



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Communication Technology Laboratory
Wireless Communications Group

Optimization of MIMO Channel

Subtitle of Master
Thesis



Master Thesis
Spring 2014

Marc MÃ¼ller

Advisor: Yahia Hassan
Professor: Prof. Dr. Armin Wittneben

Abstract

Here comes the abstract...

Preface

This Master Thesis is part of the graduate study at the Department of Information Technology and Electrical Engineering (D-ITET) at the Swiss Federal Institute of Technology (ETH) Zurich.

The author certifies that this Master Thesis, and the research to which it refers, are the product of the author's own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referring practices of the discipline.

Marc M \tilde{A} $\frac{1}{4}$ ller

Acknowledgements

I would like to thank blablabla

Author: Marc M \ddot{A} $\frac{1}{4}$ ller

marcmue@student.ee.ethz.ch

Advisor: Yahia Hassan

yehia.hassan@nari.ee.ethz.ch

Professor: Prof. Dr. Armin Wittneben

wittneben@nari.ee.ethz.ch

Contents

1	Information Theory	1
1.1	general Channel	1
1.2	MIMO	2
1.3	Singular Value Decomposition	3
1.4	the Receiver	4
2	CSI	5
2.1	Multiuser Case	5
2.2	Optimizing the Throughput of V-Blast	5
2.3	Comparison	6
2.4	Linear MMSE	6
2.5	Optimizing	6
2.6	Numerical Gradient	6
3	Gradient search	7
3.1	Steepest descent	7
3.2	Numerical Gradient	7
3.3	Analytical Gradient for Throughput Maximization	7
3.4	Adaptive step Size	8
3.5	Implementation	8
3.6	Gradient for power Minimization	8
3.7	Fodor Paper	9
4	results	11
	Bibliography	13

List of Figures

List of Tables

List of Algorithms

Chapter 1

Information Theory

1.1 general Channel

//by book of Tse

We use $x(t)$ as the transmitted signal and $y(t)$ as the received signal. Since the signal is affected by different objects in its path the received signal can be written as a function of the resulting differentiating paths

$$y(t) = \sum_i a_i(f, t) x(t - \tau_i(t))$$

. $a_i(f, t)$ describes the attenuation of each path which is mostly dependent on frequency and time. we can omit these dependencies if we can assume sufficient narrow bandwidth for a frequency flat channel and with relatively stationary objects a time flat channel, objects are transmitters, receivers, scatterers and so on. $\tau_i(t)$ is the delay in time of the path and is constant in a time flat channel.

To distinguish between these different channels we introduce some measures:

Delay Spread T_d describes the time difference from the first to the last arrival of the signal at the receiver.

Doppler Spread D_s describes the maximum frequency difference which is introduced by Doppler shifts.

coherence Bandwidth W_c is the inverse of the delay spread and a measure of the frequency flatness of the channel. it is the bandwidth at which we can assume that the channel is frequency independent.

coherence Time T_c is the inverse of the Doppler spread and a measure for the time flatness of the channel e.g. it is the time over which the channel is time invariant.

// image

1.1.1 The equivalent baseband representation

If we have a frequency-flat and time-flat channel we can represent our channel with a time varying channel gain h which leads to

$$y(t) = h \cdot x(t)$$

for our input output relation. But it is often simpler to use a complex baseband model. if we sample the system at a sufficiently high rate we get

$$y[m] = \sum_l h_l \cdot x[m]$$

where m describes discrete time $x[m]$ and $y[m]$ the transmitted respectively the received signal.

1.1.2 Channel Capacity

the channel capacity is a good figure of merit to describe the optimal performance of a communication system. It gives a maximum data rate at which we could transmit errorfree when using appropriate coding.

AWGN-channel: in an AWGN channel we experience no fading and the received signal consists just the sent signal with additive white Gaussian noise $y = x + n$. The capacity depends only on the transmit P versus noise Power N_0 :

$$C = \log\left(1 + \frac{P}{N_0}\right) \text{ bits/s/Hz}$$

time- and frequency-flat fading channel: in a fading channel the capacity is a function of the fading gain h :

$$C = \log\left(1 + |h|^2 \frac{P}{N_0}\right) \text{ bits/s/Hz}$$

1.2 MIMO

In the strive for higher data rates the limits of the basic single Antenna communication channels became obvious. The bandwidth limitation due to scarcity and cost and power constraints due to physics and regulations called for different approaches. MIMO uses the spatial dimension of the channel to further increase reliability and speed.

1.2.1 from SISO to MIMO

We are still using the assumptions of frequency- and time-flat channels. In the SISO case the discrete complex baseband channel tap h characterized the channel sufficiently. If we now use multiple antennas on the transmitter and receiver side of the channel and make sure that they are sufficiently spaced it establishes multiple distinct channels due to small-scale fading processes. In this case our discrete complex baseband channel tap h is no longer a single coefficient but a matrix $\mathbf{H} \in \mathbb{C}^{N \times M}$ with M number of transmit antennas and N number of receive antennas.

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1M} \\ h_{21} & h_{22} & \dots & h_{2M} \\ \vdots & \vdots & \dots & \vdots \\ h_{N1} & h_{N2} & \dots & h_{NM} \end{bmatrix} \quad (1.1)$$

//figure the input output relation of a MIMO channel is then written as

$$\mathbf{y}[m] = \mathbf{H}\mathbf{x}[m] + \mathbf{w}[m] \quad (1.2)$$

where $\mathbf{x} \in \mathbb{C}^M$, $\mathbf{y} \in \mathbb{C}^N$ and $\mathbf{w} \sim \mathcal{CN}(0, N_0 \mathbf{I}_N)$

1.2.2 MIMO Capacity

For a SISO channel in a Rayleigh fading channel we get a bit error rate off

$$P_b(|h|^2) = 1/2 \operatorname{erfc} \left(\sqrt{|h|^2 \frac{\bar{E}_b}{N_0}} \right). \quad (1.3)$$

, if we now send the same signal over N statistically independent channels we get a BER of

$$P_b(|h|^2) = 1/2 \operatorname{erfc} \left(\sqrt{\sum_{j=1}^N |h_j|^2 \frac{E_s}{N_0}} \right) \quad (1.4)$$

for the euivalent SISO channel.

1.3 Singular Value Decomposition

If we are in single user MIMO system and we can assume full CSI (Channel State Information at Transmitter and Receiver) we can fully diagonalize the Channel by using the SVD of \mathbf{H} :

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H \quad (1.5)$$

with \mathbf{U} and \mathbf{V} unitary and square and \mathbf{S} a diagonal matrix consiting of al singular Values of \mathbf{H} .

With this knowlage we can multiply our transmit and received vector as follows: $\tilde{\mathbf{y}} = \mathbf{U}^H \mathbf{y}$ and $\tilde{\mathbf{x}} = \mathbf{V} \mathbf{x}$. and our transmit equation changes as follows:

$$\tilde{\mathbf{y}} = \mathbf{U}^H \mathbf{y} \quad (1.6)$$

$$= \mathbf{U}^H (\mathbf{H} \tilde{\mathbf{x}} + \mathbf{w}) \quad (1.7)$$

$$= \mathbf{U}^H \mathbf{U} \mathbf{S} \mathbf{V}^H \tilde{\mathbf{x}} + \tilde{\mathbf{w}} \quad (1.8)$$

$$= \mathbf{S} \mathbf{V}^H \mathbf{V} \tilde{\mathbf{x}} + \tilde{\mathbf{w}} \quad (1.9)$$

$$= \mathbf{S} \tilde{\mathbf{x}} + \tilde{\mathbf{w}}. \quad (1.10)$$

Now we have created $\min(N, M)$ parallel SISO systems out of our MIMO system. this brings us to the question of how to distribute the transmit power optimally over these systems to acheave the maximal possible ergodic rate.

1.3.1 waterfilling

The well known answer to that question is the water filling Solution described in //tsebook as follows:

1.4 the Receiver

In this project we were mainly looking into optimizing on the transmitter side of the channel but the receiver structure has a huge impact at the performance of the system too. In this section we will use two different receiver structures:

1. Linear MMSE Receiver

//formula

2. V-BLAST MMSE or SIC with MMSE (MMSE-SIC)

//formula (algorithm)

Generally speaking the SIC receivers provide better Throughput but are more complicated.

1.4.1 Linear MMSE

1.4.2 MMSE-SIC

We can use the result of one decorrelator filter to aid the others which results in a successive cancellation strategy: we decode $x_1[m]$ first and then subtract the decoded stream from the received vector. So the second decoder sees only the interference of the streams x_3, \dots, x_m . But be aware of error propagation, this means that an error in decoder 2 leads to erroneous results in all subsequent decoders.

The V-BLAST gives us

$$R < \log_2 \det \left(\mathbf{I}_N + \frac{1}{N_0} \mathbf{H} \frac{\mathbf{K}_x}{\sigma^2} \mathbf{H}^H \right) \text{ bits/s/Hz} \quad (1.11)$$

with $\mathbf{K}_x = \{P_1, \dots, P_M\}$ as the covariance matrix of the transmitted signal. Which leads to a capacity of

$$C = \max_{\text{Tr} \mathbf{K}_x \leq P} \mathbb{E} \left[\log_2 \det \left(\mathbf{I}_N + \frac{1}{N_0} \mathbf{H} \frac{\mathbf{K}_x}{\sigma^2} \mathbf{H}^H \right) \right] \quad (1.12)$$

which is the capacity of the fast fading MIMO channel.

In the case of CSIT we already discussed the optimal solution as the waterfilling. But with only CSIR //tse proves that an equal Power allocation, in other words equal power on each transmit antenna gives us the best performance

$$\mathbf{K}_x = \left(\frac{P}{M} \right) \mathbf{I}_N \quad (1.13)$$

which gives us a resulting capacity of

$$C = \mathbb{E} \left[\log_2 \det \left(\mathbf{I}_N + \frac{\text{SNR}}{M} \mathbf{H} \mathbf{H}^H \right) \right]. \quad (1.14)$$

We see a power loss in comparison to full CSI of M which is achieved by using transmit beamforming.

Chapter 2

CSI

The question now is, is it possible to increase the Throughput of this channel with knowledge of the channel Matrix H . and to investigate this we use following Channel Representation and we are fully aware of its implications and imperfections.

// description of our channel

2.1 Multiuser Case

Since Water Filling diagonalizes the channel it does not matter what kind of decoder is used on the receiver side, the streams are already independent. But Waterfilling is only feasible in a single user case with full CSIT due to the fact that inter antenna data coordination is needed. So in a multi user scenario we need other methods to decode the channel. We use two well-known Receiver structures:

1. Linear MMSE Receiver

//formula

2. V-BLAST MMSE or SIC with MMSE (MMSE-SIC)

//formula (algorithm)

Generally speaking the SIC receivers provide better Throughput but are more complicated to implement, and harder to adapt.

2.2 Optimizing the Throughput of V-Blast

[Jindal] The Paper proposes an algorithm which optimizes the throughput with an SIC receiver, the papers goal is to solve the problem of beam forming in a Broadcast or Downlink Channel, so exactly the opposite direction than we are discussing in this report. But as //paper has described there is a duality of these two channels and we are able to use the optimized results of the MAC to optimize the BC. So his work is using this duality to solve the easier problem of the MAC and then transform it back to the BC problem, so his work is highly relevant to us.

// Duality description

// abstract of Jindal

// Jindal algorithm

2.3 Comparison

Numerical Gradient Search function We implemented a numerical gradient search for the SIC receiver to compare to the Jindal algorithm in terms of throughput maximization. We do not compare it in terms of complexity or speed of convergence. Convergence is guaranteed for the numerical Gradient based on the work of //paper. And the Jindal himself adds the proof of convergence to in his paper.

2.4 Linear MMSE

From there we started to investigate the LMMSE receiver.

//theory
// formulas

2.5 Optimizing

// problems //drawbacks

Fodor describes an algorithm in his //paper to optimize the throughput of a typical Cellular Network Channel, but to improve stability and feasibility of his algorithm he sacrifices accuracy and chooses to use a very weak constraint. Which leads to unusable results in comparison to an easy numerical Gradient search of the same Optimum.

2.6 Numerical Gradient

So we will only look at the numerical Gradient for Optimizing the LMMSE receiver. As shown in // Figure there is some pretty big improvement over just equal power distribution, but even for high SNR cases at least one channel is closed, which means that at least one user is not allowed to send any data.

Chapter 3

Gradient search

3.1 Steepest descent

The idea of the steepest descent algorithm is to find the global maximum or minimum of a function by searching along the gradient of the function. We used two ways to calculate the gradient at a certain point, numerical and analytical:

3.2 Numerical Gradient

The numerical Algorithm calculates the Gradient with numerical means. Which means to increase the input in each dimension one after another by a very small amount and recalculate output. With both results, the original one and the new ones, we can approximate the local gradient very precisely. Analytical Gradient The analytical Algorithm on the other hand uses analytical tools to calculate the gradient from the known function in advance. It uses the gradient function to calculate the exact gradient at each point.

3.3 Analytical Gradient for Throughput Maximization

// calculations: formulas step by step

Implementation with the norm and not just the sum allows us to use the same function in two different ways. Firstly with a //p value of 1, we just calculate the sum Rate for the throughput maximization or sum Rate maximization. Secondly we can set the //p value to something very small like -30 to create a maxmin optimizer. The Optimum would be the //min norm but this is not feasible in the numerical domain and is prone to get stuck in local minimas, here -30 is a good approximation, we tested the algorithm with -100 and achieved results which were closer to the optimum that each channel gets the same Rate but it's not save to use.

1. Start at one point 2. Calculate the Gradient 3. Use the gradient to get to the next point (Step) 4. Iterate

1. Calculate Starting Point 2. 1:K a. Increase input K b. Calculate Point 3. Calculate Gradient 4. Step to next point 5. Iterate

Pre: calculate analytical Gradient by hand 1. Calculate gradient at input Point with the precalculated gradient function 2. Step to the next point 3. Iterate

3.4 Adaptive step Size

One Problem that occurred during the simulations was that at some point the gradient search got unstable, in our case it started to fluctuate and stopped moving towards the optimum. The reason of this bad and unwanted behavior lies in the step size of the function, and so there are two ways to fix it. The first and easy way is just to reduce the step size and increase the number of iterations accordingly, this way the problem occurs later and weaker and has less impact on the end result. The second much more complicated solution is to introduce adaptive step size.

Adaptive Step size means, that for each iteration we search the optimal step size which allows us to take the biggest step towards our goal. So it should not be possible to choose a step size which gives us a worse result than we had on the iteration before. With this Algorithm we can prevent any fluctuation in the gradient and force it to go the fastest possible way. We do not have to calculate the gradient as often and can thus reduce the needed calculation time. Another nice side effect is, that we implicitly know when we reached the maximum or minimum when the step size is decreased to zero without finding any better points than the previous.

3.5 Implementation

//lots of code

3.6 Gradient for power Minimization

// calculations
 // differences to the throughput max
 // question about convecity?

Start at an arbitrary step size, calculate the next point If next Point is better than last, increase step size If next Point is worse than last, decrease step size Stop if increasing or decreasing does not improve the result anymore

3.7 Fodor Paper

3.7.1 The Algorithm

1. calculate the effective interference for MMSE processing:

$$\zeta_{k,s} = \left\{ \left(d_{k,k}^- \chi_{k,k} H_{k,k}^H \left(\sum_j P_j d_{k,j}^- \chi_{k,j} H_{k,j} T_j T_j^H H_{k,j}^H + N_t \sigma_n^2 I \right)^{-1} H_{k,k} + \frac{1}{P_k} I \right)^{-1} \right\}^{(s,s)} \quad (3.1)$$

2. calculate the optimal loading matrix

$$(T_k)^{(s,s)} = \sqrt{\frac{\zeta_{k,s} N_t}{\sum_{j=1}^{N_t} \zeta_{k,j}}} \quad \forall s \in [1, N_t] \quad (3.2)$$

3. calculate used Power

$$P_k = \frac{\zeta_{k,s}}{|(T_k)^{(s,s)}|^2} (\gamma_{tgt} + 1) \quad \forall k \quad (3.3)$$

- n. until no more change

3.7.2 closed form Solution

Fodor reuses a proposition for a closed form solution from xy. p^* is calculated from ... with

—

Chapter 4

results

Bibliography

