TECHNISCHE UNIVERSITÄT DRESDEN

Resource Allocation for Multiple-Input Multiple-Output Interference Networks

Pan Cao
geboren am 16. Februar 1984

von der Fakultät Elektrotechnik und Informationstechnik der Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs
(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender: Jun.-Prof. Dr.-Ing. Peter Birkholz

Gutachter: Prof. Dr.-Ing. Eduard A. Jorswieck
Gutachter: Prof. Dr. Björn Ottersten

Tag der Einreichung: 3. November 2014
Tag der Verteidigung: 12. Januar 2015

To my parents and Lina

Acknowledgements

I am grateful to my supervisor Professor Eduard Jorswieck for giving me the opportunity to do doctoral studies in the Chair of Communications Theory at TU Dresden. You give me the freedom and encouragement to explore whatever academic challenges I found interesting. This helps me to improve the ability to work independently. Meanwhile, your office door was always open to me when I needed advices and supports. From the insightful discussions with you, I have learned a broad knowledge of signal processing, optimization and game theory as well as the methodology, which not only help me to resolve the scientific problems over the years but also provide me a skeleton key to open the doors to future.

I would like to give special thanks to Dr. Shuying Shi for introducing me to the area of MIMO interference channels and teaching me some very useful optimization tools when I just joined the group. I am also thankful to Dr. Rami Mochaourab for our collaborations in the Pareto boundary computation in MIMO interference channels, and Dr. Alessio Zappone, Dr. Zuleita Ho, Zhijiat Chong, Ye Zhong for our collaborations in the energy efficiency in two-hop relay-aided channels.

I wish to thank Professor Eduard Jorswieck, Dr. Alessio Zappone and Dr. Pin-Hsun Lin for their proofreading and comments of my thesis. This helps me to improve the presentation of the thesis. I thank Jing Lv for sharing the room with me and giving the support in the process of writing thesis. I thank Sybille Siegel for the administrative work and Holger Hösel for the IT work. I also thank all my previous and current colleagues and friends in Dresden for those happy moments we spent together.

I wish to thank Professor Björn Ottersten for spending time to serve as the reviewer for this thesis, and also Professor Peter Birkholz and Professor Frank Fitzek for participating in the evaluation committee.

Finally, I want to thank my parents and my sister for their love and support in whatever aspect. I want to thank my wife Lina for the love and patience shown, sacrifices made and the wonderful suggestions.

Pan Cao

Dresden, October 2014

Abstract

To meet the exponentially increasing traffic data driven by the rapidly growing mobile subscriptions, both industry and academia are exploring the potential of a new generation (5G) of wireless technologies. An important 5G goal is to achieve high data rate. Small cells with spectrum sharing and multiple-input multiple-output (MIMO) techniques are one of the most promising 5G technologies, since it enables to increase the aggregate data rate by improving the spectral efficiency, nodes density and transmission bandwidth, respectively. However, the increased interference in the densified networks will in return limit the achievable rate performance if not properly managed.

The considered setup can be modeled as MIMO interference networks, which can be classified into the K-user MIMO interference channel (IC) and the K-cell MIMO interfering broadcast channel/multiple access channel (MIMO-IBC/IMAC) according to the number of mobile stations (MSs) simultaneously served by each base station (BS). The thesis considers two physical layer (PHY) resource allocation problems that deal with the interference for both models: 1) Pareto boundary computation for the achievable rate region in a K-user single-stream MIMO IC and 2) grouping-based interference alignment (GIA) with optimized IA-Cell assignment in a MIMO- IMAC under limited feedback. In each problem, the thesis seeks to provide a deeper understanding of the system and novel mathematical results, along with supporting numerical examples. Some of the main contributions can be summarized as follows.

It is an open problem to compute the Pareto boundary of the achievable rate region for a K-user single-stream MIMO IC. The K-user single-stream MIMO IC models multiple transmitter-receiver pairs which operate over the same spectrum simultaneously. Each transmitter and each receiver is equipped with multiple antennas, and a single desired data stream is communicated in each transmitter-receiver link. The individual achievable rates of the K users form a K-dimensional achievable rate region. To find efficient operating points in the achievable rate region, the Pareto boundary computation problem, which can be formulated as a multi-objective optimization problem, needs to be solved. The thesis transforms the multi-objective optimization problem to two single-objective optimization problems – single constraint rate maximization prob-

lem and alternating rate profile optimization problem, based on the formulations of the ϵ-constraint optimization and the weighted Chebyshev optimization, respectively. The thesis proposes two alternating optimization algorithms to solve both single-objective optimization problems. The convergence of both algorithms is guaranteed. Also, a heuristic initialization scheme is provided for each algorithm to achieve a high-quality solution. By varying the weights in each single-objective optimization problem, numerical results show that both algorithms provide an inner bound very close to the Pareto boundary. Furthermore, the thesis also computes some key points exactly on the Pareto boundary in closed-form.

A framework for interference alignment (IA) under limited feedback is proposed for a MIMO-IMAC. The MIMO-IMAC well matches the uplink scenario in cellular system, where multiple cells share their spectrum and operate simultaneously. In each cell, a BS receives the desired signals from multiple MSs within its own cell and each BS and each MS is equipped with multi-antenna. By allowing the inter-cell coordination, the thesis develops a distributed IA framework under limited feedback from three aspects: the GIA, the IA-Cell assignment and dynamic feedback bit allocation (DBA), respectively. Firstly, the thesis provides a complete study along with some new improvements of the GIA, which enables to compute the exact IA precoders in closed-form, based on local channel state information at the receiver (CSIR). Secondly, the concept of IA-Cell assignment is introduced and its effect on the achievable rate and degrees of freedom (DoF) performance is analyzed. Two distributed matching approaches and one centralized assignment approach are proposed to find a good IA-Cell assignment in three scenrios with different backhaul overhead. Thirdly, under limited feedback, the thesis derives an upper bound of the residual interference to noise ratio (RINR), formulates and solves a corresponding DBA problem. Finally, numerical results show that the proposed GIA with optimized IA-Cell assignment and the DBA greatly outperforms the traditional GIA algorithm.

Contents

List of Figures

List of Tables

Nomenclatur

List of Notations

$\mathbb{C}^{N\times M}$	The set of complex-valued $N \times M$ matrices.
$\mathbb{R}^{N\times M}$	The set of real-valued $N \times M$ matrices.
$\mathbb{R}_+^{N\times M}$	The set of non-negative real-valued $N \times M$ matrices.
$\mathbb{N}_+$	The set of non-negative integers.
$\mathrm{diag}\{\boldsymbol{x}\}$	The diagonal matrix with the elements of $\boldsymbol{x}$ at the diagonal.
$\boldsymbol{X}^H$	The conjugate transpose of $\boldsymbol{X}$.
$\boldsymbol{X}^{-1}$	The inverse of a square matrix $\boldsymbol{X}$.
$\boldsymbol{X}^\dagger$	The Moore-Penrose pseudo inverse of $\boldsymbol{X}$.
$\boldsymbol{\Pi}_{\boldsymbol{X}}$	The orthogonal projection matrix onto the column space of $\boldsymbol{X}$ (i.e., $\boldsymbol{\Pi}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^H\boldsymbol{X})^{-1}\boldsymbol{X}^H$).
$\boldsymbol{\Pi}_{\boldsymbol{X}}^\perp$	Projection matrix onto the orthogonal complement of the column space of $\boldsymbol{X}$ (i.e., $\boldsymbol{\Pi}_{\boldsymbol{X}}^\perp = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^H\boldsymbol{X})^{-1}\boldsymbol{X}^H$).
$\Re\{x\}$	Real part of a scalar x.
$\|x\|$	Absolute value of a scalar x.
$\angle x$	Phase of a complex-valued scalar x.
$[x]^+$	$\max\{0, x\}$ where x is a real scalar.
$\mathrm{Tr}\{\boldsymbol{X}\}$	The trace of a square matrix $\boldsymbol{X}$.
$\mathrm{rank}\{\boldsymbol{X}\}$	The rank of a matrix $\boldsymbol{X}$.
$\mathrm{Span}\{\boldsymbol{X}\}$	Orthonormal basis for the column space of $\boldsymbol{X}$.
$\mathrm{null}\{\boldsymbol{X}\}$	Orthonormal basis for the null space to the columns of $\boldsymbol{X}$.
$\mathcal{CN}(\boldsymbol{x}, \boldsymbol{R})$	The circular symmetric complex Gaussian counterpart.
$\mathbb{E}\{\boldsymbol{x}\}$	The mathematical expectation (i.e., mean) of a vector x.
$\|\|\boldsymbol{x}\|\|$	The 2-norm of vector $\boldsymbol{x}$ ($\|\|\boldsymbol{x}\|\| = (\sum_i \|x_i\|^2)^{\frac{1}{2}}$).
$\overrightarrow{\boldsymbol{x}}$	The normalized direction of $\boldsymbol{x}$ ($\overrightarrow{\boldsymbol{x}} = \frac{\boldsymbol{x}}{\|\|\boldsymbol{x}\|\|}$).
$\mathcal{S}\backslash\{k\}$	The remaning set when member k is removed.
$(\lambda_i(\boldsymbol{X}), \boldsymbol{u}_i(\boldsymbol{X}))$	The i-th largest eigenvalue of the matrix $\boldsymbol{X}$ and $\boldsymbol{u}_i(\boldsymbol{X})$ is the eigenvector associated with $\lambda_i(\boldsymbol{X})$.
$\boldsymbol{X}^\perp$	The space spanned by those eigenvectors whose corresponding eigenvalues are zeros.

$\mathcal{O}(\cdot)$	Big O notation where $f(x) = \mathcal{O}(g(x))$ means that it exist $c \in \mathbb{R}_+$ and $\boldsymbol{x}_0 \in \mathbb{R}$ such that $\lvert f(x)\rvert \leq c\lvert g(x)\rvert$ for $x \geq x_0$.
$\boldsymbol{X} \succeq \boldsymbol{Y}$	Means that $\boldsymbol{X} - \boldsymbol{Y}$ is positive semi-definite.
$\boldsymbol{I}_N$	The $N \times N$ identity matrix.
$\mathbf{0}_{N\times M}$	The $N \times M$ matrix of only zeros.

List of Symbols

2G	Second Generation of Cellular Wireless Technology
3G	Third Generation of Cellular Wireless Technology
3GPP	3rd Generation Partnership Project
4G	Fourth Generation of Cellular Wireless Technology
5G	Fifth Generation of Cellular Wireless Technology
AF	Amplify-and-Forward
APCC	Approximate Pareto Complex Characterzation
APRC	Approximate Pareto Real Characterzation
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BS	Base-Station
CDF	Cumulative Distribution Function
CDMA	Code-Division Multiple-Access
CoMP	Coordinated Multi-Point
CQI	Channel Quality Information
CSI	Channel State Information
CSIR	Channel State Information at Receivers
D2D	Device-to-Device
DBA	Dynamic Feedback Bit Allocation
DCT	Discrete Cosine Transform
DoF	Degrees of Freedom
DPC	Dirty Paper Coding
EDGE	Enhanced Data Rates for GSM Evolution
EVD	Eigenvalue Decomposition
FCA	The Forward Chaining Algorithm
FDD	Frequency-Division Duplex
FDMA	Frequency-Division Multiple-Access
GIA	Grouping-based Interference Alignment
GSM	Global System for Mobile Communications
HSPA	High Speed Packet Access
IA	Interference Alignment
IBC	Interfering Broadcast Channel
IC	Interference Channel
ICI	Inter-Cell Interference

IMAC	Interfering Multiple Access Channel
IUI	Inter-User Interference
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
ITU-R	International Telecommunication Union Radiocommunication Sector
MAC	Multiple Access Channel
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MMSE	Minimum Mean Square Error
MRT	Maximum Ratio Transmission
MS	Mobile-Station
NE	Nash Equilibrium
NP-hard	Non-Deterministic Polynomial-Time hard
OFDMA	Orthogonal Frequency-Division Multiple-Access
P2P	Point-to-Point
Q2	Second Quarter
QCQP	Quadratically Constrained Quadratic Program
QoE	Quality of Experience
QoS	Quality of Service
RINR	Residual-Interference-to-Noise Ratio
RSINR	Received-Signal-to-Interference-and-Noise Ratio
RSINR	Received-Signal-to-Noise Ratio
SDMA	Space-Division Multiple-Access
SIC	Successive Interference Cancellation
SIMO	Single-Input Multiple-Output
SISO	Single-Input Single-Output
SINR	Signal-to-Interference-and-Noise Ratio
SNR	Signal-to-Noise Ratio
SOCP	Second Order Cone Programming
SQ	Subspace Quantization
SU	Single-User
SVD	Singularvalue Decomposition
TD-SCDMA	Time Division Synchronous Code Division Multiple Access
TDMA	Time-Division Multiple-Access
ZF	Zero Forcing Transmission

Chapter 1

Introduction

1.1 Background & Motivation

Figure Total mobile subscriptions reached up to around 6.8 billion in the second quarter (Q2) of 2014 (including the Q2 2014) with an average increasing rate of around 7 percent per year. The number of mobile broadband subscriptions grew even faster, reaching 2.4 billion with a year-on-year increase of around 35 percent. The subscriptions with high data usage become more and more popular in the world. For instance, around 65 percent (more precisely, around 300 million) of all mobile phones sold in the Q2 2014 were smartphones that enable the advanced features and services, such as video streaming, web browsing, high speed internet access etc. [Eri14a, Eri14b].

Due to the development of wireless access technology and electronic technology, roughly every ten years new mobile phone technology and infrastructure experience a significant change in the fundamental nature of the service, transmission technology, data rate and frequency bands. These transitions following requirements stated by the International Telecommunication Union Radiocommunication Sector (ITU-R) are referred to as *generations*[1]. The current used mobile subscriptions can be mainly cat-

[1] Global System for Mobile Communications (GSM) commercially launched in 1991 and its enhanced version, Enhanced Data Rates for GSM Evolution (EDGE), belong to the second generation (2G) technologies (EDGE is also referred to as a pre third generation (3G) technology). Starting from around 2001, the code division multiple access (CDMA) family (e.g., CDMA2000), wideband code division multiple access (WCDMA), time division synchronous code division multiple access (TD-SCDMA) and some invariants belong to the family of 3G mobile technology. High Speed Packet Access (HSPA) is an enhanced 3G mobile communications technique dubbed 3.5G or 3G+. Long Term Evolution (LTE) (as specified in the 3rd Generation Partnership Project (3GPP) Release 8 and Release 9 document series), launched in 2008, is marketed as a fourth generation (4G) wireless service, while in fact it is a 3.9G or pre-4G technology because its enabled downlink data rate of up to 300 Mbit/s does not satisfy the technical requirements (1 Gbit/s) that the 3GPP consortium has adopted for its new standard generation. LTE-advanced (LTE-A) in 3GPP Release10 standardized in 2011 is officially announced as a 4G technology.

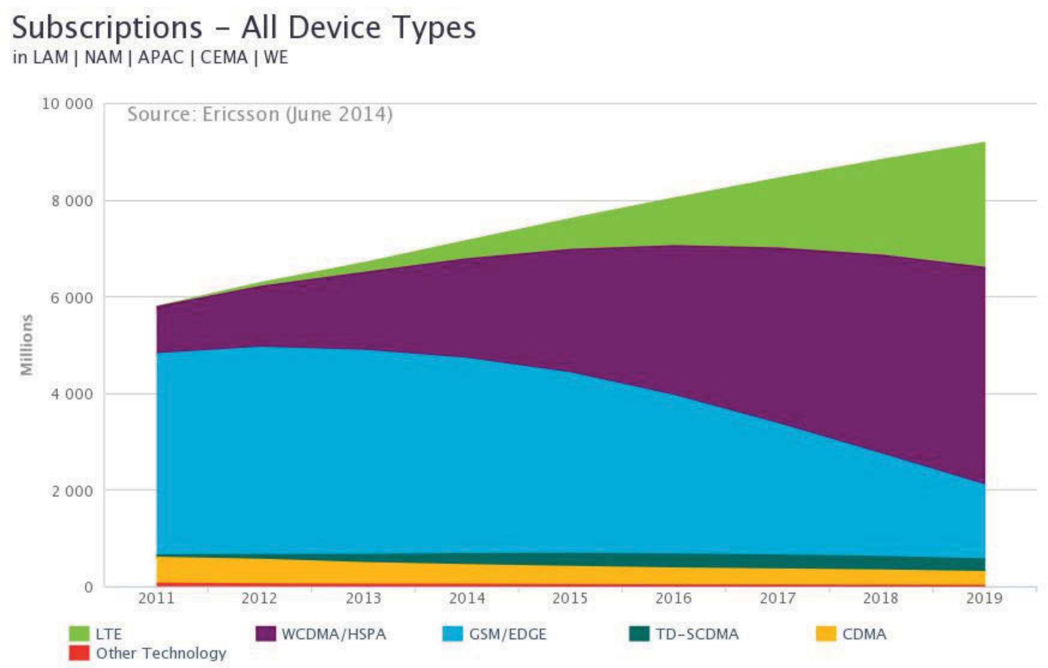


Figure 1.1: Previous, current and forecast number of mobile subscriptions worldwide partitioned by wireless access technology (Source: Ericsson traffic exploration tool, June 2014 [Eri14c]).

egorized into the 2G-only, 3G-enabled and 4G-enabled subscriptions. For instance, the 2G-only cellular phones coexist with 3G/4G-enabled smartphones. In order to have an overview of the development of the mobile subscriptions, Figure 1.1 is provided to show the previous, current and forecast number of the mobile subscriptions during the period of 2011–2019, where the mobile subscriptions are described by different technologies, i.e., GSM/EDGE (2G), WCDMA/HSPA/TD-SCDMA (3G) and LTE (4G). Figure 1.1 shows a rapid increase of global mobile subscriptions, growing to more than 9 billion in 2019. In particular, the number of global 2G subscriptions has decreased since 2012 because of the rapid migration to more advanced technologies, while it is predicted that the number of 3G and 4G subscriptions will become more and more dominant than the 2G subscriptions in the coming years. For example, the 3G and 4G mobile subscriptions will take account for around *five-sixth* of all mobile subscriptions in 2019.

Driven by the rising popularity of smart devices – smartphones, tablets and laptops, cellular networks are currently experiencing an exponential growth in data traffic along with a continuous increase in the average data volume per subscription so as to achieve the uniform quality of experience (QoE) in the whole network. The rapid growth of the data traffic from 2011 to 2019 is shown in Figure 1.2. It implies that the data traffic caused by the advanced services (e.g., video streaming) of 3G and 4G-enabled devices becomes much more heavy than 2G services (file sharing and audio). For example, it is reported in [Eri14a] that smartphone users are consuming more data than ever before – an average of about 650 MB per month in 2013. In addition, the total mobile traffic generated by mobile phones exceeded that from mobile personal computers, tablets and mobile routers for the first time in 2013. The bandwidth-heavy and rate-hungry services of the rapidly rising popularity of smart devices demand the tenfold increase of the mobile traffic during the coming years – reaching a global monthly total of almost 12 exabytes[2] by the end of 2019.

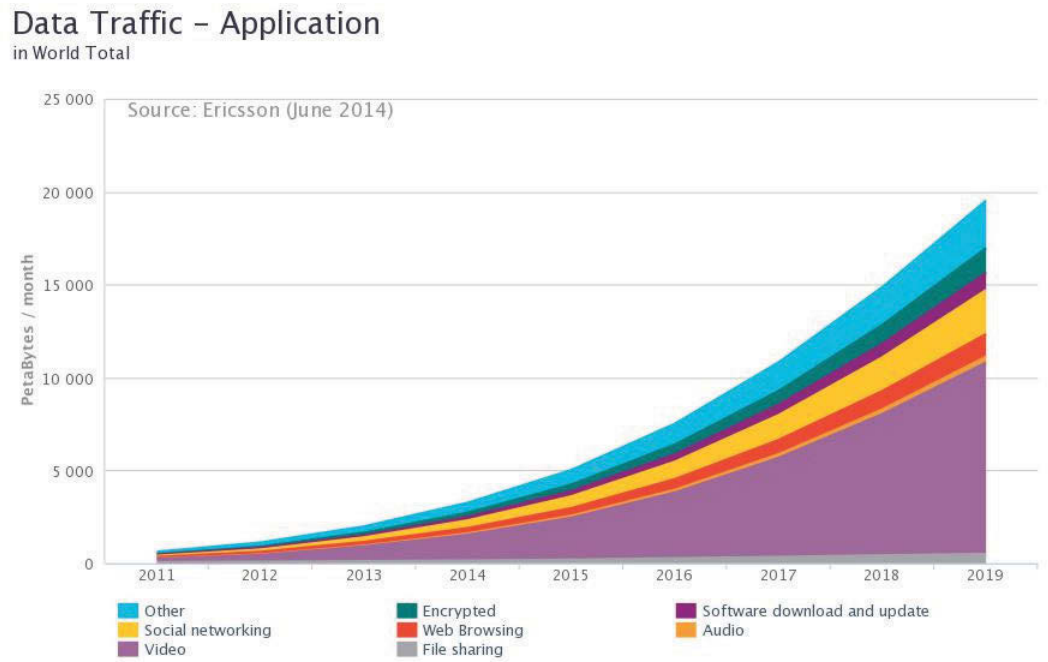


Figure 1.2: Previous, current and forecast globe mobile data traffic by different services (Source: Ericsson traffic exploration tool, June 2014 [Eri14c]; note: 1 PetaByte = 10^{15} Bytes).

To meet the exponentially increasing traffic data, both industry and academia are exploring the next generation (so-called the fifth generation, 5G) technology. The standards may be introduced approximately in the early 2020s. For instance, some mobile communications companies are now working to meet a future need for an even 1000× data increase in traffic capacity for mobile access networks, e.g., Qualcomm proposed

[2] 1 Exabytes = 10^{18} Bytes.

the concept, version and potential solutions of "The 1000× data challenge" [Qua]. Although it cannot be predicted when the 1000x traffic growth will happen, the wireless industry is currently experiencing a tremendous growth in mobile data traffic. The engineering challenges in future 5G systems are mainly on high data rate, low latency and low energy/cost [ABC+14]. This thesis focuses on the issue – *high data rate.* The data rate is usually measured in *aggregate data rate* [ABC+14]. Aggregate data rate refers to the total amount of data of all nodes in the network, characterized in units of bits/s/area, which can be defined as

$$\frac{bits/s}{Km^2} = \frac{bits/s}{Hz \cdot node} \times \frac{node}{Km^2} \times Hz. \tag{1.1}$$

To increase the aggregate data rate, it is desired to simultaneously improve all the three terms on the right hand side. In the following, the effect of these three terms on the aggregate data rate is analyzed, seperately[3].

- **Node-spectral efficiency** ($\frac{bits/s}{Hz \cdot node}$) refers to the information rate that can be transmitted per Hz and per second per node [Ver02]. Given a specific system with a fixed number of nodes, node-spectral efficiency can be directly evaluated by system-spectral efficiency ($bits/s/Hz$). Thus, we equivalently use *spectral efficiency* ($bits/s/Hz$) for convenience when we consider a system with fixed number of nodes.

  The *multiple-input and multiple-output (MIMO) technique* applies multiple antennas at both the transmitters and receivers. The channels from multiple transmit antennas to multiple receive antennas experience different multi-path propagation, which creates the so-called *diversity gain.* The best path can be selected at both the transmitter-side and receiver-side by *beamforming* – to form a direction with the strongest channel gain by combining all the possible directions in a smart way (i.e., spreading the transmit/receive power over antennas to achieve an array gain) [BO01]. This not only increases the link reliability but also improves the spectral efficiency. Furthermore, the MIMO technique also enables to transmit and receive multiple data streams independently and simultaneously. The maximum number of data streams enabled in a system is referred to as *multiplexing gain* or *degrees of freedom (DoF).* Theoretically, the sum spatial DoF for a specific system is defined as

$$d_{\sum} \triangleq \lim_{\rho \to \infty} \frac{C_{\sum}(\rho)}{\log(\rho)} \tag{1.2}$$

[3] Each of the three terms is analyzed when the others are fixed.

where $C_{\sum}(\rho)$ denotes the sum capacity at signal-to-noise ratio (SNR) ρ, which is the theoretical limit of achievable rate of a system. For a high SNR, $C_{\sum}(\rho)$ is linearly proportional to DoF, i.e.,

$$C_{\sum}(\rho) = \log(\rho) \times d_{\sum}. \tag{1.3}$$

It implies that the achievable rate can be also improved by increasing DoF. Therefore, MIMO technique should be a key technique to increase the achievable rate by enabling both high DoF and smart precoding/decoding design.

- **Nodes density ($\frac{node}{Km^2}$)** refers to the number of served nodes in a certain area. It can be increased by *network densification* – densifying the macro network and/or adding small cells (e.g., femtocells, picocells and microcells). These different ways of network densification are compatible, e.g., in heterogeneous networks with two tiers or even multiple tiers. Thus, more nodes in an area can be simultaneously served by more base-stations (BSs). In addition, the nodes density can be also increased by employing MIMO techniques. For example, a BS with multiple antennas, enabling high DoF, can simultaneously serve multiple nodes (so-called *multi-user transmission*).

- **Transmission bandwidth (Hz)** is a fundamental factor to allow data transmission. During the past three decades, cellular systems have gained the ability to operate on frequencies in the range of $0.7 - 5$ GHz and over increasing system bandwidths. This increase has been driven by an inexorable demand for spectrum that today still far exceeds that available for new exclusive licenses. The scarcity of spectrum has created a need for greater flexibility. In the following, we provide several potential ways to increase the transmission bandwidth.
  - *Spectrum sharing:* multiple cells/operators share their original allocated non-overlapped spectrum bandwidths at the same time, by which each node could occupy the whole spectrum bandwidth [JBF+14]. Thus, spectrum sharing achieves a much higher frequency reuse factor than the classical frequency division multiple access (FDMA) where the spectrum is divided into serval non-overlapped frequency bands and each node occupies a single frequency band.
  - *Carrier aggregation:* carrier aggregation enables multiple continuous or non-continuous narrow spectrum bandwidths (e.g., licensed spectrum bandwidths and unlicensed spectrum bandwidths) to form a wider transmission band-

width, e.g., the transmission bandwidth can be increased from up to 20 MHz in LTE to up to 100 MHz in LTE-A by carrier aggregation [3GP12].

- *Refarming the 2G spectrum:* as shown in Figure 1.2, the data traffic demand (mainly on voice data) for 2G networks has already dropped significantly below that for 3G/4G and surely for the future 5G networks, it is meaningful to allocate part of 2G spectrum to the 4G/5G networks. For instance, China Unicom will clear out part of 2G network in some cities in China to ensure more frequency bandwidth for 4G by upgrading the 2G BSs into 4G [C1114].
- *Employing unlicensed spectrum:* due to the limited available licensed spectrum, more and more technologies move towards to use the unlicensed spectrum to increase the spectrum bandwidth, such as Wi-Fi, Ultra Wideband, Millimeter-Wave, software-defined radio and cognitive radio etc.

Based on the above analysis, small cells with advanced MIMO and spectrum sharing techniques are considered one of the most promising version for the future 5G because of its advantage and also tractable upgrade from 4G networks [ABC+14,Qua]. This protocol enables to improve all the spectral efficiency, nodes density and spectrum bandwidth, simultaneously, thereby greatly improving the aggregate data rate. However, more BSs and nodes sharing the same spectrum also suffer from the increased interference, which in return will limit the spectral efficiency and DoF performance if not properly managed. Therefore, *interference management* – how to deal with the received interference to suppress or avoid its limit to the system performance, is becoming an important issue in the design of the future wireless communications networks [HRTA14].

1.2 MIMO Interference Networks

A MIMO interference network represents a communication network where multiple transmitters simultaneously communicate to their intended receivers through a common communication channel and each transmitter and receiver is equipped with multiple antennas. Consequently, each receiver not only receives the *desired signal* from its intended transmitter but also overhears other transmitters. The signal from the undesired transmitters corrupts the desired signal and thus is called *interference.* The multi-cell interfering MIMO protocol as mentioned above can be modeled by a MIMO interference network.

One type of MIMO interference networks consisting of K transmitter-receiver pairs, i.e., the number of transmitters is the same as that of receivers and each transmitter

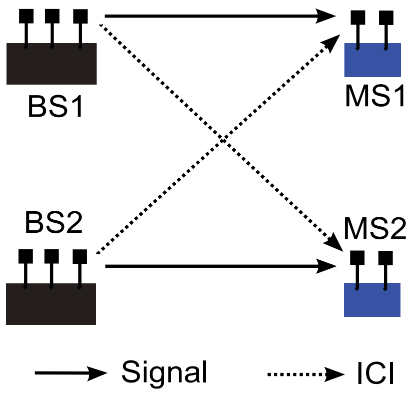


Figure 1.3: A two-user MIMO IC example: A two-cell interfering downlink scenario.

serves a single receiver, is named as *K-user MIMO interference channels (IC)* (i.e., a user refers to as a transmitter-receiver pair). It well matches a MIMO cellular system where K neighboring cells share their spectrum and each BS serves one or multiple mobile stations (MSs)[4] within its own cell, but one MS is located near the cell edge (so-called *cell-edged MS*) suffers from strong interference from other BSs, while other MSs located far away from the cell edge, e.g., near the BS, usually receive very weak interference from other BSs because of the significant path loss and fading effect. Another possible scenario shown in [LZC12] is that K neighboring cells share their spectrum but the multiple users within each cell are separated for transmission in frequency via orthogonal frequency-division multiple-access (OFDMA) or in time via time-division multiple-access (TDMA), and then the active links in different cells transmitting at the same frequency tone and in the same time slot will interfere with each other. In addition, the multiple cognitive radio links or multiple device-to-device (D2D) links operating over the same spectrum can be also formulated as the MIMO IC. For the cellular system, the MIMO IC is able to model both the downlink scenario (a BS serves as a transmitter) and uplink scenario (a BS serves as a receiver). The K-user MIMO IC only contains the interference from other cells (so-called *inter-cell interference (ICI)*). In Figure 1.3, a two-user MIMO IC example is shown to model a two-cell downlink interference network.

In a more general MIMO interference networks, the number of transmitters and receivers could be different. This general MIMO interference network model degrades to a special model, the MIMO IC model, when the number of transmitters is equal to the receivers. The interference network with more receivers than transmitters matches the

[4]Mobile stations refer to those user terminals in wireless communications, such as mobile phones, laptops etc. These devices are usually movable compared to the BSs with fixed location.

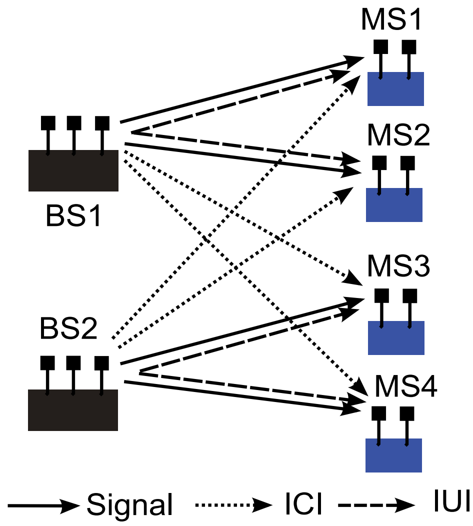


Figure 1.4: A general MIMO interference network example: A two-cell MIMO-IBC.

downlink scenario where each BS *simultaneously* transmits independent signals to multiple cell-edged MSs within its own cell. In the downlink, each cell can be modeled as a broadcast channel (BC), and thus this setup is named *multi-cell MIMO interfering BC (MIMO-IBC)*. On the other hand, the interference network is also able to model the *uplink* scenario where each BS simultaneously receives multiple signals from multiple cell-edged MSs (each cell is modeled as a multiple access channel (MAC)), which is named as *multi-cell MIMO interfering (MIMO-IMAC)*. In the multi-cell MIMO-IBC/MIMO-IMAC, there exist both ICI and *inter-user interference* (*IUI*, the interference caused by the BC/MAC within each cell). In Figure 1.4, a two-cell MIMO-IBC example is shown to illustrate a general interference network with two transmitters and four receivers.

1.2.1 Resource Allocation

In the wireless communication environment, it is well-known that some resources are usually limited, such as spectrum, transmit power, spatial resource, etc. Therefore, under spectrum sharing, it is still desired to further allocate other limited resource in an efficient way so as to improve the spectral efficiency in a MIMO interference network by the smart design of the precoding/decoding schemes.

Take the transmit power for instance, the capacity of point-to-point (P2P) channel, single-cell BC/MAC always increases with transmit power [GJJV03]. However, it is not true in interference networks because larger transmit power of one transmitter results in stronger interference to other receivers that will degrade the spectral efficiency of other receivers. Thus, transmit power allocation is a very important issue in interference networks. In addition, the transmit/receive spatial resource (related to the number of equipped antennas) is another important resource, e.g., if one transmitter sends more data streams, the dimension of the interference subspace to other receivers become higher such that the "clean/free" subspace available for their desired signals' transmission is reduced. Therefore, an interference network has multiple objectives related to the utilities of the multiple cells/MSs/BSs, and these multiple objectives usually cannot be treated independently because of a conflict between links such that improvements in one objective lead to deterioration of other objectives.

Therefore, the performance of an interference network significantly depends on two factors –*multiple users' behavior* and *interference*. There exists an inter-action between the users' behavior and interference-level, e.g., the interference level depends on the multiple users' behavior and in return the different interference levels may influence the decisions of the users. The users' behavior can be coarsely divided into *competition*, *cooperation*[5] and the trade-offs between them (e.g., coalition or partial cooperation). If each user selfishly wants to maximize its own spectral efficiency by using more resource, there will exist a utility-conflicting between these users because their objectives are usually coupled and the resource is budget-limited. This resource allocation problem can be formulated as a *competitive game* of multiple selfish users by using game theory. A solution concept is the Nash equilibrium (NE) that each user knows the equilibrium strategies of the other users and no user has anything to gain by changing only their own strategies [OR94]. Instead, if users cooperate with each other for resource allocation, it forms a *cooperative game*. The cooperative resource allocation problem enables to achieve a Pareto-optimal solution if it can be solved optimally. In fact, the NE is usually inefficient compared with Pareto-optimal solution [Coh98, LJ08] because of the selfish and competitive nature of the users in the competitive game. Therefore, cooperation is preferred in system design from the perspective of either the system or an user [LJ08]. The thesis focuses on developing coordinated algorithms based on the assumption of coordinated behavior of BSs. In the following, we briefly introduce severval types of interference management techniques to deal with interference in the coordinated interference networks.

- *Nonlinear interference processing:* the "dirty paper coding" (DPC) in the down-

[5] Some literatures, e.g., [Moc12], distinguish cooperation from coordination by implementation, where the coordination and cooperation specify the implementation in a distributed way and in a centralized way, respectively. In this thesis, coordination and cooperation are inter-changable because our designed algorithms can be applied to both cases.

link and successive interference cancellation (SIC) in the uplink are two well-known *nonlinear* interference processing techniques and enable to achieve the capacity of the degraded and MIMO BC and MAC [GJJV03], respectively. However, implementation of DPC requires significant additional complexity at both transmitter and receiver, and the problem of finding practical dirty paper codes that approach the capacity limit is still unsolved [LJ06]. Also, the implementation of SIC has a high complexity demand of the receiver.

- *Linear interference processing:* a simple way of dealing with the interference is to use linear processing techniques, i.e., by properly designing the linear precoding/decoding schemes. According to the extent of interference suppression, the linear interference processing techniques can be classified as follows.
  - *Coordinated Multi-Point (CoMP)* already standardized in long term evolution advanced (LTE-A) [V2.10] aims at turning ICI into an advantage by letting BSs share their data and perform joint precoding/decoding. In this case, multiple BSs form a *virtual* BS (like the distributed antennas system) and thus an interference network becomes a BC in the downlink or a MAC in the uplink. However, this requires the *exact* exchange of global CSI as well as (possibly) user data via high data rate backbone connections, which might be a problem when the BSs belong to different operators or have conflicting utilities [IDM+11].
  - *Interference avoidance* is a direct way to completely remove the interference. One commonly used approach is to impose a constraint that all interference terms are zero. Such a zero-forcing (ZF) approach is a good solution with simple implementation [SSH04]. The ZF precoding/decoding is widely used at BSs in the downlink/uplink if BSs have sufficient antennas to null out all the interferences.
  - *Interference alignment (IA)* in spatial domain is to align multiple interferences by properly designing the linear precoding or decoding at users, leaving more free spatial space for the useful signal [Jaf11, CJ08]. IA cannot directly mitigate the interference but enables to suppress the interference subspace into a lower-dimensional subspace. Thus, the condition on the number of BSs antennas to directly perform ZF precoding/decoding can be relaxed by employing the IA decoding/precoding at MSs. Therefore, IA is helpful to achieve the maximum DoF in MIMO interference networks.
  - *Interference control* is defined as the interference processing when interfer-

ence is not completely mitigated but controllable. One reason for adapting interference control instead of interference avoidance is that a BS has no sufficient antennas to perform the ZF scheme. Another reason is that an ill-conditioned channel matrix when inverted, will require a large normalization factor and will dramatically reduce the SNR at receivers. Therefore, allowing a limited amount of interference at each receiver facilitates a larger set of potential solutions, which can provide higher achievable rate for a given transmit power level, e.g., the linear minimum mean square error (MMSE)-type precoding/decoding controls the interference by jointly designing all transmit and receive strategies [TV05].

- *Interference employment:* some concepts to handle interference as a useful resource are introduced in [ZKM+14]: 1) exploit interference at the modulation level which leads to simple multiuser downlink precoding and provides significant energy savings; 2) use radio frequency radiation for energy harvesting and handles interference as a source of green energy [PC14]; 3) use interference as an efficient means to jam potential eavesdroppers. These new versions usually have high requirement of complexity and re-design of the receiver circuit and thus still lack practical applications. Furthermore, these ideas mainly consider the energy issue, which is not the focus of this thesis.

In the thesis, we focus on jointly designing the linear transceiver to deal with interference in MIMO interference networks. However, the proposed algorithms are not restricted to linear processing and can be also extended to the nonlinear processing case, where the downlink DPC and uplink SIC are also adopted.

1.2.2 Summary of System Assumptions

To be concise, the thesis is based on the following assumptions:

- BSs are connected – 1) each BS connects to all others through error-free backhaul links (distributed implementation is enabled) or 2) each BS connects to a central controller through error-free backhaul links (centralized implementation is enabled, e.g., Cloud RAN).

- The main processing task of transceiver design is done at the BSs or central controller, because BSs and the central controller are generally fixed, backhaul-connected and larger in size than MSs and thus they are capable to carry out a larger computational overhead.

- BSs belong to different operators and thus the ideal CoMP can not be adopted but the inter-cell coordination without global CSI and user data exchange among BSs is necessary.
- Linear transceiver is designed by treating interference as Gaussian noise, which has low complexity rand high robustness in implementations.
- Each channel matrix is *quasi-static block fading*, i.e., each channel matrix stays constant for a set consecutive discrete time instants (*coherence time*);
- Perfect local channel state information (CSI) at receivers (CSIR) is available.
- There are error-free feedback links between BSs and MSs.

1.2.3 System operations

We introduce two types of system operations – Frequency Division Duplex (FDD) and Time Division Duplex (TDD), widely used in practical mobile communication systems [CLW+06]. The main difference between them is that the downlink channel and uplink channel of a link are different in FDD systems due to sufficient separation between the uplink frequency band and downlink frequency band, while the same channel over the same frequency band switches between the downlink and uplink transmission. Based on the block fading assumption, the *closed-loop transmission*[6] becomes feasible in system operations if the system operation delay is shorter than the channel coherent time.

In the following, the basic downlink and uplink operation of FDD systems are described, respectively.

1. *FDD Downlink:*
   - 1a. **Training** & **Channel Estimation:** Each BS sends an orthogonal pilot sequence based on which each MS perfectly estimates its local downlink channels.
   - 1b. **CSI Feedback:** Each active MS quantizes and feeds back the estimated downlink channels to BSs[7].

[6] It refers to transmit strategies designed based on both the instantaneous CSI and designed transmit/receive strategies. The phases of CSI estimation, CSI feedback and transmit/receive strategies design and the feedback of transmit strategies form a closed-loop in operation.

[7] There may exist an additional feedback phase of the channel quality information (CQI) for the MSs selection. Since the procedure of MS-selection or MS-BS association is out of the scope of the thesis, we do not highlight the feedback of the CQI.

1c. **Transceiver Design at BS-side:** Based on the collection of the quantized downlink channels, BSs design the precoding and decoding strategies.

1d. **Decoders Feedback**[8]: Each BS reports the computed decoding strategies to the MSs within its own cells.

1e. **Data Transmission:** Each BS transmits the data by the designed precoders to its served MSs and each MS decodes the data based on the received decoder. The data transmission is performed with the fixed precoders and decoders until the end of the coherence time.

2. *FDD Uplink:*

2a. **Training** & **Channel Estimation:** Each active MS sends an orthogonal pilot sequence based on which each BS perfectly estimates the local uplink channels.

2b. **Transceiver Design at BS-side:** Based on the estimated CSIR, BSs design the precoding and decoding strategies.

2c. **Precoders Feedback:** Each BS reports the computed precoding strategies to the MSs within its own cells.

2d. **Data Transmission:** Each MS transmits the data by the received precoder and the BS decodes the data based on the designed decoder until the end of the channel coherence time.

The pair of frequency bands for downlink and uplink transmission are sufficiently separated by a defined frequency offset (so-called guard band). Thus, FDD systems enables parallel downlink and uplink transmissions through different channels (full-duplex communications links), thereby with low latency. In general, FDD requires the availability of a pair of frequency bands, and it can be used in licensed and license-exempt bands. Thus, FDD is widely used in cellular telephone systems.

Next, the basic operation of TDD systems is described in four steps:

3. *TDD Downlink(Uplink)*

3a. **Training** & **Channel Estimation:** Each active MS sends an orthogonal pilot sequence to BSs based on which the local uplink channels are perfectly

[8] The "decoders feedback" phase, different from the "CSI feedback" through the *backward channels*, is through the *forward channels* in the downlink scenario. The feedback of precoders/decoders is also named as *dedicated training* in some references, e.g., [CJKR10]. Please notice that the CSI feedback occupies the uplink resource, while the decoder/precoder feedback takes the downlink resource. Therefore, we consider both as feedback phases for convenience.

estimated by BSs (in the TDD uplink scenario, the downlink channels can be also obtained at BSs based on the available uplink channels by using the *channel reciprocity.*).

3b. **Transceiver Design at BS-side:** Based on the estimated uplink channels (or downlink channels in the TDD uplink scenario), BSs design the precoding and decoding strategies.

3c. **Decoders (Precoders) Feedback:** Each BS reports the computed decoders (or precoders in the TDD uplink scenario) to the MSs within its own cells.

3d. **Data Transmission:** Each BS (or MS in the TDD uplink scenario) transmits the data by the precoders to its served MSs (or BS in the TDD uplink scenario). Each MS (or BS in the TDD uplink scenario) decodes the data based on the decoders. The data transmission is performed with the fixed precoders and decoders until the end of the channel coherence time.

TDD uses a single frequency band for both downlink and uplink transmission but alternatively assigns time slots to downlink and uplink transmission, and thus has the possibility to exploit channel reciprocity. Moreover, TDD allows for the flexible allocation of throughput between the downlink and uplink transmission, making it well suited to applications with asymmetric traffic requirements, such as video surveillance, broadcast and Internet browsing. However, the switch from transmit to receive incurs a delay that causes TDD systems to have greater inherent latency than FDD systems [EXA].

Remark 1. *In TDD systems, it is usually assumed that the downlink channel is the conjugate-transpose of the uplink channel based on the ideal channel reciprocity. In this case, it is sufficient to only design the downlink precoders/decoders, since the same performance goal in the uplink can be also achieved by the same downlink strategies due to the* downlink-uplink duality*. However, the reciprocity principle usually is not fulfilled at the digital baseband interfaces in realistic system because of the different transceiver circuitries in the transmit and receive path [KJGK10]. Then, the effective downlink and uplink channels are different and the different downlink and uplink transceiver need to be designed separately.*

The above description of the FDD and TDD system operations is summarized in Figure 1.5. Without loss of generality, the system operations illustrated for the P2P channel in Figure 1.5 can be extended to multi-cell MIMO interference networks.

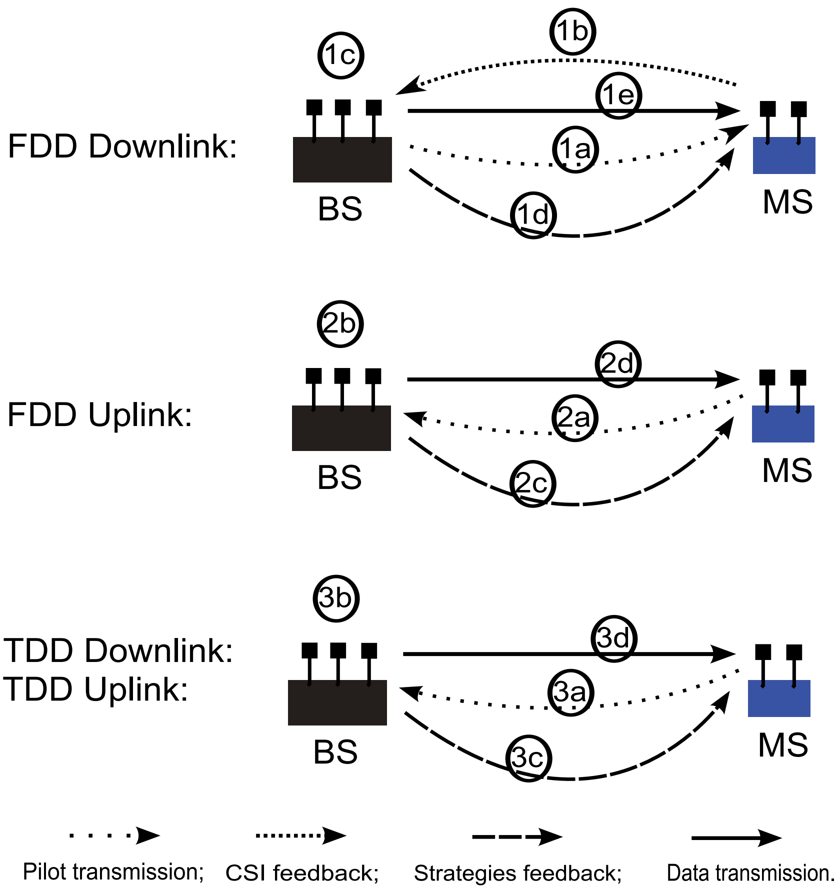


Figure 1.5: FDD vs. TDD in closed-loop system operation: A P2P channel example (the indexes denotes the steps in the system operation described before.)

Remark 2. *For the closed-loop system operation described in this section, the transceiver design is done at BS-side, which can be implemented in a distributed way. Without loss of generality, if there exists a central controller, the optimization will be done at the central controller with two more steps – 1) each BS reports its own local CSIR to the central controller (CSI collection) and 2) the central controller broadcasts the optimized transceiver to the BSs (strategies announcement).*

1.3 Problem Formulation: Pareto Boundary Computation

Chapter 2 considers a multi-user single-stream MIMO IC. The achievable rates of the multiple users form an achievable rate region. To find efficient operating points, the Pareto boundary problem needs to be solved. Therefore, we focus on computing the Pareto boundary of the achievable rate region.

1.3.1 Related Work on MIMO IC & Pareto Boundary

In information theory, a multi-user MIMO IC is characterized by its *capacity region*, i.e., the set of largest rates that can be simultaneously achieved by the users in the system while making the error probability arbitrary small (the theoretical limits of spectral efficiency). However, the capacity of a system usually cannot be achieved by the simple linear transceiver design. Furthermore, the capacity region of a MIMO IC is still unknown. Instead, an achievable rate region or inner bound of the capacity region is of great interest for both theoretical study and practical system design based on two assumptions: 1) the class of encoding strategies are constrained to use random Gaussian codebooks and 2) the decoders are restricted to treat the interference as Gaussian noise. Based on these two assumptions, we desire to find the *complete achievable rate region* by linear transceiver design.

In operation, a multi-user MIMO IC apparently has multiple objectives [BJ13], i.e., each user has a utility (e.g., achievable rate in the thesis). A *multi-objective problem* usually has a *feasible-solution set* consisting of infinite possible solutions, both optimal and non-optimal solutions. Each tuple of the values of these multiple objectives forms a joint operating point in a multi-dimensional space (achievable rate region in the thesis), which in return implies a resource allocation example for the multiple users in operation.

In optimization, a multi-objective optimization problem usually admits infinite number of *non-diminated* solutions, which form the outermost boundary of achievable rate region, so-called *Pareto boundary*[9] [Zad63]. A non-dominated solution on the Pareto boundary is considered to be Pareto-optimal in the sense that no other solution can improve some objectives without reducing other objective(s). Generally, it is hard to find the Pareto boundary efficiently, but it is significant to study it in order to determine optimal system operations based on Pareto-optimal rate tuples and their associated strategies.

In principle, the Pareto boundary can be obtained by the grid search of all the variables. However, the dimension of the variables in the MIMO IC (complex precoders and decoders) is very high such that it is too inefficient to be meaningful. How to design linear transceiver schemes to efficiently achieve the Pareto boundary has attracted intensive research for several decades. A brief, comprehensive, yet non-exhaustive review of the related works is given as follows.

The Pareto boundary can be achieved by optimally solving a multi-objective optimization problem with respect to (w.r.t.) multiple variables. Without loss of generality,

[9] When referring to Pareto boundary in the thesis, we mean the Pareto boundary of achievable rate region unless otherwise specified.

we consider the following K-objective optimization problem

$$\max_{\boldsymbol{x} \in \mathcal{X}} \ \{f_1(\boldsymbol{x}), \cdots, f_K(\boldsymbol{x})\} \tag{1.4}$$

where $\boldsymbol{x} \triangleq \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_K\}$ and $f_k(\boldsymbol{x}), \forall k \in \mathcal{K}$ with $\mathcal{K} \triangleq \{1, \cdots, K\}$ are defined as the variable set of K users (e.g., precoding/decoding strategies) and the achievable rate function of k-th user, respectively. The variable set $\boldsymbol{x}$ should be in the constraint set $\mathcal{X}$ (e.g., transmit/receive power constraints), i.e., $\boldsymbol{x} \in \mathcal{X}$.

- **Characterization:** the characterization approach is to parameterize the Pareto-optimal strategies by using less variables or lower-dimensional variables, i.e.,

$$\boldsymbol{x}^\star = g(\boldsymbol{y}), \quad \boldsymbol{y} \in \mathcal{Y} \tag{1.5}$$

  where $\boldsymbol{x}^\star$ denotes the Pareto-optimal solutions to the multi-objective optimization problem (1.4), which is parameterized by the new variable set $\boldsymbol{y} \in \mathcal{Y}$ as an expression form (function) of $g(\cdot)$, and $\boldsymbol{y} \in \mathcal{Y}$ is the parameter set (variable) with *lower dimension* than the original variable set $\boldsymbol{x}$. The characterization cannot directly determine the Pareto boundary but provides a *necessary condition* for the Pareto-optimality. Therefore, further optimization w.r.t. the parameters in $\boldsymbol{y}$ is needed. Usually, an efficient characterization enables to express the Pareto-optimal strategies by only a few parameters, and thus it is possible to find the Pareto boundary by a low-dimensional grid search of the reduced variable set.

  For the two-user multiple-input single-output (MISO) IC, [JLD08] proposed a characterization for Pareto-optimal transmit beamforming vectors, i.e., linear combinations of ZF and maximum-ratio transmission (MRT) beamforming vectors with two $[0, 1]$-parameters. By this characterization, the variables to determine the Pareto boundary are greatly reduced from the original two complex vectors (i.e., two transmit beamforming vectors) to two real parameters in the range of $[0, 1]$. The two real parameters in the characterization [JLD08] are further reduced to only a single parameter characterization in [LKL11], [MJ11b]. For the K-user MISO IC, the characterization of Pareto-optimal transmit beamforming vectors in [JLD08] requires $K(K-1)$ complex-valued parameters, which are reduced to $K(K-1)$ real-valued parameters each between 0 and a value depending on the channels in [ZC10] and each between 0 and 1 in [MJ11a], [BZGO10], respectively, and further improved by using $2K-2$ [0, 1]-parameters in [BBO12]. In [PS13], a parameterization for Pareto-optimal beam structure in a K-user MIMO IC is given by the product manifold of a Stiefel manifold and a subset of a hyperplane. However, there are still many parameters after reduction, i.e., $2(\sum_{\ell=1}^{K} M_\ell)M_k - M_k^2 + M_k - 1$ for the characterization of the Pareto-optimal beam of the k-th transmitter with M_k transmit antennas.

- **Scalarization:** instead of providing a necessary condition for the Pareto-optimal strategies by characterization, scalarization aims to transfer the original multi-objective optimization problem to a single objective optimization problem by using multiple constant scalars to combine or restrict the original objectives. Then, a Pareto boundary point can be computed directly by solving a single objective optimization problem. Some popular goal formulations are summarized as follows.
  - *ϵ-constraint optimization problem [Mie98]* aims at maximizing the objective of one user while the objectives of other users are restricted, i.e.,

$$\begin{cases} \max\limits_{\boldsymbol{x}\in\mathcal{X}} & f_k(\boldsymbol{x}) \\ \text{s.t.} & f_\ell(\boldsymbol{x}) \geq \epsilon_\ell, \quad \forall \ell \in \mathcal{K}\backslash\{k\}, \end{cases} \tag{1.6}$$

where $\{\epsilon_\ell\}_{\mathcal{K}\backslash\{k\}}$ are *feasible* constant scalars. The constraints are used to guarantee that the single objective $f_k(\boldsymbol{x})$ always increases along the same direction (determined by $\{\epsilon_\ell\}_{\mathcal{K}\backslash\{k\}}$) eventually reaching the Pareto boundary if all the constraints are active. This idea is used to compute the Pareto boundary in a two-user MISO IC [KL10].

  - *Weighted Chebyshev optimization problem [BDMS08]* aims at finding an intersection point between the Pareto boundary and a ray from the origin, i.e.,

$$\begin{cases} \max\limits_{\boldsymbol{x}\in\mathcal{X},t} & t \\ \text{s.t.} & f_k(\boldsymbol{x}) \geq \alpha_k t, \quad \forall k \in \mathcal{K}. \end{cases} \tag{1.7}$$

The scalars $\{\alpha_k\}_{\mathcal{K}}$ are chosen such that $\alpha_k \geq 0, \forall k \in \mathcal{K}$ and $\sum_{k=1}^{K} \alpha_k = 1$. Each set of $\{\alpha_k\}_{\mathcal{K}}$ corresponds to an increasing direction of the ray and determines a Pareto boundary point. In particular, Problem (1.6) becomes the well-known *max-min optimization problem* or called *max-fairness optimization problem* when $\alpha_1 = \cdots = \alpha_K$.

Given $\{\alpha_k\}_{\mathcal{K}}$, Problem (1.7) can be optimized by solving a sequence of feasibility subproblems with different t which is updated by a bisection search. If each feasibility subproblem with a fixed t can be solved optimally, the optimal solution to Problem (1.7) can be obtained with a numerical error that is introduced in the bisection search of t but could be arbitrarily small (controllable error). This scheme is renamed as *rate profile optimization* and applied to the multi-user MISO IC [ZC10] and [QZLC11].

– *Weighted sum optimization problem [GHD82]* aims at finding a Pareto boundary point corresponding to maximum weighted sum of objectives, i.e.,

$$\max_{\boldsymbol{x}\in\mathcal{X}} \quad \sum_{k=1}^{K} \alpha_k f_k(\boldsymbol{x}), \tag{1.8}$$

where $\{\alpha_k\}_{\mathcal{K}}$ satisfy $\alpha_k \geq 0, \forall k \in \mathcal{K}$ and $\sum_{k=1}^{K} \alpha_k = 1$. The formulated single objective is the weighted arithmetic mean of the original objectives.

From the system point of view, sum rate performance is an important metric in system design. Generally, the weighted sum rate maximization problem for multi-user IC is non-convex. For the multi-user SISO IC, the MAPEL algorithm proposed in [QZH09] transforms the weighted sum rate maximization into a generalized linear fractional programming problem, which can be solved optimally. In [LZC12], the authors jointly utilized the monotonic optimization and rate profile techniques to solve the weighted sum rate maximization optimally at the cost of computation load in multi-user SISO/MISO/SIMO IC. However, it is NP-hard to obtain a global optimal solution of the weighted sum rate maximization for a multi-user MIMO IC (e.g., in [CACC08]). Some literature focus on finding the suboptimal sum-rate maximum point, such as for the two-user MIMO IC based on pricing in [SSB$^+$09], for the multi-user MIMO IC based on approximation of sum rate in [CHHT12], for the single-stream MIMO IC based on balancing the egoistic and the altruistic behavior in [HG10], and for the multi-user MIMO IC based on interference alignment in [PH11]. In order to find a good suboptimal solution, a weighted minimum mean square error (WMMSE) approach to maximize the sum rate performance by alternative optimization was proposed in [CACC08] for the MIMO-BC. Based on the idea of the WMMSE, [SRLH11] proposes an algorithm to achieve a stationary solution to the original sum-rate maximization problem of the MIMO-IBC.

However, it is well-known that the weighted sum maximization method has two major drawbacks [DD97]: 1) if the Pareto boundary is not convex, there does not exist any weight corresponding to the points on the non-convex part. Increasing the number of steps of the weighting factor does not resolve this problem; 2) even if the Pareto boundary is convex, an even spread of weights does not produce an even spread of points on the Pareto boundary. Therefore, weighted sum maximization generally is not a promising method to describe *the complete Pareto boundary*, especially the non-convex boundary.

- *Weighted product optimization problem [MA04]* aims at finding a Pareto boundary point corresponding to maximum weighted product of objectives, i.e.,

$$\max_{\boldsymbol{x}\in\mathcal{X}} \quad \prod_{k=1}^{K}(f_k(\boldsymbol{x}))^{\alpha_k}, \tag{1.9}$$

  where $\{\alpha_k\}_{\mathcal{K}}$ satisfy $\alpha_k \geq 0, \forall k \in \mathcal{K}$ and $\sum_{k=1}^{K}\alpha_k = 1$. The single objective represents the weighted arithmetic mean of the original multiple objectives. In [BN10], (1.9) is used to compute the fairness of the utilities of multiple interfering users. Also, this product formulation is usually applied to compute the Nash bargaining, e.g., [CVWT12] for the MIMO IC where $\alpha_1 = \cdots = \alpha_K$.

- *Weighted min optimization problem [Mie98]* aims at finding a Pareto boundary point corresponding to maximum of minimum of weighted objectives, i.e.,

$$\max_{\boldsymbol{x}\in\mathcal{X}} \quad \min_{k\in\mathcal{K}} \quad \frac{f_k(\boldsymbol{x})}{\alpha_k}. \tag{1.10}$$

  where $\{\alpha_k\}_{\mathcal{K}}$ satisfy $\alpha_k \geq 0, \forall k \in \mathcal{K}$ and $\sum_{k=1}^{K}\alpha_k = 1$. In principle, Problem (1.10) is equivalent to Problem (1.7) and they achieve the same optimal solutions. Therefore, Problem (1.10) is usually solved by introducing an auxiliary variable t (i.e., optimizing Problem (1.7)).

Different final operating points are determined by the different choices of goal functions, e.g., the max-sum-utility point by (1.8) and the max-fairness point by (1.7) or (1.10). If a single objective optimization problem can be solved optimally for fixed weights, different Pareto boundary points can be achieved by adjusting the optimizing direction in Problem (1.6) and the weights in other problems.

For performance comparison, the operating points attained by different scalarizations are illustrated in Figure 1.6. The ϵ-constraint optimization and weight Chebyshev optimization can achieve any intersection point between the increasing ray and the Pareto boundary. For the weighted sum optimization and the weighted product optimization, the attainable operating points are those points achieved by the level curves tangent to the Pareto boundary, i.e., $\sum_{k=1}^{K} f_k(\boldsymbol{x}) = f^{\star}_{sum}$ (line when $K = 2$) and $\prod_{k=1}^{K} f_k(\boldsymbol{x}) = f^{\star}_{prod}$ (hyperbolic curve when $K = 2$) where $f^{\star}_{sum}$ and $f^{\star}_{prod}$ are the Pareto-optimal sum of the objectives and the product of the objectives, respectively.

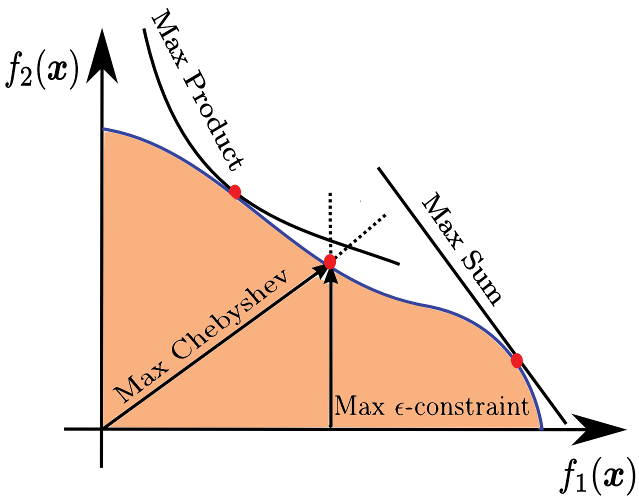


Figure 1.6: Illustration of the Pareto-optimal operating points achieved by different scalarization methods.

Remark 3. *In principle, the weighted sum formulation in Problem (1.8) cannot achieve the non-convex boundary and the weighted product formulation in Problem (1.9) may not achieve the non-convex boundary*[10]*. The other formulations in Problem (1.6), Problem (1.7) and Problem (1.10) enable to compute any point on the Pareto boundary.*

- *Game Theoretic Approaches:* game theory as a useful tool has been widely applied to resource allocation in multi-user IC by studying the competitive or cooperative behavior of the users, because it provides a systematic set of solution concepts to resolve the conflict problems between the coupled links. As we mentioned before, competitive algorithm may not converge in general or may converge to an NE, e.g., for the MIMO IC in [YB03, SPB08]. Since the best achievable performance represents the set of Pareto-optimal trade-offs among these users' conflicting objectives, the NE is *often* not Pareto-optimal [Coh98]. Therefore, to achieve the Pareto-optimal trade-offs, cooperation between users is usually required to improve their joint outcome [LJ08].

  A direct improvement from NE to Nash bargaining (NB) by cooperation for the MIMO IC has been studied in [CVWT12], where the interference-plus-noise covariance matrix of each user is assumed to approach an identity matrix and the

[10]The fact depends on whether the boundary can be touched by the level curve/surface of $\prod_{k=1}^{K} f_k(\boldsymbol{x}) = f^{\star}_{prod}$, which is influenced by the degree of non-convexity of both the boundary and level curves.

rate region approximately becomes convex. A main branch of cooperative algorithms is the interference-pricing based method, where each user updates its strategy to maximize its own utility minus the interference cost determined by the interference prices, which reflects the marginal change in utility per unit interference power. This distributed interference-pricing algorithm has been used to solve (weighted) sum-rate maximization problems for the multi-user SISO [HBH06] and MISO IC [SBH09b], multi-user single-stream MIMO IC [SBH09a], two-user MIMO IC [SSB+09]. A different pricing scheme is to balance the egoistic and altruistic strategies with different weights (i.e., prices), e.g., for the two-user MISO IC [MJ11b, JL08] and for the multi-user single-stream MIMO IC [HG10]. In [HG10], a suboptimal algorithm is provided to maximize the sum rate without convergence analysis. Most distributed cooperative algorithms for the MIMO IC focus on maximizing (weighted) sum-utility, e.g., distributed pricing based algorithms [SSB+09, SBH09a].

In principle, most approaches of characterization, scalarization can be considered as cooperative or partial cooperative algorithms, although they are not described from a viewpoint of game theory.

From the above overview of related works, some excellent references provide several approaches to achieve the Pareto boundary for the SISO/MISO/SIMO IC but there are very little results on the MIMO IC. However, the importance of MIMO technique in the future communications motivates us to investigate the Pareto boundary of a multi-user single-stream MIMO IC, although it is extremely difficult and well-known as an open problem.

For the multi-user MIMO IC, the achievable rate depends on both transmit and receive strategies, which results in a harder coupled and complicated expression than the MISO IC. Thus, it is not straightforward to extend the *implicit* or *explicit* schemes achieving the complete Pareto boundary for the MISO IC to the MIMO IC. In order to try to fill this gap, we propose two approaches to compute the complete Pareto boundary in Chapter 2. The main contributions are described as follows.

1.3.2 Contributions of the Thesis

In this section, we consider the Pareto boundary computation problem for a multi-user single-stream MIMO IC. Motivated by the overview of multi-objective optimization approaches, we first transfer the original multi-objective optimization problem into two single-objective optimization – single constraint rate maximization problem and alternating rate profile optimization problem, based on the ideas of ϵ-constraint optimization in (1.6) and weighted Chebyshev optimization in (1.7), respectively. Then, we propose two heuristic algorithms to solve both single-objective optimization problems, thereby achieving a close-to-optimal inner boundary to the strict Pareto boundary.

- We analyze the difficulty of computing the Pareto boundary for the multi-user single-stream MIMO IC – the hard-coupling of the transmit and receive beamforming vectors involved in the rate expression. Especially when the optimal linear decoding scheme, i.e., minimum mean square error (MMSE) receive beamforming, is adopted, the coupling becomes more serious regarding the transmit beamforming vectors. First, we consider a two-user single-stream MIMO IC example and derive the following new results [CSJ12, CJS13]:
  - For the two-user single-stream MIMO IC, we propose an equivalent form of the SINR expression based on the Hermitian angle, which shows the SINR is a linear combination of the single-user received-signal-to-noise ratio (RSNR) and the received-signal-to-interference-plus-noise ratio (RSINR) with two weights between 0 and 1.
  - We prove that the strict Pareto-optimal transmit power allocation policy is full power allocation at both the transmitters. That is, full power transmission is a necessary condition for Pareto-optimality.
  - We exactly compute the non-strict Pareto boundary and two ending points of strict Pareto boundary in closed-form. In addition, we also compute some ZF operating points in the rate region when the mutual interference is completely canceled.
  - We formulate a *single constraint rate maximization* problem based on the idea of the *ϵ-constraint formulation* in (1.6) to compute the strict Pareto boundary by maximizing one rate while the other rate is fixed. This problem is a non-convex optimization problem because both the objective and constraint contain the hard coupling problem. In order to make it tractable, we develop an alternating optimization algorithm to alternatively optimize the two transmit beamforming vectors until the algorithm converges [Ber99]. In each iteration, each subproblem can be solved optimally by combining semidefinite programing (SDP) and matrix rank-1 decomposition. The convergence of the algorithm is guaranteed but the global optimal solution is not guaranteed. Since the alternating optimization algorithm depends on the initializations, numerical results show the convergent points with our proposed

initialization are very close to the Pareto boundary.

- We extend the proposed single constraint rate maximization algorithm for two-user single-stream MIMO IC in [CSJ12] to the general multi-user scenario [CJS13].

- We also provide a *weighted Chebyshev optimization-based single-objective optimization* as shown in (1.7) and call it *alternating rate profile optimization* [MCJ13, MCJ14]. In the optimization, MMSE receive beamforming is not plugged in the SINR expression directly. We split the optimization problem w.r.t. both transmit beamforming vectors and receive beamforming vectors into two subproblems and each subproblem is w.r.t. only transmit beamforming vectors or receive beamforming vectors. That is, an optimization problem for the multi-user MIMO IC is transferred to one optimization problem for the multi-user MISO IC and the other one for the multi-user SIMO IC. Each subproblem can be solved optimally, and the proposed alternating optimization algorithm always converges to a stationary point of the original problem.

These novel results presented in Chapter 2 (Pareto Boundary Computation for the MIMO IC) have been published in:

- R. Mochaourab, P. Cao and E. A. Jorswieck, "Alternating rate profile optimization in single stream MIMO interference channels", *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 221–224, Feb., 2014.

- P. Cao, E. A. Jorswieck and S. Shi, "Pareto boundary of the rate region for single-stream MIMO interference channels: Linear transceiver design", *IEEE Trans. on Signal Process.*, vol. 61, no. 20, pp. 4907–4922, Oct., 2013.

- R. Mochaourab, P. Cao and E. A. Jorswieck, "Alternating rate profile optimization in single-stream MIMO interference channels", *Proc. of IEEE ICASSP*, Vancouver, BC, Canada, May, 2013.

- P. Cao, S. Shi and E. A. Jorswieck, Efficient computation of the Pareto boundary for the two-user single-stream MIMO interference channel", *Proc. of IEEE SPAWC*, Cesme, Turkey, Jun., 2012.

1.4 Problem Formulation: Interference Alignment under Limited Feedback

Interference alignment (IA) is a promising technique to efficiently mitigate interference and to enhance the capacity of a wireless communication network. In Chapter 3, we provide a framework of IA with optimized IA-Cell assignment for a multi-cell MIMO-IMAC under limited feedback.

1.4.1 Related Work on IA & Limited Feedback

The MIMO-IMAC well matches the multi-cell multi-user uplink scenario. The multiple cells share their spectrum to form a *coordinated cluster*. Each BS serves multiple MSs within its own cell and each node is equipped with multiple antennas. The uplink signal is corrupted by both ICI and IUI. In order to perform make efficient signal detection, *linear transceiver* with simple implementations is preferred to eliminate the complete interference. This was addressed in [KPSL10] by applying a coordinated ZF scheme to remove both IUI and ICI. However, ZF alone fails if a BS does not have sufficient antennas or if DoF maximization is the goal.

With this respect, a well-established technique called IA is developed [Jaf11, CJ08]. The concept of IA is to suppress the received interference by symbol extension for the time varying channel or by the precoding/decoding in spatial domain for the Gaussian MIMO channel. It has been shown that IA enables to achieve the optimal DoF of time varying K-user SISO IC through symbol extension [CJ08], deriving the "cake-cutting" result (*everyone gets half a cake*), i.e., each user achieves $\frac{1}{2}$ DoF. In the MIMO Gaussian IC, IA is not applied in the time or frequency domain but in the spatial domain by applying precoding/decoding to align the interference subspace into a lower dimensional subspace, which can reduce the required number of BS antennas to facilitate ZF reception. In [GJ10], both inner and outer bounds of the sum DoF of the K-user MIMO Gaussian IC with M antennas per transmitter and N antennas per receiver are provided, and the bounds are tight when $\frac{\max\{M,N\}}{\min\{M,N\}}$ is an integer. The optimization of both DoF and sum-rate in the K-user MIMO IC is considered in [SPLL10, SGHP10, GCJ11] by designing the linear IA precoders and decoders. Generally, it is difficult to obtain the linear IA transceiver in closed-form and thus iterative algorithms based on global CSI are usually required, except for the special case of square and invertible channel matrices in [SPLL10]. Take the K-user MIMO IC for example, an approximate IA transceiver is found by alternatively minimizing the sum interference [PH11] with the requirement of global CSI. In [GCJ11], an iterative IA scheme which exploits channel reciprocity to limit the amount of required CSI is proposed, but it only works in TDD operation systems. More recently, IA has been applied to MIMO cellular networks. In [NR10], a multi-cell MIMO downlink channel is studied and a distributed IA algorithm is proposed to minimize the interference to non-intended users. Also, [SHT11] develops an IA technique for a downlink cellular system with CSI-exchange and feedback within each cell. In [YGJK10, KLCH11], conditions for the feasibility of IA and DoF for MIMO cellular networks are investigated. However, this iterative IA algorithm comes at the expense of complexity and also of a heavy backhaul overhead (CSI/strategies-exchange in iterations).

To reduce the complexity and CSI requirement, the concept of *grouping-based* IA (GIA) is proposed for a two-cell single-stream MIMO-IBC in [SLL+11]. The idea is to let each cell align its interference to another cell, by which less antennas are required to implement ZF reception. Moreover, GIA enables to compute the optimal IA transceiver in closed-form based on the local CSI. This GIA is extended to a multi-cell MIMO-IBC in [TL13], where both the feasible condition on the GIA and a low complexity IA decoder design are studied. In [TL13], a hybrid scheme, using DPC and IA to avoid IUI and to suppress the ICI, respectively, is developed. Based on the GIA mechanism, an iterative user-selection algorithm by measuring the subspace distance is developed in [GC14].

The implementation of IA requires a closed-loop transmission. The feedback is needed in either the downlink or the uplink scenario. As illustrated in Figure 1.5, for instance, the feedback of both the CSI and precoders are needed to be considered in downlink scenario. However, most works on the limited feedback did not consider the feedback of IA decoder. Since the feedback links are usually capacity-limited in realistic systems, codebook-based feedback is widely used and already defined in modern wireless standards, e.g., in LTE [DPS11], to reduce the feedback overhead. The idea is to map a channel matrix/vector or a precoder/decoder to an index of the closest codeword in a predefined codebook known at both transmitter and receiver. The feedback of an index takes only a finite number of feedback bits, while a performance loss is inevitable because of the quantization distortion. Thus, it becomes an important issue how to control/reduce the performance loss under limited feedback [LHL+08]. A comprehensive overview of techniques and approaches for limited feedback communications in wireless networks is given in [LHL+08]. For a MIMO BC with the ZF precoder, the performance loss in limited feedback is studied in [Jin06, YJG07] and also with block diagonalization in [RJ08, SR13]. In [KMLL12], a new quantization scheme for the IA in multi-user MIMO IC is studied. The quantization distortion is reduced by designing an additional receive filter which minimizes the chordal distance metric. For a MIMO IC with heterogeneous path loss and spatial correlations, [RRL13] develops a spatial codebook design and performs dynamic quantization via feedback bit allocation. In [BT09, KV10, RG13], the feedback bits scaling law to maintain the maximum DoF for IA on general MIMO interference networks is investigated.

1.4.2 Contributions of the Thesis

Motivated by this background, we focus on applying the GIA with low complexity to the multi-cell MIMO-IMAC under limited feedback, answering the following fundamental questions [CJ14b, CZJ14].

- *How to design the optimal linear GIA transceiver with low complexity?* We further develop previous related works (e.g. [SLL+11, TL13]), providing a low-complexity *restriction-relaxation* approach to compute the optimal linear GIA transceivers which not only nulls out both ICI and IUI but also maximizes the rate performance. Moreover, we show that the restriction-relaxation process is tight such that the optimal GIA-based transceiver is obtained in closed-form.

- *How to determine a good IA-Cell assignment?* By the GIA, each cell chooses to align its interference to another cell. However, this choice clearly impacts the rate performance. Optimizing the selection of the cell to/from which a given cell provides/receives the aligned interference, is a problem which was not considered in previous works. We refer to this problem as *IA-Cell assignment* and provide three IA-Cell assignment algorithms: a centralized one, which yields global optimality but requires high complexity and overhead, and two distributed ones, which yield a stable or almost stable assignment by the theory of stable matching with limited complexity and backhaul overhead.

- *How to efficiently feed back the GIA precoders to the MSs?* In the uplink MIMO cellular scenario, the GIA precoders need to be fed back to the MSs. Let us reconsider the prior works on the IA for the MIMO IC under limited feedback. For instance, vector quantization (VQ) is employed to quantize the pre-vectorized version of a MIMO channel matrix in [RRL13], but the VQ yields a larger quantization distortion compared with channel subspace quantization (SQ) for the same feedback overhead. In [RG13], SQ is used to quantize the channel subspace. However, the receivers need to be designed with different decoding power. In addition, most previous works did not consider the feedback of IA precoders/decoders. With this respect, the contributions on limited feedback are summarized as follows.

   – We employ Grassmannian subspace quantization, developing a novel quantized subspace characterization which allows to derive a closed-form upper bound of the single-cell residual interference to noise ratio (RINR).
   – We formulate and solve in closed-form a dynamic feedback bit allocation (DBA) problem for sum-cluster RINR minimization, based on the derived upper bound of the RINR and so as to enable an efficient feedback.

  In principle, the proposed subspace feedback technique is not restricted to the GIA but the general IA, and also in more general scenarios, such as the feedback of decoders in FDD/TDD downlink and the feedback of precoders in FDD/TDD uplink.

The three main contributions above jointly provide a comprehensive holistic design of the multi-cell MIMO-IMAC system under limited feedback. Furthermore, we also discuss the potential implementation, the required backhaul overhead, and the complexity of the proposed GIA. These novel results presented in Chapter 2 (Interference Alignment under Limited Feedback) have been previously reported in:

- P. Cao, A. Zappone and E. A. Jorswieck, "Grouping-based interference alignment with IA-Cell assignment in Multi-Cell MAC under limited feedback", *submitted to IEEE Trans. on Signal Process.*, Sep., 2014.
- P. Cao and E. A. Jorswieck, "Robust Optimization for Multi-Cell Interfering MIMO-MAC under Limited Feedback", *Proc. of IEEE ICASSP*, Florence, Italy, May, 2014.

1.5 Contributions Outside the Scope of the Thesis

Considering the limited space and also to make the content consistent, the thesis does not contain all the work and results during my Ph.D. study but focuses on the resource allocation for MIMO interference networks. My Ph.D. research mainly includes resource allocation (spectral efficiency and energy efficiency) for MIMO interference networks and amplify-forward (AF) relay-aided systems. Besides of some results on MIMO interference networks presented in the thesis, we also give a brief introduction of the contributions outside of the scope of the thesis.

- *Energy efficiency for MIMO AF relay channel:* Nowadays, more and more energy is required to achieve a higher spectral efficiency. However, this usually conflicts with both the shortage of energy and high CO_2 emissions. Therefore, instead of solely targeting the spectral efficiency and energy consumption, respectively, we focus on designing linear transceiver to maximize the *energy efficiency* – the ratio between achievable rate and the total consumed power (bits/Joule). Energy efficiency also provides an insight to design the system, which is also a popular issue of the future 5G [CSS+14].

A framework for energy-efficient resource allocation in a single-user AF relay-assisted MIMO system without direct link is devised. The performance metric to optimize is the ratio between the system's achievable rate and the total consumed power. The optimization is carried out w.r.t. the source precoding and relay amplifying matrices, subject to quality-of-service (QoS) and power constraints. Such a challenging non-convex optimization problem is tackled by means of fractional programming and alternating maximization algorithms, for various CSI assumptions at the source and relay, such as perfect CSI [CCHJ12], statistical CSI for either the source-relay or the relay-destination channel [ZCJ13a, ZCJ14a, ZCJ14b] and statistical CSI for both source-relay and relay-destination channels [ZCJ14c]. An extension to multi-user AF MIMO system is studied in [ZCJ13b].

These contribution have previously published in:

- P. Cao, Z. Chong, Z. K. M. Ho and E. A. Jorswieck, "Energy-efficient power allocation for amplify-and-forward MIMO relay channel", *Proc. of IEEE CAMAD*, Barcelona, Spain, Sep., 2012.
- A. Zappone, P. Cao and E. A. Jorswieck, "Low-complexity energy efficiency optimization with statistical CSI in two-hop MIMO systems", *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1398–1402, Nov., 2014.
- A. Zappone, P. Cao and E. A. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI", *IEEE Trans. on Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan., 2014.
- A. Zappone, P. Cao and E. A. Jorswieck, "Green resource allocation in relay-assisted MIMO systems with statistical channel state information", *Proc. of IEEE ICASSP*, Florence, Italy, May, 2014.
- A. Zappone, P. Cao and E. A. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with statistical CSI", *Proc. of IEEE ChinaSIP*, Beijing, China, Jul., 2013.
- A. Zappone, P. Cao and E. A. Jorswieck, "Energy efficiency optimization

in relay-assisted multi-user MIMO systems", *Proc.of ACSSC*, Pacific Grove, CA, Nov., 2013.

- *Source energy-saving in AF relay-aided system:* Nowadays, the increasing demand for higher data rate and ubiquitous connectivity of a smart phone significantly conflicts with its limited battery lifetime. In order to prolong the battery lifetime, we desire to save the uplink transmit power of a mobile terminal with the aid of cooperative relays with higher transmit power level (e.g., powered by electrical networks), since the battery power is much more scarce.

  In [CJ13], we consider a relay-aided two-hop system consisting one source (e.g., a MS in uplink), one AF relay and one destination (e.g., a BS). We jointly optimize the transceiver strategies to minimize the source transmit power subject to a rate requirement. The optimal transceiver strategies are derived for both the perfect CSI and statistical CSI, from which we additionally find the source energy-saving region when the relay location is fixed and that for a relay when the source location is fixed. In [ZC14], we extend the single source energy saving problem to the multiple sources and multiple destinations scenario (i.e., multi-cell relay-aided cellular environment). This is formulated as a multi-buyer multi-seller power trading game and solved by an auction mechanism. These contribution have been previously appeared in:

  – P. Cao and E. A. Jorswieck, "Source energy-saving performance in amplify-and-forward relay-assisted wireless system", *Proc. of IEEE VTC-Spring*, Dresden, Germany, Jun., 2013.

  – Y. Zhong, P. Cao and E. A. Jorswieck, "Power trading in multi-cell multi-user relay-assisted uplink with private budget limits", *Proc. of the IEEE WCNC*, Istanbul, Turkey, Apr., 2014.

- *Massive MIMO and cognitive radio networks:* Some novel results on the massive MIMO system and cognitive radio MIMO networks are not included in the thesis. In [CJ14a], we consider the limited feedback problem in a multi-cell spatially-correlated massive MIMO FDD system. The very high dimension of massive MIMO channels makes the limited feedback schemes in traditional MIMO inefficient. This motivates us to explore a novel approach to reduce the feedback overhead in two stages: 1) the original channel vector is sparasified in the spatial frequency domain by the discrete cosine transform (DCT); 2) the large DCT coefficients can be further compressed by VQ. By this DCT and VQ based limited feedback approach, a very high dimensional channel vector can be indicated by only an index with an analyzable distortion. In [JC13], we propose a novel distributed two-stage resource allocation technique for MIMO cognitive radio links. Each primary link occupies exclusively part of the resources and offers the opportunity to coexistence. In the first stage, the association between secondary links and primary links is determined by the distributed stable matching. In the second phase, the optimal price and transmission strategies corresponding to a Walrasian equilibrium are obtained. These contribution have been published in:

- P. Cao and E. A. Jorswieck, "DCT and VQ based limited feedback in spatially-correlated massive MIMO systems", *Proc. of IEEE SAM*, A Coruna, Spain, Jun., 2014.
- E. A. Jorswieck and P. Cao, "Matching and exchange market based resource allocation in MIMO cognitive radio networks", *Proc. of EUSIPCO*, Marrakech, Maroc, Sep., 2013.

1.6 Outline of the Thesis

The thesis mainly focuses on resource allocation for MIMO interference networks by linear transceiver design. It consists of five chapters and is organized as follows.

In Chapter 1, we give a brief overview of the trends of mobiles communications development. In order to meet the rapidly increasing demand of the data traffic, we provide a potential version of the future generation communication networks – small cells, spectrum sharing and advanced MIMO technique so as to improve the aggregate data rate. This motivates us to focus on studying the spectral efficiency for the MIMO interference networks. More precisely, two models – a multi-user single-stream MIMO IC and a multi-cell MIMO-IMAC are considered. The problem formulations, related work, the techniques and main contributions of the thesis are given, respectively, for both the models. Furthermore, other contributions made during my Ph.D. study but excluded in the thesis are also briefly introduced.

In Chapter 2, we consider a multi-user single-stream MIMO IC. The goal is to compute the Pareto boundary of the achievable rate region, which is very important for system and strategies design and is formulated as a multi-objective optimization problem. However, this problem is a well-known open problem because of the hard coupling between the precoding and decoding strategies. In order to make the problem tractable, we first study a two-user MIMO IC example and derive some new results. For the multi-user MIMO IC, we transform the multi-objective optimization problem into two single-objective optimization problems, inspired by the ϵ-constraint optimization and the weighted Chebyshev optimization, respectively. For each single-objective optimization problem, we provide an alternating optimization algorithm to obtain a high quality suboptimal solution. The analysis and simulation of the proposed algorithms are provided, respectively.

In Chapter 3, we consider the multi-cell MIMO-IMAC (uplink scenario) analyzing the aspects of the GIA, IA-Cell assignment and limited feedback. The presentation in Chapter 3 also follows these three points. First, a complete study (including some new improvements) of the GIA with respect to the DoF and optimal linear transceiver design is performed, which allows for low-complexity and distributed implementation. Second, based on the GIA, the concept of IA-Cell assignment is introduced. Three IA-Cell assignment algorithms are proposed for the setup with different backhaul overhead and their DoF and rate performance are investigated. Third, the performance of the proposed GIA algorithms is studied under limited feedback of the GIA precoders. To enable efficient feedback, a DBA problem is formulated and solved in closed-form. The practical implementation, the required backhaul overhead, and the complexity of the proposed algorithms are analyzed. Numerical results show that our proposed algorithms greatly outperform the traditional GIA under both unlimited and limited feedback.

In Chapter 4, we summarize the contributions of the thesis, pointing out both the advantages and limits of the proposed algorithms. In addition, an outlook on future research directions is provided.

Chapter 2

Pareto Boundary Computation for MIMO IC

In this chapter, we consider an open problem – computing the Pareto boundary of the achievable rate region of a multi-user single-stream MIMO IC. Since most previous research on computing the Pareto boundary focus on the multi-user SISO/MISO/SIMO IC but no previous algorithm work well for the multi-user MIMO IC, this chapter aims to fill this gap. From the perspective of optimization, Pareto boundary computation is a multi-objective optimization problem. The problem is non-convex and even NP-hard because of the hard coupling between the transmit and receive strategies in the rate expression. In order to make the problem tractable, the multi-objective optimization problem is transferred into two single-objective optimization problems, inspired by two approaches – *ϵ-constraint optimization in (1.6)* and *weighted Chebyshev optimization in (1.7)*. For each single-objective optimization problem, it is still a non-convex problem. The alternating optimization algorithm is applied to optimize the variables separately and converges to a high-quality suboptimal solution. With different constraint values or weights, a series of achievable operating points serves a close-to-optimal inner bound of the complete Pareto boundary.

This chapter is organized as follows: the multi-user single-stream MIMO IC model is introduced and the definitions of associated achievable rate region and Pareto boundary are given in Chapter 2.1. In Chapter 2.2, we study a two-user MIMO IC example and derive some new results, such as an equivalent form of the SINR expression (Proposition 1), and the closed-form solutions to the non-strict Pareto boundary, the end points of the strict Pareto boundary (Proposition 3) and some ZF operating points. In order to compute the strict Pareto boundary, we formulate a *single constraint rate maximization* problem in Chapter 2.3 to maximize one rate while the other rate is fixed. This problem formulation is inspired by the ϵ-constraint optimization, but we replace its "inequality" in the constraint with "equality" to make sure that the achievable rate of a user is increasing always along the same direction. However, the single constraint rate maximization problem is still a non-convex optimization problem. With the aid of the SDP and matrix rank-1 decomposition techniques, the single constraint rate maximization problem is solved by alternating optimization of each transmit beamforming vector. The convergence of the proposed algorithm is guaranteed and the convergent point by the proposed initialization scheme corresponds to a good-quality suboptimal solution. This single constraint rate maximization problem is extended to the multi-user single-stream MIMO IC in Chapter 2.3.3. Inspired by the weighted Chebyshev optimization, we propose another optimization problem – *alternating rate profile optimization* problem in Chapter 2.4, which is solved by alternatively optimizing the transmit beamforming vectors in an equivalent multi-user MISO IC (for the fixed receive beamforming vectors) and the receive beamforming vectors in an equivalent multi-user SIMO IC (for the fixed transmit beamforming vectors), respectively. This algorithm is proven to converge to a stationary point of the original multi-objective optimization problem. Finally, the effectiveness of two proposed algorithms are illustrated and evaluated numerically in Chapter 2.5.

2.1 System Model

2.1.1 Signal Model

We consider a set of K transmitter-receiver pairs $\mathcal{K} = \{1, \cdots, K\}$ operating in the same spectral band. The transmitters and receivers are equipped with multiple antennas. Without loss of generality, we assume that each transmitter and each receiver have $N_T \geq 2$ and $N_R \geq 2$ antennas[1], respectively. This setup is modeled as a K-user MIMO IC as shown in Figure 2.1, which is well-matched to the downlink scenario (a transmitter serves as a BS), the uplink scenario (a receiver serves as a BS) or multiple D2D cognitive radio links. For the convenience, we do not focus on a specified application scenario, just denoting the k-th transmitter, ℓ-th receiver and their communication link by TX_k, RX_ℓ and $\mathrm{TX}_k \mapsto \mathrm{RX}_\ell$, respectively. In particular, we refer the direct link $\mathrm{TX}_k \mapsto \mathrm{RX}_k$ to be user k.

We assume that each transmitter sends a single symbol $x_k \sim \mathcal{CN}(0, 1)$ to its intended receiver. The transmit beamforming vector used at TX_k is denoted by $\boldsymbol{w}_k$ from the set

$$\mathcal{W} \triangleq \{\boldsymbol{w} \in \mathbb{C}^{N_T} : \quad ||\boldsymbol{w}||^2 \leq 1\}. \tag{2.1}$$

[1] The proposed algorithms and derived results in this chapter can be also applied to the general case where each transmitter and each receiver have different antennas.

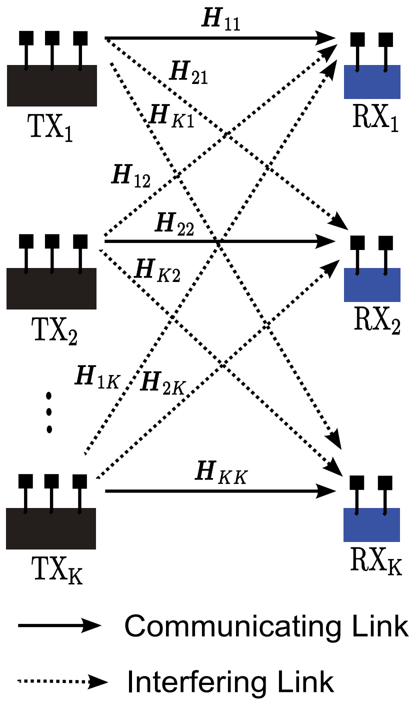


Figure 2.1: The K-user MIMO IC model: Each transmitter TX_k is paired with a single receiver RX_k, and all links are non-negligible.

Without loss of generality, the transmit power budget of each transmitter is normalized to be 1.

The received signal at RX_k is modeled as

$$\boldsymbol{y}_k = \boldsymbol{H}_{kk}\boldsymbol{w}_k x_k + \sum_{\ell \neq k} \boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell x_\ell + \boldsymbol{n}_k, \ \forall k \in \mathcal{K} \tag{2.2}$$

where $\boldsymbol{H}_{kk}, \boldsymbol{H}_{k\ell} \in \mathbb{C}^{N_R \times N_T}$ denote the flat fading channel matrices of the direct link $\mathrm{TX}_k \mapsto \mathrm{RX}_k$ and the interference link $\mathrm{TX}_\ell \mapsto \mathrm{RX}_k$, respectively. And $\boldsymbol{n}_i \sim \mathcal{CN}(0, \sigma_k^2 \boldsymbol{I}_{N_R})$ is the additive white Gaussian noise (AWGN) vector.

By the equalization with $\boldsymbol{g}_k \in \mathbb{C}^{N_R}$ at RX_k, the desired data x_k is decoded as

$$\widehat{\boldsymbol{x}} = \boldsymbol{g}_k^H \boldsymbol{y}_k. \tag{2.3}$$

Assume that the interference from the other transmitter is treated as additive Gaussian noise at each receiver. The achievable rate of user k is given by

$$R_k(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_k) = \log_2\left(1 + \mathrm{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_k)\right), \tag{2.4}$$

where $\{\boldsymbol{w}\}_{\mathcal{K}}$ denotes the set of K transmit beamforming vectors, and $\mathrm{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_k)$

denotes the SINR of the user k and is expressed as

$$\mathrm{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_k) = \frac{|\boldsymbol{g}_k^H \boldsymbol{H}_{kk} \boldsymbol{w}_k|^2}{\sigma_k^2 ||\boldsymbol{g}_k||^2 + \sum_{\ell \neq k} |\boldsymbol{g}_k^H \boldsymbol{H}_{k\ell} \boldsymbol{w}_\ell|^2}. \tag{2.5}$$

Since the SINR in (2.5) does not depend on the receive power $||\boldsymbol{g}_k||^2$, the receive beamforming vector $\boldsymbol{g}_k$ can be chosen from

$$\mathcal{G} \triangleq \{\boldsymbol{g} \in \mathbb{C}^{N_R} : \quad ||\boldsymbol{g}||^2 = 1\}. \tag{2.6}$$

2.1.2 Pareto Boundary of Achievable Rate Region

The *achievable rate region* is defined as a set of the achievable rate tuples with all the feasible transmit and receive beamforming vectors, i.e.,

$$\mathcal{R} \triangleq \bigcup_{\forall \boldsymbol{w}_k \in \mathcal{W}, \forall \boldsymbol{g}_k \in \mathcal{G}} \left(R_1(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_1), \cdots, R_K(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_K)\right). \tag{2.7}$$

Please note that the achievable rate region $\mathcal{R}$ is not the capacity region but the subset of the capacity region, which can be achieved by the simple linear transceiver. Its outermost boundary is called *Pareto boundary*, denoted by the set

$$\mathcal{R}^\star \triangleq \bigcup_{\forall \boldsymbol{w}_k \in \mathcal{W}^\star, \forall \boldsymbol{g}_k \in \mathcal{G}^\star} (R_1^\star(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_1), \cdots, R_K^\star(\{\boldsymbol{w}\}_{\mathcal{K}}, \boldsymbol{g}_K)). \tag{2.8}$$

where $\mathcal{W}^\star$ and $\mathcal{G}^\star$ denote the sets of Pareto-optimal transmit and receive beamforming vectors, respectively. Given an arbitrary operating point in the set $\mathcal{R}\backslash\mathcal{R}^\star$ (inside the rate region), we can always find at least one operating point in $\mathcal{R}^\star$ that *dominates* the given point, i.e., all users' rates can be further improved. Therefore, we focus on the operating points on the Pareto boundary.

For convenience, we denote an operating point by $\boldsymbol{r} \triangleq (R_1, \cdots, R_K) \in \mathcal{R}$. If a point is on the Pareto boundary, we say it is *Pareto-optimal*, denoted by $\boldsymbol{r}^\star$. More precisely, the *Pareto-optimality* is defined as follows.

Definition 1. *A rate-tuple $\boldsymbol{r}^\star$ is (strict) Pareto-optimal iff there does not exist another rate-tuple $\boldsymbol{r} \in \mathcal{R}^\star$ satisfying $\boldsymbol{r} > \boldsymbol{r}^\star$ ($\boldsymbol{r} \geq \boldsymbol{r}^\star$ and $\boldsymbol{r} \neq \boldsymbol{r}^\star$), where the inequality is component-wise.*

Accordingly, at a Pareto-optimal point, it is impossible to strictly improve the performance of all users simultaneously. We denote the *strict Pareto-optimal* points form a subset of the Pareto-optimal points by $\mathcal{R}_+^\star$. Those Pareto-optimal points in the set $\mathcal{R}^\star\backslash\mathcal{R}_+^\star$ are defined to be the *non-strict Pareto boundary*.

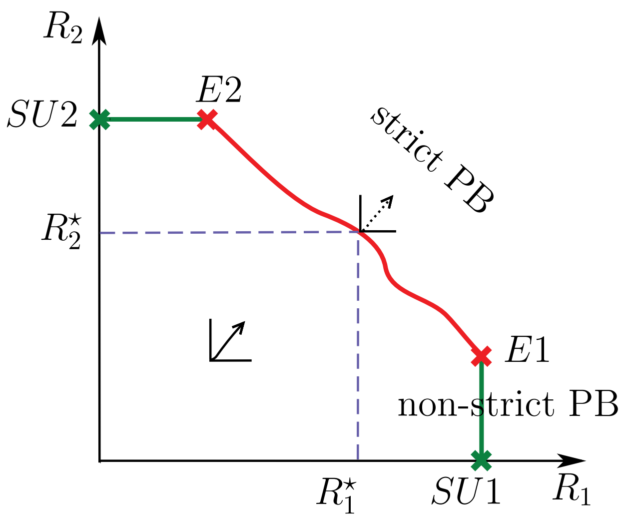


Figure 2.2: Illustration of a two-dimensional achievable rate region and the Pareto boundary

A two-dimensional rate region and its Pareto boundary are illustrated in Figure 2.2. It shows that the Pareto boundary consists of the *strict Pareto boundary* (the upper-right part graphically, denoted by "strict PB") and the *non-strict Pareto boundary* (including the vertical part and the horizontal part graphically, denoted by "non-strict PB"), divided by two points "E1" and "E2", where "E1" and "E2", "SU1" and "SU2" are defined as two *ending points of the strict Pareto boundary* and two *single user points*.

Remark 4. *For an arbitrary point on the strict Pareto boundary, it is impossible to improve one rate without simultaneously decreasing the others. For a point on the non-strict Pareto boundary, one rate can be further improved while the other rates remains unchanged.*

The Pareto boundary usually consists of both strict Pareto boundary and non-strict Pareto boundary. However, in some cases, there is only a unique strict Pareto boundary (e.g., a triangular region for the two-user SIMO IC) or no non-strict Pareto-optimal points.

2.2 New Results for Two-User Case

In this section, we consider a two-user single-stream MIMO IC and present some new results derived in [CSJ12] and [CJS13].

2.2.1 Rate with MMSE Receiver

In the linear transceiver design, it is well-known that the MMSE filter is the *optimal* receive beamforming scheme for the given transmit strategies. Thus, the MMSE filter should be employed at each receiver so as to gain maximum at receivers.

In the two-user case, the MMSE receive beamforming vector at RX_k is expressed by

$$\boldsymbol{g}_k = \left(\sigma_k^2 \boldsymbol{I}_{N_R} + \sum_{\ell=1}^{2} \boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell {\boldsymbol{w}_\ell}^H {\boldsymbol{H}_{k\ell}}^H\right)^{-1} \boldsymbol{H}_{kk}\boldsymbol{w}_k,\ \forall k \in \{1,2\} \tag{2.9}$$

Plugging (2.9) into (2.5), the SINR of user k becomes

$$\mathrm{SINR}_k(\boldsymbol{w}_1, \boldsymbol{w}_2) = \boldsymbol{w}_k^H \underbrace{\boldsymbol{H}_{kk}^H \left(\sigma_k^2 \boldsymbol{I}_{N_R} + \boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell \boldsymbol{w}_\ell^H \boldsymbol{H}_{k\ell}^H\right)^{-1} \boldsymbol{H}_{kk}}_{\triangleq \boldsymbol{A}_k(\boldsymbol{w}_\ell)} \boldsymbol{w}_k,\ \forall k \neq \ell \in \{1,2\}. \tag{2.10}$$

The complex mathematical structure (inverse of the sum of matrices and product of matrices) causes a hard-coupling problem of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in the SINR in (2.10), which makes it difficult to analyze the SINR directly. To gain more insights into this coupling problem, we propose an equivalent form of the SINR expression.

Proposition 1. *For the two-user single-stream MIMO IC, the SINR in (2.10) can be reformulated as*

$$\mathrm{SINR}_k(\boldsymbol{w}_1, \boldsymbol{w}_2) = \sin^2(\theta_{H,k}) \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2} + \cos^2(\theta_{H,k}) \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2}, \tag{2.11}$$

where $\cos(\theta_{H,k}) = \left| \overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}^H \cdot \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell} \right|$ *and* $\theta_{H,k} \in [0, \pi/2]$.

Proof. Refer to Proof 2.7.1. □

Note that the $\mathrm{SINR}_k(\boldsymbol{w}_1, \boldsymbol{w}_2)$ can be considered as a combination of $\frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2}$ and $\frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2}$ with the weights $\sin^2(\theta_{H,k})$ and $\cos^2(\theta_{H,k})$. That is, $\mathrm{SINR}_k(\boldsymbol{w}_1, \boldsymbol{w}_2)$ depends not only on the desired signal power $\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2$ and the interference power $\|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2$, but also on the Hermitian angle $\theta_{H,k}$ between the directions $\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}$ and $\overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}$. The SINR is coupled in a difficult way because of the existence of $\theta_{H,k}$. This is why it is more difficult to analyze the SINR of the MIMO IC than the MISO/SIMO IC.

2.2.2 Pareto Boundary and Some Key Points

In the two-user case, the two-dimensional Pareto boundary is illustrated in Figure 2.2. The goal is to compute some key points on/in the Pareto boundary.

A. Strict Pareto Boundary

First, we propose a necessary condition for the transmit beamforming to achieve the strict Pareto boundary.

Proposition 2. *For the two-user single-stream MIMO IC, the strict Pareto boundary can be achieved only if both the transmitters consume full power, i.e.,* $\|\boldsymbol{w}_1\|^2 = \|\boldsymbol{w}_2\|^2 = 1$.

Proof. Refer to Proof 2.7.2. □

Remark 5. *Proposition 2 provides a strict Pareto-optimal transmit power allocation policy. When both the transmitters consume full power, the two strict Pareto-optimal transmit beamforming vectors design reduces to the optimization of two transmit beamforming patterns (directions).*

Here, we define a set of all the feasible beamforming vectors with full transmit power as $\mathcal{W}_{\mathcal{FP}} \triangleq \left\{\boldsymbol{w} \in \mathbb{C}^{N_T} : \|\boldsymbol{w}\|^2 = 1\right\}$. Note that all the strict Pareto-optimal transmit beamforming vectors should be in the set $\mathcal{W}_{\mathcal{FP}}$.

B. Single-User Points

Let $SU1(\overline{R}_1, 0)$ and $SU2(0, \overline{R}_2)$ denote the two single-user points. A single-user point can be easily achieved when only one transmitter TX_k works and simultaneously operates "egoistically" to maximize its own rate. The maximum achievable rate $\overline{R}_k$ of user k and its associated "egoistic" strategy $\boldsymbol{w}_k^{Ego}$ can be determined by

$$\overline{R}_k = \log_2\left(1 + \frac{\lambda_1(\boldsymbol{H}_{kk}^H \boldsymbol{H}_{kk})}{\sigma_k^2}\right), \ \boldsymbol{w}_k^{Ego} = \boldsymbol{u}_1(\boldsymbol{H}_{kk}^H \boldsymbol{H}_{kk}), \ \forall k, \tag{2.12}$$

where $(\lambda_n(\boldsymbol{X}), \boldsymbol{u}_n(\boldsymbol{X}))$ denotes the n-th largest eigenvalue and its associated eigenvector of an arbitrary matrix $\boldsymbol{X}$.

C. Ending Points of Strict Pareto Boundary

Let $E1(\overline{R}_1, \underline{R}_2)$ and $E2(\underline{R}_1, \overline{R}_2)$ denote the two ending points of strict Pareto boundary. Each ending point of the strict Pareto boundary can be achieved when one transmitter employs an "altruistic" strategy to generate no interference to the other receiver and simultaneously to maximize its own rate and the other transmitter operates "egoistically". For $E1(\overline{R}_1, \underline{R}_2)$, we easily find from (2.11) that $\theta_{H,1} = \pi/2$ results in no interference in the interference link $\mathrm{TX}_2 \mapsto \mathrm{RX}_1$. How to find the "altruistic" strategy $\boldsymbol{w}_2^{Alt}$ is shown as follows.

Proposition 3. *The ending point of strict Pareto boundary* $E1(\overline{R}_1, \underline{R}_2)$ *can be achieved by* $(\boldsymbol{w}_1^{Ego}, \boldsymbol{w}_2^{Alt})$*, where* $\overline{R}_1$ *and* $\boldsymbol{w}_1^{Ego}$ *are in (2.12) and*

$$\underline{R}_2 = \log_2\left(1 + \boldsymbol{w}_2^{Alt,H}\boldsymbol{A}_2(\boldsymbol{w}_1^{Ego})\boldsymbol{w}_2^{Alt}\right), \tag{2.13}$$

$$\boldsymbol{w}_2^{Alt} = \frac{\Pi^{\perp}_{\boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1^{Ego}}\boldsymbol{u}_1\left(\boldsymbol{B}_1, \Pi^{\perp}_{\boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1^{Ego}}\right)}{\left\|\Pi^{\perp}_{\boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1^{Ego}}\boldsymbol{u}_1\left(\boldsymbol{B}_1, \Pi^{\perp}_{\boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1^{Ego}}\right)\right\|}, \tag{2.14}$$

with $\boldsymbol{B}_1 \triangleq \Pi^{\perp}_{\boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1^{Ego}}\boldsymbol{A}_2(\boldsymbol{w}_1^{Ego})\Pi^{\perp}_{\boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1^{Ego}}$.

Proof. Refer to Proof 2.7.3. □

The other ending point $E2(\underline{R}_1, \overline{R}_2)$ with $(\boldsymbol{w}_1^{Alt}, \boldsymbol{w}_2^{Ego})$ can be easily obtained following Proposition 3 by interchanging the indices.

D. Non-Strict Pareto-Optimal Points

Consider a non-strict Pareto-optimal point $(R_1^\star, R_2^\star) \in \mathcal{R}^\star \backslash \mathcal{R}_+^\star$. For the non-strict Pareto boundary, either the horizontal part or the vertical part starts and ends with a single user point and an ending point. Therefore, an arbitrary point $(R_1^\star, R_2^\star)$ on the non-strict Pareto boundary can be computed as

$$R_k^\star = \gamma \cdot \underline{R}_k \quad \text{and} \quad R_\ell^\star = \overline{R}_\ell, \quad \forall k,\ \ell \in \{1, 2\},\ k \neq \ell$$

where $k = 1$ and $k = 2$ correspond to the horizontal part and the vertical part, respectively. The scalar γ satisfies $\gamma \in [0, 1)$. The point $(R_1^\star, R_2^\star)$ becomes a single-user point or an ending point when $\gamma = 0$ or $\gamma = 1$, respectively. The associated non-strict Pareto-optimal transmit strategies are

$$\boldsymbol{w}_k^\star = \sqrt{\frac{2^{\gamma \underline{R}_k} - 1}{2^{\underline{R}_k} - 1}}\boldsymbol{w}_k^{Alt} \quad \text{and} \quad \boldsymbol{w}_\ell^\star = \boldsymbol{w}_\ell^{Ego}, \tag{2.15}$$

from which we find that it is not necessary for both the transmitters to simultaneously spend full power so as to achieve the non-strict Pareto boundary. Thus, the non-strict Pareto-optimal power allocation policy is different from the strict Pareto-optimal power allocation policy (full power transmission in Proposition 2).

E. Zero-Forcing Points

A ZF point refers to an operating point where there is no interference at both receivers, denoted by $ZF(R_1^{ZF}, R_2^{ZF})$. Although these points are not on the Pareto boundary, it is still interesting to study ZF strategies if there exists an additional requirement (e.g., interference temperature or secrecy constraints) that each transmitter does not leak its own signal to the unintended receivers.

Inspired by (2.11), we find that $\theta_{H,1} = \theta_{H,2} = \pi/2$ results in no interference in both interference links $\text{TX}_2 \mapsto \text{RX}_1$ and $\text{TX}_1 \mapsto \text{RX}_2$. The ZF conditions are

$$\begin{aligned}\theta_{H,1} = \pi/2 &\Leftrightarrow \overrightarrow{\boldsymbol{H}_{11}\boldsymbol{w}_1}^H \cdot \overrightarrow{\boldsymbol{H}_{12}\boldsymbol{w}_2} = 0 \\ &\Leftrightarrow \boldsymbol{H}_{11}\boldsymbol{w}_1 \perp \boldsymbol{H}_{12}\boldsymbol{w}_2 \Leftrightarrow \boldsymbol{w}_2 \perp \boldsymbol{H}_{12}^H \boldsymbol{H}_{11}\boldsymbol{w}_1, \end{aligned} \tag{2.16a}$$

$$\begin{aligned}\theta_{H,2} = \pi/2 &\Leftrightarrow \overrightarrow{\boldsymbol{H}_{22}\boldsymbol{w}_2}^H \cdot \overrightarrow{\boldsymbol{H}_{21}\boldsymbol{w}_1} = 0 \\ &\Leftrightarrow \boldsymbol{H}_{22}\boldsymbol{w}_2 \perp \boldsymbol{H}_{21}\boldsymbol{w}_1 \Leftrightarrow \boldsymbol{w}_2 \perp \boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1, \end{aligned} \tag{2.16b}$$

from which and under a sufficient condition[2], i.e., $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11}\boldsymbol{w}_1 \parallel \boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1$, to guarantee that both (2.16a) and (2.16b) hold simultaneously, we obtain some ZF transmit strategies

$$\boldsymbol{w}_1^{ZF} = \boldsymbol{u}_n(\boldsymbol{H}_{22}^H \boldsymbol{H}_{21}, \boldsymbol{H}_{12}^H \boldsymbol{H}_{11}), \forall n \in \{1, 2, ..., N_T\}, \tag{2.17a}$$

$$\boldsymbol{w}_2^{ZF} = \sum_{n=1}^{N_T - 1} c_n \boldsymbol{u}_n(\Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{ZF}}), \tag{2.17b}$$

where (2.17a) is derived from $\boldsymbol{H}_{21}^H \boldsymbol{H}_{11}\boldsymbol{w}_1 \parallel \boldsymbol{H}_{22}^H \boldsymbol{H}_{12}\boldsymbol{w}_1$, i.e., $\boldsymbol{w}_1^{ZF}$ is a joint eigenvector of matrices $\boldsymbol{H}_{22}^H \boldsymbol{H}_{12}$ and $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11}$. From (2.16), $\boldsymbol{w}_2^{ZF}$ should be in the nullspace of $\boldsymbol{H}_{21}^H \boldsymbol{H}_{11}\boldsymbol{w}_1{}^{ZF}$ or $\boldsymbol{H}_{22}^H \boldsymbol{H}_{12}\boldsymbol{w}_1{}^{ZF}$. In (2.17b), $\{c_n\}_{n=1}^{N_T - 1}$ are complex-valued numbers and satisfy $\sum_{n=1}^{N_T - 1} |c_n|^2 = 1$.

A ZF point $ZF(R_1^{ZF}, R_2^{ZF})$ can be achieved by $(\boldsymbol{w}_1^{ZF}, \boldsymbol{w}_2^{ZF})$ as

$$R_k^{ZF}(\boldsymbol{w}_1^{ZF}, \boldsymbol{w}_2^{ZF}) = \log_2 \left(1 + \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k^{ZF}\|^2}{\sigma_k^2}\right) \ \forall k. \tag{2.18}$$

Remark 6. *The ZF strategies are derived in (2.17) based on the sufficient condition* $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11}\boldsymbol{w}_1 \parallel \boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1$*. This sufficient condition is also necessary when* $N_T = N_R = 2$*. In a two-dimensional space, (2.16) requires that* $\boldsymbol{w}_2$ *is orthogonal to both* $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11}\boldsymbol{w}_1$ *and* $\boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1$*, which is equivalent to that* $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11}\boldsymbol{w}_1$ *and* $\boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1$ *must/should be aligned (sufficient and necessary condition). However, for a high-dimensional space* ($N_T \geq 3$)*,* $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11}\boldsymbol{w}_1$ *and* $\boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1$ *are not necessary to be aligned but both should lie in the null space of* $\boldsymbol{w}_2$*.*

After exactly obtaining the non-strict Pareto boundary and the ending points of strict Pareto boundary, we focus on computing the strict Pareto boundary in the following.

[2]This condition is the same as that in [CHHT12], while we derive it from a different way, i.e., by analyzing Hermitian angle in (2.11).

In principle, the computation of the Pareto boundary can be formulated as a multi-objective optimization problem, i.e.,

$$\boldsymbol{r}^\star = \arg \max_{\{\boldsymbol{w}\}_{\mathcal{K}} \in \mathcal{W}_{\mathcal{FP}}} (R_1(\{\boldsymbol{w}\}_{\mathcal{K}}), \cdots, R_K(\{\boldsymbol{w}\}_{\mathcal{K}})) \tag{2.19}$$

However, it is difficult to directly solve a non-convex multi-objective optimization problem in mathematic because the hard coupling of transmit/receiver beamforming vectors exists in the SINR. In order to make (2.19) tractable, we reformulate two different single-objective optimization problems based on *ϵ-constraint optimization* and *weighted Chebyshev optimization*, respectively, which will be solved in Chapter 2.3 and Chapter 2.4, respectively.

2.3 Pareto Boundary Computation: Single Constraint Rate Maximization

In this section, we first focus on computing the strict Pareto boundary for the two-user single-stream MIMO IC based on *ϵ-constraint optimization*. Then, we show the extension to the multi-user case.

2.3.1 Problem Formulation & Algorithm: Two-User Case

A. Problem Formulation

Since the rate region of the two-user single-beam MIMO IC is always a *normal region*[3] according to the proof of Proposition 2, there should be only one intersection point between the line $R_k(\boldsymbol{w}_1, \boldsymbol{w}_2) = R_k^\star$ where $R_k^\star \in (\underline{R}_k, \overline{R}_k)$ and the strict Pareto boundary. Thus, an arbitrary point on the strict Pareto boundary can be uniquely determined when one rate is fixed and the other rate is maximized. This motivates us to propose the following *single constraint rate maximization problem*

$$\begin{cases} \max\limits_{\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{FP}}} & \mathrm{SINR}_1(\boldsymbol{w}_1, \boldsymbol{w}_2) \\ \quad\text{s.t.} & \mathrm{SINR}_2(\boldsymbol{w}_1, \boldsymbol{w}_2) = \mathrm{SINR}_2^\star. \end{cases} \tag{2.20}$$

where $\mathrm{SINR}_2^\star \in (2^{\underline{R}_2} - 1, 2^{\overline{R}_2} - 1)$ is a SINR constraint, and $\boldsymbol{w}_1, \boldsymbol{w}_2$ should be in $\mathcal{W}_{\mathcal{FP}}$ according to Proposition 2. In order to guarantee that $\mathrm{SINR}_1(\boldsymbol{w}_1, \boldsymbol{w}_2)$ increases always along the same direction, we use *equality* in the constraint instead of the inequality

[3]A set $\mathcal{G} \subseteq \mathbb{R}_n^+$ is called a *normal region* if for any two points $\boldsymbol{x} \in \mathcal{G}, \boldsymbol{x}' \in \mathbb{R}_n^+$ such that if $\boldsymbol{x}' \leq \boldsymbol{x}$, then $\boldsymbol{x}' \in \mathcal{G}$, too.

in the ϵ-constraint optimization. If Problem (2.20) can be solved optimally, a strict Pareto-optimal point $(R_1^\star, R_2^\star) = (\log_2(1+\mathrm{SINR}_1(\boldsymbol{w}_1^\star, \boldsymbol{w}_2^\star)), \log_2(1+\mathrm{SINR}_2^\star(\boldsymbol{w}_1^\star, \boldsymbol{w}_2^\star)))$ is achieved by the optimal solution $(\boldsymbol{w}_1^\star, \boldsymbol{w}_2^\star)$ to Problem (2.20).

For Problem (2.20), direct joint optimization of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ is analytically intractable due to the hard-coupling problem of them in both the objective and the constraint. To solve (2.20), an *alternating optimization algorithm* [Ber99] is applied to optimize $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ alternatively by solving two single-beamforming vector optimization problems at each iteration. In the following, how to solve each single-beamforming vector problem is studied.

B. Optimization of $\boldsymbol{w}_1$

For a given *feasible* $\boldsymbol{w}_2$ (the feasibility of $\boldsymbol{w}_2$ will be given in Proposition 4), Problem (2.20) becomes an optimization problem w.r.t. a single beamforming vector $\boldsymbol{w}_1$. The constraint in (2.20) becomes

$$\begin{aligned}
& \boldsymbol{w}_2^H \boldsymbol{H}_{22}^H (\sigma_2^2 \boldsymbol{I} + \boldsymbol{H}_{21}\boldsymbol{w}_1{\boldsymbol{w}_1}^H {\boldsymbol{H}_{21}}^H)^{-1} \boldsymbol{H}_{22}\boldsymbol{w}_2 = \mathrm{SINR}_2^\star && (2.21\text{a})\\
\overset{(a)}{\Leftrightarrow}\ & \boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22}\boldsymbol{w}_2 - \frac{|\boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{12}\boldsymbol{w}_1|^2}{\sigma_2^2 + \|\boldsymbol{H}_{21}\boldsymbol{w}_1\|^2} = \sigma_2^2 \mathrm{SINR}_2^\star, && (2.21\text{b})\\
\overset{(b)}{\Leftrightarrow}\ & \frac{{\boldsymbol{w}_1}^H \boldsymbol{H}_{21}^H \boldsymbol{H}_{22}\boldsymbol{w}_2\boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{21}\boldsymbol{w}_1}{\boldsymbol{w}_1^H(\sigma_2^2\boldsymbol{I} + \boldsymbol{H}_{21}^H\boldsymbol{H}_{21})\boldsymbol{w}_1} = \boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22}\boldsymbol{w}_2 - \sigma_2^2 \mathrm{SINR}_2^\star, && (2.21\text{c})\\
\overset{(c)}{\Leftrightarrow}\ & \boldsymbol{w}_1^H \boldsymbol{C}(\boldsymbol{w}_2)\boldsymbol{w}_1 = 0 \text{ and } \boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22}\boldsymbol{w}_2 \geq \sigma_2^2 \mathrm{SINR}_2^\star. && (2.21\text{d})
\end{aligned}$$

The transformation (a) is based on the matrix inverse lemma. The transformation (b) is due to $\|\boldsymbol{w}_1\|^2 = 1$. In the transformation (c), the nonnegative left-hand side of (2.21c) demands $\boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22}\boldsymbol{w}_2 \geq \sigma_2^2 \mathrm{SINR}_2^\star$, and $\boldsymbol{C}(\boldsymbol{w}_2)$ is a Hermitian matrix defined as

$$\boldsymbol{C}(\boldsymbol{w}_2) \triangleq \boldsymbol{H}_{21}^H \boldsymbol{H}_{22}\boldsymbol{w}_2\boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{21} - (\boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22}\boldsymbol{w}_2 - \sigma_2^2 \mathrm{SINR}_2^\star)(\sigma_2^2 \boldsymbol{I} + \boldsymbol{H}_{21}^H \boldsymbol{H}_{21}). \tag{2.22}$$

Then, $\boldsymbol{w}_1$ can be optimized by

$$\begin{cases} \displaystyle\max_{\boldsymbol{w}_1 \in \mathcal{W}_{\mathcal{FP}}} & \boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\boldsymbol{w}_1 \\ \quad\text{s.t.} & {\boldsymbol{w}_1}^H \boldsymbol{C}(\boldsymbol{w}_2)\boldsymbol{w}_1 = 0 \end{cases} \tag{2.23}$$

where $\boldsymbol{C}(\boldsymbol{w}_2)$ and $\boldsymbol{A}_1(\boldsymbol{w}_2)$ are Hermitian matrices. Observe that Problem (2.23) is a homogeneous quadratically constrained quadratic program (QCQP), where the objective function is convex because $\boldsymbol{A}_1(\boldsymbol{w}_2)$ is a positive definite matrix but the convexity of constraints is unclear. Generally, it is difficult to solve this non-convex problem.

Note that $\boldsymbol{w}_1^H \boldsymbol{X} \boldsymbol{w}_1 = \mathrm{Tr}(\boldsymbol{X}\boldsymbol{W}_1)$ for any matrix $\boldsymbol{X}$, where $\boldsymbol{W}_1 = \boldsymbol{w}_1 \boldsymbol{w}_1^H$ is a rank-one Hermitian positive semidefinite matrix. By the semidefinite programming and rank relaxation (SDR) method [LMS+10], Problem (2.23) can be transformed to by relaxing the rank-one constraint $\mathrm{rank}(\boldsymbol{W}_1) = 1$

$$\begin{cases} \max\limits_{\boldsymbol{W}_1 \succeq \boldsymbol{0}} & \mathrm{Tr}\left(\boldsymbol{A}_1(\boldsymbol{w}_2)\boldsymbol{W}_1\right) \\ \text{s.t.} & \mathrm{Tr}\left(\boldsymbol{C}(\boldsymbol{w}_2)\boldsymbol{W}_1\right) = 0 \\ & \mathrm{Tr}\left(\boldsymbol{W}_1\right) = 1. \end{cases} \tag{2.24}$$

Observe that the SDR problem (2.24) is convex and solvable, i.e., its respective finite optimal solutions exist for a feasible $\boldsymbol{w_2}$, based on the Weierstrass' Theorem [KZ05]. Its optimal solution $\boldsymbol{W}_1^\star$ can be *efficiently* obtained by a convex optimization toolbox, e.g., SeDuMi [Pol05] or CVX [GB12]. However, the rank of $\boldsymbol{W}_1^\star$ to Problem (2.24) is not necessary to be one because we have discarded the rank constraint $\mathrm{rank}(\boldsymbol{W}_1) = 1$. Therefore, we need to extract an optimal rank-one solution $\boldsymbol{w}_1$ to Problem (2.23) from $\boldsymbol{W}_1^\star$. If $\mathrm{rank}(\boldsymbol{W}_1^\star) = 1$, it is clear that $\boldsymbol{w}_1$ is the eigenvector associated with the unique non-zero eigenvalue of $\boldsymbol{W}_1^\star$, i.e., $\boldsymbol{w}_1 = \boldsymbol{u}_1(\boldsymbol{W}_1^\star)$. Otherwise, some matrix rank-one decomposition techniques are needed. In [AHZ11], Ai et al. provide a series of matrix rank-one decomposition theorems. They imply that *the SDR problems of a large class of complex-valued homogeneous QCQPs with not more than 4 constraints are in fact tight*[4]. Since Problem (2.23) as a homogeneous QCQP with 2 constraints, an optimal $\boldsymbol{w}_1$ to Problem (2.23) can be reconstructed from $\boldsymbol{W}_1^\star$ to Problem (2.24) based on the theorem and algorithm of the matrix rank-one decomposition in [AHZ11].

Remark 7. *In Problem* (2.23)*, if* $\boldsymbol{C}(\boldsymbol{w}_2)$ *is a positive/negative semidefinite matrix and also without full rank,* $\boldsymbol{w}_1$ *should and must be in the null space of* $\boldsymbol{C}(\boldsymbol{w}_2)$ *to satisfy* $\boldsymbol{w}_1^H \boldsymbol{C}(\boldsymbol{w}_2)\boldsymbol{w}_1 = 0$*. According to Proof 2.7.3,* $\boldsymbol{w}_1$ *can be expressed by* $\boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H \boldsymbol{p}_1$ *where* $\boldsymbol{C}^\perp(\boldsymbol{w}_2) \in \mathbb{C}^{N_T \times (N_T - \mathrm{rank}(\boldsymbol{C}(\boldsymbol{w}_2)))}$ *consists of* $N_T - \mathrm{rank}(\boldsymbol{C}(\boldsymbol{w}_2))$ *eigenvectors corresponding to zero eigenvalues of* $\boldsymbol{C}(\boldsymbol{w}_2)$*. Then, Problem* (2.23) *is equivalent to*

$$\max_{\boldsymbol{p}_1 \in \mathbb{C}^{N_T \times 1}} \frac{\boldsymbol{p}_1^H \boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H \boldsymbol{A}_1(\boldsymbol{w}_2)\boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H \boldsymbol{p}_1}{\boldsymbol{p}_1^H \boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H \boldsymbol{p}_1}, \tag{2.25}$$

from which it is direct to derive the optimal solution

$$\boldsymbol{p}_1^{opt} = \boldsymbol{u}_1(\boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H \boldsymbol{A}_1(\boldsymbol{w}_2)\boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H, \boldsymbol{C}^\perp(\boldsymbol{w}_2)(\boldsymbol{C}^\perp(\boldsymbol{w}_2))^H). \tag{2.26}$$

[4]Note that the application of Theorem 2.2 and Theorem 2.3 in [AHZ11] needs $N_T \geq 3$. It implies that TX_k $\forall k$ should have $N_T \geq 3$ antennas in our scenario.

Therefore, in this case, the optimal solution to Problem (2.23) *can be solved in closed-form, i.e.,* $\boldsymbol{w}_1 = \overrightarrow{\boldsymbol{C}^{\perp}(\boldsymbol{w}_2)(\boldsymbol{C}^{\perp}(\boldsymbol{w}_2))^H\boldsymbol{p}_1^{opt}}$.

C. Optimization of $\boldsymbol{w}_2$

For a given *feasible* $\boldsymbol{w}_1$, Problem (2.20) becomes another optimization problem w.r.t. a single beamforming vector $\boldsymbol{w}_2$. Maximization of the objective function is

$$\max_{\boldsymbol{w}_2\in\mathcal{W}_{\mathcal{FP}}} \boldsymbol{w}_1^H\boldsymbol{H}_{11}^H\left(\sigma_1^2\boldsymbol{I}+\boldsymbol{H}_{12}\boldsymbol{w}_2\boldsymbol{w}_2^H\boldsymbol{H}_{12}^H\right)^{-1}\boldsymbol{H}_{11}\boldsymbol{w}_1 \tag{2.27a}$$

$$\Leftrightarrow \max_{\boldsymbol{w}_2\in\mathcal{W}_{\mathcal{FP}}} \boldsymbol{w}_1^H\boldsymbol{H}_{11}^H\boldsymbol{H}_{11}\boldsymbol{w}_1 - \frac{|\boldsymbol{w}_1^H\boldsymbol{H}_{11}^H\boldsymbol{H}_{12}\boldsymbol{w}_2|^2}{\sigma_1^2+\boldsymbol{w}_2^H\boldsymbol{H}_{12}^H\boldsymbol{H}_{12}\boldsymbol{w}_2} \tag{2.27b}$$

$$\Leftrightarrow \min_{\boldsymbol{w}_2\in\mathcal{W}_{\mathcal{FP}}} \frac{|\boldsymbol{w}_1^H\boldsymbol{H}_{11}^H\boldsymbol{H}_{12}\boldsymbol{w}_2|^2}{\sigma_1^2+\boldsymbol{w}_2^H\boldsymbol{H}_{12}^H\boldsymbol{H}_{12}\boldsymbol{w}_2} \tag{2.27c}$$

$$\Leftrightarrow \min_{\boldsymbol{w}_2\in\mathcal{W}_{\mathcal{FP}}} \frac{\boldsymbol{w}_2^H\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{w}_2}{\boldsymbol{w}_2^H\boldsymbol{C}_2\boldsymbol{w}_2} \tag{2.27d}$$

with the definitions of

$$\boldsymbol{C}_1(\boldsymbol{w}_1) \triangleq \boldsymbol{H}_{12}^H\boldsymbol{H}_{11}\boldsymbol{w}_1\boldsymbol{w}_1^H\boldsymbol{H}_{11}^H\boldsymbol{H}_{12} \tag{2.28a}$$

$$\boldsymbol{C}_2 \triangleq \sigma_1^2\boldsymbol{I}+\boldsymbol{H}_{12}^H\boldsymbol{H}_{12}. \tag{2.28b}$$

Then, $\boldsymbol{w}_2$ can be optimized by

$$\begin{cases} \min\limits_{\boldsymbol{w}_2\in\mathcal{W}_{\mathcal{FP}}} & \dfrac{\boldsymbol{w}_2^H\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{w}_2}{\boldsymbol{w}_2^H\boldsymbol{C}_2\boldsymbol{w}_2} \\ \text{s.t.} & \boldsymbol{w}_2^H\boldsymbol{A}_2(\boldsymbol{w}_1)\boldsymbol{w}_2 = \text{SINR}_2^{\star}. \end{cases} \tag{2.29}$$

Observe that Problem (2.29) is a *fractional* QCQP problem. The objective function is not even a quasi-convex function because both the nominator function and denominator function are convex. To deal with this problem, we transform it by the SDR to

$$\begin{cases} \min\limits_{\boldsymbol{W}_2\succeq\boldsymbol{0}} & \dfrac{\text{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{W}_2)}{\text{Tr}(\boldsymbol{C}_2\boldsymbol{W}_2)} \\ \text{s.t.} & \text{Tr}(\boldsymbol{A}_2(\boldsymbol{w}_1)\boldsymbol{W}_2) = \text{SINR}_2^{\star} \\ & \text{Tr}(\boldsymbol{W}_2) = 1, \end{cases} \tag{2.30}$$

which is still a non-convex problem. Fortunately, the fractional structure can be removed by a variation of the Charnes-Cooper variable transformation [?]. Define the transformed variable $\boldsymbol{Q} = s\boldsymbol{W}_2$ with $s = \frac{1}{\text{Tr}(\boldsymbol{C}_2\boldsymbol{W}_2)}$. Then, Problem (2.30) is equivalent to

$$\begin{cases} \min\limits_{\boldsymbol{Q},\, s} & \mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{Q}) \\ \text{s.t.} & \mathrm{Tr}(\boldsymbol{A}_2(\boldsymbol{w}_1)\boldsymbol{Q}) = s \cdot \mathrm{SINR}_2^\star \\ & \mathrm{Tr}(\boldsymbol{C}_2\boldsymbol{Q}) = 1,\ \mathrm{Tr}(\boldsymbol{Q}) = s \\ & \boldsymbol{Q} \succeq \boldsymbol{0},\ \dfrac{1}{\lambda_1(\boldsymbol{C}_2)} \le s \le \dfrac{1}{\lambda_{N_R}(\boldsymbol{C}_2)}. \end{cases} \tag{2.31}$$

which is a convex problem w.r.t. $\boldsymbol{Q}$ and s and solvable (see Proof 2.7.4). By a convex optimization toolbox, we can obtain the optimal solution $(\boldsymbol{Q}^\star, s^\star)$. Then, the optimal solution to Problem (2.30) can be easily obtained by $\boldsymbol{W}_2^\star = \frac{\boldsymbol{Q}^\star}{s^\star}$. Observe that Problem (2.29) is equivalent to a homogeneous QCQP with three constraints. Therefore, by the matrix rank-one decomposition method, an optimal rank-one solution $\boldsymbol{w}_2$ to Problem (2.29) can be extracted from $\boldsymbol{W}_2^\star$ when $\mathrm{rank}(\boldsymbol{W}_2^\star) > 1$.

Remark 8. *In Problem* (2.29)*, if* $\boldsymbol{D} \triangleq \boldsymbol{A}_2(\boldsymbol{w}_1) - \mathrm{SINR}_2^\star \boldsymbol{I}_{N_T}$ *is a positive/negative semidefinite matrix but also without full rank,* $\boldsymbol{w}_1$ *should and must be in the null space of* $\boldsymbol{D}$ *to satisfy* $\boldsymbol{w}_2^H \boldsymbol{D} \boldsymbol{w}_2 = 0$*. According to Proof 2.7.3,* $\boldsymbol{w}_2$ *can be expressed by* $\overrightarrow{\boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{p}_2}$ *where* $\boldsymbol{D}^\perp \in \mathbb{C}^{N_T \times (N_T - \mathrm{rank}(\boldsymbol{D}))}$ *consists of* $N_T - \mathrm{rank}(\boldsymbol{D})$ *eigenvectors corresponding to zero eigenvalues of* $\boldsymbol{D}$*. Then, Problem* (2.29) *is equivalent to*

$$\max_{\boldsymbol{p}_2 \in \mathbb{C}^{N_T}} \frac{\boldsymbol{p}_2^H \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{C}_1(\boldsymbol{w}_1) \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{p}_2}{\boldsymbol{p}_2^H \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{C}_2 \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{p}_2}, \tag{2.32}$$

from which it is easy to derive the optimal solution

$$\boldsymbol{p}_2^{opt} = \boldsymbol{u}_1(\boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{C}_1(\boldsymbol{w}_1) \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H}, \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{C}_2 \boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H}). \tag{2.33}$$

Therefore, the optimal solution to Problem (2.29) *is* $\boldsymbol{w}_2 = \overrightarrow{\boldsymbol{D}^\perp \boldsymbol{D}^{\perp,H} \boldsymbol{p}_2^{opt}}$.

D. Algorithm Description

A feasible initial $\boldsymbol{w}_2$ for optimization of TX_1 can be obtained as follows.

Proposition 4. *For a given* $\mathrm{SINR}_2^\star \in (2^{\underline{R}_2} - 1, 2^{\overline{R}_2} - 1)$*,* $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ *is always a feasible solution pair to Problem* (2.20) *if we choose* $\boldsymbol{w}_2$ *from the following set*

$$\mathcal{W}_{\mathcal{F}} \triangleq \left\{ \boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{FP}} : \boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22} \boldsymbol{w}_2 \ge \sigma_2^2 \mathrm{SINR}_2^\star,\ \lambda_1(\boldsymbol{C}(\boldsymbol{w}_2)) \cdot \lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2)) \le 0 \right\}, \tag{2.34}$$

where $\boldsymbol{C}(\boldsymbol{w}_2)$ *is defined in (2.22) and* $\boldsymbol{w}_1 \in \mathcal{W}_{\mathcal{FP}}$*.*

Proof. Refer to Proof 2.7.5. □

With a fixed $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{F}}$ in Problem (2.20), Problem (2.23) always has at least one solution $\boldsymbol{w}_1$ in $\mathcal{W}_{\mathcal{FP}}$. This guarantees that the proposed alternating optimization works.

The proposed alternating optimization algorithm for the two-user MIMO IC with any initial $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{F}}$ is described in pseudo-code as Algorithm 1:

Algorithm 1 Single constraint rate maximization algorithm for two-user case

Input: $\boldsymbol{w}_2^{Alt}$, $\boldsymbol{w}_2^{Ego}$, $\forall R_2^\star \in (\underline{R}_2, \overline{R}_2)$, and convergence threshold ϵ_{th}.

Output: a convergent point $(R_1^{(i)}, R_2^\star)$ associated with $(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)})$.

begin
- Initialization: set a feasible $\boldsymbol{w}_2^{(0)} \in \mathcal{W}_{\mathcal{F}}$, $i = 0$.
- **repeat**
  - $i \leftarrow i + 1$;
  - Given $\boldsymbol{w}_2^{(i-1)}$, obtain an optimal $\boldsymbol{W}_1$ to Problem (2.20);
  - $\boldsymbol{w}_1^{(i)} \leftarrow \boldsymbol{W}_1$ (matrix rank-1 decomposition).
  - Given $\boldsymbol{w}_1^{(i)}$, obtain an optimal $(\boldsymbol{Q}, s)$ to Problem (2.31) and an optimal $\boldsymbol{W}_2 = \frac{\boldsymbol{Q}}{s}$ to Problem (2.30);
  - $\boldsymbol{w}_2^{(i)} \leftarrow \boldsymbol{W}_2$ (matrix rank-1 decomposition).
  - Compute $R_1^{(i)} = \log_2\left(1 + \boldsymbol{w}_1^{(i),H} \boldsymbol{A}_1(\boldsymbol{w}_2^{(i)}) \boldsymbol{w}_1^{(i)}\right)$.
- **until** $|R_1^{(i)} - R_1^{(i-1)}| \leq \epsilon_{th}$

In the two-user case, Algorithm 1 alternatively optimizes $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ until some convergence criterion is satisfied. The algorithm will be analyzed in detail in next section.

2.3.2 Algorithm Analysis

In this section, we analyze the proposed alternating optimization algorithm in the following aspects: convergence behavior, quality of the solution, implementation and the complexity.

A. Convergence Analysis:

Based on the results in Section 2.3.1, both Problem (2.23) and Problem (2.29) is solved optimally at each iteration. We will show that the sequence $\left\{\mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)})\right\}_{i=1}^{\infty}$ by Algorithm 1 *monotonically increases and converges*, i.e., $\mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \boldsymbol{w}_2^{(i+1)}) \geq \mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)}), \forall i$.

Denote the optimization of $\boldsymbol{w}_1$ and the optimization of $\boldsymbol{w}_2$ by the mapping functions $\boldsymbol{w}_1 = \Phi(\boldsymbol{w}_2)$ and $\boldsymbol{w}_2 = \Theta(\boldsymbol{w}_1)$, respectively. Then, the procedure of Algorithm 1 at the

$i+1$-th iteration is shown as

$$\mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)}) \xrightarrow{\boldsymbol{w}_1^{(i+1)}=\Phi\left(\boldsymbol{w}_2^{(i)}\right)} \mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \boldsymbol{w}_2^{(i)}) \tag{2.35}$$

$$\xrightarrow{\boldsymbol{w}_2^{(i+1)}=\Theta\left(\boldsymbol{w}_1^{(i+1)}\right)} \mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \boldsymbol{w}_2^{(i+1)}), \tag{2.36}$$

For $\boldsymbol{w}_1^{(i+1)} = \Phi(\boldsymbol{w}_2^{(i)})$, since the feasible solution set of $\boldsymbol{w}_1$ to Problem (2.23) includes $\boldsymbol{w}_1^{(i)}$ and additionally the global optimal solution to Problem (2.23) can be obtained by $\Phi(\cdot)$, it implies

$$\mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \boldsymbol{w}_2^{(i)}) = \mathrm{SINR}_1(\Phi\left(\boldsymbol{w}_2^{(i)}\right), \boldsymbol{w}_2^{(i)}) \geq \mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)}). \tag{2.37}$$

Based on $\boldsymbol{w}_2^{(i+1)} = \Theta(\boldsymbol{w}_1^{(i+1)})$ and the optimality of $\Theta(\cdot)$, it is also verified that

$$\mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \boldsymbol{w}_2^{(i+1)}) = \mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \Theta\left(\boldsymbol{w}_1^{(i+1)}\right)) \geq \mathrm{SINR}_1(\boldsymbol{w}_1^{(i+1)}, \boldsymbol{w}_2^{(i)}). \tag{2.38}$$

As a consequence, the sequence of $\left\{\mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)})\right\}_{i=1}^{\infty}$ *monotonically* increases as the iteration number i increases. In addition, since the sequence $\left\{\mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)})\right\}_{i=1}^{\infty}$ is upper-bounded by the single-user SINR, i.e., $\frac{\lambda_1(\boldsymbol{H}_{11}^H \boldsymbol{H}_{11})}{\sigma_1^2}$ in (2.12), the convergence of the sequence $\left\{\mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)})\right\}_{i=1}^{\infty}$, and thus the convergence of Algorithm 1 is guaranteed for any feasible initial point $\boldsymbol{w}_2^{(0)}$.

Since the *hard-coupled* two beamforming vectors exist not only in the objective but also in the constraints in Problem (2.20), the conventional convergence analysis for the block coordinate descent algorithm [Sol98] that requires that the constraints are *separable* among the variables is not applicable to our scenario. Therefore, it is unclear whether the proposed algorithm converges to a stationary point $\left(R_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)}), R_2(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)})\right)$ where $\boldsymbol{w}_1^{(i)}$ and $\boldsymbol{w}_2^{(i)}$ satisfy the Karush-Kuhn-Tucker (KKT) conditions of the original problem (2.20).

B. Quality of Solutions:

Due to the convergence of the proposed algorithm, the limit point of sequence of $\left\{\mathrm{SINR}_1(\boldsymbol{w}_1^{(\ell)}, \boldsymbol{w}_2^{(\ell)})\right\}_{\ell=0}^{\infty}$ for an arbitrary feasible initial $\boldsymbol{w}_2^{(0)}$ can be achieved by

$$\lim_{i\to\infty} \mathrm{SINR}_1(\boldsymbol{w}_1^{(i)}, \boldsymbol{w}_2^{(i)}) = \mathrm{SINR}_1\Big(\Phi\left(\Theta\left(\ldots\left(\Phi\left(\boldsymbol{w}_2^{(0)}\right)\right)\right)\right), \Theta\left(\ldots\left(\boldsymbol{w}_2^{(0)}\right)\right)\Big). \tag{2.39}$$

It implies that the performance of the alternating optimization algorithm might depend on the initial beamforming vector $\boldsymbol{w}_2^{(0)}$. We denote the global optimal solution to Problem (2.20) by $(\boldsymbol{w}_1^\star, \boldsymbol{w}_2^\star)$. Take an extreme example, if $\boldsymbol{w}_2^{(0)} = \boldsymbol{w}_2^\star$, we can obtain the global optimum $\boldsymbol{w}_1^\star = \Phi\left(\boldsymbol{w}_2^{(0)}\right)$ directly because of the optimality of $\Phi(\cdot)$. Therefore, a good initial beamforming vector could lead to high performance. However generally, it is difficult to find a good initial point *efficiently* for such a complex multiple vector-variables optimization problem. In order to improve the performance, a common way in previous literatures is to run the alternating optimization algorithm serval times with multiple random initializations and then select the one with the best performance. In the thesis, we desire to provide a scheme to generate a good initialization *efficiently*.

Inspired by the idea in [JLD08, HG10], we heuristically propose a scheme to design a transmit beamforming vector by balancing the "egoistic" and "altruistic" strategies as

$$\boldsymbol{w}_k(\xi_{k,1}, \xi_{k,2}) = \frac{\xi_{k,1}\boldsymbol{w}_i^{Ego} + \xi_{k,2}\boldsymbol{w}_i^{Alt}}{||\xi_{k,1}\boldsymbol{w}_i^{Ego} + \xi_{k,2}\boldsymbol{w}_i^{Alt}||}, \qquad k \in \{1, 2\}, \tag{2.40}$$

where $\xi_{k,1}$ and $\xi_{k,2}$ are complex-valued parameters satisfying $|\xi_{k,1}| + |\xi_{k,2}| = 1$. In the following, we give an explanation to show that this tradeoff scheme is reasonable. The trade-off is necessary to be Pareto-optimal for the two-user MISO IC [JLD08], and its similar form still provides a good performance in sum-rate maximization for the multi-user single-stream MIMO IC [HG10]. The following simulation results show that this characterization cannot exactly achieve the whole strict Pareto boundary for the two-user MIMO IC but still has a promising performance. In particular, the two ending point of strict Pareto boundary $E1$ and $E2$ can be achieved exactly by the strategies $(\boldsymbol{w}_1(1,0), \boldsymbol{w}_2(0,1))$ and $(\boldsymbol{w}_1(0,1), \boldsymbol{w}_2(1,0))$, respectively. Therefore, (2.40) is defined as an *approximate Pareto complex characterization (APCC)*.

If the initial strategy $(\boldsymbol{w}_1^{(0)}, \boldsymbol{w}_2^{(0)})$ corresponds to the bound of the region generated by the APCC or by random beamforming, then the proposed algorithm must improve (or at least keep) the bound by the APCC or by random beamforming. However, it is not efficient to find those beamforming vector pairs achieving their bounds. Therefore, there is no guarantee to say that the proposed algorithm with *only one initial beamforming vector* always achieves an operating point outer than the bounds by the APCC and by random beamforming, but its performance will be further improved with high probability as the number of initial beamforming vectors increases (i.e., multiple initializations scheme). By the APCC, the $2N_T$-dimensional real space of each complex beamforming vector $\boldsymbol{w}_k$ is *approximately* reduced to 3-dimensional real space (i.e., $|\xi_{k,1}|$, $|\xi_{k,2}|$ and the difference of the phases of $\xi_{k,1}$ and $\xi_{k,2}$ in (2.40) without significant performance loss. Therefore, the proposed algorithm with the proposed initialization by the APCC is *more efficient or likely* to achieve a good performance compared with a random initialization.

To further enhance the efficiency of initialization by the APCC, a real constant parameter instead of two complex parameters is employed to reduce (2.40) to an *approximate*

Pareto real characterization (APRC)

$$\boldsymbol{w}_k = \frac{\zeta \boldsymbol{w}_k^{Ego} + (1-\zeta)\boldsymbol{w}_k^{Alt}}{||\zeta \boldsymbol{w}_k^{Ego} + (1-\zeta)\boldsymbol{w}_k^{Alt}||}, \quad k \in \{1,2\}, \tag{2.41}$$

where $\zeta = \frac{R_2^\star - \underline{R}_2}{\overline{R}_2 - \underline{R}_2}$ is a scalar for a given $R_2^\star$. $\zeta \in [0,1]$ approximately denotes the proportion/weight of the "egoistic" strategy. However, by (2.41), if $\boldsymbol{w}_2 \notin \mathcal{W}_{\mathcal{F}}$, we reset $\zeta \in \frac{R_2^\star - \underline{R}_2}{\overline{R}_2 - \underline{R}_2} + [-\nu, \nu]$ with $0 < \nu \leq \min\left\{\frac{R_2^\star - \underline{R}_2}{\overline{R}_2 - \underline{R}_2}, \frac{\overline{R}_2 - R_2^\star}{\overline{R}_2 - \underline{R}_2}\right\}$ until $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{F}}$. If $\boldsymbol{w}_2$ is still infeasible, we choose a randomly generated $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{F}}$ directly.

The APRC corresponds to a curve w.r.t. only a single variable ζ. Thus, our proposed algorithm with the initialization by the APRC always outperforms the bound by APRC, which can serve as a lower bound of the proposed algorithm.

C. Implementation:

A multi-user MIMO IC enables to model both the downlink and uplink in a multi-cell environment. According to the steps in Figure 1.5, several potential implementation scenarios are provided and analyzed.

- **Assumptions:** we assume perfect local CSIR (and also perfect channel reciprocity for TDD systems), perfect feedback and perfect backhaul connection between BSs.

1. ***FDD Downlink Implementation***
   - *Channel Estimation:* each BS k sends a pilot sequence, and each MS k perfectly estimates its local CSI (downlink channels $\{\boldsymbol{H}_{k\ell}\}_{\ell=1}^2$);
   - *Channel Feedback:* each MS k feeds back the local CSI to the BSs via a error-free feedback link;
   - *Optimization of both $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$:* for the optimization of $\boldsymbol{w}_1$ at BS 1, BS 1 solves (2.23) based on the updated $\boldsymbol{w}_2$ from BS 2. Similarly, BS 2 optimizes $\boldsymbol{w}_2$ by solving Problem (2.29) based on the updated $\boldsymbol{w}_1$ from BS 1. The algorithm will be terminated once BS 1 finds that convergence criterion is satisfied;
   - *Data Transmission:* BS 1 and BS 2 transmits a single symbol by the beamforming vector $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, respectively. Each MS employs the MMSE receive beamforming to decode the data.

- Total Overhead:

- *Feedback Overhead:* four $N_R \times N_T$ complex matrices $\{\{\boldsymbol{H}_{k,\ell}\}_{\ell=1}^2\}_{k=1}^2$ are needed to be fed back perfectly, i.e., $4N_RN_T$ cc, where cc denotes the unit of a complex coefficient;
- *Backhaul Overhead:* let us assume that the convergence of Algorithm 1 takes N_I iterations. Before the optimization, the two BSs need to exchange their local CSI, i.e., two $N_R \times N_T$ complex matrices for each BS. In the i-th iteration, the exchanged data are only two $N_T \times 1$ complex vectors $\boldsymbol{w}_1^{(i)}$ and $\boldsymbol{w}_2^{(i)}$. Therefore, the total backhaul overhead is $4N_RN_T + 2N_IN_T$ cc.

2. *FDD Uplink Implementation*

- *Channel Estimation:* each MS k sends a pilot sequence, and each BS k perfectly estimates its local CSI (downlink channels $\{\boldsymbol{H}_{k\ell}\}_{\ell=1}^2$);
- *Optimization of both $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$:* this procedure is the same as that in FDD downlink scenario;
- *Transmit Beamforming Vectors Feedback:* each BS k feeds back the transmit beamforming vector $\boldsymbol{w}_k$ to the MS k via a error-free feedback link;
- *Data Transmission:* MS 1 and MS 2 transmits a single symbol by the transmit beamforming vector $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, respectively. Each BS employs the MMSE receive beamforming to decode the data.

• Total Overhead:

- *Feedback Overhead:* two $N_T \times 1$ complex vectors $\{\boldsymbol{w}_k\}_{k=1}^2$ are needed to be fed back perfectly, i.e., $4N_T$ cc;
- *Backhaul Overhead:* the amount is the same as that in FDD downlink scenario, i.e., $4N_RN_T + 2N_IN_T$ cc.

3. *TDD Downlink Implementation*

- *Channel Estimation:* each MS k sends a pilot sequence. Each BS k perfectly estimates its local uplink CSI and further estimates the downlink channels based on the *channel reciprocity*;
- *Optimization of both $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$:* this procedure is the same as that in FDD downlink/uplink scenario;
- *Data Transmission:* this procedure is the same as that in FDD downlink scenario;

• Total Overhead:

- *Feedback Overhead:* the feedback is not needed;
- *Backhaul Overhead:* the amount is the same as that in FDD downlink/uplink scenario, i.e., $4N_R N_T + 2N_I N_T$ cc.

4. ***TDD Uplink Implementation***
   - *Channel Estimation:* each MS k sends a pilot sequence. Each BS k perfectly estimates its local uplink CSI;
   - *Optimization of both $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$:* this procedure is the same as that in FDD downlink/uplink scenario;
   - *Transmit Beamforming Vectors Feedback:* this procedure is the same as that in FDD uplink scenario;
   - *Data Transmission:* this procedure is the same as that in FDD uplink scenario;

- Total Overhead:
  - *Feedback Overhead:* the amount is the same as that in FDD uplink scenario, i.e., $4N_T$ cc;
  - *Backhaul Overhead:* the amount is the same as that in FDD downlink/uplink scenario, i.e., $4N_R N_T + 2N_I N_T$ cc.

In the above implementation scenarios, the optimization of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ is implemented in a distributed way. If there exists a central authority that enables to do the centralized optimization, a centralized implementation is done in the following steps: each BS forwards the local channel matrices to the central authority. Based on the collected global CSI, the optimization of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ is done at the central authority. Afterwards, the computed beamforming vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ will be broadcasted to all BSs.

D. Complexity Analysis:

For the proposed algorithm shown in Algorithm 1, each iteration involves solving two convex SDR problems –Problem (2.24) and Problem (2.31) and additional two matrix rank-one decompositions of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. In [LMS+10], it is shown that the complexity of solving the SDR is *polynomial* in the problem size (i.e., N_T) and the number of constraints (denoted by N_{con}), i,e., $\mathcal{O}(\max{(N_{con}, N_T)}^4 N_T^{1/2} \log(1/\epsilon_{th}))$ given a solution accuracy $\epsilon_{th} > 0$. In the thesis, we have $N_T \geq 3$, and $N_{con} = 2$ in Problem (2.24) and $N_{con} = 3$ in Problem (2.31). Thus, the complexity of solving Problem (2.24) and Problem (2.31) is $\mathcal{O}(N_T^{4.5} \log(1/\epsilon_{th}))$. In terms of the complexity of the matrix rank-one decomposition [PYCE10], the rank-one solution can be extracted in *polynomial-time* if $\text{rank}(\boldsymbol{W}_k) \geq 3$; If $\text{rank}(\boldsymbol{W}_k) \geq 2$, it is sufficient to seek for a rank-one solution to a sequence of linear matrix equations within a slightly expanded range space of $\boldsymbol{W}_k$. If $\text{rank}(\boldsymbol{W}_k) = 1$, only the EVD of $\boldsymbol{W}_k$ is needed.

Furthermore, the complexity is implied by the running time in numerical simulations. The average consuming time of an iteration of Algorithm 1 is 0.6180 seconds by running the MATLAB 7.10 on the computer with AMD Athlon(TM) 64 Processor 3200+, 2.01 GHZ and 2GB RAM. Additionally, the fast convergent behavior of the proposed algorithm is implied numerically (e.g., Figure 2.4(b) with 8.55 iterations on average and Figure 2.5(b) with 5.16 iterations on average). Therefore, the proposed alternating optimization algorithm has reasonable complexity.

2.3.3 Problem & Algorithm Extension: Multi-User Case

In this section, we extend the Algorithm 1 for the two-user case to the multi-user case.

A. Problem Formulation

Consider a K-user single-stream MIMO IC. With the MMSE receive filter, the achievable rate of the link $\text{TX}_k \mapsto \text{RX}_k$ is expressed as $R_k(\{\boldsymbol{w}\}_{\mathcal{K}}) = \log_2\left(1 + \text{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}})\right), \forall k \in \mathcal{K}$, where

$$\text{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}) = \boldsymbol{w}_k^H \underbrace{\boldsymbol{H}_{kk}^H (\sum_{\ell \neq k} \boldsymbol{H}_{k\ell} \boldsymbol{w}_\ell {\boldsymbol{w}_\ell}^H {\boldsymbol{H}_{k\ell}}^H + \sigma_k^2 \boldsymbol{I}_{N_R})^{-1} \boldsymbol{H}_{kk}}_{\triangleq \boldsymbol{A}_k(\boldsymbol{w}_{-k})} \boldsymbol{w}_k. \tag{2.42}$$

is the SINR expression of user k, and $\boldsymbol{w}_{-k}$ denotes $\{\boldsymbol{w}_\ell\}_{\mathcal{K}\backslash\{k\}}$.

Without loss of generality, the optimization problem (2.20) in the two-user case can be generalized to[5]

$$\begin{cases} \max\limits_{\{\boldsymbol{w}\}_{\mathcal{K}}} & \text{SINR}_1(\{\boldsymbol{w}\}_{\mathcal{K}}) \\ \text{s.t.} & \text{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}) = \text{SINR}_k^\star, \quad \forall k \in \mathcal{K}\backslash\{1\}. \\ & \boldsymbol{w}_k^H \boldsymbol{w}_k \leq 1, \quad \forall k \in \mathcal{K}, \end{cases} \tag{2.43}$$

where $\{\text{SINR}_k^\star\}_{\mathcal{K}\backslash\{1\}}$ are the selected *feasible* SINR constraint values. Problem (2.43) is a non-convex problem of $\{\boldsymbol{w}\}_{\mathcal{K}}$.

[5] For the multi-user case, an arbitrary Pareto-optimal operating point of the high-dimensional utility region can be achieved by maximizing one user's utility when the others are fixed.

B. Basics for Algorithm Extension

To extend the proposed Algorithm 1 for Problem (2.20) to Problem (2.43), by defining

$$\boldsymbol{D}(\boldsymbol{w}_{-k-k'}) \triangleq \sum_{\ell \neq k,k'} \boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\boldsymbol{w}_\ell^H\boldsymbol{H}_{k\ell}^H + \sigma_k^2\boldsymbol{I}_{N_R}, \tag{2.44}$$

we derive equivalent expressions of $\mathrm{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}})$ in (2.42) as

$$\begin{aligned}\mathrm{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}) &= \boldsymbol{w}_k^H\boldsymbol{H}_{kk}^H\left(\boldsymbol{D}(\boldsymbol{w}_{-k-k'}) + \boldsymbol{H}_{kk'}\boldsymbol{w}_{k'}\boldsymbol{w}_{k'}^H\boldsymbol{H}_{kk'}^H\right)^{-1}\boldsymbol{H}_{kk}\boldsymbol{w}_k \\ &\overset{(a)}{=} \boldsymbol{w}_k^H\boldsymbol{H}_{kk}^H\boldsymbol{D}(\boldsymbol{w}_{-k-k'})^{-1}\boldsymbol{H}_{kk}\boldsymbol{w}_k \\ &\quad - \frac{\boldsymbol{w}_{k'}^H\overbrace{\boldsymbol{H}_{kk'}^H\boldsymbol{D}(\boldsymbol{w}_{-k-k'})^{-1}\boldsymbol{H}_{kk}\boldsymbol{w}_k\boldsymbol{w}_k^H\boldsymbol{H}_{kk}^H\boldsymbol{D}(\boldsymbol{w}_{-k-k'})^{-1}\boldsymbol{H}_{kk'}}^{\triangleq \boldsymbol{F}_k(\boldsymbol{w}_{-k'})}\boldsymbol{w}_{k'}}{1 + \boldsymbol{w}_{k'}^H\underbrace{\boldsymbol{H}_{kk'}^H\boldsymbol{D}(\boldsymbol{w}_{-k-k'})^{-1}\boldsymbol{H}_{kk'}}_{\triangleq \boldsymbol{G}_k(\boldsymbol{w}_{-k'-k})}\boldsymbol{w}_{k'}}\end{aligned} \tag{2.45}$$

where the transformation (a) is based on the Sherman-Morrison Formula [Mey00] and $\boldsymbol{w}_{-k-k'} \triangleq \{\boldsymbol{w}_\ell\}_{\mathcal{K}\setminus\{k,k'\}}$.

Therefore, maximization of the objective function in Problem (2.43) w.r.t. different beamforming vectors is equivalent to

$$\max_{\boldsymbol{w}_1}\ \mathrm{SINR}_1(\{\boldsymbol{w}\}_{\mathcal{K}}) \Leftrightarrow \max_{\boldsymbol{w}_1}\ \boldsymbol{w}_1^H\boldsymbol{A}_1(\boldsymbol{w}_{-1})\boldsymbol{w}_1 \tag{2.46a}$$

$$\max_{\boldsymbol{w}_k}\ \mathrm{SINR}_1(\{\boldsymbol{w}\}_{\mathcal{K}}) \overset{(b)}{\Leftrightarrow} \min_{\boldsymbol{w}_k}\ \frac{\boldsymbol{w}_k^H\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{w}_k}{1 + \boldsymbol{w}_k^H\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{w}_k}\ \forall k \in \mathcal{K}\setminus\{1\}, \tag{2.46b}$$

and an individual SINR constraint is equivalent to

$$\mathrm{SINR}_k(\{\boldsymbol{w}\}_{\mathcal{K}}) = \mathrm{SINR}_k^\star \tag{2.47a}$$

$$\Leftrightarrow\ \boldsymbol{w}_k^H\boldsymbol{A}_k(\boldsymbol{w}_{-k})\boldsymbol{w}_k = \mathrm{SINR}_k^\star, \tag{2.47b}$$

$$\overset{(b)}{\Leftrightarrow}\ \boldsymbol{w}_{k'}^H\boldsymbol{E}_k(\boldsymbol{w}_{-k'})\boldsymbol{w}_{k'} = \gamma_k(\boldsymbol{w}_{-k'}),\ \forall k' \in \mathcal{K}\setminus\{k\} \tag{2.47c}$$

where $\boldsymbol{E}_k(\boldsymbol{w}_{-k'})$ and $\gamma_k(\boldsymbol{w}_{-k'})$ are defined as

$$\boldsymbol{E}_k(\boldsymbol{w}_{-k'}) \triangleq \boldsymbol{F}_k(\boldsymbol{w}_{-k'}) - \left(\boldsymbol{w}_k^H\boldsymbol{H}_{kk}^H\boldsymbol{D}(\boldsymbol{w}_{-k-k'})^{-1}\boldsymbol{H}_{kk}\boldsymbol{w}_k - \mathrm{SINR}_k^\star\right)\boldsymbol{G}_k(\boldsymbol{w}_{-k-k'}); \tag{2.48a}$$

$$\gamma_k(\boldsymbol{w}_{-k'}) \triangleq \boldsymbol{w}_k^H\boldsymbol{H}_{kk}^H\boldsymbol{D}(\boldsymbol{w}_{-k-k'})^{-1}\boldsymbol{H}_{kk}\boldsymbol{w}_k - \mathrm{SINR}_k^\star. \tag{2.48b}$$

The equivalence (b) in both (2.46b) and (2.47c) is based on (2.45).

C. Optimization of $\boldsymbol{w}_1$

Given the fixed $\boldsymbol{w}_{-1}$ and based on the equivalence results in (2.46) and (2.47), Problem (2.43) w.r.t. $\boldsymbol{w}_1$ is equivalent to

$$\begin{cases} \max\limits_{\boldsymbol{w}_1} & {\boldsymbol{w}_1}^H \boldsymbol{A}_1(\boldsymbol{w}_{-1})\boldsymbol{w}_1 \\ \text{s.t.} & {\boldsymbol{w}_1}^H \boldsymbol{E}_k(\boldsymbol{w}_{-1})\boldsymbol{w}_1 = \gamma_k(\boldsymbol{w}_{-1}),\ \forall k \in \mathcal{K}\backslash\{1\}. \\ & {\boldsymbol{w}_1}^H \boldsymbol{w}_1 \leq 1. \end{cases} \tag{2.49}$$

Observe that Problem (2.49) is a homogeneous QCQP. By the SDR, it is relaxed to

$$\begin{cases} \max\limits_{\boldsymbol{W}_1 \succeq \boldsymbol{0}} & \mathrm{Tr}(\boldsymbol{A}_1(\boldsymbol{w}_{-1})\boldsymbol{W}_1) \\ \text{s.t.} & \mathrm{Tr}(\boldsymbol{E}_k(\boldsymbol{w}_{-1})\boldsymbol{W}_1) = \gamma_k(\boldsymbol{w}_{-1}),\ \forall k \in \mathcal{K}\backslash\{1\} \\ & \mathrm{Tr}(\boldsymbol{W}_1) \leq 1 \end{cases} \tag{2.50}$$

where $\boldsymbol{W}_1 = \boldsymbol{w}_1\boldsymbol{w}_1^H$. Now, Problem (2.50) becomes a convex problem w.r.t. $\boldsymbol{W}_1$. The optimal $\boldsymbol{W}_1^\star$ to Problem (2.50) can be efficiently solved by a convex optimization toolbox.

If $\mathrm{rank}(\boldsymbol{W}_1^\star) = 1$, the optimal rank-one solution is $\boldsymbol{w}_1 = \boldsymbol{u}_1(\boldsymbol{W}_1^\star)$. Otherwise, we observe that Problem (2.49) as a homogeneous QCQP has K constraints, and thus an optimal $\boldsymbol{w}_1$ to Problem (2.49) can be reconstructed from $\boldsymbol{W}_1^\star$ for $K \leq 4$ by the matrix rank-one decomposition method in [AHZ11]. When $K \geq 5$, there exist several approaches (e.g., the eigenvector approximation method and the randomization method) to extract an *approximate* $\boldsymbol{w}_1$ from $\boldsymbol{W}_1^\star$. Although these approximation methods are not tight, intensive research show that they provide promising performance. The algorithm and accuracy analysis of the approximation methods are presented in [LMS+10] and the references therein.

D. Optimization of $\boldsymbol{w}_k,\ \forall k \neq 1$

Given the fixed $\boldsymbol{w}_{-k}$ and based on the equivalence results in (2.46) and (2.47), Problem (2.43) is equivalent to

$$\begin{cases} \min\limits_{\boldsymbol{w}_k} & \dfrac{{\boldsymbol{w}_k}^H \boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{w}_k}{1 + {\boldsymbol{w}_k}^H \boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{w}_k} \\ \text{s.t.} & {\boldsymbol{w}_k}^H \boldsymbol{A}_k(\boldsymbol{w}_{-k})\boldsymbol{w}_k = \mathrm{SINR}_k^\star \\ & {\boldsymbol{w}_k}^H \boldsymbol{E}_\ell(\boldsymbol{w}_{-k})\boldsymbol{w}_k = \gamma_\ell(\boldsymbol{w}_{-k}),\ \forall \ell \in \mathcal{K}\backslash\{1, k\} \\ & {\boldsymbol{w}_k}^H \boldsymbol{w}_k \leq 1. \end{cases} \tag{2.51}$$

Observe that the objective function belongs to a fractional program, while it is not a quasi-convex function due to the convexity of both the nominator function and the denominator function. To deal with this problem, we transform Problem (2.51) by the SDR to

$$\begin{cases} \min\limits_{\boldsymbol{W}_k \succeq \mathbf{0}} & \dfrac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k)}{1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k)} \\ \text{s.t.} & \mathrm{Tr}(\boldsymbol{A}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k) = \mathrm{SINR}_k^\star \\ & \mathrm{Tr}(\boldsymbol{E}_\ell(\boldsymbol{w}_{-k})\boldsymbol{W}_k) = \gamma_\ell(\boldsymbol{w}_{-k}), \forall \ell \in \mathcal{K}\backslash\{1,k\} \\ & \mathrm{Tr}(\boldsymbol{W}_k) \le 1. \end{cases} \tag{2.52}$$

where $\boldsymbol{W}_k = \boldsymbol{w}_k\boldsymbol{w}_k^H$. Ref. [MJ11a] proves that full power transmission is not always Pareto-optimal for the general multi-user MIMO/MISO IC (related to the number of users and transmit/receive antennas), which is different from the Pareto-optimal full power transmission in two-user MIMO/MISO IC. Thus, it does not always hold that

$$\frac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k)}{1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k)} \neq \frac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k)}{\mathrm{Tr}((\boldsymbol{I}+\boldsymbol{G}_k(\boldsymbol{w}_{-1-k}))\,\boldsymbol{W}_k)} \tag{2.53}$$

because $\mathrm{Tr}(\boldsymbol{W}_k)$ is not necessary to be 1. Thus, the Charnes-Cooper variable transformation used in the optimization Problem (2.30) to deal with the fractional structure cannot be applied to Problem (2.52). Instead, we observe that both the nominator function and the denominator function of the objective function are non-negative, differentiable and affine with $\boldsymbol{W}_k$. By introducing a real scalar parameter $\mu_k \ge 0$, the fractional programming problem (2.52) is equivalent to a parametric programming problem [Sch83]

$$\mathcal{F}(\mu_k) = \min_{\boldsymbol{W}_k \in \mathcal{S}_{\boldsymbol{W}_k}} \Big\{ \mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k) - \mu_k\left(1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k)\right) \Big\}, \tag{2.54}$$

where $\mathcal{S}_{\boldsymbol{W}_k}$ denotes the constraint set of $\boldsymbol{W}_k$ consisting of all the constraints in Problem (2.52), and it is obvious that $\mathcal{S}_{\boldsymbol{W}_k}$ is a convex set. Assume the optimal solution to Problem (2.52) is $\boldsymbol{W}_k^\star$. If $\mu_k^\star = \frac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k^\star)}{1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star)}$, it implies $\mathcal{F}(\mu_k^\star) = 0$. Thus, solving Problem (2.52) is equivalent to finding the root of the equation $\mathcal{F}(\mu_k) = 0$.

Given μ_k, (2.54) is a convex optimization problem w.r.t. $\boldsymbol{W}_k$, and its optimal solution $\boldsymbol{W}_k^\star(\mu_k)$ can be efficiently solved. Therefore, $\mathcal{F}(\mu_k) = 0$ can be further formulated as

$$\mathcal{F}(\mu_k^\star) = \mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k^\star) - \mu_k^\star \cdot (1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star)) = 0, \tag{2.55}$$

From [Din67], we know that $\mathcal{F}(\mu_k)$ is continuous, concave, strictly decreasing in μ_k and $\mathcal{F}(\mu_k) = 0$ has a unique solution. Additionally, we find that $-\left(1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star)\right)$ is a subgradient of $\mathcal{F}(\mu_k)$ for any μ_k. Thus, (2.55) can be solved by a generalized Newton method (also known as the Dinkelbach algorithm) described in Algorithm 2.

Algorithm 2 The Dinkelbach algorithm to solve Problem (2.54)

Input: an initial $\mu_k^{(0)}$ satisfying $\mathcal{F}(\mu_k^{(0)}) \leq 0$, convergence threshold ϵ_{th}.

Output: optimal $\mu_k^\star$ and $\boldsymbol{W}_k^\star$.

```
begin
    i = 0
    repeat
        Given μ_k^(i), solve optimal W_k^⋆(μ_k^(i)) to (2.54);
        μ_k^(i+1) = Tr(F_k(w_{-k}) W_k^⋆(μ_k^(i))) / (1+Tr(G_k(w_{-1-k}) W_k^⋆(μ_k^(i)))) 6;
        i ← i + 1.
    until |F(μ_k^(i))| ≤ ε_th
    μ_k^⋆ = μ_k^(i) and W_k^⋆ = W_k^⋆(μ_k^⋆).
```

begin

- $i = 0$
- **repeat**
  - Given $\mu_k^{(i)}$, solve optimal $\boldsymbol{W}_k^\star(\mu_k^{(i)})$ to (2.54);
  - $\mu_k^{(i+1)} = \frac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))}{1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))}$ [6];
  - $i \leftarrow i+1$.
- **until** $|\mathcal{F}(\mu_k^{(i)})| \leq \epsilon_{th}$
- $\mu_k^\star = \mu_k^{(i)}$ and $\boldsymbol{W}_k^\star = \boldsymbol{W}_k^\star(\mu_k^\star)$.

The algorithm as a Newton procedure to determine the root of the equation $\mathcal{F}(\mu_k^\star) = 0$ has superlinear convergence. By the Algorithm 2, the optimal $\mu_k^\star$ and $\boldsymbol{W}_k^\star(\mu_k^\star)$ to (2.54) is obtained. Equivalently, $\boldsymbol{W}_k^\star(\mu_k^\star)$ is an optimal solution to Problem (2.52) [Din67]. Then, a tight (for $K \leq 4$) or an approximate (for $K \geq 5$) solution $\boldsymbol{w}_k$ to Problem (2.52) can be extracted from $\boldsymbol{W}_k^\star$.

Above all, the proposed alternating optimization algorithm extended to solve Problem (2.43) is described as Algorithm 3.

The proposed alternating optimization algorithm can be extended to the K-user MIMO IC. For $K \leq 4$, it is the same as the two-user case that each optimal beamforming vector can be obtained in each iteration. For $K \geq 5$, each approximate optimal beamforming vector is obtained in each iteration. According to the convergence analysis of Algorithm 1, the convergence of the Algorithm 3 is also guaranteed because the objective function is monotonically increasing but upper bounded. In addition, the implementations and complexity of Algorithm 3 can be analyzed by directly following the same line of the analysis for the two-user case.

[6] This generalized Newton iterative update is from $\mu_k^{(i+1)} \triangleq \mu_k^{(i)} - \frac{\mathcal{F}(\mu_k^{(i)})}{-\left(1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))\right)} = \mu_k^{(i)} - \frac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))-\mu_k^{(i)}\left(1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))\right)}{-\left(1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))\right)} = \frac{\mathrm{Tr}(\boldsymbol{F}_k(\boldsymbol{w}_{-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))}{1+\mathrm{Tr}(\boldsymbol{G}_k(\boldsymbol{w}_{-1-k})\boldsymbol{W}_k^\star(\mu_k^{(i)}))}$.

[7] This condition is to make sure that a better (at least the same) solution to Problem (2.51) (only for $K \geq 5$) is always obtained in each iteration such that the objective function's non-decreasing convergence is guaranteed.

Algorithm 3 Single constraint rate maximization algorithm for K-user case

Input: feasible rate constraint values $\{R_k^{\star}\}_{\mathcal{K}\setminus\{1\}}$; convergence threshold ϵ_{th}

Output: a convergent point $(R_1^{(i)}, R_2^{\star}, ..., R_K^{\star})$ with $\{\boldsymbol{w}^{(i)}\}_{\mathcal{K}}$.

begin
  Initialization: set a feasible $\boldsymbol{w}_{-1}^{(0)}$, $i = 0$.
  repeat
    $i \leftarrow i + 1$.
    for $k = 1 \rightarrow K$ **do**
      Given $\boldsymbol{w}_{-k}^{(i-1)}$, obtain an optimal $\boldsymbol{W}_k$ to Problem (2.52);
      if $K \leq 4$ **then**
        $\boldsymbol{w}_k \leftarrow \boldsymbol{W}_k$ (exact reconstruction by matrix rank-1 decoposition);
      else
        $\boldsymbol{w}_k \leftarrow \boldsymbol{W}_k$ (approximate reconstruction by random initializations);
        if $\mathrm{SINR}_1\left(\boldsymbol{w}_k^{(i)}, \boldsymbol{w}_{-k}^{(i)}\right) < \mathrm{SINR}_1\left(\boldsymbol{w}_k^{(i-1)}, \boldsymbol{w}_{-k}^{(i)}\right)$[7] **then**
          $\boldsymbol{w}_k^{(i)} \leftarrow \boldsymbol{w}_k^{(i-1)}$;
    Compute $R_1^{(i)} = \log_2\left(1 + \mathrm{SINR}_1\left(\{\boldsymbol{w}^{(i)}\}_{\mathcal{K}}\right)\right)$.
  until $|R_1^{(i)} - R_1^{(i-1)}| \leq \epsilon_{th}$

2.4 Pareto Boundary Computation: Alternating Rate Profile Optimization

In this section, we propose another algorithm based on the *weighted Chebyshev optimization*, named by *alternating rate profile* algorithm. As shown in Figure 1.6, instead of maximizing a single rate along *a direction parallel to an axis* in the rate region by the proposed single constraint rate maximization algorithm, the proposed alternating rate profile method aims to find an intersection point between *a ray starting from the original of the rate region* and the Pareto boundary for the multi-user single-stream MIMO IC.

2.4.1 Problem Formulation

Motivated by the idea of the weighted Chebyshev optimization, the original multi-objective optimization problem (2.19) to achieve the Pareto boundary can be transformed into

$$\max_{\{\boldsymbol{g}\}_{\mathcal{K}},\{\boldsymbol{w}\}_{\mathcal{K}},R} \quad R \tag{2.56a}$$

$$\text{s.t.} \quad R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) \geq \alpha_k R, \quad k \in \mathcal{K}, \tag{2.56b}$$

$$\boldsymbol{w}_k \in \mathcal{W}, \boldsymbol{g}_k \in \mathcal{G}, \quad k \in \mathcal{K}. \tag{2.56c}$$

where the achievable rate $R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}})$ is defined in (2.4), and R denotes the weighted sum rate performance, but R is not necessary to be a maximum weighted sum rate. In (2.56), the rate profile $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$ satisfies $\alpha_k \geq 0, \forall k \in \mathcal{K}$ and $\sum_{k=1}^{K} \alpha_k = 1$. The rate profile defines the direction of a ray starting from the origin of the rate region, and the point of intersection of the ray and the Pareto boundary is a solution of Problem (2.56). Solving Problem (2.56) for all possible rate profiles, all Pareto-optimal points in $\mathcal{R}^\star$ are determined if Problem (2.56) can be optimally solved for each fixed rate profile.

The rate profile approach has been first proposed for the BC/MAC in [MZC06] and for the multi-user MISO IC in [ZC10], and the optimal solutions for both scenarios are obtained. However, Problem (2.56) is non-convex and NP-hard [LHD13], and hence no method is known that can attain its optimal solution efficiently. Motivated by the alternating optimization of different transmit beamforming vectors in [CJS13], the alternating rate profile optimization also adopts the idea of alternating optimization of multiple variables. However, instead of plugging the MMSE receive beamforming vectors into the SINR expression directly as (2.10), the transmit and receive beamforming vectors are alternatively optimized based on the SINR expression of (2.5).

We propose to decompose Problem (2.56) into two subproblems which are solved alternatively. The first problem optimizes the transmit beamforming vectors for fixed receive beamforming vectors, equivalent to a MISO IC. The second problem optimizes the receive beamforming vectors for fixed transmit beamforming vectors, equivalent to a SIMO IC. Next, we discuss and solve the two problems independently. The solutions of the two problems are used to construct the alternating algorithm.

2.4.2 Rate Profile Optimization in MISO IC

In this section, we assume the receive beamforming vectors are fixed. The considered MIMO setting reduces to a MISO IC, and its rate region is a subset of $\mathcal{R}$ in (2.1.2)

$$\mathcal{R}^{MISO} \triangleq \{\boldsymbol{r} \in \mathcal{R} : r_k = R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}), \boldsymbol{w}_k \in \mathcal{W}_k, k \in \mathcal{K}\}. \tag{2.57}$$

For fixed receive beamforming $\{\boldsymbol{g}\}_{\mathcal{K}}$, Problem (2.56) reduces to

$$\begin{aligned} \max_{\{\boldsymbol{w}\}_{\mathcal{K}}, R} \quad & R && (2.58a)\\ \text{s.t.} \quad & R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) \geq \alpha_k R, \quad k \in \mathcal{K}, && (2.58b)\\ & \boldsymbol{w}_k \in \mathcal{W}, \quad k \in \mathcal{K}. && (2.58c) \end{aligned}$$

It is shown in [QZLC11] that Problem (2.58) can be solved by a set of feasibility problems:

$$\begin{aligned} \text{find} \quad & \boldsymbol{w}_1, \cdots, \boldsymbol{w}_K && (2.59a)\\ \text{s.t.} \quad & R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) \geq \alpha_k t, \quad k \in \mathcal{K}, && (2.59b)\\ & \boldsymbol{w}_k \in \mathcal{W}, \quad k \in \mathcal{K}, && (2.59c) \end{aligned}$$

where the parameter $t > 0$ is updated based on a bisection method. In order to determine the feasibility, the problem in (2.59) is transformed in [QZLC11, Section II. D] to a second order cone programming (SOCP) and solved efficiently.

In Figure 2.3(a), solutions of Problem (2.58) are illustrated. For a rate profile ray $\boldsymbol{\alpha'}$ passing through the strict Pareto boundary $\mathcal{R}_{+}^{MISO,\star}$ (i.e., the red curve in Figure 2.3(a)), the solution of Problem (2.58) achieves a unique point on the strict Pareto boundary on the rate profile $\boldsymbol{\alpha'}$. If the rate profile ray does not pass through $\mathcal{R}_{+}^{MISO,\star}$ but the non-strict Pareto boundary as the rate profile $\boldsymbol{\alpha}$, the multiple solutions for Problem (2.58) exist corresponding to the points on green line.

2.4.3 Rate Profile Optimization in SIMO IC

In this section, we assume the transmit beamforming vectors $\{\boldsymbol{w}\}_{\mathcal{K}}$ are fixed. The setting corresponds to a SIMO IC. The rate region in the SIMO setting is a subset of the rate region $\mathcal{R}$ in (2.1.2) and has the following property.

Proposition 5. *The rate region of a K-user SIMO IC with fixed transmit beamforming is a K-dimensional box:*

$$\mathcal{R}^{SIMO} \triangleq \{\boldsymbol{r} \in \mathcal{R} : r_k \leq R_k(\boldsymbol{g}_k^\star, \{\boldsymbol{w}\}_{\mathcal{K}}), k \in \mathcal{K}\}, \tag{2.60}$$

where $\boldsymbol{g}_k^\star$ is a MMSE receive beamforming vector

$$\boldsymbol{g}_k^\star = \frac{(\sigma_k^2 \boldsymbol{I}_{N_R} + \sum_{\ell \neq k} \boldsymbol{H}_{k\ell} \boldsymbol{w}_\ell \boldsymbol{w}_\ell^H \boldsymbol{H}_{k\ell}^H)^{-1} \boldsymbol{H}_{kk} \boldsymbol{w}_k}{||(\sigma_k^2 \boldsymbol{I}_{N_R} + \sum_{\ell \neq k} \boldsymbol{H}_{k\ell} \boldsymbol{w}_\ell \boldsymbol{w}_\ell^H \boldsymbol{H}_{k\ell}^H)^{-1} \boldsymbol{H}_{kk} \boldsymbol{w}_k||}. \tag{2.61}$$

Proof. Refer to Proof 2.7.6. □

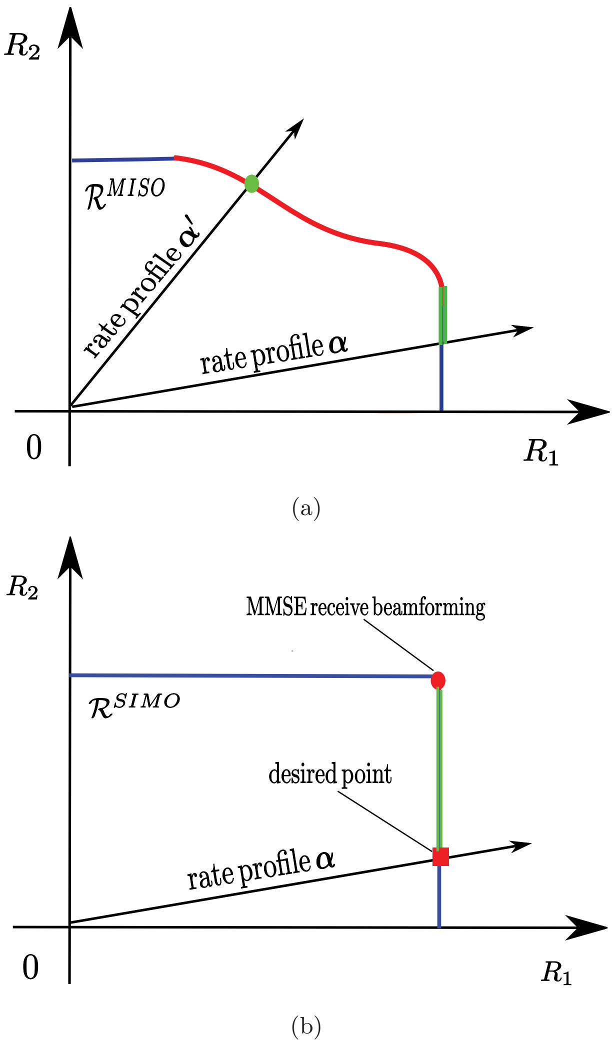


Figure 2.3: Illustration of the solutions of rate profile optimization: (a) MISO IC rate profile; (b) SIMO IC rate profile.

In Figure 2.3(b), an illustration of a two-user SIMO rate region is given. A single strict Pareto-optimal point exists corresponding to joint MMSE receive beamforming in (2.61). The rate profile optimization for fixed transmit beamforming vectors is formulated as:

$$\max_{\{\boldsymbol{g}\}_{\mathcal{K}},R} \quad R \tag{2.62a}$$

$$\text{s.t.} \quad R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) \geq \alpha_k R, \quad k \in \mathcal{K}, \tag{2.62b}$$

$$\boldsymbol{g}_k \in \mathcal{G}, \quad k \in \mathcal{K}, \tag{2.62c}$$

The receive beamforming vector that optimize (2.62) is not necessarily unique. In Figure 2.3(b), an illustration of the set of points that solve (2.62) are on the green line. One solution of (2.62) is by joint MMSE beamforming. Another special solution corresponds to the intersection point between the rate profile ray and the Pareto boundary. Note that rate profile optimization for the SIMO IC has been considered in [LZC12, Section IV. B] and [LHD13]. In comparison, we do not optimize the transmission power but only the receive beamforming by a different approach.

In order to attain the desired point on the rate profile ray, the increasing direction of R in Problem (2.62) is required to be kept. Thus, we need to solve the following problem

$$\begin{aligned} \text{find} \quad & \boldsymbol{g}_1, \cdots, \boldsymbol{g}_K && (2.63a)\\ \text{s.t.} \quad & R_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) = \alpha_k t, \quad k \in \mathcal{K}, && (2.63b)\\ & \boldsymbol{g}_k \in \mathcal{G}, \quad k \in \mathcal{K}, && (2.63c) \end{aligned}$$

where the equalities in (2.63b) guarantee the increasing direction always along the rate profile. Problem (2.63) becomes a *feasibility-checking* problems by updating the $t \geq 0$ based on a bisection method. In order to solve Problem (2.63), we reformulate the rate constraints (2.63b) to

$$\boldsymbol{g}_k^H \boldsymbol{Q}_k(t) \boldsymbol{g}_k = 0, k \in \mathcal{K}, \tag{2.64}$$

where $\boldsymbol{Q}_k(t)$ is a Hermitian matrix with the definition of

$$\boldsymbol{Q}_k(t) = \boldsymbol{H}_{kk}\boldsymbol{w}_k\boldsymbol{w}_k^H\boldsymbol{H}_{kk}^H - (2^{\alpha_k t} - 1)(\sigma_k^2 \boldsymbol{I}_{N_R} + \sum_{\ell \neq k} \boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell \boldsymbol{\ell}^H \boldsymbol{H}_{k\ell}^H). \tag{2.65}$$

The problem in (2.63) with (2.62b) replaced by (2.64) is called *inverse field of values problem* [Car09]. In order to check the feasibility of (2.63) for a chosen t, it suffices to test whether 0 lies between the smallest and largest eigenvalues of $\boldsymbol{Q}_k(t)$ in (2.65), i.e., 0 is in the field of values [HJ85] of $\boldsymbol{Q}_k(t)$. After the convergence of the bisection method which determines the optimal t, each vector $\boldsymbol{g}_k$ is determined by the algorithm from [Car09] requiring five EVDs [Car09, Section 5].

2.4.4 Algorithm Analysis

In this section, the analysis of the proposed alternating rate profile algorithm is provided.

A. Algorithm Description

The alternating rate profile algorithm is described as follows.

The measure $R^{(i)}$ at the i-th iteration is achieved progress from the origin along the rate profile ray. In each iteration i, an improvement $R^{(i)} - R^{(i-1)} \geq \epsilon_{th}$ must be achieved.

Algorithm 4 K-user alternating rate profile algorithm

Input: a rate profile $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$ and convergence threshold ϵ_{th}.

Output: the solution of Problem (2.56), i.e., $\{\boldsymbol{w}^{(i)}\}_{\mathcal{K}}, \{\boldsymbol{g}^{(i)}\}_{\mathcal{K}}$.

```
begin
    Initialization: set i = 0; choose a random {g^(0)}_K ∈ G.
    repeat
        i ← i + 1.
        Given {g^(i-1)}_K, solve Problem (2.58) to get ({w^(i)}_K, R_-^(i));
        Given {w^(i)}_K, solve Problem (2.63) to get ({g^(i)}_K, R^(i));
    until |R^(i) − R^(i-1)| ≤ ε_th
```

B. Convergence & Performance

The convergence of Algorithm 4 is proved as follows.

Proposition 6. *The alternating rate profile optimization in Algorithm 4 always converges to a stationary point of Problem* (2.56).

Proof. Refer to Proof 2.7.7. □

The solution by Algorithm 4 is not necessary global optimal. The performance of an alternating optimization algorithm depends on initializations and any normalized vector is feasible to be initial receive beamforming vectors in Algorithm 4. In order to gain a good performance, we provide an initialization method: *let the largest left singular vector of $\boldsymbol{H}_{kk}$ be the initial receive beamforming vector $\boldsymbol{g}_k^{(0)}$* (same as the simple receiver case in [BBO12]). The algorithm in [BBO12] is always with the same fixed receiver. Therefore, the proposed Algorithm 4 always outperforms the simple receiver algorithm in [BBO12] by iteratively updating the receive beamforming vectors.

C. Implementations

The proposed alternating rate profile algorithm requires a centralized implementation. First, each BS obtains its local CSI and then reports them to a central controller, e.g., a central authority or a powerful BS. Based on the collected global CSI and pre-defined rate profiles, the central controller optimizes $\{\boldsymbol{w}\}_{\mathcal{K}}$ and $\{\boldsymbol{g}\}_{\mathcal{K}}$ by running Algorithm 4. Afterwards, the central controller broadcasts the optimized transmit/receive strategies to BSs. Each BS feeds back the transmit (receive) strategies to the corresponding MSs for the uplink (downlink) transmission.

2.5 Illustrations & Discussions

In this section, the numerical results and performance analysis are presented for the proposed single constraint rate maximization algorithm and alternating rate profile optimization algorithm, separately. Then, the performance of both the proposed alternating optimization algorithms are compared.

In the following simulations, the transmit power budget is normalized as 1 for each transmitter, and the receive noise power $\sigma_k^2 = 10^{-\frac{\text{SNR}}{10}}$ where SNR refers to the transmit SNR in dB. Rayleigh fading channel model.

2.5.1 Numerical Results: Single Constraint Rate Maximization

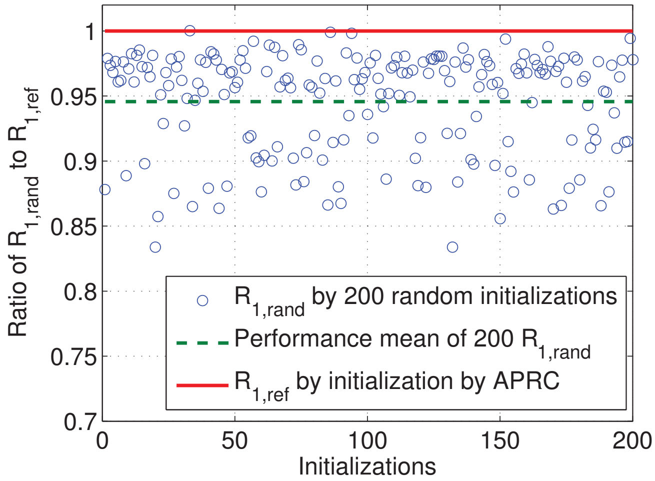
In this section, we show and discuss the numerical results of the proposed single constraint rate maximization algorithm in Algorithm 1 and Algorithm 3.

A. Convergence & Initialization

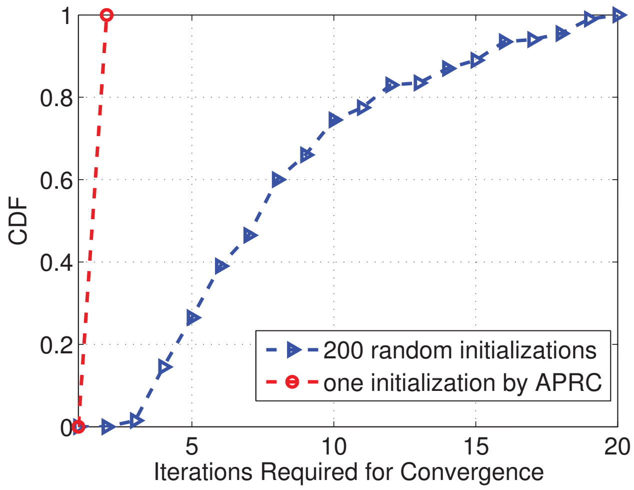
In Figure 2.4 and Figure 2.5, we illustrate the convergence speed of the alternating optimization algorithm and the effectiveness of the initialization by the APRC in (2.41) for a two-user MIMO IC. The simulation results are achieved by running Algorithm 1 with the APRC initialization and 200 random initializations[8]. The convergence threshold is $\epsilon_{th} = 10^{-3}$ and the transmit SNR is 10 dB.

In Figure 2.4, we set $R_2^{\star} = \underline{R}_2 + \frac{2}{19} \cdot (\overline{R}_2 - \underline{R}_2) = 5.6398$ bpcu, which corresponds to a strict Pareto-optimal point close to the ending point $(\overline{R}_1, \underline{R}_2)$. Figure 2.4(a) shows that the achieved value of R_1 with the APRC initialization nearly always outperforms that with 200 random initializations. Also observe that the proposed alternating algorithm even by a random initialization could also achieve a good performance with a high probability – the performance ratio mean is close to 95%, and a low variance – the performance ratios vary between 85% and 100%. Figure 2.4(b) shows the cumulative distribution function (CDF) w.r.t. the number of iterations required for convergence of Algorithm 1. It implies that the proposed algorithm generally converges fast, e.g., the CDF of 200 initializations reaches to 1 by 20 iterations. With the proposed APRC initialization, the algorithm converges much faster with only 2 iterations. In Figure 2.5, we set $R_2^{\star} = \underline{R}_2 + \frac{11}{19} \cdot \left(\overline{R}_2 - \underline{R}_2\right) = 6.2898$ bpcu, which corresponds to a strict Pareto-optimal point at the middle part of the strict Pareto boundary. Similarly, Figure 2.5(a) and Figure 2.5(b) also show that random initializations and especially the APRC initialization achieve a promising performance with a fast convergence speed. Therefore,

[8]For each random initialization, a randomly generated feasible normalized vector is set as $\boldsymbol{w}_2^{(0)}$.



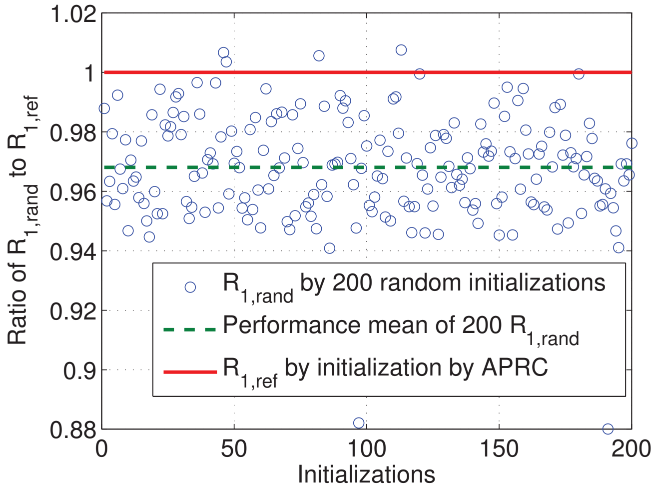
(a)



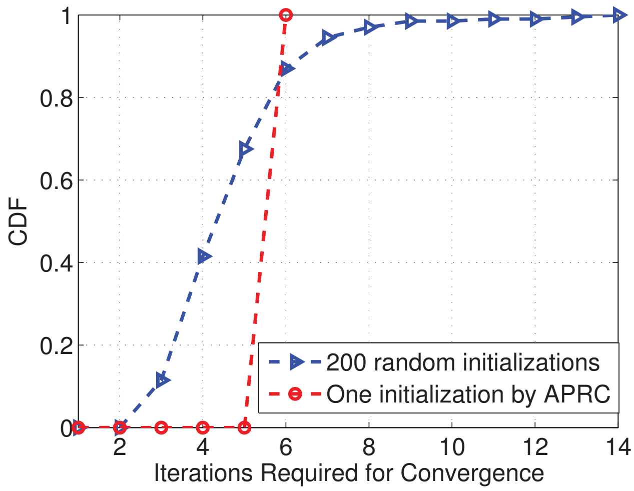
(b)

Figure 2.4: Convergence performance comparison of Algorithm 1 with different initializations at $R_2^\star = 5.6398$ bpcu: (a) Achievable rate comparison; (b) Convergence speed.

numerical results imply that the APRC is a good choice for the initialization compared with random initializations, and also that the proposed single constraint rate maximization algorithm usually has a *stable* convergence performance over initializations.



(a)



(b)

Figure 2.5: Convergence performance comparison of Algorithm 1 with different initializations at $R_2^\star = 6.2898$ bpcu: (a) Achievable rate comparison; (b) Convergence speed.

B. Achievable Rate Region Comparison

For the same channel data, we compare the proposed Algorithm 1 with some previous algorithms in the achievable rate region performance in Figure 2.6. Some benchmarks are listed as follows:

- "SCRM_1" denotes the result by Algorithm 1 with the APRC initialization and "SCRM_1+9" denotes the best result selected from the achievable rate by Algorithm 1 with the APRC initialization and nine random initializations;

- "WMMSE_1" ("WMMSE_10") denotes the result by the WMMSE sum-rate maximization algorithm in [SRLH11] with the APRC initialization (10 random initializations), where weighted sum rate is expressed as $w \cdot R_1 + (1-w) \cdot R_2$. with the weights ws in [0.05 : 0.05 : 0.95];

- "RB_10mil" denotes means the outermost boundary of the achievable rate region by 10 million random normalized transmit beamforming vectors and each receiver is the MMSE filter;

- "SR" denotes the outermost boundary of the achievable rate region achieved by the simpler receiver algorithm in [BBO12] where each receive beamforming is fixed as the largest left singular vector of the direct channel matrix;

- "APCC" denotes the outermost boundary of the achieved region by the APCC in (2.40) by 3-dimensional grid search of the three parameters;

- "APRC" denotes the curve by the APRC in (2.41) with $\zeta = \frac{n-1}{N},\ n = 1, ..., N+1$ where $N = 100$;

- "ZF points" denotes two outmost ZF points achieved by (2.18).

The SINR target in Problem (2.20) is based on the samples $\text{SINR}_2^\star = 2^{R_2^\star - 1}$s where $R_2^\star = \underline{R}_2 + \frac{n}{50} \cdot (\overline{R}_2 - \underline{R}_2),\ n = 1, 2, ..., 49$. We find that the curve "SCRM_1+9" achieves an outer boundary than the "RB_10mil" and enables to serve as a good inner bound of the Pareto boundary. Furthermore, we find that the "SCRM_1" also has a good performance with only the APRC initialization. The proposed algorithms "SCRM_1+9", "SCRM_1" and "WMMSE_10" yield a similar performance at convex parts of boundary and outperform others under the same accuracy for convergence. However, since the weighted sum maximization method cannot achieve the non-convex boundary, and even the achieved points on the convex boundary are still unevenly distributed. That is why "WMMSE_10" based on a weighted sum maximization does not achieve the part between the operating points "P1" and "P2" in Figure 2.6.

To further evaluate the performance of "WMMSE_10" and "SCRM_1+9" on illustrating the Pareto boundary, Algorithm 1 for another set of the channel data is simulated and shown in Figure 2.7. Even with fine weights ws in [0.05 : 0.005 : 0.95], we observe that there still exists a large jump between the points "P3" and "P4" by "WMMSE_10" so that it cannot effectively illustrate the rate region effectively.

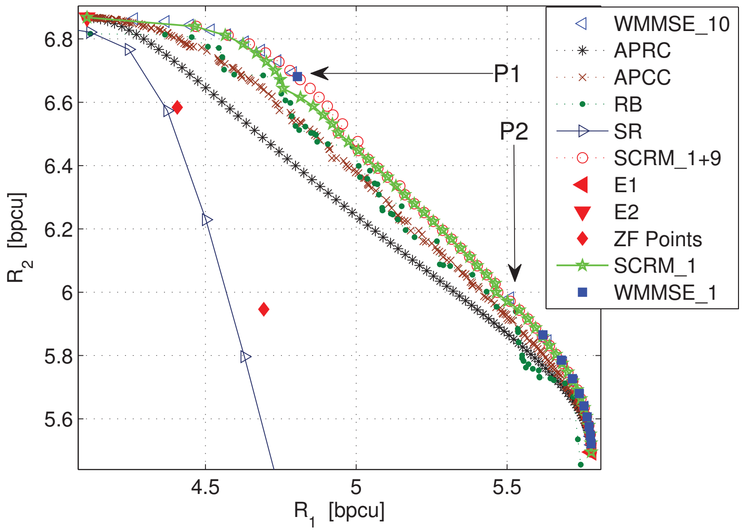


Figure 2.6: Channel realization example 1: Two-dimensional achievable rate region comparison with SNR=10 dB and $N_T = 3, N_R = 2$.

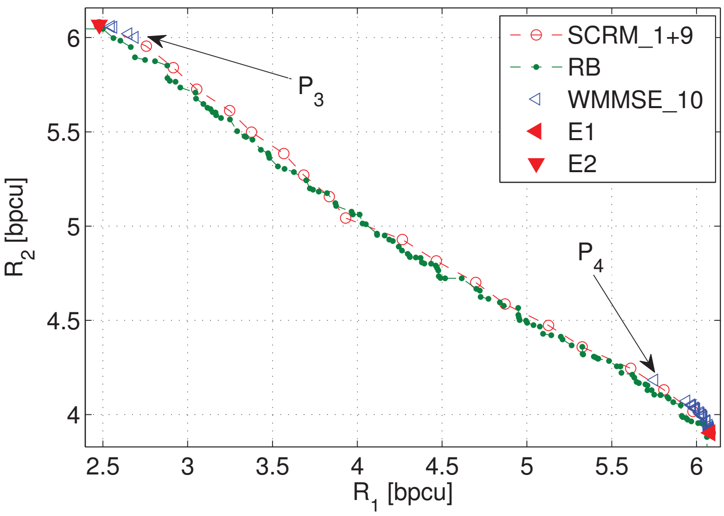


Figure 2.7: Channel realization example 2: Two-dimensional achievable region comparison with SNR=10 dB and $N_T = 3, N_R = 2$.

Based on only the result by "WMMSE_10", we do not know what the part between the operating points "P1" and "P2" in Figure 2.6 and the part between "P1" and "P2" in Figure 2.7 look like. If they are concave, the degree of concavity is still unknown. For our proposed algorithm, although its curve denoted by "SCRM_1+9" is not guaranteed to be the exact Pareto boundary, it has a even better performance than the "RB_10mil" curves in Figure 2.6 and Figure 2.7). Thus, it is able to serve as a more reasonable/complete inner bound of the whole strict Pareto boundary, especially the non-convex part. This is a main advantage of the proposed algorithm compared with the weighted sum maximization algorithms.

C. Illustration of the Multi-User Case

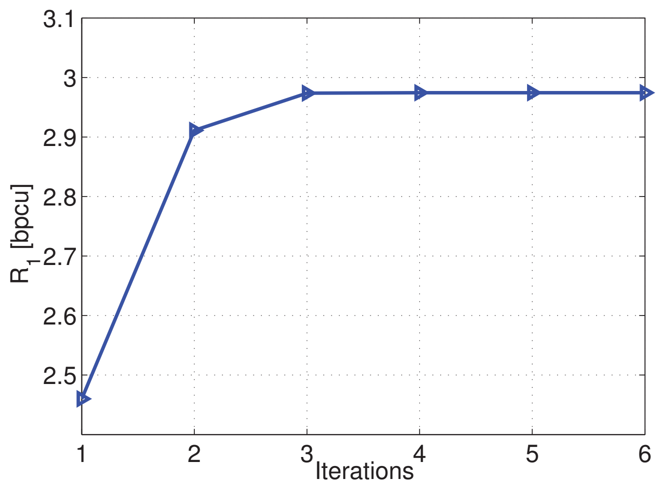


Figure 2.8: Convergence behavior of Algorithm 3 for a three-user MIMO IC with SNR=0 dB and $N_T = 3, N_R = 2$.

In order to evaluate the performance of Algorithm 3, a three-user MIMO IC example is simulated where SNR=0 dB and a convergence threshold $\epsilon_{th} = 10^{-4}$. In Figure 2.8, a convergence example for maximization of R_1 with $(R_3, R_2) = (0.2700, 2.3720)$ bpcu. It implies that the proposed algorithm still has a fast convergence behavior in the multi-user case. In Figure 2.9, a three-dimensional achievable rate region is plotted after computing the values of R_1s for the 65 samples of (R_3, R_2) and interpolation.

2.5.2 Numerical Results: Alternating Rate Profile Optimization

In this section, we show and discuss the numerical results of the proposed alternating rate profile optimization algorithm Algorithm 4.

In Figure 2.10, a two-dimensional achievable rate region is plotted. The cloud of points in Figure 2.10 corresponds to the set of operating points achieved by 1 million random beamforming realizations. By Algorithm 4, we are able to plot the operating

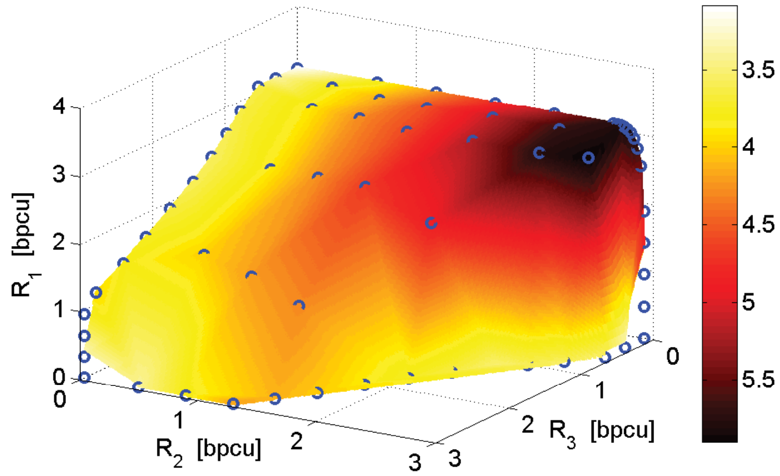


Figure 2.9: Three-dimensional achievable rate region with SNR=0 dB and $N_T = 3, N_R = 2$. The color-bar denotes the sum rate performance.

points bounding the two-dimensional achievable rate region for 50 different rate profile samples. In Figure 2.10, the points marked with cross correspond to the WMMSE algorithm in [SRLH11]. It can be observed that most of points on the strict Pareto boundary cannot be achieved because of its non-convexity.

For a selected rate profile $\boldsymbol{\alpha} = (0.79, 0.21)$ passing the non-strict Pareto boundary, the proposed alternating rate profile algorithm is compared with the following two benchmarks:

- "max-min by MMSE" denotes the max-min algorithm [LDL13, Algorithm EC-CAA], which alternatively optimizes the transmit beamforming by the MISO rate profile and employs the MMSE receive beamforming;
- "max-min by WMMSE" denotes the max-min algorithm in [RHL13], which reformulates the achievable rate expression based on the idea of WMMSE [CACC08] and the multiple variables are optimized by the alternating optimization algorithm.

In Figure 2.10, the solutions of the transmit and receive beamforming optimization problems are plotted during the procedure of the alternating optimization. The performance improvement in each iteration can be observed and the alternating optimization terminates at a convergent point. In comparison to the "max-min by MMSE" and "max-min by WMMSE", Algorithm 4 delivers a solution on the predefined rate profile ray, while the convergent points of "max-min by MMSE" and "max-min by WMMSE" are not on

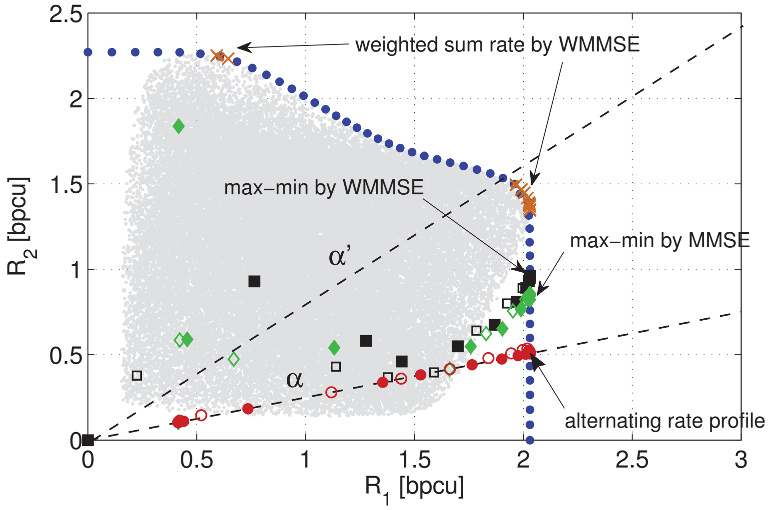
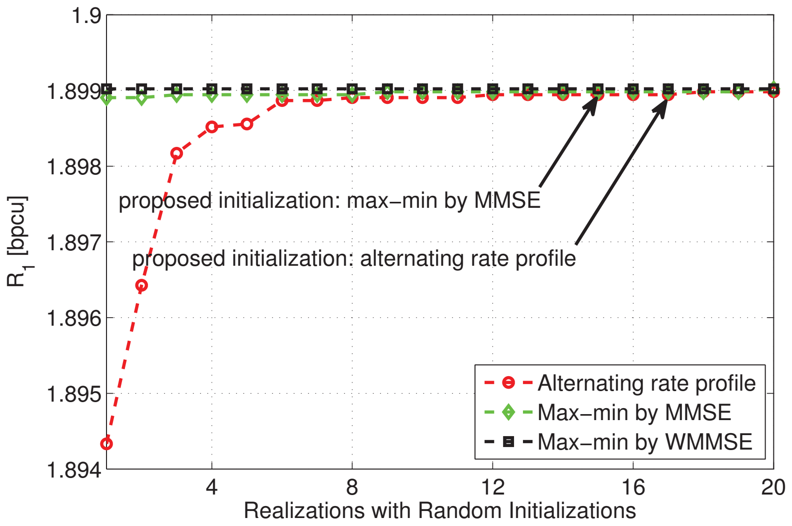


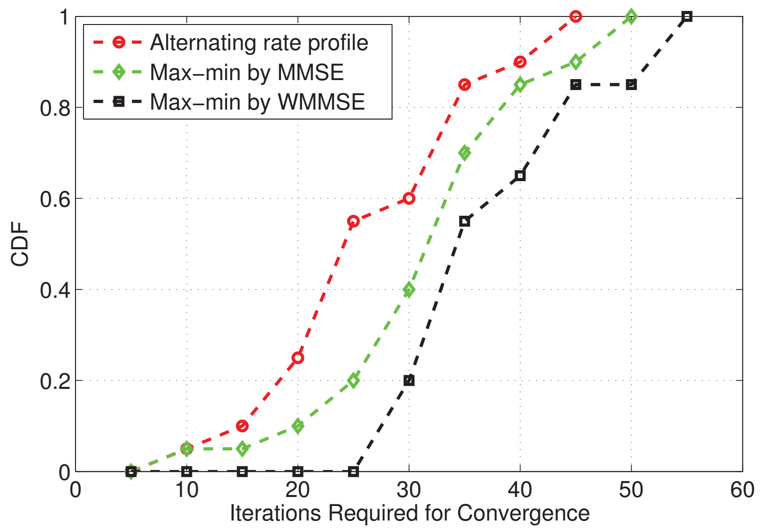
Figure 2.10: Two-dimensional achievable rate region comparison with SNR=0dB and $N_T = 2, N_R = 2$. The points marked with $\square$, $\Diamond$ and $\circ$ correspond to the rate tuples achieved in each iteration of the alternating algorithms in [RHL13], [LDL13], and Algorithm 4, respectively. The filled (unfilled) markers correspond to receive (transmit) beamforming vector optimization.

the desired rate profile. This is because both "max-min by MMSE" and "max-min by WMMSE" solve a max-min problem subject to the rate constraints with *inequalities*. Some rate constraints become inactive when the rate profile passes the non-strict Pareto boundary such that those users with inactive rate constraints may achieve higher rates. For instance, Figure 2.10 shows that $R_2^{(i)}$s in iterations by both "max-min by MMSE" and "max-min by WMMSE" are not located on the rate profile but above the rate profile. Therefore, in order to achieve an arbitrary operating point on the Pareto boundary, the rate constraints in Problem (2.63) should be with *equalities*. Then, all the rate constraints are active, and the achieved points by the MISO rate profile optimization can be pulled back to the pre-defined rate profile after solving Problem (2.63). Therefore, Algorithm 4 guarantees that the convergent point is always on the rate profile.

The proposed alternating rate profile algorithm and the max-min algorithms in [LDL13] and [RHL13] terminate on the rate profile ray if it passes the strict Pareto boundary, because all rate constraints are active in this case. For another selected rate profile $\boldsymbol{\alpha}' = (0.55, 0.45)$ passing the strict Pareto boundary in Figure 2.10, these three alter-



(a)



(b)

Figure 2.11: Convergence behavior comparison for a system with $N_T = 3, N_R = 2$ and SNR $= 0$ dB: (a) Performance comparison of 20 initializations (the values of R_1 by the proposed alternating rate profile algorithm and "max-min by MMSE" algorithm with the proposed initialization are marked); (b) Convergence speed comparison.

nating algorithms converge to the points on the rate profile. In order to show clearly the performance difference, we do not plot the convergence procedure in Figure 2.10 but plot the values of R_1 in Figure 2.11(a) for 20 initializations(one proposed initial-

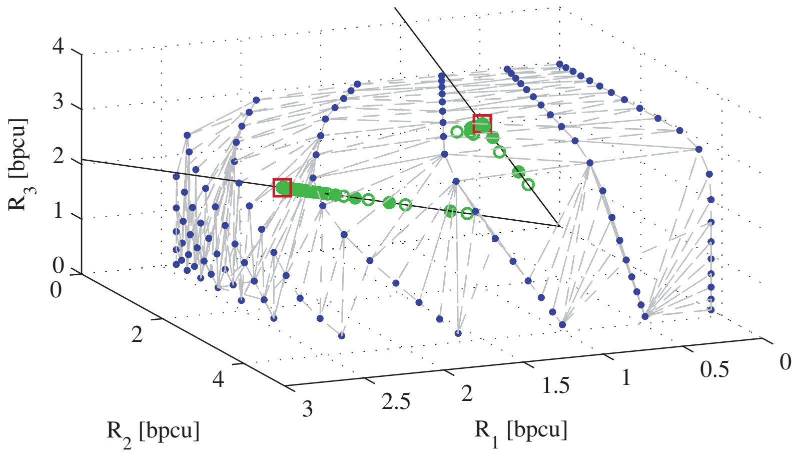


Figure 2.12: Three dimensional achievable rate region for a three-user MIMO IC with SNR=0dB and $N_T = 3, N_R = 3$. For the two rate profiles, the filled (unfilled) markers correspond to receive (transmit) beamforming optimization of Algorithm 4.

ization – the largest left singular vector of the direct channel matrix and 19 random initializations)[9]. Figure 2.11(a) shows their convergent points have a similar (almost same) achievable rate performance. Numerical results imply an order of "alternating rate profile" $\leq$ "max-min by MMSE" $\leq$ "max-min by WMMSE" in rate performance. The corresponding CDF for convergence is illustrated in Figure 2.11(b), which implies the order of "max-min by WMMSE" $\leq$ "max-min by MMSE" $\leq$ "alternating rate profile" in convergence speed.

Figure 2.11(a) and Figure 2.11(b) shows the effect of different initializations on the achievable rate performance of the weighted Chebyshev optimization-based algorithms (our proposed alternating rate profile optimization and the two max-min algorithms in [LDL13] and [RHL13]). Compared to the performance of the single constraint rate maximization algorithm in Figure 2.4 and Figure 2.5, the weighted Chebyshev optimization-based algorithms are robust to the initializations (as shown in Figure 2.11(a)) but need more iterations for convergence (as shown in Figure 2.11(b)).

We also provide a three-dimensional achievable rate region example for the three-user MIMO IC in Figure 2.12. Instead of plotting the surface of the region as Figure2.9, we plot a grid three-dimensional figure to show the optimization procedure of Algorithm 4. The terminating points of Algorithm 4, marked with a square, always achieve points on the rate profile ray.

[9] For a strict Pareto-optimal point, due to $\frac{R_1}{R_2} = \frac{0.55}{0.45}$, it is sufficient to show only the values of R_1.

2.5.3 Performance Comparison of the Two Proposed Algorithms

In this section, we compare the proposed two alternating optimization algorithms – single constraint rate maximization algorithm and alternating rate profile optimization algorithm.

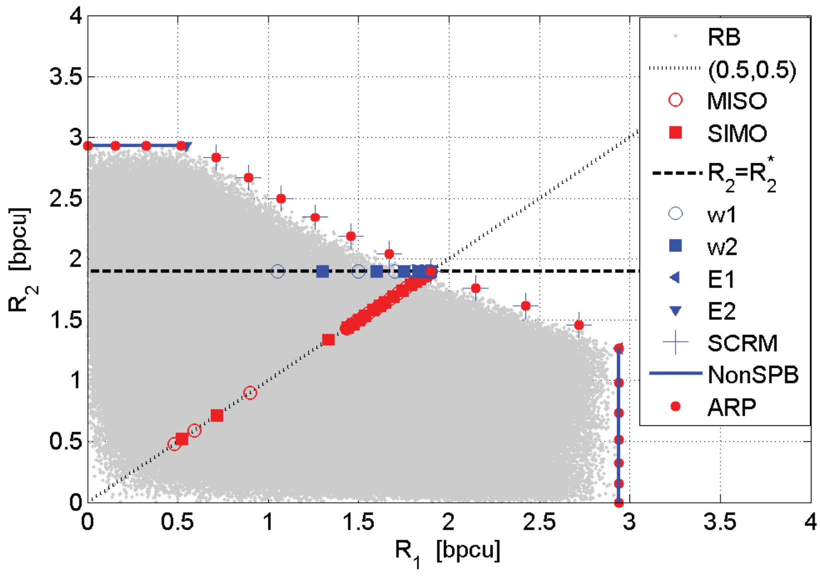


Figure 2.13: Two-dimensional achievable rate region comparison of the proposed two alternating algorithms with SNR = 0 dB and $N_T = 3, N_R = 2$. The points marked by unfilled ∘ and filled □ correspond to the rate points achieved in iterations by the two alternating optimization algorithms. The blue circles and the filled blue square correspond to the optimization of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in Algorithm 1. The red circle and filled red square correspond of the MISO rate profile optimization and the SIMO rate profile optimization in Algorithm 4.

In Figure 2.13, the cloud of points are achieved by 1 million pair of random beamforming vectors, denoted by "RB", which is bounded by the points achieved by the single constraint rate maximization algorithm (denoted by "SCRM") and the alternating rate profile algorithm (denoted by "ARP"). According to the results in Chapter 2.2, the blue lines labeled by "NonSPB" are two parts of the non-strict Pareto boundary with the ending points of "E1" and "E2" are computed. We observe that the "SCRM" and "ARP" achieve nearly the same boundary in Figure2.13.

To achieve a strict Pareto-optimal point, the alternating optimization procedure of

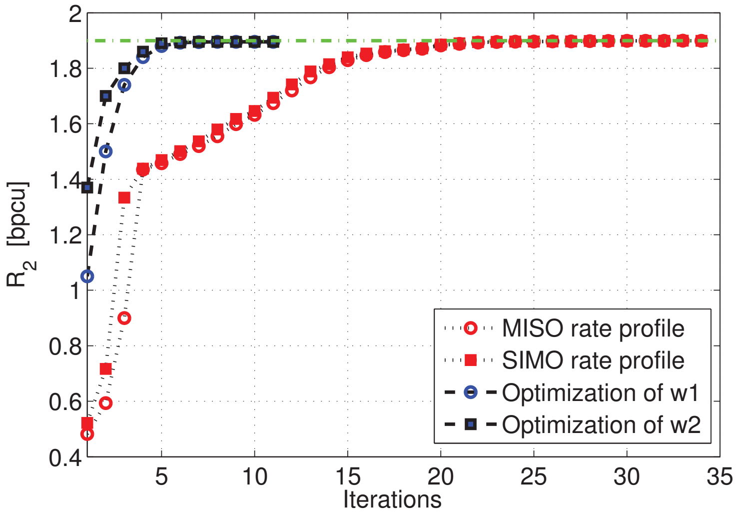


Figure 2.14: Convergence behavior comparison.

both Algorithm 1 and Algorithm 4 is shown. The rate constraint with equality in Problem (2.20) guarantees the achievable points in iterations always along the direction $R_2 = R_2^\star$ where $R_2^\star$ is the value achieved by Algorithm 4 for the rate profile $(0.5, 0.5)$. The improvement direction in Algorithm 1 is parallel to the axis of R_1, while the increasing direction in Algorithm 4 is along the pre-defined rate profile ray starting from the original. However, both the proposed algorithms are based on the same idea – *to find an intersection point between the predefined increasing direction and the Pareto boundary* and the same way – *to solve a single objective problem by the alternating optimization algorithm*. The convergence of both algorithms are guaranteed. In Figure 2.14, their convergence behavior (the monotonic increasing behavior of the objective) for the rate profile $(0.5, 0.5)$ is shown. It implies that Algorithm 1 requires less iterations to meet the same convergence criteria than Algorithm 4.

2.6 Summary

In this chapter, we propose two heuristic algorithms to compute the Pareto boundary of a multi-user single-stream MIMO IC. One is the single constraint rate maximization algorithm based on the ϵ-constraint optimization, which aims to maximize one rate while fixing other rates until a Pareto boundary point is reached. The other is the alternating rate profile optimization algorithm based on the weighted Chebyshev optimization, increasing the weighted sum rate along a predefined rate profile direction to reach the Pareto boundary. Both the algorithms are the alternating optimization algorithms, which cannot guarantee the global optimal solutions but can converge fast to high-quality suboptimal solutions. The performance of the two proposed alternating optimization algorithms depends on initializations. Numerical results verify this phenomenon and also imply that the two proposed algorithms are robust to different initializations. Furthermore, an initialization scheme for each algorithm is also proposed to achieve a good performance.

For potential implementation, the limits of the proposed algorithms is the requirement of perfect local CSI and perfect strategies feedback to MSs, which is challenging in practical systems. However, this work provides a benchmark for the algorithms that only imperfect/partial CSI is available at transmitters or the limited feedback scenario. The study of a robust alternating rate profile algorithm is one of our preparing extended work, which aims to compute the worst-case achievable rate region when each channel has a bounded channel uncertainty. Another extended work [CZJ14] studies the rate performance for a more general interference network under limited feedback, which will be presented in next chapter.

2.7 Collection of Proofs

2.7.1 Proof of Proposition 1

Given beamformers $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$, according to the matrix inversion lemma [Mey00], (2.10) can be rewritten as:

$$\mathrm{SINR}_k(\boldsymbol{w}_1, \boldsymbol{w}_2) = (\boldsymbol{H}_{kk}\boldsymbol{w}_k)^H \left(\frac{1}{\sigma_k^2}\boldsymbol{I} - \frac{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell(\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell)^H}{\sigma_k^2(\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2)} \right) \boldsymbol{H}_{kk}\boldsymbol{w}_k \tag{2.66a}$$

$$= \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2} - \frac{\left|(\boldsymbol{H}_{kk}\boldsymbol{w}_k)^H \boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\right|^2}{\sigma_k^2(\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2)} \tag{2.66b}$$

$$= \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2} \cdot \left(1 - \frac{\left|\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}^H \cdot \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}\right|^2 \cdot \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2}{\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2} \right) \tag{2.66c}$$

$$\begin{aligned} &= \left(1 - \left|\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}^H \cdot \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}\right|^2 \right) \cdot \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_\ell\|^2}{\sigma_k^2} \\ &\quad + \left|\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}^H \cdot \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}\right|^2 \cdot \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2}. \end{aligned} \tag{2.66d}$$

For two complex vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the cosine of the complex-valued angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined as [Sch01]

$$\cos(\theta_C) = \frac{\boldsymbol{a}^H\boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|} \tag{2.67}$$

where $\cos(\theta_C) = \mu e^{j\psi}$ with $\mu = |\cos(\theta_C)| \leq 1$ and $-\pi \leq \theta_C \leq \pi$ is called pseudo angle between $\boldsymbol{a}$ and $\boldsymbol{b}$.

The Hermitian angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined as

$$\cos(\theta_H) = |\cos(\theta_C)| = \frac{|\boldsymbol{a}^H\boldsymbol{b}|}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|}, \quad 0 \leq \theta_H \leq \pi/2.$$

It implies $|\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}^H \cdot \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}|^2 = \cos^2(\theta_{H,k})$ because of $\|\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}\|^2 = \|\overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}\|^2 = 1$. Thus, (2.66d) becomes

$$\mathrm{SINR}_k(\boldsymbol{w}_1, \boldsymbol{w}_2) = \sin^2(\theta_{H,k}) \cdot \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2} + \cos^2(\theta_{H,k}) \cdot \frac{\|\boldsymbol{H}_{kk}\boldsymbol{w}_k\|^2}{\sigma_k^2 + \|\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell\|^2}, \tag{2.68}$$

where $\theta_{H,k} \in [0, \pi/2]$ denotes the Hermitian angle between the desired signal direction $\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k}$ and the interference direction $\overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}$ at RX_k. Obviously, when $\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k} \parallel \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}$ (or $\overrightarrow{\boldsymbol{H}_{kk}\boldsymbol{w}_k} \perp \overrightarrow{\boldsymbol{H}_{k\ell}\boldsymbol{w}_\ell}$), we have $\theta_{H,k} = 0$ (or $\theta_{H,k} = \pi/2$).

2.7.2 Proof of Proposition 2

The idea of the proof is to show that it is not possible to achieve a strict Pareto-optimal point by the transmit beamforming vectors with less than full power. The proof works by contradiction.

Assume that a strict Pareto-optimal point $(R_1(\boldsymbol{w}_1, \boldsymbol{w}_2), R_2(\boldsymbol{w}_1, \boldsymbol{w}_2))$ is achieved by $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ where $\|\boldsymbol{w}_1\|^2 < 1$ and $\|\boldsymbol{w}_2\|^2 \leq 1$. We consider whether there exists an *outer*[10] point $(R_1(\hat{\boldsymbol{w}}_1, \boldsymbol{w}_2), R_2(\hat{\boldsymbol{w}}_1, \boldsymbol{w}_2))$ achieved by $(\hat{\boldsymbol{w}}_1, \boldsymbol{w}_2)$ where $\|\boldsymbol{w}_1\|^2 < \|\hat{\boldsymbol{w}}_1\|^2 \leq 1$ and $\|\boldsymbol{w}_2\|^2 \leq 1$. If it exists, e.g., $R_1(\hat{\boldsymbol{w}}_1, \boldsymbol{w}_2) > R_1(\boldsymbol{w}_1, \boldsymbol{w}_2)$ and $R_2(\hat{\boldsymbol{w}}_1, \boldsymbol{w}_2) = R_2(\boldsymbol{w}_1, \boldsymbol{w}_2)$, we can improve the $R_1(\boldsymbol{w}_1, \boldsymbol{w}_2)$ only by consuming more transmit power while keeping $R_2(\boldsymbol{w}_1, \boldsymbol{w}_2)$ unchanged. Thus, the existence of an *outer* operating point contradicts the assumption that $(R_1(\boldsymbol{w}_1, \boldsymbol{w}_2), R_2(\boldsymbol{w}_1, \boldsymbol{w}_2))$ is a strict Pareto-optimal point.

We define $\hat{\boldsymbol{w}}_1 \triangleq \boldsymbol{w}_1 + \boldsymbol{\delta}_p$, where $\boldsymbol{\delta}_p$ is a non-zero perturbation vector to make $||\hat{\boldsymbol{w}}_1|| \leq 1$. To guarantee the Pareto improvement, we need to show the existence of a $\boldsymbol{\delta}_p$ satisfying the following conditions:

$$(\boldsymbol{w}_1 + \boldsymbol{\delta}_p)^H \boldsymbol{A}_1(\boldsymbol{w}_2)(\boldsymbol{w}_1 + \boldsymbol{\delta}_p) > \boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\boldsymbol{w}_1 \tag{2.69a}$$

$$\boldsymbol{w}_2^H \boldsymbol{A}_2(\boldsymbol{w}_1 + \boldsymbol{\delta}_p)\boldsymbol{w}_2 = \boldsymbol{w}_2^H \boldsymbol{A}_2(\boldsymbol{w}_1)\boldsymbol{w}_2 \tag{2.69b}$$

$$\|\boldsymbol{w}_1 + \boldsymbol{\delta}_p\|^2 > \|\boldsymbol{w}_1\|^2 \tag{2.69c}$$

$$\|\boldsymbol{w}_1 + \boldsymbol{\delta}_p\|^2 \leq 1. \tag{2.69d}$$

An arbitrary nonzero $\boldsymbol{\delta}_p$ can be expressed as

$$\boldsymbol{\delta}_p = \|\boldsymbol{\delta}_p\| \cdot e^{\sqrt{-1}\phi_\delta} \cdot \overrightarrow{\boldsymbol{\delta}_p}. \tag{2.70}$$

It means that we should find $\|\boldsymbol{\delta}_p\|$, ϕ_δ and $\overrightarrow{\boldsymbol{\delta}_p}$ to satisfy all the conditions in (2.69) simultaneously. The proof of the existence of ϕ_δ is similar to that for the two-user MISO IC in [LJ08]. However, it is more difficult to find a $\overrightarrow{\boldsymbol{\delta}_p}$ for the MIMO IC because the interference channel matrix (rather than a vector in the MISO IC case) does not always have a null space for $\overrightarrow{\boldsymbol{\delta}_p}$. We give the proof in detail as follows.

1. Existence of $\overrightarrow{\boldsymbol{\delta}_p}$

By the matrix inverse lemma [Mey00], the condition in (2.69b) is equivalent to

$$\frac{|(\boldsymbol{w}_1 + \boldsymbol{\delta}_p)^H \boldsymbol{H}_{21}^H \boldsymbol{H}_{22}\boldsymbol{w}_2|^2}{\sigma_2^2 + \|\boldsymbol{H}_{21}(\boldsymbol{w}_1 + \boldsymbol{\delta}_p)\|^2} = \frac{|\boldsymbol{w}_1^H \boldsymbol{H}_{12}^H \boldsymbol{H}_{22}\boldsymbol{w}_2|^2}{\sigma_2^2 + \|\boldsymbol{H}_{21}\boldsymbol{w}_1\|^2}. \tag{2.71}$$

It is difficult to solve $\boldsymbol{\delta}_p$ directly. In fact, it is sufficient to only prove the existence of $\boldsymbol{\delta}_p$ satisfying (2.71).

[10] A point $\boldsymbol{r}' \in \mathbb{R}_K^+$ is called an outer point than $\boldsymbol{r} \in \mathbb{R}_K^+$, if $\boldsymbol{r}'$ dominates $\boldsymbol{r}$, i.e., $\boldsymbol{r}' \geq \boldsymbol{r}$ and $\boldsymbol{r}' \neq \boldsymbol{r}$ where the inequality is component-wise. The improvement from $\boldsymbol{r}$ to $\boldsymbol{r}'$ is called Pareto improvement.

Case 1. $N_R < N_T$ or $\text{rank}(\boldsymbol{H}_{21}) < N_T \le N_R$:

We always have $\boldsymbol{H}_{21}\boldsymbol{\delta}_p = \boldsymbol{0}$ if

$$\overrightarrow{\boldsymbol{\delta}_p} = \sum_{i=1}^{\text{rank}(\Pi^{\perp}_{\boldsymbol{H}_{21}^T})} a_i \boldsymbol{u}_i(\Pi^{\perp}_{\boldsymbol{H}_{21}^T}), \tag{2.72}$$

where $a_i, i = 1, ..., \text{rank}(\Pi^{\perp}_{\boldsymbol{H}_{21}^T})$ are complex-valued numbers and $\sum_{i=1}^{\text{rank}(\Pi^{\perp}_{\boldsymbol{H}_{21}^T})} |a_i|^2 = 1$. Then, (2.71) always holds because any $\boldsymbol{\delta}_p$ in the null space of the interference channel $\boldsymbol{H}_{21}$ does not cause extra interference to RX_2.

Case 2. $\text{rank}(\boldsymbol{H}_{21}) = N_T \le N_R$:

It is impossible to nullify the perturbation directly as Case 1 since there does not exist the null space for $\boldsymbol{\delta}_p$. We define $\boldsymbol{v}_1 \triangleq \boldsymbol{H}_{21}\boldsymbol{w}_1$, $\boldsymbol{v}_\delta \triangleq \boldsymbol{H}_{21}\boldsymbol{\delta}_p$ and $\boldsymbol{v}_2 \triangleq \boldsymbol{H}_{22}\boldsymbol{w}_2$. Then (2.71) becomes

$$\frac{|(\boldsymbol{v}_1 + \boldsymbol{v}_\delta)^H \boldsymbol{v}_2|^2}{\sigma_2^2 + \|\boldsymbol{v}_1 + \boldsymbol{v}_\delta\|^2} = \frac{|\boldsymbol{v}_1^H \boldsymbol{v}_2|^2}{\sigma_2^2 + \|\boldsymbol{v}_1\|^2}. \tag{2.73}$$

Assume that $\boldsymbol{v}_\delta$ is a combination of two orthogonal vectors

$$\boldsymbol{v}_\delta \triangleq \|\boldsymbol{v}_\delta\| \cdot (\sqrt{\eta} \cdot \overrightarrow{\Pi^{\perp}_{\boldsymbol{v}_2}\boldsymbol{v}_1} + \sqrt{1-\eta} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2}\boldsymbol{v}_1}), \tag{2.74}$$

Note that $\overrightarrow{\Pi_{\boldsymbol{v}_2}\boldsymbol{v}_1} = \overrightarrow{\boldsymbol{v}_2} \cdot e^{-\sqrt{-1}\phi_1}$ where $\phi_1 = \text{phase}(\boldsymbol{v}_1^H \boldsymbol{v}_2)$. Now, it remains to find whether there is a $\boldsymbol{v}_\delta$ in the plane spanned by $\overrightarrow{\Pi^{\perp}_{\boldsymbol{v}_2}\boldsymbol{v}_1}$ and $\overrightarrow{\Pi_{\boldsymbol{v}_2}\boldsymbol{v}_1}$ satisfying (2.73).

Substituting (2.74) into (2.73) yields

$$\frac{\left|\boldsymbol{v}_1^H \boldsymbol{v}_2 + \|\boldsymbol{v}_\delta\| \cdot \sqrt{1-\eta} \cdot e^{\sqrt{-1}\phi_1} \cdot \|\boldsymbol{v}_2\|\right|^2}{\sigma_2^2 + \left\|\boldsymbol{v}_1 + \|\boldsymbol{v}_\delta\|(\sqrt{\eta} \cdot \overrightarrow{\Pi^{\perp}_{\boldsymbol{v}_2}\boldsymbol{v}_1} + \sqrt{1-\eta} \cdot e^{-\sqrt{-1}\phi_1} \cdot \overrightarrow{\boldsymbol{v}_2})\right\|^2} = \frac{|\boldsymbol{v}_1^H \boldsymbol{v}_2|^2}{\sigma_2^2 + \|\boldsymbol{v}_1\|^2}. \tag{2.75}$$

Define the right-hand side and the left-hand side of (2.75) as R_{side} and $L_{side}(\eta)$, respectively. It is still hard to get a closed-form solution of η by solving (2.75) directly. Observe that the denominator of $L_{side}(\eta)$ is always positive for $\eta \in [0, 1]$, and $L_{side}(\eta)$ as a function of η is continuous over the interval $[0, 1]$. Therefore, if $(R_{side} - L_{side}(1))(R_{side} - L_{side}(0)) \le 0$, there must exist a $\boldsymbol{v}_\delta(\eta)$ with at least a certain $\eta \in [0, 1]$ satisfying (2.75).

When $\eta = 1$, the term $L_{side}(\eta)$ becomes

$$L_{side}(1) = \frac{|\boldsymbol{v}_1^H \boldsymbol{v}_2|^2}{\sigma_2^2 + \|\boldsymbol{v}_1 + \|\boldsymbol{v}_\delta\| \cdot \overrightarrow{\Pi^{\perp}_{\boldsymbol{v}_2}\boldsymbol{v}_1}\|^2} = \frac{|\boldsymbol{v}_1^H \boldsymbol{v}_2|^2}{\sigma_2^2 + \|\boldsymbol{v}_1\|^2 + \|\boldsymbol{v}_\delta\|^2 + 2\|\boldsymbol{v}_\delta\| \cdot \|\Pi^{\perp}_{\boldsymbol{v}_2}\boldsymbol{v}_1\|}. \tag{2.76}$$

Observe that $\|\boldsymbol{v}_\delta\|^2 + 2\|\boldsymbol{v}_\delta\| \cdot \|\Pi^{\perp}_{\boldsymbol{v}_2}\boldsymbol{v}_1\| > 0$. Thus, we have $L_{side}(1) < R_{side}$.

When $\eta = 0$, the term $L_{side}(\eta)$ becomes

$$\begin{aligned} L_{side}(0) &= \frac{\left|\boldsymbol{v}_1^H\boldsymbol{v}_2 + \|\boldsymbol{v}_\delta\| \cdot e^{\sqrt{-1}\phi_1} \cdot \|\boldsymbol{v}_2\|\right|^2}{\sigma_2^2 + \left\|\boldsymbol{v}_1 + \|\boldsymbol{v}_\delta\| \cdot e^{-\sqrt{-1}\phi_1} \cdot \overrightarrow{\boldsymbol{v}_2}\right\|^2} \\ &= \frac{|\boldsymbol{v}_1^H\boldsymbol{v}_2|^2 + \|\boldsymbol{v}_2\|^2 \cdot (\|\boldsymbol{v}_\delta\|^2 + 2\|\boldsymbol{v}_\delta\| \cdot |\boldsymbol{v}_1^H\overrightarrow{\boldsymbol{v}_2}|)}{\sigma_2^2 + \|\boldsymbol{v}_1\|^2 + (\|\boldsymbol{v}_\delta\|^2 + 2\|\boldsymbol{v}_\delta\| \cdot |\boldsymbol{v}_1^H\overrightarrow{\boldsymbol{v}_2}|)}, \end{aligned} \tag{2.77}$$

where $\|\boldsymbol{v}_\delta\|^2 + 2 \cdot \|\boldsymbol{v}_\delta\| \cdot |\boldsymbol{v}_1^H\overrightarrow{\boldsymbol{v}_2}| > 0$. If $L_{side}(0) > R_{side}$, we need $\|\boldsymbol{v}_2\|^2 > R_{side}$. Furthermore, R_{side} is bounded by

$$R_{side} = \frac{|\boldsymbol{v}_1^H\boldsymbol{v}_2|^2}{\sigma_2^2 + \|\boldsymbol{v}_1\|^2} \leq \frac{\|\boldsymbol{v}_1\|^2 \cdot \|\boldsymbol{v}_2\|^2}{\sigma_2^2 + \|\boldsymbol{v}_1\|^2} = \frac{\|\boldsymbol{v}_2\|^2}{\frac{\sigma_2^2}{\|\boldsymbol{v}_1\|^2} + 1} < \|\boldsymbol{v}_2\|^2.$$

Thus, we have $L_{side}(0) > R_{side}$.

Due to $L_{side}(1) < R_{side} < L_{side}(0)$, there exists at least one $\eta_0 \in (0, 1)$ satisfying $L_{side}(\eta_0) = R_{side}$. In this case, $\boldsymbol{H}_{21}$ has an inverse/Moore-Penrose pseudo-inverse matrix $\boldsymbol{H}_{21}^\dagger$. Then, we have

$$\begin{aligned} \boldsymbol{\delta}_p &= \|\boldsymbol{v}_\delta\| \cdot \boldsymbol{H}_{21}^\dagger(\sqrt{\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2}^\perp \boldsymbol{v}_1} + \sqrt{1-\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2} \boldsymbol{v}_1}), \\ &= \|\boldsymbol{\delta}_p\| \cdot e^{\sqrt{-1}\phi_\delta} \cdot \overrightarrow{\boldsymbol{\delta}_p}, \end{aligned} \tag{2.78}$$

where $\|\boldsymbol{v}_\delta\| = \frac{\|\boldsymbol{\delta}_p\|}{\left\|\boldsymbol{H}_{21}^\dagger\left(\sqrt{\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2}^\perp \boldsymbol{v}_1} + \sqrt{1-\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2} \boldsymbol{v}_1}\right)\right\|}$ depends on but has no requirement for $\|\boldsymbol{\delta}_p\|$. Therefore, $\boldsymbol{v}_\delta$ with any $\|\boldsymbol{\delta}_p\|$ and $\overrightarrow{\boldsymbol{\delta}_p} = e^{-\sqrt{-1}\phi_\delta} \cdot \overrightarrow{\boldsymbol{H}_{21}^\dagger(\sqrt{\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2}^\perp \boldsymbol{v}_1} + \sqrt{1-\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2} \boldsymbol{v}_1})}$ satisfies (2.73).

Therefore, any $\boldsymbol{\delta}_p = \|\boldsymbol{\delta}_p\| \cdot e^{\sqrt{-1}\phi_\delta} \cdot \overrightarrow{\boldsymbol{\delta}_p}$ with

$$\overrightarrow{\boldsymbol{\delta}_p} = \begin{cases} \sum_{i=1}^{\mathrm{rank}(\Pi_{\boldsymbol{H}_{21}^T}^\perp)} a_i \boldsymbol{u}_i(\Pi_{\boldsymbol{H}_{21}^T}^\perp), \quad \sum_{i=1}^{\mathrm{rank}(\Pi_{\boldsymbol{H}_{21}^T}^\perp)} |a_i|^2 = 1, \text{when } N_R < N_T \text{ or } \mathrm{rank}(\boldsymbol{H}_{21}) < N_T \leq N_R \\ e^{-\sqrt{-1}\phi_\delta} \cdot \overrightarrow{\boldsymbol{H}_{21}^\dagger(\sqrt{\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2}^\perp \boldsymbol{v}_1} + \sqrt{1-\eta_0} \cdot \overrightarrow{\Pi_{\boldsymbol{v}_2} \boldsymbol{v}_1})}, \text{ when } \mathrm{rank}(\boldsymbol{H}_{21}) = N_T \leq N_R, \end{cases} \tag{2.79}$$

satisfies the condition (2.69b).

2. Existence of ϕ_δ

Substituting (2.70) into (2.69a) yields

$$(\boldsymbol{w}_1 + \boldsymbol{\delta}_p)^H \boldsymbol{A}_1(\boldsymbol{w}_2)(\boldsymbol{w}_1 + \boldsymbol{\delta}_p) > \boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\boldsymbol{w}_1 \tag{2.80a}$$

$$\Leftrightarrow \|\boldsymbol{\delta}_p\|^2 \overrightarrow{\boldsymbol{\delta}_p}^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p} + 2\|\boldsymbol{\delta}_p\| \Re\left(\boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p} e^{\sqrt{-1}\phi_\delta}\right) > 0 \tag{2.80b}$$

$$\Leftrightarrow \frac{\|\boldsymbol{\delta}_p\|}{2} \overrightarrow{\boldsymbol{\delta}_p}^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p} + |\boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p}| \cos(\phi_\delta + \phi_2) > 0 \tag{2.80c}$$

$$\Leftrightarrow \cos(\phi_\delta + \phi_2) > -\frac{\|\boldsymbol{\delta}_p\|}{2} \cdot \frac{\overrightarrow{\boldsymbol{\delta}_p}^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p}}{|\boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p}|}, \tag{2.80d}$$

where $\phi_2 \triangleq \text{phase}(\boldsymbol{w}_1^H \boldsymbol{A}_1(\boldsymbol{w}_2)\overrightarrow{\boldsymbol{\delta}_p})$.

At the same time, substituting (2.70) into (2.69c) yields

$$\begin{aligned}
&\|\boldsymbol{w}_1 + \boldsymbol{\delta}_p\|^2 > \|\boldsymbol{w}_1\|^2 && (2.81a)\\
\Leftrightarrow\ & \|\boldsymbol{\delta}_p\|^2 + 2\|\boldsymbol{\delta}_p\|\Re\left(\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p} e^{\sqrt{-1}\phi_\delta}\right) > 0 && (2.81b)\\
\Leftrightarrow\ & |\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p}| \cos(\phi_\delta + \phi_3) > -\frac{\|\boldsymbol{\delta}_p\|}{2} && (2.81c)\\
\Leftrightarrow\ & \cos(\phi_\delta + \phi_3) > -\frac{\|\boldsymbol{\delta}_p\|}{2|\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p}|}, && (2.81d)
\end{aligned}$$

where $\phi_3 \triangleq \arg(\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p})$.

We define $\phi_\delta + \phi_2 \in [\theta_1, \theta_2]$ and $\phi_\delta + \phi_3 \in [\theta_3, \theta_4]$. Since both the right-hand side of (2.80d) and (2.81d) are negative, the range $[\theta_1, \theta_2]$ and $[\theta_3, \theta_4]$ are strictly wider than π. In addition, the intersection of two angular ranges wider than π is nonempty. Then, for arbitrary $\|\boldsymbol{\delta}_p\|$ and $\overrightarrow{\boldsymbol{\delta}_p}$, any $\boldsymbol{\delta}_p = \|\boldsymbol{\delta}_p\| \cdot e^{\sqrt{-1}\phi_\delta} \overrightarrow{\boldsymbol{\delta}_p}$ in (2.70) with $\phi_\delta \in [\theta_1, \theta_2] \cap [\theta_3, \theta_4]$ always satisfies the conditions (2.80) and (2.81) simultaneously.

3. Existence of $\|\boldsymbol{\delta}_p\|$

The condition (2.82) is equivalent to

$$\begin{aligned}
&\|\boldsymbol{\delta}_p\|^2 + 2|\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p}| \cos(\phi_\delta + \phi_3)\|\boldsymbol{\delta}_p\| + \|\boldsymbol{w}_1\|^2 - 1 \le 0,\\
\overset{(a)}{\Leftrightarrow}\ & \|\boldsymbol{\delta}_p\| \in \left(0, -|\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p}| \cos(\phi_\delta + \phi_3) + \sqrt{|\boldsymbol{w}_1^H \overrightarrow{\boldsymbol{\delta}_p}|^2 \cos^2(\phi_\delta + \phi_3) - (\|\boldsymbol{w}_1\|^2 - 1)}\right), \quad (2.82)
\end{aligned}$$

where the transformation (a) is based on $\|\boldsymbol{w}_1\| \le 1$ and $\|\boldsymbol{\delta}_p\| > 0$. For arbitrary $\overrightarrow{\boldsymbol{\delta}_p}$ and ϕ_δ, any any $\boldsymbol{\delta}_p = \|\boldsymbol{\delta}_p\| \cdot e^{\sqrt{-1}\phi_\delta} \overrightarrow{\boldsymbol{\delta}_p}$ with $\|\boldsymbol{\delta}_p\|$ in (2.82) will satisfy the condition (2.69d) .

Above all, the independent existence of $\overrightarrow{\boldsymbol{\delta}_p}$, ϕ_δ and $\|\boldsymbol{\delta}_p\|$ has been proved. That is, there always exists some $\boldsymbol{\delta}_p = \|\boldsymbol{\delta}_p\| \cdot e^{\sqrt{-1}\phi_\delta} \cdot \overrightarrow{\boldsymbol{\delta}_p}$ satisfying all the conditions in (2.69). Then, $R_1(\boldsymbol{w}_1, \boldsymbol{w}_2)$ can still be improved until $\|\boldsymbol{w}_1\|^2 = 1$, while $R_1(\boldsymbol{w}_1, \boldsymbol{w}_2)$ remains unchanged simultaneously. This contradicts the assumption that $(R_1(\boldsymbol{w}_1, \boldsymbol{w}_2), R_2(\boldsymbol{w}_1, \boldsymbol{w}_2))$ is on the strict Pareto boundary. Therefore, Proposition 2 holds.

2.7.3 Proof of Proposition 3

Consider the ending point $E1(\overline{R}_1, \underline{R}_2)$. If the link $\text{TX}_1 \mapsto \text{RX}_1$ achieves the maximum rate $\overline{R}_1$ in (2.12a), $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ should satisfy the following conditions

$$\begin{aligned}
&\boldsymbol{w}_1 = \boldsymbol{w}_1^{Ego} = \boldsymbol{u}_1(\boldsymbol{H}_{11}^H \boldsymbol{H}_{11}), && (2.83a)\\
&\theta_{H,1} = \pi/2 \Leftrightarrow \boldsymbol{w}_2 \perp \boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1. && (2.83b)
\end{aligned}$$

This means that $\boldsymbol{w}_2$ should be in the null space of $\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}$ so as to generate no interference to RX_1. Let the set of all feasible $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{FP}}$ satisfying (2.83) (i.e., the ZF strategies) be $\mathcal{W}_{\mathcal{ZF}}$. Then, any $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{ZF}}$ can be expressed by

$$\boldsymbol{w}_2 = \frac{\Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{w}_2'}{||\Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{w}_2'||}, \tag{2.84}$$

where $\boldsymbol{w}_2' \in \mathbb{C}^{N_T}$ and $\boldsymbol{w}_2' \nparallel \boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}$.

To achieve $(\overline{R}_1, \underline{R}_2)$, we need to find $(\boldsymbol{w}_2')^{opt}$ which maximizes $\mathrm{SINR}_2(\boldsymbol{w}_1^{Ego}, \boldsymbol{w}_2)$. Here, we define the optimal "altruistic" strategy $\boldsymbol{w}_2^{Alt}$ as

$$\boldsymbol{w}_2^{Alt} \triangleq \arg \max_{\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{ZF}}} \boldsymbol{w}_2^H \boldsymbol{A}_2(\boldsymbol{w}_1^{Ego}) \boldsymbol{w}_2 \tag{2.85a}$$

$$\overset{(a)}{\Leftrightarrow} (\boldsymbol{w}_2')^{opt} = \arg\max_{\boldsymbol{w}_2'} \frac{(\boldsymbol{w}_2')^H \Pi^{\perp,H}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{A}_2(\boldsymbol{w}_1^{Ego}) \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{w}_2'}{(\boldsymbol{w}_2')^H \Pi^{\perp,H}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{w}_2'} \tag{2.85b}$$

$$\overset{(b)}{=} \arg\max_{\boldsymbol{w}_2'} \frac{(\boldsymbol{w}_2')^H \overbrace{\Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{A}_2(\boldsymbol{w}_1^{Ego}) \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}}}^{\triangleq \boldsymbol{B}_1} \boldsymbol{w}_2'}{(\boldsymbol{w}_2')^H \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{w}_2'} \tag{2.85c}$$

$$= \boldsymbol{u}_1 \left(\boldsymbol{B}_1, \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \right), \tag{2.85d}$$

where transformation (a) is based on (2.84) and transformation (b) is based on the following properties

$$\Pi^{\perp,H}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} = \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \tag{2.86a}$$

$$\Pi^{\perp,H}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} = \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}}. \tag{2.86b}$$

Substituting (2.85d) into (2.84) yields the optimal "altruistic" strategy

$$\boldsymbol{w}_2^{Alt} = \overrightarrow{\Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}} \boldsymbol{u}_1(\boldsymbol{B}_1, \Pi^{\perp}_{\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{w}_1^{Ego}})}. \tag{2.87}$$

Therefore, $E1(\overline{R}_1, \underline{R}_2)$ is achieved by $(\boldsymbol{w}_1^{Ego}, \boldsymbol{w}_1^{Alt})$.

2.7.4 Proof of the Solvability of Problem (2.30)

Lemma 1. *Both the fractional problem* (2.30) *and the problem* (2.31) *are solvable.*

Proof. For the fractional problem (2.30), its constraint set

$$\Omega = \{\boldsymbol{W}_2 \succeq \boldsymbol{0} : \mathrm{Tr}(\boldsymbol{W}_2) = 1,\ \mathrm{Tr}(\boldsymbol{A}_2(\boldsymbol{w}_1)\boldsymbol{W}_2) = \mathrm{SINR}_2^{\star}\}$$

is nonempty and compact.

In Problem (2.30), the denominator of the objective over Ω satisfies

$$\sigma_1^2 + \lambda_{N_T}(\boldsymbol{H}_{12}^H \boldsymbol{H}_{12}) = \lambda_{N_T}(\boldsymbol{C}_2) \leq \mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2) \leq \lambda_1(\boldsymbol{C}_2) = \sigma_1^2 + \lambda_1(\boldsymbol{H}_{12}^H \boldsymbol{H}_{12}), \quad (2.88)$$

and the numerator obviously satisfies

$$0 \leq \mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{W}_2) \leq \lambda_1(\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{H}_{11}^H \boldsymbol{H}_{12}). \quad (2.89)$$

This implies that the objective in Problem (2.30) over Ω is bounded by

$$0 \leq \frac{\mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{W}_2)}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2)} \leq \frac{\lambda_1(\boldsymbol{H}_{12}^H \boldsymbol{H}_{11} \boldsymbol{H}_{11}^H \boldsymbol{H}_{12})}{\sigma_1^2 + \lambda_{N_T}(\boldsymbol{H}_{12}^H \boldsymbol{H}_{12})}. \quad (2.90)$$

Based on Weierstrass' Theorem, Problem (2.30) always has an optimal solution.

Assume that $\boldsymbol{W}_2^\star$ is an optimal solution to Problem (2.30), we know that $s^\star = \frac{1}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2^\star)}$ and $\boldsymbol{Q}^\star = s^\star \boldsymbol{W}_2^\star$ are feasible to Problem (2.31). Also note that the objective is bounded by (2.7.4). Similarly, Problem (2.31) is solvable according to Weierstrass' Theorem. □

Lemma 2. *The problems* (2.30) *and* (2.31) *have the same value. Furthermore, if* $\boldsymbol{W}_2^\star$ *solves Problem* (2.30)*, then* $s^\star = \frac{1}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2^\star)}$ *and* $\boldsymbol{Q}^\star = s^\star \cdot \boldsymbol{W}_2^\star$ *solves Problem* (2.31)*; if* $\boldsymbol{Q}^\star$ *and* $s^\star$ *solve Problem* (2.31)*, then* $\boldsymbol{W}_2^\star = \frac{\boldsymbol{Q}^\star}{s^\star}$ *solves Problem* (2.30)*.*

Proof. Assume that $\boldsymbol{W}_2^\star$ is an optimal solution to Problem (2.30), and $v_1^\star$ and $v_2^\star$ are the optimal values of the objective of Problem (2.30) and Problem (2.31), respectively. Thus, $s = \frac{1}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2^\star)}$ and $\boldsymbol{Q} = s\boldsymbol{W}_2^\star$ are feasible to Problem (2.31). The value of the objective of Problem (2.31) at this feasible point is

$$v_2 = \mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{Q}) \quad (2.91a)$$

$$= \mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)(s \cdot \boldsymbol{W}_2^\star)) = \frac{\mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{W}_2^\star)}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2^\star)} = v_1^\star \quad (2.91b)$$

$$\geq v_2^\star. \quad (2.91c)$$

On the other hand, suppose that $\boldsymbol{Q}^\star$ and $s^\star$ are the optimal solutions to Problem (2.31). Since $s^\star$ is always positive, $\boldsymbol{W}_2 = \frac{\boldsymbol{Q}^\star}{s^\star}$ is also feasible to Problem (2.30). Then, the value of the objective of Problem (2.30) at this feasible point is

$$v_1 = \frac{\mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{W}_2)}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{W}_2)} = \frac{\mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\frac{\boldsymbol{Q}^\star}{s^\star})}{\mathrm{Tr}(\boldsymbol{C}_2 \frac{\boldsymbol{Q}^\star}{s^\star})} \quad (2.92a)$$

$$= \frac{\mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{Q}^\star)}{\mathrm{Tr}(\boldsymbol{C}_2 \boldsymbol{Q}^\star)} = \mathrm{Tr}(\boldsymbol{C}_1(\boldsymbol{w}_1)\boldsymbol{Q}^\star) = v_2^\star \quad (2.92b)$$

$$\geq v_1^\star. \quad (2.92c)$$

Above all, we have $v_1^\star = v_2^\star$. □

2.7.5 Proof of Proposition 4

We need to find a feasible set $\mathcal{W}_{\mathcal{F}}$ such that there exists at least one solution $\boldsymbol{w}_1 \in \mathcal{W}_{\mathcal{FP}}$ to Problem (2.20) by fixing $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{F}}$.

In (2.21d), we derive that the constraint of Problem (2.20) is equivalent to

$$\boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22} \boldsymbol{w}_2 \geq \sigma_2^2 \mathrm{SINR}_2^\star \tag{2.93a}$$

$$\boldsymbol{w}_1^H \boldsymbol{C}(\boldsymbol{w}_2) \boldsymbol{w}_1 = 0. \tag{2.93b}$$

To guarantee the existence of $\boldsymbol{w}_1 \in \mathcal{W}_{\mathcal{FP}}$ in (2.93b), a feasible $\boldsymbol{w}_2$ should be determined in a way such that (2.21b) holds.

By the EVD, $\boldsymbol{C}(\boldsymbol{w}_2)$ can be rewritten as

$$\boldsymbol{C}(\boldsymbol{w}_2) = \sum_{i=1}^{N_T} \lambda_i(\boldsymbol{C}(\boldsymbol{w}_2)) \boldsymbol{u}_i(\boldsymbol{C}(\boldsymbol{w}_2)) \boldsymbol{u}_i^H(\boldsymbol{C}(\boldsymbol{w}_2)). \tag{2.94}$$

We analyze $\boldsymbol{C}(\boldsymbol{w}_2)$ for two cases.

Case 1. When $\boldsymbol{C}(\boldsymbol{w}_2)$ is a full rank matrix, i.e., $\lambda_i(\boldsymbol{C}(\boldsymbol{w}_2)) \neq 0, \forall i \in \{1, ..., N_T\}$. If $\boldsymbol{C}(\boldsymbol{w}_2)$ is a positive or negative definite matrix, it is clear that there is no nonzero vector $\boldsymbol{w}_1$ satisfying (2.93b). Otherwise, $\boldsymbol{C}(\boldsymbol{w}_2)$ has $\lambda_1(\boldsymbol{C}(\boldsymbol{w}_2)) > 0$ and $\lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2)) < 0$, a sufficient solution to (2.93b) is

$$\begin{aligned}\boldsymbol{w}_1 = &\sqrt{\frac{\lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2))}{\lambda_1(\boldsymbol{C}(\boldsymbol{w}_2)) - \lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2))}} \sqrt{-1} \cdot \boldsymbol{u}_1(\boldsymbol{C}(\boldsymbol{w}_2)) \\ &+ \sqrt{\frac{\lambda_1(\boldsymbol{C}(\boldsymbol{w}_2))}{\lambda_1(\boldsymbol{C}(\boldsymbol{w}_2)) - \lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2))}} \cdot \boldsymbol{u}_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2)).\end{aligned} \tag{2.95}$$

Case 2. When $\boldsymbol{C}(\boldsymbol{w}_2)$ is not a full rank matrix, i.e., $\lambda_i(\boldsymbol{C}(\boldsymbol{w}_2)) = 0$ for some $i \in \{1, ..., N_T\}$. In this case, $\boldsymbol{C}(\boldsymbol{w}_2)$ always has null space for $\boldsymbol{w}_1$ to satisfy (2.93b).

Above all, the sufficient and necessary condition of $\boldsymbol{w}_2$ satisfying (2.93b) is $\lambda_1(\boldsymbol{C}(\boldsymbol{w}_2)) \cdot \lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2)) \leq 0$. That is, any $\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{F}}$ is always feasible for (2.21b) where $\mathcal{W}_{\mathcal{F}}$ is

$$\mathcal{W}_{\mathcal{F}} \triangleq \Big\{\boldsymbol{w}_2 \in \mathcal{W}_{\mathcal{FP}} : \boldsymbol{w}_2^H \boldsymbol{H}_{22}^H \boldsymbol{H}_{22} \boldsymbol{w}_2 \geq \sigma_2^2 \mathrm{SINR}_2^\star, \lambda_1(\boldsymbol{C}(\boldsymbol{w}_2)) \cdot \lambda_{N_T}(\boldsymbol{C}(\boldsymbol{w}_2)) \leq 0\Big\}. \tag{2.96}$$

2.7.6 Proof of Proposition 5

The achievable rate of user k depends only on its receive beamforming vector $\boldsymbol{g}_k$ (given that $\{\boldsymbol{w}\}_{\mathcal{K}}$ are fixed). Thus, we have to show that the achievable rate of a user k takes values between $[0, R_k(\boldsymbol{g}_k^{MMSE}, \{\boldsymbol{w}\}_{\mathcal{K}})]$ for all $\boldsymbol{g}_k \in \mathcal{G}$. Since the achievable rate of user k in (2.4) is monotonically increasing in SINR, it is sufficient to analyze the SINR expression in (2.5) for the proof. We reformulate the SINR of a user k as

$$\gamma_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) = \frac{\boldsymbol{g}_k^H \boldsymbol{E}_k \boldsymbol{g}_k}{\boldsymbol{g}_k^H \boldsymbol{F}_k \boldsymbol{g}_k}, \tag{2.97}$$

where

$$\boldsymbol{E} = \boldsymbol{H}_{kk} \boldsymbol{w}_k \boldsymbol{w}_k \boldsymbol{H}_{kk} \tag{2.98}$$

$$\boldsymbol{F} = \sigma_k^2 \boldsymbol{I}_{N_R} + \sum_{\ell \neq k} \boldsymbol{H}_{k\ell} \boldsymbol{w}_\ell \boldsymbol{w}_\ell^H \boldsymbol{H}_{k\ell}^H. \tag{2.99}$$

Since $\boldsymbol{F}$ is full rank, we can transform (2.97) to a Rayleigh-Ritz ratio [HJ85, Chapter 4.2] by substituting $\boldsymbol{g}_k = \boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{z}_k$ to get

$$\gamma_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) = \frac{\boldsymbol{z}_k^H \boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{E}_k \boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{z}_k}{\boldsymbol{z}_k^H \boldsymbol{z}_k}. \tag{2.100}$$

From the Rayleigh-Ritz Theorem for Hermitian matrices [HJ85, Theorem 4.2.2] follows that

$$\gamma_k(\boldsymbol{g}_k, \{\boldsymbol{w}\}_{\mathcal{K}}) \in \mathcal{F}(\boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{E}_k \boldsymbol{F}_k^{-\frac{1}{2}}), \tag{2.101}$$

where the set $\mathcal{F}(\boldsymbol{X})$ is the field of values of a matrix $\boldsymbol{X}$ defined as [HJ91, Chapter 1]:

$$\mathcal{F}(\boldsymbol{X}) = \{\boldsymbol{x}^H \boldsymbol{X} \boldsymbol{x} \in \mathbb{R} : ||\boldsymbol{x}||^2 = 1\}. \tag{2.102}$$

The field of values $\mathcal{F}(\boldsymbol{X})$ is a compact convex set. If $\boldsymbol{X}$ is Hermitian, then $\mathcal{F}(\boldsymbol{X}) \subset \mathbb{R}$ with the smallest element and largest element corresponding to the smallest and largest eigenvalues of the matrix $\boldsymbol{X}$, respectively. Since $\boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{E}_k \boldsymbol{F}_k^{-\frac{1}{2}}$ is a rank-one positive semi-definite matrix, then the SINR in (2.100) takes values between zero and the largest eigenvalue of $\boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{E}_k \boldsymbol{F}_k^{-\frac{1}{2}}$. With $\boldsymbol{E}_k$ given in (2.98), the SINR in (2.100) is maximized by $\boldsymbol{z}_k = \boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{H}_{kk} \boldsymbol{w}_k$ which is the dominant (not normalized) eigenvector of $\boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{H}_{kk} \boldsymbol{w}_k \boldsymbol{w}_k^H \boldsymbol{H}_{kk}^H \boldsymbol{F}_k^{-\frac{1}{2}}$. Substituting $\boldsymbol{z}_k$ in $\boldsymbol{g}_k = \boldsymbol{F}_k^{-\frac{1}{2}} \boldsymbol{z}_k$ and normalizing $\boldsymbol{g}_k$ we get the expression in (2.61).

2.7.7 Proof of Proposition 6

Denote the optimization of Problem (2.58) and the optimization of Problem (2.63) by the mapping functions $\{\boldsymbol{w}\}_{\mathcal{K}} = \Psi(\{\boldsymbol{g}\}_{\mathcal{K}})$ and $\{\boldsymbol{g}\}_{\mathcal{K}} = \Xi(\{\boldsymbol{w}\}_{\mathcal{K}})$, respectively. Since both the problems can be solved optimally in each iteration in Algorithm 4, the sequence $\left\{R(\{\boldsymbol{w}\}_{\mathcal{K}}^{(i)}, \{\boldsymbol{g}\}_{\mathcal{K}}^{(i)})\right\}_{i=1}^{\infty}$ *monotonically* increases as the iteration number i increases due to the optimality of $\Xi(\cdot)$ and $\Psi(\cdot)$, and additionally is upper-bounded. The convergence of $\left\{R(\{\boldsymbol{w}\}_{\mathcal{K}}^{(i)}, \{\boldsymbol{g}\}_{\mathcal{K}}^{(i)})\right\}_{i=1}^{\infty}$ and thus the convergence of Algorithm 4 is guaranteed for any feasible initial point $\{\boldsymbol{g}^{(0)}\}_{\mathcal{K}}$.

Let $\lim_{i\to\infty} R(\{\boldsymbol{w}^{(i)}\}_{\mathcal{K}}, \{\boldsymbol{g}^{(i)}\}_{\mathcal{K}}) \triangleq \widehat{R}(\{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}, \{\widehat{\boldsymbol{g}}\}_{\mathcal{K}})$ denote the convergent point[11]. It remains to show that $(\{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}, \{\widehat{\boldsymbol{g}}\}_{\mathcal{K}}) = (\{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}, \Xi\left(\{\widehat{\boldsymbol{g}}\}_{\mathcal{K}}\right))$ is a stationary solution to Problem (2.56). Assume that $(R^\star, \{\boldsymbol{w}^\star\}_{\mathcal{K}}, \{\boldsymbol{g}^\star\}_{\mathcal{K}})$ associated with lagrange multipliers $(\{\mu_k^\star\}_{\mathcal{K}}, \{\zeta_k^\star\}_{\mathcal{K}}, \{\eta_k^\star\}_{\mathcal{K}})$ is a stationary solution to Problem (2.56), which must satisfy the following KKT conditions of Problem (2.56):

$$1 - \sum_{k\in\mathcal{K}} \mu_k^\star \alpha_k = 0, \tag{2.103a}$$

$$\sum_{k\in\mathcal{K}} \mu_k^\star \nabla_{\boldsymbol{w}_k} R_k(\boldsymbol{g}_k^\star, \{\boldsymbol{w}^\star\}_{\mathcal{K}}) - 2\zeta_k^\star \boldsymbol{w}_k^\star = 0 \quad \forall k\in\mathcal{K}, \tag{2.103b}$$

$$\mu_k^\star \nabla_{\boldsymbol{g}_k} R_k(\boldsymbol{g}_k^\star, \{\boldsymbol{w}^\star\}_{\mathcal{K}}) - 2\eta_k^\star \boldsymbol{g}_k^\star = 0 \quad \forall k\in\mathcal{K}, \tag{2.103c}$$

$$0 \le \mu_k^\star \perp R_k(\boldsymbol{g}_k^\star, \{\boldsymbol{w}^\star\}_{\mathcal{K}}) - \alpha_k R^\star \ge 0 \quad \forall k\in\mathcal{K}, \tag{2.103d}$$

$$0 \le \zeta_k^\star \perp 1 - \boldsymbol{w}_k^{\star,H} \boldsymbol{w}_k^\star \ge 0 \quad \forall k\in\mathcal{K}, \tag{2.103e}$$

$$0 < \eta_k^\star, \quad \boldsymbol{g}_k^{\star,H} \boldsymbol{g}_k^\star = 1 \quad \forall k\in\mathcal{K}. \tag{2.103f}$$

Given $\{\boldsymbol{g}\}_{\mathcal{K}} = \{\widehat{\boldsymbol{g}}\}_{\mathcal{K}}$, it is clear that $\left(\widehat{R}, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}} = \Xi(\{\widehat{\boldsymbol{g}}\}_{\mathcal{K}})\right)$ is the optimal solution to Problem (2.58). Therefore, $\left(\widehat{R}, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}\right)$ associated with lagrange multipliers $(\{\widehat{\mu}_k\}_{\mathcal{K}}, \{\widehat{\zeta}_k\}_{\mathcal{K}})$ must satisfy the following KKT conditions of Problem (2.58):

$$1 - \sum_{\mathcal{K}} \widehat{\mu}_k \alpha_k = 0, \tag{2.104a}$$

$$\sum_{\mathcal{K}} \widehat{\mu}_k \nabla_{\boldsymbol{w}_k} R_k(\widehat{\boldsymbol{g}}_k, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}) - 2\widehat{\zeta}_k \widehat{\boldsymbol{w}}_k = 0 \quad \forall k\in\mathcal{K}, \tag{2.104b}$$

$$0 \le \widehat{\mu}_k \perp R_k(\widehat{\boldsymbol{g}}_k, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}) - \alpha_k \widehat{R} \ge 0 \quad \forall k\in\mathcal{K}, \tag{2.104c}$$

$$0 \le \widehat{\zeta}_k \perp 1 - \widehat{\boldsymbol{w}}_k^H \widehat{\boldsymbol{w}}_k \ge 0 \quad \forall k\in\mathcal{K}. \tag{2.104d}$$

Similarly, $\{\widehat{\boldsymbol{g}}\}_{\mathcal{K}} = \Psi(\{\widehat{\boldsymbol{w}}\}_{\mathcal{K}})$ in Problem (2.63) corresponds to the following KKT conditions (2.104a) and

$$\widehat{\mu}_k > 0 \tag{2.105}$$

$$R_k(\widehat{\boldsymbol{g}}_k, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}) = \alpha_k \widehat{R} \quad \forall k\in\mathcal{K}, \tag{2.106}$$

$$\widehat{\mu}_k \nabla_{\boldsymbol{g}_k} R_k(\widehat{\boldsymbol{g}}_k, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}) - 2\widehat{\eta}_k \widehat{\boldsymbol{g}}_k = 0 \quad \forall k\in\mathcal{K}, \tag{2.107}$$

$$0 < \widehat{\eta}_k, \quad \widehat{\boldsymbol{g}}_k^H \widehat{\boldsymbol{g}}_k = 1 \quad \forall k\in\mathcal{K}. \tag{2.108}$$

[11] There must exist a cluster point, denoted by $\{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}$, of $\left\{\{\boldsymbol{w}^{(i)}\}_{\mathcal{K}}\right\}_{i=1}^{\infty}$ due to the compactness of the set of $\{\boldsymbol{w}\}_{\mathcal{K}}$, and the limit of $\left\{\{\boldsymbol{g}^{(i)}\}_{\mathcal{K}}\right\}_{i=1}^{\infty}$ can be expressed as $\Xi\left(\{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}\right)$ because $\Xi(\cdot)$ is a continuous function.

Combining the KKT conditions (2.104a)-(2.104d) of Problem (2.58) and (2.107)-(2.108) of Problem (2.63) and comparing with the KKT conditions (2.103a)-(2.103f), we have that $\left(\widehat{R}, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}, \{\widehat{\boldsymbol{g}}\}_{\mathcal{K}}\right)$ associated with the lagrange multipliers $(\{\widehat{\mu}_k > 0, \widehat{\zeta}_k, \widehat{\eta}_k\}_{\mathcal{K}}$ satisfy the KKT conditions of Problem (2.56), i.e., (2.103a)-(2.103f). It implies that $\left(\widehat{R}, \{\widehat{\boldsymbol{w}}\}_{\mathcal{K}}, \{\widehat{\boldsymbol{g}}\}_{\mathcal{K}}\right)$ is a stationary solution to Problem (2.56).

Chapter 3

Interference Alignment under Limited Feedback

In this chapter, we consider a more general interference network – a MIMO-IMAC, which is well matched to the uplink scenario of MIMO cellular system. IA is a promising technique to efficiently mitigate interference and to enhance the capacity of a wireless communication network. This chapter develops a framework of IA for a MIMO-IMAC with the following properties of low complexity, distributed implementation, different backhaul overhead and limited feedback. This chapter presents three main contributions [CZJ14]: 1) a complete study (including some new improvements) of the GIA with respect to the degrees of freedom (DoF) and optimal linear transceiver design is performed, which allows for low-complexity without need of iterations and distributed implementation based on the local perfect CSIR; 2) based on the GIA, the concept of IA-Cell assignment is introduced. Three IA-Cell assignment algorithms are proposed for the setup with different backhaul overhead and their DoF and rate performance are investigated; 3) the performance of the proposed GIA algorithms is studied under the limited feedback of IA precoders to MSs. To enable efficient feedback, a dynamic feedback bit allocation (DBA) problem is formulated and solved optimally.

Following the three main contributions, this chapter is organized as follows. First, we introduce the system model of a MIMO-IMAC in Chapter 3.1. The study of the GIA in terms of DoF feasibility conditions and the linear transceiver design is provided in Chapter 3.2. In Chapter 3.3, the IA-Cell assignment problem is addressed and solved. The limited feedback scenario is considered in Chapter 3.4. In Chapter 3.5, we analyze potential implementation, backhaul overhead and complexity of the proposed GIA algorithm with optimized IA-Cell assignment and under limited feedback. The numerical results in Chapter 3.6 show the effectiveness of the proposed GIA framework under unlimited and limited feedback.

3.1 System Model

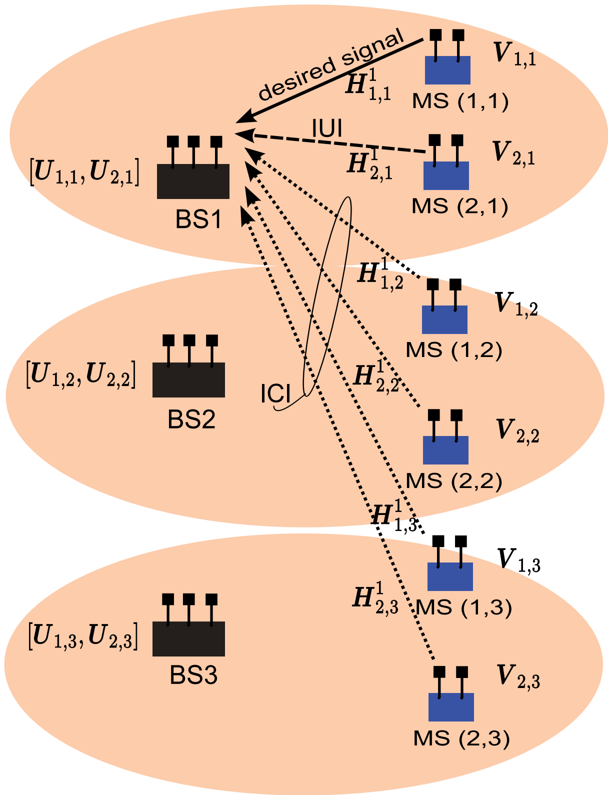


Figure 3.1: A three-cell MIMO-IMAC model where two MSs in each cell. In this example, MS $(1, 1)$ tries to convey data information to BS 1 while suffering from both the IUI and ICI.

Consider a MIMO cellular environment with K cells. In each cell, a central BS simultaneously serves L MSs in its own cell, where each BS and each MS are equipped with N_B and N_U antennas, respectively. The K cells form a *coordinated cluster* and operate over the same time-frequency resource, while the introduced IUI and ICI in return corrupt the received desired signal and limit the detection efficiency or transmission rate. Thus, interference management is required.

This chapter focuses on the uplink scenario, where the setup is modeled as a MIMO-IMAC (K, L, N_B, N_U, d_s). A MIMO-IMAC example with $K = 3$ and $L = 2$ is shown in Figure 3.1. Each MS i in cell k, denoted by MS (i, k), transmits d_s symbols $\boldsymbol{x}_{i,k} \in \mathbb{C}^{d_s \times 1}$ with $\mathbb{E}[\boldsymbol{x}_{i,k}\boldsymbol{x}^H_{i,k}] = \boldsymbol{I}_{d_s}$ to its corresponding BS k. The symbol vector $\boldsymbol{x}_{i,k}$ is precoded by a linear precoder $\boldsymbol{V}_{i,k} \in \mathbb{C}^{N_U \times d_s}$ subject to $\mathrm{Tr}(\boldsymbol{V}^H_{i,k}\boldsymbol{V}_{i,k}) \leq P_{i,k}$ where $P_{i,k}$ is the transmit power budget.

We assume that the local CSIR is perfectly estimated at each BS based on uplink pilot signals (e.g., [**?**]). The received signal at BS k for MS (i,k) is expressed as

$$\boldsymbol{y}_{i,k} = \underbrace{\boldsymbol{H}^k_{i,k}\boldsymbol{V}_{i,k}\boldsymbol{x}_{i,k}}_{desired\ signal} + \underbrace{\sum_{j=1,j\neq i}^{L} \boldsymbol{H}^k_{j,k}\boldsymbol{V}_{j,k}\boldsymbol{x}_{j,k}}_{IUI} + \underbrace{\sum_{\ell=1,\ell\neq k}^{K}\sum_{m=1}^{L} \boldsymbol{H}^k_{m,\ell}\boldsymbol{V}_{m,\ell}\boldsymbol{x}_{m,\ell}}_{ICI} + \boldsymbol{n}_k, \tag{3.1}$$

where $\boldsymbol{H}^\ell_{i,k}$ denotes the channel matrix from MS (i,k) to BS ℓ and is modeled as $\sqrt{\eta^\ell_{i,k}}\overline{\boldsymbol{H}}^\ell_{i,k}$, where $\eta^\ell_{i,k}$ denotes the effect of path-loss, and $\overline{\boldsymbol{H}}^\ell_{i,k} \in \mathbb{C}^{N_B\times N_U}$ is a Rayleigh fading channel matrix. Each channel is assumed to be quasi-static and frequency flat fading. $\boldsymbol{n}_k \in \mathbb{C}^{N_B\times 1}$ is the additive white Gaussian noise vector with zero mean and variance $\sigma_k^2\boldsymbol{I}_{N_B}$.

With the linear single-user decoding scheme, the received signal vector $\boldsymbol{y}_{i,k}$ for MS (i,k) can be decoded as $\widehat{\boldsymbol{x}}_{i,k} = \boldsymbol{U}^H_{i,k}\boldsymbol{y}_{i,k}$ by the decoder $\boldsymbol{U}_{i,k} \in \mathbb{C}^{N_B\times d_s}$. In order to make efficient detection of the desired signal, the desired signal should be linearly independent of the interference, i.e., the following conditions need to be satisfied:

$$\boldsymbol{U}^H_{i,k}\boldsymbol{H}^k_{j,k}\boldsymbol{V}_{j,k} = \boldsymbol{0}, \quad \forall j \neq i \tag{3.2a}$$

$$\boldsymbol{U}^H_{i,k}\boldsymbol{H}^k_{m,\ell}\boldsymbol{V}_{m,\ell} = \boldsymbol{0}, \quad \forall \ell \neq k,\ \forall m \tag{3.2b}$$

$$\mathrm{rank}(\boldsymbol{U}^H_{i,k}\boldsymbol{H}^k_{i,k}\boldsymbol{V}_{i,k}) = d_s, \quad \forall i,k, \tag{3.2c}$$

where (3.2a) and (3.2b) enable the mitigation of IUI and ICI, respectively, and (3.2c) guarantees the transmission of d_s data streams per MS. Then, the achievable rate for MS (i,k) is

$$R_{i,k} = \log\det(\boldsymbol{I}_{d_s} + \frac{1}{\sigma_k^2}\boldsymbol{U}^H_{i,k}\boldsymbol{H}^k_{i,k}\boldsymbol{V}_{i,k}\boldsymbol{V}^H_{i,k}\boldsymbol{H}^{k,H}_{i,k}\boldsymbol{U}_{i,k}), \tag{3.3}$$

where $\log\det(\cdot)$ denotes the operation of $\log_2(\det(\cdot))$.

For the conditions (3.2a)-(3.2c) to be fulfilled in the system (K, L, N_B, N_U, d_s), any MS (i,k) needs to satisfy

$$\boldsymbol{U}^H_{i,k}\Big[\{\boldsymbol{H}^k_{j,k}\boldsymbol{V}_{j,k}\}^L_{j=1,j\neq i},\ \{\boldsymbol{F}^k_\ell\}^K_{\ell=1,\ell\neq k}\Big] \triangleq \boldsymbol{U}^H_{i,k}\boldsymbol{F}_{i,k} = \boldsymbol{0} \tag{3.4}$$

where $\boldsymbol{F}_{i,k} \in \mathbb{C}^{N_B\times(KL-1)d_s}$ denotes the interference matrix.

Sufficient and Necessary Conditions for (3.4): (3.4) is fulfilled if and only if $N_B \geq \mathrm{rank}(\boldsymbol{F}_{i,k}) + d_s$ such that BS k could provide at least a $\mathrm{rank}(\boldsymbol{F}_{i,k})$-dimensional subspace to nullify all the interference to MS (i,k) and simultaneously guarantee d_s DoF per MS.

Due to $\mathrm{rank}(\boldsymbol{F}_{i,k}) \leq (KL-1)d_s$, it is sufficient to fulfill (3.4) by only exploiting the ZF decoding if $N_B \geq KLd_s$. In general, we have $\mathrm{rank}(\boldsymbol{F}_{i,k}) = (KL-1)d_s$ if there is no restrictions on the transmission through Rayleigh fading channels. In this paper, we study the interference cancellation in a non-trivial case $((K-1)L+1)d_s \leq N_B < KLd_s$ where the sole ZF decoding fails and IA is required. Instead of developing iterative IA algorithms, we study the problem of *low-complexity* IA transceiver design, also considering the problem of IA-Cell assignment and limited feedback.

Definitions: The channel set from MSs in cell k to BS ℓ: $\boldsymbol{H}_k^\ell \triangleq \{\boldsymbol{H}_{i,k}^\ell\}_{i=1}^L$. The local CSIR of BS ℓ: $\boldsymbol{H}^\ell \triangleq \{\boldsymbol{H}_k^\ell\}_{k=1}^K$. The interference from cell k to cell ℓ: $\boldsymbol{F}_k^\ell \triangleq [\boldsymbol{H}_{1,k}^\ell \boldsymbol{V}_{1,k}, \ldots, \boldsymbol{H}_{L,k}^\ell \boldsymbol{V}_{L,k}] \in \mathbb{C}^{N_B \times Ld_s}$. The IUI of MS (i,k): $\boldsymbol{F}_{i,k}^{IUI} \triangleq [\{\boldsymbol{H}_{j,k}^k\}_{j=1,j\neq i}^L] \in \mathbb{C}^{N_B \times (L-1)d_s}$.

3.2 Interference Alignment and Cancellation

In this section, we develop a *restriction-relaxation* two-stage algorithm based on the GIA method proposed in [SLL+11, TL13] to determine the optimal IA transceiver in closed-form.

3.2.1 Feasible Conditions for the GIA

The GIA method in [TL13] is a generalization of the non-iterative grouping scheme originally proposed in [SLL+11] to completely suppress the interference. *The basic idea of GIA method in [TL13] is to group all the MSs in one cell to generate a joint precoder aligning their interference to another cell.* Let *Cell* $k \xrightarrow{IA}$ *Cell* k' denote that cell k aligns its interference to cell k'. The feasible conditions for the GIA method and its DoF performance are shown in the following proposition.

Proposition 7. *For a MIMO-IMAC system* (K, L, N_B, N_U, d_s), *at least* d_s *DoF per MS and* KLd_s *sum DoF can be achieved by the GIA method if*

$$N_U \geq \frac{L-1}{L} N_B + \frac{1}{L} d_s \text{ and } N_B \geq ((K-1)L+1)d_s. \tag{3.5}$$

Proof. Without loss of generality, to fix ideas we consider the following scenario.

$$\textit{Cell } 1 \xrightarrow{IA} \textit{Cell } 2 \xrightarrow{IA} \ldots \xrightarrow{IA} \textit{Cell } K \xrightarrow{IA} \textit{Cell } 1. \tag{3.6}$$

In particular, the procedure of *Cell* $k \xrightarrow{IA}$ *Cell* $k+1$ can be implemented by

$$\overline{\boldsymbol{F}}_k^{k+1} \triangleq \mathrm{Span}\{\boldsymbol{H}_{1,k}^{k+1} \boldsymbol{V}_{1,k}\} = \mathrm{Span}\{\boldsymbol{H}_{2,k}^{k+1} \boldsymbol{V}_{2,k}\} = \ldots = \mathrm{Span}\{\boldsymbol{H}_{L,k}^{k+1} \boldsymbol{V}_{L,k}\}. \tag{3.7}$$

First, we *restrict* (3.7) to find those precoding matrices such that

$$\boldsymbol{H}_{1,k}^{k+1} \boldsymbol{V}_{1,k}^{in} = \boldsymbol{H}_{2,k}^{k+1} \boldsymbol{V}_{2,k}^{in} = \ldots = \boldsymbol{H}_{L,k}^{k+1} \boldsymbol{V}_{L,k}^{in}. \tag{3.8}$$

By this *restriction stage*, $\{\boldsymbol{V}_{i,k}^{in}\}_{i=1}^{L}$ in (3.8) is sufficient but not necessary to (3.7). We equivalently rewrite (3.8) as

$$\begin{bmatrix} \boldsymbol{H}_{1,k}^{k+1} & -\boldsymbol{H}_{2,k}^{k+1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{H}_{1,k}^{k+1} & \boldsymbol{0} & \boldsymbol{0} & \cdots & -\boldsymbol{H}_{L,k}^{k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_{1,k}^{in} \\ \boldsymbol{V}_{2,k}^{in} \\ \vdots \\ \boldsymbol{V}_{L,k}^{in} \end{bmatrix} \triangleq \boldsymbol{A}_k^{k+1} \boldsymbol{V}_k^{in} = \boldsymbol{0} \tag{3.9}$$

where $\boldsymbol{A}_k^{k+1} \in \mathbb{C}^{(L-1)N_B \times LN_U}$ and $\boldsymbol{V}_k^{in} \in \mathbb{C}^{LN_U \times d_s}$. To fulfill (3.9), the joint IA precoder $\boldsymbol{V}_k^{in}$ should lie in the null space of $\boldsymbol{A}_k^{k+1}$, which requires $LN_U \geq (L-1)N_B + d_s$ such that $\boldsymbol{A}_k^{k+1}$ has a at least d_s-dimensional null space.

By (3.9), the original Ld_s-dimensional interference subspace $\boldsymbol{F}_k^{k+1}$ is aligned to a d_s-dimensional subspace $\overline{\boldsymbol{F}}_k^{k+1}$ because (3.7) holds, while the interference $\boldsymbol{F}_k^{\ell}$ $\forall \ell \neq k, k+1$ is still with Ld_s dimensions. Under the assignment of (3.6), it is sufficient for each BS k to remove the complete interference for user (i,k) by the ZF decoding if $N_B \geq ((K-1)L+1)d_s$. □

Given an IA-Cell assignment, the joint IA precoder $\boldsymbol{V}_k^{in}$ for the MSs within each cell k can be determined. Let $\overrightarrow{\boldsymbol{V}}_{i,k} \triangleq \boldsymbol{T}_i \boldsymbol{V}_k (\boldsymbol{V}_k^H \boldsymbol{T}_i^H \boldsymbol{T}_i \boldsymbol{V}_k)^{-1}$ be the *precoder pattern* in (3.13) where $\overrightarrow{\boldsymbol{V}}_{i,k}^H \overrightarrow{\boldsymbol{V}}_{i,k} = \boldsymbol{I}_{d_s}$. In order to implement a closed-loop transmission, $\overrightarrow{\boldsymbol{V}}_{i,k}$ needs to be fed back to MS (i,k). Since feedback links are usually capacity-limited, *subspace quantization* is employed to reduce overhead. A subspace matrix is mapped to an index in a predefined codebook. However, the feedback of an index results in the residual interference, since the interference cannot be perfectly aligned due to the quantization distortion. Therefore, the problem of DBA to minimize sum-cluster RINR is of interest.

Remark 9. *By the feasible conditions (3.5) in Proposition 7, we gain the following insights on system design.*

1) *Given (K, L, N_B, N_U), each MS achieves at most* $\min(LN_U-(L-1)N_B, \frac{N_B}{(K-1)L+1})$ *DoF;*

2) *Given (K, L, N_B, d_s), each MS needs at least $((L-1)(K-1)+1)d_s$ antennas to guarantee its d_s DoF;*

3) *Given (K, N_B, N_U, d_s), each cell serves at most* $\min(\frac{N_B-d_s}{N_B-N_U}, \frac{N_B-d_s}{(K-1)d_s})$ *MSs;*

4) *Given (L, N_B, N_U, d_s), at most $\frac{N_B-d_s}{Ld_s}+1$ cells can be scheduled to form a cluster with the sum DoF of KLd_s if $N_U \geq \frac{L-1}{L}N_B + \frac{1}{L}d_s$.*

If the inequalities in both feasible conditions (3.5) become equalities, the required number of BS and MS antennas are the smallest.

3.2.2 Transceiver Optimization for the GIA

As in [TL13, GC14], we hereafter focus on the worst-case that $N_B = ((K-1)L+1)d_s$ and $N_U = \lceil \frac{L-1}{L} N_B + \frac{1}{L} d_s \rceil$. In this case, the optimal GIA transceiver are computed in closed-form.

Proposition 8. *Let us define*

$$\boldsymbol{T}_i \triangleq [\mathbf{0}_{N_U \times (i-1)N_U}, \boldsymbol{I}_{N_U}, \mathbf{0}_{N_U \times (L-i)N_U}] \tag{3.10}$$

$$\boldsymbol{V}_k^{in} = \left(\boldsymbol{A}_k^{k+1,H}\right)^{\perp} \tag{3.11}$$

$$\boldsymbol{F}_{i,k}^{IA,k-1} \triangleq \left[\boldsymbol{F}_{i,k}^{IUI},\ \left\{\boldsymbol{F}_{\ell}^{k}\right\}_{\ell=1,\ell\neq k,k-1}^{K},\ \overline{\boldsymbol{F}}_{k-1}^{k}\right]. \tag{3.12}$$

Considering (3.6) and the uniform power allocation policy, the achievable rate of each MS (i,k) in (3.3) is maximized by the optimal transceiver

$$\boldsymbol{V}_{i,k} = \sqrt{\frac{P_{i,k}}{d_s}} \boldsymbol{T}_i \boldsymbol{V}_k^{in} (\boldsymbol{V}_k^{in,H} \boldsymbol{T}_i^H \boldsymbol{T}_i \boldsymbol{V}_k^{in})^{-\frac{1}{2}} \tag{3.13}$$

$$\boldsymbol{U}_{i,k} = \left(\boldsymbol{F}_{i,k}^{IA,k-1}\right)^{\perp}. \tag{3.14}$$

Proof. Without loss of generality, we consider the scenario (3.6). First, since $\boldsymbol{V}_k^{in}$ lies in the null space of $\boldsymbol{A}_k^{k+1}$ to fulfill (3.9). Based on the fact $\mathrm{Span}(\boldsymbol{V}_{i,k}^{in}\boldsymbol{X}) = \mathrm{Span}(\boldsymbol{V}_{i,k}^{in})$ where $\boldsymbol{X} \in \mathbb{C}^{d_s \times d_s}$ is an arbitrary full-rank matrix, the IA precoder for each MS (i,k) is defined as

$$\boldsymbol{V}_{i,k} \triangleq \boldsymbol{V}_{i,k}^{in} \boldsymbol{V}_{i,k}^{out} = \boldsymbol{T}_i \boldsymbol{V}_k^{in} \boldsymbol{V}_{i,k}^{out} \tag{3.15}$$

where $\boldsymbol{T}_i$ is a selection matrix defined in (3.10) and $\boldsymbol{V}_k^{in}$ is an *inner precoder* defined in (3.11), and $\boldsymbol{V}_{i,k}^{out} \in \mathbb{C}^{d_s \times d_s}$ is an *outer precoder* subject to the transmit power constraint $\mathrm{Tr}(\boldsymbol{V}_{i,k}^{out,H} \boldsymbol{V}_{i,k}^{in,H} \boldsymbol{V}_{i,k}^{in} \boldsymbol{V}_{i,k}^{out}) \leq P_{i,k}$, which is used to *relax* the restriction from (3.7) to (3.8) and *this relaxation is tight.* The optimal precoder $\boldsymbol{V}_{i,k}$ can be determined by further optimizing $\boldsymbol{V}_{i,k}^{out}$.

Also due to $\mathrm{Span}(\boldsymbol{H}_{j,\ell}^{k} \boldsymbol{V}_{j,\ell}^{in} \boldsymbol{V}_{j,\ell}^{out}) = \mathrm{Span}(\boldsymbol{H}_{j,\ell}^{k} \boldsymbol{V}_{j,\ell}^{in})$, it is sufficient to design the ZF decoder $\boldsymbol{U}_{i,k}$ only based on $\boldsymbol{V}_{j,\ell}^{in}$ but without knowledge of $\boldsymbol{V}_{j,\ell}^{out}$. The ZF decoder for MS (i,k) can be designed by

$$\boldsymbol{U}_{i,k} \triangleq \boldsymbol{U}_{i,k}^{in} \boldsymbol{U}_{i,k}^{out} = \left(\boldsymbol{F}_{i,k}^{IA,k-1}\right)^{\perp} \boldsymbol{U}_{i,k}^{out}, \tag{3.16}$$

where $\boldsymbol{F}_{i,k}^{IA,k-1}$ defined in (3.12) is a $N_B \times ((K-1)L+1)d_s$ interference matrix with the aligned interference from cell $k-1$. $\boldsymbol{U}_{i,k}^{in}$ serves as an *inner decoder* to nullify interference, and $\boldsymbol{U}_{i,k}^{out} \in \mathbb{C}^{d_s \times d_s}$ is an *outer decoding* matrix with $\boldsymbol{U}_{i,k}^{out,H} \boldsymbol{U}_{i,k}^{out} = \boldsymbol{I}_{d_s}$.

With the IA transceiver in form of (3.15) and (3.16), the achievable rate of each MS (i,k) becomes[1]

$$R_{i,k}^{IA} = \log\det\Big(\boldsymbol{I}_{d_s} + \frac{1}{\sigma_k^2}\widetilde{\boldsymbol{H}}_{i,k}^{k}\boldsymbol{V}_{i,k}^{out}\boldsymbol{V}_{i,k}^{out,H}\widetilde{\boldsymbol{H}}_{i,k}^{k,H}\Big), \tag{3.17}$$

where $\widetilde{\boldsymbol{H}}_{i,k}^{k}$ denotes the effective channel from (i,k) to BS k

$$\widetilde{\boldsymbol{H}}_{i,k}^{k} \triangleq \boldsymbol{U}_{i,k}^{in,H}\boldsymbol{H}_{i,k}^{k}\boldsymbol{T}_i\boldsymbol{V}_{k}^{in}, \tag{3.18}$$

and with the constraints $\mathrm{Tr}(\boldsymbol{V}_{i,k}^{out,H}\boldsymbol{V}_{i,k}^{in,H}\boldsymbol{V}_{i,k}^{in}\boldsymbol{V}_{i,k}^{out}) \le P_{i,k}$. Under the assumption of equal transmit power allocation with practical considerations[2], maximization of $R_{i,k}^{IA}$ yields the maximum rate $\log\det\left(\boldsymbol{I}_{d_s} + \frac{P_{i,k}}{\sigma_k^2 d_s}\widetilde{\boldsymbol{H}}_{i,k}^{k}(\boldsymbol{V}_{i,k}^{in,H}\boldsymbol{V}_{i,k}^{in})^{-1}\widetilde{\boldsymbol{H}}_{i,k}^{k,H}\right)$ by the optimal solution $\boldsymbol{V}_{i,k}^{out} = \sqrt{\frac{P_{i,k}}{d_s}}(\boldsymbol{V}_{i,k}^{in,H}\boldsymbol{V}_{i,k}^{in})^{-\frac{1}{2}}$, thereby (3.13) and (3.14). □

The improvements of the derived results with respect to previous works on the GIA [SLL+11, TL13] are two-fold.

- *Lower complexity:* the complexity of the GIA mainly depends on the singular-value decomposition (SVD) of K matrices $\{\boldsymbol{A}_k^{k+1}\}$. By the *new formulation* (3.9), our GIA takes $K\mathcal{O}((L-1)^2 L N_B^2 N_U)$ arithmetic operations, since each $\boldsymbol{A}_k^{k+1}$ is a $(L-1)N_B \times LN_U$ matrix. In contrast, [TL13, Eq. (27)] (same as [SLL+11]) and [TL13, Eq. (12)-(13), (15)] have the complexity of $K\mathcal{O}(L^2N_B^2(LN_U+N_B))$ or $K\left(L\mathcal{O}(N_B^3+N_B^2N_U)+2(L+\log_2(L))\mathcal{O}(2N_B^2N_U)\right)$, respectively. It follows that the complexity of our GIA by (3.9) is always lower than [TL13, Eq. (27)] and also lower than that by [TL13, Eq. (12)-(13), (15)] when $L \le 3$.[3]

- *Optimality/tightness of the restriction-relaxation:* the procedure of *the restriction* from (3.7) to (3.8) combined with the *relaxation* by introducing the outer precoder and decoder, subject to $\mathrm{Tr}(\boldsymbol{V}_{i,k}^{out,H}\boldsymbol{V}_{i,k}^{in,H}\boldsymbol{V}_{i,k}^{in}\boldsymbol{V}_{i,k}^{out}) \le P_{i,k}$ and $\boldsymbol{U}_{i,k}^{out}$ being a unitary matrix, into the definitions (3.15) and (3.16) is tight. This property guarantees the optimality of linear transceiver design by the restriction-relaxation method, which

[1]The rate $R_{i,k}^{IA}$ is independent of the unitary matrix $\boldsymbol{U}_{i,k}^{out}$.

[2]Instead of the optimal water-filling based power allocation across the data streams, the uniform power allocation policy is adopted because of the following reasons: 1) it is known to be asymptotically optimal for large SNR [RJ08], 2) it guarantees the transmission of d_s data streams per MS (i.e., condition (3.2c)), 3) it has lower complexity compared with water-filling process and 4) it is not necessary to feed back the outer precoders to MSs.

[3]The computation of the left singular-space and the singular values of a $M \times N$ matrix where $M < N$ is $4NM^2 + 8M^3$ arithmetic operations [TU]. Based on this the complexity comparison with [TL13] is done.

does not hold in the work [TL13] where the constraints $\mathrm{Tr}(\boldsymbol{V}_{i,k}^{out,H}\boldsymbol{V}_{i,k}^{out}) \leq P_{i,k}$ and the DPC scheme are adopted.

Remark 10. *The GIA as a non-iterative algorithm determines the IA transceiver in a distributed way and with low complexity. For the distributed implementation, BSs need to exchange their inner precoders $\{\boldsymbol{V}_k^{in}\}_{k=1}^{K}$ with each other, while the outer precoder $\boldsymbol{V}_k^{out}$ and the outer decoder $\boldsymbol{U}_k^{out}$ can be designed by each MS (i,k) and BS k independently.*

3.3 IA-Cell Assignment: Problems & Solutions

In this section, we introduce the concept of IA-Cell assignment, motivate its importance for network performance and propose three algorithms for assignment optimization.

3.3.1 IA-Cell Assignment Problems

A. Observation & Motivation

For *Cell* $k \xrightarrow{IA}$ *Cell* k', we label cell k as the *IA-provider* for cell k' and cell k' as the *IA-receiver* from cell k. Clearly, this poses an assignment problem between IA-providers and IA-receivers – how should we select the IA-receiver (or IA-provider) corresponding to a given IA-provider (or IA-receiver)? From the perspective of spatial resources, a cell will waste part of its transmit spatial resources if it aligns its interference to other cells because of the IA constraint. On the other hand, a cell can save its receive spatial resource if it receives the aligned interference from other cells. Thus, providing IA and receiving IA can be considered as the *cost* and *gains*, respectively. In order to gain mutual benefits, it is expected that each cell in a coordinated cluster should simultaneously serve as an IA-provider and IA-receiver (i.e., *gains with cost*), because it is fair and motivated for multiple cells to coordinate with each other voluntarily, which allows for distributed implementations and self-organization. The mapping of K potential aligned interference to K cells in a coordinated cluster can be formulated as an *IA-Cell assignment* problem. Now, two questions arise: *Q1 – How many possible IA-Cell assignments exist in a K-cell cluster? Q2 – How to find a good IA-Cell assignment?*

B. Effect of Assignment on DoF

We first give the definitions regarding the IA-Cell assignment.

Definition 2. *(Coordinated Cell and Lone Cell) If a cell receives the aligned interference from other cells and it also aligns its own interference to others, this cell is called a* coordinated cell*; Otherwise, a cell is called a* lone cell *if it does not receive an IA from others and also it has no incentive to and will not provide its IA to others.*

Definition 3. *(Strict/Weak IA-Cell Assignment) The assignment is called a* strict IA-Cell assignment *if each cell is a coordinated cell, e.g., the example in (3.6). Otherwise, we have a* weak IA-Cell assignment*.*

For the considered system (K, L, N_U, N_B), maximum DoF can be achieved only under the strict IA-Cell assignment, which can be easily proved by contradiction. Otherwise, the lone cell has to reduce its transmit data streams because it receives $(K-1)Ld_s$-dimensional interference and thus its desired Ld_s DoF cannot be supported by $N_B = (K-1)Ld_s + d_s$ receive antennas. Under a weak IA-Cell assignment, the lone cell has only d_s DoF, while other coordinated cells are with Ld_s DoF per cell. For instance, the system $(K, L, N_U, N_B) = (3, 2, 6, 10)$ can achieve 12 sum DoF (4 DoF per cell) under a strict IA-Cell assignment, while only 10 sum DoF is achieved when there exists a lone cell (4 DoF per coordinated cell and 2 DoF of the lone cell). Therefore, *a lone cell is suboptimal* as far as either the sum DoF or fairness is concerned. Thus, the focus will be on strict IA-Cell from now on.

The question *Q1* is answered by the following lemma.

Lemma 3. *A K-cell IA-Cell assignment problem where $K \geq 3$ has $K! \sum_{k=0}^{K} \frac{(-1)^k}{k!} - 1$ strict IA-Cell assignments in total.*

Proof. Let us label K cells with the index sequence $1, 2, \ldots, K$. Under a strict IA-Cell assignment, each cell simultaneously serves as an IA-provider and IA-receiver and both for other cells. Therefore, the index sequence of K IA-providers or IA-receivers of the K cells in the sequence of $1, 2, \ldots, K$ should not share the same index at a common position. It can be formulated as a well-known *derangement problem: determine the permutations of the K elements of a set such that none of the elements appear in their original positions*, which has $K! \sum_{k=0}^{K} \frac{(-1)^k}{k!}$ derangements [Has03]. Therefore, there are $K! \sum_{k=0}^{K} \frac{(-1)^k}{k!} - 1$ strict IA-Cell assignment for a given K cells. □

According to the result in Lemma 3, there exist two strict IA-Cell assignment in total for a $K = 3$ cells IA-Cell assignment problem, which are illustrated in the following Figure 3.2.

Corollary 1. *Under different strict IA-Cell assignments, the system (K, L, N_U, N_B) has the same DoF performance.*

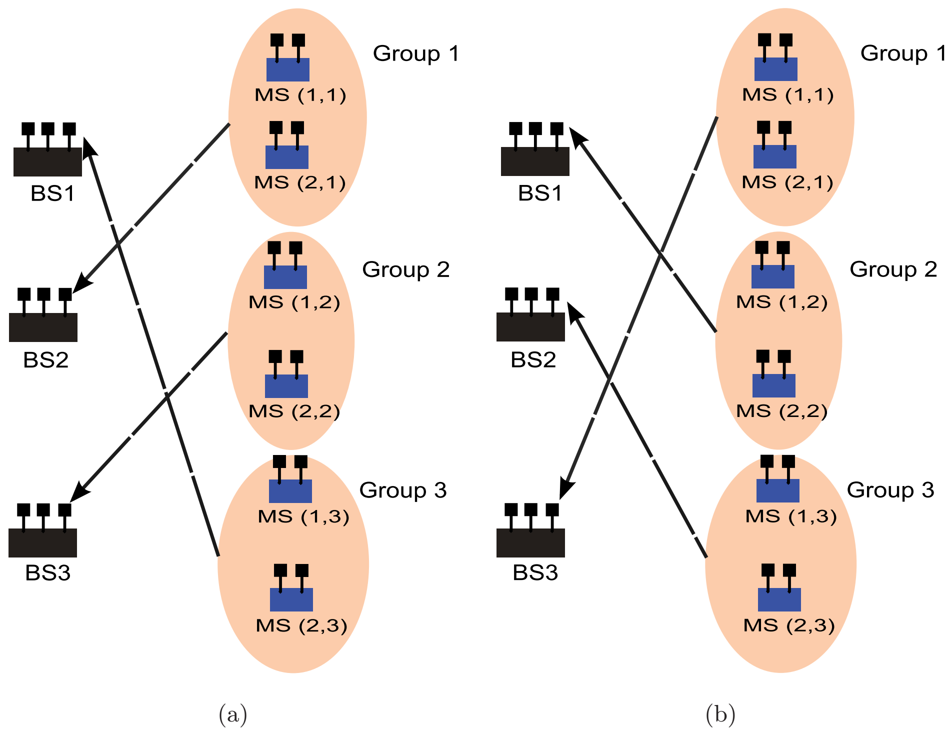


Figure 3.2: Strict IA-Cell assignment examples for $K = 3$ (two strict IA-Cell assignments in total): (a) The strict IA-Cell assignment: $Cell\ 1 \xrightarrow{IA} Cell\ 2 \xrightarrow{IA} Cell\ 3 \xrightarrow{IA} Cell\ 1$; (b) The strict IA-Cell assignment: $Cell\ 1 \xrightarrow{IA} Cell\ 3 \xrightarrow{IA} Cell\ 2 \xrightarrow{IA} Cell\ 1$.

Proof. Under an arbitrary strict IA-Cell assignment, the dimension of the space spanned by the interference to and from each BS is the same. Therefore, Corollary 1 is concluded in the homogeneous system. □

C. Effect of Assignment on Rate Performance

Different strict assignments have the same DoF, but they have different rate performance because the achievable rate (3.17) is determined by the effective channel $\widetilde{\boldsymbol{H}}_{i,k}^{k}\widetilde{\boldsymbol{H}}_{i,k}^{k,H}$ which highly depends on the IA-Cell assignment because $\boldsymbol{V}_{i,k}^{in}$ and $\boldsymbol{U}_{i,k}^{in}$ are thin matrices and could select multiple possible singular-values (or their combinations) of $\boldsymbol{H}_{i,k}^{k}$ in (3.18) and they are varying with the IA-Cell assignment.

Inspired by (3.18), each cell should have double preferences: the *IA-provider preference* and the *IA-receiver preference*, based on which each cell could find its most preferred IA-receiver and IA-provider. However, it is not possible to determine the optimal preferences before assignment because they are hardly *coupled*: 1) the preferences of one cell depend on other cells' assignment and 2) the IA-provider preference and IA-receiver preference of an individual cell depends on each other. Even if the approximate preferences are available, there is still a problem – how to balance the conflicts of multiple cells when some of them have the same preferred objective. In order to make the problem solvable and answer question *Q2*, we consider three scenarios with different practical constraints (e.g., different backhaul overhead and cooperation levels) and apply the stable matching and centralized assignment to obtain a stable or optimal strict IA-Cell assignment for each scenario.

As a desired criterion, the stability of the IA-Cell assignment can be defined as follows.

Definition 4. *(Stable Assignment) An IA-Cell assignment is* stable *if there does not exist a subset of cells consisting of more than one cell, in which the reassignment of IAs makes at least one cell better off but none worse off than their current assignment.*

3.3.2 One-Sided IA-Cell Matching

In this part, we consider the case when *no backhaul overhead is allowed between BSs before assignment.* In this case, each BS determines its assignment only based on its local CSIR.

A. Preference Generation

Since each BS k only knows its desired channels $\boldsymbol{H}_k^k$ and interference channels $\{\boldsymbol{H}_\ell^k\}_{\ell \neq k}$, it can compute $K-1$ *potential* IA precoders $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell \neq k}$ for the $K-1$ cells (*potential IA-providers*) based on $\{\boldsymbol{H}_\ell^k\}_{\ell \neq k}$. Under a strict IA-Cell assignment, each BS has only one IA-provider, and thus each BS k needs to rank the $K-1$ potential IA-providers by evaluating their corresponding interference subspace $\{\boldsymbol{F}_\ell^{IA\ k}\}_{\ell \neq k}$. However, each BS cannot construct the complete interference subspace because it does not know the IA precoders of all cells. Therefore, BS k cannot determine its IA-receiver preference but its IA-provider preference based on the $K-1$ potential aligned interference subspaces.

Let $\mathcal{P}_k^p$ with $K-1$ elements[4] arranged in decreasing order be the IA-provider preference list of BS k, i.e.,

$$\mathcal{P}_k^p = \text{arglist} \max_{\ell \neq k} \sum_{i=1}^{L} \log\det\left(\boldsymbol{I}_{N_U} + (\boldsymbol{H}_{i,k}^k)^H \boldsymbol{\Pi}_{\boldsymbol{F}_\ell^k}^{\perp} \boldsymbol{H}_{i,k}^k\right). \tag{3.19}$$

[4]Each BS has a single incomplete preference list, which excludes itself because it does not desire to become a lone cell.

The performance metric in (3.19) is to approximately measure the effect of the potential aligned interference subspace on the sum rate of cell k without knowledge of its own IA precoders.

B. Modified Residence Exchange Model based IA-Cell Matching

The one-sided matching is modeled by the *stable residence exchange model* [Yua96] in which K families wish to exchange their residences (actually, the right of renting, not the ownership of the residence) for a variety reasons. Each family has a move-in preference list consisting of up to K choices *with the last choice being its own residence without change*. The stable residence exchange demands that each family owns only one residence and each residence can only be rented by one family. This allocation involves a one-to-one matching between K families and K residences. Interpreting cells as families, IAs as residences, and IA exchange as residence exchange, our IA-Cell assignment will be well-matched to the stable residence exchange model if its *incomplete* preferences can be relaxed by allowing the existence of a lone cell. The only difference is the stable residence exchange with complete preferences but the IA-Cell assignment with *incomplete* preferences excluding itself.

a) Relaxation to Weak IA-Cell Assignment: First, we relax our strict IA-Cell assignment to the *weak IA-Cell assignment* by adding itself as the last candidate in the preference list of each BS. Then, the algorithm originally called the *Top Trading Cycle Method* in [SS74] and renamed as the *Forward Chaining Algorithm (FCA)* in [Yua96] always generates a unique stable solution for this weak IA-Cell assignment problem.

Before describing the FCA, we first introduce some basic aspects of the FCA. A chain represents subsequential IA exchange, e.g., $Cell\ k' \xrightarrow{IA} Cell\ k \xrightarrow{IA} Cell\ k''$ indicates that cell k receives IA from cell k' and cell k'' receives IA from cell k. A cycle represents a cycle chain. For instance, for the above chain, a cycle chain, denoted by $\langle Cell\ k, Cell\ k'\rangle$, is formed and $Cell\ k''$ is left if $Cell\ k \xrightarrow{IA} Cell\ k'$, or a cycle $\langle Cell\ k', Cell\ k, Cell\ k''\rangle$ is formed if $Cell\ k'' \xrightarrow{IA} Cell\ k'$. Once a cycle is formed, the IA exchange can be arranged according to the cycle. *The basic idea of the FCA is to let each cell sequently choose its current most preferred until a cycle chain is formed.* Each cycle chain is stable because its members find their *current* most preferred. The details are shown in [Yua96]. By the FCA, a stable matching of the K-cell weak IA-Cell assignment can be always obtained.

A *waiting list* is used to represent all the cells whose IA have not been provided yet, and an *arranged list* represents all the cells have been provided IA already. The FCA in the *weak IA-Cell assignment* works as follows.

Step 1. At the beginning, all cells are put into the waiting list and the arranged list is empty.

Step 2. Starting from one cell arbitrary picked (the order is irrelevant because the solution is unique [Yua96]) from the waiting list, a chain extends when each preceding cell in the chain tends to receive IA from a succeeding cell which is the best choice from its preference list. The chain will continue to grow until the last cell in the chain tends to receive IA from one already in the chain. In this case, a cycle is formed.

Step 3. The cycle will be removed from the chain and the left cells in the chain remain in the waiting list. The cells in the cycle will be removed from the waiting list and put into the arranged list, and they will be also removed from the IA-provider preference list of all other cells in the waiting list.

Step 4. Repeat **Step 2**-**Step 3** until the waiting list is empty.

Corollary 2. *For a K-cell weak IA-Cell assignment, a stable solution always exists and is unique; The solution generated by the FCA is stable; No cell can be better off by misrepresenting its true preferences, assuming other cells keep their preferences unchanged. Even when several cells try to collude by misrepresenting their true preferences, it is impossible to make at least one better off and none worse off among themselves.*

Proof. Refer to the references [SS74], [Yua96]. □

Corollary 3. *For a K-cell weak IA-Cell assignment, the stable matching by the FCA must belong to one of the two types: 1) no cell is lone cell; 2) only one cell is lone cell.*

Proof. This corollary can be easily proved by contradiction. Assume that there exist two lone cells. Since each cell has a complete IA-provider preference list where the cell itself is the last choice, these two lone cells surely prefer to exchange IA with each other rather than keep them. □

Remark 11. *If a stable matching for the weak IA-Cell assignment has no lone cell, this matching is also stable for the strict IA-Cell assignment. Otherwise, the strict IA-Cell assignment has no stable matching.*

b) "Almost Stable" Matching[5] by a Breaking Step: When the stable matching for the weak IA-Cell assignment has a lone cell, the $K-1$ coordinated cells find their preferred IA-providers and each achieves Ld_s DoF, but the lone cell with only d_s DoF may reject to join the cluster because its desired Ld_s DoF cannot be supported. This in return may degrade the $K-1$ coordinated cells' rate performance due to losing the

[5] For the assignment problem, if a stable matching does not exist, it is desired to match as many pairs as possible, i.e., to find a matching with maximum cardinality (so is "as stable as possible") [BMM10].

Table 3.1: A toy example of 4-cell assignment.

Cell	IA-Provider preference (utility)			
	1st (3)	2nd (2)	3rd (1)	4th (0)
1	3	2	4	1
2	1	3	4	2
3	2	1	4	3
4	1	2	3	4

spectrum or time resource shared by the lone cell. To circumvent this drawback, we modify the FCA algorithm by *allowing the possibility to break a cycle and insert the lone cell to form a new larger cycle (breaking step)* such that each cell achieves Ld_s DoF. In this case, an "almost stable" matching always has a better DoF performance than the matching with a lone cell. Additionally, it may also improve the sum-utility performance, as shown in the following toy example. In Table 3.1, by adding itself as the last preference (the forth preference), the original strict IA-Cell assignment is relaxed to a weak IA-Cell assignment, which yields a stable matching by the FCA: a cycle chain of $Cell\ 1 \xrightarrow{IA} Cell\ 3 \xrightarrow{IA} Cell\ 2 \xrightarrow{IA} Cell\ 1$ and $Cell\ 4 \xrightarrow{IA} Cell\ 4$. However, the stable matching does not exist in the strict IA-Cell assignment due to the existence of the lone cell $Cell\ 4$. By a breaking step, we insert the lone cell $Cell\ 4$ into the cycle chain $\langle Cell\ 1, Cell\ 3, Cell\ 2\rangle$, thereby forming an extended cycle $Cell\ 2 \xrightarrow{IA} Cell\ 4 \xrightarrow{IA} Cell\ 1 \xrightarrow{IA} Cell\ 3 \xrightarrow{IA} Cell\ 2$. This "almost stable" assignment $\langle Cell\ 1, Cell\ 3, Cell\ 2, Cell\ 4\rangle$ with sum utility of $1+3+3+3 = 10$ and $4Ld_s$ sum DoF outperforms the original matching $\langle Cell\ 1, Cell\ 3, Cell\ 2\rangle$ and $\langle Cell\ 4\rangle$ by the FCA only with the sum utility of $3+3+3+0 = 9$ and with $(3L+1)d_s$ DoF.

3.3.3 Two-Sided IA-Cell Matching

In this section, we consider another case when *low backhaul overhead is permitted before assignment*. By the GIA, each BS k can compute $K-1$ *potential* inner precoders $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell\neq k}$ for all the other cells based on $\{\boldsymbol{H}_\ell^k\}_{\ell\neq k}$, and then BS k reports the potential inner precoders to the corresponding BSs via backhaul links, e.g., sending $\boldsymbol{V}_{k'}^{in}(k)$ to BS k'.

A. Preferences Generation

In this case, each cell not only knows the potential aligned interference subspace $\{\overline{\boldsymbol{F}}_\ell^k\}_{\ell\neq k}$ (corresponding to the potential IA-providers) but also its *potential* inner precoders

$\{\boldsymbol{V}_k^{in}(k')\}_{k'\neq k}$ (corresponding to the potential IA-receivers). It is possible for each cell to compute double preferences for its IA-provider and IA-receiver.

Let $\mathcal{P}_k^p$ and $\mathcal{P}_k^r$ be the IA-provider preference list and IA-receiver preference list, and both are *incomplete preferences* with $K-1$ elements. More precisely, $\mathcal{P}_k^p$ and $\mathcal{P}_k^r$ can be generated by (3.19) and

$$\mathcal{P}_k^r = \text{arglist} \max_{\ell\neq k} \sum_{i=1}^{L} \log\det\left(\boldsymbol{I}_{d_s} + \boldsymbol{V}_{i,k}^H(\ell)\boldsymbol{H}_{i,k}^{k,H}\boldsymbol{H}_{i,k}^k\boldsymbol{V}_{i,k}(\ell)\right). \tag{3.20}$$

B. Stable Marriage Model based IA-Cell Matching

In this two-sided IA-Cell matching, each cell hopes to find its most preferred IA-provider and IA-receiver, respectively. To balance the possible preferences conflicting, the two-sided matching is required to determine a stable matching. In this case, the problem is well modeled by the well-known *stable marriage matching with unacceptable partners* [GI89] by considering each MS group and BS as a man and a woman (or reversely), respectively. Based on [GI89, Theorem 1.4.2], the following result holds.

Corollary 4. *Consider the strict IA-Cell assignment where MS group k and BS k are unacceptable to each other. The stable matching may not exist (only one pair of MS group and BS in a cell is not matched.) but is stable if it exists.*

To obtain the stable matching, following the same line of the one-sided matching, the strict two-sided IA-Cell assignment problem is first relaxed to a weak two-sided IA-Cell assignment problem. If the strict IA-Cell assignment has a stable matching, it can be efficiently determined by the basic Gale-Shapley algorithm [GS62]. Otherwise, an "almost stable" matching can be obtained by a further breaking step.

We remark that an assignment by either the one-sided or two-sided stable matching scheme does not necessarily maximize the sum-cluster rate or the single-cell rate, since the goal in distributed assignment is to find a stable matching if it exists. Otherwise, an almost stable matching if it does not exist.

3.3.4 Centralized IA-Cell Assignment

Now we consider the case when *there exists a central authority*[6] *and high backhaul overhead is permitted.* Without loss of generality, we assume BS k serves as the cluster head and performs the assignment for all cells. Each BS $k', \forall k' \neq k$ sends the $K-1$

[6]In the case of cellular networks this authority could be either a central controller (e.g., the Cloud-RAN) or a BS who serves as the cluster head and does the centralized optimization for the network. In particular, the cluster head could be a fixed one or a rotating one.

Algorithm 5 Basic Gale-Shapley algorithm [GS62] for two-sided IA-Cell assignment

Input: K cells, complete preference, low backhaul overhead.

Output: a stable K-cell IA-Cell matching.

while *some MS group k is available* **do**
- Assume BS ℓ be the first one in $\mathcal{P}_k^{rec}$ to whom MS group k has not yet asked;
- **if** *BS ℓ is free* **then**
  - assign BS ℓ as the IA-receiver for MS group k;
- **else**
  - **if** *BS ℓ prefers MS group k to its current IA-provider k'* **then**
    - assign BS ℓ as the IA-receiver for MS group k and MS group k' becomes available;
  - **else**
    - BS ℓ rejects MS group k and MS group k remains available.

potential IA precoders $\{\boldsymbol{V}_\ell(k')\}_{\ell \neq k'}$ and the direct channel matrices $\boldsymbol{H}_{k'}^{k'}$ to BS k. Then, the optimal assignment for a certain problem, e.g., sum-cluster rate maximization or minimum single-cell rate maximization, can be determined by BS k by the brute-force search and based on the collected information. Afterwards, BS k releases the assignment result to the cluster members. Also note that this optimal assignment is not guaranteed to be stable.

Remark 12. *From Lemma 3, there are few derangements for the cluster with a small number of cells, e.g.,* 2 *derangements for* $K = 3$ *and* 8 *derangements for* $K = 4$. *In this case, the brute-force search is a reasonable approach. However, as* K *increases, the number of derangements increases significantly, e.g.,* 264 *derangements for* $K = 6$, *and the resulting backhaul overhead and the computational load become too large.*

3.4 Limited Feedback: Dynamic Feedback Bit Allocation

Given an IA-Cell assignment, each BS k obtains from its IA-provider its own IA precoder $\boldsymbol{V}_k^{in}$. Let $\overrightarrow{\boldsymbol{V}}_{i,k} \triangleq \boldsymbol{T}_i \boldsymbol{V}_k (\boldsymbol{V}_k^H \boldsymbol{T}_i^H \boldsymbol{T}_i \boldsymbol{V}_k)^{-1}$ be the *precoder pattern* in (3.13) where $\overrightarrow{\boldsymbol{V}}_{i,k}^H \overrightarrow{\boldsymbol{V}}_{i,k} = \boldsymbol{I}_{d_s}$. In order to implement a closed-loop transmission, $\overrightarrow{\boldsymbol{V}}_{i,k}$ needs to be fed back to MS (i,k). Since feedback links are usually capacity-limited, subspace quantization is employed to reduce overhead. A subspace matrix is mapped to an index in a predefined codebook. However, the feedback of an index results in the residual interference, since the interference not be perfectly aligned because of the existence of quantization distortion. Therefore, the problem of DBA to minimize sum-cluster RINR is of interest.

3.4.1 Grassmannian subspace quantization

Due to $\overrightarrow{\boldsymbol{V}}_{i,k}^{H}\overrightarrow{\boldsymbol{V}}_{i,k} = \boldsymbol{I}_{d_s}, \forall i, k$, subspace quantization can be applied to quantize the precoder patterns. Here, we give a subspace quantization example of a subspace matrix $\boldsymbol{V} \in \mathbb{C}^{M\times N}$ where $M > N$ by B feedback bits. Assume that both the BS and MS know the common codebook $\mathcal{C}$, i.e.,

$$\mathcal{C} = \{\boldsymbol{C}_n \in \mathbb{C}^{M\times N} : \ \boldsymbol{C}_n^H \boldsymbol{C}_n = \boldsymbol{I}_N, n = 1, \ldots, 2^B\}, \tag{3.21}$$

which can be generated and stored offline. The quantized subspace is determined as the closest codeword in $\mathcal{C}$ by measuring the *chordal distance*

$$\begin{aligned}\widehat{\boldsymbol{V}} &\triangleq \arg\min_{\boldsymbol{C}_n\in\mathcal{C}} d_c^2(\boldsymbol{C}_n, \boldsymbol{V}) \\ &= \arg\min_{\boldsymbol{C}_n\in\mathcal{C}} M - \mathrm{Tr}(\boldsymbol{C}_n\boldsymbol{C}_n^H \boldsymbol{V}\boldsymbol{V}^H).\end{aligned} \tag{3.22}$$

The considered quantization is well-known as Grassmannian quantization on the Grassmann manifold $\mathcal{G}(M, N)$, defined as the set of the N-dimensional subspaces in the M-dimensional complex Euclidean space. The optimal Grassmann codebook $\mathcal{C}$ is designed based on the *Grassmannian subspace packing: to find 2^B subspace matrices on $\mathcal{G}(M, N)$ by maximizing the minimum pairwise subspace distance*. However, it is challenging to generate an optimal Grassmann codebook, which has attracted many research efforts [LH05, AG07, SJW09, MD14] and references therein.

Lemma 4. *(Quantized Subspace Characterization) The quantization $\widehat{\boldsymbol{V}} \in \mathbb{C}^{M\times N}$ of the subspace $\boldsymbol{V} \in \mathbb{C}^{M\times N}$ based on the subspace quantization can be characterized as*

$$\widehat{\boldsymbol{V}} = \boldsymbol{V}\boldsymbol{R}\boldsymbol{\Gamma}^{1/2}\boldsymbol{G}^H + \boldsymbol{V}^{\perp}\boldsymbol{S}(\boldsymbol{I}_N - \boldsymbol{\Gamma})^{1/2}\boldsymbol{G}^H \tag{3.23}$$

where $\boldsymbol{V}^{\perp} \in \mathbb{C}^{M\times(M-N)}$ spans the null space of $\boldsymbol{V}$, and $\boldsymbol{\Gamma} \triangleq \mathrm{diag}\{\alpha_1, \ldots, \alpha_N\}$ where $\alpha_j \in [0, 1]$ and $\sum_{j=1}^{N} \alpha_j = N - d_c^2(\widehat{\boldsymbol{V}}, \boldsymbol{V})$, and $\boldsymbol{R} \in \mathbb{C}^{N\times N}$, $\boldsymbol{G} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{S} \in \mathbb{C}^{(M-N)\times N}$ satisfy $\boldsymbol{R}^H\boldsymbol{R} = \boldsymbol{G}^H\boldsymbol{G} = \boldsymbol{S}^H\boldsymbol{S} = \boldsymbol{I}_N$.

Proof. Refer to Proof 3.8.2. □

Remark 13. *Since popular performance metrics, such as transmit power, minimum square error (MSE) and achievable rate, are functions of $\hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^H$, the quantization characterization in (3.23) can be further simplified to*

$$\widehat{\boldsymbol{V}} = \boldsymbol{V}\boldsymbol{R}\boldsymbol{\Gamma}^{1/2} + \boldsymbol{V}^{\perp}\boldsymbol{S}(\boldsymbol{I}_N - \boldsymbol{\Gamma})^{1/2}, \tag{3.24}$$

because $\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^H$ is independent of the unitary matrix $\boldsymbol{G}$ in (3.23). This quantized subspace characterization in (3.24) is more efficient than that in [RJ08, Lemma 1] where $\boldsymbol{\Gamma}^{1/2}$ is an upper triangular matrix derived based on QR decomposition instead of a diagonal matrix as in our formulation.

The quantization distortion in the Grassmannian subspace quantization problem (3.22) is defined by $d_c^2(\widehat{\boldsymbol{V}}, \boldsymbol{V})$, and its upper bound is derived in [DLR08] as

$$d_c^2(\widehat{\boldsymbol{V}}, \boldsymbol{V}) \leq c(M, N)2^{-\frac{B}{N(M-N)}}, \tag{3.25}$$

where $c(M, N)$ is a function of M and N as specified in [DLR08, Eqs. (8) and (11)] by omitting the $\mathcal{O}(1)$ term in [DLR08, Eq. (11)].

3.4.2 Dynamic IA Precoders Quantization and Feedback

By the Grassmannian subspace quantization in (3.22), each subspace matrix $\overrightarrow{\boldsymbol{V}}_{i,k}$ can be expressed by an index, which will be sent to MS (i, k) through the limited feedback link. Let $B_{i,k}$ denote the feedback bit for $\overrightarrow{\boldsymbol{V}}_{i,k}$ subject to a sum feedback bits constraint $\sum_{k=1}^{K}\sum_{i=1}^{L} B_{i,k} \leq B$.

Consider an IA-Cell assignment *Cell* $k' \xrightarrow{IA}$ *Cell* k. After subspace quantization and feedback of $\{\overrightarrow{\boldsymbol{V}}_{i,k'}\}_{i=1}^{L}$, the interference from cell k' to cell k with the quantized precoder pattern $\{\widehat{\boldsymbol{V}}_{i,k'}\}_{i=1}^{L}$, denoted by $\widehat{\boldsymbol{F}}_{k'}^{k}$, cannot be perfectly aligned into a d_s-dimensional subspace. The imperfectly aligned interference spreads into a higher dimensional subspace, which cannot be completely removed by the ZF decoding. Thus, residual interference exists.

The RINR from cell k' to cell k is defined as

$$\mathcal{I}_{k'}^{k} \triangleq \sum_{i=1}^{L} \frac{P_{i,k}}{d_s\sigma_k^2} \operatorname{Tr}(\widehat{\boldsymbol{U}}_{i,k}^{H}\boldsymbol{H}_{i,k'}^{k}\widehat{\boldsymbol{V}}_{i,k'}\widehat{\boldsymbol{V}}_{i,k'}^{H}\boldsymbol{H}_{i,k'}^{k,H}\widehat{\boldsymbol{U}}_{i,k}), \tag{3.26}$$

where the decoder $\widehat{\boldsymbol{U}}_{i,k}$ is designed as

$$\widehat{\boldsymbol{U}}_{i,k} \triangleq \left(\left[\widehat{\boldsymbol{F}}_{j,k}^{IUI},\ \left\{\widehat{\boldsymbol{F}}_{\ell}^{k}\right\}_{\ell=1,\ell\neq k'}^{K},\ \boldsymbol{H}_{i,k'}^{k}\boldsymbol{V}_{i,k'}^{in}\right]\right)^{\perp}, \tag{3.27}$$

by which the interference from other cells $\ell \neq k'$ (not the IA-provider of cell k) can be removed at BS k.

Proposition 9. *Let $\mathcal{I}^k$ denote the RINR at BS k. Without loss of generality, we consider the IA-Cell assignment Cell $\ell \xrightarrow{IA}$ Cell k. $\mathcal{I}^k$ is upper bounded by*

$$\mathcal{I}^k \leq \overline{\mathcal{I}}^k \triangleq c(N_U, d_s)\sum_{i=1}^{L}\frac{P_{i,\ell}}{\sigma_k^2 d_s}\lambda_1(\boldsymbol{\Omega}_{i,\ell}^{k})2^{-\frac{B_{i,\ell}}{ds(N_U-d_s)}}, \tag{3.28}$$

where

$$\mathbf{\Omega}_{i,\ell}^{k} \triangleq \left(\boldsymbol{V}_{i,\ell}^{in,\perp}\right)^{H} \boldsymbol{H}_{i,\ell}^{k,H} \mathbf{\Pi}_{\boldsymbol{H}_{1,\ell}^{k}\boldsymbol{V}_{1,\ell}^{in}}^{\perp} \boldsymbol{H}_{i,\ell}^{k} \boldsymbol{V}_{i,\ell}^{in,\perp}. \tag{3.29}$$

Proof. Refer to Proof 3.8.1. □

In order to efficiently exploit the limited feedback bits, it is desired to allocate the feedback bit to reduce the residual interference. Therefore, a dynamic feedback bit allocation problem is studied.

3.4.3 Dynamic Feedback Bit Allocation for Precoders

In this section, a DBA algorithm is studied to minimize the upper bound on the sum-cluster RINR.

$$\begin{aligned} \min_{\{\{B_{i,k}\}_{i=1}^{L}\}_{k=1}^{K}} \quad & \sum_{k=1}^{K} \overline{\mathcal{I}}^{k} \\ \text{s.t.} \quad & \sum_{k=1}^{K}\sum_{i=1}^{L} B_{i,k} \leq B;\ \forall B_{i,k} \in \mathbb{N}_0^{+} \end{aligned} \tag{3.30}$$

where $\overline{\mathcal{I}}^{k}$ is given in (3.28). Observe that Problem (3.30) is a jointly convex problem of $\{B_{i,k}\}$ when the non-negative integer constraint is relaxed and yields the following solutions.

Proposition 10. *(Bit Allocation Solution) Let us define*

$$\boldsymbol{a} \triangleq \operatorname{arglist} \max_{\forall i;\forall k} \{\{\log_2(\lambda_1(\mathbf{\Omega}_{i,k}^{k+1}))\}_{i=1}^{L}\}_{k=1}^{K}. \tag{3.31}$$

Given an arbitrary B, the number of active MSs whose allocated feedback bit is positive can be determined by checking

$$\sum_{n=1}^{N_a} \boldsymbol{a}(n) - N_a\boldsymbol{a}(N_a) \leq \frac{B}{d_s(N_U - d_s)} \leq \sum_{n=1}^{N_a} \boldsymbol{a}(n) - N_a\boldsymbol{a}(N_a+1), \tag{3.32}$$

where $N_a \in \{1, \ldots, KL\}$ denotes the number of active MSs. After determining N_a, the optimal solution for the N_a active MSs in Problem (3.30) is given in closed-form by

$$B_{i,k}^{\star} = \Big[d_s(N_U - d_s)\Big(\log_2(\lambda_1(\mathbf{\Omega}_{i,k}^{k+1})) - \frac{1}{N_a}\sum_{n=1}^{N_a}\boldsymbol{a}(n) + \frac{B}{N_a d_s(N_U - d_s)}\Big)\Big]_{int}. \tag{3.33}$$

And no feedback bits is allocated to those inactive MSs.

Proof. The Lagrangian function with multiplier μ for Problem (3.30) can be formulated as

$$\mathcal{L}(\{\{B_{i,k}\}_{i=1}^{L}\}_{k=1}^{K}, \mu) = \sum_{k=1}^{K}\sum_{i=1}^{L} \lambda_1(\mathbf{\Omega}_{i,k}^{k+1}) 2^{-\frac{B_{i,k}}{d_s(N_U - d_s)}} + \mu\Big(\sum_{k=1}^{K}\sum_{i=1}^{L} B_{i,k} - B\Big). \tag{3.34}$$

With the definition $\zeta \triangleq \frac{d_s(N_U-d_s)}{\ln 2}\mu$, the KKT conditions are

$$\frac{\partial \mathcal{L}}{\partial B_k} = -\lambda_1(\mathbf{\Omega}_{i,k}^{k+1})2^{-\frac{B_{i,k}}{d_s(N_U-d_s)}} + \zeta = 0 \tag{3.35}$$

$$\frac{\partial \mathcal{L}}{\partial \zeta} = \sum_{k=1}^{K}\sum_{i=1}^{L} B_{i,k} - B = 0; \quad \zeta > 0, \tag{3.36}$$

From (3.35)-(3.36), we derive

$$B_{i,k}(\zeta) = d_s(N_U - d_s)(\log_2(\lambda_1(\mathbf{\Omega}_{i,k}^{k+1})) - \log_2(\zeta)), \tag{3.37}$$

where ζ is determined such that $\sum_{k=1}^{K}\sum_{i=1}^{L} B_{i,k}(\zeta) = B$. Combining that $B_{i,k}$ is a nonnegative integer, we have

$$B_{i,k}^{\star} = [d_s(N_U - d_s)(\log_2(\lambda_1(\mathbf{\Omega}_{i,k}^{k+1})) - \log_2(\zeta))]_{int}^{+}, \tag{3.38}$$

where ζ satisfies $\sum_{k=1}^{K}\sum_{i=1}^{L} B_{i,k}^{\star} = B$.

To obtain the closed-form expression without variable ζ, the water-filling principle implies that only the active MSs are allocated to the positive feedback bits. If there are N_a active MSs where $N_a \in \{1, \ldots, KL\}$, with the definition in (3.31), the water-level satisfies

$$\boldsymbol{a}(N_a + 1) \leq \log_2(\zeta) \leq \boldsymbol{a}(N_a). \tag{3.39}$$

In the case of (3.39), plugging (3.38) into (3.36) yields

$$\log_2(\zeta) = \frac{1}{N_a}\sum_{n=1}^{N_a} \boldsymbol{a}(n) - \frac{B}{N_a d_s(N_U - d_s)}. \tag{3.40}$$

Again plugging (3.40) into (3.37) yelids (3.33) under the condition (3.32) that is obtained by combining (3.40) and (3.39). There are KL cases, i.e., $n \in \{1, \ldots, KL\}$. Given a feedback bit budget B, we can determine how many and which MSs are active by checking (3.32) and thus the closed-form bit allocation in (3.33). □

3.5 Implementation and Analysis

In this section, the proposed algorithm is analyzed in the following aspects: 1) implementation, 2) required overhead and 3) complexity.

3.5.1 Implementation

The outline of the implementation of the proposed algorithm is shown as follows, where each step could be a time slot in system operation.

- **Step 1. (CSIR estimation)**: Each BS k estimates its local CSIR $\{\boldsymbol{H}_\ell^k\}_{\ell=1}^K$ based on orthogonal uplink pilot signals;

- **Step 2. (IA percoder computation)**: Each BS k employs the GIA method to compute $K-1$ *potential* IA precoders $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell=1,\ell\neq k}^K$ for $K-1$ cells based on $\{\boldsymbol{H}_\ell^k\}_{\ell=1,\ell\neq k}^K$;

- **Step 3. (IA-Cell assignment)**:
  - *With no Backhaul Overhead Before Assignment (Distributed):* Based on only the local CSIR, each BS k computes $K-1$ possible IA strategies $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell=1,\ell\neq k}^K$ for $K-1$ potential IA-providers. One-sided matching is implemented based on $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell=1,\ell\neq k}^K$;
  - *With low Backhaul Overhead Before Assignment (Distributed):* Each BS k not only computes $K-1$ possible IA strategies $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell=1,\ell\neq k}^K$ based on its local CSIR, but also reports its computed $\{\boldsymbol{V}_\ell^{in}(k)\}_{\ell=1,\ell\neq k}^K$ to the $K-1$ corresponding BSs. Based on the collected possible IA precoders and its local CSIR, two-sided matching is implemented;
  - *With high Backhaul Overhead Before Assignment (Centralized):* Assume that BS k is the cluster head. Each BS $k' \neq k$ reports its computed $\{\boldsymbol{V}_\ell^{in}(k')\}_{\ell=1,\ell\neq k'}^K$ and its direct channels $\boldsymbol{H}_{k'}^k$ to the cluster head BS k via backhaul links. Based on the collected informations, BS k finds the optimal assignment for an arbitrary problem by the brute force search and then tells the assignment to BSs;

- **Step 4. (DBA)**: After assignment, each BS k sends the perfect IA precoder to its IA-receiver via backhaul links. Then, each BS k needs to feed back its IA precoder $\boldsymbol{V}_{i,k}^{in}$ to each MS (i,k) (broadcast feedback). In order to enable efficient feedback of $\{\{\overrightarrow{\boldsymbol{V}}_{i,k}\}_{i=1}^L\}_{k=1}^K$, the DBA is optimized by minimizing the sum RINR and yields the solution $\{\{B_{i,k}\}_{i=1}^L\}_{k=1}^K$ for the quantization of KL precoder patterns;

- **Step 5. (Quantization under limited feedback)**: Each BS k quantizes the precoder patterns $\{\overrightarrow{\boldsymbol{V}}_{i,k}\}_{i=1}^L$ to $\{\widehat{\boldsymbol{V}}_{i,k}\}_{i=1}^L$ by Grassmannian subspace codebooks with size $\{2^{B_{i,k}}\}_{i=1}^L$ and broadcasts the indexes to its MSs;

Table 3.2: Total backhaul overhead of K cells.

Algorithms	Before assignment	Assignment	After assignment
One-sided	0	$4(K+(N_C-1))$ bit	KLN_Ud_s cc $+(K-1)B$ bit
Two-sided	$K(K-1)LN_Ud_s$ cc	$4[K, K^2-K+1]$ bit	$(K-1)B$ bit
Centralized	$\begin{smallmatrix}(K-1)^2LN_Ud_s+\\(K-1)LN_UN_B\end{smallmatrix}$ cc	0	$(K-1)LN_Ud_s$ cc $+(K-1)B$ bit
Fixed	0	–	KLN_Ud_s cc

1) "cc" denotes the unit of a complex coefficient. 2) Each ask is responsed during the assignment.

- **Step 6. Data Transmission in the Uplink**: Each MS (i,k) selects one codeword in the codebook as its precoder pattern $\overrightarrow{\boldsymbol{V}}_{i,k}$ based on its received index. MS (i,k) transmits a $d_s \times 1$ data by the precoder $\sqrt{\frac{P_{i,k}}{d_s}}\overrightarrow{\boldsymbol{V}}_{i,k}$ to BS k, which will be decoded by $\widehat{\boldsymbol{U}}_{i,k}$ in (3.27).

3.5.2 Backhaul overhead

Considering the different IA-Cell assignment schemes, their required backhaul overhead (excluding the feedback overhead) are summarized in Table 3.2, where "One-sided"/"Two-sided"/"Centralized"/"Fixed" denotes the IA-Cell assignment by the one-sided/two-sided/centralized/fixed matching.

During the IA-Cell assignment by the one/two-sided matching, if BS k' asks, definitely accepts, temporarily accepts or definitely rejects BS k, it will send "01", "11", "10" or "00" to BS k, respectively, each of which can be conveyed into a QPSK symbol. In particular, the one-sided matching by the FCA takes $K+(N_C-1)$ steps where N_C denotes the number of cycle chains, and each step has one ask. while the two-sided matching takes $[1, K]$ rounds by allowing K cells to propose simultaneously. The two-sided matching by the Basic Gale-Shapley algorithm [GS62] takes $[K, K(K-1)+1]$ proposals. Consequently, the total overhead of the two/one-sided matching during the assignment are $4[K, K(K-1)+1]$, which also includes the amount of the response for each proposal. After assignment by the one-sided matching, each BS needs to send an explicit inner precoder to its corresponding IA-provider via backhaul links, while it is not necessary for the two-sided matching because it has been already exchanged before assignment. After the quantization of the precoder patterns, each BS needs to exchange the corresponding indexes with others, based on which a new ZF decoder can be designed. For the centralized assignment, we assume cell k' serves as the cluster head. BS k' first collects the information from all other cells before assignment, and tells cell $\forall k, k \neq k'$ the label of its IA-provider after doing the assignment. The resulting total backhaul overhead for these approaches is reported in Table 3.2.

3.5.3 Complexity

The complexity of the proposed GIA algorithm with three matching approaches are shown and compared as follows.

As shown in Chapter 3.2, the complexity of computing K IA precoders by the GIA is $K\mathcal{O}((L-1)^2 L N_B^2 N_U)$.

For the one-sided matching, the complexity mainly depends on the preference generation as (3.19). The generation of K *ranked* preference lists takes $K(K-1)L(\mathcal{O}(N_B N_U d_s)+2L(\mathcal{O}(N_U N_B^2)+\mathcal{O}(N_U^3))+2\mathcal{O}(N_B d_s^2)+2\mathcal{O}(d_s^3)+K\mathcal{O}(K))$ arithmetic operations. The FCA with $K+(N_C-1)$ steps has the complexity $\mathcal{O}(K)$ where N_C denotes the number of cycle chains. For the two-sided matching, besides generating (3.19) with the same complexity as the one-sided matching, the generation of K *ranked* preference lists as (3.20) requires $K(K-1)L\left(\mathcal{O}(2N_U^2 d_s)+\mathcal{O}(N_U N_B d_s)+2(\mathcal{O}(d_s^3))+K\mathcal{O}(K)\right.$ arithmetic operations. The complexity of the Basic Gale-Shapley algorithm with at most K^2-K+1 steps is upper bounded by $\mathcal{O}(K^2)$. The brute-force search in the centralized assignment needs to compute $K!\sum_{k=0}^{K}\frac{(-1)^k}{k!}-1$ possible rate performance with the complexity $(K!\sum_{k=0}^{K}\frac{(-1)^k}{k!}K-1)(L(\mathcal{O}(2N_U^2 d_s)+\mathcal{O}(N_B^2 d_s)+(L+1)\mathcal{O}(N_B N_U d_s)+(L+2)\mathcal{O}(d_s^3)+(L+2)\mathcal{O}(N_B d_s^2)))$.

Roughly speaking, the one-sided matching, the two-sided matching and the centralized assignment mainly take $K(K-1)L$, $2K(K-1)L$ and $K!\sum_{k=0}^{K}\frac{(-1)^k}{k!}-1$ "rate-like" computations[7] , respectively. Figure 3.3 shows the gross complexity of these three algorithms over the number of cells. It implies that the centralized assignment is a reasonable approach with a comparable complexity as the distributed algorithms if $K\leq 4$. Instead, when $K\geq 5$ distributed algorithms are preferable from the perspective of complexity.

3.6 Illustrations & Discussions

In this section, the performance of the GIA with optimized IA-Cell assignment under both the unlimited and limited feedback is evaluated.

[7] The computation expression is not the actual rate expression, but has the form $\log\det(\boldsymbol{I}+\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{Y}}^{\perp}\boldsymbol{X}^H)$.

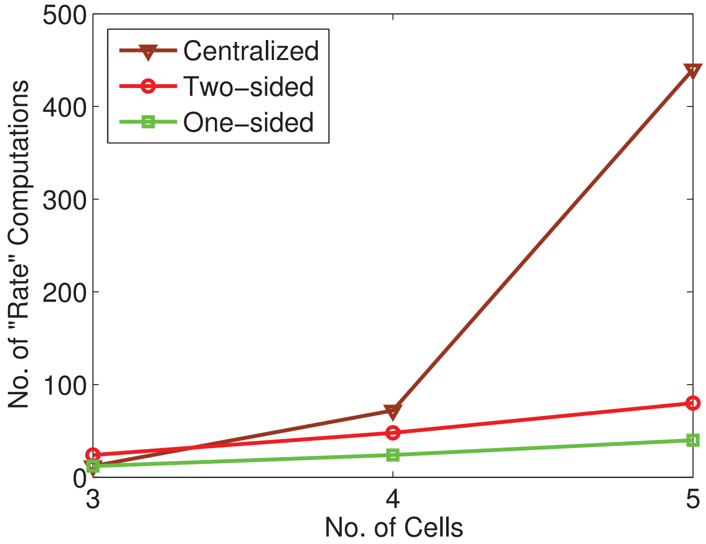


Figure 3.3: Complexity comparison of the distributed assignments and the centralized assignment.

3.6.1 System Model and Performance Metrics

We consider a $(K, L, N_B, N_U, d_s) = (4, 2, 14, 8, 2)$ MIMO-IMAC. The noise power is normalized to be $\sigma_k^2 = 1, \forall k$, and each MS is set to the same transmit power$P_{i,k} = P, \forall i, k$. Let $\text{SNR} = 10\log_{10}(P)$ denote the transmit SNR. The Rayleigh fading channel is adopted. The path loss of direct links are normalized to be 1, while the path loss of the cross links are set to satisfy randomly and uniformly distribution in $[0, 1]$, respectively.[8]

By treating residual interference as additive noise, we define the throughput under limited feedback of MS (i, k) as [RRL13]

$$\widehat{R}_{i,k} = \ln\det\Big(\boldsymbol{I}_{d_s} + \frac{\text{SNR}}{d_s}(\widehat{\boldsymbol{U}}_{i,k}^H \boldsymbol{H}_{i,k}^k \widehat{\boldsymbol{V}}_{i,k})(\widehat{\boldsymbol{U}}_{i,k}^H \boldsymbol{H}_{i,k}^k \widehat{\boldsymbol{V}}_{i,k})^H (\boldsymbol{I}_{d_s} + \boldsymbol{C}_{i,k})^{-1}\Big), \tag{3.41}$$

where $\boldsymbol{C}_{i,k} \triangleq \frac{\text{SNR}}{d_s} \sum_{(j,\ell)\neq(i,k)} \widehat{\boldsymbol{U}}_{i,k}^H \boldsymbol{H}_{j,\ell}^k \widehat{\boldsymbol{V}}_{j,\ell} (\widehat{\boldsymbol{U}}_{i,k}^H \boldsymbol{H}_{j,\ell}^k \widehat{\boldsymbol{V}}_{j,\ell})^H$ denotes the overall residual interference matrix of the MS (i, k). In the unlimited feedback case, the interference can be removed completely and (3.41) is the same as (3.3).

To properly measure the performance of the proposed approaches, we consider two following metrics in the Monte-Carlo simulations,

$$R_{sum} \triangleq \mathbb{E}\Big(\sum_{k=1}^{K}\sum_{i=1}^{L} \widehat{R}_{i,k}\Big), \quad \text{and} \quad R_{min} \triangleq \mathbb{E}\Big(\min_{k=1,\ldots,K} \sum_{i=1}^{L} \widehat{R}_{i,k}\Big),$$

[8]This is to guarantee that interference channels are not stronger than direct channels, since a MS is usually assigned to the BS who provides it the strongest link. The MS-selection and MS-BS association can be done based on the uplink CSI available at BSs.

where $\widehat{R}_{i,k}$ is given in (3.41). R_{sum} and R_{min} are the average sum-cluster rate and the average minimum single-cell rate over different channel realizations to measure the overall cluster throughput and the fairness of the cluster, respectively.

3.6.2 Performance Comparison with Unlimited Feedback

Under unlimited feedback, the effect of IA-Cell assignment on R_{sum} and R_{min} is evaluated by the following metrics.

- $\mathrm{Upper_{sum}}$ and $\mathrm{Lower_{sum}}$ ($\mathrm{Upper_{min}}$ and $\mathrm{Lower_{min}}$) denote the performance achieved by *the best* and *the worst* IA-Cell assignment for *sum cluster-rate maximization*, R_{sum} (*minimum cluster-rate maximization*, R_{min}), respectively, which are determined by the centralized assignment;

- Two/One/Fixed: Each channel realization is under the IA-Cell assignment by the two-sided/one-sided/fixed matching in (3.6);

- RB: Each channel realization is with a precoder-decoder pair: $(\sqrt{\frac{P_{i,k}}{d_s}}\boldsymbol{V}_{i,k}^{RB}, \boldsymbol{U}_{i,k}^{MF})$, where $\boldsymbol{V}_{i,k}^{RB}$ is a randomly selected subspace matrix satisfying $\boldsymbol{V}_{i,k}^{RB,H}\boldsymbol{V}_{i,k}^{RB} = \boldsymbol{I}_{d_s}$ and $\boldsymbol{U}_{i,k}^{MF}$ is set as the "matched filter" $\boldsymbol{U}_{i,k} = \boldsymbol{H}_{i,k}^{k}\boldsymbol{V}_{i,k}(\boldsymbol{V}_{i,k}^{H}\boldsymbol{H}_{i,k}^{k,H}\boldsymbol{H}_{i,k}^{k}\boldsymbol{V}_{i,k})^{-\frac{1}{2}}$;

- FDMA: Each MS occupies an un-overlapped spectrum in the uplink transmission (interference free).

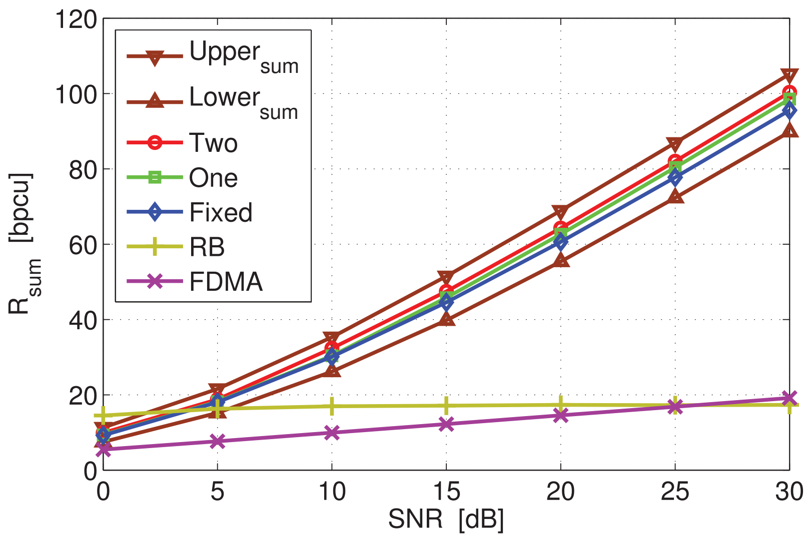


Figure 3.4: Sum-cluster rate comparison under unlimited feedback w.r.t. SNR.

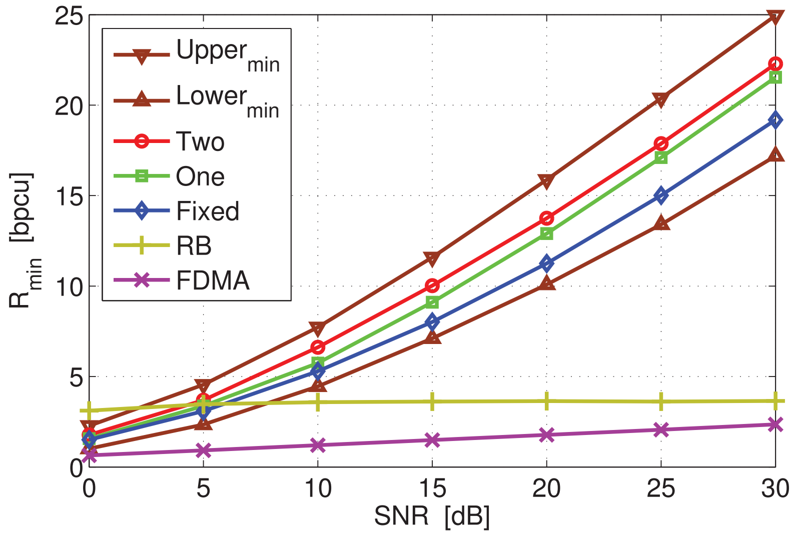


Figure 3.5: Minimum single-cell rate comparison under unlimited feedback w.r.t. SNR.

Both Figure 3.4 and Figure 3.5 show that a large performance gap exists between the best IA-Cell assignment and the worst IA-Cell assignment. It implies the IA-Cell assignment has a significant influence on both the overall throughput and the fairness. E.g., this performance gap regarding R_{sum} is as large as 5 dB and that of R_{min} is even larger than 10 dB for high SNR. Compared with the fixed matching, the two-sided and one-sided matching have a similar performance improvement, i.e., more than 1 dB for R_{sum} and more than 5 dB for R_{min}. The advantage of the GIA is significant compared with the random beamforming and FDMA, especially for high SNR.

3.6.3 Performance Comparison under Limited Feedback

Under limited feedback, the proposed DBA is evaluated by comparing with the classical EBA (plotted in dashed lines in the following figures). The random subspace codebook is adopted in the quantization.

A. Performance Comparison w.r.t. Feedback Bit

The performance w.r.t. the sum feedback bit budget is evaluated for SNR $= 25$ dB the varying sum feedback bit budget $B \in [100, 500]$ bit.

Figure 3.6 shows that the sum-cluster rate is increasing with the sum feedback bit budget for SNR $= 25$ dB, with an upper bound of the performance under unlimited feedback in Figure 3.4 and Figure 3.5. The proposed DBA outperforms the EBA in both the sum cluster-rate in Figure 3.6 and the minimum single-cell rate in Figure 3.7.

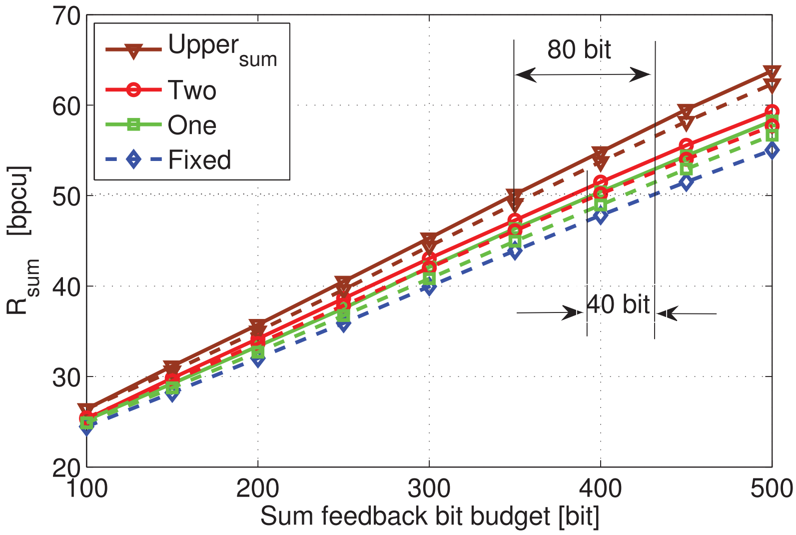


Figure 3.6: Sum-cluster rate comparison under limited feedback w.r.t. SNR.

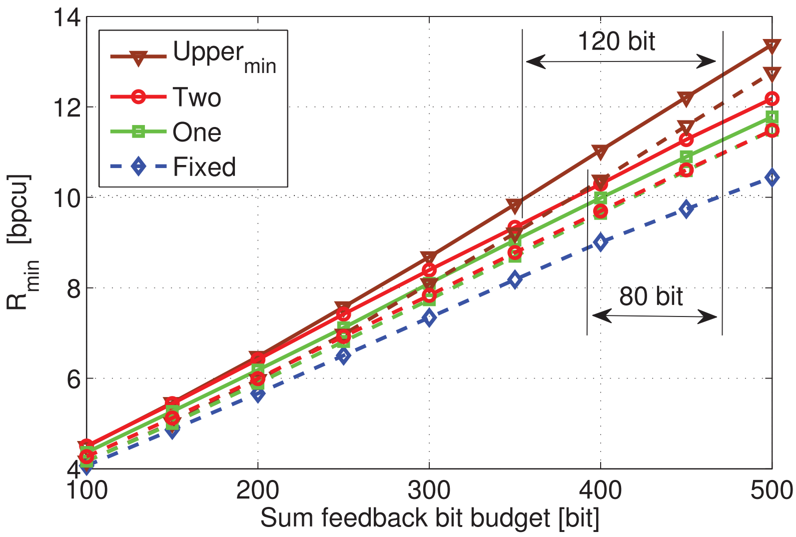


Figure 3.7: Minimum single-cell rate comparison under limited feedback w.r.t. SNR.

Compared with the fixed matching with the EBA, the proposed centralized assignment and the distributed stable matching with the DBA can save around 80 bit and 40 bit, respectively, to achieve $R_{sum} = 50$ bpcu in Figure 3.6, and around 120 bit and 80 bit, respectively, to achieve $R_{min} = 10$ bpcu in Figure 3.7. Compared with Figure 3.4 and Figure 3.5, the GIA still significantly outperforms the random beamforming and FDMA when $B \geq 100$ bit. The sum-cluster RINR in $10\log_{10}(\sum_{\ell=1}^{K} \mathcal{I}_{\ell}^{k})$ dB is decreasing with sum feedback bit budget as shown in Figure 3.8. The DBA achieves a lower RINR

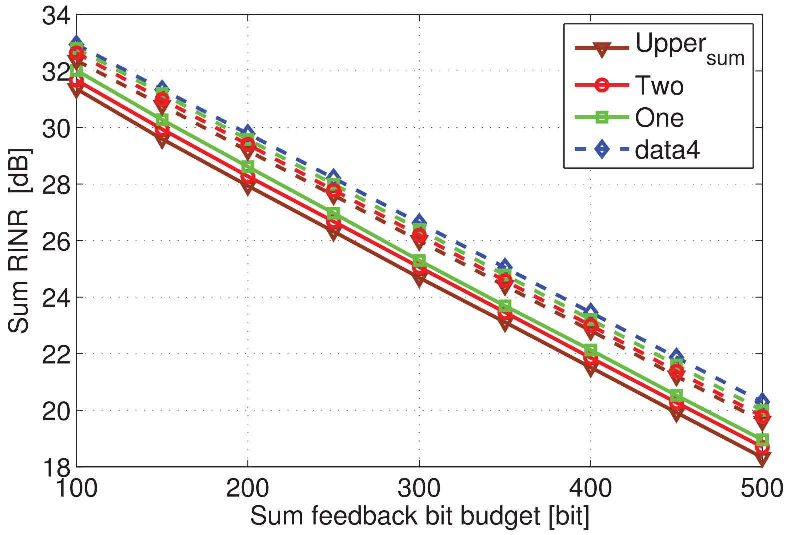


Figure 3.8: Sum-cluster RINR comparison under limited feedback w.r.t. SNR.

compared with the EBA, which implies that the effectiveness of minimizing the upper bound of sum-cluster RINR in (3.28).

B. Performance Comparison w.r.t. SNR

The proposed algorithms are evaluated by measuring the sum-cluster rate and the single-cell rate performance w.r.t. SNR for the fixed sum feedback budget $B = 300$ bit and $B = 500$ bit, respectively.

From Figure 3.9 and Figure 3.10, it is observed that the performance with $B = 500$ bits is much greater than that with $B = 300$ bits and the performance gap enlarges with the SNR. E.g., for $SNR = 30$ dB, the gap of sum-cluster rate and that of the single-cell rate are as large as around 20 bpcu and 8 bpcu, respectively. From the perspective of energy consumption, the feedback of $B = 500$ bits results in a higher complexity and more feedback energy consumption than the feedback of $B = 300$ bits, while it is still attractive when the battery power saving is focused, because MSs are usually powered by battery and BSs are powered by electric networks, and the more scarce battery energy can be saved at the cost of the BS energy. E.g., 15 dB uplink power can be saved by the stable matching to achieve $R_{sum} = 40$ bpcu with $B = 500$ bits compared with $B = 300$ bits. Compared to the fixed matching with EBA, the proposed centralized assignment and stable matching with DBA can reduce by 10 dB and 5 dB uplink power, respectively, to achieve $R_{sum} = 60$ bpcu. Furthermore, this performance improvement enlarges with SNR.

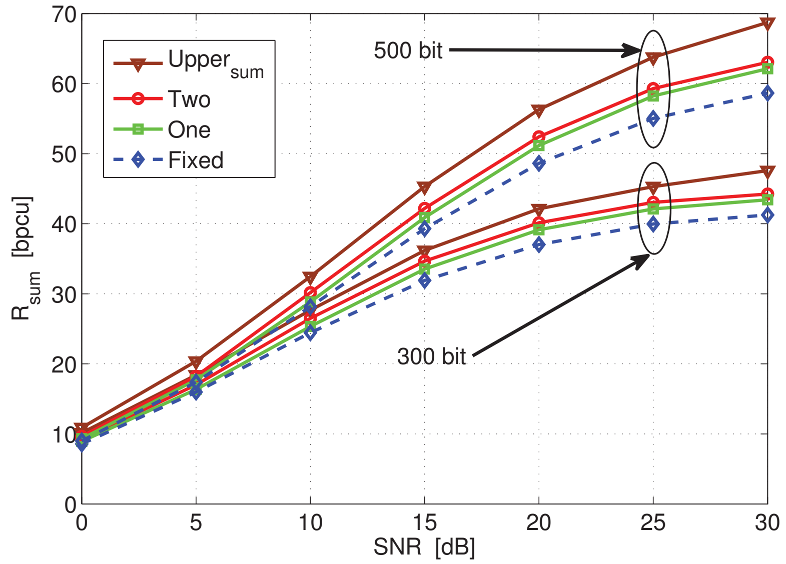


Figure 3.9: Sum cluster-rate comparison under limited feedback w.r.t. SNR.

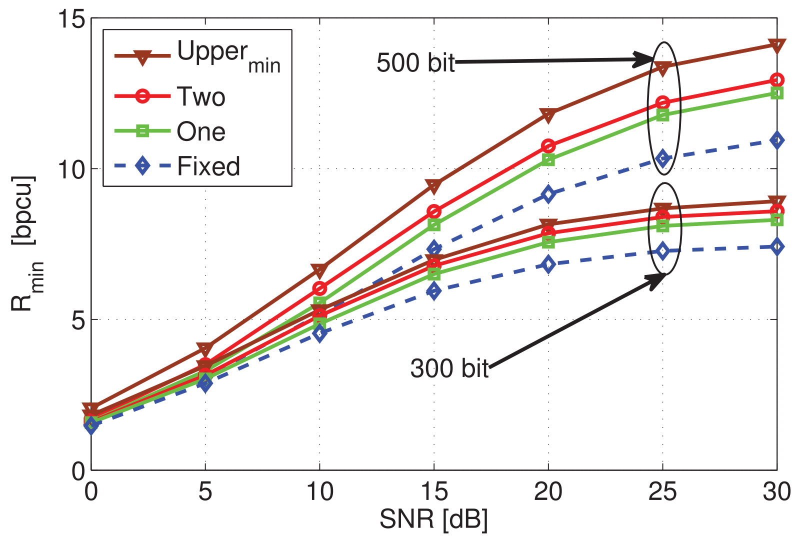


Figure 3.10: Minimum single-cell rate comparison under limited feedback w.r.t. SNR.

3.7 Summary

In this chapter, we provide a framework of the GIA with optimized IA-Cell assignment in the MIMO-IMAC network under limited feedback. This GIA algorithm enables to compute the closed-form IA precoders only based on the local CSIR at each BS. We propose the concept of the IA-Cell assignment, show its influence on the rate/DoF performance and propose three IA-Cell assignment approaches according to different backhaul overhead, which is verified by numerical results. For instance, different IA-Cell assignments achieves significantly different rate performance. The proposed two distributed matching approaches can find a stable matching if it exists, which not only increases the stability of the multi-cell cluster but also achieves an better rate performance than the fixed matching. Under limited feedback, the effectiveness of the proposed DBA problem and its solutions are also verified by numerical results. The performance of the DBA is better than that by the EBA for the high SNR. With the additional analysis on the implementations, the required backhaul overhead and the complexity, The three contributions in this chapter jointly provide a comprehensive holistic design of the multi-cell interfering system under limited feedback.

3.8 Collection of Proofs

3.8.1 Proof of Propsosition 9

Considering *Cell* $\ell \xrightarrow{IA}$ *Cell* k, we have

$$
\begin{aligned}
\mathcal{I}^k &\triangleq \sum_{m=1,m\neq k}^{K} \mathcal{I}_m^k = \mathcal{I}_\ell^k && (3.42a)\\
&\leq \sum_{i=1}^{L} \frac{P_{i,\ell}}{\sigma_k^2 d_s} \operatorname{Tr}\left(\widehat{\boldsymbol{V}}_{i,\ell}^H \boldsymbol{H}_{i,\ell}^{k,H} \boldsymbol{\Pi}^{\perp}_{\boldsymbol{H}_{1,\ell}^k \boldsymbol{V}_{1,\ell}^{in}} \boldsymbol{H}_{i,\ell}^k \widehat{\boldsymbol{V}}_{i,\ell}\right) && (3.42b)\\
&= \sum_{i=1}^{L} \frac{P_{i,\ell}}{\sigma_k^2 d_s} \operatorname{Tr}(\boldsymbol{S}_{i,\ell}^H \boldsymbol{\Omega}_{i,\ell}^k \boldsymbol{S}_{i,\ell} \boldsymbol{\Sigma}_{i,\ell}) && (3.42c)\\
&\leq \sum_{i=1}^{L} \frac{P_{i,\ell}}{\sigma_k^2 d_s} \sum_{d=1}^{d_s} \lambda_d(\boldsymbol{\Omega}_{i,\ell}^k)\beta_{i,d} && (3.42d)\\
&\leq \sum_{i=1}^{L} \frac{P_{i,\ell}}{\sigma_k^2 d_s} \lambda_1(\boldsymbol{\Omega}_{i,\ell}^k) \sum_{d=1}^{d_s} \beta_{i,d} && (3.42e)\\
&= \sum_{i=1}^{L} \frac{P_{i,\ell}}{\sigma_k^2 d_s} \lambda_1(\boldsymbol{\Omega}_{i,\ell}^k) d_c^2(\widehat{\boldsymbol{V}}_{i,\ell}, \overrightarrow{\boldsymbol{V}}_{i,\ell}) && (3.42f)\\
&\leq c(N_U, d_s) \sum_{i=1}^{L} \frac{P_{i,\ell}}{\sigma_k^2 d_s} \lambda_1(\boldsymbol{\Omega}_{i,\ell}^k) 2^{-\frac{B_{i,\ell}}{ds(N_U-d_s)}}, && (3.42g)
\end{aligned}
$$

where (3.42b) is derived based on the inequality of $||\boldsymbol{\Pi}^{\perp}_{[\boldsymbol{Y}_1,\boldsymbol{Y}_2]}\boldsymbol{Y}_3||_F^2 \leq ||\boldsymbol{\Pi}^{\perp}_{[\boldsymbol{Y}_1]}\boldsymbol{Y}_3||_F^2$. Plugging (3.24) into (3.42b) and removing the zero-valued terms and based on the definition (3.29) yield (3.42c), where $\boldsymbol{S}_{i,\ell} \in \mathbb{C}^{N_U \times d_s}$ satisfies $\boldsymbol{S}_{i,\ell}^H \boldsymbol{S}_{i,\ell} = \boldsymbol{I}_{d_s}$ and $\boldsymbol{\Sigma}_{i,\ell} = \operatorname{diag}\{\beta_{i,1}, \ldots, \beta_{i,d_s}\}, \forall i$ is with $\beta_{i,d} \in [0,1], \forall d$ and $\sum_{d=1}^{d_s} \beta_{i,d} = d_c^2(\widehat{\boldsymbol{V}}_{i,\ell}, \overrightarrow{\boldsymbol{V}}_{i,\ell})$. The upper bound (3.42d) is achieved when the truncated unitary matrix $\boldsymbol{S}_{i,\ell}$ is the eigen-subspace of the matrix $\boldsymbol{\Omega}_{i,\ell}^k$ associated with the d_s largest eigenvalues $\lambda_1(\boldsymbol{\Omega}_{i,\ell}^k), \ldots, \lambda_{d_s}(\boldsymbol{\Omega}_{i,\ell}^k)$. (3.42g) is derived by the quantization distortion upper bound (3.25).

3.8.2 Proof of Lemma 4

The quantization $\widehat{\boldsymbol{V}}$ can be exactly expressed by the N-dimensional full space $\boldsymbol{V} \cup \boldsymbol{V}^{\perp}$ as

$$\widehat{\boldsymbol{V}} = \boldsymbol{\Pi}_{\boldsymbol{V}} \widehat{\boldsymbol{V}} + \boldsymbol{\Pi}_{\boldsymbol{V}}^{\perp} \widehat{\boldsymbol{V}} = \boldsymbol{V}\boldsymbol{Q}_1 + \boldsymbol{V}^{\perp}\boldsymbol{Q}_2, \tag{3.43}$$

where $\boldsymbol{Q}_1 \in \mathbb{C}^{N \times N}$ and $\boldsymbol{Q}_2 \in \mathbb{C}^{(M-N) \times N}$ in (3.43) denote the components of $\widehat{\boldsymbol{V}}$ projected on the $\boldsymbol{V}$ and $\boldsymbol{V}^{\perp}$, respectively. From (3.43), it is derived the properties of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$

as

$$\widehat{\boldsymbol{V}}^H \widehat{\boldsymbol{V}} = \boldsymbol{I}_N \Rightarrow \boldsymbol{Q}_1^H \boldsymbol{Q}_1 + \boldsymbol{Q}_2^H \boldsymbol{Q}_2 = \boldsymbol{I}_N; \tag{3.44}$$

$$d_c^2(\widehat{\boldsymbol{V}}, \boldsymbol{V}) = N - \mathrm{Tr}(\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^H \boldsymbol{V}\boldsymbol{V}^H) \Rightarrow \mathrm{Tr}(\boldsymbol{Q}_2 \boldsymbol{Q}_2^H) = d_c^2(\widehat{\boldsymbol{V}}, \boldsymbol{V}). \tag{3.45}$$

By the singular-value decomposition (SVD), $\boldsymbol{Q}_1$ is expressed by $\boldsymbol{Q}_1 = \boldsymbol{U}_{\boldsymbol{Q}_1} \boldsymbol{\Lambda}_{\boldsymbol{Q}_1}^{1/2} \boldsymbol{V}_{\boldsymbol{Q}_1}^H$ where eigenvalues $\boldsymbol{\Lambda}_{\boldsymbol{Q}_1} \triangleq \mathrm{diag}\left\{\lambda_1(\boldsymbol{Q}_1^H \boldsymbol{Q}_1), \ldots, \lambda_N(\boldsymbol{Q}_1^H \boldsymbol{Q}_1)\right\}$ satisfy $\lambda_n(\boldsymbol{Q}_1^H \boldsymbol{Q}_1) \geq 0, \forall n$ subject to $\sum_{n=1}^N \lambda_n(\boldsymbol{Q}_1^H \boldsymbol{Q}_1) = N - d_c^2(\widehat{\boldsymbol{V}}, \boldsymbol{V})$ based on (3.44) and (3.45). From (3.44), we further derive $\boldsymbol{Q}_2^H \boldsymbol{Q}_2 = \boldsymbol{V}_{\boldsymbol{Q}_1}(\boldsymbol{I}_N - \boldsymbol{\Lambda}_{\boldsymbol{Q}_1})\boldsymbol{V}_{\boldsymbol{Q}_1}^H \succeq \boldsymbol{0}_N$, which requires $\lambda_n(\boldsymbol{Q}_1^H \boldsymbol{Q}_1) \leq 1, \forall n$. Therefore, $\boldsymbol{Q}_2$ can be expressed by

$$\boldsymbol{Q}_2 = \widetilde{\boldsymbol{U}}(\boldsymbol{I}_N - \boldsymbol{\Lambda}_{\boldsymbol{Q}_1})^{1/2} \boldsymbol{V}_{\boldsymbol{Q}_1}^H \tag{3.46}$$

where $\widetilde{\boldsymbol{U}} \in \mathbb{C}^{(M-N)\times N}$ satisfying $\widetilde{\boldsymbol{U}}^H \widetilde{\boldsymbol{U}} = \boldsymbol{I}_N$ is to select a N-dimensional subspace from the $M - N$-dimensional null space $\mathrm{Span}\{\boldsymbol{V}^{\perp}\}$.

Chapter 4

Conclusions and Future Work

4.1 Summary of Contributions

In this thesis, we mainly study how to properly manage the interference so as to suppress its influence on the achievable rate in MIMO interference networks. Two PHY resource allocation problems for MIMO interference networks are investigated: 1) Pareto boundary computation for the achievable rate region in a K-user single-stream MIMO IC and 2) IA under limited feedback for a MIMO-IMAC. For each problem, joint transceiver optimization with linear transmit and receive processing is done at BS-side, which is based on the perfect local instantaneous CSI at each BS. New algorithms for both problems are proposed and supported by numerical examples. This chapter summarizes the main conclusions of the thesis, which may provide insights on the system design of the future multi-cell interfering MIMO systems.

To find efficient operating points and their associated strategies, the Pareto boundary computation problem is studied. However, the Pareto boundary computation for a K-user single-stream MIMO IC was not studied before, since the achievable rate expressions are coupled by both the transmit and receive beamforming vectors, which is different from the MISO/SIMO IC optimization with respect to only one-side beamforming vectors. The importance of MIMO techniques motivated us to study this open problem. The main contributions in this area are concluded as follows:

- We derive an equivalent expression of the SINR by the concept of Hermitian angle for a two-user single-stream MIMO IC. Based on this new expression, we compute some key points exactly on the Pareto boundary in closed-form, such as the two ending points of the strict Pareto boundary. Furthermore, we prove that full power transmission is optimal to achieve the strict Pareto boundary for the two-user single-stream MIMO IC.

- We give an overview of the multi-objective optimization, which is helpful to understand the principle of Pareto boundary computation. Based on the formulations of ϵ-constraint optimization and weighted Chebyshev optimization, we transform the original multi-objective optimization problem for Pareto boundary computation into two single-objective problems: 1) single constraint rate maximization problem and 2) alternating rate profile optimization problem, respectively. For each single-objective optimization problem, we propose a heuristic alternating optimization algorithm. The convergence of each algorithm is guaranteed but the convergent solution is not necessary to be globally optimal.

- We show supporting numerical examples to evaluate the two proposed alternating optimization algorithms. Both algorithms have a similar rate performance, while the single constraint rate maximization algorithm usually takes less iterations than the alternating rate profile optimization algorithm. Compared with prior work, numerical results show that only the two proposed algorithms enable to compute a *complete* close-to-optimal inner bound to the Pareto boundary. The performance of the proposed algorithms depends (but not highly) on initializations. Our proposed initialization schemes usually enable better rate performance with less iterations than random initializations.

- We describe the potential implementations and complexity of the proposed algorithms. The assumptions of the perfect local CSI and unlimited feedback limit the applications in practical systems. Nevertheless, the proposed alternating optimization algorithms provide an insight or a potential way to solve the NP-hard multi-objective optimization problems. Furthermore, the proposed algorithms may be applied to the scenarios with small channel uncertainties and a long coherence time (e.g., for MSs with slow mobility), where the designed linear transceiver is effective during a long period (until the end of the coherence time) such that it is meaningful to spend time on a heuristic iterative algorithm. For the scenarios with large channel uncertainties, our proposed algorithms serve as a benchmark.

We extend the K-user MIMO IC to the K-cell MIMO-IMAC, allowing a BS serves multiple MSs simultaneously. Furthermore, we also consider some practical constraints, such as complexity, backhaul overhead and limited feedback. Thus, the studied MIMO-IMAC well matches the uplink scenario of the multi-cell MIMO system with spectrum sharing. In order to avoid the influence of interference (both the IUI and ICI), we aim to design the IA precoders and ZF decoders with low complexity, in a distributed way and under limited feedback. The main contributions in this area are concluded as follows.

- We provide a complete study of the GIA in terms of the feasible conditions for the DoF and the optimal IA precoders computation. The optimal IA transceiver is obtained in closed-form without a need of iterative computation, by jointly designing the precoders and decoders in a tight restriction-relaxation process. Furthermore, the GIA algorithm is based on only perfect local CSIR, which allows the efficient distributed implementation.

- We introduce the concept of IA-Cell assignment and analyze its effect on both the achievable rate and DoF performance. We consider three IA-Cell assignment scenarios with different backhaul overhead, for which three assignment approaches are proposed: 1) one-sided stable matching, 2) two-sided stable matching and 3) centralized assignment. The first two stable matching approaches enable to find a stable matching if it exists, since the stability and distributed implementation are important to facilitate self-organization in a multi-cell cluster. The centralized assignment approach requires a powerful authority to search an optimal assignment that satisfies an arbitrary predefined metric with a reasonable complexity when $K \leq 4$.

- We study the subspace quantization and limited feedback of IA precoders for MSs. Thereby, we develop a novel quantized subspace characterization based on the Grassmannian subspace quantization, which is more efficient than the previous result. Due to the quantization distortion, IA cannot be performed perfectly and thus residual interference exists. We derive a closed-form upper bound of the RINR. In order to enable the efficient feedback, we formulate and solve a DBA problem by minimizing the upper bound of the sum RINR subject to a sum feedback bits budget.

- We show numerical examples to illustrate that the proposed GIA algorithm with both optimized IA-Cell assignment and the DBA greatly outperforms the traditional GIA with the EBA. These main contributions above jointly form a framework to provide a comprehensive holistic design of the uplink transmission in a multi-cell interfering MIMO system. Furthermore, both the introduced IA-Cell assignment concept and approaches and the proposed subspace quantization and feedback techniques are not restricted to the GIA in the MIMO-IMAC but the general IA in the upgraded or degraded systems, such as the study of IA in heterogeneous networks or the multi-user MIMO IC.

4.2 Future Work

There is an endless road of possible improvements and generalizations to the results of the thesis. Several ideas for future work have been conceived in the process of writing the thesis. One direction is to find the global optimal solution of the non-convex optimization problems. The other direction is to relax or avoid the assumptions that may affect the generality and application of the proposed algorithms.

- *Pareto Boundary Computation:* The thesis proposes two algorithms to achieve the close-to-optimal solutions of the Pareto boundary computation problem for a K-user single-stream MIMO IC. However, the globally optimal solution is still unknown. On the one hand, since this problem is NP-hard in complexity, the computation of the global optimum is a challenging open problem. On the other hand, it may be possible to exactly compute the Pareto boundary for other utility regions, e.g., the average performance region where the transceiver is not optimized for each channel realization but for multiple channel realizations during a time period. The average performance evaluation is more practical because of the low complexity and overhead compared with the instantaneous performance.

- *Robust Transceiver Design:* The thesis is based on the assumption of the perfect local CSI, which is often unrealistic in practical systems. Therefore, it is meaningful to study the robust transceiver design when channel uncertainties exist. In this case, the channel uncertainties need to be dealt with like extra "variables" in transceiver optimization problem. For instance, one of our ongoing work is to compute the the Pareto boundary of the worst-case rate region for a K-user MIMO IC with channel uncertainties. Furthermore, the robust optimization with channel estimation uncertainties becomes more difficult if we additionally consider the uncertainties in backhaul links, or feedback links or both. How to deal with the accumulated uncertainties introduced by the practical considerations in MIMO interference networks is a novel and important topic.

- *Massive MIMO:* The traditional MIMO system can be evolved to the massive MIMO system by deploying a very large number of antennas at BSs, which not only provides more potential DoF but also achieves better performance in terms of spectral efficiency and link reliability. Most previous work assumes the independent and identically distributed (i.i.d.) complex Gaussian channel model, since it makes the channels approach to pairwise-orthogonal and the optimization problem analyzable by random matrix theory. However, this assumption does not always hold and spatial correlation usually exists in realistic channel. Therefore, the study of spatial correlation in massive MIMO is an important extended issue. For instance, how many antennas is optimal to be placed in a fixed space? This question generally introduces the effect of spatial correlation on the system performance, such as channel estimation, channel feedback, rate maximization, energy/cost efficiency etc. We have already studied the spatial-correlated channel feedback problem, where we first sparsify the long channel vectors and then quantize the sparse version. However, a complete and comprehensive study of the role of spatial correlation on the overall massive MIMO system is needed.

Bibliography

[3GP12] 3GPP, "Carrier aggregation explained," May 2012. [Online]. Available: http://www.3gpp.org/technologies/keywords-acronyms/101-carrier-aggregation-explained

[ABC+14] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[AG07] A. Ashikhmin and R. Gopalan, "Grassmannian packings for efficient quantization in MIMO broadcast systems," in *Proc. of IEEE ISIT*, June 2007.

[AHZ11] W. Ai, Y. Huang, and S. Zhang, "New results on Hermitian matrix rank-one decomposition," *Math. Program: Ser. A*, vol. 128, no. 1-2, pp. 253–283, Jun. 2011.

[BBO12] E. Björnson, M. Bengtsson, and B. Ottersten, "Pareto characterization of the multicell MIMO performance region with simple receivers," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4464–4469, Aug. 2012.

[BDMS08] J. Branke, K. Deb, K. Miettinen, and R. Slowinski, Eds., *Multiobjective Optimization: Interactive and Evolutionary Approaches.* Springer, 2008.

[Ber99] D. P. Bertsekas, *Nonlinear programming.* 2nd ed. Athena Scientific, 1999.

[BJ13] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Found. and Trends in Commun. and Inf. Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[BMM10] P. Biro, D. F. Manlove, and S. Mittal, "Size versus stability in the marriage problem," *Theoretical Computer Science*, pp. 1828–1841, 2010.

[BN10] H. Boche and S. Naik, "Revisiting proportional fairness: Anonymity among users in interference coupled wireless systems," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2995–3000, Oct. 2010.

[BO01] M. Bengtsson and B. Ottersten, "Optimum and suboptimum transmit beamforming," *Handbook of Antennas in Wireless Communications*, 2001.

[BT09] H. Bolcskei and I. Thukral, "Interference alignment with limited feedback," in *Proc. of IEEE ISIT*, June 2009.

[BZGO10] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4298–4310, Aug 2010.

[C1114] C114 news, "China Unicom will refarm 2G frequency to develop 4G in some cities," Aug. 2014. [Online]. Available: http://www.cn-c114.net/576/a854766.html

[CACC08] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wirel. Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[Car09] R. Carden, "A simple algorithm for the inverse field of values problem," *Inverse Problems*, vol. 25, no. 11, pp. 115 019–115 027, 2009.

[CCHJ12] P. Cao, Z. Chong, Z. Ho, and E. Jorswieck, "Energy-efficient power allocation for amplify-and-forward MIMO relay channel," in *Proc. of IEEE CAMAD*, Sep. 2012.

[CHHT12] C.-B. Chae, I. Hwang, R. Heath, and V. Tarokh, "Interference aware-coordinated beamforming in a multi-cell system," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 10, pp. 3692–3703, Oct. 2012.

[CJ08] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug 2008.

[CJ13] P. Cao and E. Jorswieck, "Source energy-saving performance in amplify-and-forward relay-assisted wireless systems," in *Proc. of IEEE VTC-Spring*, Jun. 2013.

[CJ14a] ——, "DCT and VQ based limited feedback in spatially-correlated massive MIMO systems," in *Proc. of IEEE SAM*, Jun. 2014.

[CJ14b] ——, “Robust optimization for multi-cell interfering MIMO-MAC under limited feedback,” in *Proc. of IEEE ICASSP*, May 2014.

[CJKR10] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser MIMO achievable rates with downlink training and channel state feedback,” *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[CJS13] P. Cao, E. Jorswieck, and S. Shi, “Pareto boundary of the rate region for single-stream MIMO interference channels: Linear transceiver design,” *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4907–4922, Oct 2013.

[CLW+06] P. Chan, E. Lo, R. Wang, E. Au, V. Lau, R. Cheng, W. H. Mow, R. Murch, and K. Letaief, “The evolution path of 4G networks: FDD or TDD?” *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 42–50, Dec 2006.

[Coh98] J. E. Cohen, “Cooperation and self-interest: Pareto-inefficiency of nash equilibria in finite random games,” *Proc. of Natl. Acad. Sci. USA*, vol. 95, pp. 9724–9731, Aug. 1998.

[CSJ12] P. Cao, S. Shi, and E. Jorswieck, “Efficient computation of the Pareto boundary for the two-user single-stream MIMO interference channel,” in *Proc. of IEEE SPAWC*, Jun. 2012.

[CSS+14] R. L. G. Cavalcante, S. Stańczak, M. Schubert, A. Eisenblätter, and U. Türke, “Toward energy-efficient 5G wireless communications technologies,” *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 24–34, Nov. 2014.

[CVWT12] Z. Chen, S. Vorobyov, C.-X. Wang, and J. Thompson, “Pareto region characterization for rate control in MIMO interference systems and Nash bargaining,” *IEEE Trans. Autom. Control*, vol. 57, no. 12, pp. 3203–3208, Dec. 2012.

[CZJ14] P. Cao, A. Zappone, and E. A. Jorswieck, “Grouping-based interference alignment with IA-Cell assignment in multi-cell MIMO MAC under limited feedback,” *submitted to IEEE Trans. Signal Process.*, Sep. 2014. [Online]. Available: http://arxiv.org/abs/1409.3686

[DD97] I. Das and J. Dennis, “A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems,” *Struct. and Multidiscip. Optim.*, vol. 14, no. 1, pp. 63–69, 1997.

[Din67] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, 1967.

[DLR08] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1108–1123, March 2008.

[DPS11] E. Dahlman, S. Parkvall, and J. Skoeld, *4G: LTE/LTE-Advanced for mobile broadband*. Academic Press, May 2011.

[Eri14a] Ericsson, "Ericsson mobility report," http://www.ericsson.com/mobility-report, Tech. Rep., Jun. 2014.

[Eri14b] ——, "Ericsson mobility report: Interim update," http://www.ericsson.com/mobility-report, Tech. Rep., Aug. 2014.

[Eri14c] ——, "Traffic exploration tool," Oct. 2014. [Online]. Available: http://www.ericsson.com/TET

[EXA] EXALT. TDD and FDD Explained. [Online]. Available: http://www.exaltcom.com/TDD-and-FDD-Explained.aspx

[GB12] M. Grant and S. Boyd. (2012) CVX: matlab software for disciplined convex programming. [Online]. Available: http://cvxr.com/cvx/

[GC14] G. Gupta and A. Chaturvedi, "User selection in MIMO interfering broadcast channels," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1568–1576, May 2014.

[GCJ11] K. Gomadam, V. Cadambe, and S. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, June 2011.

[GHD82] A. Goicoechea, D. R. Hansen, and L. Duckstein, *Multiobjective decision analysis with engineering and business applications*. Wiley, New York, 1982.

[GI89] D. Gusfield and R. Irving, *The Stable Marriage Problem: Structure and Algorithms*. MIT Press, Cambridge, Massachusetts, 1989.

[GJ10] T. Gou and S. Jafar, "Degrees of freedom of the K user MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.

[GJJV03] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of MIMO channels,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[GS62] D. Gale and L. Shapley, “College admission and the stability of marriage,” *Amer. Math. Monthly*, vol. 69, no. 1, pp. 9–15, Jan. 1962.

[Has03] M. Hassani, “Derangements and applications,” *J. Integer Seq.*, vol. 6, pp. 1–8, 2003.

[HBH06] J. Huang, R. A. Berry, and M. L. Honig, “Distributed interference compensation for wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[HG10] Z. K. M. Ho and D. Gesbert, “Balancing egoism and altruism on the interference channel: The MIMO case,” in *Proc. IEEE ICC*, Cape Town, South Africa, 2010.

[HJ85] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge Univ. Press, 1985.

[HJ91] ——, *Topics in Matrix Analysis.* Cambridge Univ. Press, 1991.

[HRTA14] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, “Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective,” *IEEE Wireless Commun. Mag.*, vol. 21, no. 3, pp. 118–127, June 2014.

[IDM+11] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[Jaf11] S. A. Jafar, “Interference alignment - A new look at signal dimensions in a communication network,” *Found. and Trends in Commun. and Inf. Theory*, vol. 7, no. 1, pp. 1–134, 2011.

[JBF+14] E. Jorswieck, L. Badia, T. Fahldieck, E. Karipidis, and J. Luo, “Spectrum sharing improves the network efficiency for cellular operators,” *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 129–136, Mar. 2014.

[JC13] E. Jorswieck and P. Cao, "Matching and exchange market based resource allocation in MIMO cognitive radio networks," in *Proc. of EUSIPCO*, Sep. 2013.

[Jin06] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov 2006.

[JL08] E. A. Jorswieck and E. G. Larsson, "The MISO interference channel from a game-theoretic perspective: A combination of selfishness and altruism achieves Pareto optimality," in *Proc. of IEEE ICASSP*, Apr. 2008.

[JLD08] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[KJGK10] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Future Network and Mobile Summit,*, Jun. 2010.

[KL10] E. Karipidis and E. G. Larsson, "Efficient computation of the Pareto boundary for the MISO interference channel with perfect CSI," in *Proc. of WIOPT*, Jun. 2010.

[KLCH11] T. Kim, D. Love, B. Clerckx, and D. Hwang, "Spatial degrees of freedom of the multicell MIMO multiple access channel," in *Proc. of IEEE GLOBECOM*, Dec. 2011, pp. 1–5.

[KMLL12] J.-S. Kim, S.-H. Moon, S.-R. Lee, and I. Lee, "A new channel quantization strategy for MIMO interference alignment with limited feedback," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 1, pp. 358–366, Jan. 2012.

[KPSL10] J. Kim, S.-H. Park, H. Sung, and I. Lee, "Sum rate analysis of two-cell MIMO broadcast channels: Spatial multiplexing gain," in *Proc. of IEEE ICC*, May 2010.

[KV10] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *Proc. of IEEE ISIT*, June 2010.

[KZ05] A. J. Kurdila and M. Zabarankin, *Convex Functional Analysis*, Springer, Ed., 2005.

[LDL13] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, “Max-min fairness linear transceiver design for a multi-user MIMO interference channel,” *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2413–2423, May 2013.

[LH05] D. Love and R. Heath, “Limited feedback unitary precoding for spatial multiplexing systems,” *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[LHD13] Y.-F. Liu, M. Hong, and Y.-H. Dai, “Max-min fairness linear transceiver design problem for a multi-user SIMO interference channel is polynomial time solvable,” *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 27–30, Jan 2013.

[LHL+08] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, “An overview of limited feedback in wireless communication systems,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[LJ06] J. Lee and N. Jindal, “Dirty Paper Coding vs. linear precoding for MIMO broadcast channels,” in *Proc. of ACSSC ’06*, Oct 2006, pp. 779–783.

[LJ08] E. G. Larsson and E. A. Jorswieck, “Competition versus cooperation on the MISO interference channel,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.

[LKL11] J. Lindblom, E. Karipidis, and E. G. Larsson, “Closed-form parameterization of the Pareto boundary for the two-user MISO interference channel,” in *Proc. of IEEE ICASSP*, May 2011.

[LMS+10] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, “Semidefinite relaxation of quadratic optimization problems,” *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[LZC12] L. Liu, R. Zhang, and K.-C. Chua, “Achieving global optimality for weighted sum-rate maximization in the K-user Gaussian interference channel with multiple antennas,” *IEEE Trans. Wirel. Commun.*, vol. 11, no. 5, pp. 1933–1945, May 2012.

[MA04] R. Marler and J. Arora, “Survey of multi-objective optimization methods for engineering,” *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, Mar. 2004.

[MCJ13] R. Mochaourab, P. Cao, and E. Jorswieck, “Alternating rate profile optimization in single stream MIMO interference channels,” in *Prof. of IEEE ICASSP*, May 2013.

[MCJ14] ——, “Alternating rate profile optimization in single stream MIMO interference channels,” *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 221–224, Feb 2014.

[MD14] A. Medra and T. Davidson, “Flexible codebook design for limited feedback systems via sequential smooth optimization on the Grassmannian manifold,” *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1305–1318, Mar. 2014.

[Mey00] C. D. Meyer, *Matrix analysis and applied linear algebra.* New York, NY, USA: Cambridge University Press, 2000.

[Mie98] K. Miettinen, *Nonlinear Multiobjective Optimization.* Springer, 1998.

[MJ11a] R. Mochaourab and E. A. Jorswieck, “Optimal beamforming in interference networks with perfect local channel information,” *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1128–1141, Mar. 2011.

[MJ11b] ——, “Walrasian Equilibrium in two-user multiple-input single-output interference channel,” in *Proc. of IEEE ICC*, Kyoto, Japan, Jun. 2011.

[Moc12] R. Mochaourab, “Game theory and microeconomic theory for beamforming design in multiple-input single-output interference channels,” Ph.D. dissertation, TU Dresden, May 2012.

[MZC06] M. Mohseni, R. Zhang, and J. Cioffi, “Optimized transmission for fading multiple-access and broadcast channels with multiple antennas,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1627–1639, Aug. 2006.

[NR10] V. Nagarajan and B. Ramamurthi, “Distributed cooperative precoder selection for interference alignment,” *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4368–4376, Nov. 2010.

[OR94] M. J. Osborne and A. Rubinstein, *A Course in Game Theory.* The MIT Press, 1994.

[PC14] J. Park and B. Clerckx, “Joint wireless information and energy transfer in a K-user MIMO interference channel,” *IEEE Trans. Wirel. Commun.*, vol. 13, no. 10, pp. 5781–5796, Oct. 2014.

[PH11] S. Peters and R. Heath, “Cooperative algorithms for MIMO interference channels,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[Pol05] I. Polik. (2005) Addendum to the SeDuMi user guide: Version 1.1. [Online]. Available: http://sedumi.ie.lehigh.edu

[PS13] J. Park and Y. Sung, “On the Pareto-optimal beam structure and design for multi-user MIMO interference channels,” *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5932–5946, Dec 2013.

[PYCE10] D. P. Palomar and E. Yonina C. Eldar, “Semidefinite programming, matrix decomposition, and radar code design,” in *Convex Optimization in Signal Processing and Communications*, Cambridge Univ. Press, New York, 2010, pp. 192–228.

[Qua] Qualcomm, “The 1000x data challenge.” [Online]. Available: https://www.qualcomm.com/1000x

[QZH09] L. P. Qian, Y. Zhang, and J. Huang, “MAPEL: Achieving global optimality for a non-convex wireless power control problem,” *IEEE Trans. Wirel. Commun.*, vol. 8, no. 3, pp. 1553–1563, Mar. 2009.

[QZLC11] J. Qiu, R. Zhang, Z.-Q. Luo, and S. Cui, “Optimal distributed beamforming for MISO interference channels,” *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5638–5643, Nov. 2011.

[RG13] M. Rezaee and M. Guillaud, “Interference alignment with quantized Grassmannian feedback in the K-user constant MIMO interference channel,” *submitted to IEEE Trans. Inf. Theory*, Jan. 2013. [Online]. Available: http://arxiv.org/abs/1207.6902

[RHL13] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness,” *Signal Processing*, vol. 93, no. 12, pp. 3327–3340, Dec. 2013.

[RJ08] N. Ravindran and N. Jindal, “Limited feedback-based block diagonalization for the MIMO broadcast channel,” *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.

[RRL13] X. Rao, L. Ruan, and V. Lau, "Limited feedback design for interference alignment on MIMO interference networks with heterogeneous path loss and spatial correlations," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2598–2607, May 2013.

[SBH09a] C. Shi, R. Berry, and M. Honig, "Local interference pricing for distributed beamforming in MIMO networks," in *Proc. of IEEE MILCOM*, Oct 2009.

[SBH09b] ——, "Monotonic convergence of distributed interference pricing in wireless networks," in *Proc. of IEEE ISIT*, June 2009.

[Sch83] S. Schaible, "Fractional programming," *Zeitschrift fuer Operations Theory and Applications*, vol. 27, no. 1, pp. 374–352, 1983.

[Sch01] K. Scharnhorst, "Angles in complex vector spaces," *Acta Appl. Math.*, vol. 69, no. 1, pp. 95–103, Oct. 2001.

[SGHP10] I. Santamaria, O. Gonzalez, R. Heath, and S. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. of IEEE GLOBECOM*, Dec. 2010.

[SHT11] C. Suh, M. Ho, and D. Tse, "Downlink interference alignment," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2616–2626, Sep. 2011.

[SJW09] K. Schober, P. Janis, and R. Wichman, "Geodesical codebook design for precoded MIMO systems," *IEEE Commun. Lett.*, vol. 13, no. 10, pp. 773–775, Oct. 2009.

[SLL+11] W. Shin, N. Lee, J.-B. Lim, C. Shin, and K. Jang, "On the design of interference alignment scheme for two-cell MIMO interfering broadcast channels," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 2, pp. 437–442, Feb. 2011.

[Sol98] M. V. Solodov, "On the convergence of constrained parallel variable distribution algorithm," *SIAM J. Optimization*, vol. 8, no. 1, pp. 187–196, Feb. 1998.

[SPB08] G. Scutari, D. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: A unified view," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1089–1103, Sep. 2008.

[SPLL10] H. Sung, S.-H. Park, K.-J. Lee, and I. Lee, "Linear precoder designs for K-user interference channels," *IEEE Trans. Wirel. Commun.*, vol. 9, no. 1, pp. 291–301, Jan. 2010.

[SR13] S. Schwarz and M. Rupp, “Subspace quantization based combining for limited feedback block-diagonalization,” *IEEE Trans. Wirel. Commun.*, vol. 12, no. 11, pp. 5868–5879, Nov. 2013.

[SRLH11] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[SS74] L. Shapley and H. Scarf, “On cores and indivisibility,” *J. Math. Econ.*, vol. 1, no. 1, pp. 23–37, Mar. 1974.

[SSB+09] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick, “Distributed interference pricing for the MIMO interference channel,” in *Proc. IEEE ICC*, Dresden, Germany, 2009.

[SSH04] Q. Spencer, A. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[TL13] J. Tang and S. Lambotharan, “Interference alignment techniques for MIMO multi-cell interfering broadcast channels,” *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 164–175, Jan. 2013.

[TU] TU Munich Lecture. 3D Computer Vision Script Draft. [Online]. Available: http://campar.in.tum.de/Chair/TeachingWs05ComputerVision

[TV05] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge: Cambridge Univ. Press, 2005.

[V2.10] G. T. V2.0.1, “Evolved Universal Terrestrial Radio Access (E-UTRA): Further advancements for E-UTRA physical layer aspects,” Mar. 2010.

[Ver02] S. Verdu, “Spectral efficiency in the wideband regime,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[YB03] S. Ye and R. Blum, “Optimized signaling for MIMO interference systems with feedback,” *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[YGJK10] C. Yetis, T. Gou, S. Jafar, and A. Kayran, “On feasibility of interference alignment in mimo interference networks,” *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[YJG07] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.

[Yua96] Y. Yuan, "Residence exchange wanted: A stable residence exchange problem," *Eur. J. of Oper. Res.*, vol. 90, no. 3, pp. 536–546, May 1996.

[Zad63] L. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Trans. Autom. Control*, vol. 8, no. 1, pp. 59–60, Jan. 1963.

[ZC10] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[ZC14] Y. Zhong and P. Cao, "Power trading in multi-cell multi-user relay-assisted uplink with private budget limits," in *Proc. of IEEE WCNC*, Apr. 2014.

[ZCJ13a] A. Zappone, P. Cao, and E. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with statistical CSI," in *Proc. of IEEE ChinaSIP*, Jul. 2013.

[ZCJ13b] ——, "Energy efficiency optimization in multiuser relay-assisted MIMO systems," in *Proc. of ACSSC*, Nov. 2013.

[ZCJ14a] ——, "Green resource allocation in relay-assisted MIMO systems with statistical channel state information," in *Prof. of IEEE ICASSP*, May 2014.

[ZCJ14b] ——, "Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan 2014.

[ZCJ14c] ——, "Low-complexity energy efficiency optimization with statistical CSI in two-hop MIMO systems," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1398–1402, Nov 2014.

[ZKM+14] G. Zheng, I. Krikidis, C. Masouros, S. Timotheou, D.-A. Toumpakaris, and Z. Ding, "Rethinking the role of interference in wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 152–158, Nov 2014.