University of Ottawa

Massive MIMO Channels Under the Joint Power Constraints

by

Mahdi Khojastehnia

A thesis submitted in fulfillment of the requirements for the degree of Masters of Applied Science in Electrical and computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering

Abstract

Massive MIMO has been recognized as a key technology for 5G systems due to its high spectral efficiency. The capacity and optimal signaling for a MIMO channel under the total power constraint (TPC) are well-known and can be obtained by the water-filling (WF) procedure. However, much less is known about optimal signaling under the per-antenna power constraint (PAC) or under the joint power constraints (TPC+PAC). In this thesis, we consider a massive MIMO Gaussian channel under favorable propagation (FP) and obtain the optimal transmit covariance under the joint constraints. The effect of the joint constraints on the optimal power allocation (OPA) is shown. While it has some similarities to the standard WF, it also has number of notable differences. The numbers of active streams and active PACs are obtained, and a closed-form expression for the optimal dual variable is given. A capped water-filling interpretation of the OPA is given, which is similar to the standard WF, where a container has both floor and ceiling profiles. An iterative water-filling algorithm is proposed to find the OPA under the joint constraints, and its convergence to the OPA is proven.

The robustness of optimal signaling under FP is demonstrated in which it becomes nearly optimal for a nearly favorable propagation channel. An upper bound of the sub-optimality gap is given which characterizes nearly (or ϵ)-favorable propagation. This upper bound quantifies how close the channel is to the FP.

A bisection algorithm is developed to numerically compute the optimal dual variable. Newton-barrier and Monte-Carlo algorithms are developed to find the optimal signaling under the joint constraints for an arbitrary channel, not necessarily for a favorable propagation channel.

When the diagonal entries of the channel Gram matrix are fixed, it is shown that a favorable propagation channel is not necessarily the best among all possible propagation scenarios capacity-wise.

We further show that the main theorems in [1] on favorable propagation are not correct in general. To make their conclusions valid, some modifications as well as additional assumptions are needed, which are given here.

Acknowledgments

I would first like to express my sincere gratitude to my supervisor, Dr. Sergey Loyka, for his guidance, motivation, knowledge, and continuous support from the very beginning of my studies at the University of Ottawa. His attention to detail helped me to learn much more and to be a much better researcher, and I would like to thank him for helping me in writing this thesis.

I would also like to thank Dr. Francois Gagnon for his help, and my appreciation goes also to École de Technologie Supérieure (Montreal), which supported this study.

I would like to thank my dear brother Ali for his support and help, and a special thank you to all my amazing friends, especially Mohsen, for helping me during these two years.

I would like to express my deepest gratitude to my lovely mother and father for their unconditional love and support. I am forever indebted to them, and I am pretty sure that without them, I could not have gotten to where I am today. My parents have always been my motivation to move forward, and I dedicate this thesis to them.

Contents

Abbreviations

4G	Fourth **G**eneration
5G	Fifth **G**eneration
AP	**A**ccess **P**oint
AFP	**A**symptotic **F**avorable **P**ropagation
AWGN	**A**dditive **W**hite **G**aussian **N**oise
BS	**B**ase **S**tation
CDF	**C**umulative **D**istribution **F**unction
CSI	**C**hannel **S**tate **I**nformation
DPC	**D**irty **P**aper **C**oding
EVD	**E**igen**v**alue **D**ecomposition
EGT	**E**qual **G**ain **T**ransmission
FP	**F**avorable **P**ropagation
i.i.d.	**i**ndependent **i**dentical **d**istribution
KKT	**K**arush-**K**uhn-**T**ucker
LOS	**L**ine **o**f **S**ight
LTE	**L**ong-**T**erm **E**volution
MAC	**M**ultiple **A**ccess **C**hannel
MIMO	**M**ultiple **I**nput **M**ultiple **O**utput
MISO	**M**ultiple **I**nput **S**ingle **O**utput
MMSE	**M**inimum **M**ean **S**quare **E**rror
MRT	**M**aximum **R**atio **T**ransmission
MRC	**M**aximum **R**atio **C**ombining
NFP	**N**early **F**avorable **P**ropagation
OFDM	**O**rthogonal **F**requency **D**ivision **M**ultiplexing

OPA	**O**ptimal **P**ower **A**llocation
PAC	**P**er-**A**ntenna power **C**onstraint
Rx	Receiver
SIC	**S**uccessive **I**nterference **C**ancellation
SINR	**S**ignal to **I**nterference and **N**oise **R**atio
SISO	**S**ingle **I**nput **S**ingle **O**utput
SLNR	**S**ignal to **L**eakage plus **N**oise **R**atio
SNR	**S**ignal to **N**oise **R**atio
SVS	**S**ingular **V**alue **S**pread
TDD	**T**ime **D**ivision **D**uplex
TPC	**T**otal **P**ower **C**onstraint
Tx	Transmitter
UCA	**U**niform **C**ircular **A**rrays
ULA	**U**niform **L**inear **A**rrays
UPA	**U**niform **P**laner **A**rrays
WF	**W**ater **F**illing
ZF	**Z**ero **F**orcing

List of Symbols

$\mathbf{A}$	Bold capitals like $\mathbf{A}$ denote the matrices
$\mathbf{a}$	Bold lower-case letter like $\mathbf{a}$ denote vectors
α_i	A coefficient that shows the allocated bandwidth and a level of priority for the i-th user
β	This controls the reduction in the step size in a backtracking line search
C	The channel Capacity [nat/second/hertz]
$C(\alpha_i, g_i)$	The capacity of the weighted rate as a function of coefficient α and channel gains g_i
$\mathbb{C}^{n,m}$	The set of all $n \times m$ complex matrices
$C_{PAC}(\mathbf{W})$	The capacity under the PAC as a function of the channel Gram matrix $\mathbf{W}$
$C_{TPC}(\mathbf{W})$	The capacity under the TPC as a function of the channel Gram matrix $\mathbf{W}$
$C(\mathbf{W})$	The capacity as a function of the channel Gram matrix $\mathbf{W}$
$C(\mathbf{W}, \mathbf{R})$	The maximum achievable rate as a function of the channel Gram matrix $\mathbf{W}$ and transmit covariance $\mathbf{R}$
$\mathbf{D}$	Duplication matrix
$D_{\mathbf{r}}$	The first-order derivative of $\mathbf{r}(\mathbf{a})$
d_0	The space for the BS antennas in the uniform linear array pattern

$\mathbf{d}_i$	A distance vector from the i-th user to all APs' antennas
$\Delta C(\mathbf{W}, \mathbf{W}_0)$	The sub-optimality gap (rate loss)
Δf	The total bandwidth of the system
Δf_i	The allocated bandwidth for the i-th user
Δ_k	The uncertainty interval at the k-th step of the bisection algorithm
$\delta \mathbf{p}_t^*$	The optimal power allocation under the TPC alone at the t-th step of the iterative water-filling algorithm.
$\delta p_{i,t}^*$	The optimal power allocation for the i-th stream under the TPC alone at the t-th step of the iterative water-filling algorithm.
Δ_t	The power allocated to streams which are removed at the end of the step t in the iterative water-filling algorithm
$\Delta \mathbf{x}$	The update of the variable in the Newton algorithm
$\mathbf{e}_i$	A vector where all components are zero except the i-th component, which is one.
η_i	The Lagrange multipliers responsible for the non-negative constraint
f_c	Carrier frequency
$\mathbf{G}$	Fast fading matrix with zero mean and unit variance
$\widetilde{\mathbf{G}}$	This matrix represents attenuation and shadow fading
$\mathbf{g}_t$	The remaining streams at the end of the t-th step of the iterative water-filling algorithm
g_i	The i-th diagonal entry of $\mathbf{W}$
γ	The SNR or SNIR at receiver

h	The fixed channel gain between transmitter and receiver in a SISO channel
$\mathbf{h}_i$	The channel gains vector from i-th transmitter antenna to each receiver antennas
$\mathbf{H}$	Channel gains matrix
$\mathbf{I}$	The identity matrix of appropriate size
$\mathbf{I}_m$	The identity matrix with size $m \times m$
$\mathcal{I}_{\overline{PAC}}$	The set of active streams with inactive PACs
$\mathcal{I}_{PAC}$	The set of streams with active PACs
$\mathcal{I}_t$	The set of streams that exceed the PACs at the t-th step of the iterative water-filling algorithm
κ	The barrier parameter t increases by the factor κ in each step of the barrier algorithm
$\mathbf{\Lambda}^*$	The diagonal matrix which the diagonal entries are determined by the water-filling (WF) procedure
λ_c	Carrier wavelength
λ_i	The Lagrange multipliers responsible for the PACs
m	Number of transmitter antennas
m_+	Number of active streams
m_{PAC}	Number of active PACs
μ	The optimal dual variable responsible for the total power constraint under the TPC+PAC constraints.
μ'_k	An estimation of μ at the k-th step of the bisection algorithm
μ_l	A lower bound of μ
μ_{lk}	The updated lower bound of μ at the k-th step of the bisection algorithm

μ_t^{-1}	The water level in the WF procedure at the t-th step in the iterative water-filling algorithm
μ_u	An upper bound of μ
μ_{uk}	The updated upper bound of μ at the k-th step of the bisection algorithm
μ_{WF}	The optimal dual variable responsible for the total power constraint in the water filling procedure.
n	Number of receiver antennas
ν	This determines the linear decreasing in the residual norm in the Newton-barrier algorithm
P_1	The identical per-antenna power constraint
P_{1i}	The power constraint for the i-th antenna
P_T	The maximum total Tx power
p_i	The i-th diagonal entry of $\mathbf{R}$
p_i^*	The optimal power allocation for the i-th stream under the total and per-antenna power constraints
$p_{i,PAC}^*$	The optimal power allocation for the i-th stream under the PACs
$p_{i,WF}^*$	The optimal power allocation for the i-th stream (the water-filling procedure)
$p_{i,t}^*$	The allocated power for the i-th stream that satisfies the PAC in the t-th step of the iterative water-filling algorithm
$\mathbf{R}$	Transmit covariance matrix
$\mathbf{R}^*(\mathbf{W})$	The optimal transmit covariance matrix as a function of the channel Gram matrix $\mathbf{W}$
$\mathbf{R}_t^*(\mathbf{W})$	The optimal point for the unconstrained problem in the Newton-barrier algorithm for a fixed t.

$\mathbf{R}_0$	An initial transmit covariance in the Newton-barrier algorithm
$\mathbf{R}_0^*$	An initial transmit covariance in the Monte-Carlo algorithm
$\mathbf{R}_i$	The feasible transmit covariance at step i in the Monte-Carlo algorithm
$\mathbf{R}_i^*$	The best feasible transmit covariance until the i-th step in the Monte-Carlo algorithm
$\mathbb{R}^{n,m}$	The set of all $n \times m$ real matrices
$\mathbf{r}(\mathbf{a})$	This shows gradient of an objective function in unconstrained problem in the Newton-barrier algorithm with respect to $\mathbf{a}$
S_R	The feasible set of Tx covariance matrices
S_p	The feasible set of power allocations under the joint total and per-antenna power constraints
s	The step size, which would be computed by a backtracking line search
s_i	A coefficient that shows a grade of service for the i-th user
$\psi_1(\mathbf{R})$	A logarithmic barrier function to absorb a constraint responsible for the positive semi-definite for Tx covariance
$\psi_1(\mathbf{R})$	A logarithmic barrier function to absorb the total power constraint
$\psi_{1i}(\mathbf{R})$	A logarithmic barrier function to absorb the per-antenna power constraints
ξ_t	The white Gaussian noise at time index t for a SISO channel
$\boldsymbol{\xi}$	The i.i.d. circularly-symmetric additive white Gaussian noise at the receiver of unit variance in a MIMO channel

t_0	An initial value of the barrier parameter t in the Newton-barrier algorithm.
t_{max}	A maximum value of the parameter t in the Newton-barrier algorithm
θ_i	The angle of arrival for the i-th user
$\mathbf{W}$	The Gram matrix of $\mathbf{H}$, i.e., $\mathbf{W} = \mathbf{H}^{+}\mathbf{H}$
$WF\{\mathbf{a}, b\}$	The standard water-filling for streams $\mathbf{a}$ under the total power constraint b
x_t	The transmitter signal at time index t for a SISO channel
$\mathbf{x}$	The transmitted signal in a MIMO channel
$\mathbf{x}_k$	The variable at the step k in the Newton algorithm.
y_t	The received signal at time index t for a SISO channel
$\mathbf{y}$	The received signal in a MIMO channel

List of Mathematical Operators

$\mathbf{A}^+$	The Hermitian conjugate of $\mathbf{A}$
$\|\mathbf{A}\|$	The determinant of $\mathbf{A}$
$\mathbf{A} \geq 0$	$\mathbf{A}$ is a positive semi-definite matrix
$\mathbf{A} > 0$	$\mathbf{A}$ is a positive definite matrix
$(\mathbf{A})_{ij}$	The ij-th entry of $\mathbf{A}$
$\|\|\mathbf{a}\|\|$	Euclidean norm of $\mathbf{a}$
$(a)_+$	This gives a maximum value between x and 0, i.e., $(a)+ = max\{a, 0\}$
$\lceil a \rceil$	It returns an integer number that is greater than or equal to a.
$c_{i,j}$	Correlation coefficient between i-th and j-th channel vectors
$\mathbf{D}_A$	It is a diagonal matrix that has the same diagonal entries as $\mathbf{A}$
$\mathbf{d}_i(\mathbf{A})$	The i-th diagonal entry of $\mathbf{A}$
$diag\{a_i\}$	It returns a diagonal matrix that the diagonal entries are a_i.
$\mathbb{E}\{\cdot\}$	Statistical expectation
$\mathbf{\Lambda}_A$	The diagonal matrix in which its diagonal entries are equal to the eigenvalues of $\mathbf{A}$
$\lambda_i(\mathbf{A})$	The i-th eigenvalue of $\mathbf{A}$. Unless stated otherwise, they are in decreasing order.
$Pr\{\cdot\}$	Probability

$rank\mathbf{A}$	Rank of matrix $\mathbf{A}$
`randn`	This operator returns a matrix whose entries are generated with a normal distribution.
$\sigma_1(\mathbf{A})$	The largest singular value of $\mathbf{A}$
$tr\mathbf{A}$	Trace of $\mathbf{A}$
$\mathbf{U}_A$	The unitary matrix consisting of the eigenvecotrs of the normal matrix $\mathbf{A}$
$\mathbf{u}_i$	The eigenvector corresponding to i-th largest eigenvalue of a square matrix
$Var\{\cdot\}$	Variance
$vec(\mathbf{A})$	This operator stacks all entries of $\mathbf{A}$ into a single column vector.
$vech(\mathbf{A})$	This operator picks up the lower-triangular entries of $\mathbf{A}$, starting from the first column to the last column.

Chapter 1

Introduction

Wireless communication technology is one of the actively growing fields today, and has become of increasing interest for both academia and industry. A lot of research has been done in recent years to find solutions for challenges in this technology as the number of users and the demand for wireless communication consistently increases [2], [3], [4]. As an example, it has been forecasted that the number of smartphone subscriptions will be around 7 billion in 2024 [5], and many of these will need high-rate wireless communication capabilities. So, providing sufficient data rates for these wireless users is an important issue.

MIMO (multiple-input multiple-output) technology is one wireless communication technology in which both receiver and transmitter devices are equipped with multiple antennas, and this technology provides significant benefits, such as providing high data rates. As an example, Fig. 1.1 shows the performance of MIMO channels in terms of data rates. In this figure, the data rates are shown for different SNR (signal-to-noise ratio) values when the fixed bandwidth of the channel is 1 MHz. It can be seen that the data rates increase significantly by increasing the number of antennas, without extra bandwidth or extra power.

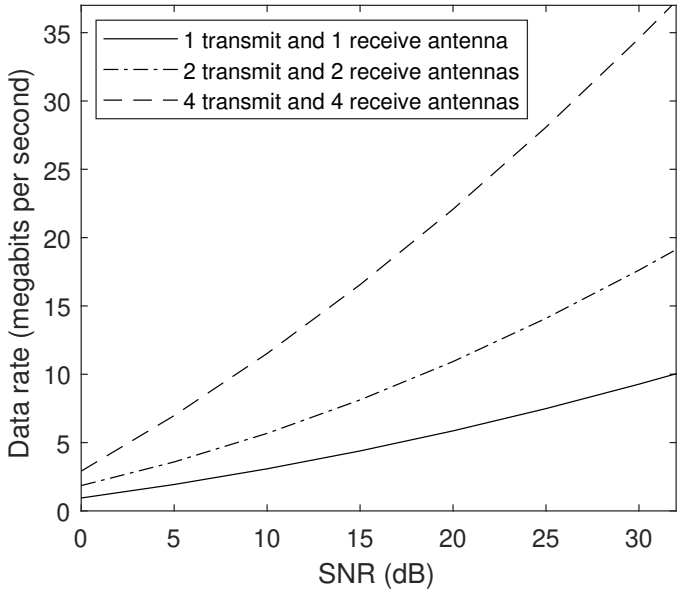

FIGURE 1.1: Data rate vs. SNR. The bandwidth of the channel is 1 MHz (reconstructed from [2]). As an example, the bandwidth of WiFi devices is around 20 MHz. In this case, the rate is scaled by 20 for this SNR.

In real systems, there are always some constraints that need to be addressed. These constraints affect the channel capacity[1]. Finite power is one of the constraints that limit the channel capacity. From a practical point of view, the total transmit power constraint (TPC) and per-antenna power constraint (PAC) should be satisfied in MIMO channels. The former constraint is determined by the limited power available from the power supply, and the latter is practically important since the amplifier for each antenna has a limited power budget. However, the optimal transmit covariance over a MIMO channel under the joint power constraints (i.e., TPC+PAC) is still unknown in general, except for some special cases (e.g., multi-input single-output (MISO) channels, full-rank optimal transmit covariance).

Massive MIMO (or large-scale antenna systems) is a new deployment for wireless communication [4]. This is considered as a key technology for 5G systems in order to provide high-rate communication to users [6]. In massive MIMO, the base station is equipped with a large number of antennas, which enables remarkable benefits. One of them is orthogonality of the wireless channel, which is due to

[1]The channel capacity is the maximum achievable data rate that reliable communication can be achieved.

favorable propagation. In this case, signal processing becomes very simple; for example, the performance of linear signal processing schemes becomes optimal [7]. Both theory and measurement-based results show that favorable propagation (or more precisely, nearly favorable propagation) is obtained for many cases as long as the number of BS antennas is sufficiently large [8], [7].

In this thesis, we study massive MIMO channels with additive white Gaussian noise (AWGN) under the joint total and per-antenna power constraints (TPC+PAC). This case is motivated by practical implementations, in which both the TPC and PAC constraints should be satisfied simultaneously.

1.1 Main Contributions

The main contributions of this thesis are summarized below.

- A closed-form expression for optimal signaling in a massive MIMO channel under favorable propagation and the joint total and per-antenna power constraints is obtained.
- The robustness of the optimal signaling under favorable propagation is established, which is nearly optimal under approximately favorable propagation.
- Monte-Carlo and Newton-barrier algorithms are developed which compute optimal signaling under the joint power constraints for general MIMO channels, not necessarily ones under favorable propagation. A bisection algorithm is developed to find the optimal dual variable.
- Reference [1] is studied in more detail, and its main theorems are shown to need some modifications to make their conclusions valid.

Based on the research reported here the following papers have been published/-submitted.

- M. Khojastehnia, S. Loyka, F. Gagnon, "Massive MIMO Channels Under the Joint Power Constraints", The Canadian Workshop on Information Theory (CWIT), Hamilton, Ontario, Canada, June 2-5, 2019.
- S. Loyka, M. Khojastehnia, "Comments on "On Favorable Propagation in Massive MIMO Systems and Different Antenna Configurations" [1]", accepted in IEEE Access.
- M. Khojastehnia, S. Loyka, F. Gagnon, "Robust Optimal Signaling for Massive MIMO Channels", submitted to 45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2020).

Additionally, the following papers have been submitted. They study optimal signaling in a wiretap MIMO channel (related to the thesis, but not included).

- M. Khojastehnia, S. Loyka, "Comments on "Precoding for Secrecy Rate Maximisation in Cognitive MIMO Wiretap Channels" [1]" accepted in Electronics Letters.
- M. Khojastehnia, S. Loyka, "Comments on "AN-Aided Secrecy Precoding for SWIPT in Cognitive MIMO Broadcast Channels "" submitted to IEEE Communications Letters.

1.2 Thesis Outline

The rest of this thesis is organized as follows.

Chapter 2: **Literature Review**

This chapter presents the literature review on MIMO channels. It is shown that considering the joint total and per-antenna power constraints is practically important, because the transmitter power supply is limited and each antenna uses its own amplifier which operates with limited power as well. Optimal signaling under

the joint power constraints (TPC+PAC) is known just for some special cases, and these are presented in this chapter. Some known results regarding favorable propagation in massive MIMO channels are discussed, and specifically, both theory and measurement-based results show that massive MIMO implies an approximate orthogonal channel for many realistic propagation scenarios.

Chapter 3: **Channel Model and Optimal Signaling**

The standard model of a discrete-time fixed AWGN MIMO channel is given, which is the basis of the analyses in this thesis. In this chapter, we consider three different power constraints: (i) total power constraint, (ii) per-antenna power constraint, and (iii) joint total and per-antenna power constraint. Optimal signaling for a MIMO channel under the total power constraint is given, which is on the channel eigenmodes. In this case, the OPA is obtained by the water-filling procedure. The optimal transmit covariance under PAC or under the joint power constrains (TPC+PAC) is not known in general, except in some special cases (MISO channels, full-rank solution). The definitions of favorable propagation, nearly FP, and asymptotically FP are given. Under favorable propagation, the channel is orthogonal. The off-diagonal entries in the channel Gram matrix are non-zero but small under nearly FP. Under asymptotically FP, the normalized inner product of each two distinct channel vectors converges to zero as the number of antennas grows.

Chapter 4: **Optimal Signaling Under Favorable Propagation**

In this chapter, a closed-form solution of the optimal signaling under favorable propagation and the joint power constraints (TPC+PAC) is obtained. It is shown that the optimal transmit covariance matrix is diagonal, and that its diagonal entries are the minimum of two terms: the first one is the per-antenna power constraints, and the second represents the WF procedure. It is shown that the optimal dual variable responsible for the TPC in the WF procedure is greater than that under the joint power constraints. Closed-form expressions for the numbers of active PACs and active streams are given. A geometric interpretation of the

OPA under the joint power constraints (TPC+PAC) is proposed. The OPA can be interpreted by pouring water into a container which has floor and ceiling profiles. Additionally, an iterative water-filling algorithm is proposed, and we show that the algorithm converges to OPA under the joint power constraints (TPC+PAC) for a favorable propagation channel. The weighted problem under the joint power constraints (TPC+PAC) is considered in which each user has a different allocated bandwidth as well as a different grade of service. In this case, the OPA is found analytically in a closed form.

Chapter 5: **Nearly Favorable Propagation and Robustness**

This chapter determines the robustness of the optimal signaling to deviations from FP. It is shown that optimal signaling under FP is nearly optimal under nearly favorable propagation. The upper bound of the sub-optimality gap becomes small when nearly favorable is obtained. Based on this, a new definition of nearly (or ϵ)-favorable propagation is given. It quantifies how close a channel is to favorable propagation one.

Chapter 6: **Numerical Algorithms**

We develop some algorithms to find the optimal signaling for a general MIMO channel under the joint power constraints (TPC+PAC). The first one is the bisection algorithm, which can find the optimal dual variable responsible for the TPC. This algorithm computes a root of a monotonic continuous function. At each step of the algorithm, the upper or lower bound of the root are redefined, and the width of the uncertainty interval around the root decreases. This process continues until this uncertainty becomes sufficiently small, and hence an estimate of the root is close enough to its actual value. Second, we develop a Newton-barrier method. This method is powerful in solving convex optimization problems such as finding the optimal transmit covariance under the joint power constraints (TPC+PAC) in the general case. In this algorithm, the logarithmic barrier function is used

to add the inequality constraints to an objective function. Third, we develop a Monte-Carlo algorithm which obtains the channel capacity by randomly sampling a large number of transmit covariance matrices and selecting the best covariance. We demonstrate via an example that CVX can give incorrect results for the capacity under the joint power constraints.

Chapter 7: **A Study of Favorable Propagation**

We consider the case that Gram matrices of orthogonal and non-orthogonal channels have the same diagonal entries. In this case, when the beamforming is optimal, it is shown that the capacity of the orthogonal channel is less than that of the receptive non-orthogonal channel. So, among all channels with the same diagonal Gram matrix entries, favorable propagation is not necessarily the best capacity-wise. Finally, we comment on [1] and show that the main theorems there are not correct in general. Some modifications and additional assumptions are proposed to make the conclusions valid.

Chapter 2

Literature Review

In this chapter, the capacity for a Gaussian channel is reviewed. We show that the joint total and per-antenna power constraints are practically important, and some known results regarding optimal signaling under these constraints are presented. Also, by using theory-based and measurement-based results, it is shown that favorable propagation is obtained for many cases of massive MIMO channels.

2.1 Point-to-Point Wireless Channel

Point-to-point communication is one type of wireless communication system in which a base station (BS) serves a user. In this deployment, both the receiver (Rx) and the transmitter (Tx) may have multiple antennas [9]. The simplest case is when both Tx and Rx have a single antenna and it is called a single-input single-output (SISO) channel. Fig. 2.1 shows a SISO channel with additive white Gaussian noise (AWGN).

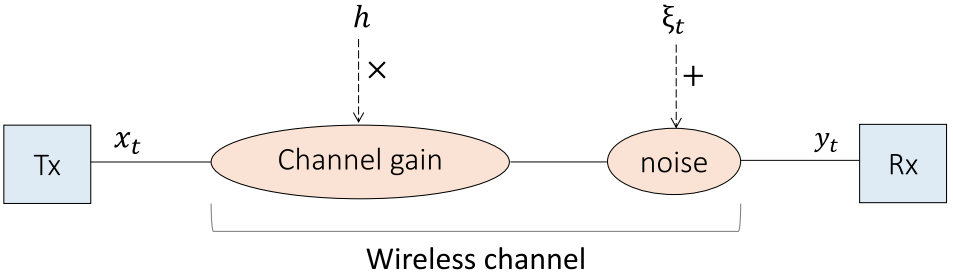

FIGURE 2.1: Gaussian SISO channel

The model for a discrete-time memoryless (frequency flat) Gaussian SISO channel for time-invariant systems is as follows[1]:

$$y_t = h \cdot x_t + \xi_t \tag{2.1}$$

where y_t and x_t are the received and transmitted signals at discrete time t, respectively; ξ_t is the white Gaussian noise, and h is the fixed channel gain between Tx and Rx (for simplicity, the time index t will be removed in later notations).

Claude Shannon showed that there is a maximum achievable rate over a noisy channel at which one can communicate with an arbitrarily small error probability, and this maximum achievable rate is called the *capacity* of the channel [2], [3], [11]. Reliable communication can be achieved for any rate less than the capacity; however, error probability cannot be arbitrary small for any rate greater than the capacity.

The channel capacity for the AWGN SISO channel is as follows:

$$C = \ln(1 + \gamma) \quad [\text{nat/s/Hz}] \tag{2.2}$$

[1]In a digital communication system, at first, the information from the source is converted into bits, in which both sampling and quantization processes are applied. At this stage, the data is discrete in time. After that, the source coding and the channel coding are used and code symbols are obtained. In the next step, discrete-time signals are converted to continuous-time signals by using the pulse-modulation process and a baseband waveform is generated. To reach the carrier frequency, bandpass modulation is used and a baseband signal converts to a bandpass signal. The achieved signal is propagated into a channel, and the receiver uses reverse processes to obtain the intended information [10].

where C is the channel capacity, γ is the SNR at Rx [2], [12], [13], and *nat* is a unit of information where natural logarithms are used instead of base 2 logarithms.

Multiple-input multiple-output (MIMO) technology has been extensively used in the wireless communication industry, since it yields benefits such as increasing the data rate without extra SNR or bandwidth [2]. MIMO technology is included in the industrial standards in already existing systems such as 4G. In particular, long-term evolution (LTE) or 4G systems are designed in which both the receiver and the transmitter may have multiple antennas [14]. Additionally, this is considered as one of the key technologies for 5G systems in the form of massive MIMO (a large number of antennas) [6], and there has been research interest in massive MIMO since about 2010.

The channel capacity increases significantly by using multiple antennas at both Tx and Rx instead of a single antenna. In this case, Gaussian signaling has been proven to be optimal for AWGN MIMO channels [2], [15]. Fig. 2.2 shows an example of a MIMO channel where m and n are the numbers of Tx and Rx antennas respectively. We denote $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_m]$ as the channel gain matrix, where $\mathbf{h}_i$ is the channel gains vector from i-th Tx antenna to each Rx antennas (see (3.1) for the details of the channel model).

FIGURE 2.2: Gaussian point-to-point MIMO channel

2.2 Power Constraints

In reality, there are some design requirements which need to be considered. Limited available power is one of them. The capacity and the optimal signaling over a fixed Gaussian MIMO channel under the total power constraint (TPC) are well-known. The optimal signaling is on the channel eigenmodes and the optimal power allocation (OPA) is obtained by the water-filling (WF) procedure [2], [3], [15]. The following reasons highlight the importance and motivation of considering the total transmit power constraint: (i) power supply is limited, (ii) the Tx power affects the battery life; (iii) for environmental safety, the total Tx power should be less than a certain amount; and (iv) high-level Tx power causes more interference. For example, in cellular communication, increasing the base station power creates more interference for other cell users [16], [17], [18].

The per-antenna power constraint (PAC) is another constraint motivated by practical design concerns, since each Tx antenna has its own amplifier with a limited power budget. This is shown in Fig. 2.3, in which each antenna can also be considered as a single-antenna user. Hence, this consideration also applies to multi-user MIMO channels.

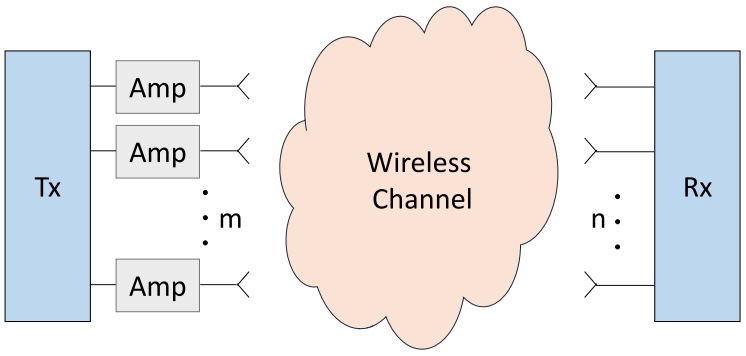

FIGURE 2.3: MIMO channel under the PAC

The optimal transmit covariance over a fixed MIMO channel under the PAC is not known in general, except for some special cases (e.g., rank-1 channels, MISO channels, full-rank solution). In the following, the results of past studies regarding

optimal signaling under the PAC will be discussed. This includes some properties of optimal signaling and some numerical algorithms to find the optimal Tx covariance[2].

Reference [19] shows that the optimal Tx covariance of a MISO channel under the PAC is rank-1 (i.e., the beamforming is optimal). Also, the beamforming is optimal for a MISO channel under the TPC[3]. In this case, the optimal Tx covariance has only one non-zero eigenvalue. This result is achieved where the channel gains are fixed and known to both the TX and the Rx. Under the TPC only, the phase of the beamforming vector matches with that of the channel vector, and its amplitude is dependent on the channel coefficient. However, under the PAC, the amplitude of the beamforming vector is not dependent on the channel vector and their phases are matched. Under the PAC, the per-antenna optimal power allocation are equal to the respective PACs (this is also seen for MISO wiretap channels under the PAC [21][4]). Also, in MISO channels, the capacity under the TPC is greater than that under the PAC when the total Tx power constraint is equal to the sum per-antenna power constraints. This is also the case for the ergodic capacity.

In a MISO channel with the i.i.d. (independent identical distribution) Rayleigh fading propagation model where the channel is only known at the Rx, independent signaling is optimal under the PAC, and the OPA is equal to the per-antenna power constraints [19]. This is also the case under the TPC only, and the power is uniformly allocated to all antennas [15]. In this propagation channel, if all the PACs are the same, then the uniform power allocation is also optimal under the PAC.

In the case of fixed MIMO channels under the PAC with full channel state information (CSI), a closed-form expression for the optimal Tx covariance is obtained

[2]The diagonal entries of the optimal transmit covariance determine the optimal power allocation for antennas, and its eigenvectors characterize the beam directions which are optimal [2] (see (3.4) for the precise definition of the optimal transmit covariance matrix).

[3]When the transmit strategy is beamforming, the signal is transmitted with a certain complex weight from each antenna, and weights are entries of the beamforming vector [20].

[4]In wiretap MIMO channels, Gaussian signaling is shown to be an optimal Tx strategy [22].

as a function of the dual variable [23]. Also, the paper proposes an iterative numerical algorithm to find the optimal dual variable. In this setting, the rank of the optimal Tx covariance is upper-bounded by that of the channel matrix. A similar result for the optimal Tx covariance rank is also used for the proof of the beamforming optimality in MISO channels under the PAC [19]. As mentioned before, the optimal Tx covariance under the TPC only is on the channel eigenmodes; however, this is not the case under the PAC [23].

The optimal signaling and capacity for MIMO channels under the PAC are obtained in [24], where both channel and Tx covariance are full-rank. The paper also obtains the optimal Tx covariance when the transmitter has two antennas and the channel is full-rank. In this case, as long as the difference between two singular values of the channel is sufficiently large, rank-1 transmission is optimal (i.e., the beamforming is optimal).

The ergodic capacity for a Gaussian fading MISO channel under the PAC is obtained in [25], where full CSI at both Rx and Tx is assumed (i.e., instantaneous realization of the channel is known for both the Rx and the Tx). It is shown that the beamforming is optimal in this setting. This result is obtained by using the Lagrange dual representation of the problem.

Reference [26] explores Rayleigh fading massive MIMO channels with independent identical distribution, and an equal gain transmission (EGT)-based scheme is proposed to satisfy the PACs. When the number of BS antennas converges to infinity for a downlink channel, the paper finds the gap for the following values: (i) the sum-rate achieved by the proposed EGT-based scheme under the PAC; and (ii) the sum-rate achieved by the maximum ratio transmission (MRT) scheme under the TPC. This gap is proportional to the number of users. Note that in massive MIMO, the channel is nearly orthogonal, which will be discussed later in Section 2.4.

Reference [27] looks at two problems in fixed MIMO channels with the per-antenna power constraint: (i) finding the optimal beamformers, which minimizes the per-antenna power under the signal to interference and noise ratio (SINR) constraint;

and (ii) maximizing the region of the achievable rate under the PAC. It is shown that both problems can be solved by their uplink dual representations with uncertain noise[5]. This duality property also holds when there is a power constraint for a subset of antennas. The paper proposes several numerical algorithms to find solutions to the above problems.

In a multiuser MIMO setting, there is a rate for each user, and the maximum of the total rate for reliable communication is the sum capacity, which shows the largest sum rate. A closed-form solution for the optimal sum capacity under the PAC is obtained in [28], where there are two users in a fixed MIMO channel. To do so, the paper shows that the following two problems are equivalent: (i) optimal downlink sum capacity under the PAC; and (ii) optimal uplink sum capacity under the TPC with uncertain noise [27], [29]. Also, the number of active users is obtained. To maximize the sum rate under the PAC, reference [28] gives a closed-form expression for the optimal downlink precoder.

Reference [30] considers two problems in a fixed multi-user MIMO under the PAC: (i) maximizing the weighted sum rate; and (ii) maximizing the minimum individual user rate. Both problems' formulations are non-convex, and hence difficult to solve analytically. For each problem, a numerical algorithm is proposed, and the algorithm's convergence is also proven. The proof is based on the convexity of equivalent problems. It is shown that the results in the paper can also be used for other optimization problems with different power constraints such as per-group power constraint.

The effect of the i-th Tx antenna on mutual information is found in [31]. Based on that, mutual information under the PAC is maximized (i.e., the capacity is achieved) when the transmitted per-antenna powers equal the respective per-antenna constraints. The paper proposes a numerical algorithm to find the capacity, and its convergence to the optimal point is proven. Simulation results show that after around 3 iterations, the algorithm reaches 99% of the capacity.

[5]Here, uncertain noise means that the noise covariance is diagonal and that the diagonal entries depend on the channel and the power constraint. Also, duality means that the gap between the optimal value of the original problem and that of the Lagrange dual problem is zero.

Reference [32] proposes an iterative numerical algorithm to find the sum capacity for a MIMO multiple access channel (MIMO-MAC) under the PAC. This algorithm is based on finding the optimal Tx covariance for a single-user MIMO under the PAC. The convergence of this algorithm to the sum capacity is proven, and simulation examples show that the complexity of the algorithm is linearly dependent on the number of users.

Reference [33] analyzes the sum rate under the PAC by using a zero-forcing (ZF) scheme. For the two-user setting, a closed-form expression for the sum capacity is achieved that is based on the water-filling (WF) procedure.

Maximization of the signal to leakage plus noise ratio (SLNR) under the PAC can also be considered, which is a non-convex problem and is more difficult to solve analytically compared to that under the TPC. Reference [34] finds a semi-closed form solution for the desired precoder, and the robustness of the solution is investigated by considering the imperfectly known CSI at the Tx.

Reference [35] looks at cooperative downlink MIMO channels under the PAC when an MMSE (minimum mean square error) beamformer is used. The paper proposes an iterative algorithm to find the optimal design for the beamformer, and has a lower complexity compared to alternative ones. The suggested algorithm solves a sequence of sub-problems in which the PAC is considered for only one antenna. The convergence of the algorithm to the optimal solution is proven.

2.2.1 Optimal Transmit Covariance Under the Joint Power Constraints

In this thesis, we consider joint power constraints, i.e., combined total and per-antenna power constraints. This is practically important, since the amplifier for each antenna has limited power and the total Tx power is also limited. In this case, the capacity and the optimal Tx covariance are not known in general, except for some special cases (MISO channels, full-rank solution). In the following, we will review some properties of the optimal signaling under the joint constraints

(TPC+PAC), and some numerical algorithms will be reviewed that can find the optimal Tx covariance.

Reference [36] shows that under full CSI, the beamforming is optimal for a fixed MISO channel under the joint power constraints (TPC+PAC). The phase of the beamforming vector matches that of the channel vector, which is also shown in [16]. This is also the case for the following cases: (i) fixed MISO channels under the PAC [19]; (ii) fixed MISO channels under the TPC, PAC, and per-group power constraint [37]; and (iii) MISO wiretap channels under the TPC, under the PAC, or under the joint power constraints (TPC+PAC) with 2 Tx antennas [21]. Reference [36] shows that for MISO channels under the joint constraints, the TPC is active as long as the total power constraint is less than the sum per-antenna power constraints, which is also shown in [16]. This is also observed for MISO wiretap channels [21] and MIMO channels [38] under joint power constraints (TPC+PAC). Reference [36] obtains a closed-form expression for the optimal signaling under the TPC+PAC constraints for the case of 2 Tx antennas. In [36], a numerical algorithm is proposed for a MISO channel in order to find the optimal signaling under the joint constraints. This algorithm is based on the following property for MISO channels: if for some antennas, the OPA under the TPC alone violates the PACs, then their OPAs under the joint power constraints are equal to their per-antenna constraints. This property also holds for a MISO wiretap channel under the joint constraint [21].

Closed-form expressions for the optimal signaling and the capacity for a fixed MISO channels under the joint constraints are given in [16], which completes the result in [36]. For active PACs, the amplitudes of the beamforming vector entries depend only on power constraints, and for the inactive PACs, they depend on the channel gains [16][6]. This observation is different from that under the PAC only, in which the amplitude of the beamforming vector does not depend on the channel vector, as discussed in [19]. In the case of the same PACs, the optimal signaling is a combination of EGT and MRC (maximum ratio combining), where the former is

[6]Active PACs means that the OPA for the receptive antennas is set to be equal to the PACs, while the rest of the antennas are considered as inactive PACs.

responsible for the active PACs and the latter is responsible for the inactive PACs. Also, sufficient and necessary conditions of the optimality of the individual MRC and EGT are found.

Isotropic signaling is optimal under the joint constraints (TPC+PAC) for a fading channel when its distribution is right unitary-invariant[7] [16]. This is the more general result for the fading channel under the PAC, in which the isotropic signaling is also optimal [19].

Reference [17] comments on the ergodic capacity for MIMO channels under the PAC and a long-term power constraint when CSI is available for both the BS and the users[8]. The condition in which the PAC or the long-term TPC are inactive is found. Also, the sub-optimal expression for the optimal signaling is found by the following procedure: to satisfy the PACs, it is assumed that the eigenvalues of the Tx covariance are limited, instead of the diagonal entries. Note that the covariance matrix is Hermitian, and the bounded eigenvalues of a Hermitian matrix indicate the bounded diagonal entries. But, the converse does not necessarily hold [39].

In the case of identical PACs, a full-rank optimal Tx covariance for a fixed MIMO channel under the joint TPC+PAC is obtained in [40]. A necessary condition for the optimality of full-rank Tx covariance is found. The off-diagonal entries of the optimal Tx covariance are the same as those for the inverse of the channel Gram matrix. The diagonal entries of the optimal transmit covariance are the minimum of two terms: (i) the OPA under the TPC only; and (ii) the OPA under the PAC only. However, the optimal dual variable responsible for the TPC under the joint power constraint is not necessarily the same as that in the WF procedure (TPC only). The number of active PACs is obtained, where this increases with the total power constraint. It is shown that in the case of a rank-deficient channel, the optimal Tx covariance may not be unique. Also, having a full-rank channel is not a necessary condition for the optimality of the full-rank Tx covariance. It is

[7]There is no Tx correlation in this model, while Rx correlation is allowed. In this model, an i.i.d. Rayleigh fading MIMO channel is a special case.

[8]Long-term power constraint means that the statistical mean of the Tx power should be less than a certain value that is equivalent to the TPC when the channel is constant.

analytically proven that isotropic signaling is an optimal Tx strategy in a high-SNR regime.

Although the optimal signaling for a fixed MIMO channel under the joint power constraints (TPC+PAC) is still unknown in the general case, reference [38] proposes a numerical algorithm to find the optimal Tx covariance. This algorithm is similar to that in [36], but it is more general and can be used for MIMO channels. The algorithm in [38] produces a decreasing array of rates which converges to capacity after a finite number of steps. The proof of convergence to the capacity is shown.

In addition to the TPC and PAC, one can also consider a per-group power constraint. In this case, to find the amplitude of the optimal beamforming vector in MISO channels, reference [37] proposes an algorithm which is similar to those in [36], and [38]. The case of the full-rank Tx covariance and the full-rank channel is studied in [37], and a closed-form expression for the off-diagonal entries of the optimal Tx covariance is obtained. As well, the diagonal entries can be found by an algorithm which is also similar to those in [36], and [38]. For the case of 2 Tx antennas, a closed-form solution of the optimal signaling is found. For general MIMO channels, an algorithm is proposed to find the optimal signaling, which is shown to have lower complexity in comparison to the known algorithms.

Reference [41] proposes an algorithm to find the optimal Tx covariance that maximizes energy efficiency for MISO channels under the joint power constraints (TPC+PAC). This case is also studied in [18], in which three constraints are assumed: PAC, TPC, and total consumption power constraints. Some algorithms are proposed to find the optimal signaling for both cases of linear and non-linear high-power amplifiers.

The optimal transmission strategy for a MIMO channel under the TPC is well-known, and is on the channel eigenmodes, and the optimal power allocation is via the water-filling procedure. However, there is no closed-form expression for optimal Tx covariance under the PAC or under the joint power constraints (TPC+PAC) in general, while some special cases have been solved (e.g., MISO channels, full-rank

optimal Tx covariance matrix). In this thesis, we consider the problem of finding the optimal Tx covariance for a massive MIMO channel under the joint constraints, which is a convex problem. In this case, convex optimization algorithms can be used, but they provide very limited insights into the problem. Most importantly, the numerical complexity of these algorithms grows in general as $O(m^6)$, where the square Tx covariance has m columns. For example, if m increases 10 times, the complexity increases 10^6 times, and hence these numerical algorithms are complex to implement for a real-time massive MIMO. Therefore, an analytical solution for the optimal Tx covariance is desired.

2.3 Massive MIMO Channels

MIMO channels are normally classified into three groups [4]. The first is point-to-point MIMO channels, in which both transmitter and receiver can have multiple antennas (one user at Tx and one user at Rx). The second is multiuser MIMO, in which a BS serves multiple users and each user may have multiple antennas (shown in Fig. 2.4, where it is assumed that each user has one antenna). In a multiuser setting, each user transmits an independent data stream. The third one is Massive MIMO, which is a type of multiuser MIMO but with the number of BS antennas being much larger than the number of users ($n \gg m$). In this case, simple linear processing techniques are nearly optimal, which is another difference between massive MIMO and conventional multiuser channels [4], [9].

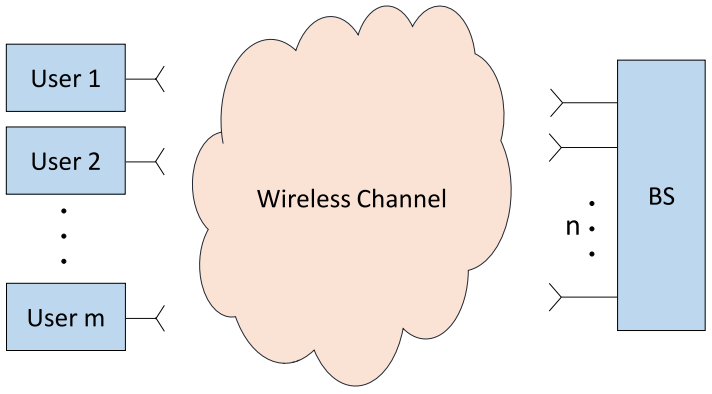

FIGURE 2.4: Gaussian multiuser MIMO channel

Massive MIMO is an indispensable technology for 5G systems [6] in order to compensate for limited bandwidth availability when the number of users of wireless channels will consistently increase [4]. References [42], [43], and [7] describe massive MIMO, where hundreds of BS antennas serve tens of users. A practical example in [43] uses 128 BS antennas for 4 users. In the following, we will review some properties and advantages of massive MIMO channels.

2.3.1 Favorable Propagation and Massive MIMO

A wireless channel is not orthogonal in general, and hence it needs advanced signal-processing schemes to eliminate multiuser interference. However, in massive MIMO and under favorable propagation, a channel is orthogonal. Specifically, favorable propagation (FP) is defined as when the inner products of the different channel vectors are zero [9], and hence simple signal-processing schemes (e.g., MRC, ZF) become optimal.

In the cases of both non-line-of-sight (non-LOS) propagation with i.i.d Rayleigh fading and LOS propagation, it is observed that FP is approximately achieved when the number of BS antennas is large [8] (i.e., massive MIMO)[9]. In a real environment, the propagation likely occurs between LOS and non-LOS, and hence we expect that FP exists in massive MIMO channels [8]. This is also observed in measurement-based results [44] (more details regarding FP will be provided in the next section).

2.3.2 Signal-Processing Schemes

In MIMO channels, non-linear signal processing techniques such as successive interference cancellation (SIC) and dirty-paper coding (DPC) are used when there exists interference between users. It is observed that when $n \approx m$, there is a

[9]In an LOS massive MIMO channel, there are some rare cases in which FP is not obtained (see Section 2.4 for more details). To resolve this problem, user selection techniques can be used, and then the FP conditions are satisfied [7].

significant mismatch between the capacity achieved by DPC/SCI and the spectral efficiency achieved by ZF (i.e., linear signal processing). However, as the number of BS antennas increases, the gap between them decreases [8][10]. Note that there is no interference between users in FP, since they are orthogonal to each other.

As mentioned before, simple signal-processing schemes become optimal in massive MIMO channels under FP. Maximum ratio combining (MRC) is one of them, which maximizes the power of the intended signal, as the MRC processor coherently combines the components of the signal. Zero forcing is another method which removes interference signals from other users with the loss of array gain. Another approach is MMSE, which is a balance between ZF and MRC [8] (see [45], [46] for more detail regarding the optimality of using linear signal processing schemes in massive MIMO).

2.3.3 Spatial Resolution

In massive MIMO channels, there exists a high level of spatial resolution, which follows from using a large number of BS antennas. Hence, in massive MIMO, separating two close users is easier than that in regular MIMO [43][11].

2.3.4 Hardware Impairments

Eliminating the sources of undesired signals at the BS is an important issue in MIMO channels. Specifically, noise at the receiver and interference signals from other users are added to the intended signal. Also, distortion of the intended signal at the receiver due to hardware impairments is undesired. Hardware impairment usually causes an additive distortion at the BS. It is shown that the additive distortion at BS disappears as the number of BS antennas grows [47]. So in

[10]Note that the time division duplex (TDD) mode is used in canonical Massive MIMO channels, which implies a duality between received-signal combining and transmission precoding [9], [8].

[11]The spatial resolution depends on the size of the BS antenna array. Hence, if the size of the antenna array increases as the number of antennas grows, then spatial resolution also increases.

massive MIMO, hardware precision is not required to be high-level compared to that in regular MIMO [8].

2.4 Favorable Propagation

As previously mentioned, channel vectors corresponding to different users are orthogonal in a favorable propagation channel. This means that $\mathbf{H}^{+}\mathbf{H}$ is diagonal, where $\mathbf{H}^{+}$ is the Hermitian conjugate of $\mathbf{H}$. In the following, we will review the literature regarding the existence of FP when a large number of antennas is used. Both theory-based and measurement-based results will be reviewed.

Reference [42] shows that the effects of fast fading can be eliminated by using a large number of BS antennas[12]. To show this, let $\mathbf{H}$ be the channel matrix, as follows:

$$\mathbf{H} = \mathbf{G}\widetilde{\mathbf{G}}^{1/2} \tag{2.3}$$

where $\mathbf{G}$ models fast fading with zero mean and unit variance, and real diagonal $\widetilde{\mathbf{G}}$ represents attenuation and shadow fading[13]. Then:

$$\frac{1}{n}\mathbf{H}^{+}\mathbf{H} = \widetilde{\mathbf{G}}^{1/2}\Big(\frac{1}{n}\mathbf{G}^{+}\mathbf{G}\Big)\widetilde{\mathbf{G}}^{1/2} \tag{2.4}$$

Based on the law of large numbers, when the number of antennas is sufficiently large, we observe that:

$$\frac{1}{n}\mathbf{G}^{+}\mathbf{G} \to \mathbf{I}, \quad \text{as} \quad n \to \infty \tag{2.5}$$

and for random variables, $\to$ denotes convergence in probability. Hence, in massive MIMO, the effect of fast fading vanishes as n grows large. This case is called

[12]In fast fading, the change of channel is significant, as the distance between a signal and receiver varies with the scale of the wavelength [42].

[13]In shadow fading, the change of channel is slow over space [42].

asymptotically favorable propagation (AFP), where $\mathbf{H}^{+}\mathbf{H}$ is approximately diagonal (see Section 3.4 for a precise definition of favorable propagation).

Reference [7] looks at two types of propagation channels: (i) independent Rayleigh fading, and (ii) uniform random LOS. It is shown that AFP is obtained in both environments except for a rare case under LOS. This unlikely case occurs when $\sin(\theta_i) - \sin(\theta_j)$ is linearly dependent on n^{-1} (θ_i denotes the angle of arrival for the i-th user). In this case, in order to achieve AFP, some users should be removed, in which each beam has one user terminal.

Reference [48] analyzes channel orthogonality for three types of antenna arrays when the number of antennas is large: (i) uniform linear arrays (ULA); (ii) uniform planer arrays (UPA); and (iii) uniform circular arrays (UCA). In a correlated fading channel[14], AFP is achieved for ULA and UPA when the sets of dominant paths do not have an intersection with each other. Here, dominant paths are defined when the path gains are linearly dependent on the number of antennas. In LOS propagation, the channel offers AFP for ULA and UPA, but does not offer it for UCA.

Note that the channel models in each paper are not necessarily the same. For example, [7] uses a model of large-scale fading and small-scale fading, and [48] uses the virtual channel model (see [50] for more details regarding the virtual channel model).

To investigate AFP conditions, we need to analyze the normalized inner product of the channel vectors. Reference [51] finds its first and second-order moments when n grows large, as follows:

$$\mathbb{E}\left\{\frac{1}{n}\mathbf{h}_i{}^{+}\mathbf{h}_j\right\} \to \frac{1}{4d_0} \tag{2.6}$$

$$Var\left\{\frac{1}{n}\mathbf{h}_i{}^{+}\mathbf{h}_j\right\} \to \frac{1}{2d_0}\left(1 - \frac{1}{8d_0}\right) - \epsilon, \quad i \neq j \tag{2.7}$$

where $\mathbb{E}\{\cdot\}$ and $Var\{\cdot\}$ denote expectation and variance respectively, correction term ϵ is proportional to d_0^{-2}, and d_0 denotes the space for the BS antennas in

[14] In this case, there exists a correlated fading between each two antennas [49].

the ULA pattern. Here, (2.6) and (2.7) are obtained under uniform random LOS propagation. It can be seen that the mean of the normalized inner product is proportional d_0^{-1}, which means that AFP is obtained as the space for BS antennas increases without limit. More precisely, if $d_0 \to \infty$ as $n \to \infty$, then AFP is achieved for massive MIMO, but if d_0 is fixed as $n \to \infty$ (i.e., the antenna array has limited space), then the channel vectors are not orthogonal to each other. It is worth mentioning that d_0 is proportional to the number of antennas in [48], and hence this observation coincides with the results in [48].

The case of co-located users is studied in [52]. In this paper, users are considered to be located in a circular region, and the authors show that AFP conditions hold as the number of BS antennas grows large. Also, the approximated closed-form expressions for the mean and variance of $n^{-1}\mathbf{h}_i{}^{+}\mathbf{h}_j$ are derived. In this case, the results reveal that a large number of antennas is needed to satisfy the AFP conditions. The following is an example for a 5m cluster radius under LOS propagation when the BS is equipped with a ULA of antennas. The space between the BS and cluster of users is 100 m. For this setting, $\mathbb{E}\{n^{-1}\mathbf{h}_i\mathbf{h}_j\} \approx 0.02$ and $Var\{n^{-1}\mathbf{h}_i\mathbf{h}_j\} \approx 0.03$ for 640 BS antennas. So, in the case of co-located users, a very large number of BS antennas is needed to achieve AFP. Additionally, a very large number of antennas implies a large size, since the spacing between the antenna elements is typically around $\lambda_c/2$, where λ_c is the carrier wavelength. Hence, in order to reduce the size, high-frequency transmission is more preferable (i.e., small λ_c).

In each cell of cellular massive MIMO, there is one BS which serves a subset of users, and the BS has a large number of antennas. Cell-free massive MIMO is another deployment, in which there is a large number of access points (APs). Also, all APs are distributed over an area which serve all users coherently. Reference [53] studies this case, and defines AFP for cell-free massive MIMO as follows:

$$\frac{\mathbf{h}_i^{+}\mathbf{h}_j}{\left(\mathbb{E}\{||\mathbf{h}_i||^2|\mathbf{d}_i\}\mathbb{E}\{||\mathbf{h}_j||^2|\mathbf{d}_j\}\right)^{1/2}} \to 0 \quad \text{as} \quad n \to \infty, \quad i \neq j \tag{2.8}$$

where $\mathbb{E}\{||\mathbf{h}_i||^2|\mathbf{d}_i\}$ denotes conditional expectation; $||\cdot||$ denotes the Euclidean norm; n is the total number of antennas (the number of APs multiplied by the number of antenna per AP); and $\mathbf{d}_i$ denotes a distance vector from the i-th user to all APs' antennas. The paper shows that the AFP condition in (2.8) is achieved as n increases (whether by increasing the number of APs or antennas per AP). The normalized inner product in (2.8) decreases as the path loss becomes smaller or the distance between users increases.

Reference [54] considers a sparse massive MIMO and analyses the normalized inner product of the different channel vectors. The analyses reveal that the first and second moments of this inner product become small for a sufficiently large number of antennas.

Reference [55] shows that AFP conditions are satisfied in most cases of semi-correlated Ricean channels (which is a more general propagation model compared to i.i.d. fading and LOS). To show this, a mean value of inter-user interference power is analyzed. However, if there exists alignment between LOS component and users' covariances, then AFP conditions do not hold in massive MIMO. In these unlikely scenarios, the alignment vanishes by removing correlated users.

To determine how close the channel is to AFP conditions, reference [56] proposes a cumulative distribution function (CDF) of the normalized inner product of the channel vectors:

$$Pr_\theta = \Pr\left\{\frac{|\mathbf{h}_i{}^{+}\mathbf{h}_j|}{n} > \theta\right\} \tag{2.9}$$

where the right-hand side is the probability that the normalized inner product is greater than or equal to θ. As n grows large, if this CDF converges to zero for all $\theta > 0$, then the channel offers AFP. The paper shows that an upper bound of this CDF becomes smaller when the spatial correlation between the users decreases[15].

[15] An upper bound of this CDF is found by using Chebychev's inequality [56], [57].

So far, we have reviewed some theory-based results which show that AFP is attained in massive MIMO. Next, we will review some measurement-based results regarding this subject.

Some papers determine the singular value spread (SVS) in order to to investigate the channel orthogonality in massive MIMO. This is defined as the ratio of the largest channel singular value to the smallest one. If this ratio is one, then the channel is orthogonal (i.e., favorable propagation). When determining a channel's orthogonality, using SVS can be misleading in some cases. For example, consider a diagonal channel matrix where the ratio of maximum and minimum diagonal entries is large. In this case, the channel is orthogonal, but the SVS is large. Another example is a rank-deficient orthogonal channel in which SVS is infinite [58] (since the minimum singular value is zero). In both examples, channels offer favorable propagation (even without $n \to \infty$), but the SVS is much larger than one. Note that to make sure that channel norms do not affect the SVS, the channel needs to be normalized [7].

Reference [44] obtains measurement-based results for massive MIMO for these settings: (i) all users are located close to each other with LOS propagation or non-LOS propagation, and (ii) LOS propagation while the distance between users is greater than that in the previous case. For both cases, it is seen that SVS decreases as the number of BS antennas increases. The smallest SVS occurs for non-LOS propagation with rich scattering, even when the users are close to each other. On the other hand, LOS propagation with near-located users has the largest SVS. With the same number of BS antennas, the measurement show that SVS has a smaller value with the ULA antenna pattern compared to UCA. This is expected, since ULA provides more angular resolution in comparison with UCA when the number of antennas and the spacing between antenna elements are the same in both ULA and UCA. However, in this case, the size of the ULA is larger than that of the UCA. The measurements in [43] also show that SVS decreases as the number of BS antennas grows.

Now let us show the results of a practical setting at 2.6 GHz for massive MIMO [59]. In this case, the antennas are in the ULA pattern, and the distance between any given two is half of the carrier wavelength. In both non-LOS and LOS, SVS decreases as the number of antennas increases from 20 to 128. The CDF of SVS in both types of propagations are almost similar to that in an i.i.d channel. However, the SVS in LOS propagation is larger than that in non-LOS. This is expected because the theory-based results for massive MIMO also show that favorable propagation is not achieved in some extreme cases of LOS propagation.

The following is another example from a real situation [60] where there are 6 single-antenna users, 3 of them indoors and 3 outdoors. The BS is indoors with a circular configuration and $f_c = 2.6$ GHz, where f_c denotes the carrier frequency. It is shown that with 6 BS antennas, the difference between the largest and smallest channel singular values is about 26dB. This difference reduces to 7dB when the BS is equipped with 128 antennas. So, the SVS decreases as the number of BS antennas increases from 6 to 128.

Reference [61] looks at massive MIMO at a millimeter-wave band, and obtains some results based on real measurement data of $f_c = 60$ GHz with 100 BS antennas. In this setting, the SVS is about one when BS serves 4 users. However, this becomes around three for 8 users. This shows that the SVS is small as long as the number of BS antennas is large compared to the number of users. The measured data in [58] also show that as the number of users grows, more BS antennas are needed in order to have a smaller SVS.

Channel orthogonality (i.e., favorable propagation) can also be determined by the value of the correlation coefficient between any given two distinct channel vectors. This is defined as follows:

$$c_{i,j} = \frac{|\mathbf{h}_i^{+}\mathbf{h}_j|^2}{||\mathbf{h}_i||^2.||\mathbf{h}_j||^2} \tag{2.10}$$

where $c_{i,j} = 0$ indicates the orthogonality of $\mathbf{h}_i$ and $\mathbf{h}_j$. Reference [62] used this coefficient for real measurement data, with ULA (vertical and horizontal) and UPA antenna patterns. In both indoor and outdoor settings, the correlation coefficient

becomes smaller as the number of BS antennas grows (this observation is also seen from the measurement data in [58]). For example, in the indoor case, $c_{i,j} \approx 0.1$ with 10 BS antennas, and reduces to 0.02 with 64 BS antennas. Also, $c_{i,j}$ from the measurement data is almost the same as that in the i.i.d. channel, which is considered as a standard reference. Based on this investigation, the horizontal antenna array yields the smallest value of correlation coefficient, and the vertical antenna array yields the largest.

It is worth mentioning that (2.6)-(2.10) are measures to quantify the orthogonality of the channel. For example, let us assume that all vectors are normalized such that $||\mathbf{h}_i|| = \sqrt{n}$ (note that $\mathbf{h}_i$ has n elements). Then:

$$\frac{\mathbf{h}_i^+ \mathbf{h}_j}{||\mathbf{h}_i||.||\mathbf{h}_j||} = \frac{1}{n}\mathbf{h}_i^+ \mathbf{h}_j \tag{2.11}$$

where the magnitude of the left-hand side in (2.11) is the square root of the correlation coefficient in (2.10), and the exception and the variance on the right-hand side are the same as in (2.6) and (2.7) respectively.

Reference [63] shows that the ergodic sum rate is maximized when the channel is orthogonal (i.e., favorable propagation). In the paper, the measured data is collected for both indoor and outdoor settings at $f_c = 5.6$ GHz in the LOS propagation. The channel matrix is normalized to get a more valid result, and it is seen that the ergodic sum rate gets close to its upper bound by increasing the number of BS antennas. Since this upper bound is achieved in FP, we can see that FP is obtained as the number of antennas increases. As a practical example, the ergodic sum rate reaches around 95% of its upper bound (95% of the ergodic sum rate under FP conditions) for the following setting: the low-SNR regime when number of BS antennas is around 6 times of the number of users. An empirical formula is proposed in order to see how close the channel is to FP.

Measurement data for the channel Gram matrix in the two-user setting is reported in [64]. It is observed that the ratio of its off-diagonal entry to the diagonal entry decreases as the number of antennas grows. In other words, the Gram matrix becomes nearly diagonal for a sufficiently large number of BS antennas (i.e., the

channel matrix becomes nearly orthogonal). In this measurement, the carrier frequency is 2.6 GHz (i.e., cellular frequency).

Favorable propagation has been studied in detail both theoretically and experimentally. In particular, the channel offers approximately FP in many cases of massive MIMO as long as the number of BS antennas (or APs antennas) is sufficiently large compared to the number of users. These cases include both LOS and non-LOS propagation as well as both millimeter-wave and cellular bands. In some scenarios such as co-located users, more BS antennas are needed to obtain approximately FP. In this case, or when n/m is not sufficiently large, a proper antenna pattern such as ULA or UPA can be used in order to have FP. These antenna patterns should have high angular resolutions. Nevertheless, based on theory and measurement-based results, when $n \geq 10m$, the channel propagation is reasonably close to favorable propagation (i.e., orthogonal channel) in a massive MIMO. There are some rare cases where FP is not achieved in a massive MIMO; for example, when two users are in exactly the same direction as seen from the BS. It is shown that to obtain FP in these scenarios, user selection techniques can be used.

2.4.1 Summary

Multiple-input multiple-output (MIMO) wireless channels bring benefits such as increased data rate. The maximum achievable rate over a noisy channel is called channel capacity, at which reliable communication is achieved. In practical systems, there are always some constraints that limit the channel capacity, of which one is limited power supply at the transmitter, which implies total transmit power constraint (TPC). Another is the per-antenna power constraint (PAC), which is practically important since each antenna has its own amplifier with a limited power budget. The capacity and optimal signaling for a fixed MIMO channel under the joint total and per-antenna power constraints are not known in general, except in some special cases (e.g., full-rank transmit covariance, MISO channels). These

known solutions, in addition to some optimal signaling properties, have been reviewed in this chapter. Also, some numerical algorithms have been reviewed that can be used to obtain the optimal signaling under the joint power constraints (TPC+PAC).

Massive MIMO is a type of multi-user MIMO in which the number of BS antennas is large. For many propagation scenarios, theory-based and measurement-based results show that the channel becomes approximately orthogonal in massive MIMO. The channel orthogonality is known as favorable propagation (FP). For example, asymptotically favorable propagation is obtained for many cases in LOS, Rayleigh fading, and Ricean massive MIMO channels. Measurement data show that favorable propagation can be obtained for massive MIMO channels for both millimeter-wave and cellular bands.

Chapter 3

Channel Model and Optimal Signaling

The model of a discrete-time memoryless fixed Gaussian MIMO channel is as follows[1]:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \tag{3.1}$$

where $\mathbf{y} \in \mathbb{C}^{n,1}$, $\mathbf{x} \in \mathbb{C}^{m,1}$ are the received and transmitted signals; $\boldsymbol{\xi} \in \mathbb{C}^{m,1}$ is the i.i.d. circularly-symmetric additive white Gaussian noise at the receiver of unit variance (hence, the signal power is also the SNR); $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_m] \in \mathbb{C}^{n,m}$ is the channel gain matrix known to both the transmitter (Tx) and the receiver (Rx) and $\mathbf{h}_i$ is the i-th column of $\mathbf{H}$; we assume that $\mathbf{h}_i \neq 0$ for all i[2]; $(\mathbf{H})_{ij}$ (i.e., the ij-th entry of $\mathbf{H}$) is the channel gain from the j-th Tx antenna to the i-th Rx antenna; the channel is fixed over the period of transmission while it may change for the next transmission; m and n are the number of antennas at the transmitter and the receiver, respectively. Fig. 3.1 shows an example of a MIMO channel when there are 2 Tx antennas and 4 Rx antennas.

[1]Here, no convolution is needed, since this is a model of memoryless (frequency flat) channel. Also, time index t is removed for the simplicity (see (2.1)).

[2]Since $\mathbf{h}_i = 0$ means that the i-th Tx antenna is disconnected, and hence it does not affect the channel capacity.

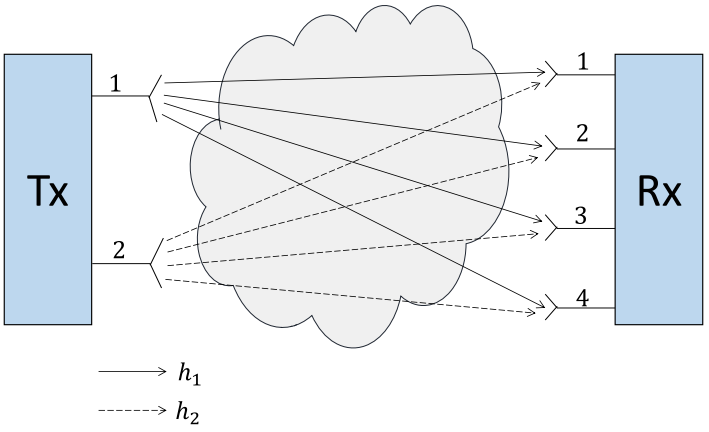

FIGURE 3.1: MIMO channels with 2 Tx antennas and 4 Rx antennas

Gaussian signaling was proven to be optimal in this case [15]. For a given channel matrix and Tx covariance matrix $\mathbf{R} = \mathbb{E}\{\mathbf{x}\mathbf{x}^+\} \in \mathbb{C}^{m,m}$, the maximum achievable rate is as follows:

$$C(\mathbf{W}, \mathbf{R}) = \ln |\mathbf{I} + \mathbf{W}\mathbf{R}| \tag{3.2}$$

where $\mathbf{W} = \mathbf{H}^+\mathbf{H} \in \mathbb{C}^{m,m}$ is the Gram matrix of $\mathbf{H}$. The channel capacity for a given $\mathbf{W}$ is an optimization problem over $\mathbf{R}$, as follows:

$$C(\mathbf{W}) = \max_{\mathbf{R} \in S_R} \ln |\mathbf{I} + \mathbf{W}\mathbf{R}| \tag{3.3}$$

where S_R is the feasible set of Tx covariance matrices. The optimal Tx covariance $\mathbf{R}^*(\mathbf{W})$ for channel $\mathbf{W}$ is:

$$\mathbf{R}^*(\mathbf{W}) = \arg\max_{\mathbf{R} \in S_R} \ln |\mathbf{I} + \mathbf{W}\mathbf{R}| \tag{3.4}$$

The eigenvectors of $\mathbf{R}^*(\mathbf{W})$ determine the optimal beam directions and its eigenvalues give the optimal power allocation for these beams [2]. The optimal Tx covariance matrix $\mathbf{R}^*(\mathbf{W})$ also determines a linear precoder at the transmitter (see [65] for more details).

3.1 Total Power Constraint (TPC)

The consideration of the TPC is important, since there is a limited power supply, and also the transmit power affects the battery life. The feasible set under the TPC is:

$$S_R = \{\mathbf{R} : \mathbf{R} \geq 0, tr\mathbf{R} \leq P_T\} \tag{3.5}$$

where P_T is the maximum total Tx power. In this case, the optimal signaling is on the right singular vectors of $\mathbf{H}$, and the optimal power allocation (OPA) is obtained by the standard water-filling (WF) procedure [15] that will be reviewed below.

To find the optimal Tx covariance under the TPC, we use eigenvalue decomposition (EVD). The EVD of $\mathbf{W}$ is as follows:

$$\mathbf{W} = \mathbf{U}_W \mathbf{\Lambda}_W \mathbf{U}_W^+ \tag{3.6}$$

where $\mathbf{U}_W$ is a unitary matrix that consists of the eigenvectors of $\mathbf{W}$ (which are also the right singular vectors of $\mathbf{H}$), and $\mathbf{\Lambda}_W$ is the diagonal matrix, its diagonal entries are equal to the eigenvalues of $\mathbf{W}$, as follows:

$$\mathbf{\Lambda}_W = diag\{\lambda_1(\mathbf{W}), \lambda_2(\mathbf{W}), ..., \lambda_m(\mathbf{W})\} \tag{3.7}$$

where $\lambda_i(\mathbf{W})$ is the i-th eigenvalue of $\mathbf{W}$ and we assume that they are in decreasing order. The optimal Tx covariance matrix $\mathbf{R}^*(\mathbf{W})$ under the TPC is as follows:

$$\mathbf{R}^*(\mathbf{W}) = \mathbf{U}_W \mathbf{\Lambda}^* \mathbf{U}_W^+ \tag{3.8}$$

where $\mathbf{\Lambda}^* = diag\{p_{i,WF}^*\}$ is the diagonal matrix in which the diagonal entries are determined by the water-filling (WF) procedure, as follows:

$$p_{i,WF}^* = \left(\mu_{WF}^{-1} - \lambda_i^{-1}(\mathbf{W})\right)_+ \tag{3.9}$$

where $(x)_+ = \max(x, 0)$, $p^*_{i,WF}$ is the OPA for the i-th stream under the TPC and μ_{WF} can be obtained from the TPC:

$$tr\mathbf{R}^*(\mathbf{W}) = \sum_{i=1}^{m} \left(\mu_{WF}^{-1} - \lambda_i^{-1}(\mathbf{W})\right)_+ = P_T \tag{3.10}$$

Finding μ_{WF} from the above equality can be illustrated as pouring water into a container (see Fig. 3.2). The container has a floor profile in which the height of each level is determined by $\lambda_i^{-1}(\mathbf{W})$, and the total amount of water is determined by P_T. As can be seen from this figure, the deeper levels get more water compared to the shallower ones. In other words, larger channel eigenmodes (which correspond to stronger streams) get more power in comparison to the smaller ones (which represent weaker streams). The optimal power allocation can be represented by the height of the water as measured from the surface of the stairs, and the "water level" μ_{WF}^{-1} is found by the height of water from the ground [66].

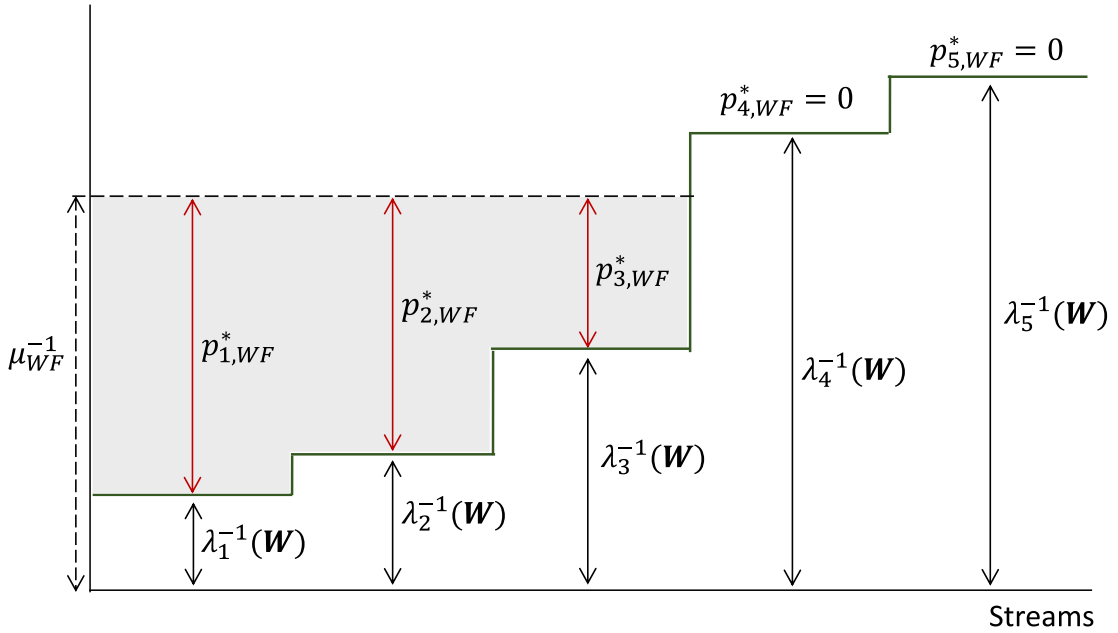

FIGURE 3.2: Geometric interpretation of the water-filling procedure

3.2 Per-Antenna Power Constraint (PAC)

Each antenna at the transmitter is connected to its own amplifier, which has limited power. This motivates the consideration of the PAC. In this case, the feasible set takes the form

$$S_R = \{\mathbf{R} : \mathbf{R} \geq 0, (\mathbf{R})_{ii} \leq P_{1i}\} \tag{3.11}$$

where P_{1i} is the power constraint for the i-th antenna. While the problem has a deceivingly simple appearance, no closed-form solution for optimal Tx covariance is known in general, but only in some special cases (e.g., MISO channels, rank-1 channels, full-rank solution).

3.3 Joint Power Constraints (TPC+PAC)

The case of the joint constraints (TPC+PAC) is motivated by practical issues. As mentioned in the previous chapter, the TPC is important to consider since there is a limited power budget at the transmitter. Here, the PAC is important since each antenna has its own amplifier with limited power. Here, we consider the joint power constraints with the following feasible set

$$S_R = \{\mathbf{R} : \mathbf{R} \geq 0, tr\mathbf{R} \leq P_T, (\mathbf{R})_{ii} \leq P_{1i}\} \tag{3.12}$$

In this thesis, we use S_R as in (3.12), unless stated otherwise.

The optimal Tx covariance for the problem in (3.3) under the joint power constraints in (3.12) is still an open problem, and a solution is known only for some special cases such as MISO channels and full-rank solution. Since the problem is convex, the optimal solution can be found by using iterative algorithms such as the Newton-barrier method [66] or the ones that were discussed in the literature review in the previous chapter. However, these numerical algorithms give us

very limited engineering insights into the problem, and their numerical complexity grows in general as $O(m^6)$, meaning that they have high complexity for large m.

3.4 Favorable Propagation (FP)

Massive MIMO is a new technology that has promising advantages, including high-rate communication for multi-user systems [4]. In particular, simple linear processing schemes become optimal under favorable propagation [9]. Specifically, favorable propagation is defined as follows [9]:

$$\mathbf{h}_i^+ \mathbf{h}_j = 0, \quad i \neq j \tag{3.13}$$

In other words, the column vectors of the channel matrix $\mathbf{H}_{n\times m}$ are orthogonal to each other, and hence $\mathbf{W}$ becomes diagonal, and $\mathbf{W} = \mathbf{H}^+\mathbf{H} = \mathbf{D}_W$. In practice, these conditions are rarely satisfied exactly, but only approximately, which we term nearly (or approximately)-favorable propagation (NFP):

$$\mathbf{h}_i^+ \mathbf{h}_j \approx 0, \quad i \neq j \tag{3.14}$$

The main difference between FP and NFP is characterized as follows: under FP, all off-diagonal entries of $\mathbf{W}$ are zero, while under NFP they are non-zero but small. While (3.13) and (3.14) are very restrictive for standard MIMO (small n), they are not restrictive but quite typical for massive MIMO, as demonstrated in the literature. Both theory-based and measurement-based results show the existence of approximately favorable propagation for many cases of massive MIMO channels, as discussed in the literature review. Specifically, the measurement-based results in [8], [44], [63], and [64] show the existence of approximately favorable propagation in cellular and millimeter-wave bands for massive MIMO, and the theory-based results in [42], [7], [52], and [55] show that favorable propagation is approximately satisfied for many cases in Rayleigh fading, LOS, and Ricean massive MIMO

channels. Hence, while considering an orthogonal channel $\mathbf{W} = \mathbf{D}_W$ for regular MIMO is a significant restriction, this is not the case for massive MIMO.

Favorable propagation can also be characterized asymptotically in a massive MIMO setting:

$$\frac{1}{n}\mathbf{h}_i^+\mathbf{h}_j \to 0, \quad i \neq j, \quad \text{as} \quad n \to \infty \tag{3.15}$$

which is called asymptotically favorable propagation (AFP) [9], where $\mathbf{W}$ becomes approximately diagonal for sufficiently large but finite n.

3.5 Summary

In this chapter, we described the model for discrete-time fixed Gaussian MIMO channels that is the basis of our analyses. Three different power constraints are considered. The first is the total power constraint (TPC), which is motivated by a limited transmit power budget. The optimal signaling under the TPC is on the channel eigenmodes, and the optimal power allocation can be obtained by the standard water-filling procedure. The second is the per-antenna power constraint (PAC), which is motivated by practical design, in which each Tx antenna has its own amplifier with a limited power budget. The optimal Tx covariance under the PAC is not known in general, except in some special cases (e.g., MISO channel, full-rank optimal Tx covariance). The third is the combined TPC and PAC, which is practically important and is motivated by the limited power in transmitter and per-antenna amplifiers. The optimal Tx covariance under the joint power constraints (TPC+PAC) is not known in general, except in some special cases (e.g, full-rank optimal signaling, MISO channels). We formulated the problem, which involves finding the optimal Tx covariance under favorable propagation and the joint total and per-antenna power constraints.

In favorable propagation (FP), the channel becomes orthogonal. In practice and in massive MIMO channels, the favorable propagation conditions can hardly be

satisfied exactly, but only approximately. We term this case as nearly-favorable propagation (NFP), in which the off-diagonal entries of the channel Gram matrix are small but non-zero. Additionally, FP can be imposed asymptotically, which is known as asymptotically favorable propagation (AFP). In this case, the normalized inner product of each two distinct channel vectors converges to zero as the number of antennas grows.

Chapter 4

Optimal Signaling Under Favorable Propagation

In this chapter, we consider a massive MIMO channel under favorable propagation and obtain a closed-form expression for the optimal transmit covariance under the joint power constraints (i.e., the total power constraint and the per-antenna power constraints). We discuss the difference between the optimal power allocation (OPA) under the joint constraints with that under the TPC alone (water-filling procedure). Under the joint constraints, a geometric interpretation of the OPA is proposed, as well as an iterative water-filling algorithm in order to compute the OPA.

4.1 Optimal Signaling Under the Joint Total and Per-Antenna Power Constraints

Here, we find the optimal Tx covariance under favorable propagation as well as the joint constraints (TPC+PAC). To be more precise, we consider the following

problem:

$$C(\mathbf{D}_W) = \max_{\mathbf{R}\in S_R} \ln |\mathbf{I} + \mathbf{D}_W\mathbf{R}| \tag{4.1}$$

where

$$S_R = \{\mathbf{R} : \mathbf{R} \geq 0,\, tr\mathbf{R} \leq P_T,\, (\mathbf{R})_{ii} \leq P_{1i}\} \tag{4.2}$$

We assume the channel to be orthogonal (i.e., $\mathbf{W} = \mathbf{H}^{+}\mathbf{H} = \mathbf{D}_W$) because of the FP, and S_R is the feasible set of Tx covariance matrices (TPC+PAC). The next theorem gives the optimal Tx covariance for the problem in (4.1). To this end, let $g_i = ||\mathbf{h}_i||^2 > 0$ and $p_i \geq 0$ be the i-th diagonal entry for $\mathbf{W}$ and $\mathbf{R}$ respectively[1].

Theorem 1. *Consider an orthogonal channel, $\boldsymbol{W} = \boldsymbol{D}_W = diag\{g_i\}$. The unique optimal Tx covariance $\boldsymbol{R}^*(\boldsymbol{D}_W)$ under the joint power constraints in* (4.2) *for the problem in* (4.1) *is the diagonal matrix, as follows:*

$$\boldsymbol{R}^*(\boldsymbol{D}_W) = diag\{p_i^*\} \tag{4.3}$$

where p_i^* is the optimal power allocation for the i-th Tx antenna:

$$p_i^* = \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} \tag{4.4}$$

where $\mu \geq 0$ is the Lagrange multiplier responsible for the total Tx power: (i) if $P_T \geq \sum_{i=1}^{m} P_{1i}$, then $\mu = 0$ and all PACs are active, i.e. $p_i^* = P_1$; (ii) if $P_T < \sum_{i=1}^{m} P_{1i}$, and then $\mu > 0$ can be found as a solution of the following equation (the TPC):

$$\sum_{i=1}^{m} \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} = P_T \tag{4.5}$$

[1]The results which are based on channel orthogonality also apply for an orthogonal frequency division multiplexing (OFDM) system.

The channel capacity is:

$$C(\boldsymbol{D}_W) = \sum_{i=1}^{m} \ln(1 + g_i p_i^*) \tag{4.6}$$

Proof. See Appendix. □

From Theorem 1, we can also find the optimal Tx covariance under the individual TPC or PAC when $\mathbf{W} = \mathbf{D}_W$. To obtain the OPA under the TPC only from Theorem 1, we need to consider $P_{1i} \geq P_T$, and then the PACs become redundant. Therefore, the effect of the PACs is removed. To see this, we consider the OPA in (4.4) and take the limit for $P_{1i} \to \infty$. Then, the first term in min in (4.4) (i.e., P_{1i}) disappears. Hence, under the TPC, the diagonal entries of the optimal Tx covariance matrix are found by:

$$p_{i,WF}^* = (\mu_{WF}^{-1} - g_i^{-1})_+ \tag{4.7}$$

where $p_{i,WF}^*$ is the OPA for the i-th stream under the TPC only (the standard WF procedure [15], [66]), and μ_{WF} is the optimal dual variable responsible for the total power constraint under the TPC.

On the other hand, to obtain the optimal power allocation under the PAC only, we need to consider $P_T \geq \sum_{i=1}^{m} P_{1i}$ in order to eliminate the effect of the TPC. In this case, $\mu = 0$ and $\min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} = P_{1i}$, and hence the TPC becomes redundant (i.e., the second term in min disappears). Therefore, under the PAC, the diagonal terms for the optimal Tx covariance matrix are $p_{i,PAC}^* = P_{1i}$. Here, $p_{i,PAC}^*$ is the OPA for the i-th stream under the PAC only.

From Theorem 1, we observe that independent signaling is optimal under the joint power constraints (TPC+PAC) and FP. Note that both $\mathbf{R}^*(\mathbf{D}_W)$ and $\mathbf{D}_W$ are diagonal matrices, so that the optimal Tx covariance and the Gram matrix of the channel have the same eigenvectors. While this is not the case under the joint power constraints for general MIMO channels [16], it is always the case under the TPC only. Theorem 1 shows that the eigenvectors of the optimal Tx covariance

matrix are the standard basis vectors, so the need for feedback is significantly reduced since the optimal power allocation p_i^* is only required at Tx.

It is seen that the OPA in (4.4) is the minimum of two terms; one represents the PACs (i.e., P_{1i}), and the second represents the standard WF procedure (i.e., under the TPC alone $(\mu^{-1} - g_i^{-1})_+$). However, the water level μ^{-1} under the joint power constraints (TPC+PAC) is not necessarily the same as that under the TPC, except when the optimal signaling under the joint power constraints coincides with the water-filling procedure (i.e., all PACs are inactive).

The non-linear equation in (4.5) can be solved via a bisection algorithm, since its left-hand side is decreasing in μ. We will provide more details about the bisection algorithm in Chapter 6.

The next corollary gives the actual transmitted power under the joint power constraints, i.e., the value of $tr\mathbf{R}^*(\mathbf{D}_W)$.

Corollary 1. *In Theorem 1, the actual transmitted power is as follows:*

$$\sum_{i=1}^{m} p_i^* = \min\left\{P_T, \sum_{i=1}^{m} P_{1i}\right\} \tag{4.8}$$

Proof. Follows from Theorem 1 by considering two possible cases: $P_T \geq \sum_{i=1}^m P_{1i}$, in which the TPC is inactive, and $\sum_{i=1}^m p_i^* = \sum_{i=1}^m P_{1i} \leq P_T$ or $P_T < \sum_{i=1}^m P_{1i}$, in which the TPC is active, and $\sum_{i=1}^m p_i^* = P_T < \sum_{i=1}^m P_{1i}$. □

In the standard WF procedure under the TPC alone, the actual transmitted power is always equal to P_T. However, it is not necessarily the case under the joint power constraints.

The next proposition gives the number of active streams (i.e., streams with $p_i^* > 0$).

Proposition 1. *Let $\{g_i\}$ be in decreasing order, i.e., $g_1 \geq g_2$.... For the problem in* (4.1)*, the number m_+ of active streams is as follows:*

$$m_+ = \max\{j : u_j < P_T\} \tag{4.9}$$

where:

$$u_j = \sum_{i=1}^{j} \min\left\{P_{1i}, (g_j^{-1} - g_i^{-1})\right\}, \quad 1 \leq j \leq m, \tag{4.10}$$

and $p_i^* > 0$ *for* $1 \leq i \leq m_+$ *and* $p_i^* = 0$ *for* $i > m_+$.

Proof. See Appendix. □

It should be pointed out that u_j is increasing in j. So, (4.9) facilitates an algorithmic implementation for finding the number m_+ by verifying the inequality for increasing j. Then, the algorithm stops at the largest j, satisfying the inequality, starting from $j = 1$.

Since $\mathbf{R}^*(\mathbf{D}_W)$ is a diagonal matrix, its rank is equal to the number of non-zero diagonal entries. Hence, $rank\big(\mathbf{R}^*(\mathbf{D}_W)\big) = m_+$. From Proposition 1, active streams are stronger than inactive ones, i.e., g_i is larger for active streams. This is also seen in the WF procedure (see Chapter 3 for more details regarding the WF procedure).

In the next proposition, a sufficient and necessary condition for the optimality of beamforming is given.

Proposition 2. *Let* $\{g_i\}$ *be in decreasing order. For the problem in* (4.1)*, the rank-1 transmission (beamforming) is optimal, i.e.,* $m_+ = 1$ *if and only if:*

$$P_T \leq \min\left\{P_{11}, (g_2^{-1} - g_1^{-1})\right\} \tag{4.11}$$

In this case, $\boldsymbol{R}^*(\boldsymbol{D}_W) = P_T \boldsymbol{e}_1 \boldsymbol{e}_1^T$.

Proof. Follows from Proposition 1. □

Under (4.11), $p_1^* = P_T$, $p_i^* = 0$ for any $i \geq 2$. This shows that only one stream is active, and that is the strongest one, i.e., $g_1 = ||\mathbf{h}_1||^2$.

Under the TPC only, the beamforming is optimal if and only if:

$$P_T \leq (g_2^{-1} - g_1^{-1}) \tag{4.12}$$

and this follows from (4.11) when $P_{1i} \geq P_T$ (i.e., the PACs become redundant and the first term in min in (4.11) disappears). By comparing (4.12) and (4.11), we can see that the optimality of rank-1 transmission under the joint power constraints is similar to that under the TPC only, and both occur at low-SNR (small P_T). However, the SNR threshold under the joint power constraint is lower than that under the TPC alone. If the beamforming is optimal under the joint power constraints, it is also optimal under the TPC only, but the converse is not true in general (i.e., (4.11) leads to (4.12), but not vice versa in general).

The next proposition gives a sufficient and necessary condition for the optimality of full-rank transmission.

Proposition 3. *For the problem in* (4.1)*, the full-rank transmission is optimal, i.e.,* $m_+ = m$ *(*$p_i^* > 0$ *for any* i*) if and only if:*

$$P_T > \sum_{i=1}^{m} \min\{P_{1i}, g_m^{-1} - g_i^{-1}\} \tag{4.13}$$

Proof. Follows from Proposition 1. □

Under the TPC only, the full-rank transmission is optimal if and only if:

$$P_T > \sum_{i=1}^{m} (g_m^{-1} - g_1^{-1}) \tag{4.14}$$

where (4.14) can be obtained from (4.11) when $P_{1i} \geq P_T$. By comparing (4.13) and (4.14), it is observed that the optimality of full-rank transmission occurs at high SNR for both scenarios. However, the SNR threshold is lower for the case of the joint power constraints. If all streams are active under the TPC alone (the WF procedure), then they are also active under the joint power constraints (TPC+PAC), but the converse is not necessarily true. However, at low SNR, if the beamforming is optimal under the joint power constraints, then it is also optimal under the TPC only.

In the next proposition, the number of active PACs is determined in a closed form.

Proposition 4. *Consider active streams, i.e., $p_i^* > 0$, and arrange them so that $\{g_i^{-1} + P_{1i}\}$ is in increasing order, i.e., $g_1^{-1} + P_{11} \leq ... \leq g_{m_+}^{-1} + P_{1m_+}$. For the problem in (4.1), the number m_{PAC} of active PACs is as follows:*

$$m_{PAC} = \max\left\{j : v_j \geq \sum_{i=1}^{m_+} P_{1i} - P_T\right\} \tag{4.15}$$

where:

$$v_j = \sum_{i=j}^{m_+} \big((g_i^{-1} + P_{1i}) - (g_j^{-1} + P_{1j})\big), \quad 1 \leq j \leq m_+, \tag{4.16}$$

and $p_i^* = P_{1i}$ for $1 \leq i \leq m_{PAC}$ and $p_i^* < P_{1i}$ for $i > m_{PAC}$. No PAC is active, i.e., $p_i^* < P_{1i}$ for any i, if and only if:

$$v_1 < \sum_{i=1}^{m_+} P_{1i} - P_T \tag{4.17}$$

and $p_i^* = P_{1i}$ for any i, i.e. $\boldsymbol{R}^*(\boldsymbol{D}_W) = diag\{P_{1i}\}$, if and only if:

$$P_T \geq \sum_{i=1}^{m} P_{1i} \tag{4.18}$$

Proof. See Appendix. □

Note that v_j is decreasing in j, so that m_{PAC} can be obtained by verifying the inequality for increasing j.

Proposition 4 shows that by increasing P_T, at first the PACs become active for the streams with smaller $g_i^{-1} + P_{1i}$, and hence no more power can be given to those. Then, the excess power is re-distributed to other streams with larger $g_i^{-1} + P_{1i}$. This shows the main difference between the standard WF procedure and that in (4.4), which we call here capped WF (see Figs. 4.2-4.4).

Next, we show an unusual property of the optimal dual variable μ under the joint power constraints (TPC+PAC).

Corollary 2. *If $P_T < \sum_{i=1}^{m} P_{1i}$, the dual variable μ is not necessarily unique, while the optimal power allocation is always unique.*

Proof. The proof of the uniqueness of the optimal solution is based on the strict convexity of the problem (see Appendix for the proof of Theorem 1 regarding the strict convexity of the problem). To show the non-uniqueness of μ, let us consider the following example: $\{g_i\} = \{200, 100, 1, 0.5\}$, $P_T = 1$, and $P_{1i} = 0.5$ for any i. Here, $u_2 = 1/200 < P_T$ and $u_3 = 1 = P_T$, so that (4.9) implies $m_+ = 2$. To find m_{PAC} from Proposition 4, note that $\sum_{i=1}^{m_+} P_{1i} - P_T = 0$ and $v_2 = 0$, so that (4.15) implies $m_{PAC} = 2$. In this example, the number of active streams equals the number of active PACs, i.e., $m_+ = m_{PAC} = 2$, so the unique optimal power allocation for this channel is as follows:

$$p_i^* = \begin{cases} 0.5 & \text{for} \quad i = 1, 2 \\ 0 & \text{for} \quad i = 3, 4 \end{cases} \tag{4.19}$$

and from Theorem 1, observe that:

$$\min\left\{0.5, (\mu^{-1} - 0.005)_+\right\} = \min\left\{0.5, (\mu^{-1} - 0.01)_+\right\} = 0.5 \tag{4.20}$$

$$\min\left\{0.5, (\mu^{-1} - 1)_+\right\} = \min\left\{0.5, (\mu^{-1} - 2)_+\right\} = 0 \tag{4.21}$$

so that:

$$\mu^{-1} - 0.01 \geq 0.5 \tag{4.22}$$

$$\mu^{-1} - 1 \leq 0 \tag{4.23}$$

Hence, μ^{-1} can be any in the interval $[0.51, 1]$. □

The optimal dual variable responsible for the total power constraint is always unique in the standard WF (under the TPC alone), while this is not the case under the joint power constraints (TPC+PAC).

The optimal dual variable μ is unique under a certain condition. In this case, μ can be expressed in a closed-form (see Proposition 5). To do so, consider active streams and let $\mathcal{I}_{PAC}$ be the set of streams with active PAC and $\mathcal{I}_{\overline{PAC}}$ be the set of streams with inactive PAC, so that:

$$\mathcal{I}_{PAC} = \{i : p_i^* = P_{1i}\} \tag{4.24}$$

$$\mathcal{I}_{\overline{PAC}} = \{i : 0 < p_i^* < P_{1i}\} \tag{4.25}$$

where $\mathcal{I}_{\overline{PAC}}$ and $\mathcal{I}_{PAC}$ have $(m_+ - m_{PAC})$ and m_{PAC} elements respectively. To find $\mathcal{I}_{\overline{PAC}}$, we find all $p_i^* > 0$ from Proposition 1. Then, we remove those streams with $p_i^* = P_{1i}$ by using Proposition 4. Furthermore, $\mathcal{I}_{PAC}$ is found by Proposition 4. The next proposition gives a closed-form expression for the optimal dual variable μ when it is unique.

Proposition 5. *If $m_+ \neq m_{PAC}$, i.e., $m_+ > m_{PAC}$ (for active antennas, at least one PAC is inactive), then μ in* (4.5) *for the problem in* (4.1) *is unique and can be expressed as:*

$$\mu^{-1} = \frac{1}{m_+ - m_{PAC}} \left(P_T + \sum_{i \in \mathcal{I}_{\overline{PAC}}} g_i^{-1} - \sum_{i \in \mathcal{I}_{PAC}} P_{1i} \right) \tag{4.26}$$

where $\mathcal{I}_{\overline{PAC}}$ and $\mathcal{I}_{PAC}$ are as in (4.25) *and* (4.24).

Proof. Here, the TPC is active since $m_+ > m_{PAC}$. Then,

$$P_T = \sum_{i=1}^{m} p_i^* = \underbrace{\sum_{i \in \mathcal{I}_{PAC}} P_{1i}}_{p_i^* = P_{1i}} + \underbrace{\sum_{i \in \mathcal{I}_{\overline{PAC}}} (\mu^{-1} - g_i^{-1})}_{0 < p_i^* < P_{1i}} \tag{4.27}$$

Hence, (4.26) follows after some simple manipulations. □

The following proposition shows the effect of the joint constraint (TPC+PAC) on the dual variable μ.

Proposition 6. *Let μ and μ_{WF} be the dual variables responsible for the total power constraint under the joint TPC+PAC constraints and the TPC alone, respectively.*

Then, for the same P_T, the following holds:

$$\mu \leq \mu_{WF} \tag{4.28}$$

The strict inequality occurs if at least one PAC is active.

Proof. See Appendix. □

In a convex problem, the sensitivity of the optimal value with respect to changes in a constraint is determined by its receptive dual variable [66]. A smaller dual variable shows less sensitivity, and a larger one shows more sensitivity. Therefore, $\mu \leq \mu_{WF}$ implies that the OPA under the joint constraints (TPC+PAC) is less sensitive to the variations of P_T compared to that under the TPC alone (the standard WF).

Next, we show that if some streams exceed P_{1i} in the standard WF, then for these streams, the transmitted per-antenna powers equal the respective PACs.

Corollary 3. *Under the same P_T, if $p^*_{i,WF} \geq P_{1i}$, then $p^*_i = P_{1i}$.*

Proof. The proof is based on Proposition 6, as follows:

$$\mu^{-1} \geq \mu_{WF}^{-1} \tag{4.29}$$

$$(\mu^{-1} - g_i^{-1})_+ \geq (\mu_{WF}^{-1} - g_i^{-1})_+ = p^*_{i,WF} \tag{4.30}$$

Hence, if $p^*_{i,WF} \geq P_{1i}$, then $(\mu^{-1} - g_i^{-1})_+ \geq P_{1i}$. So,

$$p^*_i = \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} = P_{1i} \tag{4.31}$$

□

This property was also shown in [36], [38], [37], and [21]. However, here we present a simpler proof which gives us additional insight, i.e., the optimal dual variable responsible for the total Tx power constraint under the joint constraints (TPC+PAC) is less than that under the TPC alone (see Proposition 6).

4.2 Same Per-Antenna Power Constraints

In this part, we consider identical PACs, i.e., $P_{1i} = P_1$. This is motivated by the fact that all Tx antennas are usually equipped with the same amplifiers. Here, Theorem 1 and Propositions 1-4 apply directly with $P_{1i} = P_1$.

From Proposition 4 and in the case of identical PACs with $P_{1i} = P_1$, the isotropic signaling $\mathbf{R}^*(\mathbf{D}_W) = P_1\mathbf{I}_m$ is optimal under the joint power constraints if and only if:

$$P_T \geq mP_1 \tag{4.32}$$

In this case, no feedback is needed, since P_1 is known at Tx. This is in contrast to the standard WF, where the uniform power allocation is not optimal in general at any finite value of P_T (unless all g_i are the same).

The next proposition gives a simpler closed-form expression for the dual variable μ when $P_{1i} = P_1$.

Proposition 7. *Consider identical PACs and let $\{g_i\}$ be in descending order. If $m_+ \neq m_{PAC}$, then μ is unique and can be expressed as follows:*

$$\mu^{-1} = \frac{1}{m_+ - m_{PAC}}\left(P_T + \sum_{i=m_{PAC}+1}^{m_+} g_i^{-1} - m_{PAC}P_1\right) \tag{4.33}$$

where m_+ and m_{PAC} are as in (4.9) *and* (4.15).

Proof. Follows via the same approach as in the proof for Proposition 5. □

Fig. 4.1 shows the OPA of Theorem 1 for the following example:

$$\{g_i\} = \{30, 1, 0.7, 0.5\}, \quad P_1 = 0.8 \tag{4.34}$$

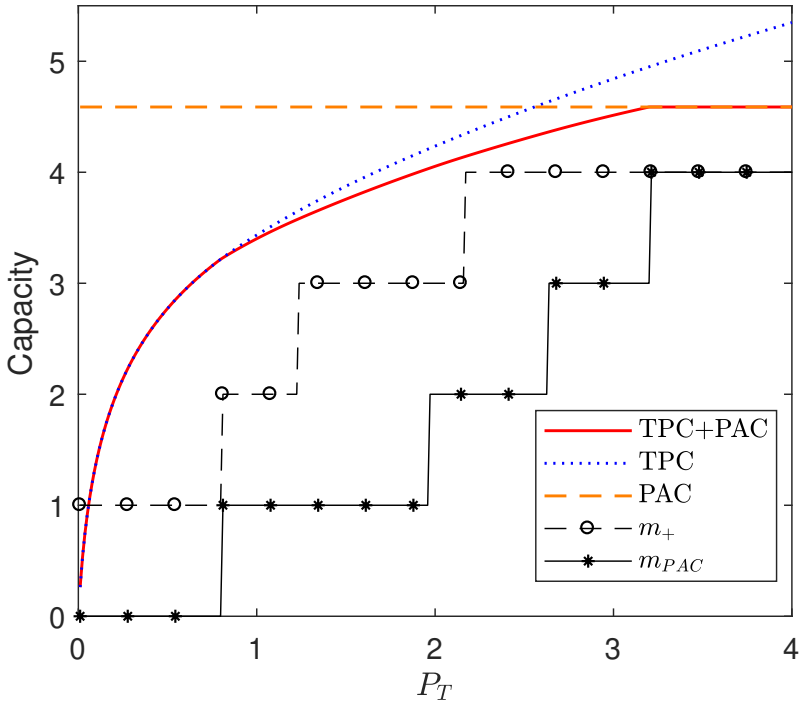

FIGURE 4.1: The capacity [nat/s/Hz] of the orthogonal MIMO channel in (4.34) under the PAC, the TPC, and the joint power constraints. The numbers of active streams and active PACs under the joint power constraints are also shown.

It is observed that when P_1 is fixed and P_T increases, the numbers of active streams and active PACs also increase. In the case of joint constraints, the capacity is upper-bounded, as follows:

$$C(\mathbf{D}_W) \leq \min\left\{C_{TPC}(\mathbf{D}_W), C_{PAC}(\mathbf{D}_W)\right\} \tag{4.35}$$

where $C_{TPC}(\mathbf{D}_W)$ and $C_{PAC}(\mathbf{D}_W)$ are the capacities under the TPC alone and the PAC alone, respectively. Fig. 4.1 shows that the upper bound is tight almost everywhere except for the transition region. This upper bound is also tight when the TPC or the PAC is inactive, i.e., for small P_T when $m_{PAC} = 0$ and $C(\mathbf{D}_W) = C_{TPC}(\mathbf{D}_W)$, or for large P_T when $m_{PAC} = m$ and $C(\mathbf{D_W}) = C_{PAC}(\mathbf{D_W})$. Also, $C(\mathbf{D}_W) \approx C_{TPC}(\mathbf{D}_W)$ when $P_T \leq 2$, and $C(\mathbf{D}_W) \approx C_{PAC}(\mathbf{D}_W)$ when $P_T \geq 3$. This implies that at low-SNR, the TPC is the dominant constraint, while the PAC is dominant at high-SNR. Hence, we can use the upper bound in (4.35) as

an approximation for any P_T, as follows:

$$C(\mathbf{D_W}) \approx \min\left\{C_{TPC}(\mathbf{D}_W), C_{PAC}(\mathbf{D}_W)\right\} \tag{4.36}$$

4.3 Capped Water-Filling Interpretation

As mentioned in Chapter 3, the standard WF procedure can be compared to pouring water into a container with a certain floor profile which is determined by the inverse of the channel gains. The optimal power allocation in Theorem 1 can also be compared to pouring water into a container which has a ceiling profile in addition to a floor profile. Fig. 4.2 shows this geometric interpretation. The floor profile is the same as for the standard WF, and is given by $\{g_i^{-1}\}$. In the case of identical PACs (i.e., $P_{1i} = P_1$), the ceiling profile is the same as the floor profile but lifted up by P_1. Adding the ceiling to the container yields a re-distribution of the extra water as follows: when one part of the container is full, the water goes to other higher parts. The water level μ^{-1} is determined by the total available water (i.e., P_T) and the floor and ceiling profiles. Specifically, in Fig. 4.2, μ^{-1} is measured by the distance between the ground and the water surface in the second and third parts. Hence, the water level is affected by the presence of the ceiling, except when all PACs are inactive. As is shown in Fig. 4.2, the OPA for each stream is given by the height of the water from the bottom of the respective parts, and cannot exceed P_1.

When the water is poured into the container, the lowest part is filled first, where the lowest part corresponds to the strongest channel gains (i.e., g_1). This shows the behavior of the OPA in the low-SNR regime. After that, water goes to the second-lowest part. This process continues until all the available water (i.e., P_T) is used. When the lowest part is full, the excess water goes to other higher parts. Hence, there is more water in the higher parts (corresponding to weaker streams) compared to the standard WF. The behavior of the capped WF is exactly like the

standard WF if all PACs are inactive, i.e., when no levels are full. This condition is given in (4.17).

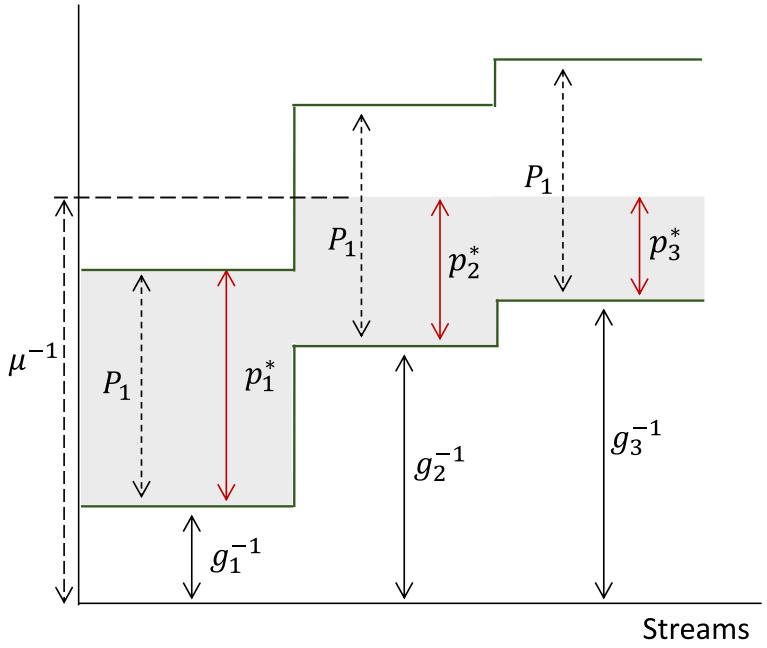

FIGURE 4.2: Geometric interpretation of the capped water-filling procedure. The water level is affected by the presence of the ceiling as well as the floor profiles. Here, the water has reached the ceiling of the first part, and so the excess water goes to the second and third parts.

Now, let us explore the geometric interpretation of the example in Corollary 2, in which the dual variable μ is not unique. As can be observed from Fig. 4.3, the first and second parts are full, while there is no water on the other parts. In this case, μ^{-1} can be anything between the ceiling height of the second part and the floor height of the third part, i.e., $\mu^{-1} \in [0.51, 1]$. So, choosing μ^{-1} from this interval does not change the water allocation in the container. For example, let $\mu^{-1} = 0.75$. Then, the first and second parts are full, while the other parts are empty. If we reduce (or increase) P_T by a small value, then μ^{-1} is unique and is determined by the elevation of the water surface at the second (or third) level. This example is a degenerate case in which for all active streams, all PACs are active (i.e., $m_+ = m_{PAC}$).

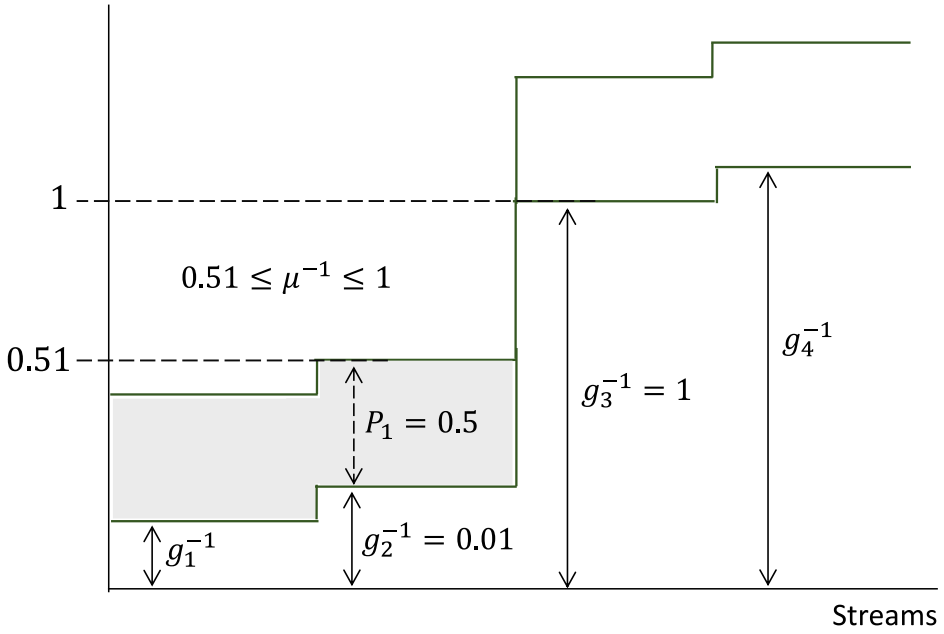

FIGURE 4.3: Capped water-filling interpretation when μ is not unique. For active streams, all PACs are active, i.e., $m_+ = m_{PAC} = 2$.

In the case of different PACs, the OPA in (4.4) can also be interpreted by the capped WF procedure. Here, the ceiling profile is not the same as the floor profile (see Fig. 4.4), and for each part, it is lifted from the floor profile by the value of P_{1i}. It is observed that lower levels do not necessarily get more water, and this depends on the height of the ceiling. For example, in Fig. 4.4, we can observe that $p_1^* < p_2^*$, but the first part is deeper than the second one. Once the third-lowest part (corresponding to the weakest stream) is full, water goes back to the second-lowest part (representing a stronger stream). However, in the case of identical PACs, water always goes from lower parts to higher parts.

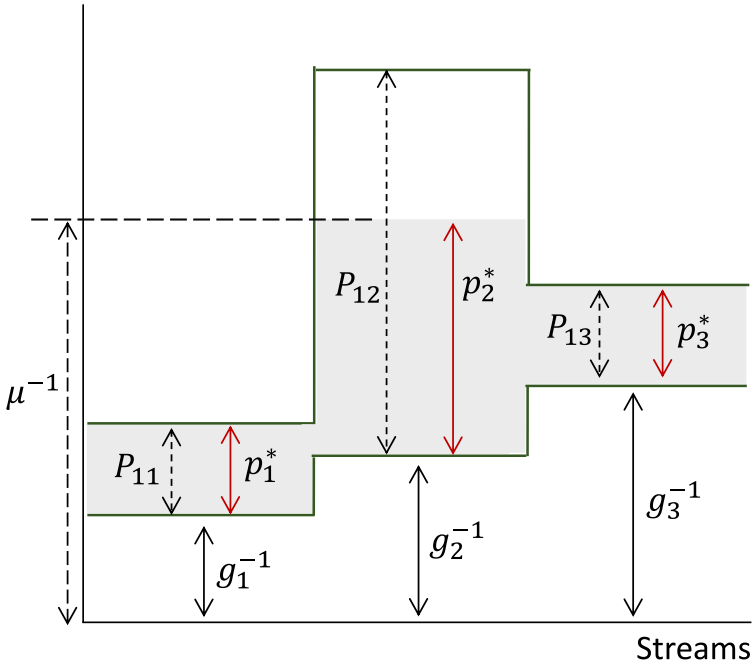

FIGURE 4.4: Capped water-filling for the case of different PACs. Also, $p_1^* < p_2^*$, but $g_1 > g_2$.

4.4 Iterative Water-Filling Algorithm

In this section, we wish to answer the following question: how can the optimal signaling under the joint power constraints in Theorem 1 be found by using the standard WF procedure? First, note that if $P_T \geq \sum_{i=1}^{m} P_{1i}$, then $p_i^* = P_{1i}$, which is the trivial solution. Otherwise, we can find the OPA in Theorem 1 via Algorithm 1 below.

In this algorithm, the channel gains $\mathbf{g} = \{g_i\}$, and the joint power constraints (i.e., $\{P_{1i}\}$ and P_T) are the inputs. First, the optimal signaling under the TPC alone (i.e., P_T) is computed via the standard water-filling procedure. After that, we check whether the optimal signaling under the TPC alone satisfies the PACs or not. If yes, then the OPA under the joint power constraints (TPC+PAC) is found, which is the same as the standard WF. Otherwise, some streams in the standard WF procedure (TPC alone) violate the PACs. For these streams, the OPA under the joint constraints follows the PACs (see Corollary 3). We can remove these

streams, since p_i^* is known for them. Hence, there is a new optimization problem with a smaller total power constraint, while the number of streams is also reduced. This process continues until all the PACs are satisfied which we here call the "iterative water-filling" algorithm.

The proposed algorithm is summarized below, where $WF\{\mathbf{g}, P_T\}$ is the standard WF for streams $\mathbf{g}$ under the total power constraint P_T. Let us consider the t-th step of the algorithm, in which the OPA under the TPC alone (the standard WF) is computed. $\delta\mathbf{p}_t^* = \{\delta p_{i,t}^*\}$ is the OPA under the TPC alone, and $\mathcal{I}_t = \{i : \delta p_{i,t}^* > P_{1i}\}$ is the set of streams that exceed the PACs. At the end of step t, some streams are removed, as mentioned before. The power allocated to them is denoted by $\Delta_{t+1} = \sum_{i \in \bigcup_{j=1}^{t} \mathcal{I}_j} P_{1i}$, and the remaining streams belong to $\mathbf{g}_{t+1} = \left\{g_i : i \notin \bigcup_{j=1}^{t} \mathcal{I}_j\right\}$.

Algorithm 1 Iterative water-filling (TPC + PAC)

Require: $\mathbf{g}$, P_{1i}, P_T

if $P_T \geq \sum_{i=1}^{m} P_{1i}$ **then**

1. $p_{i,1}^* = P_{1i}$ for any i

else

2. $\mathbf{g}_1 = \mathbf{g}$, $\Delta_1 = 0$ and $t = 0$

repeat

3. $t := t + 1$
4. $\delta\mathbf{p}_t^* = WF\{\mathbf{g}_t, P_T - \Delta_t\}$
5. $\mathcal{I}_t = \{i : \delta p_{i,t}^* > P_{1i}\}$
6. $p_{i,t}^* = P_{1i}$, if $i \in \bigcup_{j=1}^{t} \mathcal{I}_j$. Otherwise, $p_{i,t}^* = \delta p_{i,t}^*$.
7. $\mathbf{g}_{t+1} = \left\{g_i : i \notin \bigcup_{j=1}^{t} \mathcal{I}_j\right\}$
8. $\Delta_{t+1} = \sum_{i \in \bigcup_{j=1}^{t} \mathcal{I}_j} P_{1i}$

until $\mathcal{I}_t = \emptyset$

end if

In the t-th step, $p_{i,t}^*$ is the allocated power for the i-th stream that satisfies the PAC. When the algorithm stops, all allocated powers satisfy the PAC and the

TPC (see Proposition 8). There are m streams, which is a finite value, and so the algorithm stops (i.e., $\mathcal{I}_t = \emptyset$) at most after m steps.

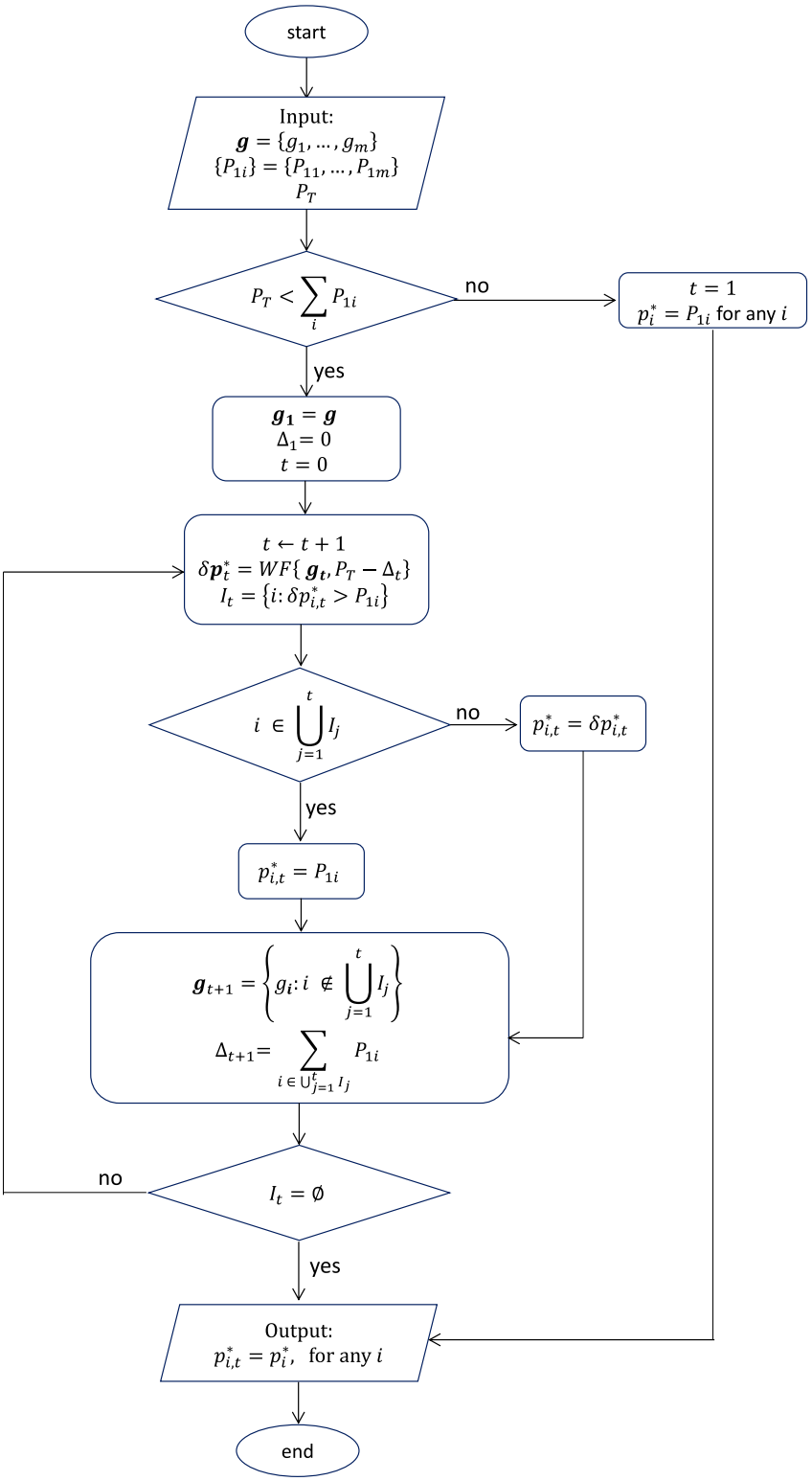

FIGURE 4.5: Iterative water-filling algorithm

To prove the convergence of this algorithm to the OPA under the joint power constraints, we use the following Lemmas. To this end, let k be the number of iterations in Algorithm 1.

Lemma 4.1. *Let μ_t^{-1} be the water level in the WF procedure at the t-th step in Algorithm 1. Then, the following holds if $P_T < \sum_{i=1}^{m} P_{1i}$,*

$$\mu_1^{-1} \leq \mu_2^{-1} \leq ... \leq \mu_k^{-1} \tag{4.37}$$

i.e., $\{\mu_t^{-1}\}$ is an increasing sequence.

Proof. See Appendix. □

The next Lemma shows that full power P_T is indeed used when the algorithm stops under active TPC.

Lemma 4.2. *For Algorithm 1, the following holds if $P_T < \sum_{i=1}^{m} P_{1i}$.*

$$\sum_{i=1}^{m} p_{i,k}^{*} = P_T \tag{4.38}$$

Proof. See Appendix. □

The next proposition shows that the iterative WF algorithm converges to the OPA under the joint power constraints in (4.4).

Proposition 8. *When Algorithm 1 stops after k iterations, then $\{p_{1,k}^{*}, p_{2,k}^{*}, ..., p_{m,k}^{*}\}$ is the optimal power allocation under the joint total and per-antenna power constraints in* (4.4) *(i.e., $p_{i,k}^{*} = p_i^{*}$ for any i).*

Proof. See Appendix. □

After this work had been finished, we became aware of similar algorithms introduced in [36], [38], and [37]. Reference [36] developed a similar algorithm for MISO channels in order to find the optimal signaling under the joint power constraints

(TPC+PAC), while [38] introduced a similar iterative algorithm for MIMO channels at high SNR. Reference [37] developed a similar algorithm to find the optimal signaling under the TPC, PAC, and per-group power constraints for three cases: (i) MISO channels; (ii) full-rank MIMO channels with full-rank optimal Tx covariance; and (iii) MIMO channels with 2 Tx antennas. Reference [67] also proposed a version of Algorithm 1 for the weighted problem (see Section 4.5 for the problem statement). In [67], it is shown that the KKT (Karush-Kuhn-Tucker) conditions for the convex problem (capacity under joint TPC+PAC constraints for orthogonal channels) are satisfied by using the algorithm, and hence the algorithm finds the solution to the optimization problem. In this thesis, however, our proof gives us insights into the problem as follows: the optimal dual variable responsible for the TPC in the standard WF decreases at each step. This continues until all power allocations satisfy the joint total and per-antenna power constraints (see Lemma 4.1).

The performance of the algorithm was evaluated for many cases, and here we present a representative example. To do so, let us consider the following channel:

$$\{g_i\} = \{2.1, 1.4, 0.9\}, \quad \{P_{1i}\} = \{0.11, 0.31, 0.34\}, \quad P_T = 0.7 \tag{4.39}$$

Table 4.1 shows the algorithm's results for the example (4.39).

Step 1	$\delta p^*_{i,1}$:	0.47	0.23	0	$\mu_1^{-1} = 0.95$	$\mathcal{I}_1 = \{1\}$
	$p^*_{i,1}$:	0.11	0.23	0		
Step 2	$\delta p^*_{i,2}$:		0.49	0.1	$\mu_2^{-1} = 1.2$	$\mathcal{I}_2 = \{2\}$
	$p^*_{i,2}$:	0.11	0.31	0.1		
Step 3	$\delta p^*_{i,3}$:			0.28	$\mu_3^{-1} = 1.39$	$\mathcal{I}_3 = \emptyset$
	$p^*_{i,3}$:	0.11	0.31	0.28		

TABLE 4.1: Results of the iterative WF for the example (4.39)

The example shows that the algorithm stops after 3 steps, i.e., $\mathcal{I}_3 = \emptyset$. In Step 1, the OPA under the TPC alone (the standard WF) for the first stream violates the PAC, i.e., $\delta p^*_{1,1} > P_{1i}$. Hence, P_{11} is allocated to that stream, and then it is removed for Step 2. After that, the WF procedure is applied to the remaining streams, i.e., Streams 2 and 3, under the residual power $P_T - \Delta_1 = 0.59$. This process is repeated for the third step. Finally, we observe that all streams satisfy the PACs, and the OPA is attained under the joint power constraints (TPC+PAC). As expected, $\{\mu_t^{-1}\}$ is an increasing sequence.

4.5 Weighted Rate Maximization

In massive MIMO channels, each user may have a different bandwidth Δf_i. Also, some users may have a higher grade of service compared to others. Let α_i be a coefficient that shows the allocated bandwidth and a level of priority for the i-th

user, and be as follows:

$$\alpha_i = s_i \frac{\Delta f_i}{\Delta f} \tag{4.40}$$

where Δf is the total bandwidth of the system and a larger s_i means that the i-th user has a higher grade of service. Thus, we consider a weighted rate maximization problem, as follows:

$$C(\alpha_i, g_i) = \max_{p_i \in S_p} \sum_{i=1}^{m} \alpha_i \ln(1 + g_i p_i) \tag{4.41}$$

where S_p is the feasible set (TPC+PAC), and:

$$S_p = \left\{ \{p_i\} : p_i \geq 0, \sum_{i=1}^{m} p_i \leq P_T, p_i \leq P_{1i} \right\} \tag{4.42}$$

If all users have the same bandwidth and priority allocated, then the α_i are the same, which was explored before in Theorem 1. The optimal power allocation for this problem under the TPC alone was obtained in [67]. Here, however, the OPA under the joint power constraints (TPC+PAC) is obtained in the next Theorem.

Theorem 2. *The optimal power allocation for the problem in* (4.41) *under the joint power constraints in* (4.42) *is as follows:*

$$p_i^* = \min \left\{ P_{1i}, (\alpha_i \mu^{-1} - g_i^{-1})_+ \right\} \tag{4.43}$$

where $\mu \geq 0$. *If* $P_T \geq \sum_{i=1}^{m} P_{1i}$, *then* $\mu = 0$; *if* $P_T < \sum_{i=1}^{m} P_{1i}$, *then* $\mu > 0$ *can be obtained as a solution of* $\sum_{i=1}^{m} p_i^* = P_T$. *The channel capacity is as in* (4.41), *with* $p_i = p_i^*$.

Proof. Follows via the same approach as that in Theorem 1. □

The next propositions give the closed-form expressions for m_+ and m_{PAC} for the problem in (4.41).

Proposition 9. *Let* $\{\alpha_i g_i\}$ *be in descending order. The number* m_+ *of active streams for the problem in* (4.41) *is as follows:*

$$m_+ = max\{j : u_j < P_T\} \tag{4.44}$$

where:

$$u_j = \sum_{i=1}^{j} \alpha_i \min\left\{\frac{P_{1i}}{\alpha_i}, \left((g_j\alpha_j)^{-1} - (g_i\alpha_i)^{-1}\right)\right\} \tag{4.45}$$

and $p_i^* > 0$ *for* $i \leq m_+$ *and* $p_i^* = 0$ *for* $i > m_+$.

Proof. The optimal power allocation p_i^* in (4.43) can be expressed as follows:

$$p_i^* = \alpha_i \min\left\{\frac{P_{1i}}{\alpha_i}, \left(\mu^{-1} - (\alpha_i g_i)^{-1}\right)_+\right\} \tag{4.46}$$

Then, (4.44) follows via the same approach as in the proof for Proposition 1. □

Proposition 10. *To find the active PACs, we consider the active streams obtained from the previous proposition, i.e.,* $p_i^* > 0$. *Let* $\{c_i\}$ *be as follows:*

$$c_i = \frac{g_i^{-1} + P_{1i}}{\alpha_i} \tag{4.47}$$

and let $\{c_i\}$ *be in ascending order, i.e.,* $c_1 \leq c_2 \leq ... \leq c_{m_+}$. *Then, the number* m_{PAC} *of active PACs for the problem in* (4.41) *is determined as follows:*

$$m_{PAC} = max\left\{j : v_j \geq \sum_{i=1}^{m_+} P_{1i} - P_T\right\} \tag{4.48}$$

where:

$$v_j = \sum_{i=j}^{m_+} \alpha_i (c_i - c_j) \tag{4.49}$$

and $p_i^* = P_{1i}$ *for* $1 \leq i \leq m_{PAC}$, $p_i^* < P_{1i}$ *for* $i > m_{PAC}$. *No PAC is active if and only if:* $v_1 < \sum_{i=1}^{m_+} P_{1i} - P_T$.

Proof. The optimal power allocation p_i^* in (4.43) for the active streams can be expressed as follows:

$$p_i^* = \alpha_i \min\left\{0, \left(\mu^{-1} - \frac{g_i^{-1} + P_{1i}}{\alpha_i}\right)\right\} + P_{1i} \tag{4.50}$$

Then, (4.48) follows via the same approach in the proof for Proposition 4. □

Note that u_j is increasing in j and v_j is decreasing in j. Hence, they can be obtained by the approaches in Propositions 1 and 4. If $m_+ \neq m_{PAC}$, then μ^{-1} is uniquely determined, as follows:

$$\mu^{-1} = \left(\sum_{i \in \mathcal{I}_{\overline{PAC}}} \alpha_i\right)^{-1} \left(P_T + \sum_{i \in \mathcal{I}_{\overline{PAC}}} g_i^{-1} - \sum_{i \in \mathcal{I}_{PAC}} P_{1i}\right) \tag{4.51}$$

where $\mathcal{I}_{\overline{PAC}}$ and $\mathcal{I}_{PAC}$ are as in (4.25) and (4.24). This expression can be obtained via the same approach as in Proposition 5.

For the problem in (4.41), rank-1 transmission (beamforming) is optimal if and only if $u_2 \geq P_T$ and all Tx power is allocated to the largest $\{\alpha_i g_i\}$. All streams are active if and only if $u_m < P_T$ and the OPA is $p_i^* = \min\{P_{1i}, \alpha_i \mu^{-1} - g_i^{-1}\}$. If:

$$u_m < P_T \quad \text{and} \quad v_1 < \sum_{i=1}^{m} (P_{1i}) - P_T \tag{4.52}$$

then all streams are active and none exceeds the PACs. Hence, the optimal power allocation reduces to $p_i^* = \alpha_i \mu^{-1} - g_i^{-1}$.

4.6 Summary

The optimal Tx covariance matrix under the total power constraint (TPC) and per-antenna power constraints (PAC) is obtained for a favorable propagation channel, and it is diagonal, i.e., independent signaling is optimal. The optimal power allocation is the minimum of two terms: the first one represents the PACs and the

second represents the water-filling (WF) procedure. The dual variable responsible for the TPC is less than that for the standard WF. In this case, the optimal Tx covariance under the joint TPC and PAC constraints is less sensitive to the variations of the total power constraint compared to that in the standard WF. This follows from the fact that in a convex problem, a smaller dual variable shows less sensitivity of the optimal value with respect to the variations of the imposed constraint [66].

The numbers of active streams and active PACs are found analytically. It is shown that active streams are stronger than inactive ones (which is also the case in the standard WF). The conditions of optimality of beamforming and full-rank transmission are obtained.

To gain more insights, a geometric interpretation of the optimal power allocation is proposed. It is similar to the standard WF, but the container has a ceiling in addition to the floor. The ceiling profile is elevated from the floor profile by the values of the PACs, and this results in the capped water-filling interpretation.

An iterative WF algorithm is proposed in order to find the optimal power allocation under the joint power constraints (TPC+PAC) by using the standard WF. The convergence to the optimal solution is proven.

Finally, the weighted problem under the joint power constraints is considered. The optimal power allocation is analytically attained in a closed form.

Chapter 5

Nearly Favorable Propagation and Robustness

In the previous chapter, the optimal Tx covariance under the joint power constraints was obtained for favorable propagation channels (i.e., $\mathbf{W} = \mathbf{D}_W$). In practice, the conditions of favorable propagation can be only approximately satisfied. Specifically, large but finite numbers of antennas result in non-zero and small off-diagonal entries of $\mathbf{W}$. So, it is important to note the effect of non-zero and small off-diagonal entries of $\mathbf{W}$ on capacity and optimal signaling. In this chapter, we consider the case that off-diagonal entries of $\mathbf{W}$ are not zero, and quantify the sub-optimality of signaling in Theorem 1 by establishing an upper bound for the rate loss. We show that as long as off-diagonal entries of $\mathbf{W}$ are small enough, the rate loss (sub-optimality gap) of using $\mathbf{R}^*(\mathbf{D}_W)$ instead of true optimal signaling $\mathbf{R}^*(\mathbf{W})$ is also small. So, the independent signaling in Theorem 1 is robust with respect to small off-diagonal entries of $\mathbf{W}$. To do so, we consider nearly-favorable propagation as defined in (3.14), where non-zero but small values are allowed. Also, the channel gain is always bounded due to the law of energy conversion, i.e., $\sigma_1(\mathbf{W}) = \sigma_1^2(\mathbf{H}) < \infty$. Further note that the Tx covariance matrix is bounded since the total Tx power is limited, i.e.,

$$\sigma_1(\mathbf{R}) \leq tr(\mathbf{R}) \leq P_T < \infty \tag{5.1}$$

where $\sigma_1(\mathbf{R})$ is the spectral norm of $\mathbf{R}$.

To show the robustness property, we first obtain the general bound for the sub-optimality gap when the covariance $\mathbf{R}^*(\mathbf{W}_0)$ (which is optimal signaling for the channel $\mathbf{W}_0$) is applied to channel $\mathbf{W}$. To this end, let $C(\mathbf{W}) = C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W})\big)$ be the capacity of channel $\mathbf{W}$ with its optimal covariance $\mathbf{R}^*(\mathbf{W})$, and $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big)$ is the maximum rate achieved by the Tx covariance $\mathbf{R}^*(\mathbf{W}_0)$ on the channel $\mathbf{W}$. Then, the sub-optimality gap (rate loss) is as follows:

$$\Delta C(\mathbf{W}, \mathbf{W}_0) = C\big(\mathbf{W}\big) - C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big) \tag{5.2}$$

The next proposition gives the upper-bound of the sub-optimality gap in (5.2).

Proposition 11. *Let* $\Delta \boldsymbol{W} = \boldsymbol{W} - \boldsymbol{W}_0$ *and* $\sigma_1\big(\boldsymbol{R}^*(\boldsymbol{W})\big), \sigma_1\big(\boldsymbol{R}^*(\boldsymbol{W}_0)\big) \leq P_T$. *If* $\sigma_1(\Delta \boldsymbol{W})P_T < 1$, *then the sub-optimality gap* $\Delta C(\boldsymbol{W}, \boldsymbol{W}_0)$ *is bounded as follows:*

$$0 \leq \Delta C(\boldsymbol{W}, \boldsymbol{W}_0) \leq m \ln\Big(\frac{1 + \sigma_1(\Delta \boldsymbol{W})P_T}{1 - \sigma_1(\Delta \boldsymbol{W})P_T}\Big) \tag{5.3}$$

where $\sigma_1(\Delta \boldsymbol{W})$ *is the largest singular value of* $\Delta \boldsymbol{W}$. *Hence:*

$$\Delta C(\boldsymbol{W}, \boldsymbol{W}_0) \to 0 \quad \textit{as} \quad \sigma_1(\Delta \boldsymbol{W}) \to 0 \tag{5.4}$$

Proof. See Appendix. □

This is a generic result, and the upper bound can be applied for any two MIMO channels (not only massive MIMO channels). This bound can be considered for the channel estimation where $\mathbf{W}_0$ is a channel estimate and $\mathbf{W}$ is the true channel. It shows that as long as the difference $\Delta\mathbf{W}$ between two channels is small, the rate loss $\Delta C(\mathbf{W}, \mathbf{W}_0)$ is also small.

Proposition 11 implies that if $\mathbf{W}_0$ is close to $\mathbf{W}$ (i.e., $\sigma_1(\Delta\mathbf{W}) \to 0$), then $\mathbf{R}^*(\mathbf{W}_0)$ is the sub-optimal Tx covariance for channel $\mathbf{W}$. However, it can be shown that

$\Delta C(\mathbf{W}, \mathbf{W}_0) \to 0$ does not result in $\mathbf{R}^*(\mathbf{W}_0) \to \mathbf{R}^*(\mathbf{W})$, which means that in general, $\mathbf{R}^*(\mathbf{W})$ is not a continuous function[1].

The continuity property of $C(\mathbf{W})$ is shown in the next Lemma.

Lemma 5.1. *For the bounded channel $\boldsymbol{W}_0$ and the bounded Tx covariance matrix, the following holds:*

$$\lim_{\boldsymbol{W} \to \boldsymbol{W}_0} C(\boldsymbol{W}) = C(\boldsymbol{W}_0) \tag{5.5}$$

It follows that: (i) $C(\boldsymbol{W})$ is a continuous function; and (ii) $C(\boldsymbol{W}, \boldsymbol{R})$ is jointly uniformly-continuous for bounded $\boldsymbol{W}$ and $\boldsymbol{R}$ (see (9.72) *for the definition of the joint uniform continuity).*

Proof. See Appendix. □

This Lemma shows that if two channels converge to each other, their capacities also converge.

The next lemma shows the robustness property of the optimal signaling under favorable propagation in Theorem 1.

Corollary 4. *The following hold for any bounded Tx covariance and $\boldsymbol{D}_W$:*

$$C(\boldsymbol{W}) \to C(\boldsymbol{D}_W), \quad \Delta C(\boldsymbol{W}, \boldsymbol{D}_W) \to 0 \quad as \quad \sigma_1(\boldsymbol{W} - \boldsymbol{D}_W) \to 0 \tag{5.6}$$

Furthermore, if $\sigma_1(\boldsymbol{W} - \boldsymbol{D}_W) P_T < 1$, then:

$$\Delta C(\boldsymbol{W}, \boldsymbol{D}_W) < \epsilon \quad if \quad \sigma_1(\boldsymbol{W} - \boldsymbol{D}_W) < \delta_\epsilon = \frac{1}{P_T} \frac{e^{\epsilon/m} - 1}{e^{\epsilon/m} + 1} \tag{5.7}$$

Proof. Follows from Proposition 11 and Lemma 5.1. □

[1]Here, $\mathbf{W}_0 \to \mathbf{W}$ means that the norm of $\Delta\mathbf{W}$ tends to zero, and we use the spectral norm (the largest singular value), since all matrix norms are equivalent when the size of a matrix is finite [68]. This is the case in the considered problem because $\mathbf{W}, \mathbf{W}_0 \in \mathbb{C}^{m,m}$ and $m < \infty$.

Further note that (5.7) implies that $\delta_\epsilon P_T < 1$, which is consistent with the assumption of $\sigma_1(\mathbf{W} - \mathbf{D}_W)P_T < 1$.

This corollary shows that the rate loss $C(\mathbf{W}, \mathbf{D}_W)$ is small as long as $\sigma_1(\mathbf{W}-\mathbf{D}_W)$ is small enough[2]. Hence, $\mathbf{R}^*(\mathbf{D}_W)$ is robust with respect to small off-diagonal entries of $\mathbf{W}$ (i.e., $\mathbf{W}$ is nearly orthogonal).

Fig. 5.1 shows that δ_ϵ is decreasing in m and P_T. When $P_T = 10$ and for any $\mathbf{W} > 0$, the figure shows that $\Delta C(\mathbf{W}, \mathbf{D}_W) < 10^{-2}$ as long as $\sigma_1(\mathbf{W}-\mathbf{D}_W) < 10^{-5}$.

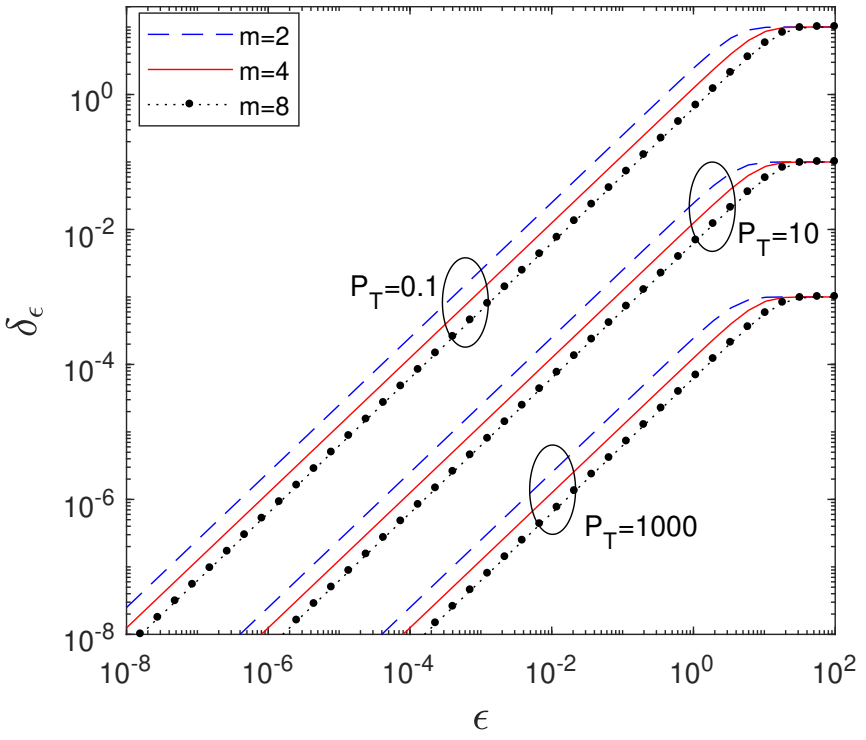

FIGURE 5.1: δ_ϵ vs. ϵ for different values of m and P_T. In all cases, $\delta_\epsilon P_T < 1$.

In fact, the bound in (5.7) can be considered as a quantitative definition of nearly-favorable propagation, which makes (3.14) precise; i.e., the channel offers nearly (or ϵ)-favorable propagation if $\sigma_1(\mathbf{W} - \mathbf{D}_W) < \delta_\epsilon$.

Corollary 5. *If $\sigma_1(\boldsymbol{W} - \boldsymbol{D}_W) < \delta_\epsilon$, then $|\boldsymbol{h}_i^+ \boldsymbol{h}_j| < \delta_\epsilon$ for any $i \neq j$, i.e., small $\sigma_1(\boldsymbol{W} - \boldsymbol{D}_W)$ implies small off-diagonal entries of $\boldsymbol{W}$*

[2]Here, $\mathbf{W} - \mathbf{D}_W$ is a Hermitian matrix with zero diagonal and its off-diagonal entries are the same as in $\mathbf{W}$.

Proof. See Appendix. □

This Corollary justifies the definition of nearly-favorable propagation.

In the next section, we will consider some examples in which the exact value of the rate loss will be numerically obtained.

5.1 Examples

Here, we present some examples in which the signaling in Theorem 1 is applied to non-orthogonal MIMO channels (i.e., $\mathbf{W} \neq \mathbf{D}_W$). We compute $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W})\big)$ and $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{D}_W)\big)$ in order to demonstrate the near-optimality of $\mathbf{R}^*(\mathbf{D}_W)$ when it is applied to $\mathbf{W}$. The high-SNR and low-SNR regimes are considered, as well as different numbers m of Tx antennas.

5.1.1 Example for $m = 2$

First, we consider the low-SNR regime (small P_T) for the following channel:

$$\mathbf{W} = \begin{bmatrix} 3 & a \\ a & 2 \end{bmatrix}, \quad P_T = 1, \quad \{P_{1i}\} = \{0.8, 0.6\} \tag{5.8}$$

Fig. 5.2 shows the capacity of channel $\mathbf{W}$ and the rate achieved by Tx covariance $\mathbf{R}^*(\mathbf{D}_W)$ (signaling in Theorem 1) on channel $\mathbf{W}$. The signaling $\mathbf{R}^*(\mathbf{W})$ is obtained by the Newton-barrier algorithm (see Chapter 6 for more details). This shows that when the value of a becomes smaller, the difference between the capacity $C(\mathbf{W})$ and the rate $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{D}_W)\big)$ also becomes smaller. Fig. 5.3 shows the exact value of the rate loss [%] versus the off-diagonal entry a of $\mathbf{W}$. We observe that the rate of $\mathbf{R}^*(\mathbf{D}_W)$ is close to the capacity when a is not too large. For $a \leq 1.6$, the rate loss is less than 10%, which means that even for a large value of a, about 90% of the capacity is achieved by the signaling $\mathbf{R}^*(\mathbf{D}_W)$ (the largest possible value of a is $\sqrt{6} \approx 2.4$, for which $\mathbf{W} \geq 0$). Fig. 5.3 shows that as $a \to 0$, the rate loss gets

close to zero, which was expected from (5.6). This example illustrates that the rate loss is small as long as the off-diagonal entries of $\mathbf{W}$ are not too large, which shows the robustness of $\mathbf{R}^*(\mathbf{D}_W)$ for channel $\mathbf{W}$ when a is not too large.

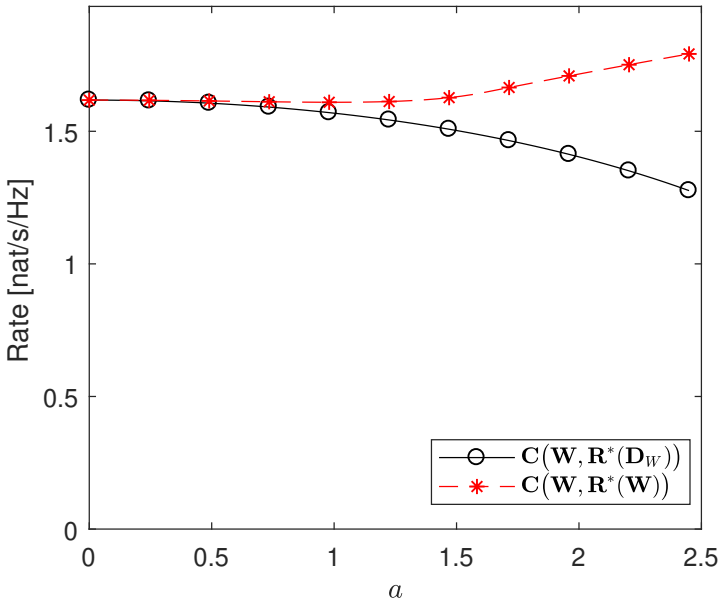

FIGURE 5.2: The capacity $C(\mathbf{W})$ and the rate $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{D}_W)\big)$ at low-SNR vs. off-diagonal entry a when $m = 2$. The channel is as in (5.8).

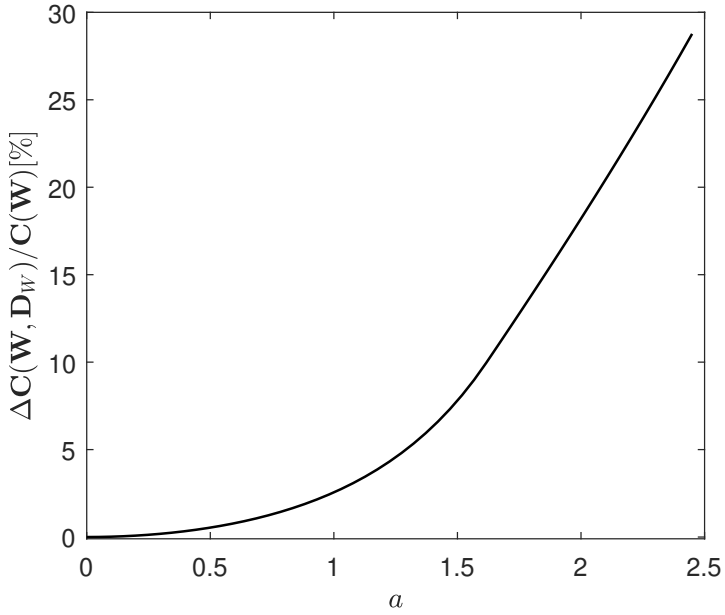

FIGURE 5.3: The sub-optimality gap (rate loss) [%] when signaling $\mathbf{R}^*(\mathbf{D}_W)$ is applied to $\mathbf{W}$. The gap is small as long as a is not too large. The channel is as in (5.8).

Next, we consider the same channel $\mathbf{W}$ in (5.8) at high SNR (large P_T) for the following setting:

$$\mathbf{W} = \begin{bmatrix} 3 & a \\ a & 2 \end{bmatrix}, \quad P_T = 20, \quad \{P_{1i}\} = \{15, 10\} \tag{5.9}$$

Fig. 5.4 shows that the rate $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{D}_W)\big)$ is almost the same as the capacity $C(\mathbf{W})$ for $a \leq 2$, which includes the large values of a. Fig. 5.5 shows that the rate loss is less than 1% for $a \leq 2$ and is less than 15% for the largest possible value of a (i.e., $a \approx 2.4$). Comparing Fig. 5.3 with Fig. 5.5, we can see that signaling $\mathbf{R}^*(\mathbf{D}_W)$ is more robust with respect to a at high SNR compared to low SNR. However, in both cases, the rate loss is small as long as the off-diagonal entries of $\mathbf{W}$ are not too large.

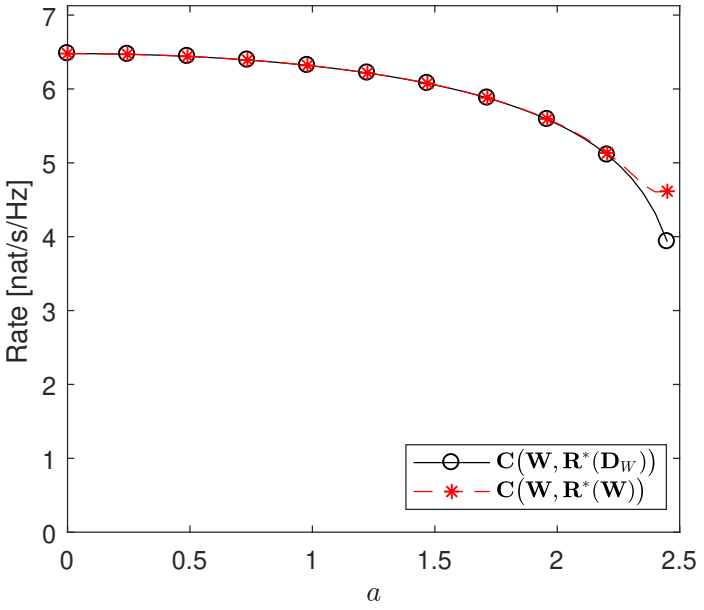

FIGURE 5.4: The capacity $C(\mathbf{W})$ and the rate $C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{D}_W)\big)$ vs. off-diagonal entry a when $m = 2$. The channel is as in (5.9).

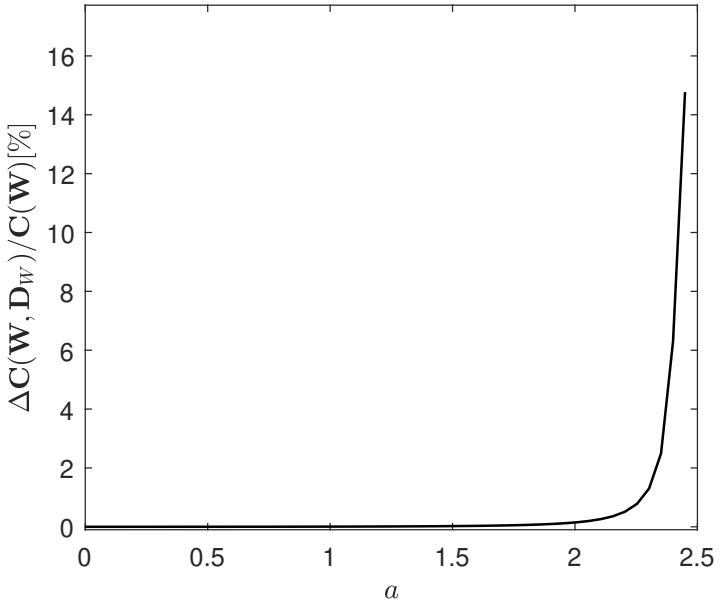

FIGURE 5.5: The rate loss [%] when signaling $\mathbf{R}^*(\mathbf{D}_W)$ is applied to $\mathbf{W}$. The channel is as in (5.9). Observe that even when a is large (i.e. $a = 2$), the rate loss is small.

5.1.2 Example for $m = 4$

Fig. 5.6 shows the rate loss for the channel in (5.10) for different values of P_T and P_{1i}.

$$\mathbf{W} = \begin{bmatrix} 2 & -0.3a & 0.5a & -1.3a \\ -0.3a & 4 & 1.6a & 0.2a \\ 0.5a & 1.6a & 5 & -0.7a \\ -1.3a & 0.2a & -0.7a & 3 \end{bmatrix}, \quad \{P_{1i}\} = P_T[0.4, 0.2, 0.3, 0.7] \quad (5.10)$$

Here, the sub-optimality gap is shown for both positive and negative values of a. In this example, the off-diagonal entries of $\mathbf{W}$ are not the same, and $\mathbf{W} \geq 0$ for $a \in [-1.5, 1.8]$. To compare the rate loss at low SNR with that at high SNR, we consider different values of P_T and P_{1i}. As can be seen from Fig. 5.6, the rate loss is small as long as a is not large. In particular, when $-0.4 \leq a \leq 0.4$, the rate loss is less than 10% for any values of P_T. This shows that as P_T increases (i.e., the SNR increases), the rate loss decreases. So, the signaling $\mathbf{R}^*(\mathbf{D}_W)$ is more robust

with respect to off-diagonal entries of $\mathbf{W}$ at high SNR compared to low SNR. Also, the rate loss is less than 10% for $-1 \leq a \leq 1$ and $P_T \geq 1$.

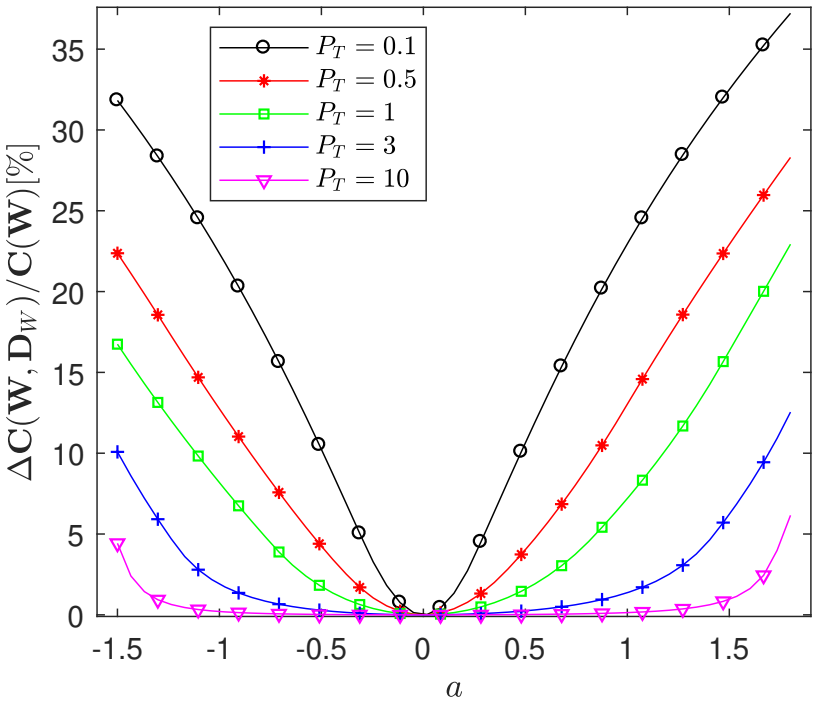

FIGURE 5.6: The rate loss $\Delta C(\mathbf{W}, \mathbf{D}_W)$ for the channel in (5.10) for different off-diagonal entry values of $\mathbf{W}$, P_T, and P_{1i}.

5.1.3 The Distribution of Sub-Optimality Gap for Different Channel Realizations

Here, the rate loss is obtained for different channel realizations. In this section, we consider the scenario where the off-diagonal entries of $\mathbf{W}$ are not necessarily small but the sum of their magnitudes is less than the respective diagonal entry. This is quantified as follows:

$$(\mathbf{W})_{ii} \geq \sum_{j \neq i} |(\mathbf{W})_{ij}|, \quad \text{for any } i \tag{5.11}$$

where $\mathbf{W}$ is diagonally dominant[3].

[3]Note that strictly diagonally dominant matrices are full-rank [68]. Hence, if $\mathbf{W}$ is strictly diagonally dominant, i.e., $(\mathbf{W})_{ii} > \sum_{j \neq i} |(\mathbf{W})_{ij}|$, then it is full-rank.

In order to obtain different diagonally dominant matrices, matrix $\mathbf{A} \geq 0$ is generated, in which the entries have a Gaussian distribution of unit variance and zero mean. Then, $\mathbf{W}$ is set such that:

$$(\mathbf{W})_{ij} = (\mathbf{A})_{ij}, \quad i \neq j \tag{5.12}$$

$$(\mathbf{W})_{ii} = \max\left\{(\mathbf{A})_{ii}, \sum_{j \neq i} |(\mathbf{A})_{ij}|\right\} \tag{5.13}$$

so that $\mathbf{W}$ is diagonally dominant, which can also be represented as follows:

$$\mathbf{W} = \mathbf{A} + diag\{a_i\} \tag{5.14}$$

where:

$$a_i = \max\left\{0, \sum_{j \neq i} |(\mathbf{A})_{ij}| - (\mathbf{A})_{ii}\right\} \tag{5.15}$$

Hence, $\mathbf{W} \geq 0$, since both $\mathbf{A}$ and $diag\{a_i\}$ are positive semi-definite matrices.

Fig. 5.7 shows the rate loss [%] for 1,000 randomly-generated $\mathbf{W} \in \mathbb{R}^{5,5}$, where $P_T = 3$ and $\{P_{1i}\} = \{1.5, 0.5, 1, 0.5, 1\}$. The rate loss is always less than 6%, and is less than 1% for around 700 channel realizations. Fig. 5.8 shows the rate loss for the same setting, where $(\mathbf{W})_{ii} \geq 1.2 \sum_{j \neq i} |(\mathbf{W})_{ij}|$. In this case, the rate loss is less than 1% for around 900 channel realizations, which means that the number of channel realizations with a very small rate loss increases as $(\mathbf{W})_{ii} / \sum_{j \neq i} |(\mathbf{W})_{ij}|$ increases.

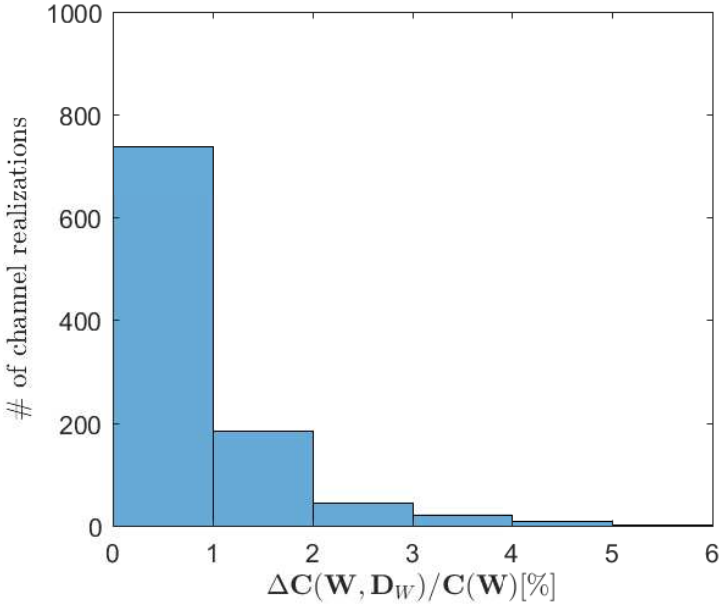

FIGURE 5.7: The rate loss [%] for 1,000 channel realizations with randomly-generated $\mathbf{W} \in \mathbb{R}^{5,5}$ where: $(\mathbf{W})_{ii} \geq \sum_{j \neq i} |(\mathbf{W})_{ji}|$, $P_T = 3$ and $\{P_{1i}\} = \{1.5, 0.5, 1, 0.5, 1\}$.

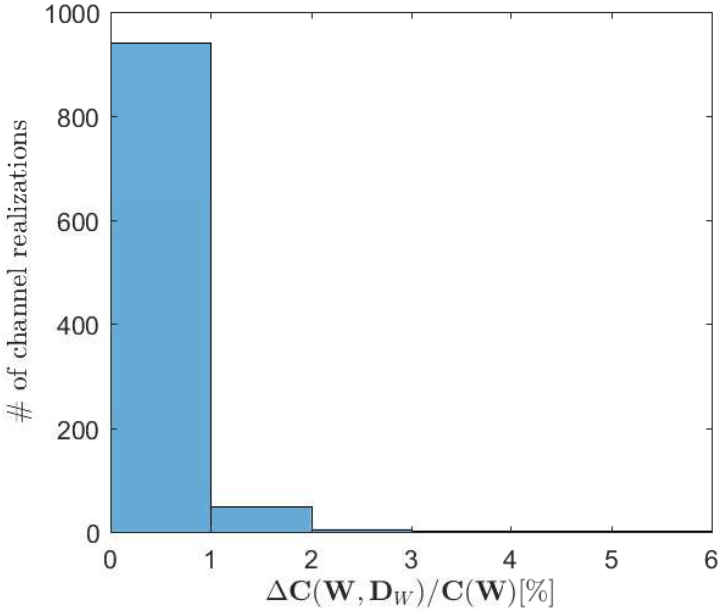

FIGURE 5.8: The rate loss [%] for 1,000 channel realizations with randomly-generated $\mathbf{W} \in \mathbb{R}^{5,5}$ where: $(\mathbf{W})_{ii} \geq 1.2\sum_{j \neq i} |(\mathbf{W})_{ji}|$, $P_T = 3$ and $\{P_{1i}\} = \{1.5, 0.5, 1, 0.5, 1\}$.

5.2 Summary

The robustness of optimal signaling under favorable propagation (FP) is studied. In particular, we study the case of "nearly-favorable propagation" (NFP) in which the channel is nearly orthogonal. We show that optimal signaling under favorable propagation (orthogonal channels) is nearly-optimal under NFP. The upper bound of the sub-optimality gap is obtained, which becomes small under NFP. Nearly (or ϵ)-favorable propagation is analytically defined, which can be used as a new measure of favorable propagation. This measure quantifies how close the channel is to FP. Some examples are presented which verify that the optimal Tx covariance under FP is nearly optimal under NFP.

Chapter 6

Numerical Algorithms

In this chapter, we develop some numerical algorithms that can be used to compute the followings: (i) the optimal dual variable μ, and (ii) the optimal transmit covariance under the joint constraints, in which the channel is not necessarily orthogonal. The first one is the bisection algorithm, which is useful to find the dual variable μ with desired accuracy. Then, the Newton-barrier algorithm will be reviewed, which can be used to attain the optimal solution for convex optimization problems. The last one is the Monte-Carlo algorithm, which searches for the optimal signaling among a large number of feasible Tx covariance matrices. Also, it is shown that CVX can give the incorrect results for the capacity under the joint power constraints.

6.1 Bisection Algorithm

In addition to the closed-form expression for the dual variable μ in (4.51), we can also numerically find it. Recall the optimal power allocation for an orthogonal MIMO channel under the total and per-antenna power constraints, i.e., $p_i^* = \min\left\{P_{1i}, (\alpha_i\mu^{-1} - g_i^{-1})_+\right\}$ (see Theorem 2). As previously mentioned, if $P_T \geq \sum_{i=1}^m P_{1i}$, then $\mu = 0$. Otherwise $\sum_{i=1}^m p_i^* = P_T$ and the dual variable $\mu > 0$ can be numerically obtained by the bisection algorithm, described below.

Let $f(x)$ be as follows:

$$f(x) = \sum_{i=1}^{m} \min\left\{P_{1i}, (\alpha_i x^{-1} - g_i^{-1})_+\right\} - P_T \tag{6.1}$$

where $f(x)$ is a continuous monotonic function and is decreasing in x, and the bisection algorithm works with monotonic functions. Let μ be the root of $f(x)$ (i.e., $f(\mu) = 0$). To find the root, let us take any lower and upper bound of μ such that $\mu_l \leq \mu \leq \mu_u$. Hence, for any x between the lower and upper bounds, we observe:

$$\mu_u - \mu_l \geq |\mu - x| \tag{6.2}$$

We find the value of $f(x)$, where x is as follows:

$$x = \frac{\mu_l + \mu_u}{2} \tag{6.3}$$

If $f(x) = 0$, then $\mu = x$. Otherwise, we can redefine the upper or lower bound of μ as follows:

- $\mu_l = x$ if $f(x) > 0$
- $\mu_u = x$ if $f(x) < 0$

Hence, the difference between the new lower and upper bounds decreases, and the new uncertainty interval around μ is reduced. So, an estimate x of the root gets closer to the actual root μ with each iteration of the bisection algorithm. This process continues until the desired uncertainty interval is achieved, i.e., $\mu_u - \mu_l \leq \epsilon$, and hence from (6.2), $|\mu - x| \leq \epsilon$. This shows that the difference between actual root and its estimation is not greater than ϵ.

Fig. 6.1 shows the first 3 steps of the bisection algorithm. In this figure, μ_{lk} and μ_{uk} are the improved lower and upper bounds of μ at the k-th step of the algorithm, and $\Delta_k = \mu_{uk} - \mu_{lk}$ and $\mu'_k = (\mu_{lk} + \mu_{uk})/2$ are the uncertainty interval

and an estimation of the root at the k-th step, respectively. In step one, we see that the value of the function at $(\mu_{l1} + \mu_{u1})/2$ is less than zero, and hence the upper bound of the root is updated as follows:

$$\mu_{u2} = \frac{\mu_{l1} + \mu_{u1}}{2} \tag{6.4}$$

and $\mu_{l2} = \mu_{l1}$. So, the uncertainty interval is decreased by a factor of two in one step, and after the k-th step, the uncertainty interval becomes:

$$\Delta_k = \frac{\Delta_1}{2^k} \tag{6.5}$$

where Δ_1 is the uncertainty interval in step one. So, the convergence of the algorithm is exponential.

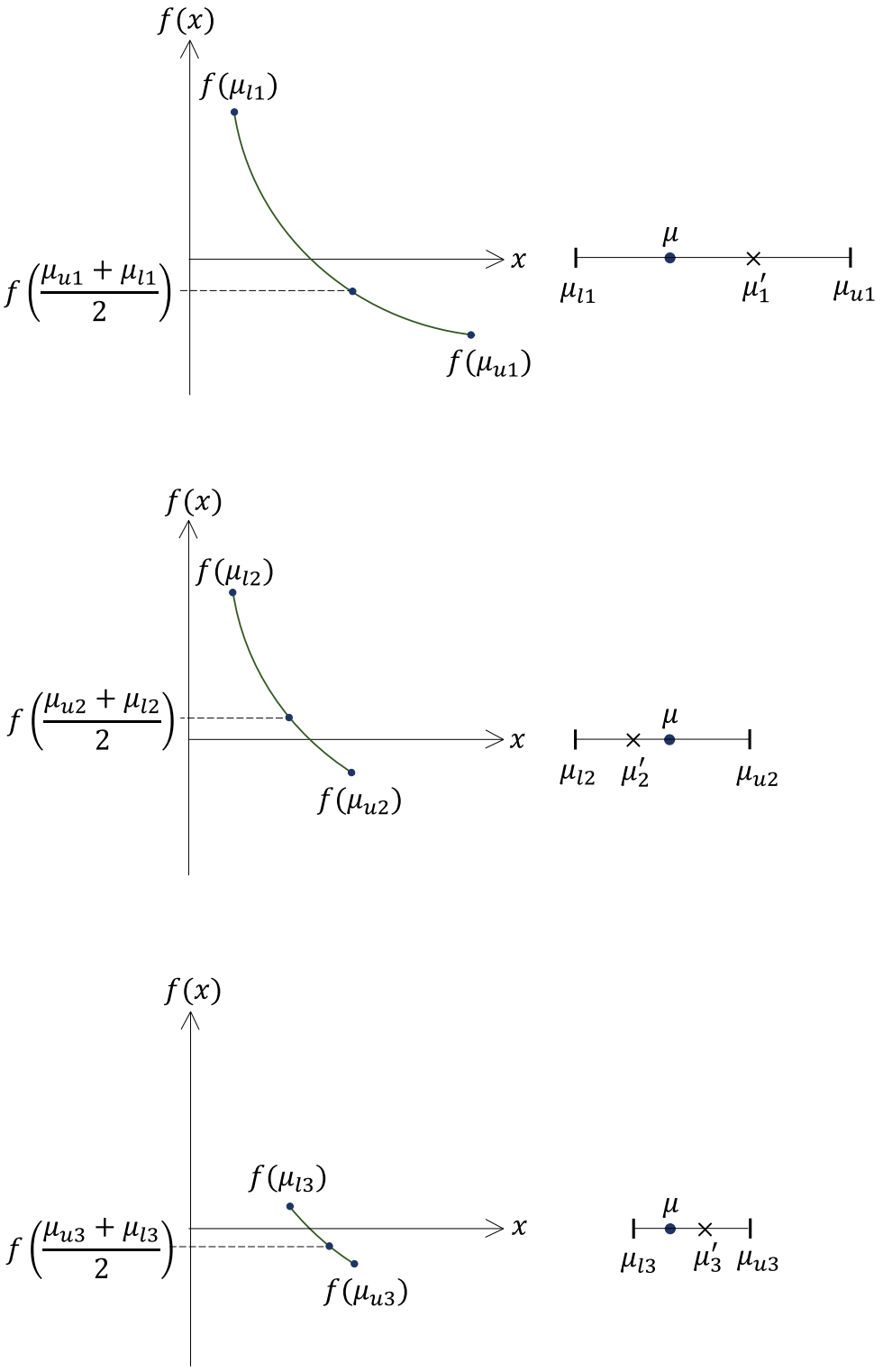

FIGURE 6.1: The first 3 steps of the bisection algorithm

Reference [66] shows that to attain ϵ-accuracy, N steps are necessary, where:

$$N = \left\lceil \log_2 \left(\frac{\mu_u - \mu_l}{\epsilon} \right) \right\rceil \tag{6.6}$$

and $\lceil \cdot \rceil$ denotes the ceiling [69]. To initialize the bisection algorithm, we need lower and upper bounds for μ (i.e., the values of μ_l and μ_u for the first step in the bisection algorithm), which are given in the next proposition.

Proposition 12. *If $P_T < \sum_{i=1}^{m} P_{1i}$, then μ is bounded as follows:*

$$0 < \mu < \max_i \{\alpha_i g_i\} \tag{6.7}$$

Proof. The lower bound follows from the dual feasibility. The upper bound follows from (4.43) since there exists at least one active stream. Hence, $\alpha_i \mu^{-1} - g_i^{-1} > 0$ for some i. □

In the following, we summarize the bisection algorithm.

Algorithm 2 Bisection algorithm

Require: $f(x)$, $\mu_l = 0$, $\mu_u = \max_i\{\alpha_i g_i\}$

repeat

1. $x = (\mu_l + \mu_u)/2$
2. If $f(x) > 0$, then $\mu_l = x$. If $f(x) < 0$, then $\mu_u = x$. Terminate if $f(x) = 0$

until $|\mu_u - \mu_l| < \epsilon$

We can also use $|f(x)| < \epsilon$ as a stopping criteria [69].

Now, we illustrate the performance of the bisection algorithm for the following example:

$$\{g_i\} = \{10, 2, 3, 1.5\}, \quad \{\alpha_i\} = \{1, 1.5, 0.5, 1\}, \quad \{P_{1i}\} = \{1, 1.2, 0.3, 0.8\} \tag{6.8}$$

Here, the bisection algorithm is used when the TPC is active (i.e., $P_T < \sum_i P_{1i} = 3.3$), otherwise $\mu = 0$. Also, $|f(x)| < \epsilon$ is the stopping criteria, with $\epsilon = 10^{-8}$. As

is shown in Fig. 6.2, both the bisection algorithm and the closed-form expression in (4.51) give the same value of μ. Fig. 6.3 shows that the bisection algorithm gives the same results for the capacity, m_+, and m_{PAC} as compared with their analytical expressions. We can see that $m_+ \neq m_{PAC}$ when the TPC is active, and hence $\mu > 0$ is unique.

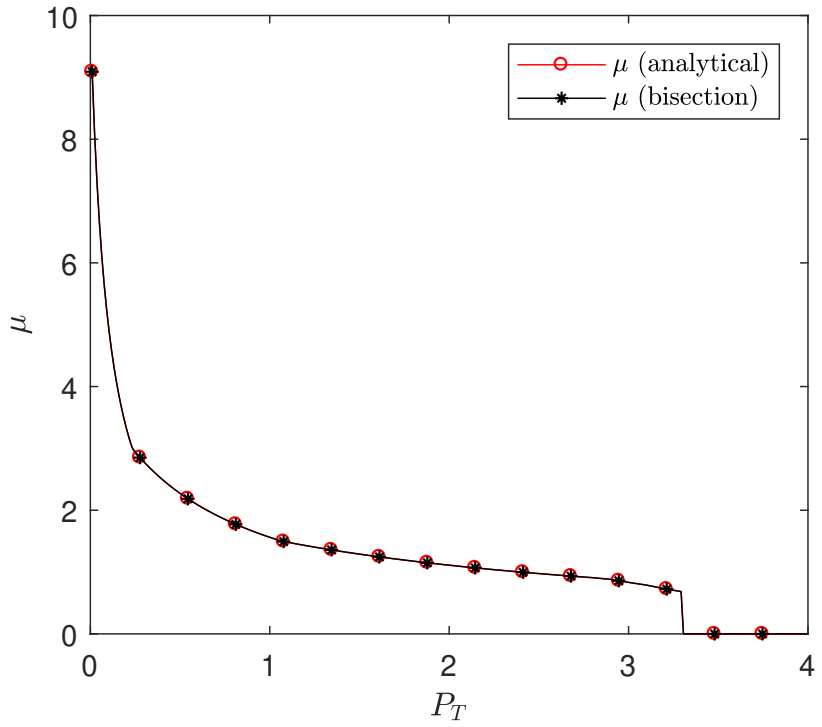

FIGURE 6.2: The dual variable μ for the example in (6.8) that is achieved by its analytical expression and the bisection algorithm.

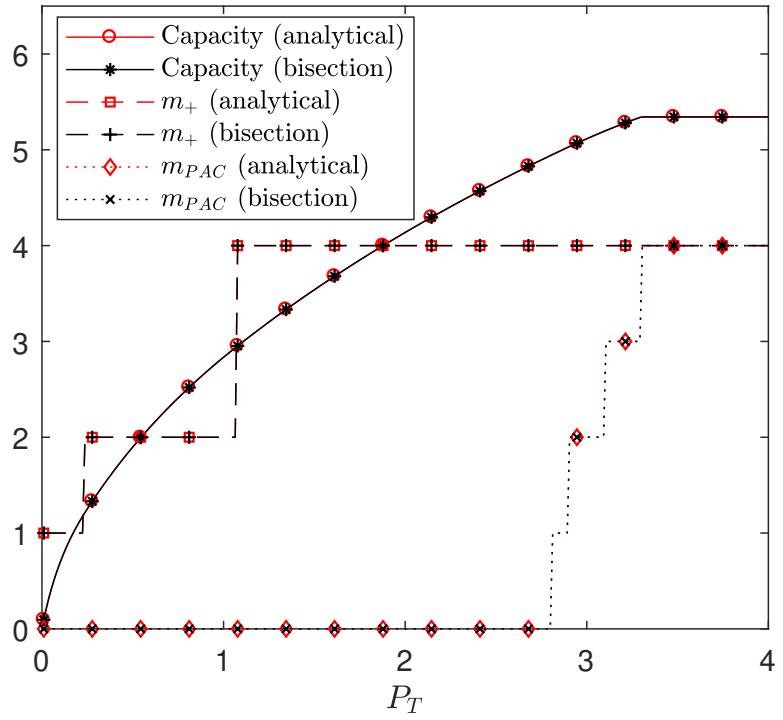

FIGURE 6.3: The capacity [nat/s/Hz], m_+, and m_{PAC} for the example in (6.8). Both the closed-form expressions and the bisection algorithm give the same results.

6.2 Newton-Barrier Algorithm

In this section, we briefly explain the Newton-barrier method, which is a powerful tool for solving convex problems [66]. Recall the optimization problem in (3.3) under the joint power constraints (TPC+PAC):

$$\text{P1:} \quad C(\mathbf{W}) = \max_{\mathbf{R}\geq 0, tr\mathbf{R}\leq P_T, (\mathbf{R})_{ii}\leq P_{1i}} \ln|\mathbf{I}+\mathbf{WR}| \tag{6.9}$$

In (6.9), the objective is concave function and inequality constraints are convex functions, and hence (6.9) has the standard form of convex problems. So, we can use the Newton-barrier method to find the optimal signaling in (6.9) for any $\mathbf{W} \geq 0$ [66]. In this method, the firm inequality constraints are converted to soft constraints, and then they are added to the objective function [70]. To do so, the inequality constrained problem P1 is transformed into unconstrained problem, as

follows:

$$\text{P2:} \quad \max_{\mathbf{R}} f_t(\mathbf{R}) \tag{6.10}$$

where

$$f_t(\mathbf{R}) = \ln|\mathbf{I} + \mathbf{WR}| + \psi_1(\mathbf{R}) + \psi_2(\mathbf{R}) + \sum_{i=1}^{m} \psi_{1i}(\mathbf{R}) \tag{6.11}$$

$$\psi_1(\mathbf{R}) = \frac{1}{t} \ln|\mathbf{R}| \tag{6.12}$$

$$\psi_2(\mathbf{R}) = \frac{1}{t} \ln(P_T - tr\mathbf{R}) \tag{6.13}$$

$$\psi_{1i}(\mathbf{R}) = \frac{1}{t} \ln\big(P_{1i} - (\mathbf{R})_{ii}\big) \tag{6.14}$$

where t determines the gap between the optimal values of the unconstrained problem P2 and the original problem P1 (see (6.16) below). Here, $\psi_1(\mathbf{R})$, $\psi_2(\mathbf{R})$, and $\psi_{1i}(\mathbf{R})$ are logarithmic barrier functions, and as the variables approach the boundary of the feasible set, values of these functions go to $-\infty$ (see [66] for more detail). The KKT (Karush-Kuhn-Tucker) condition for the unconstrained problem P2 is:

$$\nabla_{\mathbf{R}} f_t = 0 \tag{6.15}$$

Let $\mathbf{R}_t^*(\mathbf{W})$ be the optimal point for the unconstrained problem P2 for a fixed t. The gap between the solution for P1 and that for P2 is upper bounded as follows:

$$|f_t\big(\mathbf{R}_t^*(\mathbf{W})\big) - C(\mathbf{W})| \leq \frac{b}{t} \tag{6.16}$$

where b is a certain constant depending on the inequality constraints in P1 [66]. Hence, the gap can be made arbitrarily small by selecting a large value for t. So, $C(\mathbf{W})$ can be determined with any desirable accuracy [66].

Here, since $\mathbf{R}_t^*(\mathbf{W})$ is a Hermitian matrix, its lower-triangular entries and upper-triangular entries are complex conjugate pairs. So, to reduce the number of variables, only the lower-triangular entries of $\mathbf{R}_t^*(\mathbf{W}) \geq 0$ are considered. This can be obtained by $\mathbf{x} = vech(\mathbf{R})$, where operator $vech$ picks up the lower-triangular

entries of $\mathbf{R}$ from the first column to the last column [70], [71].

The KKT condition in (6.15) for the variable $\mathbf{x}$ becomes:

$$\mathbf{r}(\mathbf{x}) = \nabla_x f_t = 0 \tag{6.17}$$

The solution for $\mathbf{r}(\mathbf{x}) = 0$ can be obtained by iteratively solving the 1st-order approximation of $\mathbf{r}(\mathbf{x})$, which is expressed as follows:

$$\mathbf{r}(\mathbf{x}_k + \Delta\mathbf{x}) \approx \mathbf{r}(\mathbf{x}_k) + D\mathbf{r}\Delta\mathbf{x} = 0 \tag{6.18}$$

where $\mathbf{x}_k$ is the current feasible variable, $\Delta\mathbf{x}$ is its update, and $D\mathbf{r}$ is the 1st-order derivative of $\mathbf{r}(\mathbf{x})$:

$$D\mathbf{r} = \frac{\partial \mathbf{r}(\mathbf{x})}{\partial \mathbf{x}} = \nabla^2_{xx} f_t \tag{6.19}$$

After computing the values of $\nabla^2_{xx} f_t$ and $\mathbf{r}(\mathbf{x}_k)$, the value of $\Delta\mathbf{x}$ is obtained as a solution of the following equation:

$$-\mathbf{r}(\mathbf{x}_k) = \nabla^2_{xx} f_t \Delta\mathbf{x} \tag{6.20}$$

which follows from (6.18) and (6.19). Here, the non-singularity of $\nabla^2_{xx} f_t$ can be proved by using a method similar to that in [70]. Hence, the system of linear equations in (6.20) has a unique solution $\Delta\mathbf{x}$. Having both $\mathbf{x}_k$ and the step $\Delta\mathbf{x}$, the variable $\mathbf{x}_{k+1}$ for the next step is as follows:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + s\Delta\mathbf{x} \tag{6.21}$$

where s is the step size, which can be computed by a backtracking line search [66]. The residual norm $||\mathbf{r}(\mathbf{x}_k)||$ reduces at each step because of the following property [66]:

$$\frac{d}{ds}||\mathbf{r}(\mathbf{x}_k + s\Delta\mathbf{x})|| = -||\mathbf{r}(\mathbf{x}_k)|| < 0, \tag{6.22}$$

Hence, for sufficiently small s, the residual norm gets close to zero as the algorithm iterates (i.e., k increases). So, the optimal point $\mathbf{R}_t^*(\mathbf{W})$ for the optimization problem in (6.10) can be determined with any desired accuracy.

In practice, at first, the unconstrained problem is solved with an certain initial feasible covariance $\mathbf{R}$ and initial value of t (for instance, $t_0 = 100$ in order to initialize the Newton-barrier method for the example in (6.26)). Then, t increases and the unconstrained problem is solved again for the new t. In this step, the previous solution for the covariance is considered as the current initial variable. In this algorithm, t increases until it reaches t_{max}, in which t_{max} is large enough to obtain any desired gap according to the upper bound of the gap in (6.16) [66].

Since the gap between the solutions for the unconstrained problem P2 and the original one P1 is proportional to t^{-1}, we observe that:

$$\ln|\mathbf{I} + \mathbf{W}\mathbf{R}_t^*(\mathbf{W})| \to C(\mathbf{W}) \quad \text{as} \quad t \to \infty \tag{6.23}$$

The gradient and Hessian of the first and second terms of $f_t(\mathbf{R})$ (i.e., $\ln|\mathbf{I}+\mathbf{W}\mathbf{R}|$ and $\ln|\mathbf{R}|$) are analytically determined in [70]. Also, determining them for $\ln\big(P_{1i} - (\mathbf{R})_{ii}\big)$ is straightforward, since:

$$\frac{\partial}{\partial x}\ln\big(P_{1i} - x\big) = \frac{1}{x - P_{1i}} \tag{6.24}$$

So, we need to find only the expressions for the gradient and Hessian of $\ln|P_T - tr\mathbf{R}|$ with respect to $\mathbf{x} = vech(\mathbf{R})$, which are analytically determined in the Appendix[1].

To this end, we need a variable to initialize the algorithm, i.e., $\mathbf{x}_0 = vech(\mathbf{R}_0)$, which has to belong to the feasible set (the joint total and per-antenna power constraints). Here, we propose $\mathbf{R}_0 = diag\{r_{ii}\}$, where r_{ii} is as follows:

$$r_{ii} = a\min\left\{P_{1i}, \frac{P_T}{m}\right\} \tag{6.25}$$

[1]In addition to analytical expressions for the gradient and Hessian, one can also use numerical computations to obtain them.

where $\mathbf{R}_0 \in S_R$, and hence a can be any value in the interval [0,1]. Here, we select $a = 1/2$, since with this constant, the initial value is not too close to the boundary of the feasible set. In the results from simulations with $a = 1/2$, we observed that the algorithm converges reasonably fast. The mentioned Newton-barrier algorithm is briefly summarized below.

Algorithm 3 Newton-barrier algorithm

Require: $\mathbf{x}_0$, $\epsilon > 0$, $t_0 > 0$, $\kappa > 1$, $0 < \nu < 0.5$, $0 < \beta < 1$, $t_{max} > t$.

1. Set $t = t_0$.

repeat (barrier method)

2. Set $k = 0$.

repeat (Newton method)

3. Find $\Delta\mathbf{x}$ via (6.20) using $\mathbf{r}(\mathbf{x}_k)$ and computing $\nabla^2_{xx} f_t$ at $\mathbf{x}_k$

4. Set $s = 1$

repeat (backtracking line search to the find step size s)

5. $s := \beta s$

until $|\mathbf{r}(\mathbf{x}_k + s\Delta\mathbf{x})| \leq (1 - \nu s)|\mathbf{r}(\mathbf{x}_k)|$ & $\mathbf{R}_k \in S_R$

6. Update variable: $\mathbf{x}_{k+1} = \mathbf{x}_k + s\Delta\mathbf{x}$

7. Set $k := k + 1$

until $|\mathbf{r}(\mathbf{x}_k)| \leq \epsilon$.

8. Set the new initial value, i.e. $\mathbf{x}_0 := \mathbf{x}_k$

9. $t := \kappa t$ (a larger t)

until $t \geq t_{max}$.

The last $\mathbf{x}_k$ is the solution for the optimization problem with a desired accuracy.

In each iteration of the backtracking line search, the residual norm is decreased linearly, which is determined by ν [%], and the reduction in s is controlled by the parameter β. At each step of the barrier method, the value of t increases by the factor κ. Also, t_0 is the initial value for t.

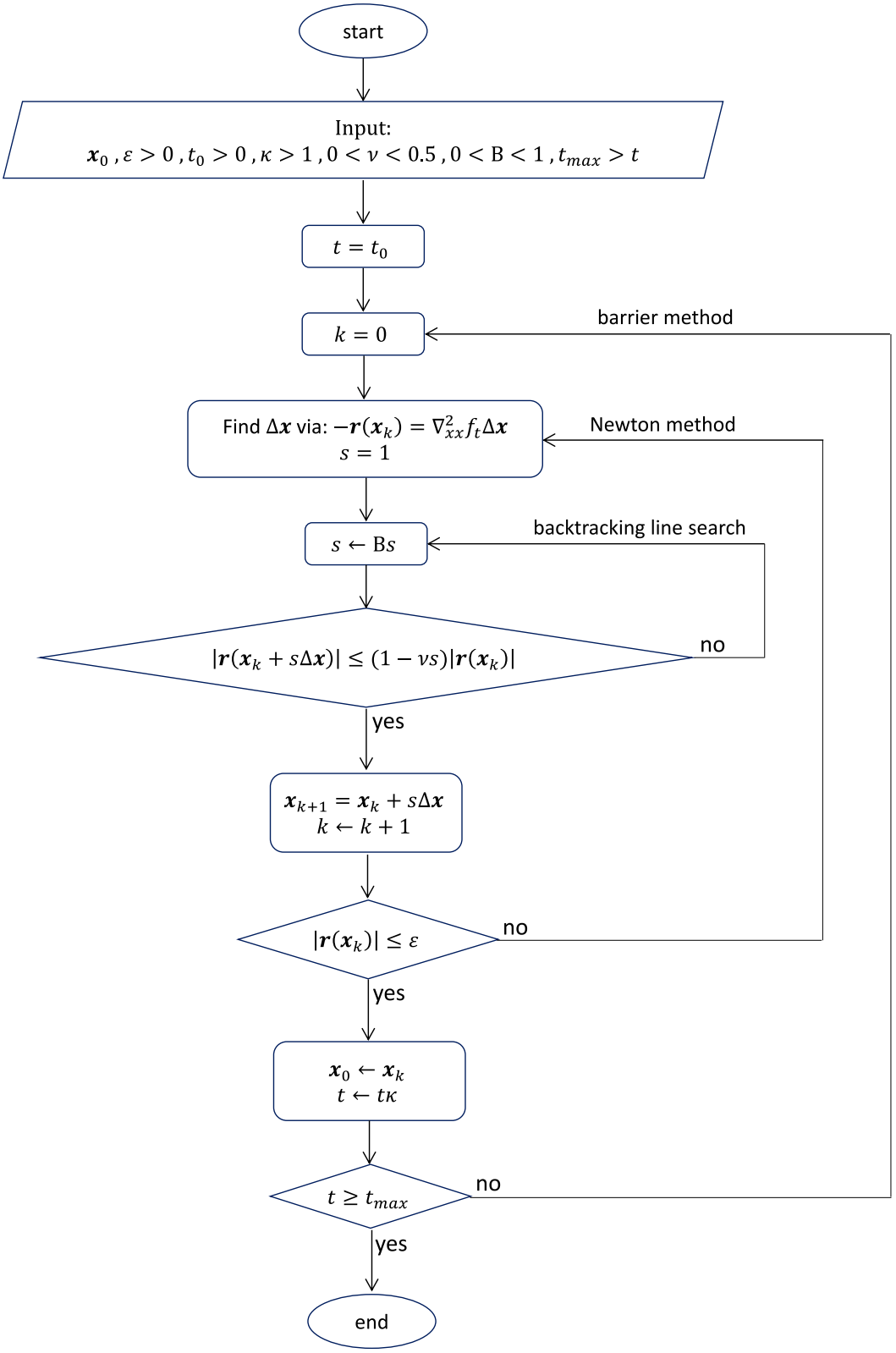

FIGURE 6.4: Newton-barrier algorithm

We ran many simulations in order to examine the performance of the Newton-barrier algorithm, and the following example is a representative case. Recall the example in (5.8) with $a = 1$:

$$\mathbf{W} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \quad P_T = 1, \quad \{P_{1i}\} = \{0.8, 0.6\} \tag{6.26}$$

Fig. 6.5 shows the residual norm for increasing t. For each value of t, the residual norm decreases quickly as more Newton steps are completed. This means that after several Newton steps, the KKT condition for the unconstrained problem in (6.15) is satisfied with ϵ-accuracy. In particular, $\mathbf{R}_t^*(\mathbf{W})$ is attained after around 5 to 10 steps. Fig. 6.6 shows that the algorithm converges to the rate 1.609 [nat/s/Hz] after around 10 Newton steps. This convergence rate also coincides with the convergence rate achieved by the Monte-Carlo algorithm in the next section.

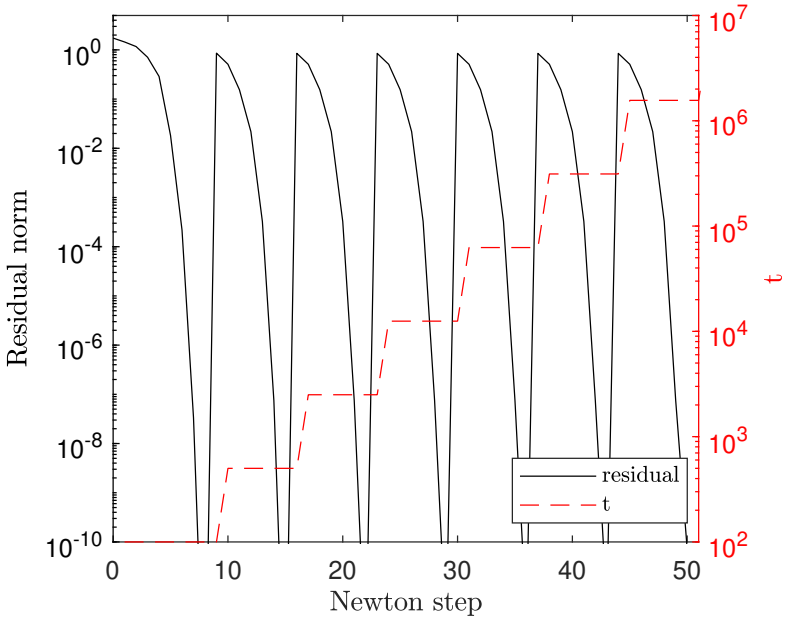

FIGURE 6.5: Convergence of the Newton-barrier algorithm for the example in (6.26). Here, $\nu = 0.3$, $\beta = 0.5$, $t_0 = 100$, $\kappa = 5$, $t_{max} = 10^7$, and $\epsilon = 10^{-10}$. For each value of t, the residual norm decreases to ϵ after around 5 to 10 steps.

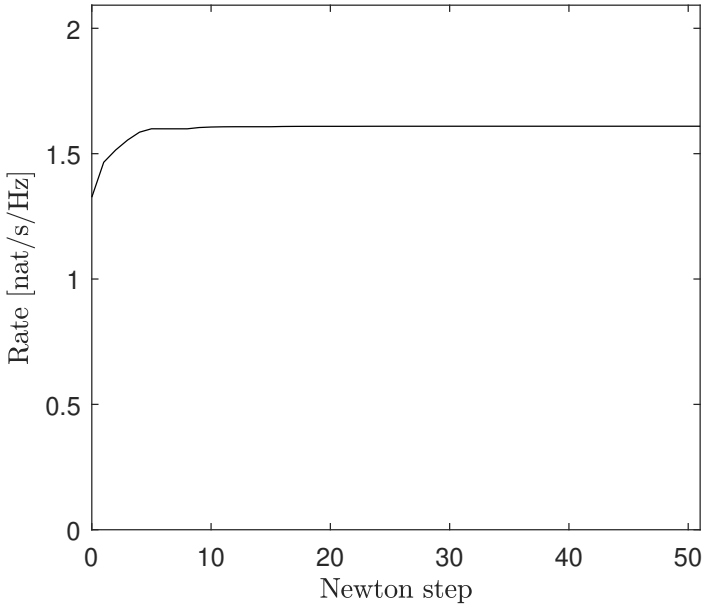

FIGURE 6.6: The rates versus Newton step for the example in (6.26). The rate converges to 1.609 after around 10 steps.

6.3 Monte-Carlo Algorithm

In this section, we develop the Monte-Carlo algorithm to obtain the optimal signaling under the joint total and per antenna power constraints (TPC+PAC) for arbitrary channel $\mathbf{H}$, which is not necessarily an orthogonal channel. In the following, we briefly review the Monte-Carlo algorithm (see [72], [73] for more details regarding the concept of the Monte-Carlo method).

In this algorithm, the feasible Tx covariance $\mathbf{R}$ is randomly and independently generated many times N, i.e., $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, ..., \mathbf{R}_N$. At first, we compute $C(\mathbf{W}, \mathbf{R}_1)$ and compare it with $C(\mathbf{W}, \mathbf{R}_0^*)$, where $\mathbf{R}_0^*$ is an initial feasible Tx covariance. If $C(\mathbf{W}, \mathbf{R}_1) > C(\mathbf{W}, \mathbf{R}_0^*)$, then $\mathbf{R}_1^* = \mathbf{R}_1$, otherwise $\mathbf{R}_1^* = \mathbf{R}_0^*$. Likewise, in each step, we determine $\mathbf{R}_i^*$ by comparing $C(\mathbf{W}, \mathbf{R}_i)$ with $C(\mathbf{W}, \mathbf{R}_{i-1}^*)$. Hence, $\left\{C(\mathbf{W}, \mathbf{R}_i^*)\right\}$ is an increasing sequence:

$$C(\mathbf{W}, \mathbf{R}_0^*) \leq C(\mathbf{W}, \mathbf{R}_1^*) \leq ... \leq C(\mathbf{W}, \mathbf{R}_N^*) \tag{6.27}$$

Also, $C(\mathbf{W}, \mathbf{R}_N^*)$ is always upper-bounded by the capacity of $\mathbf{W}$, i.e., $C(\mathbf{W}, \mathbf{R}_N^*) \leq C(\mathbf{W})$. Hence, this is an increasing sequence which is upper-bounded, and therefore it converges.

To generate a random matrix $\mathbf{S}_{m\times m}$, we use the `randn` generator in MATLAB[2]. By taking into account that $\mathbf{S}^+\mathbf{S}$ is a positive semi-definite matrix, we determine random Tx covariance $\mathbf{R} \geq 0$ as follows:

$$\mathbf{R} = a\mathbf{S}^+\mathbf{S} \tag{6.28}$$

where $a > 0$ should be selected such that $tr\mathbf{R} \leq P_T$ and $(\mathbf{R})_{ii} \leq P_{1i}$ for any i. So,

$$a(tr\mathbf{S}^+\mathbf{S}) \leq P_T \tag{6.29}$$

$$a(\mathbf{S}^+\mathbf{S})_{ii} \leq P_{1i}, \quad i = 1, ..., m \tag{6.30}$$

Hence, a is restricted by $m + 1$ inequalities, which are expressed as follows:

$$a < \frac{P_T}{tr\mathbf{S}^+\mathbf{S}} \tag{6.31}$$

$$a < \frac{P_{1i}}{(\mathbf{S}^+\mathbf{S})_{ii}}, \quad i = 1, ..., m \tag{6.32}$$

Since $a > 0$, we obtain:

$$0 < a \leq \min\left\{\frac{P_T}{tr\mathbf{S}^+\mathbf{S}}, \frac{P_{11}}{(\mathbf{S}^+\mathbf{S})_{11}}, ..., \frac{P_{1m}}{(\mathbf{S}^+\mathbf{S})_{mm}}\right\} = a^* \tag{6.33}$$

We set $a = a^*$, since in this case increasing the power (i.e., increasing the value of a) implies higher rate[3]. Here, we use an initial feasible Tx covariance $\mathbf{R}_0^* = diag\big\{\min\{P_T/m, P_{1i}\}\big\}$, where independent signaling (diagonal Tx covariance) is used and all power constraints are satisfied. Also, each antenna transmits. The whole algorithm is summarized below.

[2]This returns a matrix whose entries are generated with a normal distribution.

[3]In other words, $C(\mathbf{W}, \mathbf{R}) = \ln|\mathbf{I} + \mathbf{WR}|$ is increasing in $\mathbf{R}$. Hence, $\mathbf{R} = a^*\mathbf{S}^+\mathbf{S}$ is selected.

Algorithm 4 Monte-Carlo algorithm

Require: $\mathbf{W}_{m\times m}$, P_T, P_{1i}, N

1. Set $\mathbf{R}_0^* = diag\big\{\min\{P_T/m, P_{1i}\}\big\}$

for $i := 1$ to N **do**

2. $\mathbf{S}$ =`randn(m)`
3. Set $\mathbf{R}_i = a^*\mathbf{S}^+\mathbf{S}$
4. If $C(\mathbf{W},\mathbf{R}_i) > C(\mathbf{W},\mathbf{R}_{i-1}^*)$, then $\mathbf{R}_i^* = \mathbf{R}_i$. Otherwise, $\mathbf{R}_i^* = \mathbf{R}_{i-1}^*$.
5. Determine $C(\mathbf{W},\mathbf{R}_i^*)$

end for

A lot of simulations were run to see the performance of the algorithm, and here we show a representative case in Fig. 6.7. This figure illustrates the performance of the Monte-Carlo algorithm for the example in (6.26), in which $\big\{C(\mathbf{W},\mathbf{R}_i^*)\big\}$ is an increasing sequence. The Monte-Carlo algorithm reaches the rate 1.6 [nat/s/Hz] after around 100 steps, which is equal to the rate achieved by the Newton-barrier algorithm.

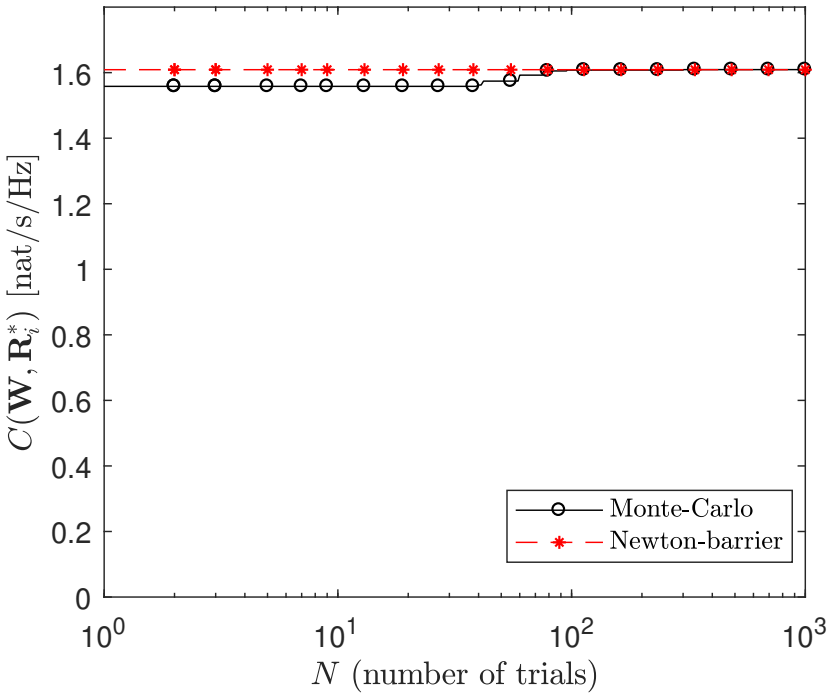

FIGURE 6.7: Performance of the Monte-Carlo algorithm for the example in (6.26)

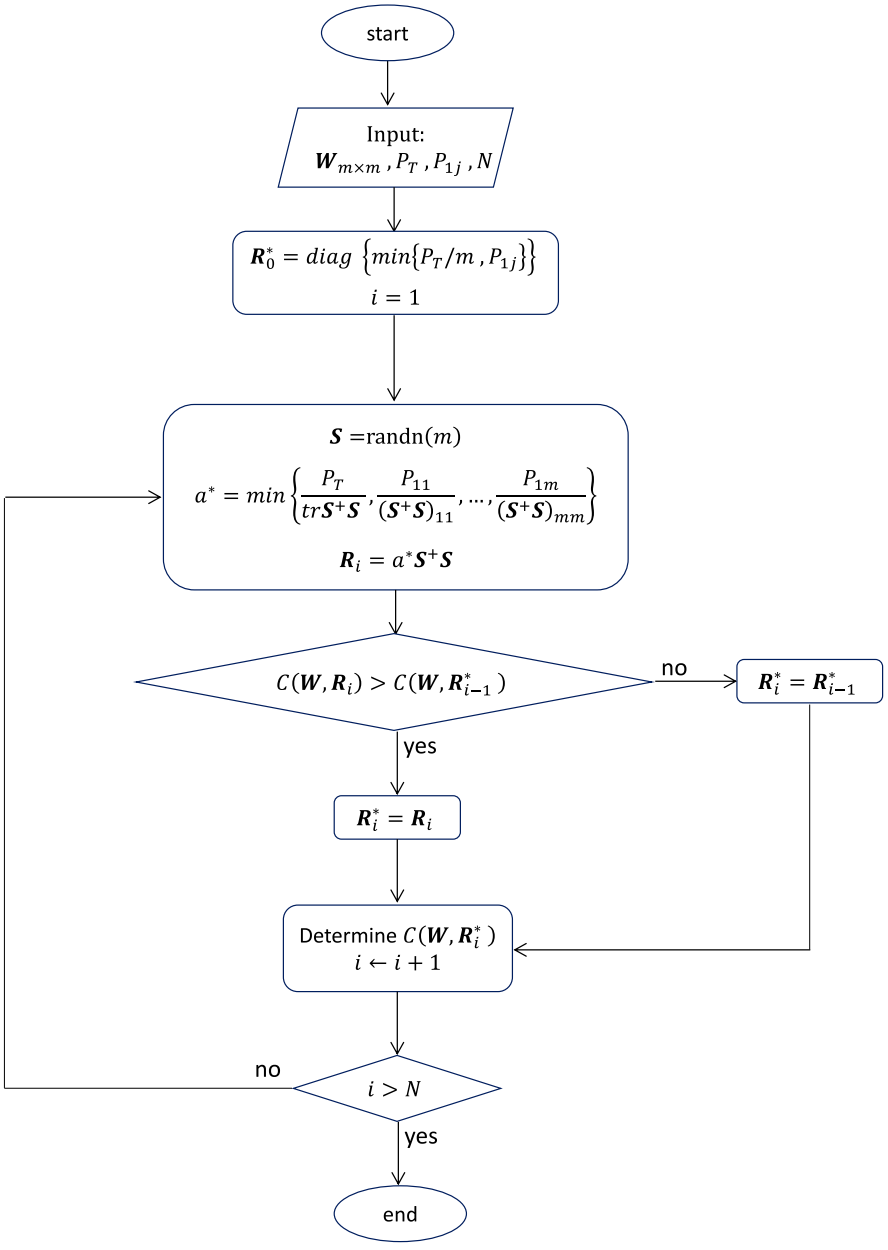

FIGURE 6.8: Monte-Carlo algorithm

6.4 CVX

CVX is an optimization tool that is widely used to compute the optimal value of a convex problem [74], [75]. Here, we show that CVX can give incorrect results for the channel capacity under the joint TPC and PAC constraints. To do so, let

us consider the following example:

$$\mathbf{W} = \begin{bmatrix} 10 & a \\ a & 2 \end{bmatrix}, \quad P_T = 1, \quad \{P_{1i}\} = \{0.3, 1\} \tag{6.34}$$

Fig. 6.9 shows the capacity of the channel $\mathbf{W}$ in (6.34) under the joint power constraints. For $1.5 \leq a \leq 4.4$, it is seen that the capacity obtained with CVX are less than those obtained with the Newton-barrier and Monte-Carlo algorithms. Note that the Monte-Carlo and Newton-barrier algorithms give the feasible Tx covariance. Also, although CVX was used with different solvers and different levels of accuracy, we observed that CVX can give incorrect results for the channel capacity in several examples[4].

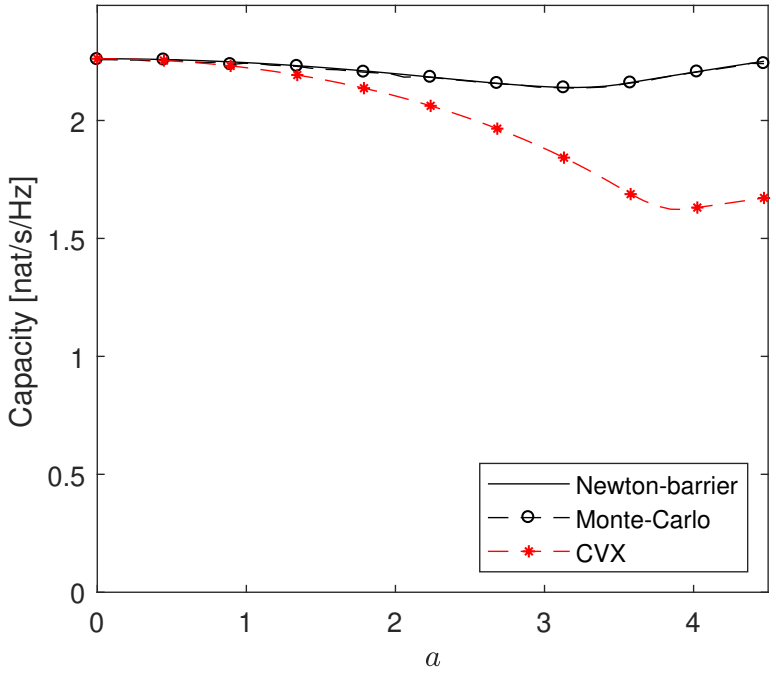

FIGURE 6.9: The channel capacity for the example in (6.34) obtained from the CVX tool as well as the Monte-Carlo and Newton-barrier algorithms.

[4]Under the joint constraints and for several examples, 360 types of CVX script were used, and none of them could give correct results for the capacity.

6.5 Summary

We develop some algorithms to find the optimal signaling. First, we develop the bisection algorithm which can obtain the optimal dual variable responsible for the TPC. This algorithm works with a monotonic function and the purpose is to find its root. In each step of this algorithm, the upper or lower bound of the root is redefined, and then the width of the uncertainty interval around the root is divided by two. Hence, an estimate of the root gets closer to its actual value as more steps are completed.

Second, the Newton-barrier method is developed, which is a powerful tool for solving convex optimization problems (such as the problem of finding optimal signaling under the TPC+PAC constraints). In this method, inequality constraints are added to an objective function by using the logarithmic barrier function.

Third, we develop the Monte-Carlo algorithm, which is a simple method for finding the channel capacity. The algorithm iteratively samples a large number of Tx covariances at random, and produces an increasing sequence of rates that get close to the capacity.

Both the Newton-barrier and Monte-Carlo algorithms compute the channel capacity under the joint TPC+PAC constraints for an arbitrary channel, which is not necessarily orthogonal. We compare the results of the Newton-barrier and Monte-Carlo algorithms, and a good agreement is found. Under the joint total and per-antenna power constraints, CVX cannot give correct results for the channel capacity.

Chapter 7

A Study of Favorable Propagation

We consider orthogonal and non-orthogonal channels with identical diagonal entries in their Gram matrices. An orthogonal channel (favorable propagation) is the best channel when the Tx covariance is diagonal (e.g., when different antennas' terminals are orthogonal to each other) [9]. Is this still the case when the Tx covariance is not diagonal? In this chapter, we wish to answer this question. In addition, we provide some corrections to the main theorems in [1] regarding favorable propagation conditions.

7.1 Capacity of Orthogonal and Non-Orthogonal Channels

In this section, we aim to answer the following question: is an orthogonal channel is the best among all channels with the same diagonal entries of their channel Gram matrices? To do so, we compare the capacity of channel $\mathbf{W}$ with that of orthogonal channel $\mathbf{D}_W$[1] (favorable propagation) under the total power constraint. Corollary 6 gives an answer to this question when the beamforming (rank-1 transmission) is optimal in both channels.

[1]The Gram matrix associated with an orthogonal channel is diagonal.

Corollary 6. *Let $d_i(\boldsymbol{W})$ and $\lambda_i(\boldsymbol{W})$ be the i-th diagonal entry and i-th eigenvalue of $\boldsymbol{W}$ respectively, and we assume that they are in decreasing order. Let the beamforming (rank-1 transmission) be optimal in both $\boldsymbol{W}$ and $\boldsymbol{D}_W$ under the TPC, i.e.,*

$$P_T \leq \min\left\{\left(d_2^{-1}(\boldsymbol{W}) - d_1^{-1}(\boldsymbol{W})\right), \left(\lambda_2^{-1}(\boldsymbol{W}) - \lambda_1^{-1}(\boldsymbol{W})\right)\right\} \tag{7.1}$$

which shows the low-SNR regime. Then:

$$C(\boldsymbol{W}) = \ln\left(1 + \lambda_1(\boldsymbol{W})P_T\right) \tag{7.2}$$

$$C(\boldsymbol{D_W}) = \ln\left(1 + d_1(\boldsymbol{W})P_T\right) \tag{7.3}$$

and $C(\boldsymbol{D}_W) \leq C(\boldsymbol{W})$, and the inequality is strict if $d_1(\boldsymbol{W}) < \lambda_1(\boldsymbol{W})$.

Proof. The beamforming is optimal in $\mathbf{W}$ under TPC if $P_T \leq \lambda_2^{-1}(\mathbf{W}) - \lambda_1^{-1}(\mathbf{W})$. As well, this is the case in $\mathbf{D}_W$, if $P_T \leq d_2^{-1}(\mathbf{W}) - d_1^{-1}(\mathbf{W})$. These conditions can be shown by the water-filling procedure (see Chapter 3). Hence, (7.1) follows. The proofs of (7.2) and (7.3) also follow from the water-filling procedure. Additionally, the maximum diagonal entry is less than or equal to the maximum eigenvalue for any positive semi-definite matrices [39]. Hence, $C(\mathbf{D}_W) \leq C(\mathbf{W})$ follows from $d_1(\mathbf{W}) \leq \lambda_1(\mathbf{W})$. □

This corollary shows that the orthogonal channel $\mathbf{D}_W$ does not necessarily provide more capacity compared to the non-orthogonal channel $\mathbf{W}$. Hence, an orthogonal channel (i.e., favorable propagation) is not necessarily the best channel among all channels with the same diagonal entries of $\mathbf{W}$. However, if the Tx covariance is caused to be diagonal, an orthogonal channel is the best [9].

Here, we present an example to show that an orthogonal channel can have lower

capacity compared to a non-orthogonal channel. We consider the following channel:

$$\mathbf{W} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad P_T \leq 0.75 \tag{7.4}$$

In this example, $\lambda_1(\mathbf{W}) = 5, \lambda_2(\mathbf{W}) = 0$ and $d_1(\mathbf{W}) = 4, d_2(\mathbf{W}) = 1$. For both $\mathbf{W}$ and $\mathbf{D}_W$ under the TPC, the beamforming (rank-1 transmission) is optimal, since $P_T \leq 0.75$ satisfies the condition in (7.1). We observe:

$$C(\mathbf{D}_W) = \ln(1 + 4P_T) \tag{7.5}$$

$$C(\mathbf{W}) = \ln(1 + 5P_T) \tag{7.6}$$

which together imply $C(\mathbf{D}_W) < C(\mathbf{W})$.

The next proposition gives a sufficient and necessary condition when $d_1(\mathbf{W}) = \lambda_1(\mathbf{W})$. In this case, if the beamforming is optimal under the TPC in both $\mathbf{D}_W$ and $\mathbf{W}$, then $C(\mathbf{W}) = C(\mathbf{D}_W)$.

Proposition 13. *Let $\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_m$ be the standard basis vectors for $\mathcal{R}^m$. There exists at least one i which $d_1(\boldsymbol{W}) = (\boldsymbol{W})_{ii}$. Then, $d_1(\boldsymbol{W}) = \lambda_1(\boldsymbol{W})$ if and only if:*

$$\boldsymbol{e}_i \in \{\boldsymbol{x} : \boldsymbol{W}\boldsymbol{x} = \lambda_1(\boldsymbol{W})\boldsymbol{x}\} \tag{7.7}$$

where $\{\boldsymbol{x} : \boldsymbol{W}\boldsymbol{x} = \lambda_1(\boldsymbol{W})\boldsymbol{x}\}$ is an eigenspace associated with $\lambda_1(\boldsymbol{W})$[2]. When there are multiple maximum diagonal entries, then $d_1(\boldsymbol{W}) = \lambda_1(\boldsymbol{W})$ if and only if (7.7) holds for at least one solution of $i = \arg\max_j (\boldsymbol{W})_{jj}$.

Proof. See Appendix. □

[2]The dimension of this eigenspace is equal to the number of repeated maximum eigenvalues [68].

7.2 On Favorable Propagation Conditions

Here, some corrections are provided for the theorems in [1]. This paper studies favorable propagation in massive MIMO channels. We show that Theorems 1 and 2 in [1] are not true under the stated conditions, and provide some modifications for the conditions to make their conclusions valid.

In this section, we use the same notations as in [1]. Also, we assume that all exception terms are bounded.

Paper [1] obtains the exception of the inner product of channel vectors as follows:

$$\frac{1}{M}\mathbb{E}\{\mathbf{g}_i^H\mathbf{g}_k\} = \sum_{r=1}^{L}\sum_{s=1}^{L}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^* w_{ms}\alpha_{ri}^*\alpha_{sk}\}\right) \tag{7.8}$$

where this is normalized by M, which is the number of BS antennas; $\mathbf{g}_i^H$ represents the Hermitian transpose of $\mathbf{g}_i$; K is the number of single-antenna users; L is the number of paths from all users to the BS; $\mathbf{g}_i$ is the i-th channel vector; w_{mr} is the mr-th entry of a steering matrix; and α_{sk} shows the signal change for the k-th user over the s-th path.

Theorem 1 in [1] is as follows:

$$\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\} \to 0, \quad \text{if} \quad \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^* w_{ms}\} \to 0 \tag{7.9}$$

when $\alpha_{ri}^*\alpha_{sk}$ and $w_{mr}^* w_{ms}$ are uncorrelated.

To see the gap in Theorem 1 in [1], we observe that:

$$||\mathbf{w}_r||^2 = \sum_{m=1}^{M}|w_{mr}|^2 = M \tag{7.10}$$

where $\mathbf{w}_r = [w_{1r}, ..., w_{Mr}]^T$, $||\cdot||$ denotes the Euclidean norm and the last equality follows from the fact the $w_{mr} = e^{j\theta}$ (e.g., (23), (28) in [1]). Hence:

$$\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{|w_{mr}|^2\} = 1 \tag{7.11}$$

So, the condition in (7.9) does not hold, since

$$\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^* w_{mr}\} = \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{||\mathbf{w}_r||^2\} = 1 \tag{7.12}$$

for arbitrary M, whether it is large or not. Note that the terms for $r = s$ exist in (7.9) (e.g., (12) and (13) in [1]).

To fix the gap in (7.9) (Theorem 1 in [1]), we need a modification and an extra assumption, which are given in the next proposition.

Proposition 14. *Let us assume that the propagation channels vectors corresponding to different users are orthogonal to each other, i.e.:*

$$\mathbb{E}\Big\{\boldsymbol{v}_i^H \boldsymbol{v}_k\Big\} = \sum_r \mathbb{E}\{\alpha_{ri}^* \alpha_{rk}\} = 0, \quad i \neq k \tag{7.13}$$

where $\boldsymbol{v}_k = [\alpha_{1k}...\alpha_{Lk}]^T$. *Then, the following condition:*

$$\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^* w_{ms}\} \to 0, \quad r \neq s \tag{7.14}$$

implies that the asymptotic favorable propagation exists "on average", i.e.:

$$\mathbb{E}\Big\{\frac{1}{M}\boldsymbol{g}_i^H \boldsymbol{g}_k\Big\} \to 0, \quad i \neq k \tag{7.15}$$

when $\alpha_{ri}^* \alpha_{sk}$ *and* $w_{mr}^* w_{ms}$ *are uncorrelated.*

Proof. We observe that:

$$\begin{aligned}\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\} &\overset{\text{(a)}}{=} \frac{1}{M}\sum_{r=1}^{L}\sum_{s=1}^{L}\mathbb{E}\{\alpha_{ri}^*\alpha_{sk}\}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^*w_{ms}\}\\ &\overset{\text{(b)}}{=} \frac{1}{M}\sum_{r\neq s}\mathbb{E}\{\alpha_{ri}^*\alpha_{sk}\}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^*w_{ms}\}+\sum_{r}\mathbb{E}\{\alpha_{ri}^*\alpha_{rk}\}\\ &\overset{\text{(c)}}{=} \frac{1}{M}\sum_{r\neq s}\mathbb{E}\{\alpha_{ri}^*\alpha_{sk}\}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^*w_{ms}\}\overset{\text{(d)}}{\to}0\end{aligned} \tag{7.16}$$

where (a) follows from (7.8) when $\alpha_{ri}^*\alpha_{sk}$ and $w_{mr}^*w_{ms}$ are uncorrelated; (b) follows from (7.11); and (c) and (d) follow from (7.13) and (7.14) respectively. □

The differences between Proposition 14 and Theorem 1 in [1] are as follows. First, the condition in (7.13) is added. Second, the orthogonality in (7.14) exists for $r \neq s$ only.

Theorem 2 in [1] shows that (7.9) holds if $|\alpha_{lk}|$ is bounded for any l and k, i.e., $|\alpha_{lk}| \leq C_\alpha$ (in this theorem, no assumption is considered regarding the correlation between $\alpha_{ri}^*\alpha_{sk}$ and $w_{mr}^*w_{ms}$). To show this, the following inequality is given in [1]:

$$\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\} \leq \sum_{r=1}^{L}\sum_{s=1}^{L}C_\alpha^2\left(\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^*w_{ms}\}\right) \tag{7.17}$$

Here, we show that Theorem 2 in [1] is not true in several different ways.

1. We observe that both sides of (7.17) are complex numbers in general, but complex numbers cannot be compared in the same way as real numbers[3]. So, the inequality in (7.17) is not correct in general.

2. Now, we show that even if both sides of (7.17) are real numbers (or their difference is real), then (7.9) does not hold if α_{lk} is bounded. To do so, note that $\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\}$ can be negative, e.g., $\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\} = -1$. This can happen even if

[3]If the difference between two complex numbers is a real number, then one can compare them. However, this is not necessarily the case in (7.17).

the upper bound is zero. In other words, to show $\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\} \to 0$, both its upper and lower bounds should converge to zero, not just the upper bound.

3. Here, we show that if α_{lk} is bounded, then the inequality in (7.17) is not true even when all terms are real. To do so, we consider the following example for arbitrary L, M:

$$w_{1r} = w_{2r} = ... = w_{Mr} = \alpha_{ri} = \alpha_{rk}, \quad r = 1, ..., L \tag{7.18}$$

To simplify the notations, we define a new random variable q_r where:

$$w_{1r} = w_{2r} = ... = w_{Mr} = \alpha_{ri} = \alpha_{rk} = q_r, \quad r = 1, ..., L \tag{7.19}$$

and we assume that the random variables q_r and q_k are independent of each other for $r \neq k$. We also assume that $q_r = \pm 1$ with equal probability, which shows that α_{ri} are bounded and $C_\alpha = 1$. Observe that:

$$\begin{aligned}\mathbb{E}\left\{\frac{1}{M}\mathbf{g}_i^H\mathbf{g}_k\right\} &= \frac{1}{M}\sum_{r=1}^{L}\sum_{\substack{s=1\\s\neq r}}^{L}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}w_{ms}\alpha_{ri}\alpha_{sk}\} + \frac{1}{M}\sum_{r=1}^{L}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^2\alpha_{ri}\alpha_{rk}\} \\ &= \frac{1}{M}\sum_{r=1}^{L}\sum_{\substack{s=1\\s\neq r}}^{L}\sum_{m=1}^{M}\mathbb{E}\{q_r^2q_s^2\} + \frac{1}{M}\sum_{r=1}^{L}\sum_{m=1}^{M}\mathbb{E}\{q_r^4\} \\ &\overset{(a)}{=} L(L-1) + L = L^2\end{aligned} \tag{7.20}$$

where (a) follows from the fact that q_r and q_s are independent of each other and $q_r^2 = q_s^2 = q_r^4 = 1$. The right-hand side of (7.17) with $C_\alpha = 1$ is as follows:

$$\begin{aligned}\sum_{r=1}^{L}\sum_{s=1}^{L}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{w_{mr}^*w_{ms}\}\right) &= \sum_{r=1}^{L}\sum_{\substack{s=1\\s\neq r}}^{L}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{q_rq_s\}\right) + \sum_{r=1}^{L}\left(\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\{q_r^2\}\right) \\ &= 0 + L = L\end{aligned} \tag{7.21}$$

By comparing (7.20) with (7.21), we can see that the right-hand side of (7.17) is less than the left-hand side for $L \geq 2$. Hence, the inequality in (7.17) is not correct in general even when all terms are real.

Following the above discussion, the next proposition shows that favorable propagation can occur when α_{ri} and w_{ms} are dependent on each other.

Proposition 15. *Let us assume that $\alpha_{ri} = a_r b_i$ for all r and i. Also, the mean is zero for random variables b_i, which are all independent of each other as well as independent of a_r and w_{mr} for all r, i, and m. Then, favorable propagation exists "on average":*

$$\mathbb{E}\left\{ \frac{1}{M} \boldsymbol{g}_i^H \boldsymbol{g}_k \right\} = 0, \quad i \neq k \tag{7.22}$$

Proof. Observe that:

$$\begin{aligned} \mathbb{E}\left\{ \frac{1}{M} \mathbf{g}_i^H \mathbf{g}_k \right\} &= \sum_{r=1}^{L} \sum_{s=1}^{L} \left(\frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\{ w_{mr}^* w_{ms} \alpha_{ri}^* \alpha_{sk} \} \right) \\ &= \sum_{r=1}^{L} \sum_{s=1}^{L} \left(\frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\{ w_{mr}^* w_{ms} a_r^* a_s \} \mathbb{E}\{ b_i^* b_k \} \right) = 0 \end{aligned} \tag{7.23}$$

where the second equality follows from the fact that b_i and b_k are independent from the rest of the variables. Also, b_i and b_k are independent of each other and their mean values are zero. Hence, the last equality follows from $\mathbb{E}\{b_i^* b_k\} = \mathbb{E}\{b_i^*\}\mathbb{E}\{b_k\} = 0$. □

The model in Proposition 15 is shown in Fig. 7.1, where b_i and a_r represent scattering factors around the user and around the BS respectively. This factorization of propagation coefficients is also confirmed experimentally for some environments [76].

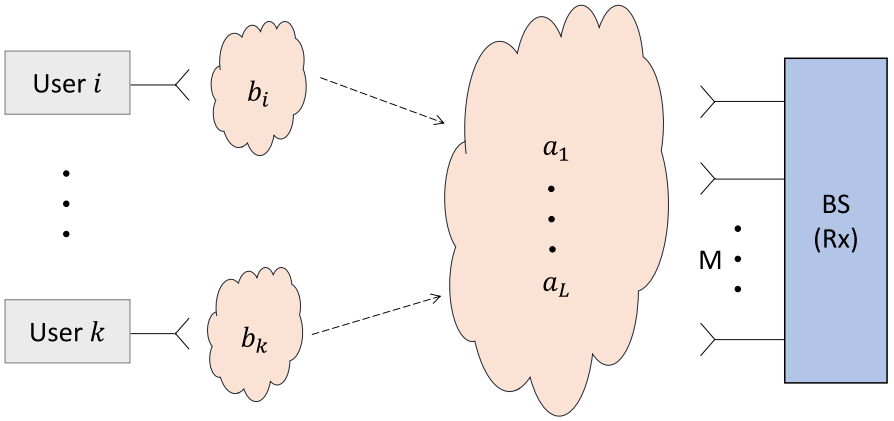

FIGURE 7.1: Illustration of the channel model in Proposition 15 where $\alpha_{ri} = a_r b_i$.

7.3 Summary

We consider orthogonal and non-orthogonal channels for which Gram matrices have the same diagonal entries. We show that when the beamforming (rank-1 transmission) is optimal in both channels under the total power constraint, then the capacity of the the non-orthogonal channel is greater than that of the receptive orthogonal channel (i.e., favorable propagation). This shows that an orthogonal channel is not necessarily the best among all channels with the same diagonal entries in their Gram matrices.

We comment on [1] that its main theorems are not correct in general, and we propose some modifications to make the conclusions valid. We consider the case where the steering matrix and the orthogonal propagation channel matrix are uncorrelated. In this case, if the normalized inner product of the different steering vectors converges to zero, then favorable propagation exists.

Chapter 8

Conclusion

8.1 Thesis Summary

In this thesis, the capacity and optimal signaling for a massive MIMO channel are studied. The optimal signaling under the total power constraint (TPC) is on the channel eigenmodes, and the optimal power allocation (OPA) is obtained by the water-filling (WF) procedure. This case is motivated by the limited power at the transmitter power supply. In addition to the TPC, the per-antenna power constraints (PACs) are practically important, since there is a power constraint on each per-antenna amplifier. The optimal signaling under the joint total and per-antenna power constraints is not known in general, except in some special cases (e.g., MISO channels, full-rank optimal transmit covariance). The considered problem involves finding the optimal transmit covariance under the joint TPC+PAC constraints for a favorable propagation (FP) channel. Under this propagation, different channel vectors become orthogonal to each other, and the channel Gram matrix becomes diagonal. Both measurement and theory-based results verify that FP (or more precisely, nearly favorable propagation) is observed for many massive MIMO scenarios.

A closed-form solution for optimal signaling for an orthogonal channel under the joint power constraints (TPC+PAC) is obtained. It is diagonal, which shows the

optimality of the independent signaling, and the diagonal entries are the minimum of two terms: one is the per-antenna power constraints, and the other represents the WF procedure. However, under the joint power constraints (TPC+PAC), the optimal dual variable responsible for the TPC is different from that in the WF procedure (TPC only). Since the considered problem is convex and strong duality holds, then the dual variable shows sensitivity with respect to constraint changes. This reveals that the optimal signaling under the joint total and per-antenna power constraints is less sensitive to the TPC variations than that in the WF procedure. The numbers of active streams and active PACs are determined analytically in closed forms, which shows that active streams are those that are stronger than inactive ones. Optimality conditions of rank-1 transmission (i.e., beamforming) and full-rank transmission under the joint power constraints (TPC+PAC) are determined. These are somewhat similar (but not identical) to those in the WF procedure (TPC alone), with a different SNR threshold.

We show that the OPA under the joint TPC+ PAC constraints can be interpreted as pouring water into a container which has ceiling and floor profiles, i.e., as "capped WF". The floor profile is determined by the inverse of the channel gains, and the ceiling profile is lifted from the floor by the receptive per-antenna power constraints. To compute the OPA under the joint constraints, an iterative WF algorithm is proposed. It is proven that the algorithm converges to the OPA after a finite number of steps.

We also consider the case where each user has a different allocated bandwidth and that the grade of service is not the same for all users. In this case and under the TPC+PAC constraints, the OPA and the numbers of active streams and active PACs are found in closed forms.

Next, we study the robustness of the optimal signaling under FP. To do so, we consider nearly favorable propagation (NFP), in which the values of off-diagonal entries in the channel Gram matrix are small but non-zero (nearly orthogonal channel). It is observed that the optimal transmit covariance under FP is nearly optimal under NFP. An upper bound of the sub-optimality gap (rate loss) is found,

and we see that this upper bound becomes small under NFP. A new definition of nearly (or ϵ)-favorable propagation is given, which quantifies how close the channel is to favorable propagation.

Three algorithms are developed which compute optimal signaling under the joint constraints (TPC+PAC). The first one is the bisection algorithm, which computes the optimal dual variable responsible for the TPC. This algorithm works with monotonic continuous functions, and in each step, the width of the uncertainty interval around its root reduces by a factor of two. To do so, the upper or lower bound of the root is updated in each step. Hence, as more iterations are completed, this width becomes smaller, and the difference between an estimate of the root and its actual value also becomes smaller. Second, we develop the Newton-barrier method which uses logarithmic-barrier functions to handle inequality constraints (i.e., TPC+PAC). The third one is a Monte-Carlo algorithm which randomly samples a considerable number of feasible transmit covariances. At each step, the best transmit covariance is selected among the existing ones. Then, an increasing sequence of rates is obtained in order to achieve the channel capacity under the joint power constraints (TPC+PAC).

In this thesis, favorable propagation (i.e., orthogonal channel) is studied in more detail. We consider orthogonal and non-orthogonal channels under the TPC when their Gram matrices have the same diagonal entries. In this case, if the beamforming (rank-1 transmission) is optimal in both cases, then the capacity of a non-orthogonal channel is greater than that in an orthogonal channel. So, considering the fixed diagonal entries of the channel Gram matrix, an orthogonal channel is not necessarily the best channel to deliver more capacity.

Finally, we show that the main theorems in [1] are not correct in general, and that some modifications are needed to make their conclusion valid, which are given in this thesis.

8.2 Possible Future Research Topics

This study can be expanded in the future, as follows.

- In addition to the total and per-antenna power constraints, one can also consider more constraints, i.e., (i) per-group antenna power constraints, (ii) per-antenna power constraints at both transmitter and receiver antennas, and (iii) an interference power constraint. So, optimal signaling under the mentioned constraints for an orthogonal channel can be considered as a new research topic.

- Optimal transmit covariance for an orthogonal wiretap channel under the TPC is known; however, optimal signaling under the joint TPC and PAC is still an open problem, and can thus be pursued.

- This study is motivated by some practical considerations such as the per-antenna power constraint. However, there are more factors which can also be considered. One of them is impairment at the transmitter and receiver hardware. So, this element can be explored following the problem considered in this thesis.

Chapter 9

Appendix

9.1 Proof of Theorem 1

First, we show that an optimal Tx covariance is diagonal under the joint power constraints (TPC+PAC) for an orthogonal channel $\mathbf{W} = \mathbf{D}_W$ (favorable propagation). This is obtained via an approach similar to that in [15] under the TPC only. Here, we use Hadamard inequality [39], as follows:

$$\begin{aligned}
\ln|\mathbf{I}+\mathbf{D}_W\mathbf{R}| &\leq \sum_{i=1}^{m} \ln\left(1+(\mathbf{D}_W\mathbf{R})_{ii}\right) \\
&= \sum_{i=1}^{m} \ln\left(1+(\mathbf{D}_W)_{ii}(\mathbf{R})_{ii}\right) \\
&= \sum_{i=1}^{m} \ln\left(1+(\mathbf{D}_W)_{ii}(\mathbf{D}_R)_{ii}\right) \\
&= \ln|\mathbf{I}+\mathbf{D}_W\mathbf{D}_R| \qquad (9.1)
\end{aligned}$$

where $\mathbf{D}_W$ and $\mathbf{D}_R$ are diagonal matrices that have the same diagonal entries as $\mathbf{W}$ and $\mathbf{R}$, respectively. Here, we assume that $\mathbf{D}_W > 0$ (the full-rank matrix)[1]. The inequality is strict when $\mathbf{R} \neq \mathbf{D}_R$, and the upper bound is achieved when

[1]Note that $(\mathbf{D}_W)_{ii} = 0$ corresponds to $\mathbf{h}_i = 0$, and those sub-channels do not affect to the capacity. Hence, they can be removed from consideration. If these sub-channels are not removed and $\mathbf{D}_W$ is singular, then $\mathbf{R}^*(\mathbf{D}_W)$ may not be unique and there may exist non-diagonal $\mathbf{R}^*(\mathbf{D}_W)$, while there always exists a diagonal optimal Tx covariance for a singular $\mathbf{D}_W$.

$\mathbf{R} = \mathbf{D}_R$ [39]. Considering the feasible set S_R in (4.2), $\mathbf{R} \in S_R$ implies $\mathbf{D}_R \in S_R$, because $tr(\mathbf{R}) = tr(\mathbf{D}_R)$ and $(\mathbf{R})_{ii} = (\mathbf{D}_R)_{ii}$. The former shows that $tr(\mathbf{D}_R) \leq P_T$ and the latter shows that $(\mathbf{D}_R)_{ii} \leq P_{1i}$, and hence $\mathbf{D}_R$ is in the feasible set of Tx covariance matrices. So, an optimal covariance matrix has to be diagonal, i.e., $\mathbf{R}^* = \mathbf{D}_{R^*}$.

Therefore, the problem in (4.1) can be expressed as follows:

$$C = \max_{p_i \in S_p} \sum_{i=1}^{m} \ln(1 + g_i p_i) \tag{9.2}$$

where

$$S_p = \left\{ \{p_i\} : p_i \geq 0,\ \sum_{i=1}^{m} p_i \leq P_T,\ p_i \leq P_{1i} \right\} \tag{9.3}$$

The problem in (9.2) is (strictly) convex since this is a maximization of the (strictly) concave function, and the inequality constraints are also convex functions. Also, Slater's condition holds (we assume that $P_T, P_{1i} > 0$). Hence, the KKT (Karush-Kuhn-Tucker) conditions are both sufficient and necessary for optimality [66]. The optimal Tx covariance is unique, since the objective function is strictly concave. The Lagrangian of the problem in (9.2) is as follows:

$$L = -\sum_i \ln(1 + g_i p_i) + \mu\Big(\sum_i p_i - P_T\Big) + \sum_i \big(\lambda_i(p_i - P_1) - \eta_i p_i\big) \tag{9.4}$$

where μ and λ_i are the Lagrange multipliers responsible for the TPC and PACs respectively, and η_i is the Lagrange multiplier responsible for the non-negative constraint $p_i \geq 0$. The KKT conditions are as follows:

$$\frac{\partial L}{\partial p_i} = -\frac{g_i}{1 + g_i p_i} + \mu - \eta_i + \lambda_i = 0 \tag{9.5}$$

$$\mu\Big(\sum_i p_i - P_T\Big) = 0, \quad \eta_i p_i = 0, \quad \lambda_i(p_i - P_{1i}) = 0 \tag{9.6}$$

$$\sum_i p_i \leq P_T, \quad p_i \geq 0, \quad p_i \leq P_{1i} \tag{9.7}$$

$$\mu \geq 0, \quad \eta_i \geq 0, \quad \lambda_i \geq 0 \tag{9.8}$$

where (9.5) and (9.6) are stationary and complementary slackness conditions respectively, and (9.7) and (9.8) are primal and dual feasibility constraints respectively. Using (9.5),

$$p_i = (\mu - \eta_i + \lambda_i)^{-1} - g_i^{-1} \tag{9.9}$$

An active stream $p_i > 0$ implies $\eta_i = 0$, which follows from the complementary slackness in (9.6). Hence, (9.9) simplifies to the following for active antennas:

$$0 < p_i = (\mu + \lambda_i)^{-1} - g_i^{-1} \leq P_{1i} \tag{9.10}$$

where the inequality follows from the PAC. For active PACs, this becomes equality, and $p_i = P_{1i}$. Hence, in this case:

$$\lambda_i = \left(P_{1i} + g_i^{-1}\right)^{-1} - \mu \geq 0 \tag{9.11}$$

Therefore, $p_i = P_{1i}$ if:

$$\mu \leq \left(P_{1i} + g_i^{-1}\right)^{-1} \tag{9.12}$$

An active stream with an inactive PAC implies $\lambda_i = \eta_i = 0$. So, for this case, (9.10) implies:

$$0 < p_i = \mu^{-1} - g_i^{-1} < P_{1i} \tag{9.13}$$

Thus, $p_i = \mu^{-1} - g_i^{-1}$ if:

$$\left(P_{1i} + g_i^{-1}\right)^{-1} < \mu < g_i \tag{9.14}$$

Finally, an inactive stream $p_i = 0$ yields $\lambda_i = 0$ and hence by (9.9),

$$\mu \geq g_i \tag{9.15}$$

Combining all these cases and after some manipulation, p_i is obtained as follows:

$$p_i = \begin{cases} 0 & \mu^{-1} - g_i^{-1} \leq 0 \\ \mu^{-1} - g_i^{-1} & 0 < \mu^{-1} - g_i^{-1} < P_{1i} \\ P_{1i} & \mu^{-1} - g_i^{-1} \geq P_{1i} \end{cases} \tag{9.16}$$

Simply put, we obtain:

$$p_i = \begin{cases} (\mu^{-1} - g_i^{-1})_+ & \mu^{-1} - g_i^{-1} < P_{1i} \\ P_{1i} & \mu^{-1} - g_i^{-1} \geq P_{1i} \end{cases} \tag{9.17}$$

So, the optimal power allocation can be expressed as $p_i^* = \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\}$.

When the TPC is inactive, $P_T \geq \sum_{i=1}^{m} P_{1i}$ and $\mu = 0$. Hence, from (9.9), $\lambda_i > 0$ and therefore $p_i^* = P_{1i}$, i.e., each PAC is active. If $P_T < \sum_{i=1}^{m} P_{1i}$, the TPC is active and there exists at least one inactive PAC. Therefore, $\lambda_i = 0$ for one particular i, and (9.9) implies $\mu > 0$. So, from the complementary slackness, $\mu > 0$ is obtained as a solution of $\sum_{i=1}^{m} p_i^* = P_T$.

The channel capacity in (4.6) is obtained with $p_i = p_i^*$ in (9.2).

9.2 Proof of Proposition 1

Here, we consider the case that the TPC is active, and hence the total Tx power is equal to P_T. Then, in Remark 9.1, we will show that the Proposition 1 also holds when the TPC is inactive.

Consider the OPA in (4.4): $p_i^* = \min\{P_{1i},\ (\mu^{-1} - g_i^{-1})_+\}$. This implies $\mu^{-1} - g_i^{-1} > 0$ for active streams and $\mu^{-1} - g_i^{-1} \leq 0$ for inactive ones. Since g_i are in decreasing order,

$$\mu^{-1} - g_1^{-1} \geq \mu^{-1} - g_2^{-1} \geq ... \geq \mu^{-1} - g_m^{-1} \tag{9.18}$$

thus:

$$\mu^{-1} - g_{m_+}^{-1} > 0, \quad \mu^{-1} - g_{m_++1}^{-1} \leq 0 \tag{9.19}$$

and hence:

$$g_{m_+}^{-1} < \mu^{-1} \leq g_{m_++1}^{-1} \tag{9.20}$$

It follows from (9.20) that (see Remark 9.1):

$$\sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})_+\right\} < \sum_{i=1}^{m} \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} = P_T \tag{9.21}$$

$$\leq \sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_++1}^{-1} - g_i^{-1})_+\right\} \tag{9.22}$$

Also, g_i^{-1} are in increasing order. Hence, $(g_j^{-1} - g_i^{-1})_+ = (g_j^{-1} - g_i^{-1})$ for any $i \leq j$, and otherwise $(g_j^{-1} - g_i^{-1})_+ = 0$. So:

$$\sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})_+\right\} = \sum_{i=1}^{m_+} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})\right\} + \sum_{i=m_++1}^{m} \min\left\{P_{1i}, 0\right\}$$

$$= \sum_{i=1}^{m_+} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})\right\} \tag{9.23}$$

Following the same approach, we obtain:

$$\sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_++1}^{-1} - g_i^{-1})_+\right\} = \sum_{i=1}^{m_++1} \min\left\{P_{1i}, (g_{m_++1}^{-1} - g_i^{-1})\right\} \tag{9.24}$$

Applying (9.23) and (9.24) to (9.22):

$$\sum_{i=1}^{m_+} \min\left\{P_{1i},\ g_{m_+}^{-1} - g_i^{-1}\right\} < P_T \leq \sum_{i=1}^{m_++1} \min\left\{P_{1i},\ g_{m_++1}^{-1} - g_i^{-1}\right\} \tag{9.25}$$

Hence, $u_{m_+} < P_T \leq u_{m_++1}$. Also, u_j are in increasing order because of the increasing order of g_i^{-1}. Hence:

$$u_1 \leq ... \leq u_{m_+} < P_T \leq u_{m_++1} \tag{9.26}$$

and (4.9) follows.

Remark 9.1. In the following, we demonstrate that the inequality in (9.21) is always strict. Here, we consider three possible cases: (i) $P_T < \sum_{i=1}^{m} P_{1i}$ and $m_+ > m_{PAC}$; (ii) $P_T < \sum_{i=1}^{m} P_{1i}$ and $m_+ = m_{PAC}$; and (iii) $P_T \geq \sum_{i=1}^{m} P_{1i}$.
(i) Consider $P_T < \sum_{i=1}^{m} P_{1i}$ and $m_+ > m_{PAC}$. Then, for active streams, at least one PAC is inactive, and let it be the t-th one. So:

$$\begin{aligned} \min\{P_{1t}, (g_{m_+}^{-1} - g_t^{-1})_+\} &\leq (g_{m_+}^{-1} - g_t^{-1})_+ \\ &< \mu^{-1} - g_t^{-1} \\ &= \min\{P_{1t}, (\mu^{-1} - g_t^{-1})_+\} \end{aligned} \tag{9.27}$$

where the strict inequality follows from (9.20), and (9.20) implies:

$$\sum_{i=1, i\neq t}^{m} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})_+\right\} \leq \sum_{i=1, i\neq t}^{m} \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} \tag{9.28}$$

Thus, using (9.27) and (9.28):

$$\sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})_+\right\} < \sum_{i=1}^{m} \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} = P_T \tag{9.29}$$

which shows the strict inequality in (9.21).
(ii) Consider $P_T < \sum_{i=1}^{m} P_{1i}$ and $m_+ = m_{PAC}$. Then, for active streams, all PACs

are active. This implies $P_T = \sum_{i=1}^{m_+} P_{1i}$. Then:

$$\begin{aligned} P_T &= P_{1m_+} + \sum_{i=1}^{m_+-1} P_{1i} \\ &\geq P_{1m_+} + \sum_{i=1}^{m_+-1} \min\{P_{1i}, g_{m_+}^{-1} - g_i^{-1}\} \\ &= P_{1m_+} + \sum_{i=1}^{m_+} \min\{P_{1i}, g_{m_+}^{-1} - g_i^{-1}\} \\ &> \sum_{i=1}^{m_+} \min\{P_{1i}, g_{m_+}^{-1} - g_i^{-1}\} \\ &= \sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})_+\right\} \end{aligned} \tag{9.30}$$

where the last equality follows from (9.23). Hence, $\sum_{i=1}^{m} \min\left\{P_{1i}, (g_{m_+}^{-1} - g_i^{-1})_+\right\} < P_T$, and the strict inequality in (9.21) follows.

(iii) Consider $P_T \geq \sum_{i=1}^{m} P_{1i}$. Here, $m_+ = m$, and there is no upper bound for μ^{-1} in (9.20) because g_{m+1} does not exist. In this case, each PAC is active. Observe that:

$$\begin{aligned} \sum_{i=1}^{m} \min\left\{P_{1i}, (g_m^{-1} - g_i^{-1})_+\right\} &= \sum_{i=1}^{m-1} \min\left\{P_{1i}, (g_m^{-1} - g_i^{-1})_+\right\} \\ &\leq \sum_{i=1}^{m-1} P_{1i} \\ &< P_T \end{aligned} \tag{9.31}$$

This implies $u_m < P_T$, and the strict inequality in (9.21) follows.

9.3 Proof of Proposition 4

Here, we obtain a proof when the TPC is active. Then, in Remark 9.2, we will show that Proposition 4 also holds when the TPC is inactive.

To find the active PACs, we consider the streams with $p_i^* > 0$ obtained from Proposition 1. Then, (4.4) for the active streams becomes:

$$p_i^* = \min\{0, \mu^{-1} - (g_i^{-1} + P_{1i})\} + P_{1i} \tag{9.32}$$

This implies $\mu^{-1} - (g_i^{-1} + P_{1i}) \geq 0$ for active PACs and $\mu^{-1} - (g_i^{-1} + P_{1i}) < 0$ for inactive ones. Since $g_i^{-1} + P_{1i}$ is in increasing order:

$$\mu^{-1} - (g_1^{-1} + P_{11}) \geq ... \geq \mu^{-1} - (g_{m_+}^{-1} + P_{1m_+}) \tag{9.33}$$

so that:

$$g_{m_{PAC}}^{-1} + P_{1m_{PAC}} \leq \mu^{-1} < g_{m_{PAC}+1}^{-1} + P_{1m_{PAC}+1} \tag{9.34}$$

and hence:

$$\begin{aligned}\sum_{i=1}^{m_+} \Big(\min\left\{0, (g_{m_{PAC}}^{-1} + P_{1m_{PAC}}) - (g_i^{-1} - P_{1i})\right\} + P_{1i}\Big) \\ \leq \sum_{i=1}^{m_+} \min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})\right\} = P_T \\ < \sum_{i=1}^{m_+} \Big(\min\left\{0, (g_{m_{PAC}+1}^{-1} + P_{1m_{PAC}+1}) - (g_i^{-1} - P_{1i})\right\} + P_{1i}\Big)\end{aligned} \tag{9.35}$$

Using the same approach as in Proposition 1, we obtain:

$$\sum_{i=1}^{m_+} P_{1i} + \sum_{i=m_{PAC}}^{m_+} \left((g_{m_{PAC}}^{-1} + P_{1m_{PAC}}) - (g_i^{-1} - P_{1i})\right) \leq P_T \tag{9.36}$$

$$< \sum_{i=1}^{m_+} P_{1i} + \sum_{i=m_{PAC}+1}^{m_+} \left((g_{m_{PAC}+1}^{-1} + +P_{1m_{PAC}+1}) - (g_i^{-1} - P_{1i})\right) \tag{9.37}$$

So:

$$\sum_{i=1}^{m_+} P_{1i} - v_{m_{PAC}} \leq P_T < \sum_{i=1}^{m_+} P_{1i} - v_{m_{PAC}+1} \tag{9.38}$$

Hence, $v_{m_{PAC}+1} < \sum_{i=1}^{m_+} P_{1i} - P_T \leq v_{m_{PAC}}$, and the lower bound holds if $m_{PAC} < m_+$. As well, v_j are in decreasing order because g_i^{-1} are in increasing order. Therefore (4.15) follows.

Remark 9.2. Here, we show that the result for Proposition 4 holds for the special case of $m_{PAC} = m_+$, whether the TPC is active or not.

If $m_{PAC} = m_+$, then there is no upper bound for μ^{-1} in (9.34). In this case, if the TPC is active, then $P_T = \sum_{i=1}^{m_+} P_{1i}$. If not, then all streams are active (i.e., $m_+ = m$), and hence $P_T \geq \sum_{i=1}^{m_+} P_{1i} = \sum_{i=1}^{m} P_{1i}$. Hence, if $m_{PAC} = m_+$, then:

$$1. \quad \text{active TPC:} \quad \sum_{i=1}^{m_+} P_{1i} - P_T = 0 \tag{9.39}$$

$$2. \quad \text{inactive TPC:} \quad \sum_{i=1}^{m_+} P_{1i} - P_T \leq 0 \tag{9.40}$$

and hence:

$$\sum_{i=1}^{m_+} P_{1i} - P_T \leq 0 \quad \text{if} \quad m_{PAC} = m_+ \tag{9.41}$$

Applying (9.41) to $v_{m_+} = 0$, we obtain $\sum_{i=1}^{m_+} P_{1i} - P_T \leq v_{m_+}$ which shows that (4.15) holds for $m_+ = m_{PAC}$.

9.4 Proof of Proposition 6

If $P_T \geq \sum_{i=1}^{m} P_{1i}$, the TPC is redundant under the joint power constraints, and hence, $\mu = 0$. Since $\mu_{WF} \geq 0$, then (4.28) follows.

If $P_T < \sum_{i=1}^{m} P_{1i}$, the TPC is active under the joint power constraints. Also, it is always active for the standard WF (TPC only). Hence:

$$\begin{aligned}\sum_{i=1}^{m}(\mu_{WF}^{-1} - g_i^{-1})_+ &= P_T \\ &= \sum_{i=1}^{m} \min\left(P_{1i}, (\mu^{-1} - g_i^{-1})_+\right) \\ &\leq \sum_{i=1}^{m}(\mu^{-1} - g_i^{-1})_+ \end{aligned} \tag{9.42}$$

where the inequality follows from:

$$\min\left\{P_{1i}, (\mu^{-1} - g_i^{-1})_+\right\} \leq (\mu^{-1} - g_i^{-1})_+ \tag{9.43}$$

so that:

$$\sum_{i=1}^{m}(\mu_{WF}^{-1} - g_i^{-1})_+ \leq \sum_{i=1}^{m}(\mu^{-1} - g_i^{-1})_+ \tag{9.44}$$

Then, $\mu \leq \mu_{WF}$ follows, since the right-hand side is decreasing in μ. The inequality in (9.42) is strict if at least one PAC is active (i.e., $P_{1i} < (\mu^{-1} - g_i^{-1})_+$ for at least one i). Then, $\mu < \mu_{WF}$ follows.

9.5 Proof of Lemma 4.1

Here, we show $\mu_1^{-1} \leq \mu_2^{-1}$, and it can be extended to all μ_i^{-1}. The first and second steps of the algorithm imply:

$$\sum_{i=1}^{m}(\mu_1^{-1} - g_i^{-1})_+ = P_T \tag{9.45}$$

$$\sum_{\substack{i=1 \\ i\notin\mathcal{I}_1}}^{m}(\mu_2^{-1} - g_i^{-1})_+ = P_T - \sum_{i\in\mathcal{I}_1} P_{1i} \tag{9.46}$$

using (9.45) and (9.46):

$$\sum_{i\in\mathcal{I}_1} P_{1i} = \sum_{i=1}^{m}(\mu_1^{-1} - g_i^{-1})_+ - \sum_{\substack{i=1\\ i\notin\mathcal{I}_1}}^{m}(\mu_2^{-1} - g_i^{-1})_+ \tag{9.47}$$

and since $P_{1i} < (\mu_1^{-1} - g_i^{-1})_+$ for $i \in \mathcal{I}_1$, one obtains:

$$\sum_{i\in\mathcal{I}_1} P_{1i} < \sum_{i\in\mathcal{I}_1}(\mu_1^{-1} - g_i^{-1})_+ \tag{9.48}$$

and applying this to (9.47):

$$\sum_{\substack{i=1\\ i\notin\mathcal{I}_1}}^{m}(\mu_1^{-1} - g_i^{-1})_+ < \sum_{\substack{i=1\\ i\notin\mathcal{I}_1}}^{m}(\mu_2^{-1} - g_i^{-1})_+ \tag{9.49}$$

and $\mu_1^{-1} \leq \mu_2^{-1}$ follows because the right-hand side is increasing in μ_2^{-1}.

9.6 Proof of Lemma 4.2

In the last step of the algorithm (i.e., the k-th step), we observe:

$$\begin{aligned} P_T - \Delta_k &= \sum \delta p_{i,k}^* \\ &= \sum_{i\in\left\{\{1,\ldots,m\}-\bigcup_{j=1}^{k-1}\mathcal{I}_j\right\}} p_{i,k}^* \end{aligned} \tag{9.50}$$

Applying this to:

$$\begin{aligned} \Delta_k &= \sum_{i\in\bigcup_{j=1}^{k-1}\mathcal{I}_j} P_{1i} \\ &= \sum_{i\in\bigcup_{j=1}^{k-1}\mathcal{I}_j} p_{i,k}^* \end{aligned} \tag{9.51}$$

we observe:

$$P_T - \sum_{i \in \bigcup_{j=1}^{k-1} \mathcal{I}_j} p^*_{i,k} = \sum_{i \in \left\{ \{1,\ldots,m\} - \bigcup_{j=1}^{k-1} \mathcal{I}_j \right\}} p^*_{i,k} \tag{9.52}$$

Hence, (4.38) follows.

9.7 Proof of Proposition 8

Proposition 4.1 implies:

$$(\mu_1^{-1} - g_i^{-1})_+ \leq \ldots \leq (\mu_k^{-1} - g_i^{-1})_+ \quad \text{for} \quad i \in \{1, 2, \ldots, m\} \tag{9.53}$$

So, for $t \leq k-1$:

$$\begin{aligned} i \in \mathcal{I}_t \Rightarrow p^*_{i,k} = p^*_{i,t} &= P_{1i} \\ &< \delta p^*_{i,t} \\ &= (\mu_t^{-1} - g_i^{-1})_+ \\ &\leq (\mu_k^{-1} - g_i^{-1})_+ \end{aligned} \tag{9.54}$$

where the first inequality follows from the definition of $\mathcal{I}_t$ in Stage 5 of the algorithm. The second inequality follows directly from (9.53).

Hence, (9.54) implies the following for $i \in \bigcup_{j=1}^{k-1} \mathcal{I}_j$:

$$\begin{aligned} p^*_{i,k} &= P_{1i} \\ &< (\mu_k^{-1} - g_i^{-1})_+ \end{aligned} \tag{9.55}$$

In the last step of the algorithm, the following holds for $i \in \left\{\{1,..,m\} - \bigcup_{j=1}^{k-1} \mathcal{I}_j\right\}$ [2]

$$\begin{aligned} p_{i,k}^* &= \delta p_{i,k}^* \\ &= (\mu_k^{-1} - g_i^{-1})_+ \\ &\leq P_{1i} \end{aligned} \tag{9.56}$$

Using (9.55) and (9.56), $p_{i,k}^*$ is obtained for any i, as follows:

$$p_{i,k}^* = \min\left\{P_{1i}, (\mu_k^{-1} - g_i^{-1})_+\right\} \tag{9.57}$$

Full power P_T is used (see Lemma 4.2). Hence, (4.38) and (9.57) imply $p_{i,k}^* = p_i^*$. This means that the algorithm finds the OPA under the joint power constraints (TPC+PAC).

9.8 Proof of Proposition 11

To obtain a proof, we will need the following Lemma, which shows that $C(\mathbf{W}, \mathbf{R})$ is uniformly continuous in $\mathbf{W}$ when $\mathbf{R}$ is bounded.

Lemma 9.3. *If* $tr\boldsymbol{R} \leq P_T$ *and* $\sigma_1(\Delta \boldsymbol{W})P_T < 1$*, then the following holds for any Hermitian* $\Delta \boldsymbol{W}$*:*

$$m \ln\left(1 - \sigma_1(\Delta \boldsymbol{W})P_T\right) \leq C(\boldsymbol{W} + \Delta \boldsymbol{W}, \boldsymbol{R}) - C(\boldsymbol{W}, \boldsymbol{R}) \leq m \ln\left(1 + \sigma_1(\Delta \boldsymbol{W})P_T\right) \tag{9.58}$$

Proof. Let ΔC be as follows:

$$\Delta C = C(\mathbf{W} + \Delta\mathbf{W}, \mathbf{R}) - C(\mathbf{W}, \mathbf{R}) \tag{9.59}$$

[2]Note that $i \in \left\{\{1,..,m\} - \bigcup_{j=1}^{k-1} \mathcal{I}_j\right\}$ denotes the elements that belong to $\{1,..,m\}$ but not to $\bigcup_{j=1}^{k-1} \mathcal{I}_j$.

so that:

$$\begin{aligned}
\Delta C &= \ln|\mathbf{I} + (\mathbf{W} + \Delta\mathbf{W})\mathbf{R}| - \ln|\mathbf{I} + \mathbf{W}\mathbf{R}| \\
&\overset{(a)}{=} \ln|\mathbf{I} + \mathbf{R}^{1/2}(\mathbf{W} + \Delta\mathbf{W})\mathbf{R}^{1/2}| - \ln|\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2}| \\
&\overset{(b)}{\leq} \ln|\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2} + \sigma_1(\Delta\mathbf{W})\mathbf{R}| - \ln|\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2}| \\
&= \ln|\mathbf{I} + \sigma_1(\Delta\mathbf{W})(\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2})^{-1}\mathbf{R}| \\
&\overset{(c)}{\leq} \ln|\mathbf{I} + \sigma_1(\Delta\mathbf{W})\mathbf{R}| \\
&\overset{(d)}{\leq} \ln|\mathbf{I} + \sigma_1(\Delta\mathbf{W})P_T\mathbf{I}| \\
&= m\ln\big(1 + \sigma_1(\Delta\mathbf{W})P_T\big)
\end{aligned} \tag{9.60}$$

where (a) follows from $|\mathbf{I} + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|$; (b) is obtained from the fact that $\Delta\mathbf{W} \leq \sigma_1(\Delta\mathbf{W})\mathbf{I}$; (c) is due to $\mathbf{I} \geq (\mathbf{I} + \mathbf{R}^{\frac{1}{2}}\mathbf{W}\mathbf{R}^{\frac{1}{2}})^{-1}$; and (d) follows from $\mathbf{R} \leq \sigma_1(\mathbf{R})\mathbf{I}$ and $\sigma_1(\mathbf{R}) \leq P_T$.

To attain the lower bound, we use the same approach:

$$\begin{aligned}
\Delta C &\overset{(e)}{\geq} \ln|\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2} - \sigma_1(\Delta\mathbf{W})\mathbf{R}| - \ln|\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2}| \\
&= \ln|\mathbf{I} - \sigma_1(\Delta\mathbf{W})(\mathbf{I} + \mathbf{R}^{1/2}\mathbf{W}\mathbf{R}^{1/2})^{-1}\mathbf{R}| \\
&\geq \ln|\mathbf{I} - \sigma_1(\Delta\mathbf{W})P_T\mathbf{I}| \\
&= m\ln\big(1 - \sigma_1(\Delta\mathbf{W})P_T\big)
\end{aligned} \tag{9.61}$$

where (e) follows from $\Delta\mathbf{W} \geq -\sigma_1(\Delta\mathbf{W})\mathbf{I}$. Here, $\sigma_1(\Delta\mathbf{W})P_T < 1$ implies that all determinant values are non-negative, and hence all values are real. □

To prove (5.3), let ΔC_1 and ΔC_2 be as follows:

$$\Delta C_1 = C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W})\big) - C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W})\big) \tag{9.62}$$

$$\Delta C_2 = C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W})\big) - C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big) \tag{9.63}$$

so that $\Delta C(\mathbf{W}, \mathbf{W}_0) = \Delta C_1 + \Delta C_2$. From Lemma 9.3, ΔC_1 is upper bounded as follows:

$$\begin{aligned}\Delta C_1 &= C\big(\mathbf{W}_0 + \Delta\mathbf{W}, \mathbf{R}^*(\mathbf{W})\big) - C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W})\big) \\ &\le m \ln\big(1 + \sigma_1(\Delta\mathbf{W})P_T\big) \end{aligned} \tag{9.64}$$

where $\Delta\mathbf{W} = \mathbf{W} - \mathbf{W}_0$. Likewise, ΔC_2 is upper bounded as follows:

$$\begin{aligned}\Delta C_2 &= C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W})\big) - C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big) \\ &\le C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W}_0)\big) - C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big) \\ &= -\Big(C\big(\mathbf{W}_0 + \Delta\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big) - C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W}_0)\big)\Big) \\ &\le -m \ln(1 - \sigma_1(\Delta\mathbf{W})P_T) \end{aligned} \tag{9.65}$$

where the first inequality follows from the fact that $\mathbf{R}^*(\mathbf{W}_0)$ is the optimal covariance under $\mathbf{W}_0$, and hence $C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W}_0)\big) \ge C\big(\mathbf{W}_0, \mathbf{R}^*(\mathbf{W})\big)$; the second inequality follows from Lemma 9.3. Hence, the upper bound in (5.3) follows from (9.64) and (9.65):

$$\Delta C(\mathbf{W}, \mathbf{W}_0) \le m \ln\Big(\frac{1 + \sigma_1(\Delta\mathbf{W})P_T}{1 - \sigma_1(\Delta\mathbf{W})P_T}\Big) \tag{9.66}$$

and the lower bound in (5.3) follows from $C\big(\mathbf{W}\big) \ge C\big(\mathbf{W}, \mathbf{R}^*(\mathbf{W}_0)\big)$.

9.9 Proof of Lemma 5.1

First, we establish the following property of $C(\mathbf{W}, \mathbf{R})$ (to our best knowledge, this property has not been given in the literature).

Lemma 9.4. *$C(\boldsymbol{W}, \boldsymbol{R})$ is a jointly uniformly-continuous function for any bounded $\boldsymbol{W}$ and $\boldsymbol{R}$:*

$$\sigma_1(\boldsymbol{R}) \le P_T < \infty, \quad \sigma_1(\boldsymbol{W}) \le A < \infty \tag{9.67}$$

Proof. First, we prove the uniform continuity of $C(\mathbf{W}, \mathbf{R})$ in $\mathbf{W}$ (under fixed $\mathbf{R}$), and then we extend it to both $\mathbf{W}$ and $\mathbf{R}$.

The following shows the definition of uniform continuity of $C(\mathbf{W}, \mathbf{R})$ in $\mathbf{W}$, i.e., for any $\epsilon > 0$, there exists $\delta(\epsilon)$ such that:

$$|\Delta C_W| < \epsilon \quad \text{if} \quad \sigma_1(\Delta \mathbf{W}) < \delta(\epsilon) \tag{9.68}$$

where $\Delta C_W = C(\mathbf{W} + \Delta \mathbf{W}, \mathbf{R}) - C(\mathbf{W}, \mathbf{R})$. Note that $\delta(\epsilon)$ is independent of $\mathbf{W}$ and $\mathbf{R}$. Now, using Lemma 9.3:

$$m \ln\left(1 - \sigma_1(\Delta \mathbf{W}) P_T\right) \leq \Delta C_W \leq m \ln\left(1 + \sigma_1(\Delta \mathbf{W}) P_T\right) \tag{9.69}$$

where we assume $\sigma_1(\Delta \mathbf{W}) P_T < 1$. This assumption can always be satisfied by considering a small $\Delta \mathbf{W}$, in which $\sigma_1(\Delta \mathbf{W}) < P_T^{-1}$. Hence, $|\Delta C_W|$ is upper bounded as follows:

$$\begin{aligned} |\Delta C_W| &\leq m \max\left\{ \ln\left(1 + \sigma_1(\Delta \mathbf{W}) P_T\right), \left| \ln\left(1 - \sigma_1(\Delta \mathbf{W}) P_T\right) \right| \right\} \\ &= m \ln \max\left\{ 1 + \sigma_1(\Delta \mathbf{W}) P_T, \frac{1}{1 - \sigma_1(\Delta \mathbf{W}) P_T} \right\} \\ &\overset{\text{(a)}}{=} -m \ln\left(1 - \sigma_1(\Delta \mathbf{W}) P_T\right) \end{aligned} \tag{9.70}$$

where (a) follows from the fact that $1 + x \leq (1 - x)^{-1}$ if $0 \leq x \leq 1$. Also, $-m \ln\left(1 - \sigma_1(\Delta \mathbf{W}) P_T\right)$ is increasing in $\sigma_1(\Delta \mathbf{W})$. So, after some manipulation, (9.70) implies that:

$$|\Delta C_W| < \epsilon \quad \text{if} \quad \sigma_1(\Delta \mathbf{W}) < \delta(\epsilon) = \frac{1 - e^{-\epsilon/m}}{P_T} \tag{9.71}$$

Hence, $C(\mathbf{W}, \mathbf{R})$ is uniformly-continuous in $\mathbf{W}$. Also, the bound for $\sigma_1(\Delta \mathbf{W})$ in (9.71) is consistent with the assumption of $\sigma_1(\Delta \mathbf{W}) P_T < 1$.

To establish the joint uniform continuity, we show its definition, i.e., for any $\epsilon > 0$, there exists $\delta(\epsilon)$ such that:

$$|\Delta C_{W,R}| < \epsilon \quad \text{if} \quad \sigma_1(\Delta\mathbf{R}), \sigma_1(\Delta\mathbf{W}) < \delta(\epsilon) \tag{9.72}$$

where:

$$\Delta C_{W,R} = C(\mathbf{W} + \Delta\mathbf{W}, \mathbf{R} + \Delta\mathbf{R}) - C(\mathbf{W}, \mathbf{R}) \tag{9.73}$$

Note that $\delta(\epsilon)$ and ϵ do not depend on $\mathbf{W}$ and $\mathbf{R}$. Let:

$$\Delta C_R = C(\mathbf{W} + \Delta\mathbf{W}, \mathbf{R} + \Delta\mathbf{R}) - C(\mathbf{W} + \Delta\mathbf{W}, \mathbf{R}) \tag{9.74}$$

Then, we observe that:

$$\begin{aligned} |\Delta C_{W,R}| &= |\Delta C_R + \Delta C_W| \\ &\leq |\Delta C_R| + |\Delta C_W| \\ &\leq -m \ln\big(1 - \sigma_1(\Delta\mathbf{R})A\big) - m \ln\big(1 - \sigma_1(\Delta\mathbf{W})P_T\big) \end{aligned} \tag{9.75}$$

where the first inequality is due to the triangle inequality. If $\sigma_1(\Delta\mathbf{W})P_T < 1$, $\sigma_1(\Delta\mathbf{R})A < 1$, $\sigma_1(\mathbf{R}) \leq P_T < \infty$, and $\sigma_1(\mathbf{W} + \Delta\mathbf{W}) \leq A < \infty$, then $|\Delta C_W|$ is upper bounded as in (9.70) and $|\Delta C_R|$ can be upper bounded with a similar approach. Hence, the second inequality follows. So, we observe that if:

$$-m \ln\big(1 - \sigma_1(\Delta\mathbf{R})A\big) < \epsilon/2 \tag{9.76}$$

$$-m \ln\big(1 - \sigma_1(\Delta\mathbf{W})P_T\big) < \epsilon/2 \tag{9.77}$$

then $|\Delta C_{W,R}| < \epsilon$, which shows that $C(\mathbf{W}, \mathbf{R})$ is jointly uniformly continuous:

$$|\Delta C_{\mathbf{W},\mathbf{R}}| < \epsilon \quad \text{if} \quad \sigma_1(\Delta\mathbf{R}), \sigma_1(\Delta\mathbf{W}) < \delta(\epsilon) = \frac{1 - e^{-\epsilon/2m}}{\max(P_T, A)} \tag{9.78}$$

where (9.78) is consistent with the assumptions of $\sigma_1(\Delta\mathbf{W})P_T < 1$ and $\sigma_1(\Delta\mathbf{R})A < 1$. □

Note that the first bound in (9.67) always holds because of the TPC, i.e., $\sigma_1(\mathbf{R}) \leq tr\mathbf{R} \leq P_T < \infty$, and the second bound follows from the fact that the channel is always bounded in practice. Now, to obtain a proof, we observe that:

$$\begin{aligned}\lim_{\mathbf{W}\to\mathbf{W}_0} C(\mathbf{W}) &= \lim_{\mathbf{W}\to\mathbf{W}_0} \max_{\mathbf{R}} C(\mathbf{W},\mathbf{R}) \\ &= \max_{\mathbf{R}} \lim_{\mathbf{W}\to\mathbf{W}_0} C(\mathbf{W},\mathbf{R}) \\ &= \max_{\mathbf{R}} C(\mathbf{W}_0,\mathbf{R}) = C(\mathbf{W}_0) \end{aligned} \tag{9.79}$$

where max and lim can be swapped because of the joint uniform continuity of $C(\mathbf{W},\mathbf{R})$. Note that $\lim\max \neq \max\lim$ in general, and hence we need the joint uniform continuity to show (5.5).

9.10 Proof of Corollary 5

The maximum singular value of any matrix $\mathbf{A}$ can be obtained from the following optimization problem [39]:

$$\sigma_1(\mathbf{A}) = \max_{||\mathbf{x}||=||\mathbf{y}||=1} |\mathbf{x}^+\mathbf{A}\mathbf{y}| \tag{9.80}$$

where $||\mathbf{x}||$ is the Euclidean norm of $\mathbf{x}$. Observe that:

$$\begin{aligned}\sigma_1(\mathbf{W}-\mathbf{D}_W) &= \max_{||\mathbf{x}||=||\mathbf{y}||=1} |\mathbf{x}^+(\mathbf{W}-\mathbf{D}_W)\mathbf{y}| \\ &\geq |\mathbf{e}_i^+(\mathbf{W}-\mathbf{D}_W)\mathbf{e}_j| \\ &= |(\mathbf{W}-\mathbf{D}_W)_{i,j}| \\ &= |\mathbf{h}_i^+\mathbf{h}_j|, \quad \text{for } i \neq j \end{aligned} \tag{9.81}$$

where the first equality follows from (9.80) and $\mathbf{e}_i$ represents the standard basis in which all the components are zero except the i-th component, which is one. Hence, $\sigma_1(\mathbf{W}-\mathbf{D}_W) < \delta_\epsilon$ implies $|\mathbf{h}_i^+\mathbf{h}_j| < \delta_\epsilon$. The converse is not true in general.

9.11 Gradient and Hessian

Here, we use some techniques for matrix differential calculus [70], [77], [78]. Let us consider the following function:

$$f(\mathbf{X}) = \ln\big(P_T - tr\mathbf{X}\big) \tag{9.82}$$

where $\mathbf{X}_{m\times m} > 0$. Observe that:

$$\begin{aligned} f(\mathbf{X}+d\mathbf{X}) &= \ln\big(P_T - tr(\mathbf{X}+d\mathbf{X})\big) \\ &= \ln\left((P_T - tr\mathbf{X})\left(1 - \frac{tr(d\mathbf{X})}{P_T - tr\mathbf{X}}\right)\right) \\ &= f(\mathbf{X}) + \ln\left(1 - \frac{tr(d\mathbf{X})}{P_T - tr\mathbf{X}}\right) \\ &= f(\mathbf{X}) - \frac{tr(d\mathbf{X})}{P_T - tr\mathbf{X}} - \frac{tr(d\mathbf{X})tr(d\mathbf{X})}{2(P_T - tr\mathbf{X})^2} + O\big(tr(d\mathbf{X})^2\big) \end{aligned} \tag{9.83}$$

where the last equality is due to $\ln(1-x) = -\sum_{n=1}^{\infty}(n^{-1})x^n$. Here, we use operator $vec(\mathbf{X})$, which accumulates all entries of $\mathbf{X}$ into a single column vector and duplication matrix $\mathbf{D}$ that is characterized by $vec(\mathbf{X}) = \mathbf{D}vech(\mathbf{X})$ [78]. We observe that:

$$\begin{aligned} tr(d\mathbf{X}) &= tr(d\mathbf{X}\mathbf{I}) \\ &\overset{(a)}{=} vec(d\mathbf{X})^{+}vec(\mathbf{I}) \\ &= vech(d\mathbf{X})^{+}\mathbf{D}^{+}vec(\mathbf{I}) \\ &\overset{(b)}{=} d\mathbf{x}^{+}\mathbf{D}^{+}vec(\mathbf{I}) \end{aligned} \tag{9.84}$$

where (a) follows from the fact that $d\mathbf{X}$ is Hermitian and $tr(\mathbf{AB}) = vec(\mathbf{A}^{+})^{+}vec(\mathbf{B})$; (b) is due to $d\mathbf{x} = vech(d\mathbf{X})$. Also, (9.84) implies:

$$tr(d\mathbf{X}) = vec(\mathbf{I})^{+}\mathbf{D}d\mathbf{x} \tag{9.85}$$

applying (9.84) and (9.85) to (9.83), one obtains:

$$f(\mathbf{X}+d\mathbf{X}) = f(\mathbf{X}) - d\mathbf{x}^{+}\frac{\mathbf{D}^{+}vec(\mathbf{I})}{P_T - tr\mathbf{X}} - \frac{1}{2}d\mathbf{x}^{+}\frac{\mathbf{D}^{+}vec(\mathbf{I})vec(\mathbf{I})^{+}\mathbf{D}}{(P_T - tr\mathbf{X})^2}d\mathbf{x} + O\big(tr(d\mathbf{X})^2\big) \tag{9.86}$$

Hence, the gradient and Hessian expressions are as follows:

$$\nabla_x f = -\frac{\mathbf{D}^{+}vec(\mathbf{I})}{P_T - tr\mathbf{X}} \tag{9.87}$$

$$\nabla_{xx} f = -\frac{\mathbf{D}^{+}vec(\mathbf{I})vec(\mathbf{I})^{+}\mathbf{D}}{(P_T - tr\mathbf{X})^2} \tag{9.88}$$

9.12 Proof of Proposition 13

Let $\mathbf{u}_1$ be the eigenvector corresponding to λ_1, so $\lambda_1(\mathbf{W}) = \mathbf{u}_1^{+}\mathbf{W}\mathbf{u}_1$. Also, there exists at least one i which $d_1(\mathbf{W}) = (\mathbf{W})_{ii}$, and this implies $d_1(\mathbf{W}) = \mathbf{e}_i^{+}\mathbf{W}\mathbf{e}_i$. If $d_1(\mathbf{W}) = \lambda_1(\mathbf{W})$, then:

$$\mathbf{e}_i^{+}\mathbf{W}\mathbf{e}_i = \mathbf{u}_1^{+}\mathbf{W}\mathbf{u}_1 \tag{9.89}$$

$$\mathbf{e}_i^{+}\mathbf{W}\mathbf{e}_i = \max_{||\mathbf{x}||=1} \mathbf{x}^{+}\mathbf{W}\mathbf{x} \tag{9.90}$$

$$\mathbf{e}_i \in \arg\max_{||\mathbf{x}||=1} \mathbf{x}^{+}\mathbf{W}\mathbf{x} \tag{9.91}$$

$$\mathbf{e}_i \in \{\mathbf{x} : \mathbf{W}\mathbf{x} = \lambda_1(\mathbf{W})\mathbf{x}\} \tag{9.92}$$

Here, $\{\mathbf{x} : \mathbf{W}\mathbf{x} = \lambda_1(\mathbf{W})\mathbf{x}\}$ is a set of vectors which maximize the quadratic expression $\mathbf{x}^{+}\mathbf{W}\mathbf{x}$. These vectors belong to an eigenspace corresponding to $\lambda_1(\mathbf{W})$. Also, it is easy to see that (9.92) implies (9.89), i.e., $\mathbf{e}_i \in \{\mathbf{x} : \mathbf{W}\mathbf{x} = \lambda_1(\mathbf{W})\mathbf{x}\}$ implies $d_1(\mathbf{W}) = \lambda_1(\mathbf{W})$.

There is no need for i to be unique in the above proof. Hence, when there are repeated maximum diagonal entries, $\lambda_1(\mathbf{W}) = d_1(\mathbf{W})$ if and only if at least one solution of $i = \arg\max_j (\mathbf{W})_{jj}$ satisfies the condition in (9.92).

9.13 MATLAB Code

9.13.1 Iterative Water-Filling Algorithm

```
% Iterative water-filling (TPC+PAC)
% capacity under the joint TPC+PAC constraints for a massive MIMO channel under
% favorable propagation
% inputs:
% W : channel gains
% PT : total transmit power constraint
% P1 : per-antenna power constraints


function iterative_water_filling(W,PT,P1)
if PT >= sum(P1)
    final_power=P1;
else
    ff=1;
    ttt=2;
    py=1;
    while(ttt>1)
        k=1;
        m=length(W);
        bisectionerror=1;
        % To find mu, we can use an analytical solution or the bisection
    algorithm.
        % Here, the bisection algorithm is used.
        xl=0;
        xu=max(W);
        f=0;
        clear r
        while(bisectionerror>1e-8)
            midpoint=(1/2)*(xl+xu);
            for k=1:m
                a=max(0,(midpoint)^(-1)-(W(k)^(-1)));
                r(k)=a;
            end
            f_value=sum(r)-PT;
            if f_value<0
                xu=midpoint;
            elseif f_value>0
                xl=midpoint;
            end
            bisectionerror=abs(f_value);
        end

        % set of streams that exceed the PACs
```

```
43          b=0;
44          pac=0;
45          sdsd=1;
46          clear delete_W
47          for i=1:m
48              if r(i)>P1(i)
49                  r(i)=P1(i);
50                  pac=1+pac;
51                  final_W(py)=W(i);
52                  final_power(py)=r(i);
53                  delete_W(sdsd)=i;
54                  sdsd=1+sdsd;
55                  py=py+1;
56              end
57          end
58
59          if pac>0
60              pac_completed=sum(P1(delete_W));
61              W(delete_W)=[];
62              P1(delete_W)=[];
63          end
64
65          % pac is the number of elements in the set of streams that exceed the
        PACs
66          if pac==0
67              ttt=0;
68              u_EWF=midpoint;
69              for i=1:m
70                  final_W(py)=W(i);
71                  final_power(py)=r(i);
72                  py=py+1;
73              end
74          end
75
76          if (ff==1 && pac==0)
77              ttt=0;
78          else
79              PT=PT-pac_completed;
80          end
81          ff=ff+1;
82      end
83      end
84  final_W;
85  final_power;
86  save('PAC_TPC_EWF.mat')
87  end
```

9.13.2 Bisection Algorithm

```
% bisection algorithm
% capacity under the joint TPC+PAC constraints for a massive MIMO channel under
% favorable propagation
% inputs:
% W : channel gains
% alpha : coefficient for grade of service
% PT : total transmit power constraint
% P1 : per-antenna power constraints


function BA_TPC_PAC(W,alpha,P1,PT)
m=length(W);
W_alpha=W.*alpha;

% desired uncertainty interval
bisectionerror=1;

% upper bound and lower bound for mu
xl=0;
xu=max(W_alpha);

while(bisectionerror>1e-8)

        % midpoint
        midpoint=(1/2)*(xl+xu);

        % redefine upper bound or lower bound
        for k=1:m
        a=max(0,(midpoint)^(-1)-(W_alpha(k)^(-1)));
        r(k)=alpha(k)*min(((P1(k))/(alpha(k))),a);
        end

        f_value=sum(r)-PT;
        if f_value<0
        xu=midpoint;
        elseif f_value>0
        xl=midpoint;
        end

        bisectionerror=abs(f_value);
        %if f_value=0, then bisectionerror=0, hence bisectionerror>1e-8 is false.

end
save('PAC_TPC_BA.mat')
end
```

9.13.3 Newton-Barrier Algorithm

```
% Newton_barrier algorithm
% capacity under the joint TPC+PAC constraints for a MIMO channel
% inputs:
% W : Channel Gram matrix
% PT : total transmit power constraint
% P1 : per-antenna power constraints


function Newton_barrier_TPC_PAC(W,PT,P1)

% t00 : the initial value for t
t00=100;
t=t00;
t_barrier(1)=t;

% the value of t increases by the factor u
u=5;

diag_P1=diag(P1);
[m,~]=size(W);

% initial Tx covariance matrix
for i=1:m
    initial_entry_R(i)=(1/2)*min(P1(i),PT/m);
end
R_Newton{1}=diag(initial_entry_R);
x_Newton(:,1)=vech(R_Newton{1});

% the value of t_max
tmax=1e7;

zzz=1;
ii=1;

%the barrier method
while(t<tmax)

    R_before_newton{ii}=R_Newton{1};

    % gradient
    Z{1}=(((eye(m,m)+W*R_Newton{1}))^(-1))*W;
    diag_R{1}=diag(diag(R_Newton{1}));
    barrier_P1{1}=(diag_R{1}-diag_P1)^(-1);
    gradient_R_Ft{1}=Z{1}+(1/t)*((R_Newton{1})^(-1))+(1/t)*barrier_P1{1};
    gradient_x_Ft{1}=(dup_n(m))'*(veC(gradient_R_Ft{1})+(1/t)*...
    veC(eye(m,m))*(1/(trace(R_Newton{1})-PT)));
```

```
47
48      x_Newton(:,1)=vech(R_Newton{1});
49      r_w{1}=(gradient_x_Ft{1})';
50
51      k=1;
52      Norm_r_w_k=1;
53      zzz=5*zzz*u;
54
55      % Newton algorithm
56      condition_newton=1;
57
58      % accuracy of the Newton algorithm
59      epsilon=10^(-8);
60      while (condition_newton>epsilon)
61
62          R_Newton{k};
63          R_P1=R_Newton{k};
64
65          % Hessian
66          for i=1:m
67              err=zeros(m,m);
68              brr=zeros(m,m);
69              err(i,i)=1/((R_P1(i,i)-P1(i))^2);
70              brr(i,i)=1;
71              er{i}=kron(brr,err);
72          end
73          [mk,nk]=size(er{i});
74          b_P1=zeros(mk,nk);
75          for i=1:m
76              b_P1= b_P1+er{i};
77          end
78          hessian_x_Ft{k}=-(dup_n(m))'*(kron(Z{k},Z{k})+(1/t)...
79          *kron((R_Newton{k})^(-1),(R_Newton{k})^(-1))+(1/t)*b_P1+...
80          +(1/t)*(1/((PT-trace(R_Newton{k}))^(2)))*veC(eye(m,m))...
81          *(veC(eye(m,m)))')*(dup_n(m));
82
83          KKT_mat{k}=hessian_x_Ft{k};
84          delta_w{k}=((-r_w{k})/KKT_mat{k});
85          delta_x{k}=(delta_w{k})';
86
87
88          % backtracking line search
89
90          % residual norm is decreased by (a)
91          a=0.3;
92
93          % reduction in s is controlled by (B)
94          B=0.5;
```

```
95
96          s=1;
97          condition_backtrack=10;
98          Norm_r_new=100;
99          aqq=0;
100         while ((Norm_r_new>condition_backtrack)||(aqq>1))
101
102             x_Newton(:,k+1)=x_Newton(:,k)+s*delta_x{k};
103             R_Newton{k+1}=invvech(x_Newton(:,k+1),m);
104             R_Newton{k+1};
105             Z{k+1}=((eye(m,m)+W*R_Newton{k+1})^(-1))*W;
106             diag_R{k+1}=diag(diag(R_Newton{k+1}));
107             barrier_P1{k+1}=(diag_R{k+1}-diag_P1)^(-1);
108             gradient_R_Ft{k+1}=Z{k+1}+(1/t)*((R_Newton{k+1})^(-1))+...
109             (1/t)*(barrier_P1{k+1});
110             gradient_x_Ft{k+1}=(dup_n(m))'*(veC(gradient_R_Ft{k+1})+...
111             (1/t)*veC(eye(m,m))*(1/(trace(R_Newton{k+1})-PT)));
112             r_w{k+1}=(gradient_x_Ft{k+1})';
113             Norm_r_new=norm(r_w{k+1});
114             condition_backtrack=(1-a*s)*norm(r_w{k});
115             s=B*s;
116             yuyu=eig(R_Newton{k+1});
117             aqq=0;
118
119             % checking the feasibility of covariance matrix
120             if ((R_Newton{k+1})')==R_Newton{k+1}
121                 for jljl=1:length(yuyu)
122                     if (yuyu(jljl))>0
123                         aqq=0+aqq;
124                     else
125                          aqq=2+aqq;
126                     end
127                 end
128             else
129                 aqq=2.5;
130             end
131
132             new_R=R_Newton{k+1};
133
134             for yryr=1:m
135                 if new_R(yryr,yryr)<=P1(yryr)
136                     aqq=0+aqq;
137                 else
138                     aqq=2+aqq;
139                     R_Newton{k+1};
140                 end
141             end
142
```

```
143             if trace(new_R)<=PT
144                 aqq=0+aqq;
145             else
146                 aqq=2+aqq;
147                 R_Newton{k+1};
148             end
149
150
151         end
152
153         condition_newton=norm(r_w{k+1});
154         residual_norm(k+1)=norm(r_w{k+1});
155         norm(r_w{k+1});
156         tyty(k)=norm(r_w{k+1});
157         c_capacity(k)=log(det(eye(m,m)+W*R_Newton{k+1}));
158         log(det(eye(m,m)+W*R_Newton{k+1}));
159         k=k+1;
160     end
161
162 cc_capacity{ii}=c_capacity;
163 clear c_capacity
164 ttt{ii}=tyty;
165 clear tyty
166 R_Newton{1}=R_Newton{k};
167 x_Newton(:,1)=vech(R_Newton{1});
168 t=u*t;
169 C(ii)=log(det(eye(m,m)+W*R_Newton{1}));
170 t_barrier(ii+1)=t;
171 ii=ii+1;
172 end
173 save('PAC_TPC_NB.mat')
174 end
```

```
1  % duplication matrix
2  % This function is needed for the Newton_barrier algorithm.
3  % vec(A)=dup_n vech(A)
4  % input = n ( A is n*n)
5
6  function duplicationmatrix=dup_n(n)
7      duplicationmatrix=zeros(n*n,n*(n+1)/2);
8          for i=1:n
9          for j=1:n
10             if i>=j
11                 a=zeros(n*n,1);
12                 a((j-1)*n+i)=1;
13                 a((i-1)*n+j)=1;
14                 duplicationmatrix(:,((j-1)*(n-(j/2))+i))=a;
15             end
```

```
        end
        end

end
```

```
% vec function
% This function is needed for the Newton_barrier algorithm.
% input: matrix A

function vec_output=veC(A)
        [m,n]=size(A);
        if m==n
        for i=1:n
            for j=1:m
                vec_outputt(j+(i-1)*n)=A(j,i);
            end
        end
        vec_output=conj(vec_outputt');

        end
end
```

```
% vech function
% This function is needed for the Newton_barrier algorithm.
% input : hermitian matrix A

function vech_output=vech(A)
    [mq,nq]=size(A);
        if mq==nq
        vech_outputt=[];
        for i=1:nq
            clear a
            for j=i:mq
                a(j-i+1)=A(j,i);
            end
            vech_outputt=[vech_outputt,a];
        end
        vech_output=conj(vech_outputt');
    else
        error('Input must be a square matrix.')
        end

end
```

```
% inverse of vech function
% This function is needed for the Newton_barrier algorithm.
% note : vech(A)=xq and A is m*m
% inputs : xq and m
```

```
5
6  function  invvech_output=invvech(xq,m)
7      invvech_output=zeros(m,m);
8      for i=1:m
9          invvech_output((i-1)+1:end,i)=xq((i-1)*m-((i-2)*(i-1)/2)+1:i*m-((i-1)*i
       /2));
10     end
11         for i=1:m
12         for j=1:m
13             if i<j
14                 invvech_output(i,j)=conj(invvech_output(j,i));
15             end
16         end
17         end
18 end
```

9.13.4 Monte-Carlo Algorithm

```
1  % Monte-Carlo algorithm
2  % capacity under the joint TPC+PAC constraints for a MIMO channel
3  % inputs:
4  % W : Channel Gram matrix
5  % PT : total transmit power constraint
6  % P1 : per-antenna power constraints
7
8
9  function MC_TPC_PAC(W,PT,P1)
10
11 [Nt,~]=size(W);
12
13 % initial Tx covariance matrix
14 for i=1:Nt
15     initial_entry_R(i)=min(P1(i),PT/Nt);
16 end
17 R{1}=diag(initial_entry_R);
18 R_star=R{1};
19 capacity(1)=log(det(eye(Nt,Nt)+W*R{1}));
20
21 % number of trials
22 number_of_trials=1e3;
23
24 for j=2:number_of_trials
25         clear H
26         var=1;
27
```

```
    % H is a random matrix
        H=sqrt(var)*(randn(Nt,Nt));
        bb=(H'*H);

    % random feasible transmit covariance
        for i=1:Nt
        rii(i)=bb(i,i);
        a(i)=(P1(i)/(rii(i)));
    end
        a=min(a);
        aa=PT/trace(H'*H);
    a=min(a,aa);
    R{j}=a.*H'*H;

    % The best transmit covariance
        C(j)=log(det(eye(Nt,Nt)+W*R{j}));
        if C(j)>capacity(j-1)
        R_star=R{j};
        end

    capacity(j)=log(det(eye(Nt,Nt)+W*R_star));
end
save('PAC_TPC_MC.mat')
end
```

9.13.5 CVX

```
% CVX: 360 types of scripts
% capacity under the joint TPC+PAC constraints for a MIMO channel
% inputs:
% W : Channel Gram matrix
% PT : total transmit power constraint
% P1 : per-antenna power constraints


function CVX_TPC_PAC_all_types(W,PT,P1)
stepp=1;
 for pre=1:5
     for ttype=1:6
         for begintype=1:2
             for solverr=1:3
                 for mmax=1:2

%-------------------------------------------------
% CVX_precision
```

```
19 if pre==1
20 cvx_precision('best')
21 end
22
23 if pre==2
24 cvx_precision('default')
25 end
26
27 if pre==3
28 cvx_precision([1e-7 1e-7 1e-5])
29 end
30
31 if pre==4
32 cvx_precision([1e-4 1e-4 1e-3])
33 end
34
35 if pre==5
36 cvx_precision('high')
37 end
38
39 %------------------------------------------------
40 % begin_type
41 if begintype==1
42 cvx_begin sdp
43 end
44
45 if begintype==2
46 cvx_begin
47 end
48
49 %------------------------------------------------
50 % CVX_solver
51 if solverr==1
52  cvx_solver  SDPT3
53 end
54 if solverr==2
55  cvx_solver  SeDuMi
56 end
57 if solverr==3
58  cvx_solver  Mosek
59 end
60 cvx_quiet (true);
61
62 %------------------------------------------------
63 % different types of variables
64 [m,n]=size(W);
65 if ttype==1
66         variable R(m,n) symmetric
```

```
67          R == semidefinite(m,n);
68
69      if  mmax==1
70      maximize log_det(eye(m,n)+W*R)
71      end
72      if  mmax==2
73      maximize  det_rootn(eye(m,n)+W*R)
74      end
75
76          subject to
77      trace(R) <= PT ;
78      R(1,1) <= P1(1) ;
79      R(2,2) <= P1(2) ;
80      % R(i,i) <= P1(i) for any i=1:m
81      % Here, we consider m=2
82      cvx_end
83  end
84
85
86  if ttype==2
87          variable R(m,n) semidefinite
88
89      if  mmax==1
90      maximize log_det(eye(m,n)+W*R)
91      end
92      if  mmax==2
93      maximize  det_rootn(eye(m,n)+W*R)
94      end
95
96          subject to
97      trace(R) <= PT ;
98      R(1,1) <= P1(1) ;
99      R(2,2) <= P1(2) ;
100     cvx_end
101 end
102
103
104 if ttype==3
105     variable R(m,n) complex semidefinite
106
107     if  mmax==1
108     maximize log_det(eye(m,n)+W*R)
109     end
110     if  mmax==2
111     maximize  det_rootn(eye(m,n)+W*R)
112     end
113
114         subject to
```

```
115          trace(R) <= PT ;
116          R(1,1) <= P1(1) ;
117          R(2,2) <= P1(2) ;
118      cvx_end
119  end
120
121
122  if ttype==4
123      variable R(m,n) hermitian semidefinite
124
125      if  mmax==1
126      maximize log_det(eye(m,n)+W*R)
127      end
128      if  mmax==2
129      maximize  det_rootn(eye(m,n)+W*R)
130      end
131
132          subject to
133          trace(R) <= PT ;
134          R(1,1) <= P1(1) ;
135      R(2,2) <= P1(2) ;
136      cvx_end
137  end
138
139
140  if ttype==5
141     variable R(m,n) symmetric
142
143      if  mmax==1
144      maximize log_det(eye(m,n)+W*R)
145      end
146      if  mmax==2
147      maximize  det_rootn(eye(m,n)+W*R)
148      end
149
150          subject to
151          R == semidefinite(m,n);
152          trace(R) <= PT ;
153          R(1,1) <= P1(1) ;
154      R(2,2) <= P1(2) ;
155          cvx_end
156  end
157
158
159  if ttype==6
160      variable R(m,n) hermitian;
161
162      if  mmax==1
```

```
163     maximize log_det(eye(m,n)+W*R)
164     end
165     if  mmax==2
166     maximize  det_rootn(eye(m,n)+W*R)
167     end
168
169     subject to
170         R == semidefinite(m,n);
171         trace(R) <= PT ;
172         R(1,1) <= P1(1) ;
173     R(2,2) <= P1(2) ;
174     cvx_end
175 end
176
177
178 R_cvx=R;
179 C_cvx=log(det(eye(m,m)+W*R_cvx));
180 C_cvx_cvx(stepp)=C_cvx;
181 R_cvx_cvx{stepp}=R_cvx;
182 stepp=stepp+1;
183 %--------------------------------------------------
184                     end
185                 end
186             end
187         end
188  end
189  save('PAC_TPC_CVX_360.mat')
190  end
```

References

[1] X. Wu, N. C. Beaulieu, and D. Liu, "On favorable propagation in massive MIMO systems and different antenna configurations," *IEEE Access*, vol. 5, pp. 5578–5593, May 2017.

[2] E. Biglieri *et al.*, *MIMO Wireless Communications*. Cambridge University Press, New York, 2007.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[4] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, March 2015.

[5] Ericsson Mobility Report, June 2019. http://www.ericsson.com/en/mobility-report.

[6] M. Shafi *et al.*, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice," *IEEE JSAC,, Part I*, vol. 35, no. 6, pp. 1201–1221, June 2017.

[7] H. Q. Ngo *et al.*, "Aspects of favorable propagation in Massive MIMO," *22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal*, September 2014.

[8] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Comm. Mag.*, vol. 54, no. 2, pp. 114–123, February 2016.

[9] T. L. Marzetta *et al.*, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[10] B. Sklar, *Digital Communications; Fundamentals and Applications (2nd edition)*. Prentice Hall, 2001.

[11] C. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, second edition ed. John Wiley & Sons, 2006.

[13] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communications*, third edition ed. Kluwer, 2003.

[14] C. Cox, *An Introduction to LTE : LTE, LTE-advanced, SAE, VoLTE and 4G Mobile Communications*, second edition ed. John Wiley & Sons, 2014.

[15] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Labs, Internal Tech. Memo*, June 1995, (European Trans. Telecom., vol.10, no.6, December 1999).

[16] S. Loyka, "The capacity of Gaussian MIMO channels under total and per-antenna power constraints," *IEEE Trans. Comm.*, vol. 65, no. 3, pp. 1035–1043, March 2017.

[17] M. Khoshnevisan and J. N. Laneman, "Power allocation in multi-antenna wireless systems subject to simultaneous power constraints," *IEEE Trans. Comm.*, vol. 60, no. 12, pp. 3855–3864, December 2012.

[18] Z. Wang and L. Vandendorpe, "Power allocation for energy efficient multiple antenna systems with joint total and per-antenna power constraints," *IEEE Trans. Comm.*, vol. 66, no. 10, pp. 4521–4535, October 2019.

[19] M. Vu, "MISO capacity with per-antenna power constraint," *IEEE Trans. Comm.*, vol. 59, no. 5, pp. 1268–1274, May 2011.

[20] A. Goldsmith, *Wireless Communications.* Cambridge University Press, New York, 2005.

[21] P. L. Cao and T. J. Oechtering, "Optimal trade-off between transmission rate and secrecy rate in Gaussian MISO wiretap channels," *21th International ITG Workshop on Smart Antennas, Berlin, Germany*, June 2017.

[22] F. Oggier and B. Hassibi, "the secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, August 2011.

[23] M. Vu, "MIMO capacity with per-antenna power constraint," *IEEE Globecom, Houston, USA*, December 2011.

[24] D. Tuninetti, "On the capacity of the AWGN MIMO channel under per-antenna power constraints," *IEEE International Conference on Communications (ICC), Sydney, NSW, Australia*, June 2014.

[25] D. Maamari, N. Devroye, and D. Tuninetti, "The capacity of the ergodic MISO channel with per-antenna power constraint and an application to the fading cognitive interference channel," *IEEE Int. Symp. Information Theory, Hawaii, USA*, July 2014.

[26] S. Zhang, R. Zhang, and T. J. Lim, "Massive MIMO with per-antenna power constraint," *GlobalSIP, Atlanta, GA, USA*, pp. 642–646, December 2014.

[27] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraint," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, June 2007.

[28] J. Park, W. Sung, and T. M. Duman, "Precoder and capacity expressions for optimal two-user MIMO transmission with per-antenna power constraints," *IEEE Comm. Lett.*, vol. 14, no. 11, pp. 996–998, November 2010.

[29] T. Lan and W. Yu, "Input optimization for multi-antenna broadcast channels with per-antenna power constraints," *IEEE Globecom, Dallas, TX, USA*, pp. 420–424, 29 November- 3 December 2004.

[30] S. Shi, M. Schubert, and H. Boche, "Per-antenna power constrained rate optimization for multiuser MIMO systems," *Int. ITG Workshop on Smart Antennas, Vienna, Austria*, pp. 270–277, February 2008.

[31] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," *IEEE International Conference on Communications (ICC), Ottawa, ON, Canada*, pp. 3779–3784, June 2012.

[32] Y. Zhu and M. Vu, "Iterative mode-dropping for the sum capacity of MIMO-MAC with per-antenna power constraint," *IEEE Trans. Comm.*, vol. 60, no. 9, pp. 2421–2426, September 2012.

[33] F. Boccardi and H. Huang, "Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints," *IEEE Int. Workshop on Signal Process. Advances in Wireless Comm., Cannes, France*, pp. 1–5, July 2006.

[34] H. Shen *et al.*, "Transmitter optimization for per-antenna power constrained multi-antenna downlinks: An SLNR maximization methodology," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2712–2725, May 2016.

[35] T. M. Kim, F. Sun, and A. J. Paulraj, "Low-complexity MMSE precoding for coordinated multipoint with per-antenna power constraint," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 395–398, April 2013.

[36] P. L. Cao *et al.*, "Optimal transmit strategy for MISO channels with joint sum and per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4296–4306, August 2016.

[37] R. Chaluvadi, S. S. Nair, and S. Bhashyam, "Optimal multi-antenna transmission with multiple power constraints," *IEEE Trans. Wireless Comm.*, vol. 18, no. 7, pp. 3382–3394, April 2019.

[38] P. L. Cao and T. J. Oechtering, "Optimal transmit strategy for MIMO channels with joint sum and per-antenna power constraints," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA*, pp. 3569–3573, March 2017.

[39] F. Zhang, *Matrix theory: basic results and techniques.* Springer, 2011.

[40] S. Loyka, "On the capacity of Gaussian MIMO channels under the joint power constraints," *IEEE Wireless Comm. Lett.*, vol. 8, no. 2, pp. 332–335, April 2019.

[41] Y. Dong *et al.*, "Energy efficiency maximization for MISO systems with total and per-antenna power constraints," *11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP), Budapest, Hungary*, July 2018.

[42] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Comm.*, vol. 9, no. 11, pp. 3590–3600, November 2010.

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

[44] X. Gao *et al.*, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Comm.*, vol. 14, no. 7, pp. 3899–3911, July 2015.

[45] L. Lu *et al.*, "An overview of massive MIMO: benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, October 2014.

[46] H. Q. Ngo, D. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Comm.*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[47] E. Björnson *et al.*, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Info. Theory*, vol. 60, no. 11, pp. 7112–7139, November 2014.

[48] J. Chen, "When does asymptotic orthogonality exist for very large arrays?" *GLOBECOM, Atlanta, GA, USA*, pp. 4146–4150, December 2013.

[49] C.-N. Chuah *et al.*, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, March 2002.

[50] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, October 2002.

[51] C. Masouros and M. Matthaiou, "Space-constrained massive MIMO: Hitting the wall of favorable propagation," *IEEE Comm. Lett.*, vol. 19, no. 5, pp. 771–774, May 2015.

[52] Y. Sun *et al.*, "On asymptotic favorable propagation condition for massive MIMO with co-located user terminals," *International Symposium on Wireless Personal Multimedia Communications (WPMC), Sydney, NSW, Australia*, pp. 706–711, September 2014.

[53] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Comm.*, vol. 66, no. 11, pp. 5205–5219, November 2018.

[54] Z. Gao *et al.*, "Asymptotic orthogonality analysis of time-domain sparse massive MIMO channels," *IEEE Comm. Lett.*, vol. 19, no. 10, pp. 1826–1829, October 2015.

[55] M. Matthaiou *et al.*, "Does massive MIMO fail in Ricean channels," *IEEE Wireless Comm. Lett.*, vol. 8, no. 1, pp. 61–64, February 2019.

[56] S. E. Hajri, J. Denis, and M. Assaad, "Enhancing favorable propagation in Cell-Free massive MIMO through spatial user grouping," *19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece*, June 2018.

[57] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

[58] J. Hoydis *et al.*, "Channel measurements for large antenna arrays," *International Symposium on Wireless Communication Systems (ISWCS),Paris, France*, pp. 811–815, August 2012.

[59] S. Payami and F. Tufvesson, "Channel measurements and analysis for very large array systems at 2.6 ghz," *6th Eur. Conf. Antennas Propag. (EUCAP), Prague, Czech Republic*, pp. 433–437, March 2012.

[60] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, January 2013.

[61] S. L. H. Nguyen *et al.*, "On the mutual orthogonality of millimeter-wave massive MIMO channels," *81st Vehicular Technology Conference (VTC Spring),Glasgow, UK*, pp. 1–5, May 2015.

[62] M. Gauger *et al.*, "Channel measurements with different antenna array geometries for massive MIMO systems," *International ITG Conference on Systems, Communications and Coding, Hamburg, Germany,*, pp. 3855–3864, February 2015.

[63] J. Li and Y. Zhao, "Measurement-based asymptotic user orthogonality analysis and modelling for massive MIMO," *IEEE Comm. Lett.*, vol. 21, no. 12, pp. 2762–2765, December 2017.

[64] X. Gao *et al.*, "Linear pre-coding performance in measured very-large MIMO channels," *IEEE Vehicular Technology Conference (VTC Fall),San Francisco, CA, USA*, pp. 1–5, September 2011.

[65] S. W. Peters and R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, January 2011.

[66] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[67] P. He *et al.*, "Water-filling: a geometric approach and its application to solve generalized radio resource allocation problems," *IEEE Trans. Wireless Comm.*, vol. 12, no. 7, pp. 3637–3647, July 2013.

[68] R. A. Horn and C. R. Johnson, *Matrix Analysis*, second edition ed. Cambridge University Press, 2012.

[69] S. Loyka, "On optimal signaling over Gaussian MIMO channels under interference constraints," *IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada*, pp. 220–223, November 2017.

[70] S. Loyka and C. D. Charalambous, "An algorithm for global maximization of secrecy rates in Gaussian MIMO wiretap channels," *IEEE Trans. on Comm.*, vol. 63, no. 6, pp. 2288–2299, June 2015.

[71] L. Dong, S. Loyka, and Y. Li, "An algorithm for optimal secure signaling over cognitive radio MIMO channels," *IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada*, pp. 126–130, November 2017.

[72] G. S. Fishman, *Monte Carlo Concepts, Algorithms, and Applications.* Springer, 1996.

[73] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control.* John Wiley & Sons, 2003.

[74] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," version 2.1, march 2014. http://cvxr.com/cvx.

[75] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[76] J. P. Kermoal *et al.*, "A stochastic mimo radio channel model with experimental validation," *IEEE J. Sel. Areas Comm.*, vol. 20, no. 6, pp. 1211–1226, August 2002.

[77] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus WithApplications to Statistics and Econometrics*. Hoboken, NJ, USA:Wiley, 1999.

[78] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York, NY, USA: Springer-Verlag, 1997.