# NONCONVEX AND GAME THEORY OPTIMIZATION FOR RESOURCE ALLOCATION IN WIRELESS CCOMMUNICATIONS 

by<br>Kandasamy Illanko<br>B.Sc. (Honours) (Elect. and Electronics Eng.) University of Peradeniya, Sri Lanka, 1985 M.A.Sc. (Comp. Eng.) University of Toronto, 1991

A Dissertation<br>presented to Ryerson University<br>in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Program of Electrical and Computer Engineering

Toronto, Ontario, Canada, 2014
(c) Kandasamy Illanko 2014

## AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Abstract<br>Kandasamy Illanko, 2014<br>Doctor of Philosophy<br>Electrical and Computer Engineering<br>Ryerson University

Designing wireless communication systems that efficiently utilize the resources, frequency spectrum and electric power, leads to problems in mathematical optimization. Most of these optimization problems are difficult to solve because the objective functions are nonconvex. While some problems remain unsolved, the solutions proposed in the literature for the others are of somewhat limited use because the algorithms are either unstable or have too high a computational complexity. This dissertation presents several stable algorithms, most of which have polynomial complexity, that solve five different nonconvex optimization problems in wireless communication. Two centralized and two distributed algorithms deal with the power allocation that maximizes the throughput in the Gaussian interference channel (GIC) with various constraints. The most valuable of these algorithms, the one with the minimum rate constraints became possible after a significant theoretical development in the dissertation that proves that the throughput of the GIC has a new generalized convex structure called invexity. The fifth algorithm has linear complexity, and finds the power allocation that maximizes the energy efficiency (EE) of OFDMA transmissions, for a given subchannel assignment. Some fundamental results regarding the power allocation are then used in the genetic algorithm for determining the subchannel allocation that maximizes the EE. Pricing for channel subleasing for ad-hoc wireless networks is considered next. This involves the simultaneous optimization of many functions that are interconnected through the variables involved. A composite game, a strategic game within a Stackelberg game, is used to solve this optimization problem with polynomial complexity. For each optimization problem solved, numerical results obtained using simulations that support the analysis and demonstrate the performance of the algorithms are presented.

## Acknowledgment

I am forever grateful to my supervisors, Profs. Alagan Anpalagan and Dimitrios Androutsos, for giving me an opportunity to do my PhD , after a lengthy departure from academic work. I am indebted to Prof. Anpalagan for his tireless enthusiasm, constant support, and critical insights over the last few years, which made this dissertation possible. I must also thank him for his style of supervising, and the supportive and friendly atmosphere of his lab. My heartfelt gratitude goes to Prof. Dimitrios Androutsos for his support and guidance, in particular, beyond the work of this dissertation.

I must acknowledge my lab mates for their support during the early years of my PhD . In particular, I fondly recall the stimulating discussions I had with Dr. Senthuran about physics and engineering. My sincere appreciation also goes to Dr. Naeem for introducing me to the topic of energy efficiency.

My work today is the result of several other key people who deserve recognition.
Prof. Safwat Zaky, my Masters thesis supervisor at University of Toronto, was my first anchor into the rough waters of graduate studies when I first arrived in Canada. His kindness, support, enthusiasm, as well as interest in my work allowed me to explore the world of computer engineering. I am grateful for his support, encouragement, and critical advice on my career over the years including my doctoral studies at Ryerson.

It is not an exaggeration to say Prof. Gunawardena from University of Peradenya (Department of Electronics and Electrical Engineering, Sri Lanka) shaped me as an engineer. I am incredibly fortunate to have learned the fundamentals of electricity, electronics, and computer science from him.

When I was in Grade 11, my uncle Arumugam introduced me to Murugaiya master - a math genius. Watching him teach mathematics can only be described as poetry in motion. What I learned from him in six months made me strong enough for a life time of adventure in mathematics. Among my high school teachers, I must thank Mr. Balasubramaniam who taught me science in Grade 9. The stories he told us and the questions he asked us inspire
me even today.
I am indebted to my uncles, Kanapathypillai who introduced me to many topics in science using stories and anecdotes during family gatherings, and Arumugam who taught me the fundamentals of math and physics. More than anyone else, it is Arumugam who shaped my scientific thinking. He is the first person who demonstrated to me that different people understand the same concept to different depths. My uncles had an intuition first approach, and insisted that solutions to problems must be elegant and graceful. Because of them, to this day, I have a healthy dislike for tedious solutions and algorithms. Arumugam also did many experiments and projects with me to fuel my interest in science, and because of him I am equally interested in both theoretical and practical aspects of engineering today.

I would like to thank my high school friends who now live in Canada. I am fortunate to be able to create a home away from home because of my friends, and I am grateful for their ongoing support for me and my family. I am thankful for the supportive advice I received throughout my PhD from my friend, Prof. Ravindiran Vaithespara (Department of History, University of Manitoba).

Most importantly, I must express my deepest gratitude towards my wife, Prof. Sepali Guruge (Department of Nursing, Ryerson University), first of all, for believing in me and making me do my PhD. I am grateful to her for her love, encouragement, and unwavering support without which I would not have been able to complete this work. I am indebted to her for reading and giving me feedback on all my conference papers and journal articles, at times, even when she was in the middle of crucial deadlines in her early (yet, shining) academic career.

## Dedication

This dissertation is dedicated to my two year old daughter, Apsara, who surprises me every day with her curiosity and eagerness to learn new things and master new skills.

## Contents

Abstract ..... iii
Acknowledgment ..... iv
Dedication ..... vi
List of Tables ..... xv
List of Figures ..... xvi
List of Abbreviations ..... xix
1 Introduction ..... 1
1.1 Chronology of cellular communications ..... 2
1.2 Physical layer ..... 3
1.3 Shannon's upper bound ..... 5
1.3.1 Throughput ..... 7
1.3.2 Power allocation ..... 8
1.3.3 Frequency assignment ..... 8
1.3.4 Energy efficiency ..... 9
1.4 System models ..... 9
1.4.1 GIC ..... 10
1.4.2 OFDMA ..... 12
1.5 Thesis statement and expected contributions ..... 13
1.5.1 Optimization problems to be solved ..... 14
1.5.2 Thesis statement ..... 16
1.5.3 Expected original contributions ..... 16
1.6 Computational complexity measure ..... 17
1.7 Background ..... 18
1.7.1 Throughput maximization in GIC ..... 18
1.7.2 Energy efficiency maximizing power allocation (for fixed subchannel assignment) in OFDMA ..... 19
1.7.3 Energy efficiency maximizing frequency and power allocation for OFDMA ..... 19
1.7.4 Pricing for subleasing channels for ad-hoc networks ..... 20
1.8 Organization ..... 20
1.8.1 Part I: Theory and Techniques ..... 20
1.8.2 Part II: Application and Algorithms ..... 22
I Theory and Techniques ..... 25
2 Mathematical optimization ..... 26
2.1 Introduction ..... 26
2.2 An optimization problem ..... 27
2.3 The dual problem ..... 28
2.4 Convex sets and convex functions ..... 29
2.5 Convex structure ..... 30
2.6 Convex optimization problem ..... 31
2.7 The Karush Kuhn Tucker (KKT) Conditions ..... 31
2.8 Generalized convex functions ..... 32
2.8.1 Quasiconvex functions ..... 33
2.8.2 Pseudoconvex functions ..... 33
2.8.3 Invex functions ..... 34
2.9 Game theory ..... 34
2.10 Real time problems ..... 35
3 Cross product in higher dimensions ..... 37
3.1 Introduction ..... 37
3.2 Traditional cross product ..... 38
3.3 Extended cross product ..... 38
4 Invexity ..... 40
4.1 Introduction ..... 40
4.2 Concavity ..... 41
4.3 Incavity ..... 43
4.4 A sufficient condition for incavity ..... 45
4.5 Being incave with respect to the same function ..... 46
5 Convexifying, increasing transformations ..... 48
5.1 Introduction ..... 48
5.2 The exponential transformation ..... 49
5.3 Charnes-Cooper transformation (CCT) ..... 50
6 Genetic algorithm ..... 52
6.1 Introduction ..... 52
6.2 Genetic algorithm with discrete variables ..... 53
6.2.1 Chromosome ..... 53
6.2.2 In the beginning ..... 56
6.2.3 The birth of children ..... 56
6.2.4 Genetic mutation ..... 57
6.2.5 The last generation ..... 57
6.2.6 Numerical results ..... 57
6.3 Genetic algorithm with continuous variables ..... 58
6.3.1 Numerical results ..... 59
7 Game theory ..... 60
7.1 Introduction ..... 60
7.2 Optimization and game theory ..... 61
7.3 Strategic games in continuous space ..... 61
7.3.1 Nash equilibrium ..... 64
7.3.2 Existence of Nash equilibrium ..... 64
7.3.3 Optimality of Nash equilibrium ..... 65
7.4 Stackelberg game ..... 66
7.4.1 Two player Stackelberg game ..... 66
II Applications and Algorithms ..... 68
8 Throughput maximization in the GIC with proportional rate constraints ..... 69
8.1 Introduction ..... 69
8.2 Related work ..... 70
8.3 The proposed approach ..... 72
8.4 System model and the problem statement ..... 76
8.5 Analysis of the optimization problem ..... 79
8.5.1 Objective and constraints as hyper surfaces ..... 79
8.5.2 Methodology of the analysis ..... 80
8.5.3 Normal vectors of the surfaces ..... 81
8.5.4 Tangent vector to curve $\mathbf{C}$ and its last component ..... 81
8.5.5 Curve $\mathbf{C}$ is increasing ..... 84
8.5.6 Curve $\mathbf{C}$ intersects the power constraint plane at a unique point ..... 84
8.5.7 Two linear constraints on the powers ..... 87
8.6 Centralized algorithm for power allocation ..... 88
8.6.1 Algorithm ..... 89
8.6.2 Complexity ..... 92
8.7 Distributed algorithms for power allocation ..... 93
8.7.1 Distributed algorithm-1 ..... 93
8.7.2 Distributed algorithm-2 ..... 99
8.7.3 Complexity ..... 102
8.8 Numerical results ..... 103
8.8.1 Centralized algorithm ..... 104
8.8.2 Distributed algorithms ..... 108
8.9 Conclusion ..... 110
9 Throughput maximization in the GIC with minimum rate constraints ..... 112
9.1 Introduction ..... 112
9.2 Background ..... 113
9.3 System model and the problem statement ..... 115
9.4 Generalized convex structure of the sum rate of the GIC ..... 118
9.5 Algorithm development ..... 124
9.5.1 Extension of traditional algorithms to locate the point of KKT condition124
9.5.2 Proposed algorithm to locate the point of KKT Condition - the algo- rithm for sum rate maximization with total power constraint ..... 125
9.5.3 A important note on the origin of the algorithm ..... 126
9.5.4 Algorithm with rate constraints ..... 127
9.5.5 Modification to work with individual power constraints ..... 130
9.5.6 Convergence of the algorithms ..... 131
9.5.7 Computational complexity of PGARC ..... 132
9.6 Numerical results ..... 133
9.7 Conclusion ..... 141
10 Fundamental aspects of energy efficient power allocation in OFDMA ..... 142
10.1 Introduction ..... 142
10.2 Background ..... 143
10.3 System model and problem statement ..... 144
10.4 The single user, single carrier Case ..... 145
10.5 Single user, multi carrier case (OFDM) ..... 148
10.5.1 Why the inequality $t>0$ can be ignored ..... 149
10.6 Multi user multi carrier case (OFDMA) with proportional rate requirements ..... 154
10.7 Multi user multi carrier case (OFDMA) with minimum rate requirements ..... 161
10.8 Multi user multi carrier case (OFDMA) with total power constraint ..... 164
10.9 EE maximizing power allocation for OFDMA with minimum rate and total power Constraints ..... 165
10.10Complexity ..... 168
10.11Conclusion ..... 169
11 Energy efficient frequency and power allocation for OFDMA ..... 170
11.1 Introduction ..... 170
11.2 Background ..... 171
11.3 System model and the optimization problem ..... 172
11.4 Channel assignment using genetic algorithm ..... 174
11.4.1 Choice of a fitness function ..... 175
11.4.2 The channel assignment solution ..... 177
11.5 Power allocation ..... 177
11.6 Numerical results ..... 178
11.7 Discussion and Conclusion ..... 181
12 Competitive pricing for spectrum subleasing for future wireless ad hoc networks ..... 182
12.1 Introduction ..... 182
12.2 Background ..... 183
12.3 Related work ..... 185
12.4 System model and the problem statement ..... 186
12.5 Analysis ..... 189
12.6 Numerical results ..... 194
12.7 Conclusion ..... 197
13 Conclusion and future work ..... 199
13.1 Contributions ..... 199
13.1.1 Theory ..... 199
13.1.2 Development of algorithms ..... 201
13.2 Future directions for research ..... 203
13.2.1 EE in GIC ..... 203
13.2.2 EE in OFDMA ..... 204
13.2.3 Is the throughput in GIC actually concave? ..... 204
Appendices ..... 205
A Review of power allocation in CDMA and GIC using game theory ..... 206
A. 1 Power control for CDMA ..... 207
A. 2 Power control for early cognitive radio ..... 209
A.2.1 Sigmoid function based utility ..... 210
A.2.2 Game with fairness ..... 211
A. 3 Power control in the GIC ..... 213
A.3.1 Iterative water filling ..... 214
B List of Publications ..... 215
B. 1 Journal Publications ..... 215
B. 2 Conference Publications ..... 216
Bibliography ..... 219

## List of Tables

6.1 The Genetic Algorithm ..... 55
8.1 Notations ..... 78
8.2 Centralized algorithm compared with geometric programming ..... 108
9.1 Notations ..... 117
9.2 Scenario-6: Comparison ..... 140
10.1 Notations ..... 155
11.1 Notations ..... 173
12.1 Notations ..... 188

## List of Figures

1.1 The OSI model. Courtesy [12] ..... 4
1.2 The $N$-user Gaussian Interference Channel ..... 10
1.3 The connection of the theory to application ..... 21
4.1 Two functions having the same $\boldsymbol{\eta}$. ..... 47
$5.1 z=g(x, y)=x^{2} \sqrt{y}$ ..... 50
$5.2 \quad h(X, Y)=\exp (2 X+(1 / 2) Y)$ ..... 50
$5.3 \quad f(x)=\ln (x) / x$ ..... 51
$5.4 \quad g(y, t)=t \ln (y / t)$ ..... 51
6.1 (a) The Chromosome (b) One point cross-over ..... 54
6.2 Genetic algorithm with population size $3 \mathrm{P}=600$ ..... 58
6.3 Genetic algorithm with population size $3 \mathrm{P}=90$ ..... 59
8.1 The locus of $f_{1}: f_{2}=1: 1$ ..... 74
8.2 The curve $\mathbf{C}$ of intersection of $f_{1}: f_{2}=1: 1$ and the sum. ..... 74
8.3 Intersection of $\mathbf{C}$ and $x+y=2$. ..... 75
8.4 The $N$-user Gaussian interference channel. ..... 76
8.5 An increasing space curve. ..... 77
8.6 A decreasing space curve. ..... 77
8.7 Solution with two linear constraints. $P^{*}$ and $P^{* *}$ are labels (not coordinates) for points in the Euclidean space. ..... 88
8.8 Worst-case crossing ..... 98
8.9 Scenario-1: Simulation details (only two of the 6 transceivers are shown). ..... 103
8.10 Transmission rates and power levels for Scenario-1. ..... 104
8.11 Scenario-2: Simulation details (only two of the 6 transceivers are shown). ..... 105
8.12 Transmission rates and power levels for Scenario-2. ..... 106
8.13 Transmission rates and power levels for Scenario-3. ..... 106
8.14 Distributed Algorithm-1: Convergence of the Fairness Quotients. ..... 109
8.15 Distributed Algorithm-1: Convergence of the proportional rates. ..... 109
8.16 Distributed Algorithm-2: Convergence of the proportional rates. ..... 110
9.1 The $N$-user Gaussian Interference Channel ..... 115
9.2 Example: The sum rate in the partially symmetrical case. ..... 118
9.3 Simulation details. ..... 134
9.4 Scenario-1: The sum rate in the symmetrical case. ..... 135
9.5 Scenario-1: The power levels in the symmetrical case. ..... 135
9.6 Scenario-2: The power levels in the partially symmetrical case ..... 136
9.7 Scenario-3: The sum rate for the example in Section 8.4. ..... 136
9.8 Scenario-3: The power levels for the example in Section 8.4. ..... 137
9.9 Scenario-4: The sum rate of eight users ..... 137
9.10 Scenario-4: The power levels of eight users ..... 138
9.11 Scenario-5: The sum rate, and the individual rates at key points. ..... 140
11.1 Genetic algorithm parameters. ..... 178
11.2 Simulation details. ..... 179
11.3 Scenario-1: Convergence of GA with $F F_{1}=-\sum_{n} R_{n} w_{n}$. ..... 179
11.5 Scenario-2: Comparing the EE values. ..... 180
12.1 System Model ..... 186
12.2 Simulation details (Only one of the 4 transceivers is shown here.) ..... 194
12.3 Utility function of Customer 1 ..... 195
12.4 Convergence of the fixed point algorithm ..... 195
12.5 Dependence of equilibrium power on the distance from the measuring point. ..... 196
12.6 Utility of the operator ..... 196
13.1 The optimization problems solved ..... 201
A. 1 A summary of game theory work on power control ..... 212

# List of Abbreviations 

CCT Charnes-Cooper Transformation<br>CDMA Code Division Multiple Access<br>DSL Digital Subscriber Line<br>EE Energy Efficiency<br>G Generation<br>GIC Gaussian Interference Channel<br>OFDM Orthogonal Frequency Division Multiplexing<br>OFDMA Orthogonal Frequency Division Multiple Access<br>KKT Karush Kuhn Tucker<br>SINR Signal to Interference and Noise Ratio

## Chapter 1

## Introduction

No one doubts that we are living amidst the "wireless revolution." There have been many discussions about the impact of this revolution on our society. Egalitarian scientists notice an aspect that is often absence in these discussions. It is the dissemination of "privileged" information, from the world's most advanced learning institutions, to anyone in the world. Top researchers at the frontiers of knowledge are making their course notes and books freely available on the internet. More than $95 \%$ of the knowledge required to finish this desertation was obtained via wireless internet at various locations in the city of Toronto. And the same knowledge, with the exception of the IEEE database, is only a laptop away from someone in a developing country. The engineers and the computer scientists who laid the foundations for the internet believed that knowledge should be free [1-3]. A quarter century later, the marriage of internet with wireless communication - the wireless internet - is making their wish almost a reality.

The core enabling technology of this revolution, the wireless communication has been around for over a century. Radio and Television can be considered as the first two waves of the wireless communication. The wireless revolution had its origin in the third wave, the availability of the cellular mobile phones to the masses. The defining feature of cellular communication, which later became part of most modern wireless communication systems, is
frequency reuse. A set of frequencies are used to serve customers in a particular neighborhood of a city. The same set of frequencies are re-used in another neighborhood of the same city.

Since its introduction in the early 1980's, cellular technology has undergone many changes. A constant throughout has been the unrelenting thirst of the customers for higher data rates. This is the principal motive for this dissertation. In order to deliver higher data rates, the engineers have to squeeze out every ounce of performance from the two main resources frequency spectrum and electric power - of wireless communication. Designing communication systems that efficiently utilize these resources leads to problems in mathematical optimization.

How the demand for higher and higher data rates shaped the evolution of cellular communication is outlined in Section 1.1. The connection of the mobile data rate to the fundamentals of communications is explained in the next two sections - the OSI model and the physical layer in Section 1.2, and Shannon's upper bound on the transmission rate in Section 1.3. Section 1.4 describes the system models used in this dissertation. The optimization problems to be solved in this dissertation are listed in Section 1.5. A short review of the literature related to these optimization problems is given in Section 1.7. More detailed, in depth, literature reviews are presented in the chapters dealing with each optimization problem. The chapter concludes with a detailed outline of the organization of the dissertation in Section 1.8.

### 1.1 Chronology of cellular communications

The first generation (1G) mobile phones were analogue devices and only carried voice communication at an average transmission rate of about $2.4 \mathrm{kbits} / \mathrm{s}$ [4]. The 2G mobile devices were digital, had text capabilities, and led to an explosive growth in the use of mobile phones worldwide in the early 1990's [4]. In North America, 2G cellular phones adopted the IS95 (CDMA) standard [5,6], while the rest of the world adopted the GSM (OFDMA) [7,8]. The
average transmission rate in 2 G was about $64 \mathrm{kbits} / \mathrm{s}$.
Operators were surprised to learn that the cellular customers were more adventurous than they thought. 3G wireless technology was rushed in to give the data rate starving customers $2 \mathrm{Mbits} / \mathrm{s}$ speeds. 3 G also saw the change from circuit switching to packet switching for data transmission; voice transmissions in 3G still used circuit switching. The arrival of 4G technology in 2010 completely eliminated circuit switching [9]. 4G mobile devices adopted the OFDMA transmission technology worldwide [10], and average data rate jumped to 100 Mbits/s. For the first time, the cellular mobile systems adapted the multiple input-multiple output (MIMO) communication technology [11].

The more than 1,000 fold increase in data rates from 2 G to 4 G was accompanied by a million fold increase in mobile customers [4], putting enormous pressure on the engineers to push the limits on wireless transmission rates. The question of how far can we push the transmission rates brings us to the fundamentals of wireless communications.

### 1.2 Physical layer

The open systems interconnection (OSI) model [12] breaks down the communication between two devices into seven layers (Fig. 1.1). This dissertation is concerned with some of the optimization problems that arise in the bottom most layer, the physical layer.

The bit stream from the packets obtained from the layer above are used to modulate an electrical signal at the radio frequency and fed to the transmit antenna. The electromagnetic wave emanating from the transmit antenna travels through air, gets reflected on roads and buildings, if necessary penetrates a few walls and reaches the receiving antenna. At the receiver, the process is reversed and the bit stream is recovered with possible errors due to noise. We will refer to everything between the output of the power amplifier at the transmitter and the input to the receiving preamplifier as the channel. Is there a way to theoretically model the happenings in the physical layer and answer the question: Given the

|  | Data unit | Layer | Function |
| :---: | :---: | :---: | :---: |
| Host layers | Data | 7. Application | Network process to application |
|  |  | 6. Presentation | Data representation, encryption and decryption, convert machine dependent data to machine independent data |
|  |  | 5. Session | Interhost communication, managing sessions between applications |
|  | Segments | 4. Transport | Reliable delivery of packets between points on a network. |
| Media layers | Packet/ <br> Datagram | 3. Network | Addressing, routing and (not necessarily reliable) delivery of datagrams between points on a network. |
|  | Bit/Frame | 2. Data link | A reliable direct point-to-point data connection. |
|  | Bit | 1. Physical | A (not necessarily reliable) direct point-to-point data connection. |

Figure 1.1: The OSI model. Courtesy [12]
parameters of the channel, what is the maximum speed with which communication can take place? What is the maximum possible data transmission rate?

But what is data transmission rate? We already mentioned that in the physical layer, at the receiver, the bits are collected with possible errors. Obviously, there is no point in calculating transmission rate with errors. By the 1920's, there was considerable knowledge about how square waves get corrupted by noise. Given the probability distribution of the noise and the transmitted square wave, by 1928, the researchers knew how to calculate the error rate [13]. Yet no one could answer the question: What is the maximum possible data rate. It turned out that we were not asking the right question. About 20 years later, Claude Shannon asked the right question and the "Mathematical Theory of Communications" was born [14].

Suppose the bits are transmitted at a rate of $100 \mathrm{bits} / \mathrm{s}$ and the bit error rate is $1 \%$. Can we say the error free transmission rate is 99 bits/s? As Shannon explains in his seminal work [14], this is wrong becasue we don't know exactly which bits are in error. The correct answer to this question cannot be arrived at without doing some serious thinking about the nature of communications. What are we trying to communicate? We know the answer to that question; it is information. But what is information? Do we have a scientific definition of information? This is the question that one must ask and answer, before concerning ourselves with the maximum speed of communication.

### 1.3 Shannon's upper bound

Shannon defined information regarding an event as a quantity that is proportional to the (probabilistic) uncertainty involved in that event [14]. The purpose of the communication channel is to reduce this uncertainty about the information at the source, in the mind of the receiver. Shannon showed that how much a channel can reduce this uncertainty depends not only on the nature of the channel but also on the relationship of the channel to the
the source. Given a channel, Shannon defined capacity of that channel as the maximum reduction in the uncertainty achievable over all possible sources. He then went on to prove two theorems that are important to this dissertation. The first one says that communication with arbitrarily small bit error rates is possible as long as the bits are transmitted at a rate that is less than or equal to the capacity of the channel. The second theorem expresses the capacity of a channel in terms of the bandwidth of the signal and the signal to noise ratio at the receiver. Suppose the signal that leaves the transmitter has a bandwidth of $B \mathrm{~Hz}$ and carries a power of $p$ Watts as it leaves the power amplifier. Then the capacity $C$ of the channel in bit/s is:

$$
\begin{equation*}
C=B \log _{2}\left(1+\frac{a p}{\sigma^{2} B}\right), \tag{1.1}
\end{equation*}
$$

where $a$ is called the gain of the channel and is basically a fraction indicating the decay in the amplitude of the signal from the transmitter to the receiver. $\sigma^{2}$ is the noise power density in Watts/Hz. This is the sum power of all the unwanted signals that have found their way to the receiver. It may include transmission signals from neighboring systems in the same frequency band.

Thus, the maximum error free transmission rate at the physical layer is upper bounded by the capacity of the channel $C$ above. Henceforth, we will refer to the error free transmission rate at the physical layer as the transmission rate. In order to achieve this upper bound in the transmission rate, that is, to transmit at capacity, one must add additional bits to the bits carrying the information we are trying to communicate. The elementary way of looking at these additional bits is to consider them as part of error correcting codes. The advanced way of looking at this is to think that these additional bits are necessary to improve the relationship of the source to the channel, or "matching" the source to the channel. As mentioned before, this has the potential to move up the "reduction in uncertainty" towards the capacity of the channel. This procedure is known as channel coding and is performed at the data link layer just above the physical layer. Advances in the last 25 years in channel
coding have enabled communication very close to the capacity of the channel [15,16]. In other words, we will assume that the transmissions are taking place at capacity of the channel. Under this assumption, the transmission rate $R$ and the capacity $C$ are equal and we have:

$$
\begin{equation*}
R=B \log _{2}\left(1+\frac{a p}{\sigma^{2} B}\right) \tag{1.2}
\end{equation*}
$$

### 1.3.1 Throughput

There are many ways in which a wide band frequency channel can be used by a number of users for communication. In the next few subsections, we will use the method used in the 4G mobile systems to explain the objective functions and the optimization variables considered in this dissertation. Typically, a wide band channel is divided into a number of subchannels (or subcarriers) and each subchannel is assigned exclusively to one user in the system. Often, the subchannels that are in use in the same neighborhood are separated by "guard bands" to make sure that the electromagnetic waves transmitted on one subchannel does not spill over into the next subchannel and cause interference. Subchannels like these are referred to as being orthogonal. If $a_{k}$ is the gain experienced by the user to whom the $k$ th subchannel is assigned and $p_{k}$ is the transmission power used on that subchannel, then the sum transmission rate over the $k$ subchannels in bits/s, expressed per Hz of the subchannels, will be:

$$
\begin{equation*}
R=\sum_{k=1}^{K} \log _{2}\left(1+\frac{a_{k} p_{k}}{\sigma_{k}^{2}}\right) \tag{1.3}
\end{equation*}
$$

where $\sigma_{k}^{2}$ is the noise power density on the $k$ th subchannel in Watts/Hz. Historically, the sum transmission rate $R$ given above has been the most important performance measure of a wireless communication system. Note that the unit of $R$ is bits $/ \mathrm{s} / \mathrm{Hz}$. This $R$ is also referred to as spectral efficiency, throughput or simply the sum rate.

### 1.3.2 Power allocation

Equation (1.3) shows that the sum rate or throughput can be increased by increasing the power $p_{k}$ on each subchannel. Until recently, electric power was not considered a limited resource by engineers except in the case of battery operated devices. This was because of two reasons: First, electric power is relatively inexpensive in north America. The second, the transmission power at a cellular base station is small compared to the power expenditure in other use of electricity such as heating and cooling. Even then, there was a reason to limit the transmission power on the subchannels. Too high a transmission power in one system can have a detrimental effect on the transmissions on a neighboring system through interference. As seen in (1.3), transmissions from a neighboring system in the subchannel $k$ can increase $\sigma_{k}^{2}$ and thereby reduce $R$. Because of this, it is customary to limit the total transmission power. The task of distributing the available power among the subchannels in a way that increases the sum rate is referred to as power allocation.

### 1.3.3 Frequency assignment

The channel gain $a_{k}$ depends on the terrain between the transmitter and the receiver. In urban environments, a signal leaving a base station antenna, for example, goes through many reflections and may have to penetrate few walls before reaching the mobile device. What fraction of the original amplitude of the signal survives this adventure mainly depends on the distance from the transmitter to the receiver. It also depends on the frequency of the signal because the loss during reflections and penetrations depends on the frequency. This shows that the transmission rate depends on which subchannel is assigned to a particular user. This brings us to the second variable, the channel gain. It is important to note that channel gain is not a continuous variable; it changes when we assign a different subchannel to a user. This shows that in fact the channel gain is a discrete variable. For example, it may change from 0.001 to 0.0004 .

Another way to increase the throughput is to increase the bandwidth $B$ in (1.2) (or equivalently $K$ in (1.3). The frequency bandwidth available for wireless systems does increase over time. However, this happens once in every 2 decades or so. At any given time, the bandwidth available for a particular wireless system is finite. Hence, the only way to increase throughput is to use the available bandwidth efficiently, by carefully distributing the subchannels among the users. This task is referred to as frequency assignment.

A typical optimization problem in resource allocation in wireless communication is concerned with the frequency assignment and power allocation that maximizes a performance measure such as system throughput.

### 1.3.4 Energy efficiency

It has been reported that, by 2010, the global carbon footprint of the information technology (IT) sector has become equal to that of the aviation industry, and that the fastest growing section within the IT sector is mobile communications [17]. This shows that we can no longer ignore the energy consumption in wireless communications, even if we are not concerned with the financial aspect. Independently of this, battery operated communication devices always have to be concerned with their energy expenditure.

Concerns with the energy expenditure of wireless transmissions have given rise to a new performance measure called energy efficiency. Energy efficiency (EE) is defined as throughput per unit of power, and measured in bits/s/Hz/Watts, or bits/Joule/Hz. A detailed expression for EE in terms of the transmission powers and the channel gains will be given in the next section after we describe the system models.

### 1.4 System models

The various optimization problems undertaken in this dissertation are unified by two aspects. First, they all arise from resource allocation in wireless communications. Second, they cannot
be solved using convex optimization techniques. In these optimization problems, the physical layer of a wireless communication system is represented using various models depending on the transmission technology used.

Orthogonal Frequency Division Multiple Access (OFDMA) is the transmission technology used in the current 4G mobile phones [10]. Gaussian Interference Channel (GIC) is used to model femto cells and ad-hoc networks [18-24]. As far as the work in dissertation is concerned, code division multiple access (CDMA) can be considered as a special case of the GIC. CDMA was the cellular transmission technology in North America from early 1990's to late 2000's [6]. We now give detailed descriptions of the GIC and OFDMA models.

### 1.4.1 GIC



Figure 1.2: The $N$-user Gaussian Interference Channel

Gaussian interference channel (GIC) is used to model CDMA uplinks, digital subscriber lines (DSL), as well as situations where geographical separation of users allow many users
to use the same channel for transmission.
GIC with $N$ users is shown in Fig. 1.2. User $i$ employs transmitter $i$ to communicate with receiver $i$ but receiver $i$ experiences interference from all other transmitters. $h_{i j}$ denotes the channel gain between transmitter $j$ and receiver $i$, and $p_{i}$ denotes the power used by transmitter $i$. If $\sigma_{i}^{2}$ is the spectral density of the additive white Gaussian noise, the transmission rate $r_{i}$ of user- $i$ in bits $/ \mathrm{s} / \mathrm{Hz}$ can be written as:

$$
\begin{equation*}
r_{i}=\log _{2}\left(1+\frac{h_{i i} p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} h_{i j} p_{j}}\right), \tag{1.4}
\end{equation*}
$$

The sum transmission rate or throughput $R$ of the GIC in bits $/ \mathrm{s} / \mathrm{Hz}$ can be written as:

$$
\begin{equation*}
R=\sum_{i=1}^{N} \log _{2}\left(1+\frac{h_{i i} p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} h_{i j} p_{j}}\right) \tag{1.5}
\end{equation*}
$$

Chapters 8 and 9 are concerned with the power levels $p_{i}$ that maximize the throughput $R$ subject to a total power constraint and, either a proportional rate constraint or a minimum rate constraint.

## CDMA

Code division multiple access (CDMA) was the multi-access technology for cellular communication in North America during the 2G and part of the 3G periods. In CDMA, each user is assigned a code. Each user, instead of transmitting a bit, would transmit a series of bits representing the code assigned to it. The codes are not truly orthogonal, they are only pseudo-orthogonal. While this keeps the interference between the users to a minimum, it does not eliminate it completely.

Consider the uplink of a cell that uses CDMA. While the signal from User-i will have an amplitude of $h_{i i} p_{i}$, it will be corrupted not only by the Gaussian noise, but also by the signals from the other users, because the codes are not truly orthogonal. The effect of the
other signals can be sumed up as $\sum_{j \neq i} h_{j j} p_{j}$. The end result is that the throughput of the uplink of a CDMA system can be written as:

$$
\begin{equation*}
R=A \sum_{i=1}^{N} \log _{2}\left(1+\frac{h_{i i} p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} h_{j j} p_{j}}\right) \tag{1.6}
\end{equation*}
$$

where $A$ is a constant [8]. This shows that the throughput model for the CDMA uplink is a special case of the GIC. The reader interested in the implementation details of CDMA is referred to [5].

This dissertation does not concern itself with any problems in the CDMA. However, since CDMA is a special case of the GIC, any solution to a GIC problem in dissertation readily applies to the same problem in the CDMA system.

There are far more research done on the throughput maximization in the CDMA than that on GIC. The related work section of Chapter 8 provides a review of the work on CDMA. We emphasize that while these work does not apply to our problems, our work does apply to their problems.

### 1.4.2 OFDMA

As the name implies, the subchannels in OFDMA are orthogonal. This means transmissions in one subchannel do not interfere with transmissions in another subchannel. Implementation details of OFDMA are irrelevant to this dissertation, but the interested reader is referred to [8] for more information.

Consider the downlink of a single cell with $N$ users and $K$ orthogonal subchannels. Since the channels are orthogonal, transmissions on one subchannel do not interfere with the transmissions on another subchannel. Each subchannel is assigned exclusively to one user. If the channel gain on the $k$ th subchannel is $a_{k}$, the transmission power $p_{k}$, the noise spectral density $\sigma^{2}$, and $h_{k}=a_{k} / \sigma^{2}$, then sum transmission rate or system throughput in bits $/ \mathrm{s} / \mathrm{Hz}$
can be written as:

$$
\begin{equation*}
R=\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right) \tag{1.7}
\end{equation*}
$$

The system energy efficiency (EE) of the transmissions over all $K$ subchannels in bits/Joule/Hz can be written as:

$$
\begin{equation*}
E E=\frac{R}{p_{c}+\psi \sum_{k=1}^{K} p_{k}}=\frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}} \tag{1.8}
\end{equation*}
$$

where $\psi$ is the reciprocal of the efficiency of the transmit power amplifier. $p_{c}$ is called the circuit power [25] of the downlink transmitter, and it represents the total power spent on preprocessing the signal before it enters the power amplifier.

Assume a yet-to-be determined subchannel assignment protocol is used to distribute the subchannels among the users. Suppose a total of $K_{1}$ subchannels - subchannel 1 through subchannel $k_{1}$ - are assigned to User-1. A total of $K_{2}$ subchannels - subchannel $k_{1}+1$ through subchannel $k_{2}$ - are assigned to User- 2 and so on. A total of $K_{n}$ subchannels - subchannel $k_{n-1}+1$ through subchannel $k_{n}$ - are assigned to User- $n$. Then the rate $r_{n}$ of User- $n$ can be written as:

$$
\begin{equation*}
r_{n}=\sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right), \tag{1.9}
\end{equation*}
$$

### 1.5 Thesis statement and expected contributions

We start with a description of the five optimization problems that will be solved in this dissertation. The thesis statement is given next. A summary of the expected original contributions follows.

### 1.5.1 Optimization problems to be solved

## Throughput of GIC with proportional rate constraints

Chapter 8 solves the problem of finding the power allocation that maximizes the throughput of the GIC with proportional rate and linear power constraints. This is achieved without any knowledge of the geometrical structure (convex or otherwise) of the throughput of the GIC. A formal description of the problem is given below. In addition to the variables and parameters defined in the last section, $\beta_{i}$ 's stands for the proportional rates, $g_{i}$ 's are the coefficients in the linear power constraint, and $P$ is the maximum (interference) power allowed.

$$
\underset{p_{1}, p_{2}, \ldots p_{K}}{\operatorname{maximize}} \sum_{i=1}^{N} \log _{2}\left(1+\frac{h_{i i} p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} h_{i j} p_{j}}\right)
$$

subject to

## Throughput of GIC with minimum rate constraints

Chapter 9 establishes that the throughput of the GIC has a generalized convex structure called invexity. This together with the understanding of the nature of the gradient vector of the throughput of the GIC gained in Chapter 8 is used to tackle the harder problem of finding the power allocation that maximizes the throughput of the GIC with minimum rate and linear power constraints. The minimum rates can be translated to minimum signal to interference and noise ratios (SINR) $\gamma_{i}$ 's.

$$
\operatorname{maximize}_{p_{1}, p_{2}, \ldots p_{K}} \sum_{i=1}^{N} \log _{2}\left(1+\frac{h_{i i} p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} h_{i j} p_{j}}\right)
$$

subject to

$$
\begin{align*}
\frac{p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} a_{i j} p_{j}} & \geq \gamma_{i} \text { for } i=1,2, \ldots N .  \tag{1.11}\\
\sum_{i=1}^{N} g_{i} p_{i} & \leq P
\end{align*}
$$

## Energy efficiency of OFDMA with minimum rate constraints

A two step, near optimal, solution to the problem of finding the frequency and power allocation that maximizes the energy efficiency of OFDMA transmissions from a base station, with minimum rate and total power constraints, is solved in Chapter 11. Genetic algorithm [26] is used for subchannel assignment, and is followed by the optimal power allocation obtained in the previous chapter by solving the problem below:

$$
\underset{p_{1}, p_{2}, \ldots p_{K}}{\operatorname{maximize}} \frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}}
$$

subject to

$$
\begin{align*}
& \sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right) \geq R_{n} \text { for } n=1,2, \ldots, N .  \tag{1.12}\\
& \sum_{k=1}^{K} p_{k} \leq P
\end{align*}
$$

The same notations for the variables and parameters from the OFDMA system model subsection 1.4.2 are used here. $R_{n}$ is the minimum rate requirement of User- $n$.

## Pricing for subleasing of channels for ad-hoc networks

An operator that subleases spectrum to agencies that use ad-hoc networks divides a city into cells. Unlike in the traditional cells, the operator allows frequency re-use in adjacent cells. The price structure for the subleasing has a part discouraging too much interference to an agency in the adjacent cell using the same subchannel. The optimization process involved in setting the best price for this part is formulated as a composite game - a strategic game inside a Stackelberg game [27]. Chapter 12 solves this game and determines the optimal price the operator would like to set in order to maximize his revenue. Transmissions from users in a cell is modeled using GIC. For the payoff functions used please see Chapter 12.

### 1.5.2 Thesis statement

The goal of this dissertation is to develop polynomial time algorithms that solve the five optimization problems listed in the last subsection.

### 1.5.3 Expected original contributions

A concise description of the ten original contributions of this dissertation are given below.

## Throughput maximization in GIC

- C-1: A new way of solving throughput maximization with proportional rate constraints.
- C-2: A centralized polynomial time algorithm.
- C-3: A distributed linear time algorithm.
- C-4: A distributed constant time algorithm (shared memory).
- C-5: The proof that the throughput has the generalized convex structure called invexity.
- C-6: The projection algorithm for throughput maximization with minimum rate constraints.


## EE maximization in OFDMA

- C-7: A linear time algorithm that obtains the power allocation by solving a single non-linear equation.
- C-8: Derivation of approximate fitness functions for evolutionary approaches to channel assignment for EE maximization.
- C-9: A two-step frequency and power allocation algorithm that achieves the highest EE in the literature.


## Subchannel pricing for ad-hoc networks

- C-10: Derivation of pricing in terms of the solution to a linear system of equations using a composite game framework.


### 1.6 Computational complexity measure

The most popular complexity measure used to evaluate the relative merits of optimization algorithms concerns itself with how the execution time of the algorithm scales with the size of the problem it solves. For example, the complexity of the fastest algorithm for sorting an array is $O\left(n \log _{2}(n)\right)$, where $n$ is the number of elements in the array. The complexity of a modern linear system solver is $O\left(N^{2} L\right)$, where $N$ is the number of equations and $L$ is the number of variables [28].

The size of a typical wireless communication resource allocation problem can be described by its two dimensions: the number of users $N$ in the system and the number of subchannels $K$ in the system. For example, a channel allocation protocol from the literature which we use for comparison purposes later in the dissertation has a complexity of $O\left(N K \log _{2}(K)\right)$.

### 1.7 Background

This dissertation deals with five optimization problems in resource allocation in wireless communications. This section presents what is state of the art in these problems, in brief. A more detailed account of the related work is given in the chapters dealing with each of these problems.

### 1.7.1 Throughput maximization in GIC

The power allocation that maximizes the throughput of the GIC with a total power constraint is an open problem in wireless communication. The existing work [29-35] on throughput maximization in GIC uses individual power constraints. While individual power constraints are important in many applications, there are also situations where a total (interference) power constraint is more appropriate. In these situations, dividing a total power constraint into equal parts into individual power constraints is inefficient because the maximum throughput under a total power constraint may occur outside these individual power constraints. Furthermore, the algorithms proposed by most work in the literature have too high a complexity to be of significant value to the industry.

Another group of work [36-61] maximizes individual transmission rates of a number of users in a system simultaneously. When a user increases its power it will increase his transmission rate but will decrease the rates of all the other users. Strategic game theory [27,62-66] is used to find a competitive (Nash) equilibrium power level for all users. This equilibrium point has the property that a unilateral deviation from this power level by a single user cannot increase his transmission rate. These work use individual power constraints, and do not maximize the system (sum) transmission rate.

This dissertation provides polynomial time algorithms for the throughput maximization problem in the GIC with total power constraint. The algorithm in this dissertation can be easily modified to accommodate individual power constraints.

### 1.7.2 Energy efficiency maximizing power allocation (for fixed subchannel assignment) in OFDMA

The objective of the optimization problem, the energy efficiency (EE), is not concave in the powers. The Charnes-Cooper Transformation has been used in many work to change it into a concave function $[25,67-71]$. However, the solution to the problem, the power allocation that maximizes the EE is obtained using an iterative algorithm.

A limitation with this type of solution is that it does not help us in any way with the other part of the optimization: the frequency allocation. In a two step solution to the frequency and power allocation that maximizes the EE, the subchannels are assigned first and then the powers are allocated in an optimal manner. In order to assign the subchannels in a near optimal way, we need insights into the question: what type of subchannel allocation leads to the highest maximum EE after the optimal power allocation. This is not possible with a power allocation that is obtained via an iterative algorithm. A closed form solution to the problem of optimal power allocation would be ideal, but that may not be possible.

We obtain the power levels that produce the maximum EE through a single non-linear equation. We also obtain some fundamental closed form results concerning the EE maximizing power allocation. These results enable us to approximate the maximum EE in closed form and use it in frequency assignment.

### 1.7.3 Energy efficiency maximizing frequency and power allocation for OFDMA

In the absence of knowledge about the type of subchannel allocation that leads to the highest maximum EE after the optimal power allocation, the existing two step solutions use a heuristic approach to frequency allocation [25, 72].

Armed with the results from the EE maximizing power allocation, we use genetic algorithm for frequency assignment.

### 1.7.4 Pricing for subleasing channels for ad-hoc networks

Pricing of spectrum in the context of a cognitive radio scenario has been studied using strategic games in [73-77]. A number of researchers use a Stackelberg game [27, 78] to analyze the pricing of spectrum by primary networks that leverage the secondary access points as cooperative relays $[79,80]$.

To model the scenario considered in this dissertation, a single game would not suffice. Hence, in this dissertation, we use a composite game: a strategic game inside a Stackelberg game.

### 1.8 Organization

This dissertation is divided into two main parts, titled, Theory and Techniques, and Applications and Algorithms. The relationship between these two parts is illustrated in Fig. 1.3.

### 1.8.1 Part I: Theory and Techniques

Part I starts by introducing the basic definitions and terminology of mathematical optimization, in Chapter 2, and goes on to explain relatively new ideas such as cross products in higher dimensions and invexity in later chapters. This part also introduces the genetic algorithm and game theory.

Please note that because of the diversity of the optimization problems considered in this dissertation, a single table summarizing all the notations is not feasible. Instead, a table summarizing the notations used in each of the application chapters is provided in each of those chapters.

Chapter 2 begins with categorizing the optimization problems into what we are successful in solving and what we are hopeless at. It then introduces the basic definitions and termi-


Figure 1.3: The connection of the theory to application
nology essential to understanding any exposition on optimization, by presenting a concise account of convex optimization. The chapter ends with a glimpse into generalized convexity.

Chapter 3 explains why the idea of the traditional cross product breaks down in higher dimensions and introduces an extended version of the cross product.

One of the most exciting ideas in this dissertation, the one that led to a breakthrough in an optimization problem that is known to be difficult to solve, invexity, is introduced in Chapter 4. While most of this chapter contains a re-organization of invex analysis found in [88-90], it also has an original theorem that is necessary for our application of invex analysis in the later Chapter 9.

Chapter 5 covers convexifying transformations that can be used to transform nonconvex functions into convex functions. A transformation introduced in this chapter is used later in Chapter 10 to convert the energy efficiency function in OFDMA into a concave function.

Genetic algorithm (GA) is explained in detail in Chapter 6. GA is used for frequency assignment in Chapter 11.

Finally, Chapter 7 presents the fundamentals of strategic game theory in continuous spaces and the Strackelberg game theory. A composite game, a strategic game inside a Strackelberg game is used to come up with the optimal spectrum pricing in Chapter 12.

### 1.8.2 Part II: Application and Algorithms

Part II uses the theory and techniques from Part I to analyze and solve 5 different nonconvex optimization problems in resource allocation for wireless communication systems. Each problem is the subject of a different chapter. Each chapter also develops algorithms that solve the respective problem and presents numerical results supporting the analysis and demonstrating the performance of the algorithm/s.

Chapter 8 solves the problem of finding the power allocation that maximizes the throughput of the GIC with proportional rate and total power constraints. Analytic geometry and the cross product in higher dimensions introduced in Chapter 3 are used here. One central-
ized algorithm (polynomial complexity) and two distributed algorithms that find the solution are developed. Apart from the usefulness of these results in themselves, there is another motivation for the work in this chapter. It is the analysis of the gradient of the throughput of the GIC in this chapter that enabled the proof of the central theorem of the next chapter.

Chapter 9 has two purposes. Establishing the generalized convex structure of the throughput of the GIC, and finding the power allocation that maximizes the throughput of the GIC with minimum rate and total power constraints. The throughput is shown to be incave (Chapter 4) in the transmission powers. A centralized algorithm of polynomial complexity is developed. This chapter and the chapter before solve optimization problems with the same objective but with different constraints. Despite this similarity, the material is presented in two different chapters because of two reasons. First, the theory used in this chapter is different from the one used in the last chapter. Second, as mentioned before, this chapter has the important result that the sum rate of the GIC is incave in the powers, which may have other applications outside of the optimization problem undertaken in this chapter.

The fundamental aspects of the power allocation that maximizes the energy efficiency of OFDM and OFDMA transmissions are the material for Chapter 10. The Charnes-Cooper Transformation (CCT) from Chapter 5 is used to convert the nonconvex objective EE into a convex one. It is shown that whether we have a single user and single carrier, or multiusers and multicarriers, for a fixed channel assignment, the optimal power allocation (without any constraints) can be obtained through the same simple equation: $\mathrm{EE}(w)=1 /(w \ln 2)$. The cases with a total power constraint as well as proportional and minimum rate constraints are also analyzed.

Chapter 11 provides a two-step solution to the frequency and power allocation that maximizes the energy efficiency of the OFDMA transmissions with minimum rate and total power constraints. Genetic algorithm from Chapter 6 is used for subchannel assignment and an algorithm from Chapter 10 is used for power allocation.

The optimum price that an operator must set when subleasing subchannels to future
ad-hoc networks is the subject of Chapter 12. Strategic and Stackelberg game theories from Chapter 7 are used to model this problem as a composite game.

## Part I

## Theory and Techniques

## Chapter 2

## Mathematical optimization

### 2.1 Introduction

Engineers, computer scientists, and applied mathematicians all agree that we do not know how to solve a vast majority of the optimization problems [81, 82]. There are two type of problems, however, that we have become experts. They are linear programing and the least square minimization [81-84]. Regardless of the number of variables (millions or more) involved, we have very reliable and very fast methods for these two type of problems [81,82]. These methods and their implementations in software have advanced to such heights that we call them mature technologies. Linear programming and least square minimization have even found their way into embedded systems [81, 82]. There is another type of problems for which our methods of solutions have improved rapidly in the last two decades or so. Our methods for these problems are approaching the level of the previously mentioned two types. This is the type called convex optimization [28]. Convex optimization has been extensively used in the design and analysis of engineering systems in the last three decades. The new frontier however, consists of nonconvex problems - both old unsolved problems and new design problems.

We now use an arbitrary optimization problem to introduce the terminology, explain
duality, and then go onto to a convex optimization problem. Generalized convex functions is discussed next and a brief introduction to game theory follows. We finish the chapter with a discussion on what type of solutions are acceptable to the wireless communication industry. All normal faced letters represent real numbers or real valued functions. Bold faced letters represent vectors of real numbers or vector valued functions. $\mathbf{a} \succeq \mathbf{0}$ means $\mathbf{a}$ has non-negative components.

### 2.2 An optimization problem

Suppose we want to find the minimum value of the function $f_{0}(\mathbf{x})$ for $\mathbf{x}$ values in the domain $D$, subject to the inequality constraints $f_{i}$ and equality constraints $h_{i}$. Note that at this point we do not make any assumptions about $f_{0}$ or $f_{i}$ (such as convex or otherwise).

$$
\begin{align*}
\operatorname{minimize} & f_{0}(\mathbf{x})  \tag{2.1}\\
\text { subject to } & f_{i}(\mathbf{x}) \leq 0, \quad i=1,2, \ldots, m  \tag{2.2}\\
& h_{i}(\mathbf{x})=0, \quad i=1,2, \ldots, n  \tag{2.3}\\
& \mathbf{x} \in D \tag{2.4}
\end{align*}
$$

The problem above is called the primal problem and its solution, namely the global optimum value of $f_{0}$, is denoted by $p^{*}$. In the optimization literature the word global is omitted and the phrase optimal solution is used to mean the global optimal solution. The set of all $\mathbf{x} \in D$ that satisfies the constraints is called the feasible set.

If $D$ is a convex set, and the objective function $f_{0}$ and all the constraints $f_{i}$ are convex functions, then the above problem is said to be a convex optimization problem. We will define a convex set, a convex function, and take a closer look at a convex optimization problem in a short while. In the mean time, lets not make any assumptions about the set $D$ or the functions $f_{0}$ or $f_{i}$, convex or otherwise.

### 2.3 The dual problem

A popular technique for solving an optimization problem uses theLagrange Dual Function. The Lagrangian, $L$, is formed in the following way where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the dual variables.

$$
\begin{equation*}
L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})=f_{0}(\mathbf{x})+\sum_{i=1}^{m} \lambda_{i} f_{i}(\mathbf{x})+\sum_{i=1}^{n} \mu_{i} h_{i}(\mathbf{x}) \tag{2.5}
\end{equation*}
$$

The infimum (greatest lower bound) of the Lagrangian over the variable $\mathbf{x}$ is called the Lagrange dual function.

$$
\begin{align*}
g(\boldsymbol{\lambda}, \boldsymbol{\mu}) & =\inf _{\mathbf{x} \in D} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})  \tag{2.6}\\
& =\inf _{\mathbf{x} \in D}\left(f_{0}(\mathbf{x})+\sum_{i=1}^{m} \lambda_{i} f_{i}(\mathbf{x})+\sum_{i=1}^{n} \mu_{i} h_{i}(\mathbf{x})\right) \tag{2.7}
\end{align*}
$$

The following result holds for any $\boldsymbol{\lambda} \succeq 0$ and any $\boldsymbol{\mu}$, and is easily proven by substituting the inequalities in the primal problem into (2.7).

$$
\begin{equation*}
g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq p^{*} \tag{2.8}
\end{equation*}
$$

Since $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a lower bound of the solution to the primal problem, it is natural to ask how high $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ can be, or, what is the maximum value of $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$. The problem of finding the maximum value of $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is called the Dual Problem.

$$
\begin{array}{ll}
\text { maximize } & g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\
\text { subject to } \boldsymbol{\lambda} \succeq 0 \tag{2.10}
\end{array}
$$

The solution of the dual problem is denoted by $d^{*}$. It can be shown easily, that regardless of the primal problem, the dual problem is always convex [81]. Now obviously (2.8) would
imply that

$$
\begin{equation*}
d^{*} \leq p^{*} \tag{2.11}
\end{equation*}
$$

This is called the Weak Duality. Weak duality holds always. Note that so far we have not made any assumptions about the functions we are dealing with.

Under certain conditions $d^{*}=p^{*}$, and this is called the Strong Duality.

$$
\begin{equation*}
d^{*}=p^{*} \tag{2.12}
\end{equation*}
$$

One of the simple sufficient conditions under which strong duality holds is called Slater's Condition [81]. It says that if the primal problem is convex and there is a strictly feasible point satisfying the constraints then Strong Duality holds. It should be noted that for almost all engineering problems this condition on feasibility holds. But convexity may not.

Slater's is a sufficient condition but it is not necessary. There are numerous other less stringent conditions under which Strong Duality holds but all of them are complicated and not useful for engineers [81].

When the primal problem is not convex (hence Slater's condition does not hold), the experts $[81,82]$ suggest that 1) It is best to analyze the problem on its own. 2) Make compromises in the primal problem formulation and see if strong duality holds approximately.

In the following we start with convex sets and functions and go on to explain what is a convex optimization problem.

### 2.4 Convex sets and convex functions

A set $D \subseteq \mathbf{R}^{n}$ is convex if for every pair of points X and Y in $D$, every point on the line segment XY also lies in $D$.

Definition 2.4.1. $D \subseteq \mathbf{R}^{n}$ is convex if for every $\mathbf{x}, \mathbf{y} \in D$ and $0<\lambda<1$, we have

$$
\begin{equation*}
\lambda \mathbf{x}+(1-\lambda) \mathbf{y} \in D \tag{2.13}
\end{equation*}
$$

Two definitions of a convex function follows. The first one says that the graph restricted to a line segment in the domain is under the secant line and the second one says that it is above the tangent line.

Definition 2.4.2. $f: \mathbf{R}^{n} \supseteq D \longrightarrow \mathbf{R}$ is convex, if $D$ is convex and if for any $\mathbf{x}, \mathbf{y} \in D$ and $0<\lambda<1$, we have

$$
\begin{equation*}
f(\lambda \mathbf{x}+(1-\lambda) \mathbf{y}) \leq \lambda f(\mathbf{x})+(1-\lambda) f(\mathbf{y}) \tag{2.14}
\end{equation*}
$$

Definition 2.4.3. A differentiable function $f: \mathbf{R}^{n} \supseteq D \longrightarrow \mathbf{R}$ is convex, if $D$ is convex and for all $\mathbf{x}, \mathbf{y} \in D$,

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{y}) \geq(\mathbf{x}-\mathbf{y}) \bullet \nabla f(\mathbf{y}) \tag{2.15}
\end{equation*}
$$

Note that the above definition immediately leads to the first order optimality condition for a convex function: $\nabla f(\mathbf{y})=\mathbf{0}$ implies $f(\mathbf{y})$ is the global minimum. More importantly, any local minimum is a global minimum. A concave function is defined by reversing the inequality above. An equivalent definition is: $f$ is concave if $-f$ is convex.

### 2.5 Convex structure

The phrase convex structure may refer to a convex or concave structure. It merely states that one of these structures is present. In other words, a function is said to have a convex structure if the function is either convex or concave.

### 2.6 Convex optimization problem

The optimization problem:

$$
\begin{align*}
\operatorname{minimize} & f_{0}(\mathbf{x})  \tag{2.16}\\
\text { subject to } & f_{i}(\mathbf{x}) \leq 0, \quad i=1,2, \ldots, m  \tag{2.17}\\
& h_{i}(\mathbf{x})=0, \quad i=1,2, \ldots, n  \tag{2.18}\\
& \mathbf{x} \in D \tag{2.19}
\end{align*}
$$

is convex if the set $D$, the objective function $f_{0}$, and all the constraints $f_{i}$ are convex. As mentioned before, the strong duality holds for this problem, and a popular method of solving this type of problems is through the Lagrangian dual. Instead of solving the above problem, one solves the dual problem.

### 2.7 The Karush Kuhn Tucker (KKT) Conditions

There is a set of optimality conditions, famously known as the KKT conditions that can be used to theoretically locate the solution to a constrained convex optimization problem. Referring to the problem in (2.16), we first form the Lagrangian:

$$
\begin{equation*}
L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})=f_{0}(\mathbf{x})+\sum_{i=1}^{m} \lambda_{i} f_{i}(\mathbf{x})+\sum_{i=1}^{n} \mu_{i} h_{i}(\mathbf{x}) \tag{2.20}
\end{equation*}
$$

Then the KKT conditions are as follows:

$$
\begin{align*}
\lambda_{i} f_{i}(\mathbf{x}) & =0, & & i=1,2, \ldots, m  \tag{2.21}\\
\mu_{i} h_{i}(\mathbf{x}) & =0, & & i=1,2, \ldots, n  \tag{2.22}\\
\frac{\partial L}{\partial x_{i}} & =0, & & i=1,2, \ldots, m  \tag{2.23}\\
f_{i}(\mathbf{x}) & \leq 0, & & i=1,2, \ldots, m  \tag{2.24}\\
h_{i}(\mathbf{x}) & =0, & & i=1,2, \ldots, n \tag{2.25}
\end{align*}
$$

We said KKT conditions locate the optimal point theoretically, because solving these equations in itself could be a hard problem. Note that except in the case of a quadratic objective and quadratic constraints, the KKT conditions lead to a system of non-linear equations. There are no efficient algorithms for solving systems of non-linear equations. The most popular algorithm for this purpose, the multivariate Newton-Raphson, is not only computationally expensive but highly unstable $[28,85]$.

### 2.8 Generalized convex functions

Many optimization problems encountered in engineering and economics cannot be considered as convex optimization problems because the objective is either known to be nonconvex (nonconcave) or no one was able to prove the objective is convex (concave). Slater's is a sufficient condition but it is not necessary for the duality gap to be zero. Convexity is a sufficient condition for the KKT conditions to give the optimum point but it is not necessary. The combination of these three facts has led many researchers to question if there is any broader class of functions than convex for which KKT conditions hold or duality gap is zero. There have been several proposals to extend or loosen the definition of convexity [86, 87]. In the following, we discuss three of these extensions, the first two of which are well known
while the last one is relatively new and unknown. This last extension of convexity, known as invexity is later used in this dissertation to solve a difficult optimization problem regarding throughput maximization in the Gaussian interference channel.

### 2.8.1 Quasiconvex functions

In 1949, de Finetti modified [86] the right side of the Definition 2.4.2 of convex functions to introduce what became known as quasiconvex functions.

Definition 2.8.1. A function $f: \mathbf{R}^{n} \supseteq D \longrightarrow \mathbf{R}$ is quasiconvex if for all $\mathbf{x}, \mathbf{y} \in D$ and for any $\lambda \in[0,1]$, we have

$$
\begin{equation*}
f(\lambda \mathbf{x}+(1-\lambda) \mathbf{y}) \leq \max \{f(\mathbf{x}), f(\mathbf{y})\} \tag{2.26}
\end{equation*}
$$

For differentiable functions, quasiconvexity can be defined as follows.

Definition 2.8.2. A differentiable function $f: \mathbf{R}^{n} \supseteq D \longrightarrow \mathbf{R}$ is quasiconvex if

$$
\begin{equation*}
f(\mathbf{x}) \leq f(\mathbf{y}) \Longrightarrow(\mathbf{x}-\mathbf{y}) \bullet \nabla f(\mathbf{y}) \leq 0, \text { for all } \mathbf{x}, \mathbf{y} \in D \tag{2.27}
\end{equation*}
$$

### 2.8.2 Pseudoconvex functions

Tuy (1964) and Mangasarian (1965) introduced differentiable pseudoconvex functions [86] by switching the sides and reversing the inequalities in the definition of a quasiconvex function.

Definition 2.8.3. A differentiable function $f: \mathbf{R}^{n} \supseteq D \longrightarrow \mathbf{R}$ is pseudoconvex if

$$
\begin{equation*}
(\mathbf{x}-\mathbf{y}) \bullet \nabla f(\mathbf{y}) \geq 0 \Longrightarrow, f(\mathbf{x}) \geq f(\mathbf{y}) \text { for all } \mathbf{x}, \mathbf{y} \in D \tag{2.28}
\end{equation*}
$$

One of the remarkable things about differentiable pseudoconvex functions is that the first order optimality condition for the convex functions also applies to pseudoconvex functions.

Note that if $\nabla f(\mathbf{y})=\mathbf{0}$, then the definition above implies that $f(\mathbf{y})$ is the global minimum. In other words, a local minimum is always a global minimum. In case of constrained optimization, pseudoconvexity is sufficient for the KKT conditions.

### 2.8.3 Invex functions

Even after the introduction of pseudoconvexity, many problems in economics and engineering remained unsolved. In 1981, Hanson introduced a much wider form of generalization of convexity, that came to be known as invexity, by replacing the ( $\mathbf{x}-\mathbf{y}$ ) term in the Definition 2.4.3 of a convex function with an arbitrary vector function [88-90]. He went on to show that in case of a constrained invex problem, the KKT conditions are sufficient for optimality. We give a short but comprehensive account of invex analysis in Chapter 4 and use invexity to solve the throughput maximization problem in the GIC in Chapter 9.

### 2.9 Game theory

Certain design problems in engineering can be modeled as the simultaneous optimization of a number of multivariable functions in the same variables, where different variables are controlled by different agents. Strategic and Stackelberg game theories can then be used to find the solution to these problems.

An important question is whether game theory belongs in convex optimization or in nonconvex optimization. An argument can be made for each side. If the payoff functions involved in the game are convex then we can say we are performing convex optimization. However, the most famous concept in the strategic game theory, the Nash equilibrium, can exist even when the payoff functions are nonconvex. The same is true for the existence of the outcome of a Stackeberg game. Therefore, one can consider game theory as part of nonconvex optimization too. This is particularly true when we prove the existence of the equilibrium outcome directly, without involving the theorems that peg the existence of the
equilibrium on the convexity of the payoff functions.

### 2.10 Real time problems

We saw that the optimization problems can be classified into linear programming, convex optimization or nonconvex optimization. There is another way to categorize optimization problems that is perhaps not important to mathematicians but certainly very important to engineers. This categorization is based on whether the solution to the optimization problem is required on a real time basis or not. Consider the flight scheduling problem at an airport. Flights are scheduled at least months ahead of time. This means the method used to optimally use the airport runways and time can take months to arrive at the solution. We are not optimizing any function when a plane is approaching the airport. This means the flight scheduling is not a real time application. If the software used for this optimization crashes once in a while, this is not a big problem.

Now consider the power allocation problem at a cellular base station. The users and their channel conditions are changing every second. Even if we decide to optimize the power allocation only every 10 seconds or so, there is not enough time for a software engineer to run a computer program in the computer and then relay the solution to the manager. This is an example of a real time optimization problem. The optimization has to be performed automatically, at very high speeds, and at the chip level at the base station. The wireless industry will accept only a solution that can be embedded in the chip that controls the power transmitter at the base station.

This brings us to the question of which type of optimization solutions or algorithms can be embedded in chips. First of all, the reason we are going to the chip level is speed. The algorithm has to be very fast. Only polynomial time algorithms are suitable. Second, the algorithm has to be stable. An algorithm that converges to the optimum result for all values of the parameters involved is said to be stable. For example, a power allocation algorithm has
to converge for all values of the channel gains it could possibly encounter in the application. We need a theoretical proof that the algorithm will converge for all values of the channel gains as well as extensive numerical results through simulations that demonstrate the convergence.

## Chapter 3

## Cross product in higher dimensions

### 3.1 Introduction

Consider a resource allocation problem that maximizes the sum output of a number of users under a total constraint on the resource. If the user's proportional outputs are admitted as additional constraints, in some cases, it may be possible to locate the point of maximum sum output, under a total resource constraint, using analytical geometry whether the sum output is concave or not. This technique is used in Chapter 8 to solve the problem of finding the power allocation that maximizes the sum transmission rate (throughput) of the Gaussian interference channel with proportional rate and total power constraints.

Suppose there are $N$ users. The idea is to follow the curve of intersection of the sum rate and the proportional rate constraints in the $N+1$-dimensional Euclidean space until it intersects the power constraint plane. Crucial in this approach is the increasing/decreasing behavior of the said curve of intersection. This brings us to the tangent line of this curve. In 3-dimensions, the direction of the tangent line of the curve of intersection of two surfaces can be found by taking the cross-product of the normal vectors of the two surfaces. This idea falls apart in dimensions higher than three because the traditional cross product is not defined in higher dimensions. In what follows, we explain this in detail and propose a
solution.

### 3.2 Traditional cross product

In 3-dimensions, two non-parallel vectors can be used to define a unique direction that is perpendicular to both vectors, using the familiar cross product. In higher dimensions, the idea of a cross product of two vectors falls apart because of the following reason. In 4-dimensions, given two non-parallel vectors, there are an infinite number of vectors that are perpendicular to both these vectors. In fact, these vectors will form a plane that is perpendicular to the given two vectors. Suppose we use $X, Y, Z, W$ to label the axes in 4-dimensions. Consider the $X$ and $Y$ axes. $Z$ axis is perpendicular to both $X$ and $Y$, but so is $W$. In fact, every direction in the $Z W$-plane is perpendicular to both $X$ and $Y$ directions. Similarly, in 5-dimensions, the vectors perpendicular to given two vectors will form a 3 -dimensional subspace.

In light of this, mathematicians have defined various advanced "products" of two vectors in higher spaces such as Wedge product and Clifford product. While Wedge product and Clifford product have found numerous important applications in Physics and Engineering, it turns out that for our present work, we could use an extension of the familiar cross product after all.

### 3.3 Extended cross product

There is a way to extend the cross product into higher dimensions [91]. In 4-dimensions, the idea is to use three linearly independent vectors to define a unique direction perpendicular to these three. For example, $W$ axis is the only direction perpendicular to $X, Y$ and $Z$ in 4-dimensions. Moreover, this new direction can be found by using the familiar determinant style formulation.

Let us consider two illustrative examples from the 4-dimensional space. Suppose we use $\theta_{1}, \theta_{2}, \theta_{3}$, and $\theta_{4}$ to denote the unit vectors in the 4 coordinate directions. Let us cross $(1,0,0,0),(0,1,0,0)$, and $(0,0,1,0)$ together. Note that these are the unit vectors in the coordinate directions $X, Y$ and $Z$. We expect the result to be the unit vector in the $W$ direction. Note that in the following, the result is obtained simply by calculating the determinant.

$$
\left|\begin{array}{cccc}
\theta_{1} & \theta_{2} & \theta_{3} & \theta_{4}  \tag{3.1}\\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{array}\right|=(0,0,0,1)
$$

as expected. As a second example, consider the cross product of $\mathbf{v}_{\mathbf{1}}=(1,1,0,0), \mathbf{v}_{\mathbf{2}}=$ $(0,1,1,0)$ and $\mathbf{v}_{\mathbf{3}}=(0,0,1,1)$.

$$
\left|\begin{array}{cccc}
\theta_{1} & \theta_{2} & \theta_{3} & \theta_{4}  \tag{3.2}\\
1 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1
\end{array}\right|=(1,-1,1,-1)=\mathbf{v}
$$

We can easily verify that the result $\mathbf{v}$ is perpendicular to each of the $\mathbf{v}_{\mathbf{i}}$ 's by taking the dot product. That is, $\mathbf{v} \cdot \mathbf{v}_{\mathbf{i}}=0$ for $i=1,2,3$.

In general, in the $N$-dimensional space, we can use $N-1$ linearly independent vectors to define a unique direction that is perpendicular to all $N-1$ vectors, using the same determinant formulation of the extension of the 3-dimensional cross product.

## Chapter 4

## Invexity

### 4.1 Introduction

Optimizing non-convex functions is difficult mainly for two reasons. First is that the simple characterization of the critical point such as the KKT conditions may not be applicable. The second is the possible existence of many local optima; a search algorithm that is looking for the global optimum point might get stuck at a local optimum point. Thankfully, convexity is not the only condition under which a local extremum necessarily becomes a global extremum. As mentioned in Chapter 2, the famous example is pseudo-convexity.

In the 1980's a more general geometrical structure called invexity (incavity) was proposed [88-90]. What is impressive about invexity is that while it is more general than pseudo-convexity and quasi-convexity, invexity guarantees that any local extremum is a global extremum. Furthermore, under certain conditions, the solution to an invex, constrained, optimization problem can be obtained via the KKT conditions. Invexity is applicable only to differentiable functions. For engineers, this is not a serious limitation, because most functions we attempt to optimize are differentiable.

We present a short treatment of invex analysis with a perspective from our application of the theory. The early development here parallels what is in [88-90], although some theorems
there are definitions here and some definition there are theorems here. Towards the end of this section however, we deviate from the development in [88-90], and provide an original theorem that is better suited to our application.

### 4.2 Concavity

We start with a definition and two theorems from convex analysis, which we purposefully present here instead of an appendix because the theory of invex analysis and the proofs of the theorems closely follow of that of the convex analysis. Unless stated otherwise, all functions are assumed to be differentiable. Bold face letters represent vectors or vector functions. a•b indicates the inner product of $\mathbf{a}$ and $\mathbf{b}$.

Definition 4.2.1. Concavity: A scalar function $f: D \subseteq \mathbf{R}^{n} \longrightarrow \mathbf{R}$ is said to be concave in domain $D$ if

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{y}) \leq(\mathbf{x}-\mathbf{y}) \bullet \nabla f(\mathbf{y}) \text { for all } \mathbf{x}, \mathbf{y} \in D \tag{4.1}
\end{equation*}
$$

Theorem 4.2.1. Suppose $f$ is concave in $D$. Then $\nabla f(\mathbf{a})=\mathbf{0}$ for an $\mathbf{a} \in D$ implies $f(\mathbf{a})$ is the global maximum of $f$ in $D$.

Proof: Suppose $\nabla f(\mathbf{a})=\mathbf{0}$. Then for all $\mathbf{x} \in D$ we have

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{a}) \leq(\mathbf{x}-\mathbf{a}) \bullet \nabla f(\mathbf{a})=(\mathbf{x}-\mathbf{a}) \bullet \mathbf{0}=0 \tag{4.2}
\end{equation*}
$$

Note that the converse is not true. That is, " $\nabla f(\mathbf{a})=\mathbf{0}$ implies $f(\mathbf{a})$ is the global maximum" does not imply that $f$ is concave.

Remark 4.2.1. Note that in order to prove that $f(\mathbf{a})$ is the maximum, we do not necessarily need (4.1) to be valid at $\mathbf{x}=\mathbf{a}$. As long as (4.1) is valid for every point $\mathbf{x} \neq \mathbf{a}$, the proof
stands. This shows that $f(\mathbf{a})$ will be the maximum even if (4.1) is valid everywhere but not at $\mathbf{a}$.

The following theorem is about what we know traditionally about constraint optimization in its simplest form. It deals with the familiar Kuhn-Tucker condition [88] for the maximum of $f(\mathbf{x})$ subject to the constraint $g(\mathbf{x}) \geq 0$.

Theorem 4.2.2. Suppose $f$ and $g$ are concave scalar functions defined in domain $D$ and there is a point $\mathbf{a}$ in $D$ such that $\nabla f(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$ with scalar $\lambda>0$ and $g(\mathbf{a})=0$. Then $f(\mathbf{a})$ is the global maximum of $f(\mathbf{x})$ in $D$ subject to the constraint $g(\mathbf{x}) \geq 0$.

Proof: The concavity of $f$ and $g$ allows us to write that for all $\mathbf{x} \in D$,

$$
\begin{align*}
& f(\mathbf{x})-f(\mathbf{a}) \leq(\mathbf{x}-\mathbf{a}) \bullet \nabla f(\mathbf{a}) \text { and }  \tag{4.3}\\
& g(\mathbf{x})-g(\mathbf{a}) \leq(\mathbf{x}-\mathbf{a}) \bullet \nabla g(\mathbf{a})
\end{align*}
$$

Multiplying the second inequality by $\lambda$ and adding to the first:

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{a})+\lambda g(\mathbf{x})-\lambda g(\mathbf{a}) \leq(\mathbf{x}-\mathbf{a}) \bullet[\nabla f(\mathbf{a})+\lambda \nabla g(\mathbf{a})] \tag{4.4}
\end{equation*}
$$

Substituting for quantities that are zero and re-arranging:

$$
\begin{equation*}
f(\mathbf{x})+\lambda g(\mathbf{x}) \leq f(\mathbf{a}) \tag{4.5}
\end{equation*}
$$

Since $\mathbf{x}$ is constraint to $g(\mathbf{x}) \geq 0$ and $\lambda>0$, we have

$$
\begin{equation*}
f(\mathbf{x}) \leq f(\mathbf{a}) \tag{4.6}
\end{equation*}
$$

Remark 4.2.2. Note that $f(\mathbf{a})$ will be the maximum even if (4.6) is valid everywhere but
not at $\mathbf{x}=\mathbf{a}$. In other words, $f(\mathbf{a})$ will be the maximum even if (4.3) is valid everywhere but not at $\mathbf{x}=\mathbf{a}$.

### 4.3 Incavity

We now start our exposition on incavity. Hanson [88] noticed that the term $(\mathbf{x}-\mathbf{a})$ in the definition of concavity in (4.1) is not special and that it can be replaced by any other vector function. He went on to define Incavity as follows.

Definition 4.3.1. Incavity: A scalar function $f: D \subseteq \mathbf{R}^{n} \longrightarrow \mathbf{R}$ is said to be incave in the domain $D$ if there exists a vector function $\boldsymbol{\eta}: D \times D \longrightarrow \mathbf{R}^{n}$ such that

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{y}) \leq \boldsymbol{\eta}(\mathbf{x}, \mathbf{y}) \bullet \nabla f(\mathbf{y}) \text { for all } \mathbf{x}, \mathbf{y} \in D \tag{4.7}
\end{equation*}
$$

Invexity is defined by reversing the inequality. In the following, we will use $\boldsymbol{\eta}$ to denote $\boldsymbol{\eta}(\mathrm{x}, \mathrm{y})$.

Notice that just like concavity, incavity is a linear property. That is, if $f$ and $g$ are incave then so is $f+k g$ for any positive $k$.

A function that is concave (convex) is necessarily incave (invex). This can be easily proven by taking $\boldsymbol{\eta}=\mathbf{x}-\mathbf{a}$. It is important to note that the converse is not true. A linear function is both concave and convex, and therefore, both incave and invex.

The analogue of Theorem 4.2.1 is stronger; the converse is true too.
Theorem 4.3.1. $f$ is incave in $D$ if and only if $\nabla f(\mathbf{a})=\mathbf{0}$ for an $\mathbf{a} \in D$ implies $f(\mathbf{a})$ is the global maximum of $f$ in $D$.

Proof: The proof of the forward direction is exactly the same as the proof of Theorem 4.2.1 but $(\mathbf{x}-\mathbf{a})$ replaced with $\boldsymbol{\eta}$. To prove the reverse direction, assume that the statement
" $\nabla f(\mathbf{a})=\mathbf{0}$ implies $f(\mathbf{a})$ is the global maximum of $f$ " is true. Let us refer to this statement as the hypothesis. We need to prove that $f$ is incave. We shall produce an $\boldsymbol{\eta}$ and demonstrate that (4.7) holds. We have two cases: Either $\nabla f(\mathbf{y})=\mathbf{0}$ or $\nabla f(\mathbf{y}) \neq \mathbf{0}$.

If $\nabla f(\mathbf{y})=\mathbf{0}$, by the hypothesis we know $f(\mathbf{y})$ is the global maximum and therefore we have,

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{y}) \leq 0 \tag{4.8}
\end{equation*}
$$

If we take $\boldsymbol{\eta}=\mathbf{0}$, (4.7) will hold and therefore $f$ is incave.
If $\nabla f(\mathbf{y}) \neq \mathbf{0}$, take

$$
\begin{equation*}
\boldsymbol{\eta}=\frac{f(\mathbf{x})-f(\mathbf{y})}{\|\nabla f(\mathbf{y})\|^{2}} \nabla f(\mathbf{y}) \tag{4.9}
\end{equation*}
$$

Then

$$
\begin{equation*}
\eta \bullet \nabla f(\mathbf{y})=\frac{f(\mathbf{x})-f(\mathbf{y})}{\|\nabla f(\mathbf{y})\|^{2}} \nabla f(\mathbf{y}) \bullet \nabla f(\mathbf{y})=f(\mathbf{x})-f(\mathbf{y}) \tag{4.10}
\end{equation*}
$$

Notice that the above equation implies the inequality in (4.7) and therefore $f$ is incave.
The next theorem is the analogue of Theorem 4.2.2.
Theorem 4.3.2. Suppose $f$ and $g$ are incave with respect to the same function $\boldsymbol{\eta}$ everywhere in $D$ except possibly at point $\mathbf{a}$ where we have: $\nabla f(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$ with scalar $\lambda>0$ and $g(\mathbf{a})=0$. Then $f(\mathbf{a})$ is the global maximum of $f(\mathbf{x})$ in $D$ subject to the constraint $g(\mathbf{x}) \geq 0$.

Proof: Straightforward from the proof of Theorem 4.2 .2 by replacing $(\mathbf{x}-\mathbf{a})$ with $\boldsymbol{\eta}$.

In convex analysis, the component $\mathbf{x}-\mathbf{y}$ in the definition of concavity is common to all functions that are concave, by definition. In contrast, in incave analysis, two functions that are incave in the same domain may be incave with respect to two different $\boldsymbol{\eta}$ functions. Theorem 4.3.2 however, requires that the functions be incave with respect to the same $\boldsymbol{\eta}$ function. This raises the question: what is the condition two incave functions have to satisfy in order to be incave with respective to the same $\boldsymbol{\eta}$ function? Before we can answer that
question, we must answer the following question: is there a sufficient condition that ensures the incavity of a function?

### 4.4 A sufficient condition for incavity

The following theorem asserts that a function without any stationary points is necessarily incave and invex.

Theorem 4.4.1. Suppose $f(\mathbf{x})$ has the property $\nabla f(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$ in $D$. Then $f$ is incave and invex in $D$.

Proof: Since $\nabla f(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$ in $D$, we have $\|\nabla f(\mathbf{y})\| \neq 0$ for all $\mathbf{y}$ in $D$. This enables us to write

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{y})=\left[\frac{f(\mathbf{x})-f(\mathbf{y})}{\|\nabla f(\mathbf{y})\|^{2}} \nabla f(\mathbf{y})\right] \bullet \nabla f(\mathbf{y}) \tag{4.11}
\end{equation*}
$$

For each $\mathbf{y} \in D$, define $\mathbf{u}(\mathbf{y})$ as any vector with components that have the same sign as that of the components of $\nabla f(\mathbf{y})$. Then $\mathbf{u} \bullet \nabla f(\mathbf{y}) \geq 0$ for all $\mathbf{y} \in D$. We can now write

$$
\begin{equation*}
f(\mathbf{x})-f(\mathbf{y}) \leq\left[\frac{f(\mathbf{x})-f(\mathbf{y})}{\|\nabla f(\mathbf{y})\|^{2}} \nabla f(\mathbf{y})+\mathbf{u}\right] \bullet \nabla f(\mathbf{y}) \tag{4.12}
\end{equation*}
$$

Comparing the above equation with the definition of incavity in (4.7), we see that $f$ is incave everywhere in $D$ with respect to the function

$$
\begin{equation*}
\boldsymbol{\eta}=\frac{f(\mathbf{x})-f(\mathbf{y})}{\|\nabla f(\mathbf{y})\|^{2}} \nabla f(\mathbf{y})+\mathbf{u} \tag{4.13}
\end{equation*}
$$

Note that we can prove $f$ is invex by choosing $\mathbf{u}$ as any vector with components that have the opposite sign to that of the components of $\nabla f(\mathbf{y})$.

### 4.5 Being incave with respect to the same function

The function $\boldsymbol{\eta}$ in (4.13) can be written as

$$
\begin{equation*}
\boldsymbol{\eta}=t \nabla f(\mathbf{y})+\mathbf{u} \tag{4.14}
\end{equation*}
$$

where $t$ is a scalar. Note that this $\boldsymbol{\eta}$ is one possible functions among many that will make the proof work. For given $\mathbf{x}$ and $\mathbf{y}$, what are the possible choices for the vector $\mathbf{u}$, that would make the proof that $f$ is incave valid? Under which condition a number of functions with no stationary points can have a common $\boldsymbol{\eta}$ function at a point? The reader is referred to [88] and [90] for the answers to these questions in the most general context. In what follows, we give an original theorem whose scope is narrower than the theorems in [88] and [90], but the proof is much simpler. The scope of this theorem is sufficient for our application of invex analysis in this article.

Theorem 4.5.1. Suppose $f$ and $g$ are two functions with $\nabla f(\mathbf{x}) \neq \mathbf{0}$ and $\nabla g(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$ in $D$ and hence incave in $D$. Then $f$ and $g$ are incave with respect to a common $\boldsymbol{\eta}$ at every point $y$ in $D$ where $\nabla f(\mathbf{y})$ and $\nabla g(\mathbf{y})$ do not point in exactly opposite directions.

Proof: Since $\nabla f(\mathbf{x}) \neq \mathbf{0}$ and $\nabla g(\mathbf{x}) \neq \mathbf{0}$, by Theorem 4.4.1, and by the observation at the beginning of the paragraph above, $f$ and $g$ are incave with respect to $\boldsymbol{\eta}_{\boldsymbol{1}}$ and $\boldsymbol{\eta}_{\boldsymbol{2}}$ given by:

$$
\begin{align*}
& \boldsymbol{\eta}_{\mathbf{1}}=t \nabla f(\mathbf{y})+\mathbf{u}_{\mathbf{1}}  \tag{4.15}\\
& \boldsymbol{\eta}_{\mathbf{2}}=s \nabla g(\mathbf{y})+\mathbf{u}_{\mathbf{2}}, \tag{4.16}
\end{align*}
$$

where $t$ and $s$ are scalars and $\mathbf{u}_{\mathbf{1}}$ and $\mathbf{u}_{\mathbf{2}}$ are vectors with $\mathbf{u}_{\mathbf{1}} \bullet \nabla f(\mathbf{y}) \geq 0$ and $\mathbf{u}_{\mathbf{2}} \bullet \nabla g(\mathbf{y}) \geq 0$. We use Fig. 4.1 to show that it is possible to choose $\mathbf{u}_{\boldsymbol{1}}$ and $\mathbf{u}_{\mathbf{2}}$ so that $\boldsymbol{\eta}_{\boldsymbol{1}}=\boldsymbol{\eta}_{\boldsymbol{2}}$, as long as $\nabla f(\mathbf{y})$ and $\nabla g(\mathbf{y})$ do not face exactly opposite directions.

Given a vector $\boldsymbol{a}$, the possible directions for a vector $\mathbf{b}$ that satisfies the condition $\boldsymbol{a} \bullet \boldsymbol{b} \geq 0$ is shown in Fig. 4.1(a). Given the directions of $\nabla f(\mathbf{y})$ and $\nabla g(\mathbf{y})$, how we can choose $\mathbf{u}_{1}$ and $\mathbf{u}_{\mathbf{2}}$ so that we end up with the same $\boldsymbol{\eta}$ is illustrated in Fig. 4.1(b) and (c). Two different relative orientations for $\nabla f(\mathbf{y})$ and $\nabla g(\mathbf{y})$ are shown in Fig. 4.1(b) and (c). It is clear from these two diagrams that the only orientation of $\nabla f(\mathbf{y})$ and $\nabla g(\mathbf{y})$ where there might not be a common $\boldsymbol{\eta}$ is when they point in exact opposite directions.


Figure 4.1: Two functions having the same $\boldsymbol{\eta}$.

## Chapter 5

## Convexifying, increasing transformations

### 5.1 Introduction

If a non-convex objective function can be transformed into a convex function in such a manner that the order of the local extreme values is preserved, the tools of convex analysis can then be used to find the optimum value of this function. A moment of reflection tells us that the order of the local extreme values will be preserved if the transformation is increasing ${ }^{1}$. Suppose $T$ transforms $f(\mathbf{x})$ into $g(\mathbf{X})$. $T$ is said to be increasing if $g(\mathbf{X})$ increases whenever $f(\mathbf{x})$ increases. Note that $g(\mathbf{X})$ will increase with $f(\mathbf{x})$ if each component of $\mathbf{X}$ increases with $\mathbf{x}$.

The concept of an increasing transformation mapping a single variable non-convex function into a single variable convex function is mystifying. An increasing transformation will preserve all the local extrema; how can the resulting function then be convex? One can safely conclude no such transformation exists. The multi-variable context is not this simple.

[^0]Furthermore, a transformation can increase the number of variables.
A natural question is what type of functions can be transformed into convex functions in this manner. Despite considerable research, the results in this area are disappointing. We only have a necessary condition for convexifyability. In other words, we know when a non-convex function is not convexifyable. A theorem due to Fenchel asserts that only quasi-convex functions are convexifyable [87]. This leaves us with only isolated examples of quasi-convex functions and the corresponding confexifying transformations. In what follows we present two such transformations, the second of which is later used in energy efficient power allocation for OFDMA transmissions.

### 5.2 The exponential transformation

Suppose $f$ and $g$ are two single variable scalar functions. It is straight forward to prove that if $f$ and $g$ are convex and if $f$ is increasing then $f \circ g$ is convex. Now consider the multi-variable scalar function $f: \mathbf{R}^{n} \longrightarrow \mathbf{R}$ defined by $f\left(x_{1}, x_{2}, \ldots x_{n}\right)=\exp \left(a_{1} x_{1}+a_{2} x_{2}+\ldots+a_{n} x_{n}\right)$ where $a_{1}, a_{2}, \ldots a_{n}$ are real numbers. Since $a_{1} x_{1}+a_{2} x_{2}+\ldots+a_{n} x_{n}$ is convex (linear) and $\exp (x)$ is increasing and convex, it follows that $f\left(x_{1}, x_{2}, \ldots x_{n}\right)=\exp \left(a_{1} x_{1}+a_{2} x_{2}+\ldots+a_{n} x_{n}\right)$ is convex.

Suppose the function one wishes to optimize is of the form $g\left(x_{1}, x_{2}, \ldots x_{n}\right)=\prod_{i} x_{i}^{a_{i}}$, which in general, may not be convex. The exponential transformation $x_{i}=\exp \left(y_{i}\right)$ maps $g$ to $h\left(y_{1}, y_{2}, \ldots y_{n}\right)=\exp \left(a_{1} y_{1}+a_{2} y_{2}+\ldots+a_{n} y_{n}\right)$, which is convex. The fact that $x_{i}$ increases with $y_{i}$ ensures that the exponential transformation is increasing.

Example 5.2.1. Consider $z=g(x, y)=x^{2} \sqrt{y}$. Letting $y=c$, where $c$ is a constant, shows that the cuts parallel to the $Z X$-plane are convex. Letting $x=c$ shows that the cuts parallel to the YZ-plane are concave. Hence, $g$ is not convex (Fig. 5.1). Under the transformation $x=\exp (X), y=\exp (Y), g(x, y)$ becomes $h(X, Y)=\exp (2 X+(1 / 2) Y)$, which is convex as proven earlier (Fig. 5.2).


Figure 5.1: $z=g(x, y)=x^{2} \sqrt{y}$


Figure 5.2: $h(X, Y)=\exp (2 X+(1 / 2) Y)$

### 5.3 Charnes-Cooper transformation (CCT)

Consider an objective that is the ratio of two functions where the numerator is concave and the denominator is affine. Such an objective, in general, may not be concave. CCT can be used to transform this objective into a concave one. The CCT transformation increases the number of variables by one. Naturally, the new variable's relationship to at least one of the original variables now becomes part of the optimization problem. However, as far as the objective function is concerned, all the variables are treated as independent. The relationship between the new variable and the original variables becomes a constraint. In this way, CCT transformation changes an unconstrained optimization problem into a constrained optimization problem. In case of constrained optimization, CCT increases the number of constraints by one.

Suppose the objective is $f(\mathbf{x}) / g(\mathbf{x})$, where $f$ is concave and $g>0$ is affine. The CCT transformation is given by: $t=1 / g(\mathbf{x})$ and $\mathbf{y}=t \mathbf{x}$. The first equation becomes a constraint in the optimization problem after the transformation. The objective $f(\mathbf{x}) / g(\mathbf{x})$ becomes $t f(\mathbf{y} / t)$. The new optimization problem is: maximize $t f(\mathbf{y} / t)$ subject to the constraints $t=1 / g(\mathbf{x})$ and $t>0$.

The reason why $t f(\mathbf{y} / t)$ is concave is best understood through a well known result from convex analysis. Consider the numerator $f(\mathbf{x})$ of the original objective and the new objective $t f(\mathbf{y} / t)$. The transformation from any function $h(\mathbf{x})$ to $t h(\mathbf{x} / t)$ for a fixed $t$ is called the perspective transformation in convex analysis [28]. Perspective transformation is one of the standard transformations studied under the umbrella of "convexity preserving" transformations. In other words, if $f(\mathbf{y})$ is convex then $t f(\mathbf{y} / t)$ is convex. In our case, we know that $f(\mathbf{x})$ is concave, which means $f(\mathbf{y})$ is concave, and hence, $t f(\mathbf{y} / t)$ is concave.

To show that CCT is an increasing transformation, note that $t f(\mathbf{y} / t)$ can be written as $f(\mathbf{y} / t) / g(\mathbf{y} / t)=(f / g)(\mathbf{y} / t)$. In other words, CCT transforms $(f / g)(\mathbf{x})$ into $(f / g)(\mathbf{y} / t)$. The relationship $\mathbf{y}=t \mathbf{x}$ can be re-written as $\mathbf{x}=\mathbf{y} / t$. This shows that each component of $\mathbf{y} / t$ is increasing with the corresponding component of $\mathbf{x}$.

Example 5.3.1. Consider the single variable function $f(x)=\ln (x) / x$. The numerator $\ln (x)$ is concave while the denominator $x$ is affine. As Fig. 5.3 shows $f$ is not concave. Applying the CCT transformation to $f$, we let $t=1 / x$ and $x=y / t$. This produces a concave function $g(y, t)=t \ln (y / t)$, which is shown in Fig. 5.4.


Figure 5.3: $f(x)=\ln (x) / x$


Figure 5.4: $g(y, t)=t \ln (y / t)$

## Chapter 6

## Genetic algorithm

### 6.1 Introduction

Genetic algorithm is one of the algorithms from a class of biology inspired algorithms called evolutionary algorithms that mimic the nature to solve optimization problems. Ant colony and particle swarm optimizations are two other examples of evolutionary algorithms.

Genetic algorithm (GA) is the culmination of many researchers' work [92]. It was popularized in North America in the 1970's by John Holland [26]. GA can be used to solve almost any optimization problem. It can be used to optimize a function with continuous, discrete or mixed variables. GA has been used to obtain near optimal solutions to many industrial problems that are too complex for analytical methods. Examples include flight scheduling in airports and design of engine parts [93,94]. GA uses a chromosome to represent a particular value of a variable in the optimization problem. The algorithm then creates improved values of that variable by mimicking the birth of children in the natural world with better chromosomes than their parents in terms of surviving the environment. Consider a maximization problem. A number of chromosomes (or individuals) are randomly created in the beginning. These individuals form the initial population. The individuals are then ranked by evaluating a measure called the fitness function. A fraction of the initial individuals with the highest
rank are allowed to produce new chromosomes (or children). The children and a fraction of the high ranking individuals form the next generation. The process is repeated for many generations. Naturally, the fitness function is often none other than the objective function of the optimization problem itself.

In the following we use two specific examples to illustrate GA in detail. The first one has discrete variables whereas the second involves continuous variables.

### 6.2 Genetic algorithm with discrete variables

The combinatorial optimization problem considered in the following example has an obvious solution. We deliberately chose this problem to illustrate how close to the optimal solution can the solution from GA can be.

Example 6.2.1. Throughput maximizing OFDMA channel assignment with equal power allocation and without rate constraints. There are $K$ orthogonal subchannels that are to be assigned to $N(N<K)$ Users so that one subchannel is assigned to only one User. The channel gain for User-n on subchannel-k, $a_{k n}$ is known for all $n$ and $k$.

The obvious solution to this problem is to assign each subchannel to the user with the highest channel gain on that subchannel. In the following, we will pretend we do not know this and let genetic algorithm decide the subchannel assignment.

### 6.2.1 Chromosome

Central in genetic algorithm is the idea of a chromosome. Associated with each variable in the optimization problem there will be a chromosome. During the execution of the algorithm, the chromosome will hold the valuation of the variable it is associated with. In our case, there is only one variable. It is an array indicating the subchannel assignment. Consider a one-dimensional array consisting of $K$ cells, each cell representing a subchannel. The number


Figure 6.1: (a) The Chromosome (b) One point cross-over.

Table 6.1: The Genetic Algorithm

| Step 1 | Randomly generate 3P number of <br> chromosomes or individuals. |
| :---: | :---: |
| Step 2 | Evaluate the fitness function <br> of each individual. |
| Step 3 | Pick the P number of individuals <br> with the highest fitness value <br> and refer to them as the Elites. |
| Step 4 | Pick the P number of individuals <br> with the lowest fitness values <br> and erase them. |
| Step 5 | Call the rest of the P number <br> of individuals Middles. |
| Step 6 | Put the P number of Elites and the <br> P number of Middles in one group and <br> randomly pair them in groups of two. |
| Step 7 | For each pair, use the chromosomes <br> of the two individuals to create two <br> new chromosomes or children using <br> one-point crossover (Fig. 6.1(b)). |
| Step 8 | Occasionally, mutate a child's <br> chromosome by interchanging the <br> numbers in two randomly selected <br> cells in the chromosome. |
| Step 9 | Put the P number of Elites and <br> the 2P number of Children together <br> to form the new generation of <br> $3 P$ individuals. |
| Step 10 | Stop if the number of generations <br> is equal to Ngen. <br> Otherwise go to Step 2. |

in each cell indicates the User to whom that subchannel is being assigned. For example, in Fig. 6.1(a), cell number 5 contains the number 7 . This means that subchannel 5 is assigned to user 7. A different subchannel assignment will be represented by a different chromosome. Each chromosome or subchannel assignment will lead to a particular value for the throughput after (equal) power allocation. Using $\rho(k)$ to denote the array representing the chromosome, the throughput resulting from that chromosome can be written as:

$$
\begin{equation*}
R=\sum_{k} \log _{2}\left(1+\frac{a_{k \rho(k)} p}{\sigma_{k}^{2}}\right) \tag{6.1}
\end{equation*}
$$

where $\sigma_{k}^{2}$ is the background noise power per Hz and $p$ is the power. The higher the throughput is, the better that channel assignment or better that chromosome will be. In what follows we will associate each chromosome with an individual. The function to be optimized, the throughput, will be referred to as the fitness function. The steps involved in the GA, in the context of the problem considered in this example, are outlined in Table I.

### 6.2.2 In the beginning

We start with the creation of a number of chromosomes or channel assignments using completely random selection. In other words, for each chromosome, each cell is assigned to a User who is selected randomly. We repeat this process $3 P$ times to create $3 P$ chromosome or individuals. Steps 2 to 5 in Table 6.1 are self explanatory.

### 6.2.3 The birth of children

In Step 6, a pool of 2P individuals consisting of the Elites and the Middles is randomly paired to form P couples. Let us call the individuals who form a couple Father and Mother. We split the Father's and Mother's chromosomes at a randomly chosen but identical location. We swap the chromosomes as shown in Fig. 6.1(b) to create two new chromosomes. In the

GA terminology, this is called one-point cross-over. In this way, each couple will create two offspring (Step 7). In total we will have 2P children.

### 6.2.4 Genetic mutation

In order to prevent the algorithm from converging to a local maximum, occasionally we mutate the chromosomes of the children at birth. This is accomplished by including Step 8 which occurs with a low probability. This step is included only if necessary.

### 6.2.5 The last generation

The algorithm continues until a predetermined number of $N$ gen generations. $N g e n$ is determined by trial and error in the first few runs of the program by observing the convergence of the throughput value. The chromosome of the individual with the highest fitness function value in the last generation represents the sub-channel allocation that leads to the near-optimum throughput value given by the fitness function.

### 6.2.6 Numerical results

Numerical results obtained using GA for the above example problem with $N=6$ users and $K=128$ subchannels, and using a population size $3 P=600$ are shown in Fig. 6.2. The dotted line represents the optimal throughput obtained by assigning each subchannel to the user with the highest gain. The near optimal solution obtained using the GA is within $3 \%$ of the optmal solution. Note that the solution from GA can be improved by using a larger population size.


Figure 6.2: Genetic algorithm with population size $3 \mathrm{P}=600$

### 6.3 Genetic algorithm with continuous variables

In case of an objective function with continuous variables, the real numbered values of the variables are represented using the binary notation. The 0's and 1's of the binary digits become the entries in the chromosome.

Example 6.3.1. Consider the problem of finding the absolute maximum value of $f(x, y)=$ $15-2(x-1.125)^{2}-3(y-2.300)^{2}$.

We know the maximum occurs at $x_{\mathrm{opt}}=1.125, y_{\mathrm{opt}}=2.300$ and it is 15 . We will demonstrate how good GA is in solving this problem.

There are two variables and hence there will be two chromosomes. Suppose we decided to represent the real number $x$ using 6 binary digits in the form $b_{6} b_{5} b_{4} \cdot b_{3} b_{2} b_{1}$. This onedimensional array of 6 cells will form one of the two chromosome. Each cell will have one of the binary digits, either a 0 or 1 . A similar array representing the variable $y$ in binary
notation will form the other chromosome. Note that the last three digits come after the "point." The largest number that can be represented is 7.875 and the smallest is 0.125 .

GA repeats the steps in Table 6.1 for each chromosome separately and in parallel. The fitness function will be the $f(x, y)$ given above.

### 6.3.1 Numerical results

The results from GA with a population size of $3 P=90$ is shown in Fig. 6.3. The throughput solution of 14.9925 from GA is practically the same as the optimal solution of 15 . Note that the near optimal $y_{\mathrm{GA}}=2.250$ value obtained by GA is the closest one can get to the optimal $y_{\text {opt }}=2.300$ when one is representing fractions using 3 binary digits. Naturally this can be improved by using more digits to represent a fraction. However, the near optimal value obtained, 14.9925 , is so close the optimal value 15 , there is no need for improvements.


Figure 6.3: Genetic algorithm with population size $3 \mathrm{P}=90$

## Chapter 7

## Game theory

### 7.1 Introduction

Game theory is simply the analysis of games in terms of strategy. We have all analyzed games starting from elementary school probably with tic-tac-toe. We came to the conclusion that if the player who starts first occupies the central node in tic-tac-toe, then he/she cannot loose. Analyzing games in sports this manner has limited use because the outcome of a sport game often depends also on other factors such as players's skill and psychology. What makes game theory such a valuable tool is that many conflicting situations that arise between organizations or nations can be modeled as a game. Each party involved in such a situation can analyze the game under the assumption that the opposing party chooses the best strategy available to it, and select the best strategy for itself. An example that is perhaps surprising is the application of game theory in the arbitration process between the management and the union. When the talks between the union and the management breaks down, the arbitrator models the "offers and counter-offers" by the two parties as a repeated game and calculates the theoretical outcome under the assumption that the negotiations go on forever [62]. The arbitrator then imposes the outcome of this infinitely-repeated-game on the parties involved.

Broadly speaking, games can be classified in two ways. One is into sequential games
and strategic games. Tic-tac-toe and chess are two examples of sequential games. Prisoner's Dilemma and Battle of the Sexes are two examples of strategic games that are often mentioned in books on game theory $[27,63,64]$. The second is to classify games based on their strategy spaces as games in discrete space or in continuous space. The strategy spaces of Prisoner's Dilemma and Battle of the Sexes are discrete. They can be used to model Social or Economic behavior. In contrast, almost every design decision an engineer makes is about a continuous variable that takes a real value. The strategy spaces of games useful for engineers are continuous.

### 7.2 Optimization and game theory

In a game, each side or user attempts to use all the resources available to it to maximize its chance of winning the game. This shows that "winning the game" has a connection to optimizing. But we see a complication immediately. Every game has at least two players who attempts to win. Every game involves at least two optimizing processes. No game in game theory can be used to optimize a single function. However, game theory is useful in an optimization process that involves several interconnected multi-variable functions, if the situation satisfies certain conditions. Certain type of interconnection and conditions enable one to model the problem as a strategic game. Another type of interconnection and conditions lead to a Stackleberg game. In Chapter 12, we use a composite game - a strategic game inside a Stackleberg game to decide the price of subchannels for an operator that subleases channels for ad-hoc networks.

### 7.3 Strategic games in continuous space

The strategic game and its central idea of a Nash equilibrium were discovered by John Nash during his doctoral work at Princeton university. For three of four years Nash did
not publish any papers, while among his classmates were a few who were very prolific in publishing. When taunted by these classmates for not publishing, Nash apparently retorted by saying that while their work was derivative research, he was waiting for that one big idea [115]. It turns out that the big idea was the Nash Equilibrium. In the following, we first describe a two player strategic game and then go on to multi-player strategic game and the definition of a Nash equilibrium.

Consider two individuals, $X$ and $Y$, playing a game. $X$ is supposed to pick a real number from the closed interval $[a, b]$. Without knowing what $X$ picked, $Y$ has to pick a real number from the interval $[c, d] .[a, b]$ is called the Strategy Space of $X$ and $[c, d]$ is called the Strategy Space of $Y$. The order in which they do their picking does not matter but it is important that they do this without the knowledge of the number picked by the other. Let us denote the number picked by $X$ by $x$ and the number picked by $Y$ by $y . x$ is called the Strategy of $X$ and $y$ is called the Strategy of $Y$. Each individual is now awarded a point based on his/her and the other player's pick. Suppose $X(x, y)$ is the function that decides the point awarded to $X$, and $Y(x, y)$ decides the point for $Y$. We will assume $X(x, y)$ and $Y(x, y)$ are continuous and bounded. These functions are fixed before the start of the game and are called Payoff (Utility) Functions: $X(x, y)$ is the payoff function for $X$ and $Y(x, y)$ is the payoff function for $Y$. Before the start of the game each player not only has the knowledge of his/her payoff function but also has knowledge of the other player's payoff function. The game is played only once and the player who gets the most points wins. Assuming each player wants to win and is capable of reasoning, can we predict the outcome of the game? In other words, can we predict the numbers the players would choose?

It is clear that $X$ will try to pick the $x$ that maximizes $X(x, y)$ and $Y$ will try to pick the $y$ that maximizes $Y(x, y)$. But the $x$ that maximizes $X(x, y)$, in general, will depend on $y$, which is decided by $Y$ alone; $X$ has no knowledge or control over this $y$. Suppose that $X$ knows that for sure $Y$ will play $y_{1}$. Then $X$ will pick the $x$ that maximizes $X\left(x, y_{1}\right)$. In other words, $X$ will pick $x_{y_{1}}=\arg \max _{x} X\left(x, y_{1}\right)$. If $X$ thinks that for sure $Y$ will play $y_{2}$,
then $X$ will pick $x_{y_{2}}=\arg \max _{x} X\left(x, y_{2}\right)$. This dependence of $X$ 's pick of $x$ on the choice made by $Y$ can be described by a function called the best response function of $X$ :

$$
\begin{equation*}
B R_{X}(y)=\arg \max _{x} X(x, y) \tag{7.1}
\end{equation*}
$$

Similarly if $Y$ knows for sure that $X$ will play $x_{1}$, then $Y$ will pick the $y_{x_{1}}=\arg \max _{y} X\left(x_{1}, y\right)$, and so on. The best response function of $Y$ will be given by:

$$
\begin{equation*}
B R_{Y}(x)=\arg \max _{y} Y(x, y) \tag{7.2}
\end{equation*}
$$

Suppose that there are two numbers $x^{*}$ and $y^{*}\left(x^{*} \in[a, b]\right.$ and $\left.y^{*} \in[c, d]\right)$ such that the $x$ that maximizes $X\left(x, y^{*}\right)$ is $x^{*}$ and the $y$ that maximizes $Y\left(x^{*}, y\right)$ is $y^{*}$. Suppose further that there is only one such pair of numbers $\left(x^{*}, y^{*}\right)$. Since each player has knowledge of both player's payoff functions, and is capable of reasoning, we can conclude that the outcome of the game would be $\left(x^{*}, y^{*}\right)$. Are the players satisfied with this outcome? How happy is each player? These questions will be addressed in Section 7.3.3.

What if there are more than one such pairs of points? Before we discuss this, we will tighten our argument by replacing the word maximum by local maximum. Suppose that there are two such pairs $\left(x_{1}^{*}, y_{1}^{*}\right)$ and $\left(x_{2}^{*}, y_{2}^{*}\right)$, such that $X\left(x_{1}^{*}, y_{1}^{*}\right)$ is a local maximum of $X\left(x, y_{1}^{*}\right)$ and $Y\left(x_{1}^{*}, y_{1}^{*}\right)$ is a local maximum of $Y\left(x_{1}^{*}, y\right)$, and like wise $X\left(x_{2}^{*}, y_{2}^{*}\right)$ is a local maximum of $X\left(x, y_{2}^{*}\right)$ and $Y\left(x_{2}^{*}, y_{2}^{*}\right)$ is a local maximum of $Y\left(x_{2}^{*}, y\right)$. Predicting the outcome might not be possible now.

In any event, points like $\left(x_{1}^{*}, y_{1}^{*}\right)$ and $\left(x_{2}^{*}, y_{2}^{*}\right)$ are called Nash equilibria. Suppose that in one instant of playing the above game the players picked $\left(x_{2}^{*}, y_{2}^{*}\right)$. Now imagine we bend the rules and say that while $Y$ is not allowed to change his mind, we give $X$ a chance to pick another number if she chooses to. Is their any motivation for $X$ to move away from $x_{2}^{*}$ ? The answer is no, because for $y$ fixed at $y_{2}, x_{2}$ maximizes $f(x, y)$ and their is no point in deviating from this. This illustrates one of the important characteristics of a Nash equilibrium. If for
some reason the players find themselves selecting strategies that produces a Nash equilibrium, then no player has the motive to move away from that strategy unilaterally, as long as all the other players choose not to deviate from their strategies that lead to that particular Nash equilibrium.

### 7.3.1 Nash equilibrium

Based on our discussion in the last section, it is not hard to motivate a definition of Nash equilibrium for two player games. For a strategic game with closed and bounded Euclidean strategy spaces $X$ and $Y$, and continuous payoff functions $X(x, y)$ and $Y(x, y)$, respectively, a point $\left(x^{*}, y^{*}\right)$ is a Nash equilibrium if and only if $X\left(x^{*}, y^{*}\right)$ is a local maximum of $X\left(x, y^{*}\right)$ for $x \in X$ and $Y\left(x^{*}, y^{*}\right)$ is a local maximum of $Y\left(x^{*}, y\right)$ for $y \in Y$.

Now consider an $N$ player game where we denote the closed and bounded Euclidean strategy space of the $i$ th player by $P_{i}$, and the continuous payoff function for the $i$ th player by $u_{i}(\mathbf{p})$. Here, $\mathbf{p}=\left[p_{1}, p_{2}, \ldots p_{N}\right]$. It is convenient to have a notation for the vector that contains the strategies of all players except the $i$ th player. Let $\mathbf{p}_{-\mathbf{i}}=\left[p_{1}, p_{2}, \ldots p_{i-1}, p_{i+1}, \ldots p_{N}\right]$. This allows us to rewrite $u_{i}(\mathbf{p})$ as $u_{i}\left(p_{i}, \mathbf{p}_{-\mathbf{i}}\right)$ when we prefer to.

Definition 7.3.1. A point $\mathbf{p}^{*}$ is a Nash equilibrium if and only if $u_{i}\left(p_{i}^{*}, \mathbf{p}_{-\mathbf{i}}^{*}\right)$ is a local maximum of $u_{i}\left(p_{i}, \mathbf{p}_{-\mathbf{i}}^{*}\right)$ for $p_{i} \in P_{i}$, for each $i$.

A natural question now would be: Do all games have Nash equilibria? Or more relevantly, do all payoff functions produce Nash equilibria?

### 7.3.2 Existence of Nash equilibrium

Well established theorems on the existence of the Nash equilibrium for continuous strategy spaces require some sort of convexity structure for the the payoff functions. It may be possible to loosen these requirements, but we cannot get rid of them altogether [65]. The
most well known of these theorems also requires the strategy spaces to be compact. In Euclidean spaces, compactness is equivalent to closed and boundedness.

Theorem 7.3.1. (Debreu, Glicksberg, and Fan) A game with closed and bounded strategy spaces $P_{i}$ and continuous payoff functions $u_{i}\left(\mathbf{p}_{-\mathbf{i}}, p_{i}\right)$ has a Nash equilibrium if each $u_{i}\left(\mathbf{p}_{-\mathbf{i}}, p_{i}\right)$ is quasi concave in $p_{i}$ [64].

One of the theorems that accepts a loosened form of structure requires super-modularity of the payoff function.

Definition 7.3.2. A twice differentiable function $f\left(x_{1}, x_{2}, \ldots, x_{n}\right): R^{n} \rightarrow R$ is said to be super-modular if the mixed partial derivatives $\frac{\partial f^{2}}{\partial x_{i} \partial x_{j}} \geq 0$ for all $i \neq j$.

Definition 7.3.3. A game with super-modular payoff functions is called a super-modular game.

Theorem 7.3.2. Every super-modular game has a Nash equilibrium.
It is important to note that all the theorems about the existence of the Nash equilibrium deal with sufficient conditions. These conditions are not necessary for the existence of the Nash equilibrium.

### 7.3.3 Optimality of Nash equilibrium

Going back to the two player game, assume that the payoff functions are such that there was only one Nash equilibrium $\left(x^{*}, y^{*}\right)$. The payoff for player $X$ would then be $X\left(x^{*}, y^{*}\right)$. We know that $X\left(x^{*}, y^{*}\right)$ is the maximum of $X\left(x, y^{*}\right)$; But $X\left(x^{*}, y^{*}\right)$ is not the maximum of $X(x, y)$. If for some reason the game has ended in another outcome ( $\mathrm{x}, \mathrm{y}$ ), payoff $X(x, y)$ for $X$ could possibly be higher than $X\left(x^{*}, y^{*}\right)$. This shows that Nash equilibrium has no relation to the maximum values of individual payoff functions.

Consider two arbitrary points $A=\left(x_{1}, y_{1}\right), B=\left(x_{2}, y_{2}\right)$. If the payoff at $A$ is better than the payoff at $B$, for both players, that is, $X(A) \geq X(B)$ and $Y(A) \geq Y(B)$, we say point $A$

Pareto Dominates point $B$. A point $\left(x_{p}, y_{p}\right)$ that Pareto Dominates every other point $(x, y)$, is said to be Pareto Optimal.

### 7.4 Stackelberg game

The main deference between the strategic game and the Stackelberg game is that in the latter, one of the players (the leader) gets to decide the value of his variable first. All the others choose the values of their variables simultaneously, but after the leader has decided on his variable. Unlike the strategic game and its idea of Nash equilibrium, the mathematics involved in the Stackelberg game is quite simple. In the following, we will describe a two player Stackelberg game in detail.

### 7.4.1 Two player Stackelberg game

Consider two players $A$ and $B$. Each player has a function associated with it called a payoff (utility) function. Both payoff functions are functions of two variables $a$ and $b$. Let us denote the payoff function of $A$ by $A(a, b)$ and that of $B$ by $B(a, b)$. Each player not only knows its payoff function but has knowledge of the other player's payoff function. Variable $a$ is controlled only by player $A$, and variable $b$ is decided only by $B$. The objective of the game for each player is to maximize its payoff function. Which player gets to decide the value of its variable first is decided before the start of the game and that player is called the leader.

Let us assume $A$ is the leader. Suppose that $A$ chooses the value $a_{1}$ for its variable $a$. Then the follower, $B$, can choose the $b$ that maximizes $B\left(a_{1}, b\right)$. That is, the follower can choose $b_{1}=\arg \max _{b} B\left(a_{1}, b\right) . b_{1}$ is called the best response of $B$ to the choice $a_{1}$ by $A$. If instead, $A$ chooses $a_{2}, B$ can choose $b_{2}=\arg \max _{b} B\left(a_{2}, b\right)$. This dependency of the follower's choice for $b$ in the leader's choice for $a$ can be described by what is called the best response function of $B$. For notational simplicity, let us denote this best response function by $b=f(a)$.

Since the leader has knowledge of the follower's utility function $B(a, b)$, the leader can compute the above mentioned best response function $b=f(a)$. The leader can then substitute this function into its own payoff function to arrive at $A(a, f(a))$ which depends only on its own variable $a$. It can now find the $a$ that maximizes $A(a, f(a))$ and announce $a^{*}=\arg \max _{a} A(a, f(a))$ as its choice. Given this, $B$ substitutes this $a^{*}$ into $b=f(a)$ (just like $A$ anticipated) to arrive at its best choice $b^{*}=f\left(a^{*}\right)$ for its variable $b$. The point $\left(a^{*}, b^{*}\right)$ is called the Stackelberg equilibrium of the game.

If the payoff functions $A(a, b)$ and $B(a, b)$ are continuous and the variables $a$ and $b$ are defined on a closed interval, then extreme value theorem guarantees that $A(a, b$ and $B(a, b)$ will posses maximum values. If in addition, the functions $f(a)$ and $A(a, f(a))$ are sufficiently smooth, there are theorems that peg the existence of the Stackelberg Equilibrium on the concavity of these functions. Note, however, that the conditions imposed by these theorems are sufficient but they are not necessary for the existence of the Stackelberg equilibrium.

## Part II

## Applications and Algorithms

## Chapter 8

## Throughput maximization in the GIC with proportional rate constraints

### 8.1 Introduction

This chapter presents the solution to the optimization problem of finding the power allocation that maximizes the throughput (sum rate) of the Gaussian interference channel with any linear power constraint and proportional rate constraints. It is proved that the sumrate of the Gaussian interference channel restricted to proportional rate constraints does not have a critical point and the maximum sum-rate subject to said constraints occurs at the boundary of the domain formed by the plane representing the linear power constraint. This is accomplished by using analytic geometry in higher dimensions to show that the curve of intersection of the sum-rate and the proportional rate constraints is always increasing, and intersects the boundary plane representing the linear power constraint at a unique point. A polynomial time (in the number of users) centralized algorithm that finds this point of optimal power allocation is proposed. This is a significant improvement over existing algorithms for related power allocation problems which have exponential time complexity in the number of users. Two distributed algorithms with linear and constant complexities are also
presented. Simulation results supporting the analysis and demonstrating the performances of the algorithms are presented.

### 8.2 Related work

Gaussian interference channel (GIC) has been used to model the uplinks of code-division multiple access (CDMA) systems, digital subscriber line (DSL) systems, and more recently, ad-hoc networks, and small cell networks such as femtocell networks. Almost 50 years after the first investigation by Shannon [98], an exact description of the capacity region of the interference channel that covers all values of the channel parameters continues to elude us [98-101]. Researchers concerned with power allocation that maximizes the system capacity have sidestepped this difficulty by optimizing what is called the sum-rate. This sumrate is obtained by applying Shannon's original formula for capacity to each user separately while considering the interference from the other users as noise. Even then, the problem of finding the power allocation that maximizes the sum-rate has remained a difficult problem.

The early 1990's saw the arrival of the 3G cellular systems, and the transmission technology for the 3G in North America was Code division multiple access (CDMA). As indicated in Chapter 1, the channel model for the CDMA uplink is a special case of the GIC. Consequently, a vast number of research related to the problem undertaken in this chapter came from work on CDMA. It should be noted that while any work on power allocation in GIC is directly applicable to CDMA, any work on CDMA would need considerable extension before it can be applied to GIC.

The early works in power allocation on CDMA, focused on each user achieving its target SINR and offered distributed solutions based on fixed point algorithms [102-106]. More recently, strategic game theory has been used to find a competitive equilibrium among users who attempt to maximize their utility functions [36-61]. A user's utility function typically includes its transmission rate and a cost function. The strategy space of a user is the range
of power available to it. This meant that game theory approaches used individual power constraints. It is important to note that the sum utility or the sum-rate is not maximized in the game theory based solutions. Game theory also offers distributed solutions based on fixed point algorithms or their variants.

In light of the big body of research in this area [36-61], we give a review of this work in Appendix A.

The difficulty in determining the power allocation that maximizes the sum-rate of the GIC arises due to the fact that the convex structure of the sum-rate is unknown. The only theoretical insight into this problem we have is that the power allocation that maximizes the sum-rate under a total power constraint occurs at the boundary plane formed by the power constraint $[107,108]$. For the problem of maximizing the sum-rate with any fairness or quality-of-service (QoS) constraints, no studies have been reported that exploits the structure of the feasible set. However, many search algorithms have been proposed to determine the optimal power allocation under various constraints. Oh et al. [29] model the uplink of a CDMA based cellular system as a GIC and consider the sum-rate maximization problem under the minimum individual SINR, total interference, and individual power constraints. They, then, propose a systematic search that finds the optimal power allocation in $O\left(t^{N}\right)$ computations, where $N$ is the number of mobile users and $t$ is the number of points to be checked for optimality. Abadpour et al. [30] report that the technique proposed by Oh et al. often produces a power allocation that is unfair to some users. To rectify this Abadapour et al. introduce maximum individual SINR constraints and modify Oh et al.'s algorithm.

Dai et al. [33] consider the problem of maximizing the minimum uplink rate of mobile users with individual power constraints as well as minimum and maximum rate constraints. They use Majorization theory to reduce this optimization problem into a search problem for a real number in a closed interval. Gjendemsj et al. [34] propose a suboptimal solution to the power allocation problem that maximizes the system throughput with individual power constraints that is based on an extension of the solution of the two-user case. Qian et
al. $[31,32]$ consider the problem of maximizing the weighted sum-rate subject to individual minimum rate and power constraints. They transform this problem into multiplicative linear fractional programming, and propose an algorithm that constructs a sequence of polyblocks of decreasing size and searches their vertexes for the optimal solution. Chiang et al. [109, 110] approximate the Shannon's formula $\log _{2}(1+$ SINR $)$ with $\log _{2}($ SINR $)$ to transform the throughput maximization problem into a geometric program.

If our goal is to maximize the system throughput with a total power constraint, then SINR balancing and game theory cannot be used. The other existing algorithms also work with individual power constraints. Total power constraint is important in practical wireless systems to limit the interference to the neighboring systems [111]. Dividing a total power constraint into equal individual power constraints cannot be efficient. None of the algorithms that maximize the sum-rate uses any knowledge of the structure of the optimization problem. Because of this they fall into to the category of search algorithms which are computationally expensive.

### 8.3 The proposed approach

This chapter takes an analytical approach to obtain the solution of the power allocation problem that maximizes the sum-rate of the GIC under any linear power constraint and proportional rate constraints. The goal is to produce an algorithm with a complexity that is practical (polynomial). Proportional rate constraints have been used in resource allocation problems in OFDMA before [95,112-114]. Proportional rate constraints can be mapped to proportional delay constraints (in case of saturated traffic scenarios). Furthermore, since existing (minimum rate constraint) algorithms have exponential complexity, a polynomial time solution with a slightly different QoS constraint can be useful. For example, suppose we have a situation where we have to satisfy minimum rate demands and the demands are feasible. The optimum solution using the existing work takes exponential time. However, we
can take the ratios of the minimum rates and use our proportional rate algorithm to get a near optimal solution in polynomial time. This is worthwhile because the difference between exponential and polynomial time is extremely large.

The optimization problem undertaken in this chapter is challenging because of two reasons. The first is that the objective function is not concave. This excludes many conventional methods from convex analysis such as dual methods and KKT conditions [28]. This difficulty is further exacerbated by the second reason that the proportional rate constraints are nonlinear. Because of this non-linearity, the maximum sum-rate subject to the proportional rate and linear power constraints could occur at any point in the domain and not necessarily on the power constraint plane (which is the case with minimum rate constraints). Furthermore, in general, the curve of intersection of the proportional rate constraints and the sum could intersect the power constraint plane at more than one point, as illustrated by the following example.

Example 8.3.1. Consider the sum $f=f_{1}+f_{2}$ of two functions $f_{1}(x, y)=2 x^{2}+2 x y-2 x+3 / 4$ and $f_{2}(x, y)=y^{2}+2 x y-4 y+4$. Suppose the proportional constraint we are interested is 1:1, and the linear constraint $x+y=2$.

Figures 8.1-8.3 show that the curve of intersection of the proportional constraint $f_{1}: f_{2}=$ 1:1 and the surface $z=f_{1}+f_{2}$ of the sum intersects the linear constraint plane $x+y=2$ at two points.

Analytically, the equations $f_{1}=f_{2}$ and $x+y=2$ can be simultaneously solved to obtain the $(x, y)$ coordinates of these two points. They turn out to be $(1 / 2,3 / 2)$ and $(3 / 2,1 / 2)$.

In the GIC however, we prove that the curve of intersection of the proportional rate constraints and the sum-rate is always increasing and intersects the power constraint plane at a unique point. Therefore this point is the maximum.

The contributions of this chapter are as follows:

1. The proof that the power allocation that maximizes the sum-rate under the propor-


Figure 8.1: The locus of $f_{1}: f_{2}=1: 1$.


Figure 8.2: The curve $\mathbf{C}$ of intersection of $f_{1}: f_{2}=1: 1$ and the sum.


Figure 8.3: Intersection of $\mathbf{C}$ and $x+y=2$.
tional rate and any linear power constraints occurs at a unique point on the power constraint plane.
2. A fast, simple, and stable centralized algorithm that finds this point in polynomial time (specifically, the time complexity of the algorithm is $O\left(N^{3}\right)$ ).
3. A distributed algorithm that converges to this point with linear time complexity, but does not require any communication between the users, or a user and a central controller.
4. A distributed algorithm with constant time complexity that conforms to the Shared Memory model [116] in distributed decision making.

The rest of the chapter is organized as follows. Section 8.4 presents the system model and the problem formulation. Analysis of the optimization problem and the summary of the analysis can be found in Section 8.5. A centralized algorithm that solves the optimization
problem is presented in Section 8.6, and two distributed algorithms are developed in Section 8.7. Numerical results obtained through simulations are provided in Section 8.8, followed by conclusion in Section 8.9.


Figure 8.4: The $N$-user Gaussian interference channel.

### 8.4 System model and the problem statement

Consider the $N$-user GIC shown in Fig. 8.4. User $i$ employs transmitter $i$ to communicate with receiver $i$ but receiver $i$ experiences interference from all other transmitters. $h_{i j}$ denotes the channel gain between transmitter $j$ and receiver $i$, and $p_{i}$ denotes the power used by transmitter $i$. The channel gains are assumed to remain constant in the time period in which the power allocation algorithm is applied. The transmission rate $R_{i}$ of user $i$ is given by


Figure 8.5: An increasing space curve.


Figure 8.6: A decreasing space curve.

$$
\begin{equation*}
R_{i}=\log _{2}\left(1+\frac{h_{i i} p_{i}}{\sigma_{i}^{2}+\sum_{j \neq i} h_{i j} p_{j}}\right) \tag{8.1}
\end{equation*}
$$

where $\sigma_{i}^{2}$ stands for the received additive white Gaussian noise power per Hz. Letting $N_{i}=\sigma_{i}^{2} / h_{i i}$ and $a_{i j}=h_{i j} / h_{i i}$,

$$
\begin{equation*}
R_{i}=\log _{2}\left(1+\frac{p_{i}}{N_{i}+\sum_{j \neq i} a_{i j} p_{j}}\right) . \tag{8.2}
\end{equation*}
$$

The sum-rate $R$ of the $N$-user GIC is

$$
\begin{equation*}
R=\sum_{i=1}^{N} R_{i} . \tag{8.3}
\end{equation*}
$$

A summary of the notations is given in Table 8.1.
Our objective is to solve the optimization problem that determines the power allocation that maximizes the sum-rate $R$ subject to two constraints. The first one is a linear constraint
in the transmit powers as follows:

$$
\begin{equation*}
\sum_{i=1}^{N} g_{i} p_{i} \leq P \tag{8.4}
\end{equation*}
$$

This constraint could arise, for example, as an interference constraint in ad-hoc or sensor networks. $g_{i}$ is the channel coefficient from transmitter $i$ to the interference measuring point. $P$ represents the acceptable interference threshold. If all the $g_{i}$ 's are equal to 1 , then this could represent a total power constraint.

The second constraint deals with proportional transmission rates

$$
\begin{equation*}
R_{1}: R_{2}: R_{3}: \ldots R_{N}=\beta_{1}: \beta_{2}: \beta_{3}: \ldots \beta_{N} \tag{8.5}
\end{equation*}
$$

where $\beta_{i}$ 's are non-zero positive real numbers. Letting $\alpha_{i}=\beta_{i} / \beta_{1}$ for $i=2,3,4, \ldots, N$, the latter can be re-written as the following $N-1$ equations:

$$
\begin{equation*}
\alpha_{i} R_{1}=R_{i} \text { for } i=2,3,4, \ldots, N \tag{8.6}
\end{equation*}
$$

| $N$ | number of users in the system |
| :---: | :---: |
| $h_{i j}$ | the channel gain between transmitter $j$ and receiver $i$ |
| $p_{i}$ | power used by transmitter $i$ |
| $R_{i}$ | transmission rate of User- $i$ |
| $\sigma_{i}^{2}$ | additive white Gaussian noise density |
| $N_{i}$ | $\sigma_{i}^{2} / h_{i i}$ |
| $a_{i j}$ | $h_{i j} / h_{i i}$ |
| $R$ | sum rate or throughput |
| $g_{i}$ | channel gain from transmitter $i$ to the interference measuring point |
| $\beta_{i}$ | proportional rate of User- $i$ |
| $\alpha_{i}$ | $\beta_{i} / \beta_{1}$ |

Table 8.1: Notations

Often in practice, there may be a total power as well as an interference constraint. This however, does not mean that we have to formulate the optimization problem with both
constraints. Suppose the solution to the problem with total power constraint is $P^{*}$ and the solution to the problem with the interference constraint is $P^{* *}$. Then as we prove in the next section, the solution to the problem with both constraints is the one among these two points that is closest to the origin.

### 8.5 Analysis of the optimization problem

In this section, we interpret the objective and the constraints of our optimization problem as hyper surfaces in the $N+1$-dimensional Euclidean space and use analytic geometry in higher dimensions to prove that the solution to the optimization problem exists, is unique, and it occurs on the hyper plane formed by the linear power constraint.

### 8.5.1 Objective and constraints as hyper surfaces

Consider first the two-user case and the 3-dimensional Euclidean space with the Cartesian system of coordinates. We can plot $p_{1}$ and $p_{2}$ along the first two axes and the sum-rate $R\left(p_{1}, p_{2}\right)$ along the third axis. In this way, $R=R\left(p_{1}, p_{2}\right)$ will form a 2-dimensional surface in the 3 -dimensional space. In this case, there will be only one proportional rate constraint, $\alpha_{2} R_{1}\left(p_{1}, p_{2}\right)-R_{2}\left(p_{1}, p_{2}\right)=0$, and it will also form a 2-dimensional surface in the 3 -dimensional space. Since the third variable $R$ is absent from the rate constraint equation, this latter surface will rise parallel to the $R$ axis and will intersect the surface formed by the sum-rate $R\left(p_{1}, p_{2}\right)$ in a space curve. In summary, the intersection of two 2-dimensional surfaces in the 3-dimensional space forms a space curve.

In the case of three users, we need 4 dimensions. The sum-rate $R\left(p_{1}, p_{2}, p_{3}\right)$ will form a 3-dimensional hyper surface in the 4 -dimensional space. There will be two rate constraints, each forming a 3-dimensional surface. The intersection of the three 3-dimensional surfaces in the 4-dimensional space will form a hyper space curve.

Consider now the $N$-user case and the $N+1$ dimensional space. We can plot the powers
$p_{i}$ 's along the first $N$ axes and the sum-rate $R\left(p_{1}, p_{2}, \ldots, p_{N}\right)$ along the last axis. In this way, in the $N+1$ dimensional space, $R$ will form an $N$-dimensional hyper surface. Each of the $N-1$ proportional rate constraints in (8.6) will form a $N$-dimensional surface. The intersection of all of these surfaces - the sum-rate and the constraints - a total of $N, N$-dimensional surfaces in the $N+1$-dimensional space will form a hyper space curve $\mathbf{C}$.

### 8.5.2 Methodology of the analysis

We first wish to prove that this space curve $\mathbf{C}$ is always increasing, and therefore, the maximum sum-rate restricted to this curve can only occur on the boundary plane $\sum_{i} g_{i} p_{i}=$ $P$. We then prove that this curve $\mathbf{C}$ indeed intersects the boundary hyper plane $\sum_{i} g_{i} p_{i}=P$, and that the point of intersection is unique.

A single-variable function $f: R \longrightarrow R$ is said to be increasing if $f(a)>f(b)$ whenever $a>b$. We say a space curve $\mathbf{c}(t)$ in $N$-dimensions defined by the parametric form $\left\{x_{1}(t), x_{2}(t), \ldots, x_{n}(t)\right\}$ is increasing if the single variable function $x_{n}(t)$ is increasing. Note that it immediately follows that a space curve $\mathbf{c}(t)$ is increasing if the $N$-th dimensional component $x_{n}^{\prime}(t)$ of its tangent vector $\mathbf{c}^{\prime}(t)$ is positive. Two examples of space curves in 3 -dimensions, one increasing and the other decreasing are shown in Fig. 8.5 and Fig. 8.6, respectively.

The goal is to prove that the space curve $\mathbf{C}$ is increasing by finding its tangent vector. Since C is the intersection of $N, N$-dimensional surfaces, a tangent vector to $\mathbf{C}$ can be obtained by the cross product of the $N$ normal vectors of the surfaces. For this purpose, we will use an extension of the familiar cross product of two vectors in three dimensions. An outline of this cross product in higher dimensions was given in Chapter 3. The orientation of the tangent vector obtained this way would depend on the way we orient the $N$ normal vectors, and the order we cross or write them in the determinant. This is further complicated by the fact that when we evaluate the determinant we need to consider whether $N$ is odd or even. We circumvent these difficulties by first concentrating on the magnitude of the tangent
vector and then establishing the orientation using an indirect argument.

### 8.5.3 Normal vectors of the surfaces

We start by finding the normal vectors of the $N$ surfaces, namely, the sum-rate surface and the $N-1$ surfaces from the proportional rate constraints. This can be done by rewriting the equations (8.3) and (8.6) as follows:

$$
\begin{gather*}
F=\sum_{i=1}^{N} R_{i}-R=0,  \tag{8.7}\\
H_{i}=\alpha_{i} R_{1}-R_{i}=0 \text { for } i=2,3,4, \ldots, N \tag{8.8}
\end{gather*}
$$

and finding the gradients. Before proceeding, we introduce the notation $R_{i j}$ for the partial derivative of user rate $R_{i}$ with respect to power $p_{j}$. That is, $R_{i j}=\frac{\partial R_{i}}{\partial p_{j}}$. The normal vector of the sum-rate surface is

$$
\begin{equation*}
\nabla F=\left(\sum_{i} R_{i 1}, \sum_{i} R_{i 2}, \sum_{i} R_{i 3}, \ldots, \sum_{i} R_{i N},-1\right) \tag{8.9}
\end{equation*}
$$

The normal vectors of the $N-1$ surfaces in (8.8) are

$$
\begin{gather*}
\nabla H_{i}=\left(\alpha_{i} R_{11}-R_{i 1}, \alpha_{i} R_{12}-R_{i 2}, \ldots, \alpha_{i} R_{1 N}-R_{i N}, 0\right)  \tag{8.10}\\
\text { for } i=2,3,4, \ldots, N
\end{gather*}
$$

### 8.5.4 Tangent vector to curve C and its last component

Crossing the $N$ vectors in (8.9) and (8.10), we obtain the tangent vector $\mathbf{T}$ of the curve of intersection $\mathbf{C}$ of the sum-rate and the rate constraints, shown in (8.11), where $\theta_{1}, \theta_{2}, \theta_{3}, \ldots, \theta_{N}, \theta_{N+1}$
are the unit vectors along the $N+1$ coordinate directions.

$$
\mathbf{T}=\left|\begin{array}{ccccc}
\theta_{1} & \theta_{2} & \ldots & \theta_{N} & \theta_{N+1}  \tag{8.11}\\
\sum_{i} R_{i 1} & \sum_{i} R_{i 2} & \ldots & \sum_{i} R_{i N} & -1 \\
\alpha_{2} R_{11}-R_{21} & \alpha_{2} R_{12}-R_{22} & \ldots & \alpha_{2} R_{1 N}-R_{2 N} & 0 \\
\alpha_{3} R_{11}-R_{31} & \alpha_{3} R_{12}-R_{32} & \ldots & \alpha_{3} R_{1 N}-R_{3 N} & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\alpha_{N} R_{11}-R_{N 1} & \alpha_{N} R_{12}-R_{N 2} & \ldots & \alpha_{N} R_{1 N}-R_{N N} & 0
\end{array}\right| .
$$

The $N+1$-dimensional component of $\mathbf{T}$, which we denote by $t_{N+1}$, is shown in (8.12).

$$
t_{N+1}= \pm\left|\begin{array}{cccc}
\sum_{i} R_{i 1} & \sum_{i} R_{i 2} & \ldots & \sum_{i} R_{i N}  \tag{8.12}\\
\alpha_{2} R_{11}-R_{21} & \alpha_{2} R_{12}-R_{22} & \ldots & \alpha_{2} R_{1 N}-R_{2 N} \\
\alpha_{3} R_{11}-R_{31} & \alpha_{3} R_{12}-R_{32} & \ldots & \alpha_{3} R_{1 N}-R_{3 N} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{N} R_{11}-R_{N 1} & \alpha_{N} R_{12}-R_{N 2} & \ldots & \alpha_{N} R_{1 N}-R_{N N}
\end{array}\right|= \pm\left|\mathbf{M}_{0}\right| .
$$

The determinant in (8.12) can be shown to be row equivalent to the determinant below.

$$
\left|\begin{array}{cccc}
R_{11} & R_{12} & R_{13} & \ldots R_{1 N}  \tag{8.13}\\
R_{21} & R_{22} & R_{23} & \ldots R_{2 N} \\
R_{31} & R_{32} & R_{33} & \ldots R_{3 N} \\
\vdots & \vdots & \vdots & \ddots . \vdots \\
R_{N 1} & R_{N 2} & R_{N 3} & \ldots R_{N N}
\end{array}\right|
$$

From (8.2), the partial derivatives are calculated as

$$
R_{i j}= \begin{cases}\frac{1}{p_{i}+I_{i}}, & \text { if } j=i  \tag{8.14}\\ -\frac{a_{i j} p_{i}}{\left(p_{i}+I_{i}\right)\left(I_{i}\right)}, & \text { if } j \neq i\end{cases}
$$

where

$$
\begin{equation*}
I_{i}=N_{i}+\sum_{j \neq i} a_{i j} p_{j} \tag{8.15}
\end{equation*}
$$

Substituting in (8.13), we obtain:

$$
\left\lvert\, \begin{array}{ccccc}
\frac{1}{p_{1}+I_{1}} & \frac{-a_{12} p_{1}}{I_{1}\left(p_{1}+I_{1}\right)} & \frac{-a_{13} p_{1}}{I_{1}\left(p_{1}+I_{1}\right)} & \cdots & \frac{-a_{1 N} p_{1}}{I_{1}\left(p_{1}+I_{1}\right)}  \tag{8.16}\\
\frac{-a_{21} p_{2}}{I_{2}\left(p_{2}+I_{2}\right)} & \frac{1}{p_{2}+I_{2}} & \frac{-a_{23} p_{2}}{I_{2}\left(p_{2}+I_{2}\right)} & \cdots & \frac{-a_{2 N} p_{2}}{I_{2}\left(p_{2}+I_{2}\right)} \\
\frac{-a_{31} p_{3}}{I_{3}\left(p_{3}+I_{3}\right)} & \frac{-a_{32} p_{3}}{I_{3}\left(p_{3}+I_{3}\right)} & \frac{1}{p_{3}+I_{3}} & \cdots & \frac{-a_{3 N} p_{3}}{I_{3}\left(p_{3}+I_{3}\right)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{-a_{N 1} p_{N}}{I_{N}\left(p_{N}+I_{N}\right)} & \frac{-a_{N 2} p_{N}}{I_{N}\left(p_{N}+I_{N}\right)} & \frac{-a_{N 3} p_{N}}{I_{N}\left(p_{N}+I_{N}\right)} & \cdots & \frac{1}{p_{N}+I_{N}}
\end{array} .\right.
$$

At this point, we emphasize that none of the powers $p_{i}$ 's can be zero, for if any of the $p_{i}$ 's is zero, then that user's rate would be zero, and that would violate our condition that none of the $\beta_{i}$ is zero in (8.5). By using a series of row and column operations, it can be then shown that the matrix of the determinant above is equivalent to the matrix $\mathbf{M}$ below:

$$
\mathbf{M}=\left[\begin{array}{ccccc}
I_{1} & -a_{12} p_{2} & -a_{13} p_{3} & \cdots & -a_{1 N} p_{N}  \tag{8.17}\\
-a_{21} p_{1} & I_{2} & -a_{23} p_{3} & \cdots & -a_{2 N} p_{N} \\
-a_{31} p_{1} & -a_{32} p_{2} & I_{3} & \cdots & -a_{3 N} p_{N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
-a_{N 1} p_{1} & -a_{N 2} p_{2} & -a_{N 3} p_{3} & \cdots & I_{N}
\end{array}\right] .
$$

From (8.15), since $I_{i}=N_{i}+\sum_{j \neq i} a_{i j} p_{j}, I_{i}>\sum_{j \neq i} a_{i j} p_{j}, \quad \forall i$. This shows that the matrix M satisfies the condition

$$
\begin{equation*}
\left|m_{i i}\right|>\sum_{j \neq i}\left|m_{i j}\right|, \quad \forall i \tag{8.18}
\end{equation*}
$$

and therefore, is diagonally dominant. The determinant of a diagonally dominant matrix cannot be zero [117]. Hence $\operatorname{det}(\mathbf{M}) \neq 0$, and by extension, $t_{N+1} \neq 0$. Since $t_{N+1}$ is clearly continuous in $p_{i}$ 's, by intermediate value theorem, $t_{N+1}$ must be either always positive or
always negative.

### 8.5.5 Curve C is increasing

We now consider the special case where $\beta_{i}=1$ for all $i$ and the users all experience the same channel conditions. That is, $a_{i k}=a_{j k}$ for all $i, j, k$. It is clear that in this case the rate constraint (8.5) would imply that the powers are all equal. Letting $p_{i}=p$ for all $i$ in (8.2),

$$
\begin{equation*}
\frac{d R_{i}}{d p}=\frac{N_{i}}{\left(N_{i}+p \sum_{j \neq i} a_{i j}\right)^{2}+p\left(N_{i}+p \sum_{j \neq i} a_{i j}\right)}>0 . \tag{8.19}
\end{equation*}
$$

Hence, $\frac{d R}{d p}>0$. This shows that the tangent line to the curve of intersection of the sum-rate and the rate constraint points in the direction of increasing $R$. In other words, $t_{N+1}>0$, in this special case. Combining this with the result that $t_{N+1}$ is either always positive or always negative, we conclude that $t_{N+1}$ is always positive.

The fact that $t_{N+1}$ is always positive implies that the curve of intersection of the sum-rate and the rate constraints is always increasing and the maximum sum-rate restricted to this curve $\mathbf{C}$ can only occur at the boundary plane $\sum_{i=1}^{N} g_{i} p_{i}=P$.

### 8.5.6 Curve C intersects the power constraint plane at a unique point

We now focus on showing that the curve $\mathbf{C}$ does indeed intersect the plane $\sum_{i=1}^{N} g_{i} p_{i}=P$, and that the point of intersection is unique. This is accomplished by proving that $\mathbf{C}$ never becomes parallel to $\sum_{i=1}^{N} g_{i} p_{i}=P$. Note that a normal vector to $\sum_{i=1}^{N} g_{i} p_{i}=P$ is $\mathbf{n}=$ $\left(g_{1}, g_{2}, g_{3}, \ldots g_{N}, 0\right)$. For $\mathbf{C}$ to become parallel to $\sum_{i=1}^{N} g_{i} p_{i}=P$, the dot product T.n must vanish.

Writing T.n explicitly, in (8.20),

$$
\text { T.n }=\left|\begin{array}{cccccc}
g_{1} & g_{2} & g_{3} & \ldots & g_{N} & 0  \tag{8.20}\\
\sum_{i} R_{i 1} & \sum_{i} R_{i 2} & \sum_{i} R_{i 3} & \ldots & \sum_{i} R_{i N} & -1 \\
\alpha_{2} R_{11}-R_{21} & \alpha_{2} R_{12}-R_{22} & \alpha_{2} R_{13}-R_{23} & \ldots & \alpha_{2} R_{1 N}-R_{2 N} & 0 \\
\alpha_{3} R_{11}-R_{31} & \alpha_{3} R_{12}-R_{32} & \alpha_{3} R_{13}-R_{33} & \ldots & \alpha_{3} R_{1 N}-R_{3 N} & 0 \\
\alpha_{4} R_{11}-R_{41} & \alpha_{4} R_{12}-R_{42} & \alpha_{4} R_{13}-R_{43} & \ldots & \alpha_{4} R_{1 N}-R_{4 N} & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\alpha_{N} R_{11}-R_{N 1} & \alpha_{N} R_{12}-R_{N 2} & \alpha_{N} R_{13}-R_{N 3} & \ldots & \alpha_{N} R_{1 N}-R_{N N} & 0
\end{array}\right| .
$$

and expanding by the last column and letting $S_{i j}=\alpha_{i} R_{1 j}-R_{i j}$, we obtain

$$
\text { T.n }= \pm\left|\begin{array}{ccccc}
g_{1} & g_{2} & g_{3} & \ldots & g_{N}  \tag{8.21}\\
S_{21} & S_{22} & S_{23} & \ldots & S_{2 N} \\
S_{31} & S_{32} & S_{33} & \ldots & S_{3 N} \\
S_{41} & S_{42} & S_{43} & \ldots & S_{4 N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_{N 1} & S_{N 2} & S_{N 3} & \ldots & S_{N N}
\end{array}\right|
$$

Letting $\gamma_{j}=g_{j} / g_{1}$, taking away $\gamma_{j}$ times the first column from column $j$, and expanding by the first row, we arrive at (8.22).

$$
\mathbf{T . n}= \pm g_{1}\left|\begin{array}{cccc}
S_{22}-\gamma_{2} S_{21} & S_{23}-\gamma_{3} S_{21} & \ldots & S_{2 N}-\gamma_{N} S_{21}  \tag{8.22}\\
S_{32}-\gamma_{2} S_{31} & S_{33}-\gamma_{3} S_{31} & \ldots & S_{3 N}-\gamma_{N} S_{31} \\
S_{42}-\gamma_{2} S_{41} & S_{43}-\gamma_{3} S_{41} & \ldots & S_{4 N}-\gamma_{N} S_{41} \\
\vdots & \vdots & \ddots & \vdots \\
S_{N 2}-\gamma_{2} S_{N 1} & S_{N 3}-\gamma_{3} S_{N 1} & \ldots & S_{N N}-\gamma_{N} S_{N 1}
\end{array}\right|
$$

At this point we go back to (8.17) and observe that $\mathbf{M}$ is a $Z$-matrix that is diagonally
dominant. Therefore, $\mathbf{M}$ must be positive definite [118]. By extension, the matrix $\mathbf{M}_{\mathbf{0}}$ in the determinant that defined $t_{N+1}$ in (8.12) must be either positive or negative definite. Using our earlier shorthand $S_{i j}=\alpha_{i} R_{1 j}-R_{i j}$, matrix $\mathbf{M}_{\mathbf{0}}$ can be re-written as

$$
\mathbf{M}_{\mathbf{0}}=\left[\begin{array}{ccccc}
\sum_{i} R_{i 1} & \sum_{i} R_{i 2} & \sum_{i} R_{i 3} & \ldots & \sum_{i} R_{i N}  \tag{8.23}\\
S_{21} & S_{22} & S_{23} & \ldots & S_{2 N} \\
S_{31} & S_{32} & S_{33} & \ldots & S_{3 N} \\
S_{41} & S_{42} & S_{43} & \ldots & S_{4 N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_{N 1} & S_{N 2} & S_{N 3} & \ldots & S_{N N}
\end{array}\right] .
$$

This matrix can be shown to be equivalent to the matrix $\mathbf{M}_{\mathbf{2}}$ shown in (8.24).

$$
\mathbf{M}_{\mathbf{2}}=\left[\begin{array}{ccccc}
S_{22}-\gamma_{2} S_{21} & S_{23}-\gamma_{3} S_{21} & \ldots & S_{2 N}-\gamma_{N} S_{21} & S_{21}  \tag{8}\\
S_{32}-\gamma_{2} S_{31} & S_{33}-\gamma_{3} S_{31} & \ldots & S_{3 N}-\gamma_{N} S_{31} & S_{31} \\
S_{42}-\gamma_{2} S_{41} & S_{43}-\gamma_{3} S_{41} & \ldots & S_{4 N}-\gamma_{N} S_{41} & S_{41} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_{N 2}-\gamma_{2} S_{N 1} & S_{N 3}-\gamma_{3} S_{N 1} & \cdots & S_{N N}-\gamma_{N} S_{N 1} & S_{N 1} \\
\sum_{i} R_{i 2}-\gamma_{2} \sum_{i} R_{i 1} & \gamma_{3} \sum_{i} R_{i 3}-\sum_{i} R_{i 1} & \cdots & \sum_{i} R_{i N}-\gamma_{N} \sum_{i} R_{i 1} & \sum_{i} R_{i 1}
\end{array}\right] .
$$

Note that $\mathbf{M}_{\mathbf{2}}$ must also be either positive or negative definite. We now notice that the determinant in (8.22) is a principal minor of the matrix $\mathbf{M}_{\mathbf{2}}$ above. Since $\mathbf{M}_{\mathbf{2}}$ is either positive or negative definite, any principal minor of $\mathbf{M}_{\mathbf{2}}$ must be non-zero. Hence, T. $\mathbf{n} \neq 0$. This shows that $\mathbf{C}$ never becomes parallel to $\sum_{i=1}^{N} g_{i} p_{i}=P$. Note that $\mathbf{C}$ cannot intersect any of the coordinate planes because such an intersection would force one of the $p_{i}$ 's to be zero and that would violate our condition that none of the $\beta_{i}$ is zero. This proves $\mathbf{C}$ must intersect $\sum_{i=1}^{N} g_{i} p_{i}=P$.

Now suppose that $\mathbf{C}$ intersects $\sum_{i=1}^{N} g_{i} p_{i}=P$ at two points. Then, after the first intersection, $\mathbf{C}$ must turn back towards $\sum_{i=1}^{N} g_{i} p_{i}=P$ at some point. At this turning point, $\mathbf{C}$ must become parallel to $\sum_{i=1}^{N} g_{i} p_{i}=P$. This is a contradiction.

We now summarize the result of our analysis in the theorem below.

Theorem 8.5.1. The power allocation that maximizes the sum-rate $R$ of the GIC subject to the constraint $\sum_{i=1}^{N} g_{i} p_{i} \leq P$ and the proportional rate constraints in (8.6) exists, unique, and lies at the hyper plane $\sum_{i=1}^{N} g_{i} p_{i}=P$.

### 8.5.7 Two linear constraints on the powers

Proposition 8.5.1. Suppose the solution to the optimization problem with total power constraint is $P^{*}$ and the solution with the interference constraint is $P^{* *}$. Then the solution with both constraints is the one among these two points that is closest to the origin.

Proof: Note that the feasible set with both constraints is the intersection of the feasible sets with each of the constraints alone. Therefore both points $P^{*}$ and $P^{* *}$ are upper bounds for the new feasible set. By the proof of Theorem 8.5.1, both lie on the intersections of curve $\mathbf{C}$ with the respective constraint planes (Fig. 8.7). When we travel along curve $\mathbf{C}$ starting from the origin, we will encounter one of these points first. This point is the only one among the two points that is guaranteed to lie on the intersection of the two original feasible sets. This point will be the solution to the new optimization problem, because an upper bound that lies on the feasible set is the maximum. Note that this point has the shortest distance from the origin measured along curve $\mathbf{C}$.

Distance along the curve $\mathbf{C}$ from the origin can be replaced with distance from the origin unless curve $\mathbf{C}$ "turns too much," or more precisely, unless curve $\mathbf{C}$ becomes parallel to one of the planes. The proof of Theorem 8.5.1 established that this is not the case.


Figure 8.7: Solution with two linear constraints. $P^{*}$ and $P^{* *}$ are labels (not coordinates) for points in the Euclidean space.

### 8.6 Centralized algorithm for power allocation

In the last section, we proved that the power allocation that maximizes the sum-rate of the GIC subject to proportional rate constraints and $\sum_{i=1}^{N} g_{i} p_{i} \leq P$ lies on a unique point in the plane $\sum_{i=1}^{N} g_{i} p_{i}=P$. In this section, we first develop a polynomial time algorithm that finds this optimal point. Then we compare the complexity of this algorithm to a typical power allocation algorithm for GIC.

The straightforward way of determining the power levels at the optimal point is to solve the $N-1$ proportional rate constraints equations with the power constraint equation $\sum_{i=1}^{N} g_{i} p_{i}=P$. The $N-1$ proportional rate constraints equations are non-linear. This means we have to solve a system of non-linear equations. The most popular method to solve a system of non-linear equations is the multi-variable Newton-Raphson method. However, multi-variable Newton-Raphson method is highly unstable. That is, it is extremely sensitive to the initial guess for the solution, and depending on this initial point, may not converge
at all. Because of this, we abandoned this approach and investigated other ways to find the solution, and discovered the following method which is guaranteed to converge as well as faster than the multi-variable Newton-Raphson method.

### 8.6.1 Algorithm

Given an analytical or algebraic characterization of a point, there are many ways to write an algorithm so that a computing machine might be used to find the point. Some are direct and some are indirect. However, not all these methods would necessarily converge to that point. Some that do converge to the required point might take an extremely long time to reach that point.

In our case, the algebraic characterization of the optimal point was: "the intersection of the rate constraints and the power constraint plane." The most straight forward way of finding this point is to solve the $N-1$ non-linear equations simultaneously with the power constraint equation. This would form a system of $N$ non-linear equations. The most popular method of solving non-linear equations is the multi-variable Newton-Raphson algorithm. In fact, this is what we first attempted. The algorithm found the optimal point for some values of the channel gains, but diverged for the others. The multi-variable Newton-Raphson algorithm is highly unstable. We know very well that the single variable Newton-Raphson will fail to find the root if there is an extrema between the initial guess and the actual root. In multi-variable Newton-Raphson, this issue is magnified even more and this is why it is not suitable for our purpose.

Another example where not every algebraic definition will yield a converging algorithm is found in the popular problem of finding the fixed point. Suppose we know that a fixed point $x^{*}$ exists for the function $f(x)$. That is, there exists an $x^{*}$ such that $f\left(x^{*}\right)=x^{*}$. The simplest way of finding this point $x^{*}$ is through the algorithm: $x_{n+1}=f\left(x_{n}\right)$. But that will work only sometimes - only if the condition $\left|f^{\prime}\left(x^{*}\right)\right|<1$ is satisfied. If this condition is not satisfied, we must re-arrange the equation to another form until this condition is satisfied.

As an example, consider the fixed point of $f(x)=x^{2}$. Now $x=f(x)$ becomes $x=x^{2}$ and we see $x=1$ is an obvious fixed point. However the algorithm $x_{n+1}=x_{n}^{2}$ will not converge to the fixed point 1 , because $f^{\prime}(1)=2 \geq 1$. We can easily see this by starting at $x_{1}=0.9$ which will go zero (which is another fixed point) or by starting at $x_{1}=1.1$ which will go to infinity. On the other hand, if we re-write that equation as $x=\sqrt{x}$ and use the algorithm $x_{n+1}=\sqrt{x_{n}}$, it will converge to 1 , because for the new function $f(x)=\sqrt{x}, f^{\prime}(1)=1 / 2<1$.

In our problem, instead of using the multi-variable Newton-Raphson method, we decompose the task at hand into two problems, for each of which there is a "mature technology" [28] available to solve that problem. We first write the proportional rate equations in terms of User-1's rate $R_{1}$ to obtain a linear system in the powers $p_{i}$ 's. We then substitute the "solution" of this system into the power constraint equation to form a single non-linear equation in $R_{1}$.

If we let

$$
\begin{equation*}
S_{i}=2^{R_{i}}-1=2^{\alpha_{i} R_{1}}-1, \tag{8.17}
\end{equation*}
$$

(8.1) can be re-written as

$$
\begin{equation*}
h_{i i} p_{i}-S_{i} \sum_{j \neq i} h_{i j} p_{j}=S_{i} \sigma_{i}^{2} \text { for } i=1,2,3, \ldots, N . \tag{8.18}
\end{equation*}
$$

This is a linear system of $N$ equations in the solution variables $p_{i}$. This linear system can be written as

$$
\begin{equation*}
\mathbf{A p}=\mathbf{q} \tag{8.19}
\end{equation*}
$$

where the entries of the matrix $\mathbf{A}$ can be written in terms of $h_{i j}$ 's, $\aleph_{i}, \mathrm{~s}$ and $R_{1}$. Once we know the channel gains and proportional rates, these entries can be written in terms of $R_{1}$ alone. Similarly, $\mathbf{q}$ is the column vector with entries $S_{i} \aleph_{i}$ and these numbers can also be written in terms of $R_{1}$ alone. We use subscripts to denote this dependency on $R_{1}$ and re-write the
linear system:

$$
\begin{equation*}
\mathbf{A}_{R_{1}} \mathbf{p}=\mathbf{q}_{R_{1}} \tag{8.20}
\end{equation*}
$$

If we know $R_{1}$ then the above linear system can be used to find the $p_{i}$ 's. The idea is to find the $R_{1}$ that leads to $p_{i}$ 's that satisfy

$$
\begin{equation*}
\sum_{i=1}^{N} g_{i} p_{i}=P \tag{8.21}
\end{equation*}
$$

We now show that this can be accomplished using a linear system solver and a linear equation solver. Most platforms in which we do our programming these days come with these built-in functions or modules. We shall describe our algorithm using the Matlab built-in functions linsolve and fzero. How linsolve works will be clear from the way it is used in our algorithm. Suppose we wish to solve the non-linear equation $x^{5}-\sin (x)=0$ using fzero. We first define a function named "hello" as:

$$
\begin{align*}
& \text { function } y=\operatorname{hello}(x)  \tag{8.22}\\
& y=x^{5}-\sin (x) \tag{8.23}
\end{align*}
$$

Then we call fzero using the syntax fzero(@(t) hello( t$)$, $[\mathrm{a}, \mathrm{b}]$ ), where $[\mathrm{a}, \mathrm{b}]$ is an interval containing the root.

We now present our low-complexity algorithm that finds the power allocation that maximizes the sum-rate of the GIC subject to proportional rate constraints as well as a linear constraint on the transmit powers.

## Centralized Algorithm

1. Define the function called "pconstr" given below.
(a) function $y=\operatorname{pconstr}\left(R_{1}\right)$
(b) $\mathbf{p}=\operatorname{linsolve}\left(\mathbf{A}_{R_{1}}, \mathbf{q}_{R_{1}}\right)$
(c) $\mathbf{p}$
(d) $y=\sum_{i=1}^{N} g_{i} p_{i}-P$
2. Call the built-in function "fzero" by entering: fzero(@(x) pconstr(x), $[\mathrm{a}, \mathrm{b}]$ )
fzero may call the function "pconstr" several times until $y$ becomes close enough to zero. At every iteration, at step (c), "pconstr" will output $\mathbf{p}$. The $\mathbf{p}$ value at the last iteration is our power solution.

Note that the convergence of this algorithm to the global optimal point is guaranteed because of the following. Theorem 8.5.1 asserts that for a particular $P$ and a particular set of proportional rate constraints, there exists only a unique set of power levels and those power levels actually globally maximize the sum-rate. This unique set of power levels corresponds to a unique $R_{1}$ as given by (8.1). This shows that $R_{1}$ exists and is unique for a given P .

### 8.6.2 Complexity

Even though fzero might call "pconstr" several times, this number does not depend on the number of users $N$. There is only one non-linear equation to solve even if there are 25 users. Therefore the number of times the outer-loop is executed does not increase with $N$. Inside the function "pconstr, there is a step involving solving a system of $N$ by $N$ linear equations. Solving linear systems is considered a matured technology, meaning there are extremely reliable software packages that can solve systems with a very large number of variables very accurately within a very short time [28]. Since the time complexity of solving an $N$ by $N$ system is $O\left(N^{3}\right)$ [28], the complexity of our algorithm is $O\left(N^{3}\right)$, where $N$ is the number of users.

### 8.7 Distributed algorithms for power allocation

The Centralized Algorithm presented in the last section requires a duplex control channel between each user and a central controller. Each user must estimate its channel coefficient and report it to the central controller via the control channel and then obtain the optimum power level from the central controller. In this section, we provide two distributed algorithms, one of which completely eliminates the need for this control channel and the other eliminates the need for a central controller. It should be noted that even though the structure and the design of these algorithms seem fairly intuitive on their own, the proof that these algorithms actually converge to the power allocation that maximizes the sum-rate while maintaining the proportional rate constraints relies on Theorem 8.5.1 of Section 8.5.

### 8.7.1 Distributed algorithm-1

This algorithm assumes no communication between the users, or between a user and a central controller. The only requirement is that the users have access to synchronized clocks and that each user is aware of the total interference constraint $P$. The time period allocated to power control during the control part of the frame is divided into time slots and the power levels are updated on each time slot until convergence.

The total power or the interference constraint places considerable limitations on the possible distributed algorithms. Since there is no way to check if this constraint is met in the middle of the algorithm, the only choice is to start with a power allocation that satisfies this constraint and then preserve this during each iteration. Hence, the algorithm starts with User- $i$ employing a power $p_{i}=P /\left(N g_{i}\right)$.

After this, all users update their powers at the aforementioned synchronized time slots. During each time slot only one user increases its power by $s / g_{i}$, while all the others decrease their power by $s /\left[g_{i}(N-1)\right]$, where $s$ is a predetermined step size. Note that this keeps the sum interference at $P$. Which user needs to increase its power is determined in the following
way. Suppose $\gamma_{i}$ is the SINR of User- $i$. We define Fairness Quotient $F Q_{i}$ of User- $i$ as follows:

$$
\begin{equation*}
F Q_{i}=\left(1+\gamma_{i}\right)^{1 / \alpha_{i}} . \tag{8.24}
\end{equation*}
$$

The proportional rate constraints in (8.6) can be now re-written as

$$
\begin{equation*}
F Q_{i}=K \text { for } i=1,2,3, \ldots, N \tag{8.25}
\end{equation*}
$$

where $K$ is a number that is independent of $i$ and $\alpha_{1}$ is defined as unity. By Theorem 8.5.1, for a fixed interference constraint $P$, there is only one power allocation and hence one $K$ that satisfies this equation. This implies that for a fixed interference constraint $P$, there is only one $K$. To emphasize this one-to-one relationship between $K$ and $P$, we re-write the last equation as

$$
\begin{equation*}
F Q_{i}=K_{P} \text { for } i=1,2,3, \ldots, N . \tag{8.26}
\end{equation*}
$$

Note that the above equation will be valid only at the optimal power allocation. For an arbitrary power allocation satisfying the interference constraint, such as at the beginning of the proposed distributed algorithm, some $F Q_{i}$ 's will be lower than $K_{P}$ while the others greater than $K_{P}$. Note also that while the value of $K_{P}$ can be calculated using the Centralized Algorithm when all users' channel gains are known, in the current distributed decision making scenario, users have no way of determining $K_{P}$. However, each user can calculate its own current Fairness Quotient from its SINR. The proposed distributed algorithm works by allowing the user with the lowest $F Q_{i}$ value to increase its power in each time slot. At the beginning of each time slot, each user starts a timer that expires after $c F Q_{i}$ time units, where $c$ is an appropriate constant that was previously agreed upon. The user whose timer expires first increases its power by $s / g_{i}$. At this point, other users will observe a reduction in their SINR. This is the signal for these users to reduce their power by $s /\left[g_{i}(N-1)\right]$, stop their timers, and wait for the beginning of the next time slot.

At the beginning of the algorithm, a larger step size $s$ will be useful so that the gap in $F Q$ values can be bridged sooner. However, as can be seen from the proof of the Theorem 8.7.1 below, at the end of the algorithm, a larger value of $s$ will make the $F Q$ values oscillate with a large amplitude. A step size that decreases during the run of the algorithm would be the best one; and for this reason, a function $f(t)$, which we call the accelerating factor, is introduced, where $t$ is the number of iterations. $f(t)$ must be an increasing function in $t$; but exactly what function to choose is best decided by simulations. The algorithm is terminated when the change in the SINR in successive iterations becomes too small to be of any practical value, that is, when it is smaller than a predetermined number $\gamma_{\text {stop }}$.

## Distributed Algorithm-1

1. $p_{i}:=P /\left(N g_{i}\right), \forall i$.
2. Compute $F Q_{i}=\left(1+\gamma_{i}\right)^{1 / \alpha_{i}}$ and at the beginning of the next time slot set the timer to expire exactly after $c F Q_{i}$ time units.
3. The User- $i$ whose timer expires first increases its power by $s / g_{i}$.
4. If any user's SINR $(\gamma)$ decreases before its timer expires, it decreases its power by $s /\left[g_{i}(N-1)\right]$.
5. Go to step 2 with $s:=s / f(t)$ unless change in SINR is smaller than $\gamma_{\text {stop }}$.

The following Lemmas and definitions are necessary for the proof that this distributed algorithm converges.

Lemma 8.7.1. For a given interference constraint $P$, there cannot be any power distribution that makes $F Q_{i}>K_{P}$ for all $i$.

Proof: Suppose there is such a power distribution. For ease of explanation, let us consider the two-user case first. Without loss of generality, we can assume there is a power distribution such that $F Q_{1}>F Q_{2}>K_{P}$. By incrementally reducing User-1's power but increasing the power of User-2, we can make $F Q_{1}=F Q_{2}>K_{P}$, while satisfying the interference constraint $P$. This would imply that there are two different power distributions both satisfying a particular proportional rate and interference constraints. This contradicts Theorem 8.5.1. The argument for the case with more than two users follows the same path. The contrary assumption implies that by a re-distribution of power levels we can make $F Q_{1}=F Q_{2}=F Q_{3}=\ldots=F Q_{N}>K_{P}$ while satisfying the interference constraint $P$. This contradicts Theorem 8.5.1.

Definition 8.7.1. Increases to within $\delta: A$ sequence $\left\{a_{i}\right\}$ is said to be increasing to within $\delta$ if $a_{i+1}>a_{i}-\delta$ for all $i$, where $\delta$ is a fixed positive number whose magnitude is small compared to any $a_{i}$.

Note that Decreases to within $\delta$ can be defined in a similar manner.
Definition 8.7.2. Converges to $C$ within $\delta$ from below: $A$ sequence $\left\{a_{i}\right\}$ is said to converge to $C$ within $\delta$ from below if there exists an $M$ such that $C-\delta \leq a_{i} \leq C$ for $i>M$.

Converges to $C$ within $\delta$ from above is defined in a similar way.
Lemma 8.7.2. A sequence that is bounded above (below) and increases (decreases) to within $\delta$ converges to its least upper bound (greatest lower bound) within $\delta$ from below (above).

Proof: Follows from the well-known theorem that a sequence that is bounded above (below) and increases (decreases) converges to its least upper bound (greatest lower bound).

Theorem 8.7.1. Distributed Algorithm-1 converges to the unique, optimum power levels mentioned in Theorem 8.5.1.

Proof: The behavior of the algorithm with a fixed step size $s$ is established first. In other words, assume $f(t)=1$ for now. Consider two sequences of numbers selected from the $F Q$ values from successive iterations of the algorithm. First one, which we call the Lower Sequence consists of the smallest $F Q$ value during each iteration. Suppose during the first iteration User-3 has the lowest $F Q$ value, during the second iteration User-1 has the lowest $F Q$ value and during the third iteration User-5 has the lowest $F Q$ value and so on. With a slight abuse of notation, the lower sequence would then look like $F Q_{3}, F Q_{1}, F Q_{5}, \cdots$.

At first it would appear that the Lower Sequence is always increasing. This is because at any iteration, the user with the lowest $F Q$ is allowed to increase its power and hence its $F Q$. However, a closer examination shows that this is true only at the beginning. Sooner or later these increasing values will cross over the $F Q$ value of a user who is decreasing its power. The new lowest $F Q$ may not be necessarily higher than the previous one. As can be seen from Fig. 8.8, the new $F Q$ value can be lower than the old $F Q$ value.

When a user decreases its power its $F Q$ value will decrease accordingly. How much $F Q$ decreases will depend on the channel gains as well as the current power levels. Since the changes in power levels are the same from iteration to iteration, the magnitude of the change in $F Q$ for a particular User- $i$ will remain approximately the same throughout the run of the algorithm. However, it will be different from user to user. Let $\Delta F Q_{M}$ stand for the magnitude of the maximum change in the $F Q$ over all users. That is, let

$$
\begin{equation*}
\Delta F Q_{M}=\max _{i} \Delta F Q_{i} \tag{8.27}
\end{equation*}
$$

As illustrated in Fig. 8.8, the new $F Q$ value can be lower than the old $F Q$ value by at most $\Delta F Q_{M}$. This shows that the Lower Sequence is increasing to within $\Delta F Q_{M}$.

The Lower Sequence gets arbitrarily close to $K$ but cannot exceed $K$. It cannot exceed $K$ because of the following reason. Suppose at the end of an iteration the Lower Sequence exceeds $K$. This would imply that the $F Q$ values at the end of this iteration are all greater


Figure 8.8: Worst-case crossing.
than $K$. This contradicts Lemma 8.7.1. Thus $K$ is the least upper bound of the Lower Sequence which increases to within $\Delta F Q_{M}$. By Lemma 8.7.2, it converges to $K$ within $\Delta F Q_{M}$ from below.

Now consider the Upper Sequence which consists of the highest $F Q$ values on successive iteration. It can be shown, using the arguments similar to the ones we used for the Lower Sequence, that the upper sequence decreases to within $\triangle F Q_{M}$ and gets arbitrarily close to $K$ but never goes below $K$. Hence, the Upper Sequence converges to $K$, to within $\Delta F Q_{M}$ from above.

The sequence of $F Q_{i}$ values of any User- $i$ through the iterations is trapped between the Lower and Upper sequences and hence will eventually oscillate about $K_{P}$ to within $2 \Delta F Q_{M}$.

The accelerating factor $f(t)$ will make the effective step size approach zero as the iteration number increases. This will have the effect of making $\Delta F Q_{M}$ approach zero. Thus, the $F Q_{i}$ values will all eventually approach $K_{P}$.

Earlier, we explained that Theorem 8.5.1 implied that the power levels will satisfy $F Q_{i}=$ $K_{P}$ for all $i$ only at the global optimal point. Therefore, Distributed Algorithm-1 converges to the global optimum power allocation.

### 8.7.2 Distributed algorithm-2

An algorithm that conforms to the Shared Memory model in the study of distributed algorithms [116] is presented here. The proposed algorithm assumes that control channels are available between each user and a shared memory location so that it can write its rate and obtain the rates of others or the sum-rate. Note that this memory can be provided by one of the users, a leader.

This algorithm also starts with power levels that satisfy the interference constraint. Unlike the last algorithm, all users simultaneously update their power levels during each time slot. However, this is done in a way that the interference constraint is met throughout the run of the algorithm. The change in power level is based on the each user's proportion of its
rate to the sum-rate of all $N$ users. This proportional rate is defined as

$$
\begin{equation*}
r_{i}=\frac{R_{i}}{\sum_{i} R_{i}} \tag{8.28}
\end{equation*}
$$

At the beginning of each time slot the sum-rate of the $N$ users is obtained by each user from the shared memory. This enables each user to calculate its proportional rate at each iteration of the algorithm. Let $r_{i t}$ be the proportional rate of User- $i$ at the beginning of iteration number $t$ and $r_{i}$ be the proportional rate requested by User- $i$. During iteration number $t$, User- $i$ would change its power by

$$
\begin{equation*}
\Delta p_{i}=\left(k / g_{i}\right)\left(r_{i}-r_{i t}\right) \tag{8.29}
\end{equation*}
$$

where $k$ is an appropriate constant. Note that this means a user whose current proportional rate is lower than its required proportional rate will end up increasing its power while a user whose current proportional rate is higher than its required proportional rate will end up decreasing its power. The following Lemma proves that the proposed update in power levels satisfies the interference constraint.

Lemma 8.7.3. Suppose the power levels of the users satisfy the interference constraint $\sum_{i} g_{i} p_{i}=P$ at the beginning of an iteration. The power levels after a change of power levels given by $\Delta p_{i}=\left(k / g_{i}\right)\left(r_{i}-r_{i t}\right)$ will still satisfy the same interference constraint.

Proof: Since $r_{i}$ and $r_{i t}$ are ratios of user rates to the sum-rate as defined in (8.28), we have

$$
\begin{equation*}
\sum_{i} r_{i}=\sum_{i} r_{i t}=1 \tag{8.30}
\end{equation*}
$$

The change in the sum $g_{i} p_{i}$ is

$$
\begin{align*}
\Delta \sum_{i} g_{i} p_{i} & =\sum_{i} g_{i} \Delta p_{i}  \tag{8.31}\\
& =\sum_{i} k\left(r_{i r}-r_{i t}\right), \text { using (8.29) }  \tag{8.32}\\
& =k \sum_{i}\left(r_{i r}-r_{i t}\right)  \tag{8.33}\\
& =k\left(\sum_{i} r_{i}-\sum_{i} r_{i t}\right)=k(1-1)=0 . \tag{8.34}
\end{align*}
$$

## Distributed Algorithm-2

1. $p_{i}:=P /\left(N g_{i}\right), \forall i$.
2. $p_{i}:=p_{i}+\Delta p_{i}=p_{i}+\left(k / g_{i}\right)\left(r_{i}-r_{i t}\right), \forall i$.
3. Go to step 2 unless change in SINR is smaller than $\gamma_{\text {stop }}$.

The following theorem proves that Distributed Algorithm-2 converges as long as the proportionality constant $k$ is not too large. It should be noted here that the condition on $k$ is nothing new. For example, the gradient decent algorithm, which finds the location of the minimum of $f$, increments its independent variables in step sizes of $k \nabla f$, and will overshoot and oscillate about the minimum point if $k$ is too large.

Theorem 8.7.2. Distributed Algorithm-2 converges to the unique optimal point mentioned in Theorem 1 where the proportional rate and the interference constraints are satisfied.

Proof: Consider User- $i$ on iteration number $t$. Without loosing generality we may assume that its proportional rate $r_{i t}$ at this point is lower than what it requested. This means it will increase its power. Since everyone whose proportional rate is smaller than their requested proportional rate would have increased their powers while the others lowered their powers, the new proportional rate of User- $i$ would be higher. As long as $k$ is not too large, it will still be smaller than the requested proportional rate $r_{i}$. This shows the proportional rates $r_{i t}$ 's will form an increasing sequence whose least upper bound is $r_{i}$. Hence it will converge to $r_{i}$. This shows that the users' rates will converge to the proportional rates requested.

Lemma 8.7.3 showed that during every iteration of the algorithm the power levels remain on the interference constraint plane. Therefore, the point of convergence is on this plane. By Theorem 8.5.1, there is only one point on this plane that satisfies the proportional rate constraints. Hence, this point of convergence is the unique optimal point.

### 8.7.3 Complexity

Let $\Delta p_{i}$ be the deference between the initial and final power level of User- $i$ during the execution of Distributed Algorithm-1. In the worst case, this user may have to decrease its power throughout the run of the algorithm. In each iteration, it will decrease the power by $s /\left[(N-1) g_{i}\right]$, or approximately by $s /\left[\left(N g_{i}\right]\right.$, where $s$ is the step size. If $s$ is the average step size during the run of the algorithm, then it would take $\Delta p_{i} N g_{i} / s$ iterations.

Each iteration consists of one time slot of fixed duration in which all users update their powers, even though they have to wait until their timers to expire before they can increase the power or notice a decrease in their SINR and reduce the power. The important question is if the length of the time slot should depend on the number of users. The time slot should be long enough to allow each user to estimate its SINR and calculate $F Q_{i}=\left(1+\operatorname{SINR}_{i}\right)^{1 / \alpha}$. It should also be long enough to accommodate a length of time equal to $c\left[F Q_{i}\right]_{\max }$. But $\left[F Q_{i}\right]_{\text {max }}$ does not depend on how many users are there and $c$ can be chosen appropriately to fit $\left[F Q_{i}\right]_{\text {max }}$ into an appropriate time interval in which users can estimate their SINR and set


Figure 8.9: Scenario-1: Simulation details (only two of the 6 transceivers are shown).
the timers. This shows the length of the time slot does not depend on the number of users. If we denote the length of each time slot by $\tau$ and $\lambda_{1}=\left\{\Delta p_{i} g_{i}\right\}_{\text {max }}$, then the complexity order of Distributed Algorithm-1 would be $\tau \lambda_{1} N / s$.

A similar calculation shows that the time complexity of Distributed Algorithm-2 is $\tau \lambda_{2} / s$, where $\lambda_{2}$ is a constant. Its run time is independent of the number of users.

### 8.8 Numerical results

Numerical results obtained using simulations for three different scenarios are presented here. In each scenario there are 6 users. A path-loss coefficient of 4 is used to calculate the channel coefficients from the distances. The performance of the Centralized Algorithm is demonstrated first followed by that of the distributed algorithms.


Figure 8.10: Transmission rates and power levels for Scenario-1.

### 8.8.1 Centralized algorithm

The first scenario is about power allocation in GIC under a total power constraint. The details are shown in Fig. 8.9, where we deliberately place the 2nd and 5th users in unenviable positions: their transmitter to receiver distances are a bit greater than the other users. The power allocation by the Centralized Algorithm with the total power constraint of 80 micro Watts is illustrated in Fig. 8.10. Not surprisingly, all the runs of the Centralized Algorithm produce transmission rates that are exactly at the proportional rates requested: 1.0000, $1.2000,1.4000,1.6000,1.8000$ and 2.000 . We use bar charts to illustrate the patterns in power allocation that is required to produce this rate ratios. Note that in Fig. 8.10, the power levels of User-2 and User-5 are higher than what should be expected (the inclined line) for their rate demands. This should be anticipated because, by our design, their transmitter to receiver distances are greater than the others.

The second and third scenarios deal with power allocation in GIC under an interference constraint. Details of Scenario-2 are shown in Fig. 8.11, where the 6 users from the previous


Figure 8.11: Scenario-2: Simulation details (only two of the 6 transceivers are shown).
scenario now become primary users, and an interference measuring point, indicated by PR, is added. An interference tolerance level of $P=1$ pico Watts is used. The power levels prescribed by our algorithm is shown in Fig. 8.12. We notice that the rates of users for Scenario-2 are smaller compared to those for Scenario-1. This is because an interference tolerance level of 1 pico Watts and the assumed transmitter to interference measuring point distances in Scenario-2 put a stringent condition on the power levels than a total power constraint of 80 micro Watts from Scenario-1.

The third scenario is identical to the second one except that the value of $s_{5}$ is fixed at 70 meters, which is lower than the other $s_{i}$ 's. We put User- 5 in a tough position by placing it closer to the interference measuring point than the others. The power allocation


Figure 8.12: Transmission rates and power levels for Scenario-2.


Figure 8.13: Transmission rates and power levels for Scenario-3.
for this Scenario is shown in Fig. 8.13. We notice that the rates of all users are smaller compared to Scenario-2. User-5's proximity to the interference measuring point pushes its power down, which in turn has the effect of lowering the powers of all the other users through the proportional rate constraint.

The algorithms mentioned in the review of the related work all use with minimum rate constraints but ours use proportional rate constraints. Therefore, it is not possible to compare our algorithm to those algorithms directly. However, we will now demonstrate the performance of our algorithm in solving the problem for which the algorithms from the literature are designed.

As mentioned before, the algorithm by Chiang et al. [109,110] uses geometric programming. We used Boyd's geometric program (GP) software [28] to imitate Chiang's algorithm and compared the results to the that from our algorithm. We considered the sum-rate maximization problem with minimum rate constraints and a total power constraint of 80 micro Watts. We calculated the proportional rates using the minimum rate demands and used them in our proportional rate algorithm. If the minimum rates are feasible then our algorithm should produce a result that satisfies the minimum rate demands. Note that our algorithm will not provide optimal sum-rate. This is because it uses that exact proportional rates whereas minimum rates give more leeway in the feasible set. But we expect our algorithm to be faster than GP. We restricted the number of users to 4 because of the long execution time for the geometric programming. The channel gains are obtained from Scenario-1, without the exception to User-2.

Table 8.2 shows that our algorithm achieves more than $90 \%$ of the sum-rate achieved by GP but it is faster by orders of magnitude. Granted we used an application program for GP, perhaps we can knock off one zero. Our algorithm is still about 100 times faster. This is only with 4 users; the comparison will be even more impressive with 10 users.

| Minimum rate <br> demand (bits/s/Hz) |  | sum-rate <br> (bits/s/Hz) | Time <br> (Sec.) |
| :---: | :---: | :---: | :---: |
| $3,3,3,3$ | GP | 16.03 | 3.89 |
|  | Ours | 16.34 | 0.00069 |
| $4,2,2,2$ | GP | 16.03 | 3.62 |
|  | Ours | 14.60 | 0.00053 |
| $3,3,2,2$ | GP | 16.03 | 3.61 |
|  | Ours | 15.64 | 0.00059 |
| $3,3,4,4$ | GP | 16.03 | 3.60 |
|  | Ours | 15.89 | 0.00076 |
| $1.5,1.5,3,5$ | GP | 15.23 | 3.62 |
|  | Ours | 12.59 | 0.00068 |
| $5,3,1.5,1.5$ | GP | 15.77 | 3.61 |
|  | Ours | 13.11 | 0.00057 |

Table 8.2: Centralized algorithm compared with geometric programming

### 8.8.2 Distributed algorithms

The second scenario in the previous subsection with the same interference tolerance level of $P$ $=1$ pico Watts is used to evaluate the distributed algorithms. Fig. 8.14 shows the behavior of the Distributed Algorithm-1 with an accelerating factor of $f(t)=t^{1 / 30}$. To show that the algorithm behaves exactly as predicted in the convergence proof, we have plotted the $F Q$ values of the users during the execution of the algorithm. In the proof of convergence, it was mentioned that a particular $F Q$ curve will jump in the opposite direction when it crosses another curve. This can be clearly seen in Fig. 8.14. The proportional rates produced by this algorithm, shown in Fig. 8.14, are satisfactory.

For the purpose of comparison, we use the Centralized Algorithm to calculate the $K_{P}$ value and it turned out to be 4.1223 . The convergence proof of the Distributed Algorithm1 claims that the $F Q$ values should converge to $K_{P}$. In Fig. 8.14, we see the $F Q$ values converging to a number just above 4. A separate plot in Fig. 8.15 shows how the proportional rates converge to the required ratios.

The behavior of the Distributed Algorithm-2 for the same simulation scenario is shown


Figure 8.14: Distributed Algorithm-1: Convergence of the Fairness Quotients.


Figure 8.15: Distributed Algorithm-1: Convergence of the proportional rates.


Figure 8.16: Distributed Algorithm-2: Convergence of the proportional rates.
in Fig. 8.16. We see a much faster convergence to the required proportional rates. It should be noted that this performance gain is obtained at the cost of a communication channel between each user and a shared memory.

### 8.9 Conclusion

We have presented a deterministic solution to the optimization problem of finding the power allocation that maximizes the sum-rate of the Gaussian interference channel subject to any linear power constraint and proportional rate constraints. This has been accomplished using analytic geometry in higher dimensions to show that the curve of intersection of the sumrate and the proportional rate constraints is always increasing and the maximum sum-rate occurs at the unique point where this curve intersects the boundary plane formed by the linear power constraint. A polynomial time centralized algorithm as well as two distributed
algorithms that find the optimal power allocation have been proposed. Simulation results supporting the analysis and demonstrating the performances of the algorithms have been presented.

## Chapter 9

## Throughput maximization in the GIC with minimum rate constraints

### 9.1 Introduction

Finding the power allocation that maximizes the throughput of the Gaussian Interference Channel (GIC) has remained a difficult problem because it is not known if the throughput posses the classical convex (or concave) structure. This problem with a total power constraint has actually remained unsolved, except for the case of two users [107]. That is, to the best of our knowledge, there is no algorithm that finds the throughput maximizing power allocation, for more than two users, under a total power constraint.

As mentioned in Chapter 4, in the 80 's, a more general form of geometrical structure called invexity under which any local extrema automatically becomes a global extrema was proposed. This chapter establishes that while the throughput of the GIC is both incave and invex, the throughput restricted to a linear constraint on the powers is incave. It is shown that one of the KKT conditions can be used to locate the power allocation that maximizes the throughput subject to the said constraint on the powers. The existence and uniqueness of a point on the power constraint plane that satisfies the KKT condition is proven. These
theoretical results lead to the first ever algorithm that maximizes the throughput subject to total power and QoS constraints.

The proposed algorithm uses a new idea that can be used by any algorithm for solving a very common problem encountered in optimization: locating the point that satisfies the KKT condition. This problem arises in many engineering applications that requires solving a convex optimization problem. The current algorithms for this purpose are fraught with stability issues because they are sensitive to the initial values of the dual variables in the problem [108]. The proposed algorithm eliminates the dual variables from the problem and hence does not suffer from any stability issues. The algorithm uses the new idea of projections onto the linear space defined by the constraints. This algorithm is also at least twice as fast the most common algorithm [108] for finding the optimum point (for other problems, not sum rate maximization) using the KKT condition.

### 9.2 Background

Gaussian interference channel (GIC) has been used to model the uplinks of code-division multiple access (CDMA) systems, digital subscriber line (DSL) systems, and more recently, ad-hoc networks, and small cell networks such as femtocell networks. First investigated by Shannon [98], the capacity region of the GIC continues to elude us except for particular cases of channel parameters [98-101]. Researchers concerned with power allocation that maximizes the system capacity have sidestepped this difficulty by optimizing what is called the sum-rate. This sum-rate is obtained by applying Shannon's original formula for capacity to each user separately while considering the interference from the other users as noise. Even then, the problem of finding the power allocation that maximizes the sum-rate has remained a difficult problem. To the best of our knowledge, there are no techniques or algorithms that find the power allocation that maximizes the sum rate of the GIC subject to a total power constraint and any quality of service (QoS) constraints, for more than two users.

There is only one algorithm that works with a total power constraint, the one from the last chapter, but it uses proportional rate constraints. The main reason for this is that the convex structure of the sum rate was unknown.

Optimizing convex functions is relatively easy because any local extremum is necessarily the global extremum. In case of constrained optimization, convexity provides the simplest sufficient condition for optimality that characterizes the critical points: the KKT conditions. Convex optimization has been studied for well over 70 years now. The algorithms that solve convex optimization problems now belong to a category called "mature technology" because many reliable software packages are available that guarantee the solutions within specified time limits [28].

Optimizing non-convex functions is difficult mainly for two reasons. First is that the simple characterization of the critical point such as the KKT conditions may not be applicable. The second is the possible existence of many local optima; a search algorithm that is looking for the global optimum point might get stuck at a local optimum point. Thankfully, convexity is not the only condition under which a local extremum necessarily becomes a global extremum. As shown in Chapter 2, the famous example is pseudo-convexity.

What is impressive about invexity [88-90] is that while it is more general than pseudoconvexity and quasi-convexity, invexity guarantees that any local extremum is a global extremum. Furthermore, under certain conditions, the solution to an invex optimization problem can be obtained via one of the KKT conditions. Invexity is applicable only to differentiable functions. For engineers, this is not a serious limitation, because most functions we attempt to optimize are differentiable.

Please see Chapter 8 for a review of the related literature. Section 9.3 describes the system model and formulates the optimization problem. Application of the invex analysis (Chapter 4) to the sum rate maximization of the GIC is presented in Section 9.4. An algorithm that maximizes the sum rate of the GIC subject to any linear constraint on the powers as well as QoS constraints is developed in Section 9.5. Numerical results that demonstrate the
convergence of the proposed algorithm under widely varying channel conditions are given in Section 9.6. The chapter concludes in Section 9.7.


Figure 9.1: The $N$-user Gaussian Interference Channel

### 9.3 System model and the problem statement

Consider the $N$-user GIC shown in Fig. 9.1. User $i$ employs transmitter $i$ to communicate with receiver $i$ but receiver $i$ experiences interference from all other transmitters. $h_{i j}$ denotes the channel gain between transmitter $j$ and receiver $i$, and $p_{i}$ denotes the power used by transmitter $i$. If $\sigma_{i}^{2}$ is the additive white Gaussian noise power per Hz , and $N_{i}=\sigma_{i}^{2} / h_{i i}$, and $a_{i j}=h_{i j} / h_{i i}$, the sum transmission rate $R$ of the GIC can be written as:

$$
\begin{equation*}
R=\sum_{i=1}^{N} \log _{2}\left(1+\frac{p_{i}}{N_{i}+\sum_{j \neq i} a_{i j} p_{j}}\right) \tag{9.1}
\end{equation*}
$$

We wish to establish the generalized convex structure of the sum rate $R$ as well the sum rate restricted to a linear constraint on the powers:

$$
\begin{equation*}
g(\mathbf{p})=P-\sum_{i=1}^{N} g_{i} p_{i} \geq 0 \tag{9.2}
\end{equation*}
$$

where $\mathbf{p}=\left(p_{1}, p_{2}, \ldots p_{N}\right)$. We then wish to give an analytic characterization of the power allocation that maximizes the sum rate subject to the linear power constraint.

Finally, we wish to add minimum rate constraints. Note that minimum rate constraints can be translated to minimum SINR constraints:

$$
\begin{equation*}
\operatorname{SINR}_{i} \geq \gamma_{i} \text { for } i=1,2, \ldots N \tag{9.3}
\end{equation*}
$$

The constraints above can be re-written as:

$$
\begin{equation*}
\frac{p_{i}}{N_{i}+\sum_{j \neq i} a_{i j} p_{j}} \geq \gamma_{i} \text { for } i=1,2, \ldots N . \tag{9.4}
\end{equation*}
$$

This is a linear inequality in the powers $p_{i}$ and can be re-arranged to:

$$
\begin{equation*}
\sum_{j=1}^{N} b_{i j} p_{j} \geq c_{i} \text { for } i=1,2, \ldots N \tag{9.5}
\end{equation*}
$$

where

$$
b_{i j}=\left\{\begin{array}{ll}
1, & \text { if } i=j  \tag{9.7}\\
-a_{i j} \gamma_{i}, & \text { if } i \neq j
\end{array} \quad(9.6) \quad \text { and } c_{i}=\gamma_{i} N_{i}\right.
$$

A summary of the notations are given in Table 9.1.
If we change the inequalities in (9.5) into equalities we get $N$ hyper planes in $R^{N}$. All these inequalities combine to form a hyper solid bounded by the said hyper planes. User's minimum rate demands will be met only when the power allocation falls within this hyper
solid.

| $N$ | number of users in the system |
| :---: | :---: |
| $h_{i j}$ | channel gain between transmitter $j$ and receiver $i$ |
| $p_{i}$ | power used by transmitter $i$ |
| $\gamma_{i}$ | minimum SINR acceptable to User- $i$ |
| $\sigma_{i}^{2}$ | additive white Gaussian noise power density |
| $N_{i}$ | $\sigma_{i}^{2} / h_{i i}$ |
| $a_{i j}$ | $h_{i j} / h_{i i}$ |
| $R$ | sum rate or throughput |
| $g_{i}$ | channel gain from transmitter $i$ to the interference measuring point |
| $P$ | total power constraint |

Table 9.1: Notations

## A specific question on power allocation

We are also interested in answering the following question. Are there any channel conditions under which, the sum rate maximizing power distribution, under a total power constraint but without any rate constraints, allocates zero power to some users? In case of two users, it is easily seen that this indeed happens by conducting an exhaustive search on the power levels using simulation. Performing an exhaustive search over three or more users is very time consuming. Still, there is a clever way of out of this as shown by the following example.

Example: Consider a partially symmetric placement of users where (using the same notation as in Fig. 9.1) $h_{i i}=30^{-4}$ for $i=1,2,3, \quad h_{44}=140^{-4}$ and $h_{i j}=240^{-4}$ for all $i \neq j$.

The first three users all have identical and favorable channel conditions, whereas the fourth user faces adverse channel conditions. Because of the partial symmetry, we know that the sum rate maximizing power distribution will allocate equal powers to the first three users. This reduces the exhaustive search to just two variables: one power level for the first three users and another one for the fourth user. A total power of $P=100$ micro Watts is divided into two parts. One is given to the fourth user. The other part is divided into three
equal parts and given to the first three users. Fig. 9.2 shows the plot of the sum rate against the part given to the fourth user in the simulation: the maximum sum rate occurs when the last user receives no power.


Figure 9.2: Example: The sum rate in the partially symmetrical case.

We would like to know if there is any theoretical basis for this. Exactly how bad the channel gain of the last user must be for this to occur? Especially, given the channel gains, is it possible to predict this before going through the sum rate optimization?

### 9.4 Generalized convex structure of the sum rate of the GIC

The three central theorems of this chapter are presented in this section. The first one proves that the sum rate of the GIC is both invex and incave. The second one proves that the sum rate restricted to a linear power constraint is incave and that the global maximum sum rate
restricted to the said constraint can be located using one of the KKT conditions. The third theorem proves the existence and uniqueness of such a point on the power constraint plane. The notation $\mathbf{v} \succ \mathbf{0}$ is used to indicate that the components of vector $\mathbf{v}$ are all positive.

Theorem 9.4.1. The gradient of the sum rate $R$ of the GIC (with respect to the powers $p_{i}$ 's) is never zero $(\nabla R \neq \mathbf{0})$ and hence the sum rate $R$ is both invex and incave in the powers $p_{i}$. Furthermore, the global maximum of $R$ subject to the constraint $g(\mathbf{p}) \geq 0$ can only occur on $g(\mathbf{p})=0$.

Proof: Using $I_{i}$ to denote $N_{i}+\sum_{j \neq i} a_{i j} p_{j}$, the partial derivative of the sum rate $R$ in (9.1) with respect to $p_{i}$ can be written as:

$$
\begin{equation*}
R_{p_{i}}=\frac{1}{p_{i}+I_{i}}-\sum_{j \neq i} \frac{a_{j i} p_{j}}{\left(p_{j}+I_{j}\right)\left(I_{j}\right)} \tag{9.8}
\end{equation*}
$$

The gradient of the sum rate $R$ can then be written as:

$$
\begin{equation*}
\nabla R=\mathbf{M x} \tag{9.9}
\end{equation*}
$$

where the entries the matrix $\mathbf{M}$ and the vector $\mathbf{x}$ are given by

$$
m_{i j}=\left\{\begin{array}{ll}
1, & \text { if } i=j \\
\frac{-a_{j i} p_{j}}{I_{j}}, & \text { if } i \neq j \tag{9.11}
\end{array} \quad \text { (9.10) } \quad \text { and } x_{i}=\frac{1}{p_{i}+I_{i}}\right.
$$

Notice that $\mathbf{x} \neq \mathbf{0}$. We now perform the following three operations on $\mathbf{M}$ in that order to arrive at a new matrix $\mathbf{M}^{\prime}$.

1) Multiply column $j$ by $I_{j} / p_{j}$ for $j=1,2,3, \ldots, N$.
2) Multiply row $i$ by $p_{i}$ for $i=1,2,3, \ldots, N$.
3) Transpose

$$
m_{i j}^{\prime}= \begin{cases}I_{i}, & \text { if } i=j  \tag{9.12}\\ -a_{i j} p_{j}, & \text { if } i \neq j\end{cases}
$$

Recalling that

$$
\begin{align*}
& I_{i}=N_{i}+\sum_{j \neq i} a_{i j} p_{j}, \quad \text { we observe that }  \tag{9.13}\\
& I_{i}>\sum_{j \neq i} a_{i j} p_{j} \text { for all } i . \tag{9.14}
\end{align*}
$$

This shows that the matrix $\mathbf{M}^{\prime}$ satisfies the condition

$$
\begin{equation*}
\left|m_{i i}^{\prime}\right|>\sum_{j \neq i}\left|m_{i j}^{\prime}\right| \quad \text { for all } i \tag{9.15}
\end{equation*}
$$

and therefore, is diagonally dominant. The determinant of a diagonally dominant matrix cannot be zero [117]. Hence, $\mathbf{M}^{\prime}$ is invertible. By extension, $\mathbf{M}$ is invertible.

Now suppose $\nabla R=\mathbf{M x}=\mathbf{0}$. Since $\mathbf{M}$ is invertible, this would imply $\mathbf{x}=\mathbf{0}$. This is a contradiction.

The invexity and incavity of $R$ now follows from Theorem 4.4.1. The fact that $\nabla R \neq \mathbf{0}$ implies that there are no local maxima inside the region $g(\mathbf{p}) \geq 0$.

Remark 9.4.1. Note that it is possible to prove the stronger result that $\nabla R \succ \mathbf{0}$, using continuity and considering a special case, or by other methods. This might shed more light in the geometrical structure of $R$. However, we do not need this result in order to establish the key conclusions of this chapter.

The following lemma is necessary for the next theorem.

Lemma 9.4.1. Suppose $p_{i} \gg I_{i}$ for all $i$. Then when $p_{i}$ is fixed, $R_{p_{i}}$ increases with any $p_{j \neq i} . R_{p_{i}}$ decreases with $p_{i}$ when all the other powers are fixed.

Proof: If $p_{i} \gg I_{i}$, from (9.8) we can approximate $R_{p_{i}}$ as:

$$
\begin{equation*}
R_{p_{i}}=\frac{1}{p_{i}}-\sum_{j \neq i} \frac{a_{j i}}{I_{j}} \tag{9.16}
\end{equation*}
$$

Since $I_{j}$ increases with any $p_{j \neq i}$, so does $R_{p_{i}}$, when $p_{i}$ is fixed.

$$
\begin{equation*}
R_{p_{i} p_{i}}=\frac{-1}{p_{i}^{2}}+\sum_{j \neq i} \frac{a_{j i}^{2}}{I_{j}^{2}} \tag{9.17}
\end{equation*}
$$

Note that $R_{p_{i}}$ and $R_{p_{i} p_{i}}$ have the following forms

$$
\begin{equation*}
R_{p_{i}}=A-\sum_{k} B_{k} \quad R_{p_{i} p_{i}}=-A^{2}+\sum_{k} B_{k}^{2} \tag{9.18}
\end{equation*}
$$

where $A$ and $B_{k}$ 's are positive constants. We know that $R_{p_{i}}>0$ by Theorem 9.4.1 and hence

$$
\begin{equation*}
A-\sum_{k} B_{k}>0 \tag{9.19}
\end{equation*}
$$

Since all $B_{k}$ 's are positive, we have

$$
\begin{equation*}
\sum_{k} B_{k}^{2}<\left[\sum_{k} B_{k}\right]^{2} \tag{9.20}
\end{equation*}
$$

Therefore,

$$
\begin{align*}
R_{p_{i} p_{i}}=-A^{2}+\sum_{k} B_{k}^{2}< & -A^{2}+\left[\sum_{k} B_{k}\right]^{2}  \tag{9.21}\\
& =\left[A+\sum_{k} B_{k}\right]\left[-A+\sum_{k} B_{k}\right]  \tag{9.22}\\
& <0, \quad \text { using }(9.19) \tag{9.23}
\end{align*}
$$

Hence, $R_{p_{i}}$ decreases with $p_{i}$ when all the other powers are fixed.
Note that in CDMA uplinks, the pseudo-orthogonal codes will make sure, unless a very large number of customers are active, that the SINR is good enough to make $p_{i} \gg I_{i}$. The SINR in a Femtocell is typically greater than 15 dB [119]. This is a ratio of more than 30. From (9.1), we have $\operatorname{SINR}_{i}<p_{i} / I_{i}$. Hence, $p_{i} \gg I_{i}$.

Theorem 9.4.2. There exist a unique $\lambda>0$ and a unique point a such that $\nabla R(\mathbf{a})+$ $\lambda \nabla g(\mathbf{a})=\mathbf{0}$ and $g(\mathbf{a})=0$.

Proof: We shall first prove the existence. It is sufficient to prove that there is a point on $g(\mathbf{p})=0$ such that

$$
\begin{equation*}
R_{p_{1}} / g_{1}=R_{p_{2}} / g_{2}=\ldots=R_{p_{N}} / g_{N} \tag{9.24}
\end{equation*}
$$

Start at any power allocation $\mathbf{p}$ that satisfies $g(\mathbf{p})=0$. The $R_{p_{i}} / g_{i}$ values of the users may be at different levels. We now show that by adjusting users' power levels selectively, we can bring the $R_{p_{i}} / g_{i}$ values to the same level. Increase the power of the user with the highest $R_{p_{i}} / g_{i}$ incrementally, while simultaneously reducing the power of the user with the lowest $R_{p_{i}} / g_{i}$ so as to make sure we are still on the plane $g(\mathbf{p})=0$. By Lemma 9.4.1, this will reduce the $R_{p_{i}} / g_{i}$ of the highest user and increase that of the lowest. When some other user's $R_{p_{i}} / g_{i}$ crosses over either user, start working with the new highest and lowest users.

It is possible that at some point in this process the power of the user whose power we are decreasing might hit zero. This means that at the optimal point the power allocation for this user is zero. We remove this user from the system and continue the procedure. Note that during each step, the separation between the highest and the lowest $R_{p_{i}} / g_{i}$ decreases. Therefore, this process will eventually lead to a point where all the remaining users have the same $R_{p_{i}} / g_{i}$ value.

Suppose we are at a point where all $R_{p_{i}} / g_{i}$ are the same. If we increase the power of any user, we have to decrease the power of at least one other user to remain in the plane $g(\mathbf{p})=0$. By Lemma 9.4.1, this means at least two of the $R_{p_{i}} / g_{i}$ values will move in opposite directions.

Therefore, it will not be possible to bring all the $R_{p_{i}} / g_{i}$ values together at a different level.

Note that the proof above offers a partial answer to the "specific question on power allocation" we posed in Section 8.4. Some user's $R_{p_{i}} / g_{i}$ value is too low for it to be able to rise up to the equilibrium level given by (9.24), and these users might end up with zero power. We are still far away from identifying these users without actually going through the optimization process. We need further analysis in this regard, and it is beyond the scope of this article. However, one possible theoretical starting point has just been established.

Theorem 9.4.3. The sum rate $R$ subject to the linear power constraint $g(\mathbf{p}) \geq 0$ is incave in the powers. $R(\mathbf{a})$ is the global maximum of $R$ subject to the constraint $g(\mathbf{p}) \geq 0$ where $\mathbf{a}$ is defined by $\nabla R(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$ and $g(\mathbf{a})=0$.

Proof: Let D be the region of the domain of $R$ restricted to the power constraint plane. In other words, D is defined by $g(\mathbf{p})=0$. According to Theorem 9.4.1, the global maximum of $R$ can only occur on D. Furthermore, Theorem 9.4.1 proved that $\nabla R \neq \mathbf{0}$ and $R$ is incave. Since $g$ is linear, $\nabla g \neq \mathbf{0}$ and $g$ is incave. According to Theorem 9.4.2, $\nabla R(\mathbf{a})$ and $\nabla g(\mathbf{a})$ point in exact opposite directions only at the point $\mathbf{a}$. Then by Theorem 4.5.1, there exists a common $\boldsymbol{\eta}$ with respect to which both $R$ and $g$ are incave in D , except possibly at a. Hence, by Theorem 4.3.2, $R(\mathbf{a})$ is the global maximum of $R$ subject to the constraint $g(\mathbf{p}) \geq 0$.

The fact that " $\nabla[R(\mathbf{a})+\lambda g(\mathbf{a})]=\nabla R(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$ implies $R(\mathbf{a})$ is the global maximum of $R$ restricted to $g(\mathbf{p}) \geq 0$ " shows that $R(\mathbf{a})+\lambda g(\mathbf{a})$ restricted to $g(\mathbf{p}) \geq 0$ is incave, by Theorem 4.3.1. Since $g$ is incave, the linearity property of incavity implies $R$ restricted to $g(\mathbf{p}) \geq 0$ is incave.

### 9.5 Algorithm development

In this section, we develop an algorithm that finds the power allocation that maximizes the sum rate of the GIC under a total power and minimum rate constraints. In the last section, we proved that the maximum occurs at the point given by the KKT condition. $(\nabla R(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0})$. There is a traditional way to design an algorithm that locates this point. Our algorithm however, does not use the traditional approach. Rather, it is based on the new idea of projection of the gradient onto the space defined by the constraints. We will first explain the traditional approach.

### 9.5.1 Extension of traditional algorithms to locate the point of KKT condition

It is important to note here that by traditional algorithms we don't mean any algorithms that solve the sum rate maximization problem in the GIC with total power constraint. As mentioned earlier, to the best of our knowledge, there are no such algorithms. We are here talking about algorithms for other optimization problems where the location of the optimal point is characterized by the KKT condition. Consider such an maximization problem, where the optimal point $\mathbf{a}$ is known to be given by $\nabla f(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$. The algorithm to locate the optimal point a, also has to find the optimal $\lambda$.

If $f(x)$ is convex, then the duality theory from the convex analysis asserts that the optimal $\lambda$ actually minimizes $\nabla f(\mathbf{x})+\lambda \nabla g(\mathbf{x})$ for a fixed $\mathbf{x}[28]$. This duality theory can be extended for invex functions [89]. Assume for a moment that we are taking this approach. Typically, algorithms that use the dual approach perform two tasks during each iteration: a step towards minimizing $\nabla f(\mathbf{x})+\lambda \nabla g(\mathbf{x})$ over $\lambda$ for a fixed $\mathbf{x}$ using steepest gradient, and another step towards maximizing $\nabla f(\mathbf{x})+\lambda \nabla g(\mathbf{x})$ over $\mathbf{x}$ for fixed $\lambda$ using steepest gradient.

The algorithm has to start at some initial values of $\mathbf{x}$ and $\lambda$. Usually any value of $\mathbf{x}$ in the domain would be fine for this purpose. But how do we find a good initial $\lambda$ value?

Furthermore, in our experience, we find such algorithms extremely sensitive to the initial value of $\lambda$, and will fail to converge if the initial value of $\lambda$ is not close enough to the optimal value of $\lambda$ [108].

In what follows, we develop an original algorithm that avoids the need to find $\lambda$ altogether. This not only doubles the speed of the algorithm but avoids any pitfalls regarding convergence.

### 9.5.2 Proposed algorithm to locate the point of KKT Condition - the algorithm for sum rate maximization with total power constraint

The goal is to design an algorithm that starts at any point $\mathbf{p}$ in the plane $g(\mathbf{p})=0$ and finds the point $\mathbf{a}$ on this plane where $\nabla R(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$. Notice that $\mathbf{a}$ is the point on that plane where $\nabla R(\mathbf{p})$ becomes parallel to a normal of the plane $g(\mathbf{p})=0$, namely, $-\nabla g(\mathbf{p})=\left(g_{1}, g_{2}, \ldots g_{N}\right)$. The key idea behind the algorithm is that at this point $\mathbf{a}$, the projection of $\nabla R(\mathbf{p})$ onto the plane $g(\mathbf{p})=0$ is zero. This means we can use a variation of the popular steepest gradient method to find this point. The idea is to use the projection of $\nabla R(\mathbf{p})$ onto the plane $g(\mathbf{p})=0$ instead of $\nabla R(\mathbf{p})$. The required projection can be easily obtained by first finding the projection of $\nabla R(\mathbf{p})$ onto the subspace that is orthogonal to the plane $g(\mathbf{p})=0$, that is the projection of $\nabla R(\mathbf{p})$ in the direction of the normal vector $\left(g_{1}, g_{2}, \ldots g_{N}\right)$, and then subtracting it from $\nabla R(\mathbf{p})$. Let

$$
\begin{equation*}
\mathbf{g}=\left(g_{1}, g_{2}, \ldots g_{N}\right) \tag{9.25}
\end{equation*}
$$

The required projection is

$$
\begin{equation*}
\operatorname{Proj}_{g(\mathbf{p})=0} \nabla R(\mathbf{p})=\nabla R(\mathbf{p})-\frac{\nabla R(\mathbf{p}) \bullet \mathbf{g}}{\|\mathbf{g}\|^{2}} \mathbf{g} \tag{9.26}
\end{equation*}
$$

### 9.5.3 A important note on the origin of the algorithm

After developing the algorithm, we noticed that there is a "theoretical algorithm" called the "projected gradient algorithm" in the convex optimization literature [84]. This algorithm essentially uses the same idea but concerns itself with projecting points onto a convex set. In other words, our algorithm is a special case of this algorithm for linear sets. The projection algorithm in the literature has never been used in practice (to the best of our knowledge) because finding the protection of a point onto a convex set itself is an optimization problem. However, when that set is linear, the projected point can be written in closed form. This is why the special case of the "wider" projection algorithm that we "re-invented" is feasible. In our literature search, we did not find any work that reports the feasibility of the projected algorithm in the linear case.

We now present the algorithm where $s$ is the step size.

## Projected Gradient Algorithm (PGA)

1. $p_{i}:=P /\left(N g_{i}\right)$ for all $i$.
2. $\mathbf{p}:=\mathbf{p}+s \operatorname{Proj}_{g(\mathbf{p})=0} \nabla R(\mathbf{p})$
3. if the change in $R(\mathbf{p})<\delta$ stop.
4. else go to step 2.

### 9.5.4 Algorithm with rate constraints

In Section 8.4, the minimum rate constraints were converted into minimum SINR constraints which were then translated into the following linear inequalities:

$$
\begin{equation*}
\sum_{j=1}^{N} b_{i j} p_{j} \geq c_{i} \text { for } i=1,2, \ldots N \tag{9.27}
\end{equation*}
$$

The goal is to find the power allocation that maximizes the sum rate while satisfying the above inequalities and the linear power constraint:

$$
\begin{equation*}
g(\mathbf{p})=0 \tag{9.28}
\end{equation*}
$$

The first task is to make sure that the algorithm starts at a point $\mathbf{p}_{\text {initial }}$ that falls inside the hyper solid defined by the linear inequalities in (9.27) and the equation in (9.28). At first it might appear that we have to solve this system of inequalities using linear programming. However, this is not necessary as all we need is a single point satisfying this system of inequalities. We can modify the inequalities in (9.27) into equations and solve the following linear system of equalities to obtain $\mathbf{p}_{\text {initial }}$.

$$
\begin{align*}
\sum_{j=1}^{N} b_{i j} p_{j} & =c_{i} \text { for } i=1,2, \ldots N  \tag{9.29}\\
g(\mathbf{p}) & =0
\end{align*}
$$

Consider the point of maximum sum rate subject to the power constraint $g(\mathbf{p})=0$ alone. If this point happens to fall inside the hyper solid defined by the linear inequalities in (9.27), then PGA is all we need, provided that it starts at $\mathbf{p}_{\text {initial }}$ mentioned above. However, this may not always be the case. When the optimum point is outside this solid, PGA will cross over one of the hyperplanes defined by user's minimum rate demands - a hyperplane of the form $\sum_{j=1}^{N} b_{i j} p_{j}=c_{i}$. If we want to satisfy users' rate demands we need to halt PGA
temporarily at this point. We know that the optimum point with the rate constraints lies somewhere on the plane $\sum_{j=1}^{N} b_{i j} p_{j}=c_{i}$. The algorithm needs to travel along the plane $\sum_{j=1}^{N} b_{i j} p_{j}=c_{i}$ to find the optimum point. However, we have to keep in mind that we have to be still on the power constraint plane. In other words, the algorithm has to travel along the intersection of the rate constraint plane $\sum_{j=1}^{N} b_{i j} p_{j}=c_{i}$ and the power constraint plane $g(\mathbf{p})=0$.

At first, it appears that the direction of the intersection of these two planes can be found by taking the cross product of the two normal vectors of the planes. However, this works only in 3 -dimensions. In higher dimensions, there is no cross product of two vectors. This is because, given two non-parallel vectors in higher dimensions, there are an infinite numbers of directions, each of which are perpendicular to any given two vectors.

In dimensions higher than 3 , say in $R^{N}$, the intersection of two hyper planes do form a linear space, and is actually a shifted version of a subspace of $R^{N}$. Note that while a subspace of $R^{N}$ contains the origin, the intersection of our hyper planes is away from the origin. Let us call the linear space formed by the intersection of the hyper planes $\sum_{j=1}^{N} b_{i j} p_{j}=c_{i}$ and $g(\mathbf{p})=0$ by $S_{i}$. The algorithm should travel along the projection of $\nabla R$ onto this linear space $S_{i}$. Since the algorithm, at this point, is actually halted on a point on $S_{i}$, we need not worry about the location; all we need is a direction. This direction can be obtained by considering the associated subspace $S_{i 0}$ that actually contains the origin. In other words, the required projection of $\nabla R$ onto the linear space $S_{i}$ is the same as the projection of $\nabla R$ onto the subspace $S_{i 0}$.

Just like before, this projection of $\nabla R$ onto $S_{i 0}$ can be found by first finding the projection of $\nabla R$ onto the space $S_{i 0}^{\perp}$ that is orthogonal to $S_{i 0}$ and then subtracting it from $\nabla R$. In order to find the projection of $\nabla R$ onto $S_{i 0}^{\perp}$, we need an orthogonal basis for $S_{i 0}^{\perp}$. Since $S_{i 0}$ is the intersection of two hyper planes $g(\mathbf{p})=0$ and $\sum_{j=1}^{N} b_{i j} p_{j}=0$, a basis for $S_{i 0}^{\perp}$ can be
made from the two normal vectors as $\left\{\mathbf{g}, \mathbf{b}_{\mathbf{i}}\right\}$, where $\mathbf{g}$ is given in (9.25) and

$$
\begin{equation*}
\mathbf{b}_{\mathbf{i}}=\left(b_{i 1}, b_{i 2}, \ldots b_{i N}\right) \tag{9.30}
\end{equation*}
$$

An orthogonal basis $\left\{\mathbf{e}_{\mathbf{i} \mathbf{1}}, \mathbf{e}_{\mathbf{i} \mathbf{2}}\right\}$ for $S_{i 0}^{\perp}$ can be found by:

$$
\begin{equation*}
\mathbf{e}_{\mathbf{i} 1}=\mathbf{g} \text { and } \mathbf{e}_{\mathbf{i} 2}=\mathbf{b}_{\mathbf{i}}-\frac{\mathbf{b}_{\mathbf{i}} \bullet \mathbf{g}}{\|\mathbf{g}\|^{2}} \mathbf{g} \tag{9.31}
\end{equation*}
$$

Thus, the projection of $\nabla R$ onto the linear space $S_{i}$ is obtained as:

$$
\begin{equation*}
\operatorname{Proj}_{S_{i}} \nabla R(\mathbf{p})=\nabla R(\mathbf{p})-\frac{\nabla R(\mathbf{p}) \bullet \mathbf{e}_{\mathbf{i} \mathbf{1}}}{\left\|\mathbf{e}_{\mathbf{i} \mathbf{1}}\right\|^{2}} \mathbf{e}_{\mathbf{i} \mathbf{1}}-\frac{\nabla R(\mathbf{p}) \bullet \mathbf{e}_{\mathbf{i} \mathbf{2}}}{\left\|\mathbf{e}_{\mathbf{i} \mathbf{2}}\right\|^{2}} \mathbf{e}_{\mathbf{i} \mathbf{2}} \tag{9.32}
\end{equation*}
$$

While traveling along $S_{i}$ it is possible that the algorithm encounters the rate constraint plane of another user. When this happens the algorithm should change its direction and travel along the space formed by the intersection of the power constraint plane and the two rate constraint planes. In the following, $S_{j}$ represents the space formed by the intersection of the power constraint plane and all the rate constraint planes encountered until and up to the rate constraint plane of User- $j$. We now present the algorithm with rate constraints.

## Projected Gradient Algorithm with

Rate Constraints (PGARC)

1. $T=\{1,2,3, \ldots N\}$
2. Solve (9.29) to obtain $\mathbf{p}_{\text {initial }}$.
3. $\mathbf{p}:=\mathbf{p}_{\text {initial }}$.
4. $\mathbf{D}(\mathbf{p}):=\operatorname{Proj}_{g(\mathbf{p})=0} \nabla R(\mathbf{p})$
5. $\mathbf{p}:=\mathbf{p}+s \mathbf{D}(\mathbf{p})$
6. for $i \in T$
7. if $\operatorname{SINR}_{i} \leq \gamma_{i}$
8. $\quad j:=i$
9. remove $i$ from $T$.
10. $\quad \mathbf{D}(\mathbf{p}):=\operatorname{Proj}_{S_{j}} \nabla R(\mathbf{p})$
11. end if
12. end for
13. if the change in $R(\mathbf{p})<\delta$ stop.
14. else go to step 4.
$T$ is an array that contains the user numbers whose rate constraint planes have not been reached yet. Initially $T$ contains all the integers from 1 to $N . \mathbf{D}(\mathbf{p})$ is the vector that determines the direction of the algorithm. Initially it is set to the projection of $\nabla R(\mathbf{p})$ onto the power constraint plane, and remains so as long as no user's rate constraint plane is encountered. After each update in Step 5, the loop between Steps 6 and 12 checks if any user's rate constraint plane has been reached, and if so updates $T$ in Step 9, and modifies the direction of the algorithm in Step 10.

### 9.5.5 Modification to work with individual power constraints

PGARC can be easily modified to handle individual power constraints instead of a total power constraint. Each individual power constraint adds a hyper plane boundary in the form of a linear inequality to the feasible set. We can combine these inequalities to those from the minimum rate constraints, and use linear programming to find a point $p_{i}$ in the
feasible set. Steps between 4) and 11) can be modified to check if the algorithm crosses over the new hyper planes. Otherwise the algorithm remains the same.

### 9.5.6 Convergence of the algorithms

We need the following Lemma to prove that PGA and PGARC converge to the respective global maximum sum rates.

Lemma 9.5.1. The direction of the fastest increase of any function $f$ restricted to a linear space is the projection of $\nabla f$ onto that linear space, evaluated on that linear space.

Proof: We restrict our attention to three variable functions but the proof easily extends to functions of any number of variables. We will first deal with the case where the linear space is the $X Y$-plane.

Consider a three variable scalar function $f(x, y, z)$. We know that $\nabla f=\left(f_{x}, f_{y}, f_{z}\right)$ gives the direction of the fastest increase in $f$, when there are no restrictions on $x, y, z$. Suppose $x, y, z$ are restricted to the $X Y$-plane. Then $f$ becomes $g(x, y)=\left.f\right|_{z=0}=f(x, y, 0)$, and the direction of the fastest increase of $f$ restricted to the $X Y$-plane will be given by $\nabla g=\left(g_{x}, g_{y}\right)$. But it is easily seen that $g_{x}=\left.f_{x}\right|_{z=0}$ and $g_{y}=\left.f_{y}\right|_{z=0}$ and therefore, $\nabla g=\left(\left.f_{x}\right|_{z=0},\left.f_{y}\right|_{z=0}\right)$. However, $\left(\left.f_{x}\right|_{z=0},\left.f_{y}\right|_{z=0}\right)$ is nothing but the projection of $\nabla f=\left(f_{x}, f_{y}, f_{z}\right)$ onto the $X Y$ pane evaluated on the $X Y$-plane. Note that the same arguments hold when the linear space is the line formed by the $X$-axis.

By rotating the axes, we can extend the proof to include any inclined linear space.
Theorem 9.5.1. PGA and PGARC converge to the points of global maximum sum rate of the GIC subject to the respective constraints.

Proof: PGA converges to the global maximum point for the same reason as to why the steepest gradient algorithm (SGA) converges to the global maximum point. Consider first, a concave function $f(\mathbf{x})$ and the SGA that finds the point at which $f(\mathbf{x})$ attains its
maximum. The SGA starts at any point $\mathbf{x}$ and updates its position by:

$$
\begin{equation*}
\mathbf{x}:=\mathbf{x}+s \nabla f(\mathbf{x}) \tag{9.33}
\end{equation*}
$$

These updates eventually lead to the optimal point because two conditions are fulfilled. First is that at any point $\mathbf{x}$, the gradient $\nabla f(\mathbf{x})$ gives the direction in which $f(\mathbf{x})$ increases the fastest. This guarantees that the updates take the algorithm not only in the direction in which $f(\mathbf{x})$ increases but in the most efficient route towards the maximum. The second is that near the maximum point, $\nabla f(\mathbf{x})$ becomes smaller and smaller and eventually becomes zero. This makes sure that the algorithm hits the point of maximum and stays there or oscillates about the maximum point with an amplitude determined by the step size $s$.

We already proved that the maximum of $R(\mathbf{p})$ restricted to $g(\mathbf{p})=0$ occurs at the point given by $\nabla R(\mathbf{a})+\lambda \nabla g(\mathbf{a})=\mathbf{0}$, or at the point where $\operatorname{Proj}_{g(\mathbf{p})=0} \nabla R(\mathbf{p})=0$. Since the updates in our algorithm PGA use $\operatorname{Proj}_{g(\mathbf{p})=0} \nabla R(\mathbf{p})$, the second condition mentioned above is satisfied.

By Lemma 9.5.1, $\operatorname{Proj}_{g(\mathbf{p})=0} \nabla R(\mathbf{p})$ represents the direction of the fastest increase in $R(\mathbf{p})$, when $\mathbf{p}$ is restricted to $g(\mathbf{p})=0$. This shows that the first condition is satisfied too, and hence, PGA converges to the global maximum point.

The same argument applies to PGARC with $g(\mathbf{p})=0$ replaced with the "intersection of $g(\mathbf{p})=0$ and the rate constraint planes."

### 9.5.7 Computational complexity of PGARC

Note that the part of PGARC from steps 7) to 11) is identical to the steepest gradient algorithm but with a different gradient. The steepest gradient algorithm has geometrical convergence, and this does not depend on the number of variables in the problem [28]. Hence, the time taken from steps 7) to 11) does not depend on $N$. In the worst case, the steps between 7) and 11) may have to be executed $N$ times. In Step 2), we have to solve an
$N \times N$ linear system. This takes a time in the order of $N^{3}$ [28]. Steps 4) and 5) are part of the steepest gradient algorithm but with a different gradient. Therefore, PGARC has a time complexity $O\left(N^{3}\right)+O(N)$ which simplifies to $O\left(N^{3}\right)$.

The modified algorithm with individual power constraints has to solve a $2 N \times N$ linear programming problem in step 2). This will take a time in the order of $2 N^{3}$. The algorithm has to check $2 N$ number of planes between the steps 7 ) and 11). Hence, the complexity here will be $O\left(2 N^{3}\right)+O(2 N)$ which simplifies to $O\left(N^{3}\right)$ as well.

### 9.6 Numerical results

In this section, we provide considerable numerical evidence regarding three matters. The proposed algorithms converge; they converge to the global maximum sum rate; they do so with remarkable speed. The details of the simulation parameters are given in Fig. 9.3. The first test of any algorithm should involve a case where we already know the solution. With this in mind, in the first few scenarios we calculate the channel gains using path loss only, from known distances so as to keep the channel gains predictable. The later scenarios gradually allow for random variation in distances, Rayliegh fading and log-normal-shadowing.

We start with Scenario-1 in Fig. 9.3, which is a completely symmetrical system with a total power constraint. We know that the maximum sum rate for this case should occur when the power levels of the 4 users are equal. We deliberately start the PGA with unequal power levels. Fig. 9.4 shows the convergence of the sum rate. Fig. 9.5 shows the individual power levels converging to the same level.

In Scenario-2 (Fig. 9.3), we test the PGA with the interference constraint. This system is partially symmetrical; the first two users face identical conditions and so do the last two users. Hence, at the maximum sum rate, the first two users should have the same power level and so do the last two users. For this Scenario, we started the algorithm with $p_{i}=P /\left(N g_{i}\right)$. Fig. 9.6 shows the convergence of the power levels to what we expected.


Figure 9.3: Simulation details.


Figure 9.4: Scenario-1: The sum rate in the symmetrical case.


Figure 9.5: Scenario-1: The power levels in the symmetrical case.


Figure 9.6: Scenario-2: The power levels in the partially symmetrical case.


Figure 9.7: Scenario-3: The sum rate for the example in Section 8.4.


Figure 9.8: Scenario-3: The power levels for the example in Section 8.4.


Figure 9.9: Scenario-4: The sum rate of eight users


Figure 9.10: Scenario-4: The power levels of eight users

Scenario-3 is the same as the example we considered in Section 8.4. The exhaustive search we conducted at that time demonstrated that at the maximum sum rate, the power level of User-4 must be zero. The sum rate and power levels obtained through the algorithm PGA are shown in Fig. 9.7 and Fig. 9.8. As expected, the first three users get equal power levels, while the last user gets nothing. The maximum sum rate obtained using PGA shown in Fig. 9.7 ( $25.32 \mathrm{bits} / \mathrm{s} / \mathrm{Hz}$ ) agrees with the one obtained using an exhaustive search (a bit more than $25 \mathrm{bits} / \mathrm{s} / \mathrm{Hz}$ ) shown in Fig. 9.2 of Section 8.4. The bend in iteration number 8 in Fig. 9.8 is the result of the algorithm getting halted when the power level of User-4 is about to cross over zero. Instead of negative power, User-4 gets zero power, which is slightly "higher" than what the algorithm is about to give it. This in turn results in the upward bend in the sum rate curve in Fig. 9.7 at iteration number 8 .

In Scenario-4, we increase the number of users, allow some randomness in the interfering distances and include fading and shadowing in the calculation of the channel gains. The convergence of the sum rate is shown in Fig. 9.9 and the power levels in Fig. 9.10. The last
three users get no power. This should not be surprising given their channel conditions.
Scenario- 5 allows randomness in the direct as well as the interfering distances. The minimum rate constraints are taken into consideration. Fading and shadowing are superimposed on the path loss to obtain the channel gains. Fig. 9.11 shows the convergence of the sum rate as well as the individual transmission rates and the sum rate at important points during the execution of the algorithm PGARC. The numbers in the first row gives the rates obtained at Step-2 of the algorithm when the linear system is solved to find a point in the feasible set. The second row gives the rates when the algorithm is about to cross the hyper plane representing User-1's minimum rate demand. From this point on, the algorithm travels on the intersection of the power constraint plane and the afore mentioned rate demand plane of User-1, until it is about to cross the rate demand plane of User-3. The rates at this point is given in the third row. Hereafter, the algorithm travels on the intersection of three hyper planes given by: the total power constraint, the rate demand of User-1 and the rate demand of User-3.

Any comparison of our algorithm with an algorithm that considers individual power constraints by dividing the total power constraint $P$ into $N$ equal individual power constraints would be unfair to the later algorithm as the maximum sum rate might occur at a point where one or more users are allocated a power that is more than $P / N$. With this in mind, in Scenario-6, we compare the sum rate obtained using PGARC to that obtained using the only other algorithm that uses a total power constraint, albeit a different rate constraint - proportional rate constraints. This is the central algorithm from the last chapter. We understand this could be construed as unfair to the algorithm from the last chapter. We emphasize here that it is not our intention to compare the performances of the two algorithms. Rather, we would like to illustrate the advantages of an algorithm that considers minimum rate constraints over one that considers proportional rates when it comes to satisfying QoS requirements. We set the QoS requirements via the minimum rate demands shown in Fig. 9.3 and made sure the problem is feasible by running PGARC. Then we executed the algo-
rithm from the last chapter with the proportional rates in the same ratio as the minimum rate demands. As shown in Table 9.2, the algorithm from the last chapter also satisfies the minimum rate demands. However, PGARC achieves a higher sum rate.


Figure 9.11: Scenario-5: The sum rate, and the individual rates at key points.

Table 9.2: Scenario-6: Comparison

|  | Individual rates <br> $(\mathrm{bits} / \mathrm{s} / \mathrm{Hz})$ | Sum rate <br> $(\mathrm{bits} / \mathrm{s} / \mathrm{Hz})$ |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Proportional |  |  |  |  |
| rate | 1.779 | 2.135 | 2.490 |  |
| algorithm (Chapter 8) | 2.846 | 3.202 | 3.558 | 16.010 |
| Proposed <br> minimum rate <br> algorithm | 1.121 | 1.852 | 2.351 |  |
|  | 3.731 | 6.495 | 4.908 | 20.458 |

### 9.7 Conclusion

The chapter used a new type of geometrical structure called invexity (incavity) to characterize the sum rate of the GIC. It was established that while the sum rate of the GIC is both incave and invex, the sum rate restricted to a linear constraint on the powers is incave. One of the KKT conditions was then used to locate the power allocation that maximizes the sum rate subject to the said constraint. The existence and uniqueness of a point on the power constraint plane that satisfies the KKT condition was proven. An algorithm that maximizes the sum rate of the GIC subject to total power and QoS constraints is proposed. Numerical results demonstrating the performance of the algorithm were also presented.

## Chapter 10

## Fundamental aspects of energy efficient power allocation in OFDMA

### 10.1 Introduction

This chapter fills a gap in the literature regarding the foundations of energy efficient power allocation for orthogonal frequency division multiple access (OFDMA). It is shown that for fixed subchannel assignment, the power allocation that maximizes the energy efficiency (EE) of OFDMA transmissions can be obtained through a single water level that can be calculated by solving a single non-linear equation. What is more, this equation is the same for single user, single carrier case or multi user, multi carrier case. The maximum EE is shown to have an inverse relationship to the said water level. The theoretical results obtained in this chapter are used in the next chapter to find a two step solution to the problem of determining the frequency and power allocation that maximizes the EE of OFDMA.

The approach we take in this chapter to arrive at the results is rather unique. In Chapter 2, we mentioned that the KKT equations that define the critical point of a constrained optimization problem, in general, are difficult to solve. Yet, this is the approach that we take in this chapter because the alternative, using iterative algorithms to arrive at the optimum
power allocation, does not help with the other part in the resource allocation problem - the frequency assignment. To the best of our knowledge, this is the first time KKT equations from a fairly complex optimization problem have been solved analytically to express the optimum point as a solution to a single equation.

For alternative approaches to obtaining the power allocation that maximizes the energy efficiency, namely using iterative algorithms, the reader is referred to our works that use fractional programming in cooperative cognitive radios [142-146].

### 10.2 Background

The need to reduce the carbon footprint of mobile communications together with the increasing demand for data rates has necessitated research into the energy efficiency of the transmissions from base stations [17], [68]. Power allocation that maximizes the bits/Joule/Hz energy efficiency (EE) of orthogonal frequency division multiple access (OFDMA) based transmissions has been studied well [25,67-72]. It is known that for a fixed channel assignment, the power allocation that maximizes the EE, without any QoS or total power constraints, is given by a common water level for all users [25,69-72]. The best approach among these work uses Charnes-Cooper Transformation (CCT) to change the quasi-concave EE objective function into a concave function. The common water level is obtained using iterative algorithms.

Consider the bigger problem of frequency and power allocation that maximizes the EE. This problem is too complex and we can only consider two step near optimal solutions - a near optimal frequency assignment followed by optimal power allocation. In order to assign subchannels optimally, we need theoretical insights into the role the channel assignment plays in the maximum EE attainable from a given subchannel assignment. This insight is possible only if we have a closed form solution to the maximum EE in terms of the channel gains. If this is not possible then the next best thing is an approximate expression for the maximum EE. But none of this is possible with a solution based on iterative algorithms. It
is because of this reason that we decide to attack the KKT equations face on.
The next section presents the system model and the problem statements. The following sections develop the fundamentals of energy efficiency maximizing power allocation for OFDMA. Section 10.4 deals with the single user, single carrier case. Single user, multi carrier case (OFDM) is considered in Section 10.5, and multiuser, multi carrier case (OFDMA) with proportional rate constraints in Section 10.6. Multiuser, multi carrier case (OFDMA) with minimum rate constraints is considered in Section 10.7 followed by Multiuser, multi carrier case (OFDMA) with a total power constraint in Section 10.8. Section 10.9 presents the algorithm that maximizes the EE, with minimum rate constraints, for a given channel assignment. The complexity of this algorithm is discussed in Section 10.10, and the chapter concludes in Section 10.11.

### 10.3 System model and problem statement

Consider the downlink of a single cell with $N$ users and $K$ orthogonal subchannels. Each subchannel is assigned exclusively to one user. If the channel gain on the $k$ th subchannel is $a_{k}$, the transmission power $p_{k}$, the noise spectral density $\sigma^{2}$, and $h_{k}=a_{k} / \sigma^{2}$, then the system EE of the transmissions over all $K$ subchannels in bits/Joule/ Hz can be written as:

$$
\begin{equation*}
E E=\frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}} \tag{10.1}
\end{equation*}
$$

where $\psi$ is the reciprocal of the efficiency and $p_{c}$ is the circuit power [25] of the downlink transmitter.

Assume a yet-to-be determined subchannel assignment protocol is used to distribute the subchannels among the users. Suppose a total of $K_{1}$ subchannels - subchannel 1 through subchannel $k_{1}$ - are assigned to User-1. A total of $K_{2}$ subchannels - subchannel $k_{1}+1$ through subchannel $k_{2}$ - are assigned to User- 2 and so on. A total of $K_{n}$ subchannels - subchannel
$k_{n-1}+1$ through subchannel $k_{n}$ - are assigned to User- $n$. Then the rate $r_{n}$ of User- $n$ can be written as:

$$
\begin{equation*}
r_{n}=\sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right), \tag{10.2}
\end{equation*}
$$

where $k_{0}=0$. If the minimum rate acceptable to User- $n$ is $R_{n}$ then we must have

$$
\begin{equation*}
r_{n}-R_{n} \geq 0 \text { for } n=1,2, \ldots, N \tag{10.3}
\end{equation*}
$$

The goal is to maximize the EE in (10.1) subject to (10.3) and a total power constraint $P$ using a two step optimization procedure - first using the GA for subchannel assignment and then solving the following power allocation problem:

$$
\max _{p_{1}, p_{2}, \ldots p_{K}} E E
$$

subject to

$$
\begin{align*}
& \mathrm{C} 1: r_{n}-R_{n} \geq 0 \text { for } n=1,2, \ldots, N .  \tag{10.4}\\
& \mathrm{C} 2: P-\sum_{k=1}^{K} p_{k} \geq 0 .
\end{align*}
$$

The material developed in this chapter is also sufficient to solve the EE maximizing power allocation problem with proportional rate constraints. That problem can be formally described by replacing the constraint C 1 above with C 3 below:

$$
\begin{equation*}
\mathrm{C} 3: \alpha_{n+1} \sum_{k=1}^{k_{1}} \log _{2}\left(1+h_{k} p_{k}\right)-\sum_{k=k_{n}+1}^{k_{n+1}} \log _{2}\left(1+h_{k} p_{k}\right)=0, \text { for } n=1,2,3, \ldots, N-1 . \tag{10.5}
\end{equation*}
$$

### 10.4 The single user, single carrier Case

In this Section we analyze the EE of a single user. Even though the results of this section can be obtained without introducing a water level, we do introduce the water level so that
we can compare the results here with results from the multi-carrier and multi user cases. If the channel gain is $a$, the transmission power $p$, the noise spectral density $\sigma^{2}$, and $h=a / \sigma^{2}$, then the EE of the transmission in bits/Joule/Hz can be written as:

$$
\begin{equation*}
E E=\frac{\log _{2}(1+h p)}{p_{c}+\psi p} \tag{10.6}
\end{equation*}
$$

where $\psi$ is the reciprocal of the efficiency of the power amplifier and $p_{c}$ is the circuit power [25] of the transmitter. The circuit power represents the total power spent on preprocessing the signal before it enters the power amplifier. We wish to find the $p$ that maximizes the EE for fixed $h, \psi$, and $p_{c}$. The optimization problem here is:

$$
\begin{equation*}
\max _{p} E E=\frac{\log _{2}(1+h p)}{p_{c}+\psi p} \tag{10.7}
\end{equation*}
$$

The fact that the EE here is not concave in the power $p$ does not trouble us because we have a single variable. The existence of a global maximum can be established using simple analysis. We summarize the results in the following Theorem.

Theorem 10.4.1. The single user EE has only one stationary point and this point is a global maximum. The stationary point can be obtained by solving the equation $E E(w)=1 / \psi w \ln 2$, where $w$ is the water level. The maximum EE has an inverse relationship to the water level that produces it and a direct relationship to the channel gain. In other words, the maximum $E E$ is higher if the channel gain is higher.

Proof: Differentiating EE with respect to $p$ :

$$
\begin{equation*}
\operatorname{EE}^{\prime}(p)=\frac{f(p)}{\left(p_{c}+\psi p\right)^{2}} \tag{10.8}
\end{equation*}
$$

where

$$
\begin{equation*}
f(p)=\frac{h}{1+h p}\left(p_{c}+\psi p\right)-\psi \ln (1+h p) \tag{10.9}
\end{equation*}
$$

Note that $\lim _{p \rightarrow 0^{+}} f(p)=h p_{c}>0$ and $\lim _{p \rightarrow \infty} f(p)=\psi-\infty<0$. Since $f(p)$ is continuous for $p>0$, by intermediate value theorem, there is at least one $p$ such that $f(p)=0$.

Suppose there are two such $p$ 's. Since $f(p)$ is differentiable, by Rolle's theorem, $f^{\prime}(p)=0$ for some $p>0$. However, it can be shown that

$$
\begin{equation*}
f^{\prime}(p)=-\frac{h^{2}\left(p_{c}+\psi p\right)}{(1+h p)^{2}}>0 \text { for all } p>0 \tag{10.10}
\end{equation*}
$$

which is a contradiction.
Thus, $\mathrm{EE}(p)$ has only one stationary point. Note further that $\mathrm{EE}(p)$ is positive, differentiable, $\lim _{p \rightarrow 0^{+}} \mathrm{EE}(p)=0$ and $\lim _{p \rightarrow \infty} \mathrm{EE}(p)=0$. Therefore, this stationary point must be a global maximum.

From (10.9), we see that the stationary point is given by

$$
\begin{equation*}
\frac{h}{1+h p}=\psi \frac{\ln (1+h p)}{p_{c}+\psi p} . \tag{10.11}
\end{equation*}
$$

Introducing the water level

$$
\begin{equation*}
w=p+\frac{1}{h} \tag{10.12}
\end{equation*}
$$

on the left only, and rewriting:

$$
\begin{equation*}
\mathrm{EE}=\frac{1}{\psi w \ln 2} \tag{10.13}
\end{equation*}
$$

That is, the maximum EE is inversely proportional to the water level that produces it. Since the EE on the left side of (10.13) can also be written in terms of the water level $w$, this equation can be solved to obtain the optimum water level $w^{*}$.

We now investigate the relationship of the channel gain $h$ to the maximum EE. Writing the EE on the left side of (10.13) in terms of the water level $w$ and re-arranging:

$$
\begin{equation*}
p_{c}+\psi w-\frac{\psi}{h}=\psi w \ln w+\psi w \ln h \tag{10.14}
\end{equation*}
$$

Implicitly differentiating this equation with respect to $h$, it can be shown that

$$
\begin{equation*}
\frac{d w}{d h}=\frac{-p}{h \ln (1+h p)}<0 \tag{10.15}
\end{equation*}
$$

This shows that the water level that produces the maximum EE decreases with the channel gain $h$. We earlier showed that the maximum EE is inversely proportional to the water level that produces it. Combining these two results, we conclude that the maximum EE is increasing with channel gain.

### 10.5 Single user, multi carrier case (OFDM)

Consider a single user transmitting over $K$ subchannels. If the channel gain on the $k$ th subchannel is $a_{k}$, the transmission power on the $k$ th subchannel is $p_{k}$, and $h_{k}=a_{k} / \sigma^{2}$, the optimization problem we are interested is:

$$
\begin{equation*}
\max _{\mathbf{p}} E E=\frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}}, \tag{10.16}
\end{equation*}
$$

where $\mathbf{p}=\left(p_{1}, p_{2}, \ldots p_{K}\right)$.
The numerator of our objective function is concave in $\mathbf{p}$, while the denominator is affine. Such a fraction may not be concave. However, as mentioned in Chapter 5 this fraction can be transformed in to a concave function using Charnes-Cooper transformation (CCT).

Theorem 10.5.1. The power levels that maximize the total EE can be obtained through a water-filling-like solution with a common water level on all the carriers. This water level can
be obtained by solving a single non-linear equation:

$$
\begin{equation*}
E E(w)=\frac{1}{\psi w \ln 2} \tag{10.17}
\end{equation*}
$$

The maximum $E E$ is inversely proportional to the water level that produces it. Anytime a subchannel is replaced with another with a greater channel gain, the maximum EE increases.

Proof: We apply CCT transformation (Chapter 5) to the optimization problem in (10.16) by letting

$$
\begin{equation*}
t=\frac{1}{p_{c}+\psi \sum_{k=1}^{K} p_{k}} \quad \text { and } \quad y_{k}=t p_{k} \tag{10.18}
\end{equation*}
$$

to obtain the new optimization problem:

$$
\begin{align*}
& \max _{\mathbf{y} t} \quad f(\mathbf{y}, t)=t \sum_{k} \log _{2}\left(1+h_{k} y_{k} / t\right) \\
& \text { subject to }  \tag{10.19}\\
& \mathrm{C} 1: \psi \sum_{k} y_{k}+p_{c} t-1=0 \\
& \mathrm{C} 2: t>0
\end{align*}
$$

### 10.5.1 Why the inequality $t>0$ can be ignored

This inequality $t>0$ in constraint C 2 will be a recurring theme in this chapter. However, it does not affect the results in any meaningful manner. In order to avoid unnecessary complexity in the derivation, we would like to ignore this constraint in the future. We first show why constraint C 2 can be ignored.

The strict inequality in the constraint $\mathrm{C} 2: t>0$ is not part of the convex optimization theory. We can change this constraint into $\mathrm{C} 2: t-\delta \geq 0$ with the understanding that at the end of the analysis we let $\delta \longrightarrow 0$.

$$
\max _{\mathbf{y} t} \quad f(\mathbf{y}, t)=t \sum_{k} \log _{2}\left(1+h_{k} y_{k} / t\right)
$$

subject to

$$
\begin{align*}
& \mathrm{C} 1: \psi \sum_{k} y_{k}+p_{c} t-1=0  \tag{10.20}\\
& \mathrm{C} 2: t-\delta \geq 0
\end{align*}
$$

The Lagrangian can be written as follows, where $\lambda$ and $\mu$ are the dual variables.

$$
\begin{equation*}
L(\mathbf{y}, t, \lambda, \mu)=f+\lambda\left(\psi \sum_{k} y_{k}+p_{c} t-1\right)+\mu(t-\delta) \tag{10.21}
\end{equation*}
$$

The KKT conditions:

$$
\begin{align*}
\frac{\partial f}{\partial y_{k}}+\lambda \psi & =0  \tag{10.22a}\\
\frac{\partial f}{\partial t}+\lambda \psi p_{c}+\mu & =0  \tag{10.22b}\\
\psi \sum_{k} y_{k}+p_{c} t-1 & =0  \tag{10.22c}\\
\mu(t-\delta) & =0  \tag{10.22d}\\
t-\delta & \geq 0 \tag{10.22e}
\end{align*}
$$

Note that (10.22d) implies either $\mu=0$ or $t=\delta$.
Suppose $t=\delta$. Since we were going to make $\delta \longrightarrow 0$, this implies that the solution occurs as $t \longrightarrow 0$. Recalling that

$$
\begin{equation*}
t=\frac{1}{p_{c}+\psi \sum_{k} p_{k}} \tag{10.23}
\end{equation*}
$$

we conclude that the solution occurs as $\sum_{k} p_{k} \longrightarrow \infty$. This will happen only if some of the $p_{k}$ 's goes to infinity. We can stop here and go to the alternate case that $\mu=0$, because
infinite power is not feasible.
In the following, we go one step further and prove that the solution we obtain for the optimum energy efficiency, from the case that some of the $p_{k}$ 's goes to infinity, is actually zero.

Let us put the $k$ 's for which $p_{k} \longrightarrow \infty$ into the set $S$ and the rest of the $k$ 's into set $S^{c}$. The energy efficiency $f$ can be written as:

$$
\begin{equation*}
f=\frac{\sum_{k \in S} \log _{2}\left(1+h_{k} p_{k}\right)+\sum_{k \in S^{c}} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k \in S} p_{k}+\psi \sum_{k \in S^{c}} p_{k}} \tag{10.24}
\end{equation*}
$$

We wish to calculate the limit of $f$ as those $p_{k}$ 's that go to infinity approach infinity.
Note that both in the numerator and the denominator, the summations with the $k \in S$ terms dominate. After dividing the numerator and the denominator by $\sum_{k \in S} p_{k}$ and omitting the terms that approach zero,

$$
\begin{equation*}
\operatorname{limit}_{p_{k} \longrightarrow \infty} f=\operatorname{limit}_{p_{k} \longrightarrow \infty} \frac{\sum_{k \in S} \log _{2}\left(1+h_{k} p_{k}\right)}{\psi \sum_{k \in S} p_{k}} \tag{10.25}
\end{equation*}
$$

We know that all the $p_{k}$ 's here are approaching infinity. Among this $p_{k}$ 's there will be one $p_{i}$ that goes to infinity at a speed equal to or greater than the others. Then

$$
\begin{equation*}
\operatorname{limit}_{p_{k}, p_{i} \longrightarrow \infty} \frac{p_{k}}{p_{i}}=c_{k}, \tag{10.26}
\end{equation*}
$$

where $c_{k}$ is either zero or a constant. By dividing the numerator and the denominator of $f$ by this $p_{i}$,

$$
\begin{equation*}
\operatorname{limit}_{p_{k}, p_{i} \longrightarrow \infty} f=\operatorname{limit}_{p_{k}, p_{i} \longrightarrow \infty} \frac{\frac{\log _{2}\left(1+h_{i} p_{i}\right)}{p_{i}}+\frac{p_{k}}{p_{i}} \sum_{k} \frac{\log _{2}\left(1+h_{k} p_{k}\right)}{p_{k}}}{\psi\left(1+\sum_{k} p_{k} / p_{i}\right)}=\frac{0+c_{k} \cdot 0}{\psi\left(1+\sum_{k} c_{k}\right)}=0 \tag{10.27}
\end{equation*}
$$

Because, by a famous theorem, $\log x$ goes to infinity at a slower speed than $x$. Thus, we
conclude that

$$
\begin{equation*}
\operatorname{limit}_{p_{k} \longrightarrow \infty} f=0 \tag{10.28}
\end{equation*}
$$

The above argument shows that if we go with the case in the KKT equations that $t=\delta$ then the solution for the optimal energy efficiency is zero. (10.22d) implies that if $t \neq \delta$ then $\mu=0$. This is the other case, and this case essentially eliminates the last term in the Lagrangian in (10.21). This means that in this case, we can eliminate the constraint C 2 : $t>0$ from our optimization problem.

In what follows we will always go with the later case, ignore the constraint C 2 : $t>0$ from the optimization problem, and try to obtain a positive solution for the optimal energy efficiency. As long as we obtain a positive solution for the optimal value, that must be the actual maximum value, because the other solution is zero.

We re-write the optimization problem 10.20 without C 2 and proceed.

$$
\begin{align*}
& \max _{\mathbf{y} t} \quad f(\mathbf{y}, t)=t \sum_{k} \log _{2}\left(1+h_{k} y_{k} / t\right) \\
& \text { subject to }  \tag{10.29}\\
& \mathrm{C}: \psi \sum_{k} y_{k}+p_{c} t-1=0
\end{align*}
$$

The Lagrangian is formed as:

$$
\begin{equation*}
L(\mathbf{y}, t, \lambda)=f+\lambda\left[\psi \sum_{k} y_{k}+p_{c} t-1\right] \tag{10.30}
\end{equation*}
$$

The KKT conditions:
The Constraint C:

$$
\begin{equation*}
\psi \sum_{k} y_{k}+p_{c} t-1=0 \tag{10.31}
\end{equation*}
$$

The $t$ derivative of the Lagrangian in (10.30):

$$
\begin{equation*}
\sum_{k=1}^{K}\left[\frac{\ln \left(1+h_{k} p_{k}\right)}{\ln 2}-\frac{h_{k} p_{k}}{\left(1+h_{k} p_{k}\right) \ln 2}\right]+\lambda p_{c}=0 \tag{10.32}
\end{equation*}
$$

The $y_{k}$ derivatives of the Lagrangian in (10.30):

$$
\begin{equation*}
\frac{h_{k}}{\left(1+h_{k} p_{k}\right) \ln 2}+\lambda \psi=0, \quad \text { for } k=1,2, \ldots, K \tag{10.33}
\end{equation*}
$$

Eliminating $\lambda \psi$ from the $K$ equations in (10.33):

$$
\begin{equation*}
\frac{1+h_{1} p_{1}}{h_{1}}=\frac{1+h_{2} p_{2}}{h_{2}}=\frac{1+h_{3} p_{3}}{h_{3}}=\ldots=\frac{1+h_{K} p_{K}}{h_{K}} \tag{10.34}
\end{equation*}
$$

Denoting the common number above by $w$ :

$$
\begin{equation*}
\frac{1+h_{1} p_{1}}{h_{1}}=\frac{1+h_{2} p_{2}}{h_{2}}=\frac{1+h_{3} p_{3}}{h_{3}}=\ldots=\frac{1+h_{K} p_{K}}{h_{K}}=w \tag{10.35}
\end{equation*}
$$

This shows that the power levels that maximize the $E E$ are given by a common water level:

$$
\begin{equation*}
p_{k}=w-\frac{1}{h_{k}} \text { for } k=1,2, \ldots, K \tag{10.36}
\end{equation*}
$$

We now concentrate on finding this water level. Using (10.36) to eliminate $p_{k}$ from (10.32) and (10.33), we obtain two new equations.

$$
\begin{align*}
\sum_{k=1}^{K}\left[\frac{\ln \left(w h_{k}\right)}{\ln 2}-\frac{w h_{k}-1}{w h_{k} \ln 2}\right]+\lambda p_{c} & =0  \tag{10.37a}\\
\frac{1}{w \ln 2}+\lambda \psi & =0 \tag{10.37b}
\end{align*}
$$

Multiplying the $k$ th equation in (10.33) by $p_{k}$, and adding all $K$ equations:

$$
\begin{equation*}
\sum_{k=1}^{K}\left[\frac{h_{k} p_{k}}{\left(1+h_{k} p_{k}\right) \ln 2}\right]+\lambda \psi \sum_{k=1}^{K} p_{k}=0 \tag{10.38}
\end{equation*}
$$

Adding this equation to (10.32):

$$
\begin{equation*}
\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)+\lambda\left(p_{c}+\psi \sum_{k=1}^{K} p_{k}\right)=0 \tag{10.39}
\end{equation*}
$$

which implies that

$$
\begin{equation*}
f=\mathrm{EE}=-\lambda \tag{10.40}
\end{equation*}
$$

Using (10.37b) now,

$$
\begin{equation*}
\mathrm{EE}(w)=\frac{1}{\psi w \ln 2} \tag{10.41}
\end{equation*}
$$

This is exactly the same equation as (10.13) in the single carrier case. The maximum EE is inversely proportional to the water level that produces it. After writing the EE in terms of the common water level $w$, this equation can be used to solve for $w$. Since the optimization problem is concave, we expect only one solution to this equation. However, it is possible to prove that this equation has a unique solution by independent means just as in the single user case. Also, it is straight forward to proceed like in the single user case and obtain the following.

$$
\begin{equation*}
\frac{\partial w}{\partial h_{k}}=\frac{-h_{k}}{h_{k} \ln \left(1+h_{k} p_{k}\right)}<0 \text { for all } k . \tag{10.42}
\end{equation*}
$$

Combining this with the result that the maximum EE is inversely proportional to the water level, we conclude that any time a subcarrier with a greater channel gain is substituted for one with a lower channel gain, the maximum EE increases.

### 10.6 Multi user multi carrier case (OFDMA) with proportional rate requirements

Note that without any rate requirements from the users, the power allocation that maximizes the EE in this case will be identical to the last section. With this in mind in this section
we include proportional rate requirements. As mentioned earlier, the main reason for this analysis is to gain theoretical understanding of the nature of the EE maximizing power allocation. The knowledge gained in this section is used in the next chapter when using genetic algorithm for subchannel assignment. A summary of the notations is given in Table 10.1.

| $N$ | number of users in the system |
| :---: | :---: |
| $K$ | number of subchannels in the system |
| $a_{k}$ | channel gain on subchannel $k$ experienced by the user to whom it is assigned |
| $\sigma_{k}^{2}$ | additive white Gaussian noise density |
| $h_{k}$ | $a_{k} / \sigma_{k}^{2}$ |
| $p_{k}$ | power used on subchannel $k$ |
| $p_{c}$ | circuit power |
| $\psi$ | reciprocal of the power amplifier efficiency |
| $\alpha_{n}$ | User- $n$ 's transmission rate as a ratio of User-1's transmission rate |

Table 10.1: Notations

There are $N$ users and $K$ subcarriers. A total of $K_{1}$ subcarriers - starting from subcarrier 1 through subcarrier $k_{1}$ are assigned to User-1. A total of $K_{2}$ subcarriers - starting from subcarrier $k_{1}+1$ through subcarrier $k_{2}$ are assigned to User-2. In general, a total of $K_{n}$ subcarriers - starting from subcarrier $k_{n-1}+1$ through subcarrier $k_{n}$ are assigned to User-n. User-n's rate demand is $\alpha_{n}$ times that of User- 1 . For notational simplicity we use $\alpha_{1}$ as well, with the assumption that $\alpha_{1}=1$. We use the short hand $l$ to denote $\ln 2$.

The optimization problem:

$$
\max _{\mathbf{p}} \frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}}
$$

subject to

$$
\begin{equation*}
\mathrm{C} 1: \alpha_{n+1} \sum_{k=1}^{k_{1}} \log _{2}\left(1+h_{k} p_{k}\right)-\sum_{k=k_{n}+1}^{k_{n+1}} \log _{2}\left(1+h_{k} p_{k}\right)=0, \text { for } n=1,2,3, \ldots, N-1 . \tag{10.43}
\end{equation*}
$$

Theorem 10.6.1. The power levels that maximize the system EE can be obtained through a water-filling-like solution where each user has a common water level on all of its carriers. All user's water levels can be written in terms of a single User's water level, and this particular User's water level can be obtained by solving a single non-linear equation:

$$
\begin{equation*}
E E\left(w_{1}, \beta_{2} w_{1}^{\gamma_{2}}, \beta_{3} w_{1}^{\gamma_{3}}, \ldots, \beta_{N} w_{1}^{\left.\gamma_{N}\right)}=\frac{\sum_{n=1}^{N} \alpha_{n}}{(\ln 2) \sum_{n=1}^{N} \alpha_{n} \beta_{n} w_{1}^{\gamma_{n}}},\right. \tag{10.44}
\end{equation*}
$$

where $\beta_{n}$ and $\gamma_{n}$ are defined in the proof below. The maximum $E E$ is inversely proportional to the weighted average water level. Anytime a subchannel is replaced with another with a greater channel gain, the maximum EE increases.

Proof: Just like in the last section we apply the CCT transformation (Chapter 5):

$$
\begin{equation*}
t=\frac{1}{p_{c}+\psi \sum_{k=1}^{K} p_{k}} \quad \text { and } \quad y_{k}=t p_{k} \tag{10.45}
\end{equation*}
$$

The new optimization problem:

$$
\max _{\mathbf{y}, t} \quad f(\mathbf{y}, t)=t \sum_{k} \log _{2}\left(1+h_{k} y_{k} / t\right)
$$

subject to
$\mathrm{C} 1: \alpha_{n+1} \sum_{k=1}^{k_{1}} \log _{2}\left(1+h_{k} y_{k} / t\right)-\sum_{k=k_{n}+1}^{k_{n+1}} \log _{2}\left(1+h_{k} y_{k} / t\right)=0$, for $n=1,2,3, \ldots, N-1$.
$\mathrm{C} 2: \psi \sum_{k} y_{k}+p_{c} t-1=0$

The Lagrangian:

$$
\begin{align*}
L(\mathbf{y}, t, \lambda, \mu)=f+\lambda \psi & {\left[\sum_{k} y_{k}+p_{c} t-1\right] } \\
& +\sum_{n=1}^{N-1} \mu_{n}\left[\alpha_{n+1} \sum_{k=1}^{k_{1}} \log _{2}\left(1+h_{k} y_{k} / t\right)-\sum_{k=k_{n}+1}^{k_{n+1}} \log _{2}\left(1+h_{k} y_{k} / t\right)\right] \tag{10.47}
\end{align*}
$$

The KKT conditions:
Constraint C1:

$$
\begin{equation*}
\alpha_{n+1} \sum_{k=1}^{k_{1}} \log _{2}\left(1+h_{k} p_{k}\right)=\sum_{k=k_{n}+1}^{k_{n+1}} \log _{2}\left(1+h_{k} p_{k}\right), \text { for } n=1,2,3, \ldots, N-1 . \tag{10.48}
\end{equation*}
$$

Constraint C2:

$$
\begin{equation*}
\psi \sum_{k} y_{k}+p_{c} t-1=0 \tag{10.49}
\end{equation*}
$$

$t$ derivative:

$$
\begin{align*}
\sum_{k=1}^{K} \ln \left(1+h_{k} p_{k}\right) & -\frac{h_{k} p_{k}}{1+h_{k} p_{k}}-\frac{1}{t} \sum_{n=1}^{N-1} \mu_{n} \alpha_{n+1} \sum_{k=1}^{k_{1}} \frac{h_{k} p_{k}}{1+h_{k} p_{k}}+\frac{\mu_{1}}{t} \sum_{k=k_{1}+1}^{k_{2}} \frac{h_{k} p_{k}}{1+h_{k} p_{k}}  \tag{10.50}\\
& +\frac{\mu_{2}}{t} \sum_{k=k_{2}+1}^{k_{3}} \frac{h_{k} p_{k}}{1+h_{k} p_{k}}+\ldots+\frac{\mu_{N-1}}{t} \sum_{k=k_{N-1}+1}^{K} \frac{h_{k} p_{k}}{1+h_{k} p_{k}}+\lambda \psi l p_{c}=0
\end{align*}
$$

$y_{k}$ derivatives:

$$
\begin{gather*}
\left(1+\frac{1}{t} \sum_{n=1}^{N-1} \mu_{n} \alpha_{n+1}\right) \frac{h_{k}}{1+h_{k} p_{k}}+\lambda \psi l=0, \text { for } k=1,2, \ldots, k_{1}  \tag{10.51a}\\
\left(1-\frac{\mu_{1}}{t}\right) \frac{h_{k}}{1+h_{k} p_{k}}+\lambda \psi l=0, \text { for } k=k_{1}+1, k_{1}+2, \ldots, k_{2}  \tag{10.51b}\\
\left(1-\frac{\mu_{2}}{t}\right) \frac{h_{k}}{1+h_{k} p_{k}}+\lambda \psi l=0, \text { for } k=k_{2}+1, k_{2}+2, \ldots, k_{3}  \tag{10.51c}\\
\vdots  \tag{10.51d}\\
\left(1-\frac{\mu_{N-1}}{t}\right) \frac{h_{k}}{1+h_{k} p_{k}}+\lambda \psi l=0, \text { for } k=k_{N-1}+1, k_{N-1}+2, \ldots, N
\end{gather*}
$$

Eliminating $\lambda$ from (10.51a) shows that User-1 has a common water level on all his subcarriers. Similarly, (10.51b)-(10.51e) show each user has his own water level. If we denote User- $n$ 's water level by $w_{n}$, we have

$$
\begin{array}{cl}
p_{k}+1 / h_{k}=w_{1}, & \text { for } k=1,2, \ldots, k_{1} \\
p_{k}+1 / h_{k}=w_{2}, & \text { for } k=k_{1}+1, k_{1}+2, \ldots, k_{2} \\
p_{k}+1 / h_{k}=w_{3}, & \text { for } k=k_{2}+1, k_{2}+2, \ldots, k_{3} \\
\vdots & \\
p_{k}+1 / h_{k}=w_{N}, & \text { for } k=k_{N-1}+1, k_{N-1}+2, \ldots, N \tag{10.52e}
\end{array}
$$

Using (10.52) on (10.51),

$$
\begin{align*}
& 1+\frac{1}{t} \sum_{n=1}^{N-1} \mu_{n} \alpha_{n+1}=-\lambda \psi l w_{1}  \tag{10.53a}\\
& 1-\frac{\mu_{1}}{t}=-\lambda \psi l w_{2}  \tag{10.53b}\\
& 1-\frac{\mu_{2}}{t}=-\lambda \psi l w_{3}  \tag{10.53c}\\
& \vdots  \tag{10.53e}\\
& 1-\frac{\left.\mu_{( } N-1\right)}{t}=-\lambda \psi l w_{N} \\
&(10.53 \mathrm{a})+\alpha_{2}(10.53 \mathrm{~b})+\alpha_{3}(10.53 \mathrm{c})+\ldots+\alpha_{N}(10.53 \mathrm{e}) \text { gives } \\
& \sum_{n=1}^{N} \alpha_{n}=-\lambda \psi l \sum_{n=1}^{N} \alpha_{n} w_{n}
\end{align*}
$$

Multiplying (10.51a) by $p_{k}$ and summing over $k=1$ to $k=k_{1}$,

$$
\begin{equation*}
\sum_{k=1}^{k_{1}}\left(1+\frac{1}{t} \sum_{n=1}^{N-1} \mu_{n} \alpha_{n+1}\right) \frac{h_{k} p_{k}}{1+h_{k} p_{k}}+\lambda \psi l \sum_{k=1}^{k_{1}} p_{k}=0 \tag{10.55}
\end{equation*}
$$

We can obtain similar equations from the other equations in (10.51), by multiplying by $p_{k}$ and summing over appropriate $k$ values. Adding all these equations to (10.50) and rearranging give us,

$$
\begin{equation*}
f=-\lambda, \tag{10.56}
\end{equation*}
$$

a result that is exactly the same as the one in (10.40) from the last section.
Combining this with (10.54),

$$
\begin{equation*}
f=\frac{\sum_{n=1}^{N} \alpha_{n}}{\psi(\ln 2) \sum_{n=1}^{N} \alpha_{n} w_{n}} \tag{10.57}
\end{equation*}
$$

This equation is very similar to (10.13) and (10.41) from the last two sections except the water level in the last two sections is now replaced with the average water level. In other words, the maximum EE in this section is inversely proportional to the weighted average of the water levels.

Now compare (10.57) to (10.41) from last section. The position or role of any $w_{k}$ in (10.57) is similar to the role of $w$ in (10.41), as far as the increasing/decreasing behavior of the functions they are part of. In the last section we concluded that $\partial w / \partial h_{k}<0$. Based on the comparisons of (10.57) and (10.41), we deduce that $\partial w / \partial h_{k}<0$ here too. That is, the optimum water level decreases with the channel gains. Since the maximum EE decreases with the water level it produces it, we conclude that anytime a subchannel is replaced with another with a greater channel gain, the maximum EE increases.

Using (10.52) on (10.48),

$$
\begin{gather*}
\alpha_{n+1} \sum_{k=1}^{k_{1}} \log _{2}\left(h_{k} w_{1}\right)=\sum_{k=k_{n}+1}^{k_{n+1}} \log _{2}\left(h_{k} w_{n}\right), \text { for } n=1,2,3, \ldots, N-1 .  \tag{10.58}\\
{\left[\prod_{k=1}^{k_{1}}\left(h_{k} w_{1}\right)\right]^{\alpha_{n+1}}=\prod_{k=k_{n}+1}^{k_{n+1}}\left(h_{k} w_{n}\right), \text { for } n=1,2,3, \ldots, N-1 .} \tag{10.59}
\end{gather*}
$$

Letting

$$
\begin{equation*}
\beta_{n+1}=\frac{\left[\prod_{k=1}^{k_{1}}\left(h_{k}\right)\right]^{\frac{\alpha_{n+1}}{K_{n}}}}{\left[\prod_{k=k_{n}+1}^{k_{n+1}} h_{k}\right]^{\frac{1}{K_{n}}}} \text { for } n=1,2,3, \ldots, N-1 \tag{10.60}
\end{equation*}
$$

and

$$
\begin{equation*}
\gamma_{n+1}=\frac{K_{1} \alpha_{n+1}}{K_{n}} \text { for } n=1,2,3, \ldots, N-1 \tag{10.61}
\end{equation*}
$$

give us

$$
\begin{equation*}
w_{n}=\beta_{n} w_{1}^{\gamma_{n}} \text { for } n=2,3,4, \ldots, N . \tag{10.62}
\end{equation*}
$$

Note that this equation allows us to write the water levels of all users in terms of the water
level $w_{1}$ of User-1. We can now write both sides of (10.57) in terms of $w_{1}$.

$$
\begin{equation*}
f\left(w_{1}, w_{2}, w_{3}, \ldots, w_{N}\right)=f\left(w_{1}, \beta_{2} w_{1}^{\gamma_{2}}, \beta_{3} w_{1}^{\gamma_{3}}, \ldots, \beta_{N} w_{1}^{\left.\gamma_{N}\right)}=\frac{\sum_{n=1}^{N} \alpha_{n}}{(\ln 2) \sum_{n=1}^{N} \alpha_{n} \beta_{n} w_{1}^{\gamma_{n}}}\right. \tag{10.63}
\end{equation*}
$$

where we assume that $\beta_{1}=1$. This single non-linear equation will give us $w_{1}$. Once we know $w_{1}$, all the other water levels can be calculated. From the water levels, we can obtain the power levels that maximize the total EE.

### 10.7 Multi user multi carrier case (OFDMA) with minimum rate requirements

In this section we include user's minimum rate constraints. Consider, for a moment, the solution to the unconstrained EE maximization problem with multi users and multi carriers. This solution will be identical to the solution from Section 10.5 to the EE maximization problem with a single user and multi carriers - the solution based on a single water level $w$. It may happen that the power levels corresponding to this solution do not satisfy some (or all) user's (say in set $S_{\mathrm{uh}}$ minimum rate requirements. Note that $w^{*}$ will satisfy the rate demand of User- $n$ only if

$$
\begin{equation*}
\sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(w^{*} h_{k}\right) \geq R_{n} \tag{10.64}
\end{equation*}
$$

This can be re-arranged to:

$$
\begin{equation*}
w^{*} \geq\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}} \tag{10.65}
\end{equation*}
$$

We must identify all the users for whom the above condition is not satisfy and put them in set $S_{\mathrm{uh}}$. Then we must allocate enough power to users in set $S_{\mathrm{uh}}$ so as to satisfy their rate demands, remove these users from the mathematics and then re-visit the original optimization problem. Before we re-visit the original problem minus these users, we have to think
about how to allocate the power to the subchannels belonging to these users. The power distribution to subchannels assigned to users in set $S_{\mathrm{uh}}$ must maximize the sum EE of these users, while meeting their rate demands exactly. That is, we must solve the optimization problem:

$$
\begin{equation*}
\max _{\mathbf{p}} \frac{\sum_{k \in S_{\mathrm{uh}}} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k \in S_{\mathrm{uh}}} p_{k}} \tag{10.66}
\end{equation*}
$$

subject to
where $k \in S_{\text {uh }}$ means subchannels assigned to users in $S_{\mathrm{uh}}$ and $R_{n}$ is the minimum rate requirement of User- $n$.

Theorem 10.7.1. The EE maximizing power allocation for users in $S_{u h}$ is given by the water levels:

$$
\begin{equation*}
w_{n}=\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}} \text { for all } n \in S_{u h} \tag{10.67}
\end{equation*}
$$

Proof: At first it appears that we have to repeat the original optimization problem for the users in $S_{\mathrm{uh}}$. However, there is a way in which we can obtain a closed form solution. Notice that the above optimization problem is the same as

$$
\max _{\mathbf{p}} \frac{\sum_{n \in S_{\mathrm{uh}}} R_{n}}{p_{c}+\psi \sum_{k \in S_{\mathrm{uh}}} p_{k}}
$$

subject to

$$
\sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right)=R_{n} \text { for } n \in S_{\mathrm{uh}} .
$$

Furthermore, the numerator of the objective function as well as $p_{c}$ and $\psi$ are constants.

Therefore, this optimization problem reduces to

$$
\min _{\mathbf{p}} \sum_{k \in S_{\mathrm{uh}}} p_{k}
$$

subject to

$$
\begin{equation*}
\mathrm{C}: \sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right)-R_{n}=0 \text { for } n \in S_{\mathrm{uh}} . \tag{10.69}
\end{equation*}
$$

Notice that this is the opposite of the OFDMA power allocation problem that maximizes the throughput subject to a total power constraint. Here too, the solution is obtained using the Lagrange multipliers. The Lagrangian equations are:

$$
\begin{align*}
& 1=\alpha_{n} \frac{h_{k}}{1+h_{k} p_{k}} \text { for } k=k_{n-1}+1, k_{n-1}+2, \ldots k_{n}  \tag{10.70}\\
& \quad \text { and for all } n \in S_{\mathrm{uh}},
\end{align*}
$$

where $\alpha_{n}$ 's are the dual variables. Writing $p_{k}$ 's in terms of $\alpha_{n}$ 's,

$$
\begin{align*}
p_{k}= & \alpha_{n}-\frac{1}{h_{k}} \text { for } k=k_{n-1}+1, k_{n-1}+2, \ldots k_{n}  \tag{10.71}\\
& \text { and for all } n \in S_{\mathrm{uh}} .
\end{align*}
$$

This shows that $\alpha_{n}$ is in fact the water level for User- $n$. Using this to substitute for $p_{k}$ in constraint C 1 and re-arranging,

$$
\begin{equation*}
w_{n}=\alpha_{n}=\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}} \text { for all } n \in S_{\mathrm{uh}} \tag{10.72}
\end{equation*}
$$

Substituting in (10.71) for $\alpha_{n}$,

$$
\begin{align*}
p_{k}=p_{k}^{I}= & {\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}}-\frac{1}{h_{k}} }  \tag{10.73}\\
& \text { for } k=k_{n-1}+1, k_{n-1}+2, \ldots k_{n}
\end{align*}
$$

### 10.8 Multi user multi carrier case (OFDMA) with total power constraint

Often in OFDMA, there is a total power constraint, $\sum_{k} p_{k} \leq P_{T}$, on the transmission powers. When the solution to the unconstrained EE maximization occurs outside the hyper plane given by $\sum_{k} p_{k}=P_{T}$, the solution to the constrained EE maximization problem will occur on the hyper plane $\sum_{k} p_{k}=P_{T}$. It is easily seen that this happens when

$$
\begin{equation*}
w^{*}>P+\sum_{k} \frac{1}{h_{k}} \tag{10.74}
\end{equation*}
$$

In this case, we must solve the following optimization problem:

$$
\begin{equation*}
\max _{\mathbf{p}} \frac{\sum_{k} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k} p_{k}} \tag{10.75}
\end{equation*}
$$

subject to

$$
\sum_{k} p_{k}=P_{T}
$$

Once again we can obtain the solution easily, this time without doing any work at all. Note that we can replace the $\sum_{k} p_{k}$ in the denominator of the objective function with $P_{T}$
from the constraint.

$$
\begin{equation*}
\max _{\mathbf{p}} \frac{\sum_{k} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi P_{T}} \tag{10.76}
\end{equation*}
$$

subject to

Since the denominator of the objective function is a constant, the above problem reduces to:

$$
\begin{equation*}
\max _{\mathbf{p}} \sum_{k} \log _{2}\left(1+h_{k} p_{k}\right) \tag{10.77}
\end{equation*}
$$

subject to

$$
\sum_{k} p_{k}=P_{T}
$$

This is the familiar OFDMA throughput maximization (under a total power constraint) problem, for which there is the well known water filling solution [140]:

$$
\begin{equation*}
p_{k}=p_{k}^{* *}=\left[w^{* *}-\frac{1}{h_{k}}\right]^{+} \tag{10.78}
\end{equation*}
$$

where

$$
\begin{equation*}
w^{* *}=\frac{P_{T}}{K}+\frac{1}{K} \sum_{k} \frac{1}{h_{k}} \tag{10.79}
\end{equation*}
$$

### 10.9 EE maximizing power allocation for OFDMA with minimum rate and total power Constraints

This section combines the theoretical results obtained in this chapter to form a power allocation procedure. This procedure assumes that the subchannels have been assigned to the users already, and provides the optimal power allocation that maximizes the system EE with
user's minimum rate constraints and a system wide total transmission power constraint. The material developed in this chapter is also sufficient to form a power allocation procedure with proportional rate constraints. The reader to referred to our work in [139] for such a power allocation procedure.

For the convenience of the reader, the system model and the problem statement with minimum rate and total power constrains are repeated before the power allocation procedure is presented. Consider the downlink of a single cell with $N$ users and $K$ orthogonal subchannels. Each subchannel is assigned exclusively to one user. If the channel gain on the $k$ th subchannel is $a_{k}$, the transmission power $p_{k}$, the noise spectral density $\sigma^{2}$, and $h_{k}=a_{k} / \sigma^{2}$, then the system EE of the transmissions over all $K$ subchannels in bits/Joule/Hz can be written as:

$$
\begin{equation*}
E E=\frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}} \tag{10.80}
\end{equation*}
$$

where $\psi$ is the reciprocal of the efficiency and $p_{c}$ is the circuit power [25] of the downlink transmitter.

Assume a yet-to-be determined subchannel assignment protocol is used to distribute the subchannels among the users. Suppose a total of $K_{1}$ subchannels - subchannel 1 through subchannel $k_{1}$ - are assigned to User-1. A total of $K_{2}$ subchannels - subchannel $k_{1}+1$ through subchannel $k_{2}$ - are assigned to User-2 and so on. A total of $K_{n}$ subchannels - subchannel $k_{n-1}+1$ through subchannel $k_{n}$ - are assigned to User- $n$. Then the rate $r_{n}$ of User- $n$ can be written as:

$$
\begin{equation*}
r_{n}=\sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right) \tag{10.81}
\end{equation*}
$$

where $k_{0}=0$. If the minimum rate acceptable to User- $n$ is $R_{n}$ then we must have

$$
\begin{equation*}
r_{n}-R_{n} \geq 0 \text { for } n=1,2, \ldots, N \tag{10.82}
\end{equation*}
$$

The goal is to maximize the EE in (10.80) subject to (10.82) and a total power constraint
$P$ using a two step optimization procedure - first using the GA for subchannel assignment and then solving the following power allocation problem:

$$
\begin{align*}
& \max _{p_{1}, p_{2}, \ldots p_{K}} E E \\
& \text { subject to } \\
& \mathrm{C} 1: r_{n}-R_{n} \geq 0 \text { for } n=1,2, \ldots, N .  \tag{10.83}\\
& \mathrm{C} 2: P-\sum_{k=1}^{K} p_{k} \geq 0 .
\end{align*}
$$

## Feasibility

For a given subchannel allocation, it may happen that the constraints C 1 and C 2 cannot be satisfied simultaneously. However, this will come to light only after the channel allocation has been performed and power allocation is being attempted. At this point there are two options. One is to remove the users with the largest $R_{n} / \prod_{k} h_{k}$ 's one by one until the constraints are satisfied. The second is to redo the channel allocation so as to give better channels to the users whose rate demands are not met. The second option will increase the complexity of the resource allocation considerably. For simplicity, in this dissertation we will assume no feasibly issue arises.

## The Optimal Power Allocation Procedure

1. Solve (10.41) and obtain the common water level $w^{*}$.
2. If $w^{*} \geq\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{k}}}\right]^{1 / K_{n}}$ for all $n \in S$ and $w^{*} \leq P+\sum_{k \in S} \frac{1}{h_{k}}$ then $p_{k}:=w^{*}-\frac{1}{h_{k}}$ for all $k \in S$.
3. If $w^{*}>P+\sum_{k \in S} \frac{1}{h_{k}}$ then $p_{k}=p_{k}^{* *}$ from (10.78)
4. If $w^{*} \leq P+\sum_{k \in S} \frac{1}{h_{k}}$ and $w^{*} \geq\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}}$ for all $n \in S_{\mathrm{h}}$ and $w^{*}<\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n} h_{k}}}\right]^{1 / K_{n}}$ for all $n \in S_{\mathrm{uh}}$ then $p_{k}=\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{k}}}\right]^{1 / K_{n}}-\frac{1}{h_{k}}$ for $k=k_{n-1}+1, k_{n-1}+2, \ldots k_{n}$, and for all $n \in S_{\mathrm{uh}}$, and go to step 1 with $S:=S_{\mathrm{h}}$.
$S$ represents the set of all users in the system. Step 1 computes the solution to the unconstrained optimization problem. Step 2 checks if all the user's rate demands are met and if the solution lies inside power constraint plane. If so then the solution to the unconstrained optimization problem $w^{*}$ is the solution to the constrained optimization problem.

Step 3 checks if $w^{*}$ falls outside the power constraint plane and if so then the solution to the constrained optimization problem lies on the power constraint plane.

Step 4 identifies those users whose rate demands are not met (set $S_{\mathrm{uh}}$ ), allocates appropriate power to them and go back to step 1 with the reduced set of users.

### 10.10 Complexity

Note that step-1 of our power allocation procedure solves a single non-linear equation regardless of the number of users $N$ in the system. Since the equation has only one root and we have an interval in which we expect $w^{*}$ to fall, $w^{*}$ can be obtained in few iterations. Therefore, the time it takes for step-1 is independent of $N$ and is negligible. In the worst case scenario, step-4 may involve going through all $N$ users. Hence, our algorithm has the worst case complexity of $O(N)$.

### 10.11 Conclusion

This chapter showed that in OFDM, whether we have single user and single carrier or multi users with multi carriers, the power allocation that maximizes the EE for a given carrier assignment can be obtained by solving a single non-linear equation. The cases with a total power, proportional rate, and minimum rate constraints are also considered. The chapter concluded with a power allocation algorithm that maximizes the EE for a fixed channel assignment under a total power and minimum rate constraint. Note that this chapter also has sufficient material to formulate a power allocation procedure with proportional rate constraints instated of minimum rate constraints. The interested reader is referred to our work in [139].

## Chapter 11

## Energy efficient frequency and power allocation for OFDMA

### 11.1 Introduction

Another gap in the literature regarding the energy efficiency (EE) maximization, the application of genetic algorithms (GA) for channel assignment, is addressed in this chapter. GA is ideally suited to a combinatorial problem like channel assignment and is widely used for channel assignment in many other optimization problems in wireless communications. Yet, GA is conspicuously absence in problems regarding EE maximization. As explained in detail later, this is probably due to the complexity involved in calculating the maximum possible EE for each individual in each generation of the GA. In this chapter, we overcome this difficulty by using the analytical insights obtained in the last chapter regarding the nature of the EE maximizing power allocation.

A two-step near optimal solution to the problem of finding the subchannel and power allocation that maximizes the energy efficiency of the orthogonal frequency division multiple access (OFDMA) based transmissions, under minimum rate and total power constraints, is presented. Genetic algorithm is used for subchannel assignment, and is followed by optimal
power allocation derived in the last chapter via analytical methods.

### 11.2 Background

As mentioned in the last chapter, the power allocation that maximizes the bits/Joule/Hz energy efficiency (EE) of OFDMA transmissions has been well studied [25,67-72,138]. Channel assignment that maximizes the EE in OFDMA, however, has not received as much attention. Only simple, low complexity, ad-hoc protocols based on intuition, have been proposed for EE maximizing channel allocation $[25,72,138]$. A related optimization problem is the throughput maximizing channel assignment for OFDMA with QoS constraints. This is a well known, computationally challenging problem that remains unsolved. As expected, only simple, suboptimal, heuristic solutions have been proposed [95] for this problem. As demonstrated in Chapter 6, genetic algorithm (GA) is perfectly suited to this later optimization problem. In one of our earlier work [96], we used GA for finding the throughput maximizing channel assignment for OFDMA. Recently, GA has also been used for power allocation in EE maximization [97].

Despite the similarities between the throughput maximizing channel assignment and EE maximizing channel allocation, there is a notable absence of the application of GA for channel assignment in EE maximization. One of the reason for this could be the following. For each generation, and for each individual, the GA needs to evaluate the fitness function. In our present application, the fitness function would be the maximum EE possible from a given channel assignment. However, the maximum possible EE from a given channel assignment can only be evaluated by going through the optimization process that finds the optimal power allocation for a given channel assignment. As mentioned in the last chapter, the method available for determining the maximum EE for a given channel assignment in the literature is via an iterative algorithm. In the middle of GA, waiting for the convergence of this algorithm for each of the 100's of individuals in each of the 100's of generations, means
this iterative algorithm has to be executed at least a total of 10,000 times. This can only be described as "computational suicide."

In the last chapter, we made considerable progress away from the EE maximizing power allocation techniques in the literature, by showing that the maximum EE can be calculated by solving a single non-linear equation. While solving a single non-linear equation takes much shorter time compared to waiting for the convergence of the iterative algorithms from the literature, this approach too wastes precious time in the middle of GA.

This difficulty with fitness functions that are too complex to evaluate in the middle of the GA is not uncommon in the broader area of GA applications, particularly, outside of wireless communications. A common solution is to use an approximation to the fitness function. It is this approach that we take in this chapter; we use two different approximations to the maximum EE possible from a given channel assignment. These approximate fitness functions are obtained through our work in the last chapter dealing with the EE maximizing power allocation with proportional rate . After obtaining the channel assignment using GA, we use the optimal power allocation from the last chapter. Numerical results obtained show that the channel assignment using GA produces better EE than the best protocol in the literature.

Section 11.3 formulates the two-step optimization problem. Channel assignment using GA is discussed in Section 11.4, and power allocation in Section 11.5. Numerical results are presented in Section 11.6, and the chapter concludes in Section 11.7.

### 11.3 System model and the optimization problem

Consider the downlink of a single cell with $N$ users and $K$ orthogonal subchannels. Each subchannel is assigned exclusively to one user. If the channel gain on the $k$ th subchannel is $a_{k}$, the transmission power $p_{k}$, the noise spectral density $\sigma^{2}$, and $h_{k}=a_{k} / \sigma^{2}$, then the
system EE of the transmissions over all $K$ subchannels in bits/Joule/Hz can be written as:

$$
\begin{equation*}
E E=\frac{\sum_{k=1}^{K} \log _{2}\left(1+h_{k} p_{k}\right)}{p_{c}+\psi \sum_{k=1}^{K} p_{k}} \tag{11.1}
\end{equation*}
$$

where $\psi$ is the reciprocal of the efficiency and $p_{c}$ is the circuit power [25] of the downlink transmitter. A summary of the notations is given in Table 11.1.

| $N$ | number of users in the system |
| :---: | :---: |
| $K$ | number of subchannels in the system |
| $a_{k}$ | channel gain on subchannel $k$ experienced by the user to whom it is assigned |
| $\sigma_{k}^{2}$ | additive white Gaussian noise density |
| $h_{k}$ | $a_{k} / \sigma_{k}^{2}$ |
| $p_{k}$ | power used on subchannel $k$ |
| $p_{c}$ | circuit power |
| $\psi$ | reciprocal of the power amplifier efficiency |
| $r_{n}$ | transmission rate of User- $n$ |
| $R_{n}$ | minimum transmission rate acceptable to User- $n$ |
| $P$ | total transmission power constraint |

Table 11.1: Notations

Assume a yet-to-be determined subchannel assignment protocol is used to distribute the subchannels among the users. Suppose a total of $K_{1}$ subchannels - subchannel 1 through subchannel $k_{1}$ - are assigned to User-1. A total of $K_{2}$ subchannels - subchannel $k_{1}+1$ through subchannel $k_{2}$ - are assigned to User-2 and so on. A total of $K_{n}$ subchannels - subchannel $k_{n-1}+1$ through subchannel $k_{n}$ - are assigned to User- $n$. Then the rate $r_{n}$ of User- $n$ can be written as:

$$
\begin{equation*}
r_{n}=\sum_{k=k_{n-1}+1}^{k_{n}} \log _{2}\left(1+h_{k} p_{k}\right), \tag{11.2}
\end{equation*}
$$

where $k_{0}=0$. If the minimum rate acceptable to User- $n$ is $R_{n}$ then we must have

$$
\begin{equation*}
r_{n}-R_{n} \geq 0 \text { for } n=1,2, \ldots, N . \tag{11.3}
\end{equation*}
$$

The goal is to maximize the EE in (11.1) subject to (11.3) and a total power constraint $P$ using a two step optimization procedure - first using the GA for subchannel assignment and then solving the following power allocation problem:

$$
\begin{align*}
& \max _{p_{1}, p_{2}, \ldots p_{K}} E E \\
& \text { subject to } \\
& \mathrm{C} 1: r_{n}-R_{n} \geq 0 \text { for } n=1,2, \ldots, N .  \tag{11.4}\\
& \mathrm{C} 2: P-\sum_{k=1}^{K} p_{k} \geq 0 .
\end{align*}
$$

## Feasibility

For a given subchannel allocation, it may happen that the constraints C 1 and C 2 cannot be satisfied simultaneously. However, this will come to light only after the channel allocation has been performed and power allocation is being attempted. At this point there are two options. One is to remove the users with the largest $R_{n} / \prod_{k} h_{k}$ 's one by one until the constraints are satisfied. The second is to redo the channel allocation so as to give better channels to the users whose rate demands are not met. The second option will increase the complexity of the resource allocation considerably. For simplicity, in this chapter we will assume no feasibility issue arises.

### 11.4 Channel assignment using genetic algorithm

A detailed description of the steps involved in the genetic algorithm (GA) was given in Chapter 6. In the current application of GA, the objective function, the EE, has a single discrete variable - an array representing the subchannel assignment. As mentioned in Chapter 6 , this array is referred to as a chromosome or an individual. GA creates a number of these individuals randomly to form the initial population. The population is ranked using the
fitness function and the top two third is used to create two thirds of the individuals in the next generation.

### 11.4.1 Choice of a fitness function

Ideally, in order to decide the effectiveness of a particular channel assignment or chromosome, we must calculate the best possible EE resulting from that channel assignment. However, this would involve finding the power allocation that maximizes EE for that particular channel assignment. In other words, we must solve the optimization problem in (11.4). As mentioned earlier, the fastest method of accomplishing this is to solve a non-linear equation. This means if we were to take this path, for each generation of the GA, and for each individual, we have to solve a non-linear equation in order to evaluate its fitness function. Obviously, this is far too time consuming.

In what follows, we will consider two approximations that can be used as the estimates of the maximum EE resulting from a particular channel assignment. In Chapter 10, we mentioned the equation that relates the maximum EE, for a given channel assignment, to the water level that produces it.

$$
\begin{equation*}
E E(w)=\frac{1}{(\ln 2) \psi w} \tag{11.5}
\end{equation*}
$$

This shows the maximum EE is inversely related to the water level that produces it. In other words, the lower the water level, the higher the EE. This suggests that we could use the reciprocal of the common water level as the fitness function. However, the water level that maximizes the EE cannot be obtained without going through the optimization process.

In Chapter 10, we also proved that the maximum EE under proportional rate constraints is achieved when each user has its own separate water level. Furthermore, the maximum EE
in this case is related to the weighted average water level in the following way.

$$
\begin{equation*}
E E(w)=\frac{\sum w_{n}}{(\ln 2) \psi \sum \alpha_{n} w_{n}}, \tag{11.6}
\end{equation*}
$$

where $\alpha_{n}$ 's are the proportional rates of the users with respect to one of the users.
Suppose we make the simplifying assumption that the optimum EE under minimum rate requirements actually occurs at the proportional rates given by the minimum rates requested. Then the proportional rates can be substituted by minimum rates. Furthermore, each user's optimum water level can be easily calculated using (10.72) of Chapter 10.

$$
\begin{equation*}
w_{n}=\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}} \tag{11.7}
\end{equation*}
$$

We can now use the reciprocal of the weighted average water level as the fitness function.

$$
\begin{equation*}
F F=\frac{\sum_{n} R_{n}}{\sum_{n} R_{n} w_{n}} \tag{11.8}
\end{equation*}
$$

However, taking the reciprocal of a real number is computationally expensive. But maximizing the reciprocal of a function is equivalent to maximizing the negative of the function. Since $\sum_{n} R_{n}$ is a constant, the first fitness function we propose is:

$$
\begin{equation*}
F F_{1}=-\sum_{n} R_{n} w_{n} \tag{11.9}
\end{equation*}
$$

The fastest computational approach, in fact, is to use $\sum_{n} R_{n} w_{n}$ and to choose the individual with the least "fitness function." However, a graph that converges to a maximum is clearer and more demonstrative than a graph that converges to a minimum. Therefore, we will stick with maximizing $-\sum_{n} R_{n} w_{n}$. Another option is to minimize the maximum water level. This is the same as maximizing the negative of maximum water level, which brings us to the
second fitness function:

$$
\begin{equation*}
F F_{2}=-\max _{n} w_{n} \tag{11.10}
\end{equation*}
$$

### 11.4.2 The channel assignment solution

The top one-third of the original population and the children born to the top two thirds form the next generation of individuals. The process is repeated until a predetermined number of generations. The chromosome of the individual with the highest fitness function value in the last generation represents the near optimal subchannel assignment.

### 11.5 Power allocation

The optimum power allocation procedure of Chapter 10 is applied after the subchannels have been assigned using GA. For the convenience of the reader we repeat that power allocation procedure here.

## The Optimal Power Allocation Procedure

1. Solve (10.41) and obtain the common water level $w^{*}$.
2. If $w^{*} \geq\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{k}}}\right]^{1 / K_{n}}$ for all $n \in S$ and $w^{*} \leq P+\sum_{k \in S} \frac{1}{h_{k}}$ then $p_{k}:=w^{*}-\frac{1}{h_{k}}$ for all $k \in S$.
3. If $w^{*}>P+\sum_{k \in S} \frac{1}{h_{k}}$ then $p_{k}=p_{k}^{* *}$ from (10.78)
4. If $w^{*} \leq P+\sum_{k \in S} \frac{1}{h_{k}}$ and $w^{*} \geq\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}}$ for all $n \in S_{\mathrm{h}}$ and $w^{*}<\left[\frac{2^{R_{n}}}{\prod_{k=k_{n-1}+1}^{k_{n}} h_{k}}\right]^{1 / K_{n}}$ for all $n \in S_{\mathrm{uh}}$
then $p_{k}=\left[\frac{2^{R_{n}}}{\Pi_{k=k_{n-1}}^{k_{n}} h_{k} h_{k}}\right]^{1 / K_{n}}-\frac{1}{h_{k}}$ for $k=k_{n-1}+1, k_{n-1}+2, \ldots k_{n}$, and for all $n \in S_{\mathrm{uh}}$, and go to step 1 with $S:=S_{\mathrm{h}}$.

### 11.6 Numerical results

We illustrate the convergence behavior of the GA first and then compare the EE results obtained using GA to that from the literature. The details of the parameters used is the GA are given in Fig. 11.1. The channel gain simulation details are given in Fig. 11.2.

```
Number of individuals in the population = 3P,}\mathrm{ where P=100
No mutation was used
Stopping criteria: Number of generations Ngen = 100
```

Figure 11.1: Genetic algorithm parameters.

The behaviors of the GA for Scenario-1 from Fig. 11.2 with the two proposed fitness functions are shown in Fig. 11.3 and Fig. 11.4. Fig. 11.5 compares the EE results obtained using GA for channel allocation with that obtained using the best intuitive channel allocation protocol from the literature [25]. In all three cases, the optimal power allocation from last chapter was used. GA using both fitness functions perform better than the intuitive channel allocation protocol in [25]. Not surprisingly the GA that uses the weighted water level performs better than the one that uses the maximum water level.

| Users were placed randomly within a cell radius $\mathbf{R}$. <br> Rayleigh fading and log-normal shadowing were superimposed on the channel gains obtained from the distance to the transmitter using a path loss coefficient of 4 . |  |
| :---: | :---: |
| SCENARIO - <br> Minim | Users $K=128$ subchannels $\quad R=500$ m <br> rate requirements over the 128 channels (bits/s/Hz): ${ }_{1}=75 \quad R_{2}=70 \quad R_{3}=65 \quad R_{4}=60 \quad R_{5}=55 \quad R_{6}=45$ |
| SCENARIO-2: $\text { N = } 6 \text { Users }$ | $K=128$ subchannels $\quad R$ varied from 500 m to 1000 m <br> Minimum rate requirements were scaled from SCENARIO-1 according to cell radius. |

Figure 11.2: Simulation details.


Figure 11.3: Scenario-1: Convergence of GA with $F F_{1}=-\sum_{n} R_{n} w_{n}$.


Figure 11.4: Scenario-1: Convergence of GA with $F F_{2}=-\max _{n} w_{n}$.


Figure 11.5: Scenario-2: Comparing the EE values.

### 11.7 Discussion and Conclusion

A two-step near optimal solution to the problem of finding the subchannel and power allocation that maximizes the energy efficiency of the orthogonal frequency division multiple access (OFDMA) based transmissions, under minimum rate and total power constraints, was presented in this chapter. Genetic algorithm (GA)was used for subchannel assignment, and was followed by optimal power allocation derived in the last chapter via analytical methods.

A common criticism of GA is the high computational complexity involved. In spite of this, the material in this chapter is useful in two different ways. First, the chapter shows that the existing heuristic protocols for channel assignments have room to improve. Perhaps an avenue to explore is to start with the two approximations to the maximum EE that were used as fitness functions in the GA to come up with heuristic algorithms for channel assignment. Second, there is a memory - speed tradeoff in GA. By increasing the number of population $3 P$ in each generation, it is possible to reduce the number of generations $N$ gen necessary for convergence. Researchers were probably reluctant to attempt this until now because of the memory requirements in the computer. However, the latest computers have plenty of memory to handle this. Of course the memory-speed tradeoff cannot be pushed forever in one direction, as at one point the memory access latency of the computer might come into play. But how far we can push the memory to achieve speed in GA is an open problem. We should also be mindful of how much memory can be embedded in chip levels for real time applications.

## Chapter 12

## Competitive pricing for spectrum subleasing for future wireless ad hoc networks

### 12.1 Introduction

This chapter envisions a near future in which the proliferation of wireless ad hoc networks in urban centers causes excessive spectrum pollution on currently allocated unlicensed bands. One solution for this problem is for the operators to lease freshly released spectrum from the regulators and sublease it to agencies in major cities. We consider one such operator who divides an urban area into regions and subleases spectrum with the condition that the interference measured at boundary points should not exceed a threshold. The subleasing pricing structure has a fixed part, as well as a variable part that discounts the price based on the margin between the interference threshold and the actual interference. The operator has already decided on the fixed part of the spectrum subleasing price. It wishes to decide on the variable part by carrying out an optimization procedure.

The slope of the variable part is called the discount rate and is determined by a competition that is modeled as a game within a game. For a fixed discount rate, the competition between the customers forms a strategic game. The end result of this game becomes the input to the Stackelberg game between the customers as a whole on the one side and the operator on the other side. We derive the mild condition under which the strategic game of the customers has a unique Nash equilibrium, and obtain an explicit solution for the equilibrium point. This result is then used to derive the best response of the operator and the optimum (Stackelberg equilibrium) discount rate the operator would want to offer. Numerical results obtained through simulations that support the analysis are also provided.

### 12.2 Background

Wireless sensor networks are becoming widespread in urban centers because of their applications in monitoring air pollution, greenhouse temperature and humidity, and machine and structural health, to mention a few. Wireless mesh networks are being deployed because of their advantage in emergency situations and military applications. A large number of institutions are carrying out research on road safety and accident prevention based on the assumption that VANETs will be widely deployed in the future. At the moment all of these ad hoc networks have been served by the unlicensed frequency bands dedicated to the 802.11, 802.15 and 802.16 standards. Already, it has been reported that the ad hoc networks are often deployed in areas where the unlicensed bands suffer from excessive spectrum pollution [77]. It is not hard to envision these bands getting overcrowded in the near future. At the same time, the CRTC in Canada has not yet decided what to do with the newly freed TV white spaces. The FCC in United States still has many unoccupied bands that have not been leased [120].

We envision a near future, in which an operator leases spectrum from the FCC, CRTC, EU, OFTA, TRAI or similar organizations in other jurisdictions, breaks them into subchan-
nels and subleases them to city governments, health services, emergency departments and industries, or to a service provider who caters to these organizations. An urban center is divided into regions and on each region spectrum is subleased to one organization or service provider. Because of the absence of base stations and high power transmissions in ad hoc networks, unlike in the usual cellular systems, this operator allows frequency reuse in adjacent regions, but imposes the condition that the interference measured at boundary points must be less than a certain threshold.

However, in order to make sure the spectrum is utilized to the fullest and to maximize its profit, the operator uses a subleasing pricing that has a fixed part as well as a variable part. The variable part has a sliding scale that effectively discounts the price based on the margin between the interference threshold and the actual interference. The slope of the sliding scale is referred to as the discount rate ${ }^{1}$. The operator wishes to determine the optimum discount rate by weighing the revenue against the negative effects of interference experienced by the customers. This chapter shows that the situation faced by the operator can be modeled using a composite game - a game within a game. The outcome of this composite game will provide the operator with the optimum discount rate.

For a given discount rate, the competition among the customers becomes a non-cooperative strategic game. The Nash equilibrium outcome of this game becomes the input to the Stackelberg game between the operator on the one side and the customers on the other side. The operator anticipates the outcome of the strategic game as a function of the discount rate, and uses this information to arrive at the optimal discount rate.

Section 12.3 discusses related work. The system model and the game theoretic problem formulation are give in Section 12.4. Game theoretic analysis and the theorems that articulate the results are provided in Section 12.5. Numerical results that support the analysis, obtained through simulations, are presented in Section 12.6, followed by conclusion, in

[^1]Section 12.7.

### 12.3 Related work

Pricing of spectrum by a primary license holder who wishes to lease its spectrum within a subregion of its network is studied by Daoud et al. [73]. The optimization problem arising from the tradeoff between the revenue on the one side, and the reduced coverage of the original network and call blocking through interference on the other side, is analyzed using the reduced load approximation. A game theoretic frame work for dynamic spectrum leasing in a cognitive network is developed by Jayaweera et al. [74]. The primary user participates in a strategic game with the secondaries and uses the interference threshold as its strategy. The strategy of the secondaries is their power, and their utilities contain two parts: a rewarding function based on the received SNR and a penalizing term based on the difference between the actual interference and the interference threshold. Howayek et al. [75] extend this work to multiple channels and provide a distributed solution. The strategic interactions of two competing operators who lease spectrum from the owners and sublease it to secondary users is studied by Duan et al. [121]. The amount of spectrum to lease from the owner and the best price to sell it to secondary users, who want to maximize their individual profits, is modeled as a three-stage dynamic game.

Pricing by a primary user who sells part of his spectrum to a small number of secondaries is studied under an oligopoly market competition and Cournot game model by Niyato et al. [76]. The design of a secondary wireless mesh network that leases spectrum from a primary mesh network subject to interference constraints is taken up by Shakeri et al. [77]. Multiple primaries sharing their spectrum with secondaries through an intermediary broker is studied by Lin et al. [122]. The interaction between two wireless service providers and their customers is analyzed by Ya-Zhen et al. using a Wardrop Principle point of view [123]. A number of researchers use Stackelberg game to analyze the pricing of spectrum by primary
networks that leverage the secondary access points as cooperative relays [73-77,79, 80].


Only one interference path from a transmitter to another receiver is shown here in order not to crowd the diagram.

Figure 12.1: System Model

### 12.4 System model and the problem statement

The operator divides a metropolitan area into regions similar to today's wireless cells. In each region the operator subleases its spectrum only to one agency, on the condition that the interference measured at the boundary points do not exceed a threshold. The subleasing is based on the subchannel framework. An agency in a region can use the spectrum either
in an OFDMA based communication or dump all the subchannels onto a wide band and use CDMA. The subchannels in one regions are reused in the neighboring regions. The price per subchannel is $B-C\left(I_{0}-I\right)$ dollars/time-unit, where $B$ is the fixed part of the price, $I_{0}$ is interference threshold and $I$ is the total interference caused by that user on that subchannel measured at points on the boundary of the regions. $C$ is the discount rate measured in Dollars/micro Watts on that subchannel.

It is worthwhile noting here that while Jayaweera et al. [74] allow the interference to exceed the threshold but discourage it by having a negative term proportional to ( $I-I_{0}$ ) in the customers' utility, we do not allow the interference to exceed the threshold at all. The term $C\left(I_{0}-I\right)$ in our pricing scheme discourages the customers from approaching the interference threshold, unless they want to pay a higher price for the spectrum.

Consider $N$ node-pairs in neighboring regions shown in Fig. 12.1 using the same subchannel. In each pair, one node is transmitting to the other but the receiving node experiences interference from transmission by other pairs. Let $g_{i j}$ be the channel gain between the transmitting node $j$ and receiving node $i$. The equivalent channel gain between a transmitting node $i$ and the measuring points $M P$ 's is denoted by $h_{i}$. Note that at any particular moment only one node pair will be active on any given subchannel within a particular region.

Each node pair will try to strike a balance between the amount of throughput achievable in the subchannel with a certain amount of power and the cost of that power-subchannel combination. Henceforth, we will refer to node-pair $i$ as customer $i$. Using the Gaussian interference channel to model the subchannel, this trade-off can be quantified by a utility function $U_{i}$ for each customer $i$ given by,

$$
\left.\begin{array}{rl}
U_{i}=A \log _{2}\left(1+\frac{g_{i i} p_{i}}{\sigma_{i}^{2}+} \sum_{j \neq i} g_{i j} p_{j}\right. \tag{12.1}
\end{array}\right)
$$

where $\sigma_{i}^{2}$ stands for the additive Gaussian noise and $p_{i}$ the power employed by user $i$. The
constant $A$ is there to convert the transmission rate to corresponding monetary value. Letting $N_{i}=\sigma_{i}^{2} / g_{i i}$ and $a_{i j}=g_{i j} / g_{i i},(12.1)$ can be rewritten as

$$
\begin{equation*}
U_{i}=A \log _{2}\left(1+\frac{p_{i}}{N_{i}+\sum_{j \neq i} a_{i j} p_{j}}\right)-\left[B-C\left(I_{0}-h_{i} p_{i}\right)\right] \tag{12.2}
\end{equation*}
$$

A summary of the notations is given in Table 12.1.

| $g_{i j}$ | channel gain between transmitter $j$ and receiver $i$ |
| :---: | :---: |
| $p_{i}$ | power used by customer $i$ |
| $\sigma_{i}^{2}$ | additive white Gaussian noise density |
| $N_{i}$ | $\sigma_{i}^{2} / g_{i i}$ |
| $a_{i j}$ | $g_{i j} / g_{i i}$ |
| $I_{0}$ | interference threshold |
| $h_{i}$ | equivalent channel gain between a transmitting node $i$ and the measuring points MPs |
| $U_{i}$ | utility of custormer- $i$ |
| $B$ | fixed part of the spectrum subleasing price |
| $C$ | discount rate of the spectrum subleasing price |

Table 12.1: Notations

For the sake of mathematical completeness, we will assume that each $p_{i}$ belongs in a closed and bounded (compact) set of real numbers $[0, P]$, where

$$
\begin{equation*}
h_{i} P \leq I_{0} \text { for all } i . \tag{12.3}
\end{equation*}
$$

Note that the above requirement relieves us from worrying about $\left(I_{0}-h_{i} p_{i}\right)$ becoming negative in (12.2).

For a fixed $C$, each customer will try to buy the $p_{i}$ that maximizes its $U_{i}$. Unfortunately each $U_{i}$ also depends on the powers bought by the other users. This shows that the simultaneous maximization of $U_{i}$ 's by the customers in a subchannel forms a strategic game. Assuming this game has a unique Nash equilibrium, each customer will end up buying an amount of power that depends on the discount rate $C$.

The operator subleasing the subchannel must weigh the relative merit of the revenue with the negative effect of allowing frequency reuse in neighboring cells as opposed to the policy in the traditional cells. This negative effect depends on the level of interference measured. It can quantize this trade-off by defining a utility function

$$
\begin{equation*}
U_{o}=\sum_{i} B-C\left(I_{0}-h_{i} p_{i}\right)-T h_{i} p_{i} \tag{12.4}
\end{equation*}
$$

where $T$ is the constant representing the harmful effects of interference. The operator would want to choose a $C$ that will maximize $U_{o}$. However, the amount of power bought by the customers will depend on $C$. In other words there is $C$ "inside" each $p_{i}$. This shows that from the operator's point of view, the process of determining an optimal discount rate $C$ can be modeled as a Stackelberg game [27]- [78].

If the strategic game of the customers does have a unique Nash equilibrium for a fixed $C$, and if that equilibrium point can be explicitly calculated in closed form, then the operator can perform these calculations and express the power each customer will buy in terms of $C$ via functions $p_{i}=p_{i}(C)$. It can then replace the $p_{i}$ 's in (12.4) to write its utility $U_{o}$ as a function of $C$ alone. The operator can now optimize $U_{o}$ as a function of $C$ and find the $C$ that maximizes its utility.

### 12.5 Analysis

We prove that the strategic game of the customers does have a unique Nash equilibrium, provided that the channel coefficients satisfy a mild condition. We derive the best response functions of the customers and obtain a closed form solution for the equilibrium point. We then formulate the best response of the operator and derive the Stackelberg equilibrium discount rate $C^{*}$. In the forgoing, we will use the standard notation from game theory where $\mathbf{p}$ denotes the vector $\left(p_{1}, p_{2}, \ldots, p_{N}\right)$ containing all the powers, and $\mathbf{p}_{-\mathbf{i}}$ denotes the vector
containing all the powers except $p_{i}$.

Theorem 12.5.1. For a given discount rate $C$ offered by the operator, the strategic game among the customers with the utility functions in (12.2) has a unique Nash equilibrium if $\operatorname{det}(M) \neq 0$. Here $M$ is the matrix obtained from the channel matrix, by erasing all entries associated with the customers for whom $C \geq \frac{A}{h_{i} N_{i} \ln 2}$ or $C \leq \frac{A}{h_{i} \ln 2\left[N_{i}+\left(1+\sum_{j \neq i} a_{i j}\right) P\right]}$.

Proof: Differentiating $U_{i}$ in (12.2) twice with respect to $p_{i}$,

$$
\begin{equation*}
\frac{\partial^{2} U_{i}}{\partial p_{i}^{2}}=\frac{-A}{(\ln 2)\left(p_{i}+N_{i}+\sum_{j \neq i} a_{i j} p_{j}\right)^{2}}<0 . \tag{12.5}
\end{equation*}
$$

This shows that each $U_{i}$ is concave in $p_{i}$. Since each $p_{i}$ comes from a compact set, a theorem due to Debreu, Glicksberg and Fan [27] now guarantees the existence of a Nash equilibrium.

We will now derive the best responses $B R_{i}\left(\mathbf{p}_{-\mathbf{i}}\right)$ 's of the players. Starting from (12.2),

$$
\begin{equation*}
\frac{\partial U_{i}}{\partial p_{i}}=\frac{A}{(\ln 2)\left(p_{i}+N_{i}+\sum_{j \neq i} a_{i j} p_{j}\right)}-C h_{i} \tag{12.6}
\end{equation*}
$$

Letting $f(\mathbf{p})=\frac{1}{p_{i}+N_{i}+\sum_{j \neq i} a_{i j} p_{j}}$,

$$
\begin{equation*}
\frac{\partial U_{i}}{\partial p_{i}}=\frac{A}{\ln 2}\left[f(\mathbf{p})-\frac{C h_{i} \ln 2}{A}\right] \tag{12.7}
\end{equation*}
$$

Since $f(\mathbf{p})$ is decreasing in each $p_{i} \in[0, P],[f(\mathbf{p})]_{\text {Max }}=1 / N_{i}$ and $[f(\mathbf{p})]_{\text {Min }}=1 /\left[N_{i}+P(1+\right.$ $\left.\left.\sum_{j \neq i} a_{i j}\right)\right]$. Here, $[f(\mathbf{p})]_{\text {Max }}$ and $[f(\mathbf{p})]_{\text {Min }}$ denote the maximum and minimum values of $f(\mathbf{p})$, respectively.

Case: I $\quad C \geq \frac{A}{h_{i} N_{i} \ln 2}$
Eq. (12.7) shows $\frac{\partial U_{i}}{\partial p_{i}} \leq 0$ always, $U_{i}$ will be non-increasing and the maximum $U_{i}$ will occur at the boundary $p_{i}=0$. Hence, for a customer who belongs in this case, $B R_{i}\left(\mathbf{p}_{-\mathbf{i}}\right)=0$.
$\underline{\text { Case: II } \quad C \leq \frac{A}{h_{i} \ln 2\left[N_{i}+\left(1+\sum_{j \neq i} a_{i j}\right) P\right]}}$
$\frac{\partial U_{i}}{\partial p_{i}} \geq 0$ always, $U_{i}$ will be non-decreasing and the maximum $U_{i}$ will occur at the boundary $p_{i}=P$. In other words, for a customer who belongs in this case, $B R_{i}\left(\mathbf{p}_{-\mathbf{i}}\right)=P$.

Case: III $\frac{A}{h_{i} \ln 2\left[N_{i}+\left(1+\sum_{j \neq i} a_{i j}\right) P\right]}<C<\frac{A}{h_{i} N_{i} \ln 2}$
The maximum $U_{i}$ will occur at the critical point given by $\frac{\partial U_{i}}{\partial p_{i}}=0$. Using (12.7),

$$
\begin{equation*}
p_{i}=B R_{i}\left(\mathbf{p}_{-\mathbf{i}}\right)=\frac{A}{C h_{i} \ln 2}-N_{i}-\sum_{j \neq i} a_{i j} p_{j} \tag{12.8}
\end{equation*}
$$

For users who fall into Case-I or Case-II there is only one choice for best response and that will be their Nash equilibrium point. Whereas for the users in Case-III, the Nash equilibrium occurs when their best responses in (12.8) coincide. Suppose $S_{2}$ and $S_{3}$ denote the sets of integers containing the indexes of the users who fall into Case-II and Case-III, respectively. Then the Nash equilibrium power levels $\mathbf{p}^{*}$ of the users in Case-III will be given by,

$$
\begin{equation*}
p_{i}^{*}=\frac{A}{C h_{i} \ln 2}-N_{i}-P \sum_{j \in S_{2}} a_{i j}-\sum_{\substack{j \neq i \\ j \in S_{3}}} a_{i j} p_{j}^{*} \text { for all } i \in S_{3} \tag{12.9}
\end{equation*}
$$

Letting $r_{i}=\frac{A}{C h_{i} \ln 2}-N_{i}-P \sum_{j \in S_{2}} a_{i j}$,

$$
\begin{equation*}
p_{i}^{*}+\sum_{\substack{j \neq i \\ j \in S_{3}}} a_{i j} p_{j}^{*}=r_{i} \text { for all } i \in S_{3} \tag{12.10}
\end{equation*}
$$

This is a square system of linear equations in the equilibrium powers with a coefficient matrix $M$ with entries

$$
m_{i j}= \begin{cases}1, & \text { if } i=j  \tag{12.11}\\ a_{i j}, & \text { if } i \neq j\end{cases}
$$

and will have a unique solution

$$
\begin{equation*}
p_{i}^{*}=M^{-1}\left[r_{i}\right], \tag{12.12}
\end{equation*}
$$

if $\operatorname{det}(M) \neq 0$. Here, $\left[r_{i}\right]$ stands for the column matrix with entries $r_{i}$.

## Mildness of the condition $\operatorname{det}(M) \neq 0$

Note that the diagonal entries of the matrix $M$ are all 1's. Suppose we assume that the nearest interfering transmitter is at least 1.5 times as far away as your own transmitter. If we use a path loss coefficient of 4 , this would give an interfering channel coefficient of $a_{i j}=0.1975, j \neq i$. Now consider row $i$ of matrix $M$. Since $0.1975 \times 5<1$, even with 6 customers we will have $\left|m_{i i}\right|>\sum_{j \neq i}\left|m_{i j}\right|$. This shows we can go up to 6 customers in the same subchannel, and the matrix $M$ will remain diagonally dominant. The determinant of a diagonally dominant matrix cannot be zero. This argument shows how mild the condition on the uniqueness of the equilibrium is.

As usual in Stackelberg game model, the operator anticipates the optimizing process the customers went through using the strategic game, mimics the calculations, and comes to the same conclusion as Theorem 12.5.1. It can now use this conclusion to decide the discount rate that maximizes its utility. However it should be mindful of not extending the conclusion of the theorem into regions where the interference exceeds the threshold.

Theorem 12.5.2. The Stackelberg game between the operator with utility in (12.4) and the costumers with utilities in (12.2), has a competitive equilibrium discount rate $C^{*}$ given by,

$$
\begin{equation*}
C^{*}=\operatorname{Max}\left\{\sqrt{\frac{A T \sum_{i \in S_{3}} h_{i} k_{i}}{\ln 2 \sum_{i \in S_{3}} I_{0}+h_{i} l_{i}}},\left[\frac{A h_{i} k_{i}}{\ln 2\left(I_{0}+h_{i} l_{i}\right)}\right]_{M a x}\right\}, \tag{12.13}
\end{equation*}
$$

where $k_{i}$ and $l_{i}$ are defined at the beginning of the proof below.
Proof: Let $\left[1 / h_{i}\right]$ denote the column matrix with entries $1 / h_{i}$ and $k_{i}$ denote the entry in the $i$ th row of the column vector that is obtained by the matrix product $M^{-1}\left[1 / h_{i}\right]$. Similarly,
use $l_{i}$ to denote the entry in the $i$ th row of the result of the product $M^{-1}\left[N_{i}-P \sum_{j \in S_{2}} a_{i j}\right]$. We can now use (12.12) and write the dependency of the Nash equilibrium power $p_{i}$ on $C$ as

$$
\begin{equation*}
p_{i}=\frac{A}{C \ln 2} k_{i}-l_{i}, \text { where } p_{i} \in[0, P] \text { and } h_{i} P \leq I_{0} \tag{12.14}
\end{equation*}
$$

Substituting this in the utility of the operator

$$
\begin{equation*}
U_{o}=\sum_{i \in S_{3}} B-C I_{0}+\frac{A h_{i} k_{i}}{\ln 2}-C h_{i} l_{i}-\frac{A T h_{i} k_{i}}{C \ln 2}-T h_{i} l_{i} \tag{12.15}
\end{equation*}
$$

from which,

$$
\begin{equation*}
\frac{d^{2} U_{o}}{d C^{2}}=-\sum_{i \in S_{3}} \frac{2 A T h_{i} k_{i}}{C^{3} \ln 2}<0 \tag{12.16}
\end{equation*}
$$

This shows that $U_{o}$ is concave in $C$. If the domain of $U_{0}$ is unlimited, then the maximum of $U_{o}$ will occur when

$$
\begin{equation*}
\frac{d U_{o}}{d C}=\sum_{i \in S_{3}}-I_{0}-h_{i} l_{i}+\frac{A T h_{i} k_{i}}{C^{2} \ln 2}=0 \tag{12.17}
\end{equation*}
$$

from which,

$$
\begin{equation*}
C=\sqrt{\frac{A T \sum_{i \in S_{3}} h_{i} k_{i}}{\ln 2 \sum_{i \in S_{3}} I_{0}+h_{i} l_{i}}} \tag{12.18}
\end{equation*}
$$

However, the conditions in (12.14) limit the domain of $U_{0}$ and the maximum of $U_{0}$ could occur at the end point where one of the customers' interference reaches the threshold $I_{0}$. This will happen when

$$
\begin{equation*}
p_{i}=\frac{I_{0}}{h_{i}} \text { for some } i . \tag{12.19}
\end{equation*}
$$

Using (12.14),

$$
\begin{equation*}
\frac{A}{C \ln 2} k_{i}-l_{i}=\frac{I_{0}}{h_{i}} \tag{12.20}
\end{equation*}
$$

Rearranging,

$$
\begin{equation*}
C=\frac{A h_{i} k_{i}}{\ln 2\left(I_{0}+h_{i} l_{i}\right)} \tag{12.21}
\end{equation*}
$$

Note that this equation gives us many $C$ values. We must watch out for the highest of these values because the moment the $C$ value in (12.18) becomes less than this, the position of the maximum will switch from (12.18) to (12.21). Keeping this in mind, we finally arrive at the the Stackelberg discount rate $C^{*}$ as,

$$
\begin{equation*}
C^{*}=\operatorname{Max}\left\{\sqrt{\frac{A T \sum_{i \in S_{3}} h_{i} k_{i}}{\ln 2 \sum_{i \in S_{3}} I_{0}+h_{i} l_{i}}}, \quad\left[\frac{A h_{i} k_{i}}{\ln 2\left(I_{0}+h_{i} l_{i}\right)}\right]_{\operatorname{Max}}\right\} \tag{12.22}
\end{equation*}
$$



Figure 12.2: Simulation details (Only one of the 4 transceivers is shown here.)

### 12.6 Numerical results

In this section, we present numerical results obtained through simulations that support the analysis of the previous section. We placed four customer transceivers randomly and used a path loss exponent of 4 to calculate the channel gains. Further details of the simulation are given in Fig. 12.2.


Figure 12.3: Utility function of Customer 1


Figure 12.4: Convergence of the fixed point algorithm


Figure 12.5: Dependence of equilibrium power on the distance from the measuring point.


Figure 12.6: Utility of the operator

For one of the placement, the utility function of customer 1 as a function of its power, while the power levels of the others were fixed at 0.1 Watts, for four different values of the discount rate $C$ set by the operator are shown in Fig. 12.3. The discount rate numbers in this figure are in 100's of Dollars/micro Watts. We observe that the utilities are concave just as claimed in the proof of Theorem 12.5.1. The vertical dashed lines in Fig. 12.3 indicate the analytical best responses of customer 1 for each discount rate as given by (12.8). We see these values agree with the maximum values of the graphs. We also see that as the discount rate increases, the customer is buying less power, so it can stay far way from the interference threshold and pay a lesser price for the spectrum.

Next we provide a result that verifies our closed form solution for the Nash equilibrium powers of the customers. For a particular placement of the transceivers, we started with the power levels of 0.1 Watts for all, and then ran the simple fixed point algorithm $p_{i}^{*}(t+1)=$ $B R_{i}\left(\mathbf{p}_{-\mathbf{i}}^{*}(t)\right)$ for $i=1,2,3,4$. The convergence of this algorithm, to the power levels of the Nash equilibrium dictated by Theorem 12.5.1 in (12.14), is shown in Fig. 12.4.

The effect of the distance from a measuring point on a customer's equilibrium power is investigated in Fig. 12.5. As expected, the farther away from the measuring point the customer is, the greater its power is.

The utility of the operator for a particular realization of the customers is shown in Fig. 12.6. The vertical dashed line indicates the analytical Stackelberg equilibrium discount rate given by Theorem 12.5.2, which is in close agreement with the simulated maximum seen from the graph.

### 12.7 Conclusion

This chapter used the structure of a game within a game to model and analyze the pricing scheme offered by an operator who subleases spectrum to wireless ad hoc networks. The operator divides an urban center into regions and subleases spectrum to one agency in each
region on the condition that interference measured at boundary points do not exceed a threshold. The price of spectrum has a fixed part as well as a variable part. The variable part is essentially a sliding scale that offers a discount based on the margin between the interference threshold and the actual interference. The operator uses a Stackelberg game inside a strategic game to competitively decide the discount rate. For a fixed discount rate, the competition among the customers on a subchannel becomes a strategic game. We proved that this strategic game has a unique Nash equilibrium and obtained an explicit closed form solution for the equilibrium point. This result then becomes the input to the Stackelberg game between the customers as a whole against the operator. The best response of the operator was derived, and was used to obtain the optimum (Stackelberg equilibrium) discount rate the operator would want to offer. We also presented numerical results obtained through simulations that supported our analysis.

## Chapter 13

## Conclusion and future work

In Chapter 2, we opined that unless the techniques to solve nonconvex problems approach convex optimization techniques, in terms of speed, they would not be useful for the industry. This dissertation presented solutions to five non-convex problems, four of which have speeds equal to or better than convex optimization techniques. In solving the optimization problems, this dissertation contributed both in theory as well as algorithm development. The next section summarizes these contributions while the section that follows discusses possible directions for future research.

### 13.1 Contributions

### 13.1.1 Theory

Chapter 8 presented a new approach to solving a resource allocation problem with proportional output constraints. In the absence of any knowledge about the geometrical structure (convex or not) of the objective function, it used analytical geometry to locate the global extrema.

Chapter 9 established the generalized convex structure (invex) of the GIC along the way
to characterize the global optimum point using the KKT condition. To the best of our knowledge, this was the first application of invexity to a real world problem.

Another first in the optimization field was presented in Chapter 10. To the best of our knowledge, this was the first time the equations from the KKT conditions from a fairly complex asymmetrical problem were solved directly to obtain the solution to the optimization problem. Until now the common water level that led to the EE maximizing power allocation in OFDMA was obtained through iterative algorithms. Chapter 10 showed that this water level can be obtained by solving a single non-linear equation.

The first ever application (to the best of our knowledge) of genetic algorithm (GA) for channel allocation that maximizes the EE in OFDMA was presented in Chapter 11. It is not clear why other researchers did not consider genetic algorithm for this purpose. It is possible that they were not successful in overcoming a certain problem regarding the fitness function. The fitness function can only be calculated by going through another optimization procedure, and this has to be done for every individual in every generation. This is prohibitive computationally and would make the application of GA worthless. An approximation to the fitness function could have been used but, as mentioned before, the maximum possible EE was obtained through an iterative algorithm. It is not possible to approximate the end result of a such an algorithm. Our work on Chapter 10 with proportional rate constraints gave us the suitable closed form approximation.

Chapter 12 used a game within a game structure (a composite game) to simultaneously solve several interconnected optimization problems. At the time of the publication of this work we thought we were the first one to use a composite game. A few months later we stumbled upon a paper by other authors that had used a game within game structure for a different problem. Upon reading the paper we realized that the authors did not realize the theoretical significance of what they have done. The phrases "game within a game" or "a composite game" were not mentioned in that paper.

### 13.1.2 Development of algorithms

This dissertation presented six state of the art optimization algorithms; the new idea used in one of these algorithm (the projected gradient) has application beyond the current problem it solves - in the broader area of convex optimization. Table 13.1 lists these algorithms.

| Objective | Variable | Constraints | Technique | Complexity |
| :---: | :---: | :---: | :---: | :---: |
| Throughput in GIC | Power | Prop. rates <br> Total power | Analytical Geometry | O( $\mathrm{N}^{3}$ ) |
| $\begin{gathered} \text { Throughput } \\ \text { in } \\ \text { GIC } \end{gathered}$ | Power | Minimum rates <br> Total power | Invex Analysis | $O\left(N^{3}\right)$ |
| Energy Efficiency in OFDMA | Power | Minimum rates <br> Total power | CCT <br> Transform. | O(N) |
| Energy Efficiency in OFDMA | Frequency | Minimum rates <br> Total power | Genetic Algorithm | N/A |
| Spectrum Subleasing Revenue | Price | Interference | Game <br> Theory | $\mathrm{O}\left(\mathrm{N}^{3}\right)$ |

Figure 13.1: The optimization problems solved

## Spectral efficiency in GIC

Chapter 8 demonstrated how to go from an analytical description of the optimum point to an algorithm that can locate that point in practical time. Initial description of the optimal
point was then re-interpreted so that mature technologies can be used to arrive at an efficient and stable algorithm.
Chapters 8 and 9 offered the first ever throughput maximizing algorithms with total power constraints. Three of them used proportional rate and total power constraints ( $N$ is the number of users in the system):

- A centralized algorithm with complexity $N^{3}$.
- A distributed algorithm with linear complexity.
- A distributed algorithm that uses a shared memory with constant complexity.

The fourth algorithm worked with minimum rate and total power constraints. It is a centralized algorithm with complexity $O\left(N^{3}\right)$.

## Energy efficiency in OFDMA

In OFDMA, the dissertation offered the energy efficiency maximizing power allocation via the solution of a single non-linear equation. For subchannel assignment that maximizes the EE in OFDMA, the dissertation proposed a solution to the problem using genetic algorithm with an approximate fitness function.

## Subchannel pricing for ad-hoc networks

An explicit solution for the price was obtained in terms of the solution of a linear system of equations.

### 13.2 Future directions for research

### 13.2.1 EE in GIC

## Centralized approach

To the best of our knowledge, energy efficiency (EE) of transmissions in the Gaussian interference channel (GIC) has not been addressed in the literature. This is understandable given that EE is the ratio of the throughput to the sum power, and throughput maximization itself remained difficult enough. Now that we have established the generalized convex structure of the throughput in the GIC, attacking the EE may be plausible.

In the OFDMA where the throughput is concave in the powers (and the sum power is affine), the EE was shown to be quasi-concave. Charnes-Cooper transformation was then used to convert the EE into a concave function. In this dissertation, we showed that the throughput in the GIC is incave. So what structure might EE in GIC have? This is indeed an intriguing future research direction.

## Distributed approach

Consider the power allocation that optimizes the EE in GIC. Suppose we are willing to settle for a competitive equilibrium among the users who attempt to maximize their individual EE, motivated by the possibility of a distributed solution. Each user's throughput is concave and its power is linear. Hence its EE is quasi-concave, and can be converted to a concave objective by Charnes-Cooper Transformation (CCT). The strategic game with the new objectives as the payoff functions will have a Nash equilibrium. We can also experiment by creating a utility function that is the difference between the above payoff function and a function of the user's power expenditure.

### 13.2.2 EE in OFDMA

## Centralized approach

In this dissertation we had used genetic algorithm for EE maximizing subchannel allocation. Obviously, this approach for subchannel assignment has high computational complexity. The main purpose in the dissertation was first, to see if we can improve on the EE attained by simple heuristic channel assignment protocols, and second, to test the effectiveness of the closed form indicators of the maximum possible EE from a given channel assignment. Now that we know these indicators are effective in improving the EE, we can attempt to come up with new heuristic channel assignment protocols based on these indicators.

## Distributed approach

Every thing mentioned in the last section for GIC applies here verbatim with the qualification that we are talking about the power allocation that maximizes the EE for a fixed channel assignment.

### 13.2.3 Is the throughput in GIC actually concave?

A significant theoretical progress made in this dissertation was the proof that the throughput of the GIC is invex and incave in the powers. Is there more to this than it appears? Is it possible that the throughput is actually concave?

If it is concave what are the ramification? Of course, throughput maximization is already solved in this dissertation with the same speed as convex optimization. Whether the throughput restricted to the total power constraint is incave or concave, the best algorithm to find the optimal power allocation will be the same - the one presented in this dissertation. However, if the throughput is actually concave then the EE maximization in the GIC opens up. Charnes-Cooper transformation can be now applied in the GIC as well. We probably will not have common water level solution like in the OFDMA, but certainly we can have
a solution based on an iterative algorithm. Hence the question of if the throughput of the GIC is concave is worth investigating.

## Appendix A

## Review of power allocation in CDMA and GIC using game theory

In GIC, each user's throughput depends on all the users' power levels. If a user unilaterally increase its power, its transmission rate will increase but all the other user's transmission rates will decrease because of the increase in interference. However, the throughput of each user is concave in that user's power. If all users attempt to maximize their transmission rates at the same time, the resulting simultaneous optimization process becomes a strategic game. Nash equilibrium will provide the "optimum" power levels for the users. An important point to note here is that a total power constraint on the users in the system cannot be used here. Only individual power constraints can be used. The strategy space of a user could be $[0, P]$, where $P$ is the maximum transmission power that user can employ.

If each user's payoff function is its transmission rate and the strategy space is $\left[0, P_{i}\right]$, then it can be shown that the Nash equilibrium will occur at the power levels $P_{i}$. That is, each user will use the maximum transmission power it can. This will change if we use the concepts of utility and pricing from economics, and use a payoff function that is supposed to indicate the "level of satisfaction" of each user.

## A. 1 Power control for CDMA

Application of game theory to power control in wireless communication originated at Rutgers University in New Jersey. In a series of papers published between 1998 and 2002, Mandayam, Goodman, and their students developed a new framework for power control in CDMA networks using the concepts of utility and pricing from economics [36] - [39]. Mandayam and Goodman [36] - [39] considered the CDMA based uplink of a single cell, where user $i$ transmits $L$ information bits in frames of length $M$ bits, at a rate of $R$ bits/seconds using $p_{i}$ Watts of power. They formed a utility function $u_{i}$ that is indicative of the "level of satisfaction" of each user $i$.

$$
\begin{equation*}
u_{i}=\frac{L R\left(1-2 P_{e}\left(\gamma_{i}\right)\right)^{M}}{M p_{i}} \tag{A.1}
\end{equation*}
$$

where $P_{e}$ is the bit error rate and $\gamma_{i}$ is the SINR of user $i$ at the receiver. If $h_{i}$ is the path gain from user $i$ to the base station,

$$
\begin{equation*}
\gamma_{i}=\frac{W}{R}\left[\frac{h_{i} p_{i}}{N+\sum_{j \neq i} h_{j} p_{j}}\right] \tag{A.2}
\end{equation*}
$$

where $W$ is the available spread-spectrum bandwidth, and $N$ is the AWGN power at the receiver ${ }^{1}$.

Mandayam and Goodman proceeded to formulate a strategic game played by the users currently in the cell, where each user's power $p_{i}$ is its strategy, and $u_{i}$ its payoff function. Since the utility function $u_{i}$ is concave in $p_{i}$, this strategic game has a Nash equilibrium. Mandayam and Goodman proved that at the Nash equilibrium, all users end up with the same SINR at their receivers.

In order to move the equilibrium towards a point that is Pareto optimal, they modified the payoff function by introducing the pricing of the power. The new payoff function $u_{i}^{c}$

[^2]is the difference of two functions: first is the utility function $u_{i}$ above and the second is a function that is proportional to the power employed by the user.
\[

$$
\begin{equation*}
u_{i}^{c}=u_{i}-c \alpha_{i} p_{i}, \tag{A.3}
\end{equation*}
$$

\]

where $c$ is a constant for the system and $\alpha_{i}$ is a constant for user $i$. Mandayam and Goodman proceeded to note that the new payoff function is not quasi concave and, therefore, the techniques that are usually used to prove the existence of Nash equilibria under the strong assumption of convexity are no longer applicable. Instead they turned to supermodular game theory and proved that the new game has many Nash equilibria.

Sung et al. derived a particular transmission technology specific upper and lower bounds on capacity and employed them in their utility function [42]. They also changed the independent variable from transmission power to received power. Sung et al. then introduced pricing to shape the users' behavior so as to improve the performance from a system viewpoint. They suggested a pricing based on normalized interference and formulated the pricing function for user $i$ as

$$
\begin{equation*}
c_{i}=\frac{\lambda Q_{i}}{N+\sum_{j} Q_{j}} \tag{A.4}
\end{equation*}
$$

where $\lambda$ is a constant. They proceeded to prove the existence and uniqueness of the Nash equilibrium to TMG with pricing.

Ghasemi et al. went back to Mandayam et al.'s utility function and made a modification to the pricing function [43]. Instead of making the pricing function proportional to the user power, Ghasemi et al. made it proportional to the user SINR. They motivated the new pricing function in the following way. A pricing function based solely on power does not take into account the channel conditions. Whereas a pricing function proportional to the SINR would impose a lower price for users with good channel conditions. They argued this is consistent with the idea that the network efficiency and throughput are improved when terminals with bad channel conditions are forced to work with smaller SINR. Ghasemi et
al. proceeded to prove the existence of a Nash equilibrium and provided simulation results that show that pricing based on SINR leads to equilibrium points with less power but higher utility compared to pricing based on power.

## A. 2 Power control for early cognitive radio

At the beginning of the this millennium, a survey of the crowded spectrum [125] revealed that the frequency bands permanently licensed to many organizations are not utilized around the clock. This led to the proposal of a concept called cognitive radio [126-128] devices that are envisioned to eventually have the capabilities of scanning the airwaves and choosing the empty bands to be used in their communication. In the early stages of cognitive radio, a model considered by researchers called overlay envisioned cognitive (secondary) devices to use even the frequency bands that are currently in use, as long as the cognitive devices are geographically separated from the licensed (primary) devices with the condition that the interference to the licensed devices is below a threshold. At that time, many researchers considered CDMA to be the technology used by cognitive devices. It should be noted that we decided to include the review of the literature on power control for cognitive radio only because CDMA is a special case of GIC when it comes to optimizing power allocation.

The first game theoretic analysis of power control for a cognitive network appears on a work by Wang, Peng and Wang in 2006 [44]. Wang et al. formed a power control game model for the unlicensed users in a cognitive network that is not much different from the uplink of a single cell considered in Mandayam and Goodman, where, a number of cognitive transmitters all communicate with a cognitive base station. This allowed Wang et. al to directly borrow the CDMA interference model of Mandayam and Goodman with a small modification. The total interference and noise $I_{i}$ for the $i$ th cognitive receiver (in the base station) is

$$
\begin{equation*}
I_{i}=N+Q_{p}+\sum_{j \neq i} Q_{j} \tag{A.5}
\end{equation*}
$$

where $N$ is the AWGN power and $Q_{p}$ is the total interference caused by all the licensed users. They also seemed to use the same utility function as Mandayam and Goodman, however, this was not clear in their paper. The pricing function $c_{i}$ has two parts $c_{i}^{(1)}$ and $c_{i}^{(2)}$. The first part $c_{i}^{(1)}$ is identical to the one used by Sung et al [42]. The novelty is in the second part $c_{i}^{(2)}$, which discourages the interference to the licensed users.

$$
\begin{equation*}
c_{i}^{(1)}=\frac{\lambda Q_{i}}{Q_{i}+I_{i}} \quad c_{i}^{(2)}=\frac{\lambda Q_{i}}{Q_{i}+I_{i}} \exp \left(\frac{Q_{i}+I_{i}-Q_{p}-Q_{T}}{Q_{T}}\right) \tag{A.6}
\end{equation*}
$$

where $Q_{T}$ is a constant that denotes the maximum allowable interference to the licensed users. Wang et al. mentioned that the interference to the licensed users are measured at the licensed user's receivers but failed to mention where these receivers are located. Judging from the formulation of $c_{i}^{(2)}$, it appears that the licensed receivers are located at the cognitive base station. The authors concluded the paper by proving the existence and uniqueness of the Nash equilibrium for their power control game with pricing.

## A.2.1 Sigmoid function based utility

Xia and Qi formed a utility function [45] that is based on the sigmoid function , [46].

$$
\begin{equation*}
u_{i}\left(p_{i}\right)=\frac{L R}{M p_{i}\left(1+\exp \left(10-\gamma_{i}\right)\right)} \tag{A.7}
\end{equation*}
$$

where $L, M$ and $R$ are as in Mandayam and Goodman (number of information bits, packet length and transmission rate respectively). They used a pricing function that includes the power as well as the path gain and form the payoff function as

$$
\begin{equation*}
u_{i}\left(p_{i}\right)=\frac{L R}{M p_{i}\left(1+\exp \left(10-\gamma_{i}\right)\right)}-\lambda h_{i} p_{i} \tag{A.8}
\end{equation*}
$$

Xia and Qi showed that their game is supermodular and proved the existence and uniqueness of the Nash equilibrium.

## A.2.2 Game with fairness

Wang et.al proposed a power control game for cognitive radios with fairness [47], [48]. They considered the uplink of a cognitive radio network with a cognitive base station and expressed the SINR of user $i$ as

$$
\begin{equation*}
\gamma_{i}=\frac{p_{i} h_{i}}{N+Q_{p}+\sum_{j \neq i} p_{j} h_{j}} \tag{A.9}
\end{equation*}
$$

where $Q_{p}$ is the interference brought by the licensed users. They used the throughput of user $i$ as the utility and defined

$$
\begin{equation*}
u_{i}\left(\gamma_{i}\right)=R_{i} f\left(\gamma_{i}\right) \tag{A.10}
\end{equation*}
$$

where $R_{i}$ is the transmission rate and $f\left(\gamma_{i}\right)$ denotes the probability of correct reception at the receiver. The pricing function was defined as

$$
\begin{equation*}
c_{i}\left(p_{i}\right)=\lambda_{1} \frac{I_{i} \bar{h}_{i}}{h_{i}} p_{i}+\lambda_{2} \sum_{m} p_{i} h_{m}^{2} \tag{A.11}
\end{equation*}
$$

where $I_{i}$ is the total interference user $i$ suffers, $h_{m}$ is the path gain from user $i$ to licensed base station $m, h_{i}$ is the path gain to the cognitive base station, $\bar{h}_{i}$ is the average of $h_{i}$ in the last time slot and $\lambda_{1}, \lambda_{2}$ are pricing coefficients. The second term above imposes a penalty for the cognitive users for the interference they cause to the licensed users.

According to Wang et al., the first part of the pricing function takes into consideration the throughput and fairness. $I_{i} / h_{i}$ reflects the condition of the transmission channel: if channel $i$ is in a good condition, then $I_{i} / h_{i}$ will be small, and in order to use this situation opportunistically the price will be small. But this is unfair for the users who are far from the base stations or users with bad channels, and $\bar{h}_{i}$ was introduced to overcome this. They argued that if channel $i$ was in a good condition during the last time slot, $\bar{h}_{i}$ will be higher, which will impose a higher price for user $i$ during the current time slot, thus introducing some fairness. Wang et al. then established the existence and uniqueness of Nash equilibrium for their power control game.

| Researchers | Game Participants | Strategy Space | Payoff Function |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Users or Secondary Users | Primary Users | Utility Based on | Pricing Based on |
| Mandayam et al. |  | N/A | Number of bits/Joule* | Power |
| Ghasemi et al. |  | N/A | Number of bits/Joule* | SINR |
| Sung et al. |  | N/A | Bounds on Number of bits/Joule* | Normalized Interference |
| Wang et al. |  | No | Number of bits/Joule* | Normalized Interference, Interference to Primary |
| $\mathbf{X i a}$ and Qi |  | No | Sigmoid Function | Power, Path Gain |
| Zhu et al. |  | No | Number of bits/Joule* | Throughput, Fairness |
| Li and Jayaweere |  |  | Number of bits/Joule* | SINR |

* Number of bits correctly received per energy expended (transmission technology specific).

Figure A.1: A summary of game theory work on power control

## A. 3 Power control in the GIC

Competition for transmission rates in the Gaussian interference channel was modeled using a strategic game by Yu and Cioffi [40]. They considered two transmitter receiver pairs communicating over a single channel with bandwidth F , with power distributions $P_{1}(f)$ and $P_{2}(f)$ and constraints $\int_{0}^{F} P_{1}(f) d f \leq P_{1}^{\max }$ and $\int_{0}^{F} P_{2}(f) d f \leq P_{2}^{\max }$. Each user was trying to maximize its own transmission rate and the payoff functions of the two users were defined using the Shannon's formula:

$$
\begin{align*}
& u_{1}=\int_{0}^{F} \log \left(1+\frac{P_{1}(f)}{N_{1}(f)+a_{2}(f) P_{2}(f)}\right) d f  \tag{A.12}\\
& u_{2}=\int_{0}^{F} \log \left(1+\frac{P_{2}(f)}{N_{2}(f)+a_{1}(f) P_{2}(f)}\right) d f \tag{A.13}
\end{align*}
$$

Yu and Cioffi pointed out that for a fixed power distribution $P_{1}(f)$ of user one, the best response by user two is the water filling of its power with respect to the combined noise and, interference from $P_{1}(f)$. Similarly, for a fixed $P_{2}(f)$, the best response by user one is the water filling of its power with respect to the combined noise and, interference from $P_{2}(f)$. If there are power distributions $P_{1}(f), P_{2}(f)$ such that water filling is achieved simultaneously for both users, a Nash equilibrium is reached. They proved that indeed such $P_{1}(f)$ and $P_{2}(f)$ exist, if the interference coefficients $a_{1}$ and $a_{2}$ satisfy a certain condition.

Yu and Cioffi's work is best understood by transforming their channel model into the one we have been using in this document so far. Break up the channel of bandwidth F into $K$ number of subchannels of small enough bandwidth so that the fading is flat in each of the subchannels. Now we can consider two transceiver pairs trying to communicate over these $K$ subchannels. Each subchannel is used by both users and the power allocation problem is to decide how each user will distribute its power among the $K$ subchannels.

If one user fixes his powers, then the optimal power distribution for the other user is the watter filling with respect to the combined noise and interference. If there are power
distributions that watter fill each user with respect to the combined noise and interference, then it must be a Nash equilibrium as no user can increase its rate by unilaterally deviating from this power allocation. What Yu and Cioffi did is to prove that such a power allocations must exist.

## A.3.1 Iterative water filling

The procedure Yu and Cioffi used to prove the existence of the Nash equilibrium in the two user Gaussian interference channel game is actually an algorithm that can take us to this equilibrium point. In the years that followed, this algorithm has come to be known as iterative water filling algorithm (ITF). One user starts with an arbitrary power distribution. The other uses water filling to compute his power allocation based on this. The first reallocates its power using water filling based on the second user's power distribution and so on until each attains water filling with respect to the other. The convergence of the iterative watter filling algorithm for the two user case is proved by Yu and Cioffi in a later paper in 2002 [59]. Leung et al. proved that the iterative water filling algorithm would converge even if the users simultaneously updated their powers [60].

## Appendix B

## List of Publications

## B. 1 Journal Publications

1. On the power allocation problem in the Gaussian interference channel with proportional rate constraints
K. Illanko, A. Anpalagan, E. Hossain, and D. Androutsos

IEEE Transactions on Wireless Communications, vol. 13, issue 2, 2014, pp. 1101 1115.
2. Frequency and power allocation for energy efficient OFDMA systems with proportional rate constraints
K. Illanko, M. Naeem, A. Anpalagan, and D. Androutsos

IEEE Wireless Communication Letters vol. 3, issue 3, 2014, pp. 313-316
3. Energy efficient frequency and power allocation for cognitive radios in television systems
K. Illanko, M. Naeem, A. Anpalagan, and D. Androutsos

Under revision in IEEE Systems Journal.
4. Decode and forward relaying for energy-efficient multiuser cooperative cog-
nitive radio network with outage constraints
M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jaseemuddin

IET Journal of Communications, vol. 8, issue 5, 2014, pp. 578-586
5. Optimal power allocation for green cognitive radio: fractional programming approach
M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jesemuddin IET Journal of Communications, vol. 7, issue 12, 2013, pp. 1279-1286

## B. 2 Conference Publications

1. Energy efficiency of cooperative cognitive radio network with outage constraints
M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jaseemuddin Proc. IEEE Radio and Wireless Symposium, Newport Beach, CA, USA, Jan 19-23, 2014
2. Power allocation in decode and forward relaying for green cooperative cognitive radio systems
M. Naeem, K. Illanko, K. A. Karmokar, A. Anpalagan, and Jaseemuddin Proc. IEEE Wireless Communications and Networking Conference, Shanghai, China, April 7-10, 2013
3. Low complexity energy efficient power allocation for green cognitive radio with rate constraints
K. Illanko, M. Naeem, A. Anpalagan, and D. Androutsos

Proc. IEEE Global Communications Conference, Anaheim, USA, December 3-7, 2012.
4. Competitive pricing for spectrum subleasing for future wireless ad hoc networks
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. IEEE International Conference on Communications, Ottawa, Canada, June 10-15, 2012
5. Iterative power allocation for downlink green cognitive network M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jaseemuddin Proc. IEEE Global Communications Conference, Anaheim, USA, Dec. 3-7, 2012.
6. Convex structure of the sum rate on the boundary of the feasible set for coexisting radios
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. IEEE International Conference on Communications, Kyoto, Japan, June 5-9, 2011
7. Stackelberg game on the boundary of coexistence
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. IEEE International Conference on Communications, Kyoto, Japan, June 5-9, 2011
8. Dual methods for power allocation for radios coexisting in unlicensed spectra
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. IEEE Global Communications Conference, Miami, Florida, USA, December 6 10, 2010.
9. An Optimal and fair distributed algorithm for power allocation for radios coexisting in unlicensed spectra
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. IEEE International Conference on Communications, Cape Town, South Africa, May 23-27, 2010.
10. Adaptive demodulation using Raptor codes: Analytical results and extension to fading channels
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. 11th Canadian Workshop on Information Theory, Ottawa, Canada, May 13 15, 2009
11. Cooperative communication using bit-selective adaptive demodulation and Raptor codes: The Gaussian relay channel case
K. Illanko, A. Anpalagan, and D. Androutsos

Proc. IEEE Vehicular Technology Conference (VTC), Barcelona, Spain, April 26-29, 2009
12. Sub-channel and power allocation for multiuser OFDM with fairness using genetic algorithm
K. Illanko, K . Raahemifar, and A. Anpalagan

Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Aug. 23-26, 2009

## Bibliography

[1] M. A. Banks, "On the Way to the Web: The Secret History of the Internet and Its Founders," Apress, New York, 2008.
[2] B. M. Leiner, "A Brief History of the Internet," Internet Society, December, 2003.
[3] B. Stewart, "Internet History," Living Internet web book, http://www.livinginternet.com/, last accessed on August 20, 2014.
[4] Phonearena news, www.phonearena.com/news/, last accessed on July 16, 2014.
[5] H. Schulze and C. Luders, "Theory and Applications of OFDM and CDMA," John Wiley and Sons Ltd., 2005.
[6] H. H. Chen, J. F. Yeh, and N. Suehiro, "A multicarrier CDMA architecture based on orthogonal complementary codes for new generations of wideband wireless communications," IEEE Communications Magazine, vol. 39, issue 10, 2001, pp. 126-135.
[7] L. Robert, L. N. Pissinou, and S. Makki, "Third generation wireless network: the integration of GSM and Mobile IP," Proc. IEEE Wireless Communications and Networking Conference, Chicago, USA, Sept. 23-28, 2000.
[8] T. S. Rappaport, "Wireless Communications," Prentice Hall, 2002.
[9] V. Fotheringham, and C. Sharma, "Wideband 3G to Broadband 4G: Collision and Convergence of Standards," Chapter 10, Wireless Broadband: Conflict and Convergence, Wiley ebook, 2008.
[10] N. Himayat, S. Talwar, A. Rao, and R. Soni, "Interference management for 4G cellular standards [WIMAX/LTE UPDATE]," IEEE Communications Magazine, vol. 48, issue 8, 2010, pp. 86-92.
[11] E. Basar, and U. Aygolu, "Channel capacity achieving space-time block codes for MIMO systems with two, three and four transmit antennas," Proc. IEEE Signal Processing and Communications Applications Conference, Antalya, Turkey, April 9-11, 2009.
[12] International Standard ISO/IEC 74-98-1, http://standards.iso.org/ittf/licence.html, last accessed on July 17, 2014.
[13] R. V. L. Hartley, "Transmission of information", Bell System Technical Journal, vol. 7, no. 3, July 1928, pp. 535-563.
[14] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, July, October, 1948, pp. 373-423 and 623-656.
[15] N. Seshadri, V. Tarokh, and A. R. Calderbank, "Space-time codes for wireless communication: code construction," Proc. IEEE Vehicular Technology Conference, Phoenix, USA, May 4-7, 1997.
[16] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data-rate wireless communication over wideband channels," Proc. IEEE Vehicular Technology Conference, Ottawa, Canada, May 18-21, 1998.
[17] A Fehske, G. Fettweis, J. Malmodin, G. Biczk, "The global footprint of mobile communications: The ecological and economic perspective," IEEE Communications Magazine, August 2011, pp. 55-62.
[18] V. G. Douros, S. Toumpis, and G. C. Polyzos, "Power control under best response dynamics for interference mitigation in a two-tier femtocell network," Proc. International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, Paderborn, Germany, May 14-18, 2012.
[19] P. Li and Y. Zhu, "Power control of femtocell networks based on Stackelberg game," Proc. IEEE International Conference on Communications in China, Beijing, China, Aug 15-17, 2012.
[20] V. Chandrasekhar, J. G. Andrews, T. Muharemovict, S. Zukang, and A. Gatherer, "Power control in two-tier femtocell networks" Proc. IEEE Transactions on Wireless Communications, vol. 8, issue 8, 2009, pp. 4316-4328.
[21] N. Chakchouk, and B. Hamdaoui, "QoS-aware autonomous distributed power control in co-channel femtocell networks," Proc. IEEE Global Communications Conference, Anaheim, USA, Dec. 3-7, 2012.
[22] K. Illanko, A. Anpalagan, and D. Androutsos, "An Optimal and fair distributed algorithm for power allocation for radios coexisting in unlicensed spectra," Proc. IEEE International Conference on Communications, Cape Town, South Africa, May 23-27, 2010.
[23] K. Illanko, A. Anpalagan, and D. Androutsos, "Adaptive demodulation using Raptor codes: Analytical results and extension to fading channels," Proc. 11th Canadian Workshop on Information Theory, Ottawa, Canada, May 13-15, 2009.
[24] K. Illanko, A. Anpalagan, and D. Androutsos, "Cooperative communication using bitselective adaptive demodulation and Raptor codes: The Gaussian relay channel case," Proc. IEEE Vehicular Technology Conference, Barcelona, Spain, April 26-29, 2009.
[25] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," Proc. IEEE Global Communications Conference, Houston, Texas, USA, December 5-9, 2011.
[26] "Introduction to Genetic Algorithms," lancet.mit.edu/mbwall/presentations/IntoToGas/, last accessed on May 11, 2014.
[27] D. Fundenberg and J, Triole, "Game Theory," MIT Press, 1991.
[28] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, March 2004.
[29] S. J. Oh and A. C. K. Soong, "QoS-constrained information-theoretic sum capacity of reverse link CDMA systems," IEEE Transactions on Wireless Communications, vol. 5, no. 1, Jan. 2006, pp. 3-7.
[30] A. Abadpour, A. S. Alfa, and A. C. K. Soong, "Closed form solution for maximizing the sum capacity of reverse-link CDMA system with rate constraints," IEEE Transactions on Wireless Communications vol. 7, no. 4, Apr. 2008, pp. 1179-1183.
[31] L. P. Qian, Y. J. Zhang, and J. Huang, "MAPEL: achieving global optimality for a non-convex wireless power control problem," IEEE Transactions on Wireless Communications vol. 8, no. 3, Mar. 2009, pp. 1553-1563.
[32] L. P. Qian, and Y. Jun, "Monotonic optimization for non-concave power control in multiuser multicarrier network systems," Proc. IEEE International Conference on Computer Communications, Rio De Janeiro, Brazil, April 19-25, 2009.
[33] J. Dai, Z. Ye, and X. Xu, "Power allocation for maximizing the minimum rate with QoS constraints," IEEE Transactions on Vehicular Technology, vol. 58, no. 9, Nov. 2009, pp. 4989-4996.
[34] A. Gjendemsj, D. Gesbert, G. E. Yien, and S. G. Kiani, " Binary power control for sum rate maximization over multiple interfering links," IEEE Transactions on Wireless Communications, vol. 7, no. 8, Aug. 2008, pp. 3164-3173.
[35] D. Julian, M. Chiang, D. ONeill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," Proc. IEEE International Conference on Computer Communications, New York, USA, June 23-27, 2002.
[36] V. Shah, N.B. Mandayam, and D. J. Goodman, "Power control for wireless data based on utility and price," Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Boston, Massachusetts, Sept. 8-11, 1998.
[37] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pareto efficiency of pricingbased power control in wireless data networks," Proc. IEEE Wireless Communications and Networking Conference, New Orleans, USA, Sept. 21-24, 1999.
[38] D. J. Goodman and N. B. Mandayam, "Network assisted power control for wireless Data," Proc. IEEE Vehicular Technology Conference, Rhodes, Greece, May 6-9, 2001.
[39] C. U. Saraydar, N. B. Mandayam, D. J. Goodman, "Efficient power Control via pricing in wireless data networks," IEEE Transactions on Communications, vol. 50, issue 2, Feb. 2002, pp. 291-303
[40] W. Yu, and J.M. Cioffi, "Competitive equilibrium in the Gaussian interference channel," Proc. IEEE International Symposium on Information Theory, Sorrento, Italy, June 25 - 30, 2000.
[41] D. T. Ngo, L. B. Le, T. Le-Ngoc, E. Hossain, and D. I. Kim, "Distributed interference management in two-tier CDMA femtocell networks," IEEE Transactions On Wireless Communications, vol. 11, no. 3, March 2012, pp. 979-989.
[42] C. W. Sung and W. S. Wong, "A noncooperative power control game for multirate CDMA data networks" IEEE Transactions on Wireless Communications, vol 2, no. 1, Jan. 2003, pp. 186-194.
[43] A. Ghasemi, K. Faez, and M. Dehgham, "A new pricing function for power control game in wireless data networks" Proc. IEEE Conference on Vehicular Technology, Montreal, Canada, Sept. 25-28, 2006.
[44] W. Wang, T. Peng, W. Wang, "Game theoretical analysis of power control for cognitive network in licensed spectrum" Proc. First International Conference on Communications and Networking in China, Beijing, China, Oct. 25-27, 2006.
[45] W. Xia, and Z. Qi, "Power control for cognitive radio based on game theory," Proc. International Conference on Wireless Communications, Networking, and Mobile Computing, Shanghai, China, Sept. 21-25, 2007.
[46] T. M. Mitchell, "Machine Learning," WCB-McGraw-Hill, 1997.
[47] Y. Zhu, W. Wang, T. Peng, and W. Wang, "A noncooperative power control game considering utilization and fairness in cognitive radio network" Proc. IEEE MAP and EMCT International Symposium, 2007.
[48] W. Wang, Y. Cui, T. Peng, W. Wang, "Noncooperative power control game with exponential pricing," Proc. IEEE Vehicular Technology Conference, Dublin, Ireland, April 22-25, 2007.
[49] S. Subramani, T. Basar, S. Armour, D. Kaleshi, and Zhong Fan, "Noncooperative equilibrium solutions for spectrum access in distributed cognitive radio networks new frontiers in dynamic spectrum access networks," Proc. IEEE Dynamic Spectrum Access Networks Conference, Chicago, USA, Oct. 14-17, 2008.
[50] D. Wu, D. Yu, Y. Cai, "Subcarrier and power allocation in uplink OFDMA systems based on game theory" Proc. IEEE International Conference on Neural Networks and Signal Processing, Zhenjiang, China, June 8-10, 2008.
[51] T. Li and S. K. Jayaweere, "A novel primary-secondary user power control game for cognitive radios," Proc. IEEE International Symposium on Information Theory and Its Applications, Auckland, New Zealand, Dec. 7 - 10, 2008.
[52] A. Attar, M. Reza, and H. Aghvami, "Cognitive radio game for secondary spectrum access problem," IEEE Transactions on Wireless Communications, vol 8, no. 4, April 2009, pp. 2121-2131.
[53] L. Qian, X. Li, J. Attia, and Z. Gajic, "Power control for cognitive radio ad hoc networks," Proc. IEEE Workshop on Local and Metropolitan Area Networks, New York, USA, Jun 10-13, 2007.
[54] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," IEEE Journal on Selected Areas in Communications, vol. 25, no. 3, April 2007, pp. 517-528.
[55] Y. Xu, W. Chen, and Z. Chao, "Optimal power allocation for spectrum sharing in frequency-selective unlicensed bands," IEEE Communication Letters, vol. 12, no. 7, July 2008, pp. 511-513.
[56] N. Clemens and C. Rose, "Intelligent power allocation strategies in an unlicensed spectrum," Proc. First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, USA, Nov. 8-11, 2005,
[57] R. Farha, N. Abji, O. Sheikh, and A. L. Garcia, "Market based resource management for cognitive radios using machine learning," Proc. IEEE Global Telecommunications Conference, Washington, DC, USA, Nov. 26-30, 2007.
[58] F. F. Digham, "Joint power and channel allocation for cognitive radios," Proc. IEEE Wireless Communications and Networking Conference, Las Vegas, USA, March 31 April 3, 2008,
[59] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," IEEE Journal on Selected Areas in Communications, vol. 20, no. 5, June 2002, pp. 1105-1115.
[60] K. K. Leung, C. W. Sung, and K. W. Shum, "Iterative Waterfilling for Parallel Gaussian Interference Channels," Proc, IEEE International Conference in Communications, vol. 12, June 2006, pp. 5491-5494.
[61] J. Pang, G. Scutari, and C. Wang, "Distributed power allocation with rate constraints in Gaussian parallel interference channels," IEEE Transactions on Information Theory, vol. 54, no. 8, August 2008, pp. 3471-3480.
[62] C. M. Grinstead and J. L. Snell, "Introduction to Probability," Swarthmore College and Dartmouth College, American Mathematical Society, 2002.
[63] R. Gibbons, "Game Theory for Applied Economists," Princeton University Pres, 1992.
[64] M. J. Osborne and A. Rubinstein, "A Course in Game Theory," Oxford University Press, 1994.
[65] Asu Ozdaglar, MIT, Game Theory Course Notes, http://ocw.mit.edu, last accessed on March 12, 2011.
[66] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, "Algorithmic Game Theory," Cambridge University Press, 1997.
[67] R. S. Prabhu and B. Daneshrad, "An energy-efficient watter-filling algorithm for OFDM systems," Proc. IEEE International Conference on Communications, Cape Town, South Africa, May 23-27, 2010.
[68] A. Akbari, R. Hoshyar, and R. Tafazolli, "Energy-efficient resource allocation in wireless OFDMA systems," Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, September 26-29, 2010.
[69] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," Proc. IEEE Global Communications Conference, Miami, USA, December 6-10, 2010.
[70] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequencyselective channels," IEEE Transactions on Communications, vol. 58, no. 2, February 2010, pp. 545-554.
[71] G. W. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Low-complexity energy-efficient scheduling for uplink OFDMA", IEEE Transactions on Communications, vol. 60, no. 1, January 2012, pp. 112-120.
[72] F. Wu, Y. Mao, X. Huang, and S. Leng, "Low-Complexity optimal energy-efficient resource allocation in downlink OFDMA networks." Proc. Proc. IEEE International Conference on Intelligent Computing and Communications, Cluj-Napoca, Romania, August 30 - September 1, 2012.
[73] A. A. Daoud, M. Alanyali, and D. Starobinski, "Pricing strategies for spectrum lease in secondary markets." IEEE/ACM Transactions on Networking, vol. 18, no. 2, April 2010, pp. 462-475.
[74] S. K. Jayaweera, G. Vazquez-Vilar, and C. Mosquera, " Dynamic spectrum leasing: A new paradigm for spectrum sharing in cognitive radio networks," IEEE Transactions on Vehicular Technology, vol. 59, no. 5, June 2010, pp. 2328-2339.
[75] G. El-Howayek, and S. K. Jayaweera, "Distributed dynamic spectrum leasing (D-DSL) for spectrum sharing over multiple primary channels," IEEE Transactions on Wireless Communicatins, vol.10, no.1, January 2011, pp. 55-60.
[76] D. Niyato and E. Hossain, "A game-theoretic approach to competitive spectrum sharing in cognitive radio networks," Proc. IEEE Wireless Communications and Networking Conference, Hong Kong, March 11-15, 2007.
[77] S. Shakeri, A. A. Sayegh, and T. D. Todd, "Secondary wireless mesh network design using leased frequency spectra," Proc. IEEE Wireless Communications and Networking Conference, Sydney, Australia, April 18-21, 2010.
[78] K. Illanko, A. Anpalagan and D. Androutsos, "Stackelberg game on the boundary of coexistence," Proc. IEEE International Conference on Communications, Kyoto, Japan, June 5-9, 2011.
[79] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative cognitive radio networks," Proc. ACM International Symposium on Mobile Ad Hoc Networking and Computing, New Orleans, Louisiana, USA, May 18-21, 2009.
[80] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," IEEE Journal on Selected Areas in Commun., vol. 26, no. 1, January 2008, pp. 203-213.
[81] S. Boyd, "Lectures on Convex Optimization-I," Stanford University, http://academicearth.org/courses/convex-optimization-i, last accessed on June 14, 2011.
[82] S. Boyd, "Lectures on Convex Optimization-II," Stanford University, http://academicearth.org/courses/convex-optimization-ii, last accessed on June 14, 2011.
[83] M. Chiang, "Nonconvex Optimization for Communication Systems," Electrical Engineering Department, Princeton University, Princeton, New Jersey, USA. http://www.princeton.edu/ chiangm/nonconvex.pdf, last accessed on July 2, 2014.
[84] A. Ben-Tal and A. Nemirovski "Non-linear programming theory and Algorithms," Lecture Notes, The Israel Institute of Technology, Israel, http://www2.isye.gatech.edu/ nemirovs/lmco_run_prf.pdf, last accessed on July 20, 2014.
[85] K. Illanko, A. Anpalagan, E. Hossain, and D. Androutsos, "On the power allocation problem in the Gaussian interference channel with proportional rate constraints," IEEE Transactions on Wireless Communications, vol. 13, issue 2, 2014, pp. 1101-1115.
[86] S.K. Mishra et al., "Generalized Convexity and Vector Optimization, " Nonconvex Optimization and Its Applications, Springer-Verlag, Berlin, Heidelberg 2009.
[87] M. Avriel, W. E. Diewert, S. Schaible, and I. Zhang, "Generalized Concavity," SIAM Publications, 2010.
[88] M. A. Hanson, "On sufficiency of Kuhn-Tucker conditions," Journal of mathematical analysis and applications, Ser. 80, 1981, pp. 545-550.
[89] A. Ben-Israeli and B. Mond, "What is invexity," Journal of Australian mathematical society, Ser. B 28, 1986, pp. 1-9.
[90] M. A. Hanson, "Invexity and the Kuhn-Tucker Theorem," Journal of mathematical analysis and applications, Ser. 236, 1999, pp. 594-604.
[91] M. Reeder, Boston College, MT210 Linear Algebra Course Notes, Chapter 18, "A Tour of Four Dimensions," https://www2.bc.edu/ reederma/Linalg18.pdf, last accessed on January 8, 2012.
[92] Wikipedia, http://en.wikipedia.org/wiki/Genetic_algorithm, last accessed on June 10, 2014.
[93] M. Mitchell, "An Introduction to Genetic Algorithms," MIT Press, Cambridge, MA, USA, 1996.
[94] J. Holland, "Adaptation in Natural and Artificial Systems," MIT Press, Cambridge, MA, USA, 1992.
[95] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," IEEE Transactions on Wireless Communications, vol. 4, no. 6, November 2005, pp. 2726-2737.
[96] K. Illanko, K. Raahemifar, and A. Anpalagan. "Sub-Channel and power allocation for multiuser OFDM with fairness using genetic algorithm," Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, B.C., Canada, Aug. 23-26, 2009.
[97] S. F. Chien, T.O. Ting, X. S. Yang, A. K. N. Ting, and D. W. Holtby, "Optimizing energy efficiency in multi-user OFDMA systems with genetic algorithm." Proc. IEEE International Conference on Advances in Computers, Communications and Information Theory, Mysore, India, August 22-25, 2013.
[98] C. E. Shannon, "Two-way communication channels," in Proc. Fourth Berkeley Symposium on Mathematics, Statistics and Probability, vol. 1 (Univ. of Calif. Press, 1961), pp. 611-644.
[99] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," IEEE Transactions on Information Theory, vol. it-27, no. 1, 1981, pp. 49-60.
[100] I. Sason, "On achievable rate regions for the Gaussian interference channel," IEEE Transactions Information Theory, vol. 50, no. 6, Jun. 2004, pp. 1345-1356.
[101] X. Shang, G. Kramer, and B. Chen, "A new outer bound and noisy interference sumrate capacity for Gaussian interference channels," IEEE Transactions on Information Theory, vol. 55, no. 2, Feb. 2009, pp. 689-699.
[102] S. Ariyavistakul, "SIR-based power control in a CDMA system." Proc. IEEE Global Telecommunications Conference, Orlando, USA, December 6-9, 1992.
[103] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," IEEE Transactions on Vehicular Technology, vol. 41, no. l, Feb 1992, pp. 57 62.
[104] J. Zander, "Distributed co channel interference control in cellular radio systems," IEEE Transactions on Vehicular Technology, vol. 41, no. 3, Aug. 1992, pp. 305-311.
[105] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," IEEE Transactions on Vehicular Technology, vol. 42, no. 4, Nov. 1993, pp. 641-646.
[106] R. Yates, "A framework for uplink power control in cellular radio systems," IEEE Journal on Selected Areas in Communications, vo1.13 no. 7, 1995, pp. 1341-1347.
[107] N. Vucic, S. Shi, and M. Schubert, "DC programming approach for resource allocation in wireless networks," Proc. of 8th International Symposium on the Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, Avignon, France. May 31 - June 4, 2010.
[108] K. Illanko, A. Anpalagan, and D. Androutsos, "Dual methods for power allocation for radios coexisting in unlicensed spectra," Proc. IEEE Global Communications Conference, Miami, Florida, USA, December 6-10, 2010.
[109] M. Chiang, "Geometric programming for communication systems," Foundations and Trends of Communications and Information Theory, vol. 2, no. 1-2, August 2005, pp. 1-156.
[110] M. Chiang, C. W. Tang, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," IEEE Transactions on Wireless Communications, vol. 6, no. 7, July 2007, pp. 2640-2651.
[111] J. Kim and D. H. Cho, "A joint power and subchannel allocation scheme maximizing system capacity in dense femtocell downlink systems," IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Tokyo, Japan, September 13 - 16, 2009.
[112] I. C. Wong, Z. Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems," IEEE Workshop on Signal Processing Systems, Austin, Texas, USA, October 13-15, 2004.
[113] S. Sadr, A. Anpalagan, and K. Raahemifar, "A novel subcarrier allocation algorithm for multiuser OFDM system with fairness: users perspective," Proc. IEEE Vehicular Technology Conference, Baltimore, USA, Sept. 30 - Oct. 3, 2007.
[114] H. Seo and B. G. Lee, "A proportional-fair power allocation scheme for fair and efficient multiuser OFDM systems," Proc. IEEE Global Telecommunications Conference, Texas, USA, Nov. 29 - Dec 3, 2004.
[115] S. Nassar, "A Beautiful Mind: A Biography of John Forbes Nash, Jr.," Simon and Schuster, 1994.
[116] N. Lynch, "Distributed Algorithms," Morgan Kaufmann Publishers, 1996.
[117] O. Taussky, "A Recurring Theorem on Determinants," The American Mathematical Monthly, vol. 56, no. 10, Dec. 1949, pp. 672-674.
[118] R. A. Horn and C. R. Johnson, "Topics in Matrix Analysis," Cambridge University Press, 1991.
[119] Y. Haddad and D. Porrat, "Femtocell SINR performance evaluation," Proc. International Conference on Evolving Internet, Valencia, Spain, September 20-25, 2010.
[120] http://wireless.fcc.gov/licensing/index.htm?job=spectrum_leasing, last accessed on January 8, 2011.
[121] L. Duan, J. Huang, and B. Shou, "Competition with dynamic spectrum leasing," Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, Singapore, April 6-9, 2010.
[122] P. Lin, J. J. Q. Zhang, and M. Hamdi, "Dynamic spectrum sharing with multiple primary and secondary users," IEEE Transactions on Vehicular Technology, vol. 60, no. 4, May 2011, pp. 1756-1765.
[123] R. Ya-zhen, Z. Qin-yu, and C. Pei-pei, "Cooperative and non-cooperative WSPs in dynamic spectrum leasing," Proc. IEEE International Conference on Communications and Mobile Computing, Qingdao, China. April 18-20, 2011
[124] Y. Yi, J. Zhang, Q. Zhang, and T. Jiang, "Spectrum leasing to multiple cooperating secondary cellular networks," Proc. IEEE International Conference on Communications, Kyoto, Japan, June 5-9, 2011.
[125] Federal Communications Commission (FCC), "Spectrum policy task force," ET Docket no. 02-135, November 15, 2002.
[126] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," IEEE Signal Processing Magazine, vol. 24, issue 3, May 2007, pp. 79-89.
[127] J. Mitola, "Cognitive Radio: An integrated agent architecture for software defined radio," PhD dissertation, Royal Institute of Technology, 2000.
[128] S. Haykin, "Cognitive Radio: Brain-empowered wireless communications," IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, Feb 2005, pp. 201-220.
[129] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," IEEE Transactions on Communications, vol. 54, July 2005, pp. 1310-1322.
[130] M. R. Garey and D. Johnson, "Computers and intractability: A guide to the theory of NP-completeness," W.H. Freeman and Company, 1979, pp. 245-248.
[131] I. C. Wong, B. I. Evans, "Optimal downlink OFDMA resource allocation with linear complexity to maximize ergodic rates," IEEE Transactions on Wireless Communications, vol. 7, no. 3, March 2008.
[132] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," IEEE Journal on Selected Areas in Communications, vol. 21, no. 2, Feb 2003, pp. 171 - 178.
[133] W. Rhee and J. M. Cioffi, "Increasing in capacity of multiuser OFDM system using dynamic subchannel allocation," Proc. IEEE Vehicular Technology Conference, Tokyo, Japan, May 2000.
[134] I. Kim, H. L. Lee, B. Kim, and Y. H. Lee, "On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM, "Proc. IEEE Global Communications Conference,, San Antonio, Texas, USA, November 25-29, 2001.
[135] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," IEEE Transactions on Communications, vol. 54, issue 5, May 2006, pp. 922-933.
[136] D. I. Kim, L. B. Le, and E. Hossain, "Joint rate and power allocation for cognitive radios in dynamic spectrum access environment," IEEE Transactions on Wireless Communications, vol. 7, no.12, Dec. 2008, pp. 5517-5527.
[137] A. Guha, V. Ganapathy, "Power allocation schemes for cognitive radios," Proc. IEEE International Conference on Communication Systems Software and Middleware and Workshops, 2008, Bangalore, India, Jan. 6-10, 2008.
[138] K. Illanko, M. Naeem, A. Anpalagan, D. Androutsos, "Low complexity energy efficient power allocation for green cognitive radio with rate constraints," Proc. IEEE Global Communications Conference, Anaheim, USA, December 3-7, 2012.
[139] K. Illanko, M. Naeem, A. Anpalagan, and D. Androutsos, "Frequency and power allocation for energy efficient OFDMA systems with proportional rate constraints," IEEE Wireless Communication Letters vol. 3, issue 3, 2014, pp. 313-316.
[140] T. M. Cover and J. A. Thomas, "Elements of Information Theory," John Wiley, New York. 1991
[141] G. Zhang, "Subcarrier and bit allocation for real-time services in Multiuser OFDM systems," IEEE Communication Society, 2004.
[142] M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jaseemuddin, "Decode and forward relaying for energy-efficient multiuser cooperative cognitive radio network with outage constraints," IET Journal of Communications, vol. 8, issue 5, 2014, pp. 578-586.
[143] M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jesemuddin, "Optimal power allocation for green cognitive radio: fractional programming approach," IET Journal of Communications, vol. 7, issue 12, 2013, pp. 1279-1286.
[144] M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jaseemuddin, "Iterative power allocation for downlink green cognitive network," Proc. IEEE Global Communications Conference, Anaheim, USA, Dec. 3-7, 2012.
[145] M. Naeem, K. Illanko, A. Karmokar, A. Anpalagan, and M. Jaseemuddin, "Energy efficiency of cooperative cognitive radio network with outage constraints," Proc. IEEE Radio and Wireless Symposium, Newport Beach, CA, USA, Jan 19-23, 2014.
[146] M. Naeem, K. Illanko, K. A. Karmokar, A. Anpalagan, and Jaseemuddin, "Power allocation in decode and forward relaying for green cooperative cognitive radio systems," Proc. IEEE Wireless Communications and Networking Conference, Shanghai, China, April 7-10, 2013.
[147] K. Illanko, A. Anpalagan, and D. Androutsos, "Competitive pricing for spectrum subleasing for future wireless ad hoc networks," Proc. IEEE International Conference on Communications, Ottawa, Canada, June 10-15, 2012.


[^0]:    ${ }^{1}$ Strictly speaking monotonicity is sufficient. However, the existence of a monotonic transformation guarantees the existence of an increasing one.

[^1]:    ${ }^{1}$ In economics, "discount rate" means something completely different from our usage here. Our usage should not be confused with the one in economics.

[^2]:    ${ }^{1}$ Please note that the CDMA uplink model here is a spacial case of the GIC. If all the receivers in GIC are located at the same location (base station), then GIC model becomes the CDMA model

