenna. Then, if all the signals from those transmitting antennas travel through a same point, the phase difference between these signals is fully determined by the relative positions of transmitting antennas and the scattering point. Therefore, user location has no effect on the relative phase differences of the signals.

To solve the aforementioned problem, scatterers are replaced by reflectors. Instead of being points, reflectors can be viewed as small reflecting mirrors. This changes the propagation paths to be defined by surface reflections from the reflectors. Changing scatterers to reflectors has multiple benefits as well



(a) Geometry-based SISO channel model

(b) MIMO extension

Figure 6.1: Geometry-based channel model propagation paths. Red diamond is the base station, red asterisks are the users, blue dots/rectangles are the scatterers/reflectors, and magenta, cyan, green, and black lines are the propagation paths to different users.

as some drawbacks. The first benefit is that the relative phase differences of signals arriving from different antennas is no more fully determined by the locations of transmitting antennas and a single point, but rather by the position of the user relative to a reflector and the base station. The second benefit is that reflectors are not zero-dimensional objects which enables intuitive LOS modeling. Finally, the third benefit is that reflecting mirrors give arguably more realistic environment than scatterers that are able to perfectly scatter signals to any given direction.

On the downside, the main drawback is added complexity. To maintain the geometric intuition, reflectors need to have a surface normal. Therefore, only those reflections that nearly follow the law of reflection should be allowed. Also the 180 degree phase shift during a reflection [61] should be considered.

Example of propagation paths in this channel model can be found in Figure 6.1b. Here, we have allowed slight deviation from the law of reflection by allowing all the reflections that have at least 0.9 correspondence to the

direction of the true reflection. This allows using less reflectors while still having usually multiple different propagation paths to a given position. In the simulations the correlation factor was used to scale the received signal amplitude for that path. This means that the path loss is slightly increased if the reflection does no follow the law of reflection.

Finally, before discussing the simulation results, we specify the used path loss model. For all propagation paths, whether they are LOS or via a reflection, we use an urban macro environment path loss model defined in [62]. This model defines the path loss as

$$L = 37.6 \log_{10}(d) + 15.3 \tag{6.5}$$

where L is the path loss in decibels, and d is the propagation path length in meters[1]. This model assumes carrier frequency of 2.0 GHz, so this frequency is used in the simulations.

6.2 Eigen-Beamforming Results

Eigen-beamforming defines a way to control an antenna array to achieve the optimal performance with one receiving user. By simulations, we study how much gain can be obtained by using eigen-beamforming. This question can be approached from many directions, but here we will consider two of them. First, we will compare the performance of an eigen-beamformed antenna array and a single omni-directional antenna. Second, we change the reference system to be an array that has no signal preprocessing. Then, after these two considerations we will visualize eigen-beamforming in both LOS and non-LOS cases. Finally, these visualization are used to discuss channel aging.

Before we step into the results, a remark needs to be made. To allow more easy visualization, all the simulations are run with 2D environments even if the geometry-based model is capable of producing 3D environments. By a 2D environment we mean that all the reflectors and users as well as the base station lie on the same plane. However, the antenna array is perpendicular to this plane. In the simulations we use $n \times n$ square antenna arrays, but due to this flat environment, effectively roughly $1 \times n$ horizontal array is used for beamforming since almost no vertical separation needs to be made. This must to be taken into account when considering the results presented here.

[1]Remark: In the 3rd Generation Partnership Project model the distance refers to the distance between the transmitter and the receiver. Since here were focusing on multi-path propagation, we use the full propagation path distance.

6.2.1 Benefit of Large Antenna Array

Let us now consider beamforming gain in the case where the reference system is an omni-directional antenna. The question is, how much gain is obtained by using an antenna array of M antenna elements combined with eigen-beamforming compared to this single antenna configuration. We assume that the receiver has only one receiving antenna, and only the received signal powers are compared[2]. Comparing the received powers is motivated by the fact that the single stream channel capacity is maximized by maximizing the received power as was learned in Section 4.2.1. We also assume that the total transmitted power is the same for both antenna configurations. This gives a fair comparison in the sense that both systems use the same amount of energy.

Results for beamforming versus an omni-directional antenna can be found in Figure 6.2. In this simulation, beamforming was performed with a 4×4 square antenna array (i.e. $M = 16$) where each antenna element was assumed perfectly isotropic. The result indicates that beamforming performance would be always better compared to the omni-directional antenna. This is perfectly expected result since beamforming is the optimal way to control a given antenna array for a single user. Therefore, the signal power at the receiver must be at least as good as could be obtained by using any subset of antennas from the array, particularly, the performance is always at least as good as one antenna could achieve on its own. Since the system powers were assumed to be the same in both cases, the beamformed antenna array never[3] performs worse than a single omni-directional antenna.

If the channel was a plain LOS channel with no multi-path propagation, the beamforming gain would be exactly the number of transmitting antennas M. In around 45% of the cases this gain is not achieved. The main reason for this is that the channel depicts a flat environment. Therefore, only 1×4 antenna array is in use. We can see that gain of 4 is achieved roughly in 99% of the cases. In the rare occasions where this gain is not achieved occur since sometimes the propagation paths are such that no perfectly coherent signal summation can be achieved at the receiver. On the other hand, in many cases the gain is much higher than the number of antennas. This is caused by such channel configurations that the signal transmitted from the omni-

[2]If we assume that the noise level is the same for both systems, then this comparison is equivalent to comparing SNRs.

[3]This is true in theory, however in practice, in some very rare cases beamforming can perform worse because none of the antenna array elements may not be exactly in the same position than where the omni-directional antenna is. Therefore, the channel may be slightly better for the single antenna.

directional antenna is summed destructively at the receiver. Since the large antenna array has more degrees-of-freedom, it can prevent the destructive summation.

Previously we assumed that both systems have the same power consumption. While this assumption leads to very large gains in favor of beamforming, this assumption may not always be applicable. In the real world application there are emission limits given by different regulations that determine the maximum power at transmission per volume unit, e.g. [63]. Therefore, using beamforming may lead to violation of these emission limits. If we assume that both the system with beamforming and the single antenna system transmit at the maximum allowed power, then the one with beamforming consumes less energy. This is because the input power is concentrated optimally in the target direction and less power is wasted into undesired directions. Using less energy means for example that battery life of smart phones is increased.

So far, we have discussed the received signal power that is obtained by using an eigen-beamformed antenna array compared to an omni-directional antenna. However, this discussion does not consider all the benefits that are obtained by using multiple transmitting antennas. We assumed previously that the receiving user has only one receiving antenna. Nevertheless, the modern mobile phones have multiple antennas (e.g. [64]), so we can also consider a case where the receiver has multiple receiving antennas. In this case, we have learned in Chapter 4 that multiple simultaneous data streams can be sent. Considering this additional benefit of large antenna arrays, we can conclude that it is rarely advantageous to use only one transmitting antenna even if it has a large power input.

The next result demonstrates that using beamforming is necessary for large antenna arrays if a good coverage want to be achieved. Suppose we have an antenna array of M antennas. Then, the beam-pattern of this array is determined by its geometry, but the beam width depends mostly on aperture of the array[4]. In order to get the best performance from an antenna array, the aperture must increase as the number of antennas increase. This, on the other hand, means that beam width decreases as the number of antennas increase.

Figure 6.3 shows the comparison of a 4×4 element antenna array with eigen-beamforming to the same array with no antenna array processing. Figure shows an environment that is filled with reflectors, and for each position (x, y, z) the received SNR with and without beamforming is computed and compared (a single antenna user is assumed)[5]. We can see that there is little

[4]Physical dimension of an antenna array is called aperture.

[5]For illustration purposes, in all simulations we have set $z = 0$.

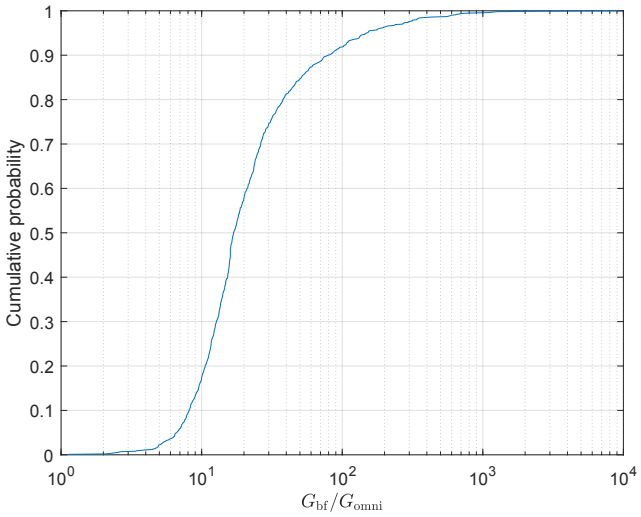


Figure 6.2: Cumulative distributive function of the beamformed signal strength (G_{bf}) over the strength of a signal coming from an omni-directional antenna (G_{omni})

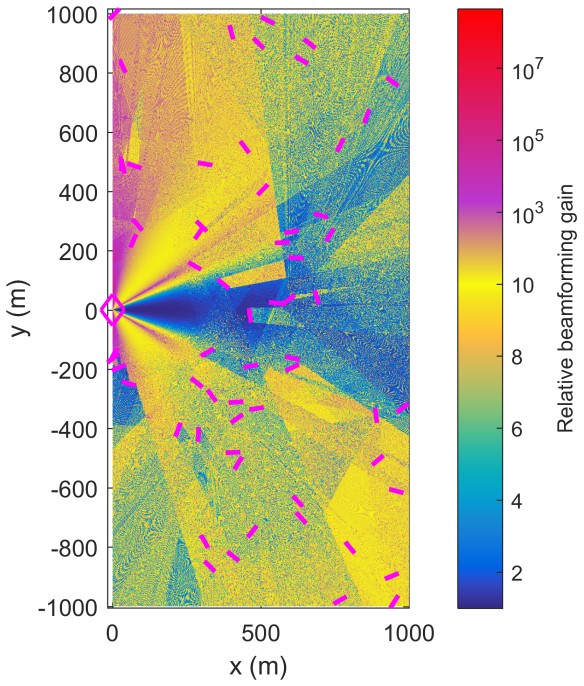


Figure 6.3: Gain of beamformed antenna array compared to one with no antenna array processing. Magenta diamond is the base station and magenta lines are the reflectors. Notice that after the relative beamforming gain of 10, the color scale is transformed from linear to logarithmic.

to no gain close to x-axis, since the non-beamformed array has the largest gain in that direction. Apart from this direction, beamforming achieves substantially better SNR than the array with no antenna array processing. The largest gains are obtained in the directions in which the non-beamformed array has the most destructive interference. Result of Figure 6.3 is not surprising, but it nicely visualizes the benefit of using beamforming.

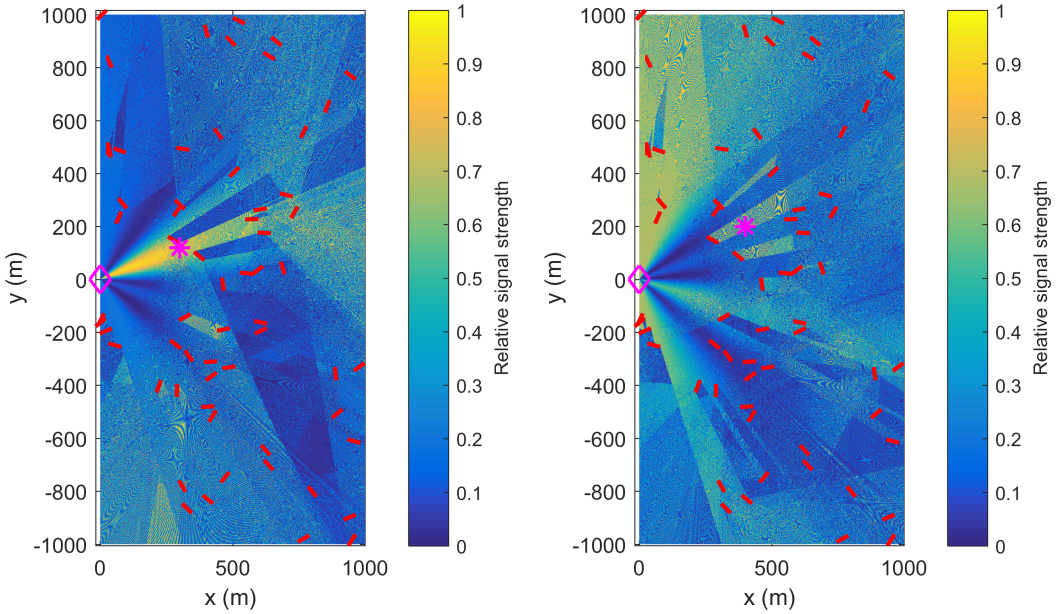
6.2.2 Eigen-Beamforming Visualization and Channel Aging

In Figures 6.4a and 6.4b, eigen-beamforming is visualized in LOS and non-LOS scenarios. In these figures, the relative signal strength across the environment is plotted when the source is an eigen-beamformed signal intended to one user. By the relative signal strength we here mean the comparison of the maximum possible signal power (i.e. signal power obtained by using eigen-beamforming to the given position) and the power of the signal leaked from the transmission to one user. There are two equivalent interpretations for these results. First, we think that relative interference experienced by other user is plotted when we fix the position of the user the signal is intended to. Another way to look these results is from the channel aging perspective. The channel is measured at the given position and the beamforming coefficients are based on that measurement, but the user is no longer in that position when the signal is received. We will now discuss both of these interpretations.

Consider first the interference interpretation. Using this perspective, we can study what would happen if eigen-beamforming is used in MU-MIMO. In both Figures 6.4a and 6.4b, we can see that interference is noticeable in most positions in the environment. This confirms the point made in Section 5.2 that in MU-MIMO scenarios there may be too much inter-user interference while using eigen-beamforming. Therefore, a scheduler should be used to careful choose the set of user that receive simultaneous transmissions in order to maintain good quality of service.

Now, let us consider the channel aging perspective. The problem of channel aging arises naturally with moving users. In section 4.2.3, we found that channel is measured before this information is used in transmission, so if the user moves, then the channel matrix is not perfectly correct when the beamforming coefficients are calculated. Fortunately, the delay is rather short, so the user can be expected to be only slightly out of position at reception compared to the original position the channel was measured in.

While the resolution of Figures 6.4a and 6.4b may not be optimal for this study, we can still see the general behavior. In Figure 6.4a, we can see that if



(a) Relative interference created by an eigen-beamformed signal intended to one user. Magenta diamond is the base station, magenta asterisk is the user, and red lines are the reflectors.

(b) The same channel geometry was used here, but the user was moved behind a reflector.

Figure 6.4: Eigen-beamforming visualization

the user has LOS and there are no or only few significant other propagation paths, then the beamforming coefficients are not very sensitive for channel aging. On the other hand, from Figure 6.4b, we can see that if there are many multi-path components, or there is no LOS, the beamforming coefficients are not anymore as robust against channel aging. In other words, these results mean that in LOS cases the performance is not much impacted even if an aged channel matrix is used for computing the beamforming coefficients, but the same does not hold for non-LOS cases.

It must be remembered that these results should not be taken as ground truth, since the channel model does not perfectly depict real world phenomena. Simulations have indicated that in LOS cases the direct signal path is so dominant that other multi-paths are insignificant. On the other hand, in non-LOS cases there usually seems to be several significant multi-path components. However, it should be noticed that these observations may be caused by the channel model. Nevertheless, the results suggest that in cases with multiple significant propagation paths, the system performance may be worse than expected due to channel aging. Field measurements are needed

to verify these indications. It is left to study whether using slightly aged beamforming coefficients should be replaced by another pre-coding method that is either more robust against channel aging or is able to predict channel changes.

6.3 Zero-Forcing Beamforming Results

In this section, the zero-forcing beamforming results are presented. First, we will visualize channel inversion similarly as in the previous section with eigen-beamforming. Then, we will consider channel aging and show it may cause severe problems in real world applications. Finally, we study how many transmitting antennas are needed to support a given number of users while using channel inversion. In Section 5.3.2, we proved that the channel inversion performance is very poor if the number of transmitting antennas is the same as the number of users. However, here we study, how much extra degrees of freedom are needed.

6.3.1 Channel Inversion Visualization and Channel Aging

Channel inversion visualization is presented in Figure 6.5. The simulation setup is as follows. There is a base station with a 4×4 square antenna array, and 4 users that all have one receiving antenna. Plain channel inversion beamforming is performed. In each sub-figure titled UE i, we have plotted relative signal strength intended to the ith user (magenta asterisk) across the environment in a similar fashion as with eigen-beamforming in Figure 6.4. In theory, we should observe zero relative strength at other users (cyan asterisks) and non-zero relative signal strength at the intended user. In the sub-figures, we can see that signals not intended to user 2, do not interfere with the intended signal at user 2 location. However, it is hard to visually verify if this applies for other users as well, since there is rapid variation of relative signal strengths near other users' locations. Nevertheless, if the computations are performed exactly at the locations of the users, then the relative strengths of signals not intended to the user in question are indeed zero as should be. Having environment with only LOS, that is, with no reflectors, would allow more clear illustration that the relative signal strengths are zero at undesired users. This visualization is left to Appendix A.2.

In contrast to the eigen-beamforming results, we may notice that the relative signal strengths at the intended users are not the best possible in this zero-forcing case. This is the result of the trade-off between interference

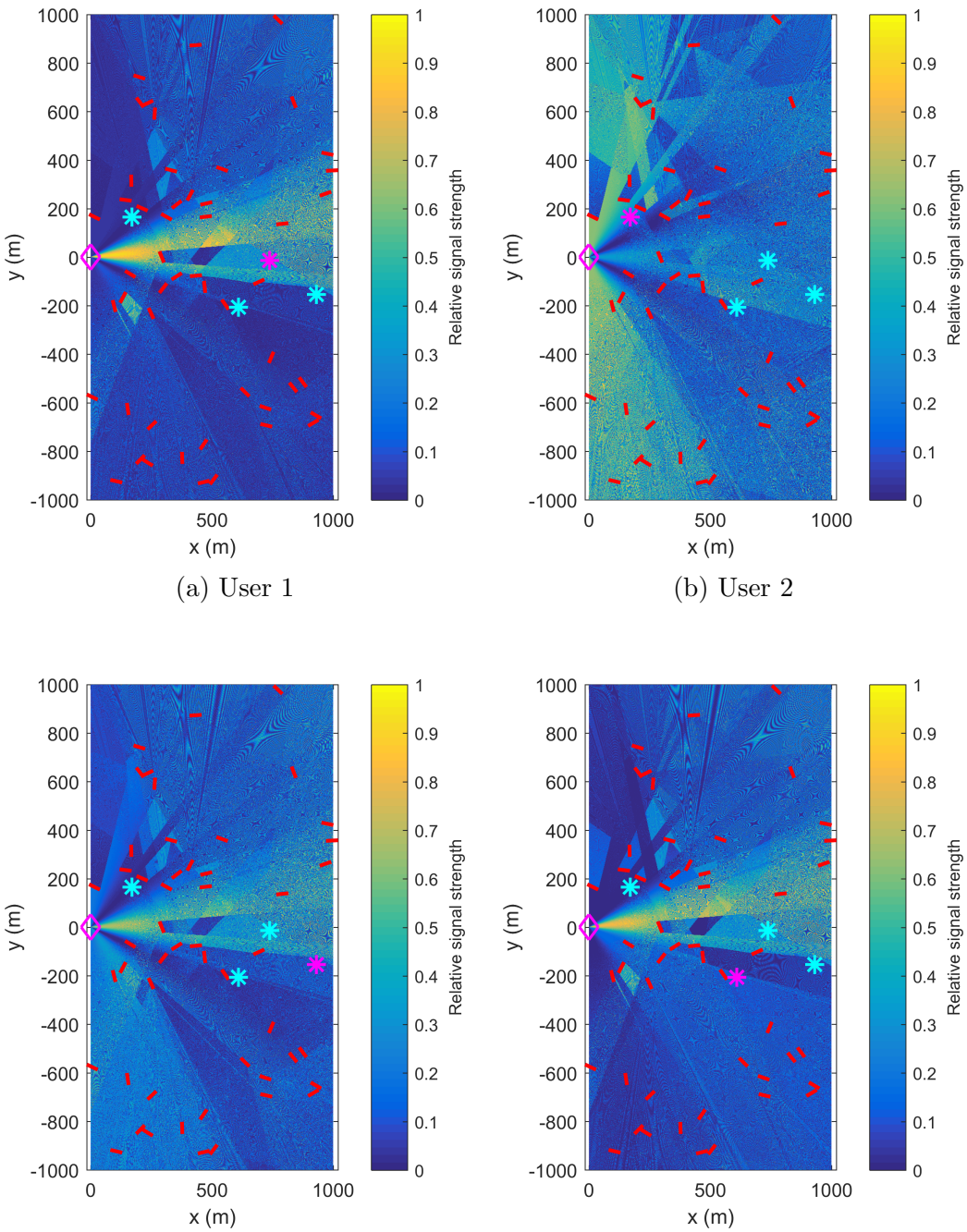
and gain. Since in this case the interference is forced to be zero for each user, the gain cannot be controlled. Some control on gains could be achieved by using block diagonalization instead of plain channel inversion. It turns out that with the same exact channel and user positions, block diagonalization with maximum throughput water-filling leaves user 3 out of service, thus reducing interference cancellation from 4 users to 3. This suggests that even if some zero-forcing method is used for beamforming, a user scheduler should be used to select reasonable sets of users to optimize system performance.

Let us now study the effect of channel aging when using plain channel inversion. In Figure 6.6, the same channel with the same user positions is used as previously, but now the view is zoomed to one user at a time. The plots describe the achievable relative capacity to the user in question.

More precisely, let $\mathbf{x}_0^{(i)} = (x_0^{(i)}, y_0^{(i)}, z_0^{(i)})$ with $z_i^{(0)} = 0$ be the original position of the user i in study. Then, for each position $\mathbf{x}^{(i)} = (x^{(i)}, y^{(i)}, z^{(i)})$ with $z^{(i)} = 0$, the capacity $C(\mathbf{x}^{(i)})$ to the user i is computed. This capacity calculation is done with the beamforming coefficients that are computed from the channel matrix that is based on the channel estimation for the original user position $\mathbf{x}_0^{(i)}$. Obtained capacity is then compared to the capacity $C(\mathbf{x}_0^{(i)})$ achievable if the user was still in the original position, that is, the plotted value is given by $C(\mathbf{x}^{(i)})/C(\mathbf{x}_0^{(i)})$. Only one user is moved at a time, and the other 3 users are in their original positions $\mathbf{x}_0^{(j)}$ with $j \neq i$. The total transmit power was normalized to be 1. If this normalization is changed, then the color scale would also change due to logarithmic relation between power and capacity.

In Figure 6.6, we can see that the capacity highly depends on the user position, i.e. channel aging has high impact on capacity. There are two main reason to explain these results. The first reason is that the intended signal may become weaker if the user moves, and the other reason is that the interference may become too large. It is rather surprising that even movement of few centimeters may cause the user capacity to drop significantly. This is extremely undesirable in real world applications. However, this rises a question whether the used channel model is accurate enough for channel aging study.

Based on various other simulation cases, the behavior observed in Figure 6.6 occurs when there are multiple signal propagation paths of similar path loss. This is similar to the non-LOS case with eigen-beamforming that was found to be sensitive to channel aging. Field testing is needed to find if this is the case in reality, or if it is just a feature of this channel model.



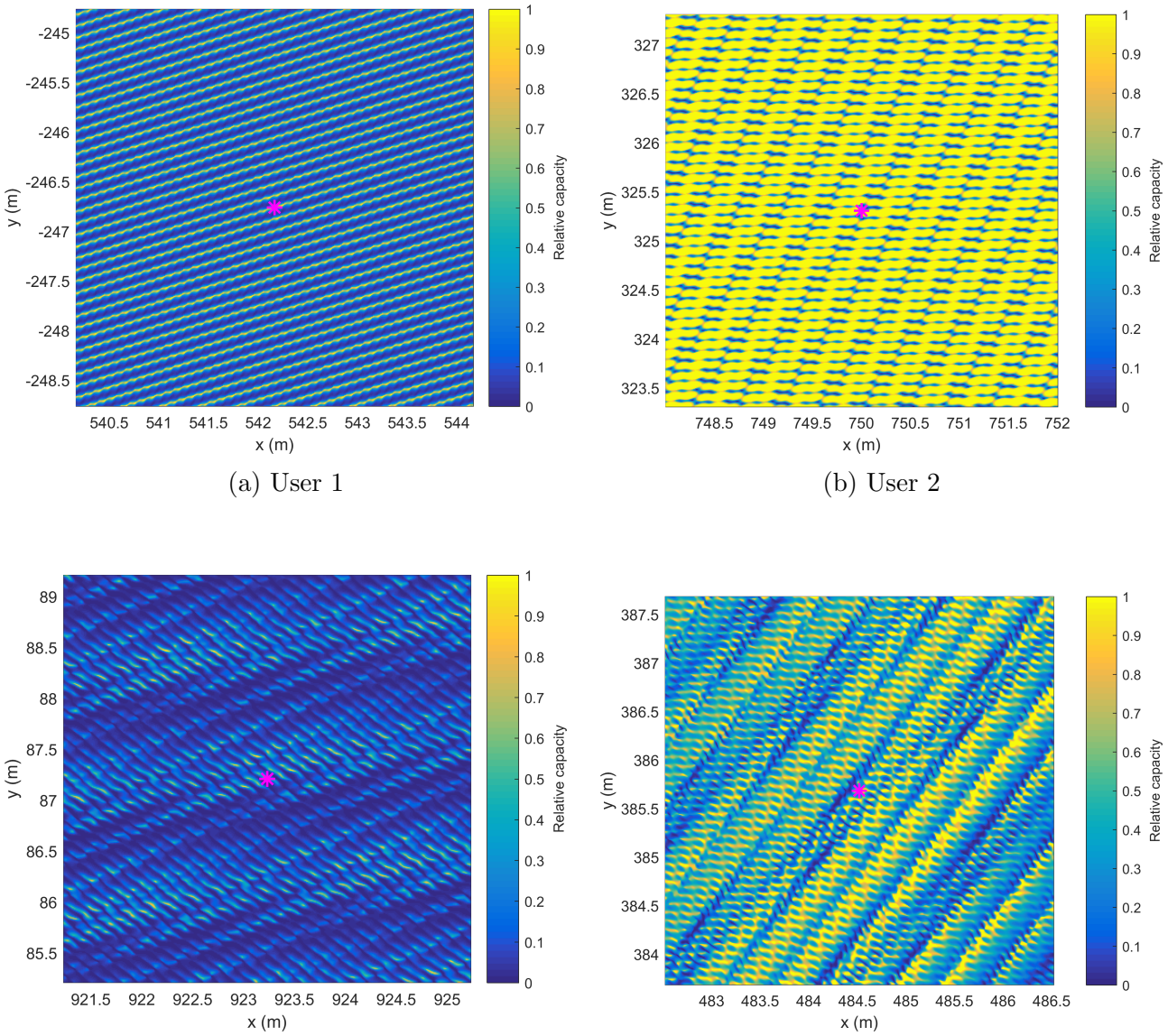
(a) User 1

(b) User 2

(c) User 3

(d) User 4

Figure 6.5: Interference patterns created by signals intended to one user subject to zero-forcing beamforming. Magenta diamond is the base station, magenta asterisk is the intended user, cyan asterisks are the other users, and red lines are the reflectors.



(a) User 1

(b) User 2

(c) User 3

(d) User 4

Figure 6.6: Relative channel capacity when the channel inversion beamforming coefficients are computed based on the user position of the magenta asterisk, but the user actually is in some other position. The channel and the users are the same as in Figure 6.5.

6.3.2 Effect of Extra Degrees of Freedom

In Chapter 5, we found that channel inversion performs poorly if the number of users is the same as the number of transmitting antennas. However, we did not discuss how much extra degrees of freedom are needed if all the users should have good relative signal strength. By studying LOS channels, we found that only a few extra degrees of freedom does not suffice. This study is left to Appendix A.2.

Figure 6.7 shows the relative signal strength of the worst users as a function of the number of transmitting antennas. In the simulations, the number of users was set to be 4 and their locations were selected randomly in a plain LOS channel. Mean performance was used to plot the results. Only square antenna arrays were used at the transmitter. Again, it must be again noted that the flat channel causes these square antenna arrays be effectively roughly $1 \times n$ antenna arrays, which should be considering when interpreting the results.

From this result, we can see that the relative signal strength at the worst user with a 3×3 transmitting antenna array is very close to 0. This is an expected result, since effectively there are less transmitting antennas than users. If we want to have relative signal strength of at least 0.5 at each user, we need to use effectively 12 to 13 transmitting antennas meaning three times the number of users. This result can be explained by the fact that in some cases two users are placed nearly on the same LOS direction, and therefore, were narrow beams (meaning large antenna array) need to be used if good relative signal strengths want to be obtained at both of these users. This result is another good indication that if channel inversion is used, then a user scheduler must be applied to prevent highly correlated user channels.

Since the flat channel biases the result, we should not pay too much attention to the exact numbers, but rather to the general trend. The key information gained form this result is that a large number of extra degrees of freedom are needed if we want to have high relative signal strength at each user when using channel inversion method. It should be remembered that this result was obtained by using a plain LOS channel, so the results may be different for more complicated environments.

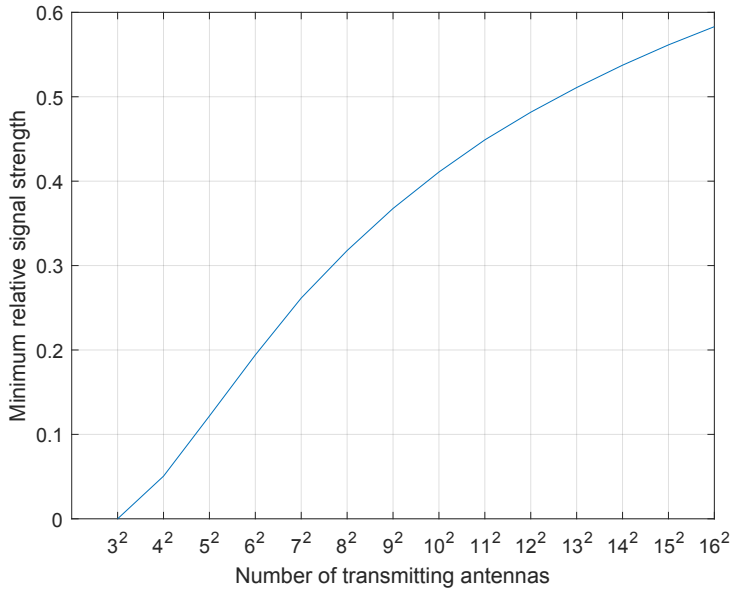


Figure 6.7: Relative signal strength at the worst of 4 users plotted against the number of transmitting antennas. A square antenna array of given size was used, and the channel was set to have no reflectors, i.e. the channel is plain LOS.

Chapter 7

Conclusion

In this thesis, we reviewed multiple beamforming methods both for SU-MIMO and MU-MIMO. Derivations for these methods were given by solving optimization problems arising from Shannon's Theorem of channel capacity. The main tools for solving these optimization problems were linear algebraic methods, especially, SVD was used intensively. Theoretic results were supported by simulations that visualized beamforming, but also initialized discussion on possible challenges in the real world applications.

For SU-MIMO we derived the well-known eigen-beamforming technique. We showed that this method achieves the maximum capacity given by Shannon's theorem. We learned that further optimization is obtained by using the water-filling principle that finds the optimal power allocations for each spatial stream. While eigen-beamforming provides capacity-wise the optimal SU-MIMO performance, we found that its use in MU-MIMO is limited due to inter-user interference.

With MU-MIMO, we stated that the optimal channel capacity is achieved by the dirty paper coding and interference alignment methods. However, their practical applications are still limited by complexity and other factors. Therefore, a set of sub-optimal common beamforming methods that arise from the zero inter-user interference constraint were studied.

We started by studying cases where each user had only one receiving antenna. It was shown that a simple channel inversion method is able to cancel all the interferences. However, this method turned out to be impractical in many cases due to singularities in channel matrices. Especially, the case where the transmitter and the receiver have both the same number of antennas proved to be problematic. This problem was solved by introducing a regularization term in the inversion. However, the regularization provided improvement only in the low SNR regime. Further improvement was discussed to be achievable with the vector perturbation method.

As a final zero-forcing technique, we studied the block diagonalization method that generalizes channel inversion to the case where users can have more than one receiving antenna. This method was shown to be the optimal precoding method subject to the zero-interference constraint. Versatility of the block diagonalization method was described by studying three use cases: maximum throughput, power control, and partial channel state information beamforming.

The theory part was concluded by discussing advanced beamforming solutions that both extended the methods we had earlier discussed, but also defined methods based on other initial assumptions. The solutions that extended the techniques we covered in this thesis seemed to follow two mainlines. Some of the solutions focused only on beamforming, improving the existing methods, while others introduced different methods for user scheduling to increase the system performance.

In the simulation part, we first described a geometric SISO channel model that was then extended to support MU-MIMO. This channel model then allowed visualization of beamforming in an environment with reflecting objects. Both eigen-beamforming and zero-forcing beamforming were visualized, but also other results were obtained as a side product.

We found that a large antenna array combined with eigen-beamforming has multiple benefits compared to a single omni-directional antenna. Then, we learned that with large antenna arrays, beamforming is necessary for good coverage. Finally, results showed that eigen-beamforming is robust to channel aging in LOS scenarios, and more generally, in scenarios with only one dominant propagation path. However, in non-LOS cases with more than one significant propagation path, we found that channel aging may cause problems in the real world applications.

Results for zero-forcing beamforming suggested that plain channel inversion may be very prone to channel aging. It was found that block diagonalization can provide improvement in performance, since it is able to automatically determine if some users are better left out of service. These results proved that a user scheduler is important in maximizing the system performance. As a final result, we found that a large number of transmitting antennas are needed if a given number of users want to be supported with high relative signal strength while using the channel inversion method.

Bibliography

[1] K. Hoffman and R. Kunze, *Linear Algebra*. Prentice-Hall, Inc., 2 ed., 1971.

[2] D. A. Harville, *Matrix algebra from a statistician's perspective*, vol. 1. Springer, 1997.

[3] D. S. Dummit and R. M. Foote, *Abstract Algebra*, pp. 543, 545, 615–617. John Wiley & Sons, Inc., 3 ed., 2004.

[4] R. Bellman, "Notes on matrix theory–ii," *The American Mathematical Monthly*, vol. 60, no. 3, pp. 173–175, 1953.

[5] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, pp. 610–617. Cambridge university press, 2005.

[7] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer, 2 ed., 1994.

[8] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta mathematica*, vol. 30, no. 1, pp. 175–193, 1906.

[9] E. F. Beckenbach and B. Richard, *Inequalities*. Springer-Verlag, 1961.

[10] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II. John Wiley & Sons, INC., 1971.

[11] J. Gipple, "The volume of n-balls," *Rose-Hulman Undergraduate Mathematics Journal*, vol. 15, no. 1, pp. 238–248, 2014.

[12] H. Minkowski, "Über Geometrie der Zahlen," *Gesammelte Abhandlungen*, vol. 1, p. 264, 1911.

[13] E. Hlawka, "Zur Geometrie der Zahlen," *Mathematische Zeitschrift*, vol. 49, no. 1, pp. 285–312, 1943.

[14] J. Cassels, "A short proof of the Minkowski-Hlawka theorem," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 49, pp. 165–166, Cambridge University Press, 1953.

[15] ETSI, *3GPP TS 36.212 version 10.0.0 Release 10*, 2011.

[16] Cisco, *802.11ac: The Fifth Generation of Wi-Fi*, 2014. Technical Wthie Paper.

[17] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[18] 3GPP, "5G; Study on New Radio (NR) access technology," Technical Report (TR) 38.912, 3rd Generation Partnership Project (3GPP, 05 2017. Version 14.0.0, Release 14.

[19] J. Boutros, G. Caire, E. Viterbo, H. Sawaya, and S. Vialle, "Turbo code at 0.03 dB from capacity limit," in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, p. 56, IEEE, 2002.

[20] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Communications letters*, vol. 5, no. 2, pp. 58–60, 2001.

[21] E. Arikan and E. Telatar, "On the rate of channel polarization," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pp. 1493–1495, IEEE, 2009.

[22] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Transactions on Emerging Telecommunications Technologies*, vol. 10, no. 6, pp. 585–595, 1999.

[23] H. Huang, C. B. Papadias, and S. Venkatesan, *MIMO Communication for Cellular Networks*, ch. 2, pp. 35–78. Springer Science & Business Media, 2011.

[24] E. T. Jaynes, "Information theory and statistical mechanics," in *Statistical Physics*, vol. 3, pp. 181–218, W. A. Benjamin, INC., 1963.

[25] G. Tauböck, "A maximum entropy theorem for complex-valued random vectors, with implications on capacity," in *Information Theory Workshop (ITW), 2011 IEEE*, pp. 375–379, IEEE, 2011.

[26] R. V. Hartley, "Transmission of information," *Bell Labs Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928.

[27] Y. Shen and E. Martinez, "Channel estimation in ofdm systems," *Freescale Semiconductor*, vol. 2, no. 6, 2006.

[28] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-advanced for mobile broadband*, pp. 35–36, 152–161, 210–221. Academic press, 2011.

[29] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on communications*, vol. 46, no. 3, pp. 357–366, 1998.

[30] S. Russ, "A translation of Bolzano's paper on the intermediate value theorem," *Historia Mathematica*, vol. 7, no. 2, pp. 156–185, 1980.

[31] M. Costa, "Writing on dirty paper (corresp.)," *IEEE transactions on information theory*, vol. 29, no. 3, pp. 439–441, 1983.

[32] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[33] O. El Ayach, S. W. Peters, and R. W. Heath, "The practical challenges of interference alignment," *IEEE Wireless Communications*, vol. 20, no. 1, pp. 35–42, 2013.

[34] S. Serbetli and A. Yener, "Time-slotted multiuser mimo systems: Beamforming and scheduling strategies," *EURASIP Journal on Wireless Communications and Networking*, vol. 2004, no. 2, pp. 286–296, 2004.

[35] H. Lee, M. Shin, and C. Lee, "An eigen-based MIMO multiuser scheduler robust to spatial channel correlation," in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, vol. 2, pp. 1134–1137, IEEE, 2005.

[36] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[37] T. Haustein, C. Von Helmolt, E. Jorswieck, V. Jungnickel, and V. Pohl, "Performance of MIMO systems with channel inversion," in *Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th*, vol. 1, pp. 35–39, IEEE, 2002.

[38] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: Channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[39] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: Perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, 2005.

[40] R. Penrose, "A generalized inverse for matrices," in *Mathematical proceedings of the Cambridge philosophical society*, vol. 51, pp. 406–413, Cambridge University Press, 1955.

[41] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," in *in Proc. 40th Annual Allerton Conf. Communications, Control and Computing*, Citeseer, 2002.

[42] A. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Meth. Dokl.*, vol. 4, pp. 1035–1038, 1963.

[43] A. Tikhonov, "Regularization of incorrectly posed problems," *Soviet Meth. Dokl.*, vol. 4, pp. 1624–1627, 1963.

[44] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics letters*, vol. 7, no. 5, pp. 138–139, 1971.

[45] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Transactions on Communications*, vol. 20, no. 4, pp. 774–780, 1972.

[46] O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Communications letters*, vol. 4, no. 5, pp. 161–163, 2000.

[47] A. Tarighat, M. Sadek, and A. H. Sayed, "A multi user beamforming scheme for downlink mimo channels based on maximizing signal-to-leakage ratios," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, vol. 3, pp. iii–1129, IEEE, 2005.

[48] D. A. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding.," in *PIMRC*, pp. 107–111, 2005.

[49] D. A. Karpuk and P. Moss, "Channel pre-inversion and max-sinr vector perturbation for large-scale broadcast channels," *IEEE Transactions on Broadcasting*, vol. 63, no. 3, pp. 494–506, 2017.

[50] C.-B. Chae, S. Shim, and R. W. Heath, "Block diagonalized vector perturbation for multiuser mimo systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, 2008.

[51] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series*, vol. 7. MIT press Cambridge, MA, 1949.

[52] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on selected areas in communications*, vol. 24, no. 3, pp. 528–541, 2006.

[53] L. Jin, X. Gu, and Z. Hu, "Low-complexity scheduling strategy for wireless multiuser multiple-input multiple-output downlink system," *IET communications*, vol. 5, no. 7, pp. 990–995, 2011.

[54] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Transactions on Signal processing*, vol. 53, no. 10, pp. 3857–3868, 2005.

[55] J. Wang, D. J. Love, and M. D. Zoltowski, "User selection with zero-forcing beamforming achieves the asymptotically optimal sum rate," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3713–3726, 2008.

[56] S. Huang, H. Yin, J. Wu, and V. C. Leung, "User selection for multiuser mimo downlink with zero-forcing beamforming," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3084–3097, 2013.

[57] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.

[58] M. Steinbauer, A. F. Molisch, and E. Bonek, "The double-directional radio channel," *IEEE Antennas and propagation Magazine*, vol. 43, no. 4, pp. 51–63, 2001.

[59] J. Salmi, "Introduction to propagation modeling." Tutorial lecture at GETA Winter School, 2012.

[60] S. Faruque, "Free space propagation," in *Radio Frequency Propagation Made Easy*, pp. 19–26, Springer, 2015.

[61] T. Doi, K. Toyoda, and Y. Tanimura, "Effects of phase changes on reflection and their wavelength dependence in optical profilometry," *Applied optics*, vol. 36, no. 28, pp. 7157–7161, 1997.

[62] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) requirements for LTE Pico Node B," Technical Report (TR) 36.931, 3rd Generation Partnership Project (3GPP, 05 2011. Version 9.0.0, Release 9.

[63] Federal Communications Comission (FCC), *Code of Federal Regulations, Title 47 , Chapter I, Subchapter A, Part 15 – Radio frequency devices*. Government Publishing Office (GPO), 2016.

[64] Samsung, "Galaxy S8 – specifications." `http://www.samsung.com/global/galaxy/galaxy-s8/specs/`. Accessed at: 19/12/2017.

[65] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical review*, vol. 32, no. 1, p. 97, 1928.

[66] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physical review*, vol. 32, no. 1, p. 110, 1928.

Appendix A

Additional Simulation Results

Here we present additional results that did not fit into the main text. This appendix consists of two sections. First, in Section A.1, we describe how the radiation pattern of an ideal antenna array can be calculated. This gives a visual motivation for the term beamforming. Then, in Section A.2, we extend the zero-forcing results that were presented in Chapter 6.

A.1 Antenna Array Radiation Pattern

This section gives a visual motivation for the term beamforming. We will first derive the optimal relative phases for each antenna element in an array so that the transmitted signal is steered to the desired direction. Then, we will use this result to plot a radiation pattern of an ideal antenna array.

Instead of studying the transmission of a signal, we will first consider its reception. Due to symmetry, this gives then the optimal relative phases that should be used while transmitting signal to the desired direction. Let $\mathbf{d}_i$ be the position of the ith antenna element with respect to the phase center[1] and suppose that it is a perfect omni-directional antenna. Then, suppose that a plane wave with frequency f approaches the antenna array from a direction $\mathbf{r}$. Assuming that the direction is given by a unit vector, then the signal must travel distance

$$\Delta_i = \mathbf{d}_i^\top \mathbf{r} \tag{A.1}$$

after reaching the phase center to reach the ith antenna element. Notice that the travel distance Δ_i can be negative meaning that the signal arrives to the ith antenna element before reaching the phase center. Since the signal phase

[1]Phase center here refers to a virtual location that is agreed to be the reference point. This means that the signal is received at phase zero at the phase center, and the phases at physical antennas are compared to this reference point.

at the phase center is zero, we can find the signal phase at the ith antenna element by

$$\varphi_i = \frac{2\pi\Delta_i}{\lambda} = 2\pi\frac{f}{c}\mathbf{d}_i^\top\mathbf{r} \tag{A.2}$$

where λ is the wave length of the signal and c is the speed of light. Therefore, the received complex signal at the ith antenna is

$$b_i = A_i \exp\left(\mathrm{i}2\pi\frac{f}{c}\mathbf{d}_i^\top\mathbf{r}\right) \tag{A.3}$$

where i is the imaginary unit and A_i is the amplitude of the signal.

To transmit the signal to the same direction as discussed with reception, need to notice that the normal vector of wave has now the opposite sign. Therefore, the desired direction is given by $-\mathbf{r}$. Using equation (A.3), we can find that the optimal relative complex unit signal at the ith antenna element is

$$w_i(\mathbf{r}, f) = \exp\left(-\mathrm{i}2\pi\frac{f}{c}\mathbf{d}_i^\top\mathbf{r}\right). \tag{A.4}$$

Using equation (A.4) we can compute how the transmitted signals are summed in any direction $\mathbf{q}$.

Let us fix the frequency f, and let $w_i(\mathbf{r}, f) = w_i(\mathbf{r})$. Denoting the relative transmit phases for the whole antenna by $\mathbf{w}(\mathbf{r}) = (w_i(\mathbf{r}))$. Then, the magnitude at direction $\mathbf{q}$ of a transmitted signal intended to direction $\mathbf{r}$ is given by

$$p(\mathbf{q}) = \left|\mathbf{w}(\mathbf{r})^\dagger\mathbf{w}(\mathbf{q})\right|^2. \tag{A.5}$$

Using equation (A.5) for each $\mathbf{q}$ in a dense unit sphere grid, we can plot the radiation pattern of an antenna array.

Figure A.1 shows an example radiation pattern of a 4×4 square antenna array. In this figure, the signal is transmitted to the directions that deviates $15°$ in both azimuth and elevation directions from the array normal. Due to symmetry of the array the radiation pattern is reflection symmetric. We can see that in the transmission direction (and near it) the gain is large, but in the other directions, the gain is rather small. Visually this creates a beam looking pattern.

To link the previous discussion to beamforming, and especially to eigenbeamforming, we may consider a plain LOS channel with one user that has one receiving antenna. Suppose the user is far away from the base station, so that a plane wave assumption can be made. Then, the channel matrix $\mathbf{H}$ is in this case

$$\mathbf{H} = L_P\mathbf{w}(\mathbf{r}) \tag{A.6}$$

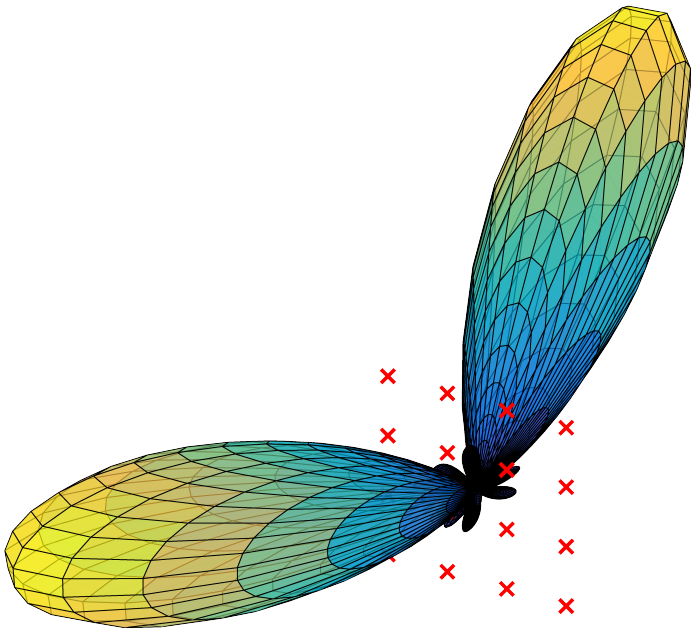


Figure A.1: Radiation pattern of an ideal 4×4 antenna array. Red crosses denote the antenna elements.

where $\mathbf{r}$ points to the direction of the user and $L_P \in \mathbb{R}_+$ is the path loss. Since the channel matrix is a vector, its only right singular vector is the channel matrix itself scaled to unit length. Therefore, the radiation pattern in Figure A.1 visualizes the beamforming gain in plain LOS channel. Due to this visual interpretation, we can see that the name beamforming is rather descriptive.

A.2 Zero-Forcing Beamforming

In this section, we present additional zero-forcing beamforming results. We will first visualize that the zero-forcing condition is satisfied by channel inversion. This was hard to verify from the results presented in Chapter 6, but now we will remove all the reflectors, which allows a more clear visualization. Then, we study the selection of the regularization term in the regularized inversion method.

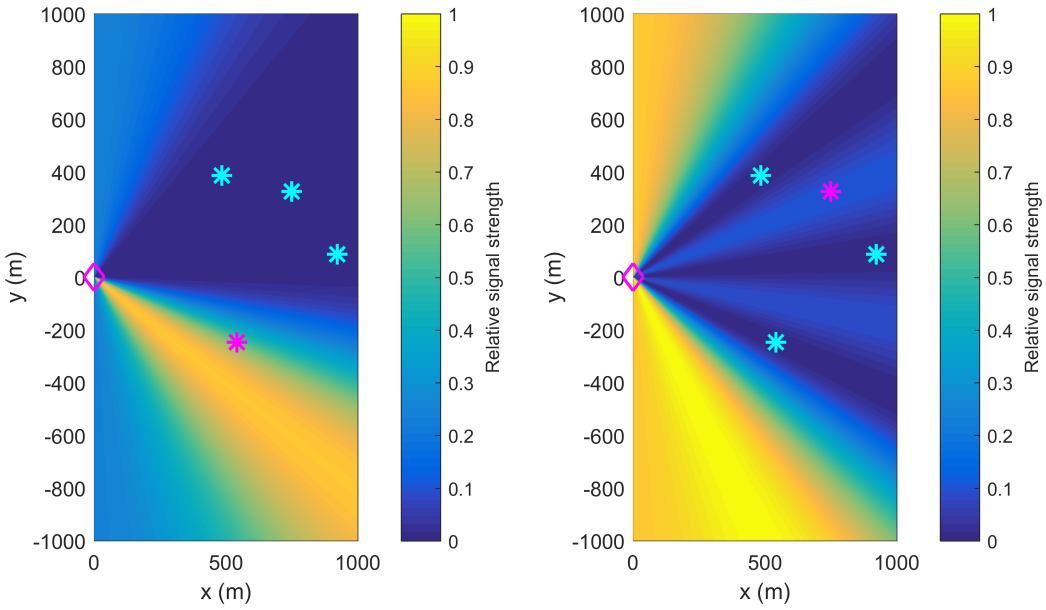
Figure A.2 visualizes channel inversion in plain LOS channel. In this simulation, the base station was equipped with a 4×4 square antenna array, and each user was assumed to have only one receiving antenna. Geometry-based channel was set to have no reflectors and each party was set to lie on

the same plane. Each figure from A.2a to A.2d visualizes the relative signal strength[2] intended to one of the users. We can clearly see that in each figure only the intended user has non-zero relative signal strength. Therefore, there exists no inter-user interference as desired. However, we can also notice that the relative signal strengths at the intended users are not very good. For example, user 2 has only 0.1 relative signal strength.

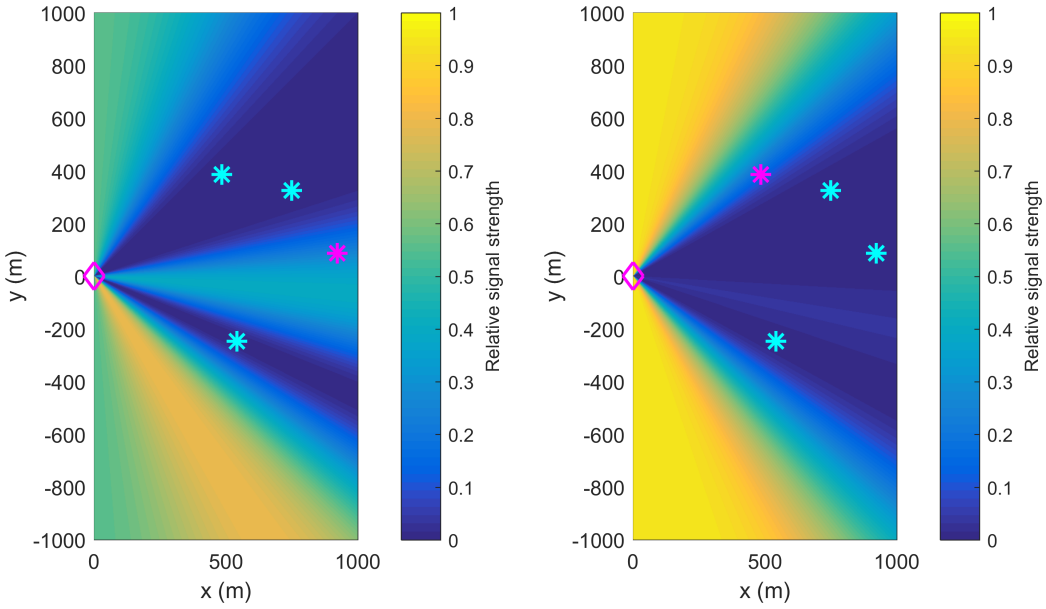
In Figure A.3, the scenario is the same in A.2b other than the transmitting antenna array is now 8×8 instead of 4×4. We can see that the relative signal power at user 2 is now close to 1. It seems that extra degrees of freedom provide better signal strengths at the intended user while still satisfying the zero-forcing condition.

We conclude this section by considering regularized channel inversion. Figure A.4 shows the channel capacity is a function of regularization term α while using regularized channel inversion. Transmitting antenna array was set to be 4×4, and 4 users were randomly placed in a geometry-based channel model. Sufficiently many channel realizations were used to stabilize the results. Figure shows the mean capacity and one standard deviation error bounds. We may notice that substantially better channel capacity can be obtained by using regularized inversion compared to plain channel inversion (corresponding $\alpha = 0$). However, too much regularization decreases the performance to the level that is worse than that of channel inversion. The optimal value of α turned out to be $\alpha \approx 2 \cdot 10^{-12}$. This aligns were well with the theoretical optimal value $\alpha_{opt} = N\sigma^2$ derived in [38] by using large-N approximation. Here N refers to the number of users and σ^2 is the noise power. In the simulation, the noise power was computed by using Johnson-Nyquist noise [65], [66] which led the value of $N\sigma^2$ to be just below 10^{-12}.

[2]Relative signal strength here means the received signal power compared to the optimal signal power that would be obtained via eigen-beamforming.



(a) User 1 (b) User 2



(c) User 3 (d) User 4

Figure A.2: Interference patterns created by signals intended to one user subject to zero-forcing beamforming. Magenta diamond is the base station, magenta asterisk is the intended user, and cyan asterisks are the other users. Reflectors were not used in this case.

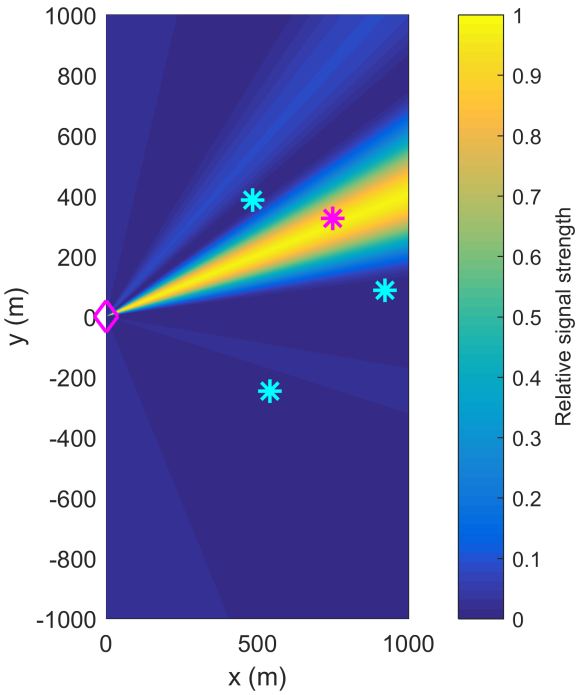


Figure A.3: Same case as in Figure A.2b, but with a 8×8 transmitting antenna array instead of 4×4.

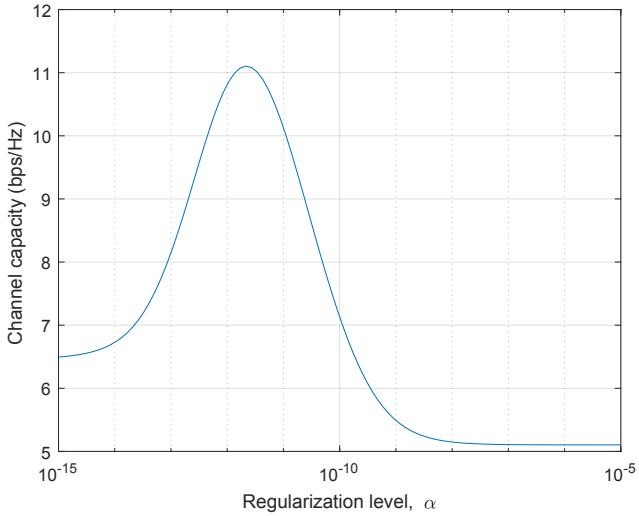


Figure A.4: Channel capacity as a function of regularization level α while using regularized channel inversion. Transmitter has a 4×4 antenna array, and there are 4 users each with one antenna. Total transmitting power was normalized to be 1.

Appendix B

Additional Proofs

B.1 Proof of Equation (3.3)

Lemma B.1. *Let* $\mathbf{y} = \mathbf{x} + \mathbf{n}$ *where* $\mathbf{x} = (x_i)$, $\mathbf{n} = (n_i) \in \mathbb{R}^n$. *Suppose* $\mathbb{E}[x_i^2] = P$ *and* $n_i \sim \mathcal{N}(0, \sigma^2)$. *Then,*

$$\mathbb{E}[\|\mathbf{y}\|] \leq \sqrt{N(P + \sigma^2)} \ .$$

Proof. Using Jensen's inequality on convex function $x \mapsto x^2$ and linearity of expectation, we have

$$\begin{aligned}
\mathbb{E}[\|\mathbf{y}\|] &\leq \sqrt{\mathbb{E}\left[\|\mathbf{y}\|^2\right]} = \sqrt{\mathbb{E}\left[\|\mathbf{x} + \mathbf{n}\|^2\right]} \\
&= \sqrt{\mathbb{E}\left[\sum_{i=1}^{N} (x_i^2 + n_i^2)\right]} \\
&= \sqrt{\sum_{i=1}^{N} \mathbb{E}[x_i^2] + \sum_{i=1}^{N} \mathbb{E}[n_i^2]} \\
&= \sqrt{N(P + \sigma^2)} \ .
\end{aligned}$$

□

B.2 Proof of Equation (4.3)

Lemma B.2. *Suppose*

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ *and* $\mathbf{y}$ *are the transmitted and received signals respectively and* $\mathbf{n}$ *is the random noise. Then,*

$$h(\mathbf{y}|\mathbf{x}) = h(\mathbf{n})$$

where h *denotes differential entropy.*

Proof. Using definitions of differential entropy and expected value, we may write the differential entropies as integrals.

$$\begin{aligned}
h(\mathbf{y}|\mathbf{x}) &= h(\mathbf{y},\mathbf{x}) - h(\mathbf{x}) \\
&= -\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}} p(\mathbf{y},\mathbf{x})\log_2 p(\mathbf{y},\mathbf{x})\mathbf{dxdy} + \int_{\mathbb{R}^{2n}} p(\mathbf{x})\log_2 p(\mathbf{x})\mathbf{dx} \\
&= -\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}} p(\mathbf{y},\mathbf{x})\log_2 p(\mathbf{y},\mathbf{x})\mathbf{dxdy} + \int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}} p(\mathbf{y},\mathbf{x})\log_2 p(\mathbf{x})\mathbf{dxdy} \\
&= -\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}} p(\mathbf{y},\mathbf{x})\log_2 \frac{p(\mathbf{y},\mathbf{x})}{p(\mathbf{x})}\mathbf{dxdy}
\end{aligned}$$

Now, we do a change of variables $\mathbf{y} \mapsto \mathbf{y} - \mathbf{Hx} = \mathbf{n}$. From the perspective of $\mathbf{y}$, this change is just setting an offset of $\mathbf{Hx}$ with fixed $\mathbf{x}$. Therefore, the Jacobian scaling factor is 1 and can be neglected. Since the transmitted signal is independent of noise, $p(\mathbf{n},\mathbf{x}) = p(\mathbf{n})p(\mathbf{x})$.

$$\begin{aligned}
&= -\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}} p(\mathbf{n},\mathbf{x})\log_2 \frac{p(\mathbf{n},\mathbf{x})}{p(\mathbf{x})}\mathbf{dxdn} \\
&= -\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}} p(\mathbf{n})p(\mathbf{x})\log_2 \frac{p(\mathbf{n})p(\mathbf{x})}{p(\mathbf{x})}\mathbf{dxdn} \\
&= -\int_{\mathbb{R}^{2n}} p(\mathbf{x})\mathbf{dx}\int_{\mathbb{R}^{2n}} p(\mathbf{n})\log_2 p(\mathbf{n})\mathbf{dn}
\end{aligned}$$

Since $p(\mathbf{x})$ is PDF of $\mathbf{x}$, its integral over whole domain is 1 by definition. Therefore, we have

$$\begin{aligned}
&= -\int_{\mathbb{R}^{2n}} p(\mathbf{n})\log_2 p(\mathbf{n})\mathbf{dn} \\
&= h(\mathbf{n})
\end{aligned}$$

which is the desired result. □

B.3 Proof of Lemma 4.4

Lemma 4.4. *Let $a_i, x_i \in \mathbb{R}$ for all $i = 1, \ldots, n$ satisfying $a_1 \geq \cdots \geq a_n \geq 0$ and $\sum_{i=1}^n x_i^2 = 1$. Then,*

$$\max_{x_1,\ldots,x_n} \sum_{i=1}^n a_i x_i^2 = a_1$$

and the maximum is achieved with $x_1 = 1$, $x_2 = \cdots = x_n = 0$.

Proof. Notice that this is an optimization problem on a surface of the unit sphere, so the maximum (and minimum) value is attained at a critical point. We use the method of Lagrange multipliers. Lagrange function is given by

$$\mathcal{L}(x, \ldots, x_n, \lambda) = \sum_{i=1}^n a_i x_i^2 - \lambda \left(\sum_{i=1}^n x_i^2 - 1 \right) .$$

We may compute the needed partial derivatives

$$\frac{\partial \mathcal{L}}{\partial x_i} = 2x_i(a_i - \lambda)$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_{i=1}^n x_i^2 .$$

Notice that $\partial \mathcal{L}/\partial x_i = 0$ if and only if $x_i = 0$ or $a_i = \lambda$. Thus, all the critical points are $x_i = 0$ for $i \notin \mathcal{D}$ and $\sum_{i \in \mathcal{D}} x_i^2 = 1$ where $\mathcal{D}$ is some set of i for which a_i is the same. Let $a_i = a_{\mathcal{D}}$ for all $i \in \mathcal{D}$. Then, the value of the objective function is

$$\sum_{i=1}^n a_i x_i^2 = \sum_{i \in \mathcal{D}} a_i x^2 = a_{\mathcal{D}} \sum_{i \in \mathcal{D}} x_i^2 = a_{\mathcal{D}} .$$

Since we assumed that $\max_i a_i = a_1$, we conclude that

$$\max_{x_1,\ldots,x_n} \sum_{i=1}^n a_i x_i^2 = a_1 .$$

Moreover, this maximum is attained at $x_1 = 1$ and $x_2 = \cdots = x_n = 0$. □

B.4 Proof of Proposition 5.2

Lemma B.3. *For all $k \in \mathbb{Z}_+$ the integral*

$$\int_{-\infty}^{\infty} \left| n^k e^{-n^2/2\sigma^2} \right| \mathbf{d}n$$

converges.

Proof. Let us prove the statement first for even k via induction. If $k = 0$, then

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-n^2/2\sigma^2} \mathbf{d}n &= \left(\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} \mathbf{d}x \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} \mathbf{d}y \right)^{1/2} \\
&= \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-x^2-y^2}{2\sigma^2}} \mathbf{d}x \mathbf{d}y \right)^{1/2} \qquad \left| \begin{matrix} x = r\cos\varphi \\ y = r\sin\varphi \end{matrix} \right. \\
&= \left(\int_0^{\infty} \int_0^{2\pi} r e^{-r^2/2\sigma^2} \mathbf{d}\varphi \mathbf{d}r \right)^{1/2} \\
&= \sqrt{2\pi\sigma^2} \left(\Big/_0^{\infty} - e^{-r^2/2\sigma^2} \right)^{1/2} \\
&= \sqrt{2\pi\sigma^2} \, .
\end{aligned}$$

Now, suppose that the integral converges for $2k$. Then,

$$\begin{aligned}
\int_{-\infty}^{\infty} n^{2(k+1)} e^{-n^2/2\sigma^2} \mathbf{d}n &= C \Big/_{-\infty}^{\infty} n^{2k+1} e^{-n^2/2\sigma^2} + C' \int_{-\infty}^{\infty} n^{2k} e^{-n^2/2\sigma^2} \mathbf{d}n \\
&= C' \int_{-\infty}^{\infty} n^{2k} e^{-n^2/2\sigma^2} \mathbf{d}n < \infty
\end{aligned}$$

where C, C' are some constants. Here we used result $\lim_{n\to\infty} n^k e^{-n^2} = 0$ for all k, which can be proved by applying l'Hospital rule k times to the limit. By induction, the integral in consideration converges for all even k.

Now, let k be an odd number. Then, we may notice that

$$n^{k+1} e^{-n^2/2\sigma^2} \geq n^k e^{-n^2/2\sigma^2}$$

for all $n \geq 1$. Let

$$h(n) = \begin{cases} 1 \, , & |n| \leq 1 \\ n^{k+1} e^{-n^2/\sigma^2} \, , & \text{otherwise} \end{cases} .$$

Since k is an odd number, $k+1$ is ever, so we have

$$\int_{-\infty}^{\infty}\left|n^k e^{-n^2/2\sigma^2}\right|\mathbf{d}n \leq \int_{-\infty}^{\infty} h(n)\mathbf{d}n = \int_{-1}^{1} 1\mathbf{d}n + \int_{\mathbb{R}\setminus[-1,1]} n^{k+1}e^{-n^2/2\sigma^2}\mathbf{d}n$$
$$\leq 2 + \int_{-\infty}^{\infty} n^{k+1}e^{-n^2/2\sigma^2}\mathbf{d}n < \infty \ .$$

Therefore, the statement holds also for the odd numbers. □

Proposition 5.2. *Let $a \in \mathbb{R}$ and $n \sim \mathcal{N}(0,\sigma^2)$. Then,*

$$\lim_{\beta\to\infty} \beta^k \mathbb{E}\left[\left\lfloor \frac{a+n}{\beta} + \frac{1}{2}\right\rfloor^{\ell} n^m\right] = 0$$

for all $k, m \in \mathbb{Z}_+$ and $\ell \in \mathbb{Z}_{\geq 1}$.

Proof. We may write the expectation as an integral.

$$\lim_{\beta\to\infty} \beta^k \mathbb{E}\left[\left\lfloor \frac{a+n}{\beta} + \frac{1}{2}\right\rfloor^{\ell} n^m\right] = C \lim_{\beta\to\infty} \int_{-\infty}^{\infty} \beta^k e^{-n^2/2\sigma^2}\left\lfloor \frac{a+n}{\beta} + \frac{1}{2}\right\rfloor^{\ell} n^m \mathbf{d}n$$

where $C = 1/\sqrt{2\pi\sigma^2}$. Let us denote

$$g_\beta(n) = \beta^k e^{-n^2/2\sigma^2}\left\lfloor \frac{a+n}{\beta} + \frac{1}{2}\right\rfloor^{\ell} n^m \ .$$

We will now use the dominant convergence theorem to show that the order of limit and integral can be interchanged.

Notice that

$$\left\lfloor \frac{a+n}{\beta} + \frac{1}{2}\right\rfloor n^m = 0$$

if $\beta > 2(a+n)$. Let $n_0 = \beta/2 - a$ and suppose β is large enough to ensure $n_0 > a$ and $\beta \geq 1$. For $n > n_0$ we have

$$\beta^k = 2^k(a+n_0)^k < 2^k(2n_0)^k < 4^k n^k$$

and

$$\left\lfloor \frac{a+n}{\beta} + \frac{1}{2}\right\rfloor^{\ell} n^m \leq \left\lfloor 2\frac{a+n}{\beta}\right\rfloor^{\ell} n^m \leq 2^\ell (a+n)^\ell n^m \leq 4^\ell n^{\ell+m} \ .$$

Let $g(n) = 4^{k+\ell} n^{k+\ell+m} e^{-n^2/2\sigma^2}$. Then, $|g_\beta(n)| \leq g(n)$ for all n (except maybe the point n_0 which can be excluded). By Lemma B.3, g is absolutely integrable over reals, so by the dominant convergence theorem, we may interchange the order of limit and integral. Hence, we have

$$\begin{aligned}
\lim_{\beta\to\infty}\left|\beta^k \mathbb{E}\left\lfloor \frac{a+n}{\beta}+\frac{1}{2}\right\rfloor^\ell n^m\right| &= C\lim_{\beta\to\infty}\int_{-\infty}^{\infty}|g_\beta(n)|\mathbf{d}n \\
&= C\int_{-\infty}^{\infty}\lim_{\beta\to\infty}|g_\beta(n)|\mathbf{d}n \\
&= C\lim_{R\to\infty}\int_{-R}^{R}|n^m|e^{-n^2/2\sigma^2}\lim_{\beta\to\infty}\left|\beta^k\left\lfloor \frac{a+n}{\beta}+\frac{1}{2}\right\rfloor^\ell\right|\mathbf{d}n \\
&\leq C\lim_{R\to\infty}\int_{-R}^{R}\lim_{\beta\to\infty}\left|\beta^k\left\lfloor \frac{a+R}{\beta}+\frac{1}{2}\right\rfloor^\ell\right|\mathbf{d}n \\
&= 0
\end{aligned}$$

for all $k, m \in \mathbb{Z}_+$ and $\ell \in \mathbb{Z}_{\geq 1}$. Here, we used the fact that the last floor function is zero for all $\beta > 2(a+R)$. This implies the desired result, so we are ready. □