Transmission Design for Reconfigurable Intelligent Surface-Aided Wireless Systems

Gui Zhou

Doctor of Philosophy

School of Electronic Engineering and Computer Science

Queen Mary University of London

May 11, 2022

Abstract

The performance benefits promised by Reconfigurable Intelligent Surface (RIS) are strongly dependent on the availability of highly accurate and up-to-date Channel State Information (CSI), which, however, is challenging to obtain. This thesis proposes efficient transceiver designs for a variety of CSI challenges such as worst channel condition in multicast systems, channel uncertainties caused by the presence of random blockages in millimeter wave systems, by the channel estimation error in downlink systems and by the presence of eavesdropper in security systems.

First, a low-complexity transceiver design scheme in the multicast systems is proposed. In order to ensure the quality of service of the user with the worst channel condition, this thesis deploys an RIS to enhance signal coverage, and proposes two novel and efficient algorithms to jointly design the Base Station (BS) and RIS beamformings. The low-complexity algorithm with closed-form solutions is proved to have the same performance as the general second-order cone programming based algorithm.

Second, novel fairness-oriented robust transceiver design schemes are proposed in RIS-aided millimeter wave systems. The channel uncertainty caused by the random blockages is analyzed, and the metric of maximum outage probability minimization is proposed. To address this problem, stochastic optimization techniques are adopted and closed-form solutions of the BS and RIS beamformings are then obtained. The proposed stochastic optimization algorithms are proved to converge to the set of stationary points.

Third, a framework of robust transceiver design scheme is proposed to address the channel uncertainty caused by the cascaded BS-RIS-user channel estimation error. Two cascaded channel error models are analyzed, and the correspondingly two robust beamforming design problems are proposed. The optimization theory is used to address the complex non-convex optimization problems. The numerical results show that the pro-

posed robust scheme can effectively resist channel uncertainty.

Finally, robust transceiver design schemes are proposed in RIS-aided physical layer security systems. The schemes analyze the channel uncertainties caused by the eavesdropper who launches an active attack, and by the eavesdropper conducting passive eavesdropping. Numerical results show that the negative effect of the eavesdropper's channel error is larger than that of the legitimate user.

Acknowledgments

Foremost, I would like to express my deepest gratitude to my primary supervisor Prof. Cunhua Pan for his prudent guidance, continuous support and patience during my PhD journey, and to my co-supervisor Dr. Kok Keong (Michael) Chai for his kind help and continuous support in the final year of my PhD. With their professional supervision, I could figure out research problems, deal with challenges, and then contribute a list of publications.

Secondly, I would like to thank Prof. Marco Di Renzo (Université Paris-Saclay), Prof. A. Lee Swindlehurst (University of California Irvine) and Prof. Petar Popovski (Aalborg University) for their technical support and research feedback. Through our collaboration, I gained special research experience and a thorough understanding of my research work.

Thirdly, I would also like to express my thanks to supportive teams at QMUL, especially Ms. Melissa Yeo, Mr. Edward Hoskins, Ms. June and other EECS staff, for their support of the student community. I am deeply grateful to my dear friend Miss Lanting Zha for always being able to enjoy every moment of life together. I would also like to thank my friends Dr. Yan Liu, Dr. Jiadong Yu, Dr. Yujian Gan, Dr. Ruikang Zhong, Dr. Yifeng Xiong, for their continuous encouragement and accompany during my PhD research.

Finally, I am very grateful to my family for their unconditional support and my own persistence. My graduate study path is tortuous (from 2015 to 2022): I once tried to study continuously for a joint degree of master and doctoral, but time has proved that this shortcut will consume more time and energy. I am very glad that I quit the previous doctoral program bravely in time, reapplied for the current one, and finally achieved the doctoral degree with satisfactory academic profile. I cherish this unusual brand of my life. The straight path is a way, the tortuous path is a way, even if going back, it is also a route of retreat, but I will never make progress if I hesitate to stand still.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5G	Fifth-Generation
6G	Sixth-Generation
ADMM	Alternating Direction Method of Multipliers
AF	Amplify-and-Forward
AN	Artificial Noise
AO	Alternating Optimization
AoA	Angle of Arrival
AoD	Angle of Departure
APG	Accelerated Projected Gradient
AWGN	Additive White Gaussian Noise
BCD	Block Coordinate Descent
BS	Base Station
BSs	Base Stations
BTI	Bernstein-Type Inequality
CBRUT	BS-RIS-User Channels at the Transmitter
CCM	Complex Circle Manifold
CCP	Convex-Concave Procedure
CDF	Cumulative Distribution Function
CoMP	COordinated MultiPoint
CPU	Central Processing Unit
CSCG	Circularly Symmetric Complex Gaussian

CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
DCSIT	Direct Channel State Information at the Transmitter
DRL	Deep Reinforcement Learning
ED	EavesDropper
FCU	Full Channel Uncertainty
FDD	Frequency Division Duplex
GA	Genetic Algorithms
INs	Interference-Plus-Noises
KKT	Karush-Kuhn-Tucker
KPIs	Key Performance Indicators
LHS	Left Hand Side
LMI	Linear Matrix Inequality
LoS	Line-of-Sight
LRE	LU-RIS-ED
LRL	LU-RIS-LU
LS	Least Square
LU	Legitimate User
LUs	Legitimate Users
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
MM	Majorization–Minimization or Minorization–Maximization
MMF	Max-Min Fairness
MMSE	Minimum Mean Sum Error
mmWave	Millimeter Wave
MU-MIMO	Multi-user Multiple-Input Single-Output
MU-MISO	Multi-user Multiple-Input Single-Output
NLoS	Non Line-of-Sight
PCU	Partial Channel Uncertainty
PSO	Particle Swarm Optimization

QoS	Quality of Service
RF	Radio Frequency
RIS	Reconfigurable Intelligent Surface
RISs	Reconfigurable Intelligent Surfaces
SAA	Sample Average Approximation
SCA	Successive Convex Approximation
SDP	Semidefinite Program
SDR	Semidefinite Relaxation
SGD	Stochastic Gradient Descent
SINR	Signal to Interference Plus Noise Ratio
SMM	Stochastic Majorization–Minimization
SOC	Second-Order Cone
SOCP	Second-Order Cone Programming
SSCA	Stochastic Successive Convex Approximation
SU-MIMO	Single-User Multiple-Input Multiple-Output
SV	Saleh-Valenzuela
SWIPT	Simultaneous Wireless Information and Power Transfer
TDD	Time Division Duplex
ULA	Uniform Linear Array
UMi	Urban Micro
UPA	Uniform Planar Array

Chapter 1

Introduction

While the fifth-Generation (5G) wireless communication system is under deployment worldwide, research interest has shifted to the future sixth-Generation (6G) wireless system [1]–[3], which targets supporting not only cutting-edge applications like multisensory augmented/virtual reality applications, wireless brain computer interactions, and fully autonomous systems, but also the wireless evolution from "connected things" to "connected intelligence". The required Key Performance Indicators (KPIs), including data rates, reliability, latency, spectrum/energy efficiency, and connection density, will be much more superior to those for 5G. For example, the energy and spectrum efficiency for 6G are expected to be 10-100 times and 5 times over that of 5G, respectively. These KPIs, however, cannot be fully achieved by the existing three-pillar 5G physical layer techniques [4], which include massive Multiple-Input Multiple-Output (MIMO), millimeter Wave (mmWave) communications, and ultra-dense heterogeneous networks. In particular, a large number of antennas along with active Radio Frequency (RF) chains are needed for massive MIMO to achieve high spectrum efficiency, which leads to high energy consumption and hardware cost. Moreover, moving to the mmWave frequency band renders the electromagnetic waves more susceptible to blockage by obstacles such as furniture and walls in indoor scenarios. In addition, more costly RF chains and sophis-

ticated hybrid precoding are necessary for mmWave communication systems. The dense deployment of small Base Stations (BSs) also incurs high maintenance cost, network energy consumption, and hardware cost due to high-speed backhaul links. Furthermore, sophisticated interference management techniques are necessary in ultra-dense networks.

Conventionally, the wireless environment is perceived as a randomly varying entity that impairs the signal quality due to uncontrolled reflections, refractions and unexpected interference. Although a plethora of physical layer techniques such as advanced modulation/demodulation and precoding/decoding schemes have been developed at the endpoints of communication links to compensate for these negative impacts, it is undeniable that a certain level of saturation has been reached in terms of achievable data rate and performance reliability. Huge performance gains are expected when regarding the wireless environment as an additional variable to optimize. This is made possible by exploiting the new and revolutionary idea of Reconfigurable Intelligent Surfaces (RISs) [5]–[9], which are capable of reconfiguring the wireless propagation environment into a transmission medium with more desirable characteristics.

Reconfigurable Intelligent Surfaces

An RIS is a planar surface composed of a large number of passive and low-cost reflecting elements, each of which can impose an independent phase shift/amplitude on the impinging electromagnetic signals in a fully customized way. Thanks to recent advances in metamaterials [10], the phase shifts imposed on the incident electromagnetic signals can be adjusted in real-time in reaction to the rapid variations in the wireless propagation environment. By judiciously tuning the phase shifts of the RIS, the signals re-radiated from the RIS can be added constructively with the signals from other paths to enhance the received signal power at the desired users, or can be combined destructively to mitigate the undesired signals at unintended users such as multiuser interference and signal leakage at the eavesdroppers.

Unlike conventional relaying techniques, an RIS is free from RF chains and amplifiers,

and thus entails much reduced power consumption and hardware cost. Furthermore, due to their passive nature, RIS can be fabricated with a low profile, light weight, and limited thickness, which enables them to be readily layered on surfaces available in the environment, including building facades, ceilings, street lamps, and so on.

Furthermore, the RIS, as a new concept beyond conventional massive MIMO systems, maintains all the advantages of massive MIMO systems, such as being capable of focusing large amounts of energy in three-dimensional space which paves the way for wireless charging, remote sensing and data transmissions. However, the differences between RIS and massive MIMO are also obvious. Firstly, the RIS can be densely deployed in indoor spaces, making it possible to provide high data rates for indoor devices in the way of near-field communications [11]. Secondly, in contrast to conventional active antenna array equipped with energy-consuming RF chains and power amplifiers, the RIS with passive reflection elements is cost-effective and energy-efficient [12], which enables RIS to be a prospective energy-efficient technology in green communications. Thirdly, as the RIS just reflects the signal in a passive way, there is no thermal noise or self-interference imposed on the received signal as in conventional full-duplex relays.

1.1 Background and Motivation

An RIS can be used to extend the coverage area, improve the channel rank, mitigate the interference, enhance the reliability, and improve the positioning accuracy. However, this comes with a price of higher restrictions imposed on the quality of the Channel State Information (CSI) required at the transmitter, specifically in the downlink mode. This stems from the necessity to deal with multiuser interference through preprocessing, as the receivers are generally decentralized and cannot cooperate. The ability to provide highly accurate and up-to-date CSI at the transmitter is questionable. Therefore, considerable effort has been devoted to characterize and improve the performance in the presence of CSI challenges. This thesis focuses on developing novel transmit beamforming strategies for some applications of RIS assisted wireless communications with a variety of CSI

challenges. Specifically, four practically relevant scenarios are considered: a multicast communication system with the worst channel condition, a mmWave communication system with CSI uncertainty caused by the presence of random blockages, a downlink communication system with CSI uncertainty caused by the cascaded channel estimation error and a secrecy communication system with CSI uncertainty caused by the presence of eavesdropper.

1.1.1 CSI Challenge in Multicast Communications

Background: Most of the contributions only investigated the performance benefits of deploying an RIS in unicast transmissions, where the Base Station (BS) sends an independent data stream to each user. However, unicast transmissions will cause severe interference and high system complexity when the number of users is large. To address this issue, the multicast transmission based on content reuse [13] (e.g., identical content may be requested by a group of users simultaneously) has attracted wide attention, especially for the application scenarios such as popular airport or video conference. From the perspective of operators, it can be envisioned that multicast transmission is capable of effectively alleviating the pressure of tremendous wireless data traffic and play a vital role in the next generation wireless networks. Therefore, it is necessary to explore the potential performance benefits brought by multiple RISs during the multigroup multicast transmission. Specifically, in multicast systems, the data rate of each group is limited by the user with the worst-channel gains. Hence, the RIS can be deployed to improve the channel conditions of the worst-case user, which can be significantly improve the system performance. A common performance metric in multicast transmissions is the Max-Min Fairness (MMF), where the minimum Signal-to-Interference-plus-Noise-Ratio (SINR) or spectral efficiency of users in each multicasting group or among all multicasting groups is maximized [14]–[18]. The existing contributions of multicast transmission in single-group and multigroup are presented in [14], [15], where the MMF problems are formulated as a fractional Second-Order Cone Programming (SOCP) and solved using SemiDefinite Relaxation (SDR) technique [19].

Motivation: Unfortunately, the popular SDR-based method incurs a high computational complexity ($\mathcal{O}(N^{4.5})$, where N denotes the number of variables) which hinders its practical implementation when the number of design parameters (e.g., precoding matrix and reflection coefficient vector) becomes large. Furthermore, the aforementioned low-complexity techniques designed for the RIS-aided unicast communication schemes cannot be directly applied in the multigroup multicast communication systems since the MMF metric is a non-differentiable and complex objective function. Chapter 3 focuses on the investigation of RIS-aided multicast transmission, and the non-differentiable and complex MMF metric is addressed by proposing two efficient algorithms, one of which has low-complexity closed-form solution. To the best of the knowledge, this is the first work exploring the performance benefits of deploying an RIS in multigroup multicast communication systems.

1.1.2 CSI Challenge in mmWave Communications

Background: mmWave communication is expected to be a promising technology to meet the growing demand for data rate in current and future wireless networks. mmWave communication systems are affected by severe signal attenuations. Thanks to the small signal wavelength, however, this can be compensated by deploying antenna-arrays with a large number of antennas at the transmitters and receivers. In addition, the high-directional beams obtained by utilizing large antenna-arrays can mitigate the inter-user interference. However, mmWave communication systems suffer from high penetration losses [20]–[22]. Hence, mmWave systems are much more susceptible to the presence of spatial blockages than sub-6 GHz systems, and the reliability of the communication links cannot always be guaranteed throughout the whole network [20]–[22].

Motivation: Although the performance advantages of deploying RISs in mmWave systems have been demonstrated in recent contributions, there still exist major open problems to solve. Examples include the following. The authors of [23] have only considered the BS-RIS-user channels and have assumed that the direct BS-user communi-

cation links are completely blocked by obstacles. However, this assumption only applies to the case of static blockages with blockage probability equal to one, but it does not apply in the presence of dynamic blockages since the blockage probability lies in [0,1] [20]–[22]. The numerical results illustrated in [24], for example, have shown that the gain from additional reflections can compensate for the performance loss caused by the presence of random blockages. However, the impact of blockages was not considered in the beamforming design. Most recently, the authors of [25] have considered the design of robust beamforming methods for RIS-aided mmWave communication systems by taking the presence of random blockages into consideration. However, the considered optimization problem is the minimization of the sum outage probability, which cannot ensure the fairness among all the users. Chapter 4 proposes a min-max outage probability problem for fairness in the RIS-aided mmWave systems, which is tackled by proposing two low-complexity stochastic optimization algorithms. To the best of the knowledge, this is the first work that introduces a robust beamforming design for RIS-aided downlink multiuser mmWave systems that relies on the knowledge of large-scale CSI and blockage probability.

1.1.3 CSI Challenge Caused by Channel Estimation

Background: The algorithms developed in most contributions [26]–[32] were based on the assumption of perfect Channel State Information at the Transmitter (CSIT). Unfortunately, it is challenging to estimate the channels for the RIS-aided wireless systems, since RIS is passive and can neither send nor receive pilot symbols. Due to the inevitable channel estimation error, it will induce system performance loss if naively treating the estimated channels as perfect ones. Hence, it is imperative to design robust transmission strategies for the RIS-aided wireless communication systems.

Motivation: To the best of the knowledge, there are only a few contributions in this area [33], [34]. Specifically, the authors of [33] first proposed a worst-case robust design algorithm by assuming that the BS only knew the imperfect RIS-user channels

in a Multi-User Multiple-Input Single-Output (MU-MISO) wireless system. Then, the authors in [34] further proposed a robust secure transmission strategy by also applying the worst-case optimization method when the channels from the RIS to the eavesdroppers were imperfect. However, to implement the above robust design algorithms in [33] and [34], one should rely on the channel estimation where the BS-RIS channels and RIS-user channels should be independently estimated. This is difficult to achieve since several active elements should be installed at the RIS. A more practical robust transmissions under the cascaded channel estimation error needs to be investigated. Chapter 5 proposes a framework of robust beamforming designs to tackle the channel uncertainties caused by the cascaded channel error. To the best of the knowledge, this is the first work to study the robust transmission design based on imperfect cascaded BS-RIS-user channels, which is more practical than the previous works in which imperfect RIS-user channels were considered.

1.1.4 CSI Challenge in Security Communications

Background: Based on the exploration of the differences in channel conditions and interference environment, the RIS can enhance the received signal of Legitimate Users (LUs) and suppress the signal received by the EavesDropper (ED) by changing the reflection direction of the incident signal [31], [34]–[37]. Thus, RIS has the potential of extending the coverage area, mitigating the interference, and improving the physical layer security communication. In general, ED works in two modes: active attacks and passive eavesdropping [38], [39]. In an active attack, in order to mislead the BS to send signals to the ED, the ED pretends to be a LU sending pilot signals to the BS during the channel estimation procedure. Nonetheless, a passive attack is more challenging to tackle since the passive ED can hide itself and its CSI is not available at the BS.

Motivation: To the best of the knowledge, all the existing contributions on the RIS-aided security enhancement were developed under the active attacks, where the BS can acquire the CSI of ED. There is no existing work studying the passive eavesdropping in

RIS-aided secure communication systems. In addition, even for the imperfect CSI under the active attacks, the methods proposed in [34], [37] are only applicable to small-size RIS (i.e., the number of the reflecting elements is less than 10) which can be observed from the numerical results. However, for practical RIS-aided communication systems, the RIS can be equipped with a large number of reflecting elements due to the passive feature of the RIS. Furthermore, more reflecting elements can capture more electromagnetic energy. In addition, RIS has advantages over the conventional massive MIMO and relay in terms of energy efficiency only when the number of the reflecting elements of the RIS is large [29]. Chapter 6 proposes an RIS-aided two-phase secrecy communication scheme to avoid the leakage of useful signals from BS to ED, and effective secure beamforming algorithms are proposed for the active attack mode with imperfect CSI and the passive eavesdropping mode with blind CSI, respectively.

1.2 Outline and Contributions

In the following, the outline of the thesis is presented, in which the key contributions of each chapter are briefly summarized.

Chapter 2 provides an overview of the primary concepts, related literature, and optimization techniques used throughout the thesis. In particular, reviews of literature on RIS assisted wireless communication systems, random blockages in mmWave communication systems and estimation of RIS related channels are first briefly presented. Mathematical preliminaries of optimization techniques used in the proposed algorithms are then provided. Specifically, the overview of RIS element design under the unit-modulus constraints in the existing literature is briefly discussed. Then, the frameworks of the Majorization–Minimization (MM) and Successive Convex Approximation (SCA) methods, which are the basis of the proposed efficient solutions to the considered problems in the thesis, are provided.

Chapter 3 considers downlink multigroup multicast communication systems assisted

by an RIS. It aims for maximizing the sum rate of all the multicasting groups by the joint optimization of the precoding matrix at the BS and the reflection coefficients at the RIS under both the power and unit-modulus constraints. To tackle this non-convex problem, two efficient algorithms under the MM algorithm framework are proposed. Specifically, a concave lower bound surrogate objective function of each user's rate has been derived firstly, based on which two sets of variables can be updated alternately by solving two corresponding SOCP problems. Then, in order to reduce the computational complexity, another concave lower bound function of each group's rate for each set of variables at every iteration is derived, and the closed-form solutions under these loose surrogate objective functions is obtained. Finally, the simulation results demonstrate the benefits in terms of the spectral and energy efficiency of the introduced RIS and the effectiveness in terms of the convergence and complexity of the proposed algorithms.

Chapter 4, in order to improve the robustness of mmWave systems in the presence of random blockages, considers the deployment of multiple RISs to enhance the spatial diversity gain, and the design of robust beamforming schemes based on stochastic optimization methods that minimize the maximum outage probability among multiple users so as to ensure fairness. Under the stochastic optimization framework, the Stochastic Majorization–Minimization (SMM) method and the Stochastic Successive Convex Approximation (SSCA) method are adopted to construct deterministic surrogate problems at each iteration, and to obtain closed-form solutions of the precoding matrix at the BS and the beamforming vectors at the RISs. Both stochastic optimization methods are proved to converge to the set of stationary points of the original stochastic problems. Simulation results show that the proposed robust beamforming for RIS-aided systems can effectively compensate for the performance loss caused by the presence of random blockages, especially when the blockage probability is high.

Chapter 5 studies the robust beamforming based on the imperfect Cascaded BS-RIS-User channels at the Transmitter (CBRUT). Specifically, the transmit power minimization problems are formulated subject to the worst-case rate constraints under the

bounded CSI error model and the rate outage probability constraints under the statistical CSI error model, respectively. After approximating the worst-case rate constraints by using the S-procedure and the rate outage probability constraints by using the Bernstein-type inequality, the reformulated problems can be efficiently solved. Numerical results show that the negative impact of the CBRUT error on the system performance is greater than that of the direct CSI error.

Chapter 6 adopts RIS to enhance the physical layer security in the Rician fading channel where the angular direction of the eavesdropper is aligned with a legitimate user. In this scenario, a two-phase communication system under the active attacks and passive eavesdropping is considered. Particularly, in the first phase, the BS avoids direct transmission to the attacked user. While, in the second phase, other users cooperate to forward signals to the attacked user with the help of RIS and energy harvesting technology. Under the active attacks, an outage constrained beamforming design problem under the statistical cascaded channel error model is investigated, which is solved by using the Bernstein-type inequality. As for the passive eavesdropping, an average secrecy rate maximization problem is formulated, which is addressed by a low-complexity algorithm. Numerical results show that the negative effect of the eavesdropper's channel error is larger than that of the legitimate user.

Chapter 7 presents the conclusion and some thoughts for future work.

1.3 Publications

The thesis is written as a monograph based on four journal papers [J1, J2, J6, J8, J11] and three related conference papers [C1, C2, C3]. All the mentioned papers have already been published. The author of this thesis had the main responsibility for developing the original ideas, formulating the mathematical problems, deriving the analytical equations and algorithms, writing the MATLAB-based simulation codes, generating the numerical results, and writing the papers. The role of all co-authors was to provide valuable

guidance, ideas, comments, criticisms, and support in developing the ideas/algorithms and writing the papers.

In addition to the papers mentioned above, the author of this thesis also published four other journal papers [J3, J4, J5, J7] and one conference paper [C5], which are not included in the thesis. The author was also the co-author of the journal papers [J9, J10, J11] and the conference paper [C4].

Published journal papers

[J1] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang and A. Nallanathan, "Intelligent Reflecting Surface Aided Multigroup Multicast MISO Communication Systems," in IEEE Transactions on Signal Processing, vol. 68, pp. 3236-3251, Apr. 2020. (Chapter 3)

[J2] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang and A. Nallanathan, "A Framework of Robust Transmission Design for IRS-Aided MISO Communications With Imperfect Cascaded Channels," in IEEE Transactions on Signal Processing, vol. 68, pp. 5092-5106, Aug. 2020. (Chapter 5)

[J3] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang, Marco Di Renzo and A. Nallanathan, "Robust Beamforming Design for Intelligent Reflecting Surface Aided MISO Communication Systems," in IEEE Wireless Communications Letters, vol. 9, no. 10, pp. 1658-1662, Oct. 2020.

[J4] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang, M. Elkashlan and Marco Di Renzo, "Stochastic Learning-Based Robust Beamforming Design for RIS-Aided Millimeter-Wave Systems in the Presence of Random Blockages," in IEEE Transactions on Vehicular Technology, vol. 70, no. 1, pp. 1057-1061, Jan. 2021.

[J5] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang and Zhangjie Peng, "Secure Wireless Communication in RIS-Aided MISO System With Hardware Impairments," in IEEE Wireless Communications Letters, vol. 10, no. 6, pp. 1309-1313, Jun. 2021.

[J6] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang, Marco Di Renzo, "Fairness-Oriented Multiple RISs-Aided MmWave Transmission: Stochastic Optimization Methods," in IEEE Transactions on Signal Processing, vol. 70, pp. 1402-1417, Mar. 2022. (Chapter 4)

[J7] **Gui Zhou**, Cunhua Pan, Hong Ren, P. Popovski, and A. L. Swindlehurst, "Channel estimation for RIS-aided multiuser millimeter-wave systems," in IEEE Transactions on Signal Processing, vol. 70, pp. 1478-1492, Mar. 2022.

[J8] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang, Kok Keong Chai, Kai-Kit Wong, "User Cooperation for IRS-aided Secure SWIPT MIMO Systems," accepted by Intelligent and Converged Networks. (Chapter 6)

[J9] Kangda Zhi, Cunhua Pan, **Gui Zhou**, Hong Ren, Maged Elkashlan, Robert Schober, "Is RIS-Aided Massive MIMO Promising with ZF Detectors and Imperfect CSI?," accepted by IEEE Journal on Selected Areas in Communications.

[J10] A. Lee Swindlehurst, **Gui Zhou**, Rang Liu, Cunhua Pan, Ming Li, "Channel Estimation with Reconfigurable Intelligent Surfaces – A General Framework," in Proceedings of the IEEE, early access.

Published conference papers

[C1] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang and A. Nallanathan, "Outage Constrained Transmission Design for IRS-aided Communications with Imperfect Cascaded Channels," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, Taipei, Taiwan, 2020, pp. 1-6. (Chapter 5)

[C2] **Gui Zhou**, Cunhua Pan, Hong Ren, Kezhi Wang and Kok Keong Chai, "RIS-Aided mmWave Transmission: A Stochastic Majorization-Minimization Approach," ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, 2021, pp. 1-6. (Chapter 4)

[C3] **Gui Zhou**, Cunhua Pan, Hong Ren, Kangda Zhi, Sheng Hong and Kok Keong Chai, "User Cooperation for RIS-aided Secure SWIPT MIMO Systems under the passive eavesdropping," 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops), Xiamen, China, 2021, pp. 171-176. (Chapter 6)

[C4] Kangda Zhi, Cunhua Pan, **Gui Zhou**, Hong Ren and Kezhi Wang, "Analysis and Optimization of RIS-aided Massive MIMO Systems with Statistical CSI," 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops), Xiamen, China, 2021, pp. 153-158.

[C5] **Gui Zhou**, Cunhua Pan, Hong Ren and Kezhi Wang, "Channel Estimation for RIS-Aided Millimeter-Wave Massive MIMO Systems : (Invited Paper)," 2021 55th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2021, pp. 698-703.

Journal papers under review

[J11] Cunhua Pan, **Gui Zhou**, Kangda Zhi, Sheng Hong, Tuo Wu, Yijin Pan, Hong Ren, Marco Di Renzo, A. Lee Swindlehurst, Rui Zhang, Angela Yingjun Zhang, "An Overview of Signal Processing Techniques for RIS/IRS-aided Wireless Systems," Dec. 2021, submitted to IEEE Journal of Selected Topics in Signal Processing. [Online]. Available: https://arxiv.org/abs/2112.05989 (Chapter 2)

1.4 Notations

The following mathematical notations and symbols are used throughout this thesis. Vectors and matrices are denoted by boldface lowercase letters and boldface uppercase letters, respectively. The symbols $\mathbf{X}^*$, $\mathbf{X}^{\mathrm{T}}$, $\mathbf{X}^{\mathrm{H}}$, and $||\mathbf{X}||_F$ denote the conjugate, transpose, Hermitian (conjugate transpose), Frobenius norm of matrix $\mathbf{X}$, respectively. The symbol $||\mathbf{x}||_2$ denotes the 2-norm of vector $\mathbf{x}$. The symbols $\mathrm{Tr}\{\cdot\}$, $\mathrm{Re}\{\cdot\}$, $|\cdot|$, $\lambda(\cdot)$, and $\angle(\cdot)$ denote the trace, real part, modulus, eigenvalue, and angle of a complex number, respectively. $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the entries of $\mathbf{x}$ on its main diagonal. $[\mathbf{x}]_m$ denotes

the m-th element of vector $\mathbf{x}$. The Kronecker product between two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{X} \otimes \mathbf{Y}$. $\mathbf{X} \succeq \mathbf{Y}$ means that $\mathbf{X} - \mathbf{Y}$ is positive semidefinite. The symbol $\mathbb{C}$ denotes the complex field, $\mathbb{R}$ denotes the real field, and $\mathrm{j} \triangleq \sqrt{-1}$ is the imaginary unit. The inner product $\langle \bullet, \bullet \rangle : \mathbb{C}^{M \times N} \times \mathbb{C}^{M \times N} \to \mathbb{R}$ is defined as $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Re}\{\mathrm{Tr}\{\mathbf{X}^{\mathrm{H}}\mathbf{Y}\}\}$.

Chapter 2

Literature Review and Mathematical Preliminaries

This chapter first summarizes a review of related works on transmission strategy designs of RIS-aided systems, random blockages in mmWave communication systems, channel estimations of RIS-aided systems and RIS-aided secure communications associated with the scope of the thesis. Then, mathematical preliminaries of optimization techniques, which lay the foundations for the development of solutions for the design problems, are provided.

2.1 Review of Related Works

2.1.1 Transmission Strategy Designs of RIS-Aided Systems

RIS has been investigated in various wireless communication systems. Specifically, the authors in [12] first formulated the joint active and passive beamforming design problem both in downlink single-user and multiple-users Multiple-Input Single-Output (MISO) systems assisted by the RIS, while the total transmit power of the BS is minimized based on the SDR [19] and Alternating Optimization (AO) techniques. In order to reduce the

high computational complexity incurred by SDR, Yu *et al.* proposed low complexity algorithms based on MM (Majorization–Minimization or Minorization–Maximization) algorithm in [30] and manifold optimization in [40] to design reflection coefficients with the targets of maximizing the security capacity and spectral efficiency communications, respectively. Pan *et al.* considered the weighted sum rate maximization problems in multicell MIMO communications [27], Simultaneous Wireless Information and Power Transfer (SWIPT) aided systems [26], Artificial Noise (AN)-aided secure MIMO communications [35], all demonstrating the significant performance gains achieved by deploying an RIS in the networks. A Deep Reinforcement Learning (DRL)-based algorithm [41] and a mobile edge computing-based algorithm [28] were proposed to jointly design the active and passive beamformings in RIS-related systems. In cognitive radio communication systems, the high rate for the secondary user can be achieved with the assistance of the RIS [42].

2.1.2 Random Blockages in mmWave Communication Systems

The spatial blockages can be divided into static blockages (e.g., buildings and other static objects), dynamic blockages (e.g., human beings, vehicles, or moving obstructions) and self-blockages (e.g., hand blocking of the user itself and blockage from other parts of the body). To account for the impact of blockages, statistical models have been proposed to characterize the properties of dynamic blockages and self-blockages [21], [22], [43]. The authors of [44] have developed a distance-dependent blockage probability model, in which the probability that a link is blocked increases exponentially with the length of the link. Furthermore, the authors of [45] and [46] have proposed an approach to predict the blockage probability via machine learning. If the blockage probability is known or is estimated, robust beamforming design strategies can be utilized to address the channel uncertainties caused by the presence of random blockages [47], [48]. Specifically, the authors of [47] have proposed a worst-case robust beamforming design for application to Coordinated MultiPoint (CoMP) systems in which all possible combinations of blockage patterns are considered. Due to the high computational complexity of the method intro-

duced in [47], an outage-minimum strategy based on a stochastic optimization method has been proposed in [48] to improve the robustness of mmWave systems against random blockages. The use of multiple BSs in CoMP systems may be an option to compensate for the performance loss caused by the presence of random blockages by exploiting spatial diversity gains. However, this solution incurs in excessive hardware cost and power consumption. Another possible solution consists of deploying cost-efficient RISs so as to create alternative and reliable communication links in mmWave systems [25].

2.1.3 Channel Estimations of RIS-Aided Systems

In RIS-aided communication systems, there are two types of channels: the direct channel spanning from the BS to the user, and the RIS-related channels. The direct channel can be readily estimated by using conventional channel estimation methods such as the least square algorithm. Hence, most of the existing contributions focused on the channel estimation for the RIS-related channels, which are composed of the channel from the BS to the RIS (BS-RIS channel), and those from the RIS to the users (RIS-user channels). In general, there are two main approaches to estimate the RIS-related channels. The first approach is to directly estimate the RIS-related channels, i.e., estimate BS-RIS channel and RIS-user channels separately [49]. Specifically, in [49], some active channel elements are installed at the RIS to estimate the individual channels. This method, however, has several drawbacks. The active elements may increase the hardware cost and consume extra power, which causes un-affordable burden on the RIS. In addition, the channel information estimated at the RIS needs to be fed back to the BS, which increases the information exchange overhead.

Fortunately, it is observed that the cascaded BS-RIS-user channels, which are the product of the BS-RIS channel and the RIS-user channels, are sufficient for the joint active and passive beamforming design [29]–[32]. As a result, most of the existing contributions focused on the second approach, i.e., the cascaded channel estimation [50]–[53]. Specifically, the channel estimation of the cascaded channel has been investigated both in

the Single-User Multiple-Input Multiple-Output (SU-MIMO) system [50] and the Multi-User Multiple-Input Single-Output (MU-MISO) system [51] by using the Least Square (LS) method. However, the pilot overhead of the LS method is prohibitively high, which is required to be larger than the dimension of cascaded channel and thus scales up with number of reflection elements. In order to reduce the pilot overhead, the authors in [52] exploited the sparse property of the channel matrix and proposed a channel estimation method based on compressed sensing technique. Furthermore, another sparsity representation of the cascaded channel has been found in [53] by using the fact that the height of the BS and the RIS are often the same.

2.1.4 RIS-Aided Secure Communications

Recently, the benefits of RIS in physical layer security under the active attacks have been investigated in the existing literature [31], [34]–[37], [54]. The performance gains of RIS in terms of security capacity was first explored in a simple model consisting of only one single-antenna LU and one single-antenna ED in [31]. Closed-form solutions of the phase shifters of RIS were obtained by leveraging the MM technique in [31], which has a better performance than the classical SDR method. The authors in [35] and [54] respectively extended the results in [31] to a MIMO system and a multicast system where AN was introduced to enhance the security performance. The results in [36] further showed that the AN-aided system without an RIS outperforms the RIS-aided system without AN when the RIS is surrounded by a large number of eavesdroppers. However, all the above contributions were based on the assumption of perfect CSI of the eavesdropping channels at the BS. This assumption is too strict and even impractical. The reasons are twofold: 1) It is challenging to estimate the RIS-related channels since RIS is passive and can neither send nor receive pilot signals. 2) The pilot transmission from the ED to the BS may not be continuous and the corresponding CSI at the BS may be outdated. To deal with the imperfect CSI of the ED, robust transmission methods for secure communication of RIS were proposed in [34], [37]. In particular, the authors in [34] proposed a worst-case robust secure transmission strategy under the assumption of imperfect CSI from the RIS

to the ED. On the other hand, the authors in [37] considered the more practical imperfect cascaded BS-RIS-ED channel and proposed an outage constrained beamforming design method under the statistical CSI error model. However, the imperfect CSI of both LU and ED was not studied in [37].

2.2 Mathematical Preliminaries

This section provides a concise overview of optimization techniques on which proposed solutions for the problems considered in the thesis are based. In this section, the key technical concepts are mainly presented, whereas the rigorous mathematical analysis, which can be found in the mentioned references, is omitted.

2.2.1 Optimization Techniques for Continuous Phase Shifts

The optimization problem with continuous phase shifts constraint can be expressed as

$$\begin{aligned} &\min_{\theta} \ f(\theta) \\ &\text{s.t.} \quad \mathcal{S} = \left\{ \theta | \, \theta = e^{j\varphi}, \varphi \in [0, 2\pi) \right\}. \end{aligned} \tag{2.1}$$

Existing techniques used in RIS-related literature for optimizing Problem (2.1) can be classified into the following categories.

(1) *Relaxation and projection [25], [55], [56]*: The unit modulus constraint on the phase shift can be rewritten as $\mathcal{S}_1 = \{\theta | \, |\theta| = 1, \theta \in \mathbb{C}\}$. The idea of this technique is first to relax the non-convex constraint $\mathcal{S}_1$ to the convex constraint $\tilde{\mathcal{S}}_1 = \{\theta | \, |\theta| \leq 1, \theta \in \mathbb{C}\}$, and then to project the obtained solution onto the unit-modulus constraint $\mathcal{S}_1$. Accordingly, given the solution θ_m of the relaxed problem, the final solution is $\theta_m^\star = e^{j\varphi_m}$, where φ_m is the phase of θ_m.

(2) *SemiDefinite Relaxation (SDR) [12], [24], [57]–[61]*: The SDR method is the most common method for optimizing the phase shifts under constraint $\mathcal{S}_1$, i.e., for continuous phase shifts. Define $\mathbf{V} = \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{H}}$. Then, the unit modulus constraint can be

equivalently written as $\mathbf{V} \succ \mathbf{0}$ and $\text{rank}\,(\mathbf{V}) = 1$. Because of the rank one constraint, the transformed problem is still non-convex. Based on the SDR method, the non-convex rank one constraint is removed. The obtained relaxed problem is a convex SemiDefinite Program (SDP), which can be readily solved by using CVX [62]. In general, the obtained relaxed problem is not a rank-one solution, i.e., $\text{rank}\,(\mathbf{V}) \neq 1$. In this case, the Gaussian randomization method [14] is utilized to obtain a rank-one solution.

(3) *Majorization-Minimization (MM) algorithm [26], [29], [31], [63]–[65]*: The MM algorithm is another widely used technique for optimizing the phase shifts of the RIS. The MM algorithm is an iterative optimization method that approximates a difficult problem as a series of more tractable subproblems that are solved iteratively. Assume that the solution of the subproblem at the t-th iteration is $\boldsymbol{\theta}^t$ and the corresponding objective function is $f(\boldsymbol{\theta}^t)$ [1]. Based on the MM algorithm, a surrogate objective function $\tilde{f}\left(\boldsymbol{\theta}|\,\boldsymbol{\theta}^t\right)$ is constructed, which fulfills the following three conditions: 1) $\tilde{f}\left(\boldsymbol{\theta}^t\big|\,\boldsymbol{\theta}^t\right) = f\left(\boldsymbol{\theta}^t\right)$; 2) $\nabla_{\boldsymbol{\theta}}\tilde{f}\left(\boldsymbol{\theta}|\,\boldsymbol{\theta}^t\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^t} = \nabla_{\boldsymbol{\theta}} f\left(\boldsymbol{\theta}^t\right)\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^t}$; 3) $\tilde{f}\left(\boldsymbol{\theta}|\,\boldsymbol{\theta}^t\right) \geq f\left(\boldsymbol{\theta}\right)$. If these conditions are fulfilled, the sequence of the solutions obtained by solving each subproblem will converge. By replacing the original objective function with the constructed function $\tilde{f}\left(\boldsymbol{\theta}|\,\boldsymbol{\theta}^t\right)$ and removing the constant terms, the subproblem to be solved in each iteration is given by

$$\max_{\boldsymbol{\theta}} \ \text{Re}\left\{\boldsymbol{\theta}^{\text{H}}\mathbf{q}^t\right\} \tag{2.2a}$$

$$\text{s.t.} \quad |\theta_m| = 1, m = 1, \cdots, M, \tag{2.2b}$$

where M is the length of vector $\boldsymbol{\theta}$, and $\mathbf{q}^t$ is a constant complex vector at the t-th iteration. The optimal solution to the optimization problem in (2.2) is

$$\boldsymbol{\theta}^{t+1} = e^{j\arg(\mathbf{q}^t)}. \tag{2.3}$$

This procedure is iterated until convergence according to any criterion of convergence. If the phase shifts of the RIS appear in the constraints of the optimization problem, the

[1] When the beamforming vectors in $\mathbf{W}$ are given, the objective function in (2.1) is denoted by $f(\boldsymbol{\theta})$, and the functions in constraint C1 are denoted by $g_i(\boldsymbol{\theta}), \forall i$.

pricing-based method can be utilized [26].

(4) *Manifold approach [27], [40], [66]–[68]*: There exist different kinds of manifold methods. This part considers the Complex Circle Manifold (CCM) method [27] as an example. The constraint space in $\mathcal{S}_1$ can be regarded as the product of M complex circles, which is a sub-manifold of $\mathbb{C}^M$ given by

$$\mathcal{S}^M \triangleq \left\{\mathbf{x} \in \mathbb{C}^M : |x_l| = 1, l = 1, 2, \cdots, M\right\}, \tag{2.4}$$

where x_l is the l-th element of vector $\mathbf{x}$. The main idea of the CCM method is to derive the gradient descent algorithm based on the manifold space given in (2.4). The optimization problem aims at optimizing the phase shifts to minimize the objective function $\hat{f}(\boldsymbol{\theta})$. The main steps can be summarized as follows.

(a) Computation of the gradient in Euclidean space: The search direction for the minimization problem is the opposite of the gradient of $\hat{f}(\boldsymbol{\theta})$, which is given by $\boldsymbol{\eta}^t = -\nabla_{\boldsymbol{\theta}} \hat{f}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^t}$;

(b) Computation of the Riemannian gradients: The Riemannian gradient of $\hat{f}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^t$ should lie in the tangent space $\mathcal{T}_{\boldsymbol{\theta}^t}\mathcal{S}^M$ [27]. Then, the Riemannian gradient of $\hat{f}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}^t$ is obtained by projecting $\boldsymbol{\eta}^t$ onto $\mathcal{T}_{\boldsymbol{\theta}^t}\mathcal{S}^M$, which yields $\mathbf{P}_{\mathcal{T}_{\boldsymbol{\theta}^t}\mathcal{S}^M}(\boldsymbol{\eta}^t) = \boldsymbol{\eta}^t - \mathrm{Re}\{\boldsymbol{\eta}^{t*} \odot \boldsymbol{\theta}^t\} \odot \boldsymbol{\theta}^t$;

(c) Update over the tangent space: Update the point $\boldsymbol{\theta}^t$ on the tangent space $\mathcal{T}_{\boldsymbol{\theta}^t}\mathcal{S}^M$ as $\bar{\boldsymbol{\theta}}^t = \boldsymbol{\theta}^t + \beta \mathbf{P}_{\mathcal{T}_{\boldsymbol{\theta}^t}\mathcal{S}^M}(\boldsymbol{\eta}^t)$, where β is a constant step size;

(d) Retraction operator: This step aims to map $\bar{\boldsymbol{\phi}}^t$ onto the manifold $\mathcal{S}^M$ using the retraction operator $\boldsymbol{\theta}^{t+1} = \bar{\boldsymbol{\theta}}^t \odot \frac{1}{|\bar{\boldsymbol{\theta}}^t|}$. Through iterating steps 1) to 4) until convergence, the final solution is obtained.

(5) *Element-wise Block Coordinate Descent (BCD) [30], [32], [69], [70]*: The idea of the element-wise BCD algorithm is simple. At the m-th iteration, one reflection coefficient θ_m is optimized by keeping fixed the other reflecting coefficients $\theta_{m'}, m' \neq m, m =$

$1, \cdots, M$. The algorithm ends after M iterations when all the reflection coefficients are optimized one-by-one while keeping the other fixed. The element-wise BCD algorithm is simple since it is simpler to optimize a single variable rather than optimizing M variables simultaneously. However, the complexity may be high when the number of reflecting elements is large.

(6) *Rank-one equivalents [71], [72]*: Similar to the SDR method, by defining $\mathbf{V} = \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{H}}$, the unit modulus constraint can be written as $\mathbf{V} \succ \mathbf{0}$ and $\mathrm{rank}\,(\mathbf{V}) = 1$. The rank-one constraint can be equivalently transformed to

$$\mathrm{tr}\,(\mathbf{V}) - \|\mathbf{V}\|_2 = 0. \tag{2.5}$$

Also, $\mathrm{tr}\,(\mathbf{V}) = \sum_{m=1}^{M} \lambda_m$ and $\|\mathbf{V}\|_2 = \lambda_1$, where λ_m denotes the m-th largest singular value of $\mathbf{V}$. Since $\mathbf{V} \succ \mathbf{0}$ and $\mathbf{V}$ is a non-zero matrix, the equality $\mathrm{tr}\,(\mathbf{V}) - \|\mathbf{V}\|_2 = 0$ holds only when $\lambda_1 > 0$ and $\lambda_m = 0, m = 2, \cdots, M$. Then, at the $(t+1)$-th iteration of the iterative algorithm, a lower-bound for $\|\mathbf{V}\|_2$ at the point $\mathbf{V}^t$ can be derived as

$$\|\mathbf{V}\|_2 \geq \left\|\mathbf{V}^t\right\|_2 + \left\langle \left(\mathbf{V} - \mathbf{V}^t\right), \partial_{\mathbf{V}} \|\mathbf{V}\|_2 |_{\mathbf{V}=\mathbf{V}^t} \right\rangle \triangleq f\left(\mathbf{V}; \mathbf{V}^t\right), \tag{2.6}$$

where $\partial_{\mathbf{V}} \|\mathbf{V}\|_2 |_{\mathbf{V}=\mathbf{V}^t}$ is a subgradient of $\|\mathbf{V}\|_2$ with respect to $\mathbf{V}$ at $\mathbf{V} = \mathbf{V}^t$, which is equal to $\mathbf{u}_1\mathbf{u}_1^{\mathrm{H}}$ with $\mathbf{u}_1$ denoting the eigenvector that corresponds to the largest singular value of $\mathbf{V}^t$.

Based on (2.6), the constraint in (2.5) can be approximated with the following convex constraint

$$\mathrm{tr}\,(\mathbf{V}) - f\left(\mathbf{V}; \mathbf{V}^t\right) \leq \varepsilon, \tag{2.7}$$

where ε is a very small positive constant. Then, using (2.6) and (2.7), one has $0 \leq \mathrm{tr}\,(\mathbf{V}) - \|\mathbf{V}\|_2 \leq \mathrm{tr}\,(\mathbf{V}) - f\left(\mathbf{V}; \mathbf{V}^t\right) \leq \varepsilon$. Hence, when ε tends to zero, $\mathrm{tr}\,(\mathbf{V})$ will approach $\|\mathbf{V}\|_2$, which ensures that the rank-one constraint is fulfilled.

(7) *Alternating Direction Method of Multipliers (ADMM) based algorithm [12], [73],*

[74]: An auxiliary variable $\boldsymbol{\omega}$ is introduced such that $\boldsymbol{\omega} = \boldsymbol{\theta}$, which can be regarded as a copy of $\boldsymbol{\theta}$. The feasible region of constraint C1 is denoted by $\mathcal{B}$, which, by using the indicator function, can be formulated as follows

$$\mathbb{I}_{\mathcal{B}}(\mathbf{W}, \boldsymbol{\theta}) = \begin{cases} f(\mathbf{W}, \boldsymbol{\theta}), & \text{if } \{\mathbf{W}, \boldsymbol{\theta}\} \in \mathcal{B} \\ \infty, & \text{otherwise.} \end{cases} \tag{2.8}$$

Similarly, the feasible region that corresponds to constraint C2, i.e., $\mathcal{S}_1$ can be written as follows

$$\mathbb{I}_{\mathcal{S}_1}(\boldsymbol{\omega}) = \begin{cases} 0, & \text{if } \boldsymbol{\omega} \in \mathcal{S}_1 \\ \infty, & \text{otherwise.} \end{cases} \tag{2.9}$$

Then, the equivalent ADMM reformulation for the optimization problem in (2.1) is

$$\begin{aligned} &\min_{\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\omega}} \ \mathbb{I}_{\mathcal{B}}(\mathbf{W}, \boldsymbol{\theta}) + \mathbb{I}_{\mathcal{S}_1}(\boldsymbol{\omega}) \\ &\text{s.t.} \quad \boldsymbol{\omega} = \boldsymbol{\theta}. \end{aligned} \tag{2.10}$$

The augmented Lagrangian of the optimization problem in (2.10) is

$$\mathcal{L}_{\xi} = \mathbb{I}_{\mathcal{B}}(\mathbf{W}, \boldsymbol{\theta}) + \mathbb{I}_{\mathcal{S}_1}(\boldsymbol{\omega}) + \frac{\xi}{2} \|\boldsymbol{\theta} - \boldsymbol{\omega} + \boldsymbol{\lambda}\|_2^2, \tag{2.11}$$

where $\xi > 0$ is a constant penalty parameter, and $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_M]^T$ is the dual variable vector of the constraint $\boldsymbol{\omega} = \boldsymbol{\theta}$. Based on the ADMM algorithm, the variables $\mathbf{W}, \boldsymbol{\theta}$ and $\boldsymbol{\omega}$ are alternately optimized.

The ADMM algorithm is an iterative approach. In the t-th iteration, given $\mathbf{W}^t, \boldsymbol{\theta}^t$ and $\boldsymbol{\omega}^t$, the variables are updated as follows.

(a) Updating $\boldsymbol{\theta}$: The subproblem for updating $\boldsymbol{\theta}$ is

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \ f(\boldsymbol{\theta}) + \frac{\xi}{2} \left\|\boldsymbol{\theta} - \boldsymbol{\omega}^t + \boldsymbol{\lambda}^t\right\|_2^2 \\ &\text{s.t.} \quad g_i(\boldsymbol{\theta}) \geq D_i, i = 1, \cdots, I. \end{aligned} \tag{2.12}$$

Note that the unit-modulus constraint for $\boldsymbol{\theta}$ is not included in this subproblem, which significantly reduces the complexity of computing $\boldsymbol{\theta}$.

(b) Updating $\mathbf{W}$*:* The subproblem for updating $\mathbf{W}$ is

$$\begin{aligned} &\min_{\mathbf{W}} \ f(\mathbf{W}) \\ &\text{s.t.} \quad g_i(\mathbf{W}) \geq D_i, i = 1, \cdots, I. \end{aligned} \tag{2.13}$$

(c) Updating $\boldsymbol{\omega}$*:* The subproblem for updating $\boldsymbol{\omega}$ is

$$\boldsymbol{\omega}^{t+1} = \arg \min_{\boldsymbol{\omega} \in \mathcal{S}_1} \left\| \boldsymbol{\theta}^{t+1} + \boldsymbol{\lambda}^t - \boldsymbol{\omega} \right\|_2^2. \tag{2.14}$$

The objective of the optimization problem in (2.14) is to project $\boldsymbol{\theta}^{t+1} + \boldsymbol{\lambda}^t$ onto the feasible set $\mathcal{S}_1$, whose solution is $\boldsymbol{\omega}^{t+1} = e^{j \arg(\boldsymbol{\theta}^{t+1} + \boldsymbol{\lambda}^t)}$.

(d) Updating $\boldsymbol{\lambda}$*:* The update of $\boldsymbol{\lambda}$ is $\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t + \boldsymbol{\theta}^{t+1} - \boldsymbol{\omega}^{t+1}$.

(8) *Penalty Convex-Concave Procedure (CCP) [75], [76]*: The unit modulus constraint can be equivalently rewritten as $1 \leq |\theta_m|^2 \leq 1, \forall m$. Using the SCA method, the non-convex constraint $1 \leq |\theta_m|^2$ can be converted into a series of convex constraints, i.e., $1 \leq 2\mathrm{Re}(\theta_m^* \theta_m^t) - \left|\theta_m^t\right|^2$, where θ_m^t is the solution in the t-th iteration. By introducing $2M$ slack variables $\mathbf{b} = [b_1, \cdots, b_{2M}]$, the phase shift optimization problem can be rewritten as

$$\begin{aligned} &\min_{\boldsymbol{\theta}, \mathbf{b} \geq \mathbf{0}} \ f(\boldsymbol{\theta}) - \lambda^t \sum\nolimits_{m=1}^{2M} b_m \\ &\text{s.t.} \quad g_i(\boldsymbol{\theta}) \geq D_i, i = 1, \cdots, I, \\ &\qquad \left|\theta_m^t\right|^2 - 2\mathrm{Re}(\theta_m^* \theta_m^t) \leq b_m - 1, \forall m, \\ &\qquad |\theta_m|^2 \leq 1 + b_{m+M}, \forall m, \end{aligned} \tag{2.15}$$

where λ^t is the regularization factor to control the feasibility of the constraints in the t-th iteration. After some transformations, the optimization problem in (2.15) can be solved by CVX, and the detailed procedure to solve this problem, which can be found

in references [75], [76], is omitted here for brevity.

(9) *Barrier function penalty [77]–[79]*: The unit modulus constraint can be equivalently written as $\mathrm{tr}\left(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{H}}\right) = M$ and $\|\boldsymbol{\theta}\|_\infty \leq 1$. Since $\|\boldsymbol{\theta}\|_\infty$ is non-differentiable, the l_p norm with large p can be used to approximate it, i.e., $\|\boldsymbol{\theta}\|_\infty = \lim_{p\to\infty} \|\boldsymbol{\theta}\|_p$. To deal with the constraint $\|\boldsymbol{\theta}\|_p \leq 1$, the log barrier function $F(x)$ (which is self-concordant and can converge quickly) can be used to approximate the penalty of violating the l_p constraint, as

$$F(x) = \begin{cases} -\frac{1}{\kappa}\ln(x), & x > 0, \\ \infty, & x \leq 0, \end{cases} \tag{2.16}$$

where $\kappa > 0$ is the barrier function penalty factor, which does not affect the solution of the approximation problem due to the fact that this merely scales the objective function. For simplicity, constraint C1 is ignored. Accordingly, the phase shift optimization problem can be reformulated as

$$\begin{aligned} \min_{\boldsymbol{\theta}} \ & G(\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + F\left(1 - \|\boldsymbol{\theta}\|_p\right) \\ \text{s.t.} \ & \mathrm{tr}\left(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{H}}\right) = M. \end{aligned} \tag{2.17}$$

Due to the non-convex constraint, the optimization problem in (2.17) is still non-convex. To circumvent this issue, a possible solution is to utilize the gradient and projection method, which provides a low complexity but suboptimal solution. Specifically, the gradient of the objective function $G(\boldsymbol{\theta})$ can be formulated as

$$\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta}) = \frac{\|\boldsymbol{\theta}\|_p^{1-p}}{2\kappa\left(1 - \|\boldsymbol{\theta}\|_p\right)}\boldsymbol{\xi} + \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}), \tag{2.18}$$

where $\boldsymbol{\xi} = [\theta_1|\theta_1|^{p-2}, \cdots, \theta_M|\theta_M|^{p-2}]^T$.

Since the problem formulation in (2.17) is a minimization problem, the search direction is opposite to the direction of the gradient in (2.18). Let $\boldsymbol{\theta}^{(i)}$ denote $\boldsymbol{\theta}$ at the i-th iteration, the search direction in the i-th iteration is $\mathbf{d}_{\mathrm{gd}}^{(i)} = -\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(i)}}$ Then, this

search direction $\mathbf{d}_{\text{gd}}^{(i)}$ is projected onto the tangent plane of $\text{tr}\left(\boldsymbol{\theta}\boldsymbol{\theta}^{\text{H}}\right) = M$, as

$$\mathbf{d}_{p}^{(i)} = \mathbf{d}_{\text{gd}}^{(i)} - \frac{\left(\mathbf{d}_{\text{gd}}^{(i)}\right)^{T}\left(\boldsymbol{\theta}^{(i)}\right)^{*}\boldsymbol{\theta}^{(i)}}{\left\|\boldsymbol{\theta}^{(i)}\right\|^{2}}. \tag{2.19}$$

Then, the update of $\boldsymbol{\theta}$ in the $(i+1)$-th iteration is

$$\boldsymbol{\theta}^{(i+1)} = (1-\alpha^{\star})\boldsymbol{\theta}^{(i)} + \alpha^{\star}\sqrt{M}\frac{\boldsymbol{d}_{p}^{(i)}}{\left\|\boldsymbol{d}_{p}^{(i)}\right\|^{2}}, \tag{2.20}$$

where the parameter $\alpha^{\star}$ is obtained by

$$\alpha^{\star} = \arg\max_{\alpha} f\left((1-\alpha)\boldsymbol{\theta}^{(i)} + \alpha\sqrt{M}\frac{\boldsymbol{d}_{p}^{(i)}}{\left\|\boldsymbol{d}_{p}^{(i)}\right\|^{2}}\right). \tag{2.21}$$

(10) *Accelerated Projected Gradient (APG) [80]–[82]*: For simplicity, constraint C1 is ignored and only the optimization of the phase shifts is considered. A projection operator $\mathbb{P}_{\mathcal{S}_1}$ is defined as

$$\hat{\boldsymbol{\theta}} = \mathbb{P}_{\mathcal{S}_1}(\boldsymbol{\theta}) \Leftrightarrow \hat{\theta}_m = \begin{cases} \theta_m/|\theta_m|, \text{if } \theta_m \neq 0 \\ 1, \qquad \text{otherwise.} \end{cases} \tag{2.22}$$

Then, the update of the phase shifts in the $(i+1)$-th iteration is given by

$$\boldsymbol{\theta}_{i+1} = \mathbb{P}_{\mathcal{S}_1}\left(\mathbf{z}_i - \frac{1}{\gamma_i}\nabla_{\boldsymbol{\theta}} f\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\mathbf{z}_i}\right), \tag{2.23}$$

where $\mathbf{z}_i = \boldsymbol{\theta}_i + \alpha_i\left(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1}\right)$, and α_i is updated as

$$\alpha_i = \frac{\xi_{i-1}-1}{\xi_i}, \xi_i = \frac{1+\sqrt{1+4\xi_{i-1}^{2}}}{2}. \tag{2.24}$$

In (2.23), γ_i is obtained by using the backtracking line search method [83].

(11) *Gradient descent approach [84], [85]*: When the objective function $f(\boldsymbol{\theta})$ is differ-

entiable, the optimization problem can be solved by using the gradient descent method. Specifically, let $\boldsymbol{\theta}^t$ be the phase shift vector at the t-th iteration. Then, the optimization variable $\boldsymbol{\theta}$ at the $(t+1)$-th iteration is updated as

$$\boldsymbol{\theta}^{t+1} = \exp\left(j \arg\left(\boldsymbol{\theta}^t - \mu \nabla_{\boldsymbol{\theta}} f\left(\boldsymbol{\theta}\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^t}\right)\right), \tag{2.25}$$

where μ is the step size and the arg operator is used for satisfying the unit-modulus constraint.

(12) *Heuristic methods [86]–[89]*: When the objective function is analytically involving, the above-mentioned algorithms may not be applicable or the computation of the gradient may be time-consuming. Possible solutions to circumvent this issue include the use of heuristic methods such the Genetic Algorithms (GA) or the Particle Swarm Optimization (PSO) methods. More details can be found in [86]–[89].

(13) *Deep reinforcement learning [41], [90], [91]*: Machine learning methods can also be applied to optimize the phase shifts of the RIS. A suitable approach is the use of deep reinforcement learning. In fact, unlike supervised learning methods that require a large number of training labels, deep reinforcement learning based methods do not need training labels and can learn and operate in an online manner. Examples of application of deep reinforcement learning to the optimization of RIS-aided communications can be found in [41], [90], [91].

Simulation results: Fig. 2.1 illustrates the performance of the different algorithms discussed in this article in terms of sum rate and Central Processing Unit (CPU) run time. All algorithms are represented by the numbers they are introduced above. The simulation setups: BS with $N = 10$ antennas communicates with $K = 4$ users via RIS equipped with $M = 100$ reflection elements at SNR of 5 dB. It can be seen that most of the algorithms for which a closed-form solution for the phase shifts can be found at each iteration (algorithms 3-5, 7, 9-11) provide a high sum rate with a low CPU time (around 100 seconds). However, the time-consuming algorithms (algorithms 1,2,6,8), which are

implemented by using CVX, are more flexible to address optimization problems with complex constraints, such as Quality of Service (QoS) constraints.

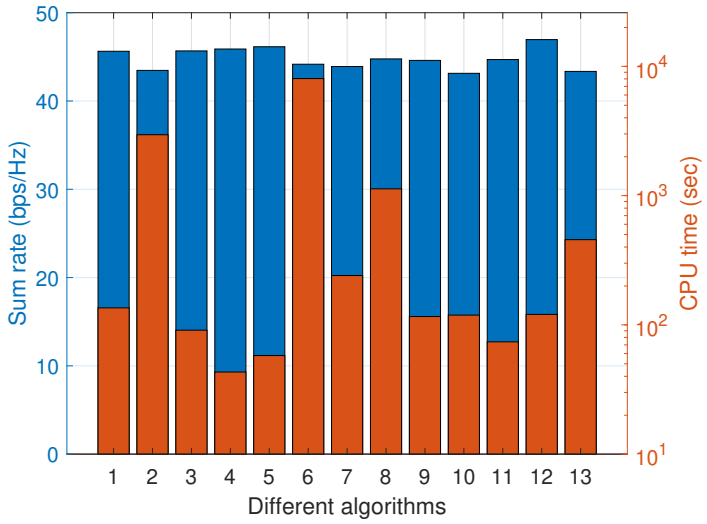


Figure 2.1: Sum rate and CPU time consumption of different algorithms.

2.2.2 Majorization-Minimization Method

The aim of the MM method [92], [93] is to find an easy-to-solve surrogate problem with a surrogate objective function, then optimize it instead of the original complex one. Specifically, suppose that $f(\mathbf{x})$ is the original objective function which needs to be minimized over a convex set $\mathcal{S}_x$. Let $\widetilde{f}(\mathbf{x}|\mathbf{x}^t)$ denote a real-valued function of variable $\mathbf{x}$ with given $\mathbf{x}^t$. The function $\widetilde{f}(\mathbf{x}|\mathbf{x}^t)$ is said to majorize $f(\mathbf{x})$ at a given point $\mathbf{x}^t$ if they satisfy the following assumptions [93]:

Assumption A

$$(\text{A1}) : \widetilde{f}(\mathbf{x}^t|\mathbf{x}^t) = f(\mathbf{x}^t), \forall \mathbf{x}^t \in \mathcal{S}_x.$$

$$(\text{A2}) : \widetilde{f}(\mathbf{x}|\mathbf{x}^t) \geq f(\mathbf{x}), \forall \mathbf{x}, \mathbf{x}^t \in \mathcal{S}_x.$$

$$(\text{A3}) : \widetilde{f}'(\mathbf{x}|\mathbf{x}^t;\mathbf{d})|_{\mathbf{x}=\mathbf{x}^t} = f'(\mathbf{x}^t;\mathbf{d}), \forall \mathbf{d} \text{ with } \mathbf{x}^t + \mathbf{d} \in \mathcal{S}_x.$$

$$(\text{A4}) : \widetilde{f}(\mathbf{x}|\mathbf{x}^t) \text{ is continuous in } \mathbf{x} \text{ for } \forall \mathbf{x}^{t-1} \in \mathcal{S}_x.$$

where $f^{'}(\mathbf{x}^t;\mathbf{d})$, defined as the direction derivative of $f(\mathbf{x}^t)$ in the direction $\mathbf{d}$, is

$$f^{'}(\mathbf{x}^t;\mathbf{d}) = \lim_{\lambda\to 0}\frac{f(\mathbf{x}^t+\lambda\mathbf{d}) - f(\mathbf{x}^t)}{\lambda}.$$

The Assumptions (A1)-(A3) indicate that the surrogate function $\hat{f}(\mathbf{x},\mathbf{x}^{t-1})$ is a local upper bound of the original function $f(\mathbf{x})$ around the feasible point $\mathbf{x}^{t-1}$. The Assumption (A3) is a consistency condition for the first-order directional derivative.

To ensure the convergence of the MM algorithm, the following conditions also need to be fulfilled [94].

Assumption: B

(B1) : The feasible set $\mathcal{S}_x$ and the channel realizations are bounded.

(B2) : The functions $\hat{f}(\mathbf{x}|\mathbf{x}^{t-1})$ and $f(\mathbf{x})$, their derivatives, and their second-order derivatives are uniformly bounded.

2.2.3 Successive Convex Approximation Method

The surrogate functions employed by the SCA method do not need to be upper bound of the original function but they only need to preserve the first-order properties of the original function. Accordingly, the surrogate function, which is denoted by $\hat{f}(\cdot)$, needs to satisfy Assumption B and the following assumptions [95].

Assumption C

$$\begin{aligned}
&\text{(C1)} : \hat{f}\left(\mathbf{x}|\mathbf{x}^{t-1}\right) \text{ is strongly convex in } \mathbf{x} \text{ for } \forall \mathbf{x}^{t-1}\in\mathcal{S}_x.\\
&\text{(C2)} : \hat{f}(\mathbf{x}^{t-1}|\mathbf{x}^{t-1}) = f(\mathbf{x}^{t-1}), \forall \mathbf{x}^{t-1}\in\mathcal{S}_x.\\
&\text{(C3)} : \nabla_{\mathbf{x}}\hat{f}(\mathbf{x}^{t-1}|\mathbf{x}^{t-1}) = \nabla_{\mathbf{x}} f(\mathbf{x}^{t-1}), \forall \mathbf{x},\mathbf{x}^{t-1}\in\mathcal{S}_x,
\end{aligned}$$

where $\nabla_{\mathbf{x}}\hat{f}(\cdot)$ denotes the gradient operation applied to complex-valued functions [96].

Assumption C cannot ensure that the sequences of the approximate objective values are monotonically decreasing at each iteration. To guarantee the convergence of the algorithm, however, the variables can be updated, at each iteration, by choosing an appropriate step size that ensures that the objective value decreases. Based on the above assumptions, the proximal gradient-like approximation is chosen to construct the surrogate function, which is [97]

$$\hat{f}\left(\mathbf{x}|\mathbf{x}^{t-1}\right) = f\left(\mathbf{x}^{t-1}\right) + \nabla_{\mathbf{x}} f(\mathbf{x}^{t-1})^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^{t-1}) + \frac{\tau^t}{2}||\mathbf{x} - \mathbf{x}^{t-1}||^2, \tag{2.26}$$

where τ^t can be any positive number.

Chapter 3

Transmission Design under Perfect Channel State Information

This chapter considers an RIS-assisted multigroup multicast transmission system in which a multiple-antenna BS transmits independent information data streams to multiple groups, and the single-antenna users in the same group share the same information and suffer from interference from those signals sent to other groups. The main contributions of this chapter are summarized as follows:

- To the best of the knowledge, this is the first work exploring the performance benefits of deploying an RIS in multigroup multicast communication systems. Specifically, the precoding matrix and the reflection coefficient vector are jointly optimized to maximize the sum rate of all the multicasting groups, where the rate of each multicasting group is limited by the minimum rate of users in the group. This formulated problem is much more challenging than previous problems considered in unicast systems since the considered objective function is non-differentiable and complex due to the nature of the multicast transmission mechanism. In addi-

tion, the highly coupled variables and complex sum rate expression aggravates the difficulty to solve this problem.

- The formulated problem is solved efficiently in an iterative manner based on the alternating optimization method under the MM algorithm framework. Specifically, the original non-concave objective function is firstly minorized by a surrogate function which is biconcave of precoding matrix and reflection coefficient vector, and then apply the alternating optimization method to decouple those variables. At each iteration of the alternating optimization method, the subproblem corresponding to each set of variables is reformulated as an SOCP problem by introducing auxiliary variables, which can help to transform the non-differentiable concave objective function into a series of convex constraints.

- To further reduce the computational complexity, the MM method is used to derive closed-form solutions of each subproblem, instead of solving the complex SOCP problems with a high complexity at each iteration. Specifically, the log-sum-exp lower bound is firstly applied to approximate the non-differentiable concave objective function, yielding a differentiable concave function. Then, a tractable surrogate objective function of the log-sum-exp function is derived, based on which the closed-form solutions of each subproblem is obtained. Finally, it is proved that the proposed algorithm is guaranteed to converge and the solution sequences generated by the algorithm converge to KKT points.

- Finally, the simulation results demonstrate the superiority of the RIS-assisted multigroup multicast system over conventional massive MIMO systems in terms of the spectral efficiency and energy efficiency. The convergence and the low complexity of the proposed algorithms have also been illustrated.

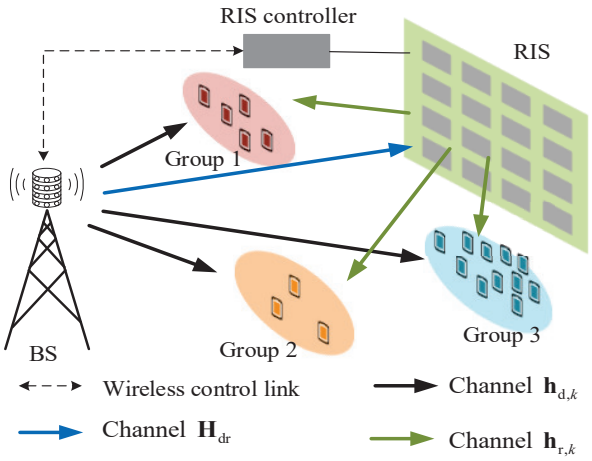


Figure 3.1: An RIS-aided multigroup multicast communication system.

3.1 System Model

3.1.1 Signal Transmission Model

As shown in Fig. 3.1 [1], an RIS-aided multigroup multicast MISO communication system is considered. There is a BS with N transmit antennas serving G multicasting groups. Users in the same group share the same information data and the information data destined for different groups are independent and different, which means there exists inter-group interference. Let us define the set of all multicast groups by $\mathcal{G} = \{1, 2, ..., G\}$. Assuming that there are $K(K \geq G)$ users in total, the user set belonging to group $g \in \mathcal{G}$ is denoted as $\mathcal{K}_g$ and each user can only belong to one group, i.e., $\mathcal{K}_i \cap \mathcal{K}_j = \emptyset, \forall i, j \in \mathcal{G}, i \neq j$. The transmit signal at the BS is

$$\mathbf{x} = \sum_{g=1}^{G} \mathbf{f}_g s_g, \tag{3.1}$$

where s_g is the desired independent Gaussian data symbol of group g and follows $\mathbb{E}[|s_g|^2] = 1$ as well as $\mathbf{f}_g \in \mathbb{C}^{N\times 1}$ is the corresponding precoding vector that maps the scalar symbol s_g to the N-antenna BS. Let us denote the collection of all precoding vectors as $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_G] \in \mathbb{C}^{N\times G}$ satisfying the power constraint $\mathcal{S}_F = \{\mathbf{F} \mid \mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right) \leq$

[1] It is better that the physical locations of multiple users in multicast group are adjacent, otherwise the interference between groups will be severe.

$P_{\mathrm{T}}\}$, where P_{T} is the maximum available transmit power at the BS.

In the multigroup multicast system, an RIS is employed with the goal of enhancing the received signal strength of users by reflecting signals from the BS to the users. It is assumed that the signal power of the multi-reflections (i.e., reflections more than once) on the RIS is ignored due to the severe path loss [12]. Denote M as the number of the reflection elements on the RIS, then the reflection coefficient matrix of the RIS is modeled by a diagonal matrix $\mathbf{E} = \mathrm{diag}([e_1, \cdots, e_M]^{\mathrm{T}}) \in \mathbb{C}^{M\times M}$, where $|e_m|^2 = 1, \forall m = 1, \cdots, M$ [12]. Please note that the design of the practical reflection amplitude which was modeled as a function of the phase shifts [98] is more complex and will be investigated in future work. The channels spanning from the BS to user k, from the BS to the RIS, and from the RIS to user k are denoted by $\mathbf{h}_{\mathrm{d},k} \in \mathbb{C}^{N\times 1}$, $\mathbf{H}_{\mathrm{dr}} \in \mathbb{C}^{M\times N}$, and $\mathbf{h}_{\mathrm{r},k} \in \mathbb{C}^{M\times 1}$, respectively.

It is assumed that the CSI is perfectly known at the BS. The BS is responsible for designing the reflection coefficients of the RIS and sends them back to the RIS controller as shown in Fig. 3.1. As a result, the received signal of user $k \in \mathcal{K}_g$ belonging to group g is

$$y_k = (\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\mathbf{E}\mathbf{H}_{\mathrm{dr}}) \sum_{g=1}^{G} \mathbf{f}_g s_g + n_k, \tag{3.2}$$

where n_k is the received noise at user k, which is an Additive White Gaussian Noise (AWGN) following Circularly Symmetric Complex Gaussian (CSCG) distribution with zero mean and variance σ_k^2. Then, its achievable data rate (bps/Hz) is given by

$$R_k = \log_2\left(1 + \frac{|(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\mathbf{E}\mathbf{H}_{\mathrm{dr}})\mathbf{f}_g|^2}{\sum_{i\neq g}^{G} |(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\mathbf{E}\mathbf{H}_{\mathrm{dr}})\mathbf{f}_i|^2 + \sigma_k^2}\right). \tag{3.3}$$

Denoting by $\mathbf{H}_k = \begin{bmatrix} \mathrm{diag}(\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}})\mathbf{H}_{\mathrm{dr}} \\ \mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} \end{bmatrix} \in \mathbb{C}^{(M+1)\times N}$ the equivalent channel spanning from the BS to user k and by $\mathbf{e} = [e_1, \cdots e_M, 1]^{\mathrm{T}} \in \mathbb{C}^{(M+1)\times 1}$ the equivalent reflection

coefficient vector, then

$$|(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\mathbf{E}\mathbf{H}_{\mathrm{dr}})\mathbf{f}_g|^2 = |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g|^2, \tag{3.4}$$

$$\sum_{i\neq g}^{G} |(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\mathbf{E}\mathbf{H}_{\mathrm{dr}})\mathbf{f}_i|^2 = \sum_{i\neq g}^{G} |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2 + \sigma_k^2. \tag{3.5}$$

Note that $\mathbf{e}$ belongs to the set $\mathcal{S}_e = \{\mathbf{e} \mid |e_m|^2 = 1, 1 \leq m \leq M, e_{M+1} = 1\}$. Then, the data rate expression in (3.3) can be rewritten in a compact form as

$$R_k(\mathbf{F}, \mathbf{e}) = \log_2\left(1 + \frac{|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g|^2}{\sum_{i\neq g}^{G} |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2 + \sigma_k^2}\right). \tag{3.6}$$

Due to the nature of the multicast mechanism, the achievable data rate of group g is limited by the minimum user rate in this group and is defined as follows

$$\min_{k\in\mathcal{K}_g} \{R_k(\mathbf{F}, \mathbf{e})\}. \tag{3.7}$$

3.1.2 Problem Formulation

This chapter aims to jointly optimize the precoding matrix $\mathbf{F}$ and reflection coefficient vector $\mathbf{e}$ to maximize the sum rate of the whole system, which is defined as the sum rate achieved by all groups. Mathematically, the optimization problem is formulated as

$$\begin{aligned} \max_{\mathbf{F},\mathbf{e}} \quad & \left\{ F(\mathbf{F}, \mathbf{e}) = \sum_{g=1}^{G} \min_{k\in\mathcal{K}_g} \{R_k(\mathbf{F}, \mathbf{e})\} \right\} \\ \text{s.t.} \quad & \mathbf{F} \in \mathcal{S}_F, \mathbf{e} \in \mathcal{S}_e. \end{aligned} \tag{3.8}$$

Problem (3.8) is a non-convex problem and difficult to solve since the objective function $F(\mathbf{F}, \mathbf{e})$ is non-differentiable and non-concave, while the unit-modulus constraint set $\mathcal{S}_e$ is also non-convex. In the following, two efficient algorithms are proposed based on the MM algorithm framework to solve Problem (3.8).

3.2 SOCP-based MM method

This section proposes an SOCP-based MM method to solve Problem (3.8). Specifically, under the MM algorithm framework, the non-convex objective function is first handled by introducing its concave surrogate function. Then, the alternating optimization method is adopted to solve the subproblems corresponding to different sets of variables alternately.

Note that $F(\mathbf{F}, \mathbf{e})$ is a composite function which is the linear combinations of some pointwise minimum with non-concave subfunction $R_k(\mathbf{F}, \mathbf{e})$. The non-concave property of $R_k(\mathbf{F}, \mathbf{e})$ is first tackled. To this end, the following lemma is introduced.

Lemma 1. *Let* $\{\mathbf{F}^t, \mathbf{e}^t\}$ *be the solutions obtained at iteration* $n-1$*, then* $R_k(\mathbf{F}, \mathbf{e})$ *is minorized by a concave surrogate function* $\widetilde{R}_k\left(\mathbf{F}, \mathbf{e}|\mathbf{F}^t, \mathbf{e}^t\right)$ *defined by*

$$\begin{aligned}\widetilde{R}_k\left(\mathbf{F}, \mathbf{e}|\mathbf{F}^t, \mathbf{e}^t\right) &= const_k + 2\mathrm{Re}\left\{a_k \mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g\right\} - b_k\sum_{i=1}^{G}|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2 \\ &\leq R_k(\mathbf{F}, \mathbf{e}),\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}a_k &= \frac{(\mathbf{f}_g^t)^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\mathbf{e}^t}{\sum_{i\neq g}^{G}|(\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i^t|^2+\sigma_k^2}, \\ b_k &= \frac{|(\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g^t|^2}{\left(\sum_{i\neq g}^{G}|(\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i^t|^2+\sigma_k^2\right)\left(\sum_{i=1}^{G}|(\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i^t|^2+\sigma_k^2\right)}, \\ const_k &= R_k\left(\mathbf{F}^t, \mathbf{e}^t\right) - b_k\sigma_k^2 - b_k\left(\sum_{i=1}^{G}|(\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i^t|^2+\sigma_k^2\right),\end{aligned}$$

at fixed point $\{\mathbf{F}^t, \mathbf{e}^t\}$.

Proof: Please refer to Appendix A.1. ■

Based on the above lemma, Problem (3.8) can be transformed into the following

surrogate problem:

$$\max_{\mathbf{F},\mathbf{e}} \left\{ \widetilde{F}\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right) = \sum_{g=1}^{G} \min_{k\in\mathcal{K}_g} \left\{ \widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)\right\}\right\}$$
$$\text{s.t. } \mathbf{F}\in\mathcal{S}_F, \mathbf{e}\in\mathcal{S}_e. \tag{3.10}$$

It notes that $\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)$ is biconcave of $\mathbf{F}$ and $\mathbf{e}$ [99] due to the fact that $\widetilde{R}_k\left(\mathbf{F}|\mathbf{F}^t\right) = \widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)$ with given $\mathbf{e}$ is concave of $\mathbf{F}$ and $\widetilde{R}_k\left(\mathbf{e}|\mathbf{e}^t\right) = \widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)$ with given $\mathbf{F}$ is concave of $\mathbf{e}$. This biconvex problem enables us to use the AO method to alternately update $\mathbf{F}$ and $\mathbf{e}$.

3.2.1 Optimizing the Precoding Matrix

This subsection aims to optimize the precoding matrix $\mathbf{F}$ with given $\mathbf{e}$. With some manipulations, $\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)$ in (3.9) can be shown to be a quadratic function of $\mathbf{F}$:

$$\begin{aligned}\widetilde{R}_k\left(\mathbf{F}|\mathbf{F}^t\right) &= \text{const}_k + 2\text{Re}\left\{a_k\mathbf{e}^{\text{H}}\mathbf{H}_k\mathbf{f}_g\right\} - b_k\sum_{i=1}^{G}|\mathbf{e}^{\text{H}}\mathbf{H}_k\mathbf{f}_i|^2 \\ &= \text{const}_k + 2\text{Re}\left\{\text{Tr}\left(\mathbf{C}_k^{\text{H}}\mathbf{F}\right)\right\} - \text{Tr}\left(\mathbf{F}^{\text{H}}\mathbf{B}_k\mathbf{F}\right),\end{aligned} \tag{3.11}$$

where $\mathbf{B}_k = b_k\mathbf{H}_k^{\text{H}}\mathbf{e}\mathbf{e}^{\text{H}}\mathbf{H}_k$, $\mathbf{C}_k^{\text{H}} = a_k\mathbf{t}_g\mathbf{e}^{\text{H}}\mathbf{H}_k$, and $\mathbf{t}_g\in\mathbb{R}^{G\times 1}$ is a selection vector in which the g^{th} element is equal to one and all the other elements are equal to zero.

By using (3.11), the subproblem of Problem (3.10) for the optimization of $\mathbf{F}$ is

$$\max_{\mathbf{F}} \sum_{g=1}^{G}\min_{k\in\mathcal{K}_g}\left\{\text{const}_k + 2\text{Re}\left\{\text{Tr}\left(\mathbf{C}_k^{\text{H}}\mathbf{F}\right)\right\} - \text{Tr}\left(\mathbf{F}^{\text{H}}\mathbf{B}_k\mathbf{F}\right)\right\}$$
$$\text{s.t. } \mathbf{F}\in\mathcal{S}_F. \tag{3.12}$$

The pointwise minimum expressions in the objective function of Problem (3.12) is then

tackled by introducing auxiliary variables $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_G]^{\mathrm{T}}$, as follows

$$\begin{aligned} \max_{\mathbf{F},\boldsymbol{\gamma}} \ & \sum_{g=1}^{G} \gamma_g \\ \text{s.t.} \ & \mathbf{F} \in \mathcal{S}_F, \\ & \text{const}_k + 2\text{Re}\left\{\text{Tr}\left(\mathbf{C}_k^{\mathrm{H}}\mathbf{F}\right)\right\} - \text{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{B}_k\mathbf{F}\right) \geq \gamma_g, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}. \end{aligned} \tag{3.13}$$

Problem (3.13) is an SOCP problem and the globally solution can be obtained by the CVX [62] solver, such as MOSEK [100].

3.2.2 Optimizing the Reflection Coefficient Vector

This subsection focuses on optimizing the reflection coefficient vector $\mathbf{e}$ with given $\mathbf{F}$, then $\widetilde{R}_k\left(\mathbf{e}|\mathbf{e}^t\right)$ can be rewritten as

$$\widetilde{R}_k\left(\mathbf{e}|\mathbf{e}^t\right) = \text{const}_k + 2\text{Re}\left\{\mathbf{a}_k^{\mathrm{H}}\mathbf{e}\right\} - \mathbf{e}^{\mathrm{H}}\mathbf{A}_k\mathbf{e}, \tag{3.14}$$

where $\mathbf{A}_k = b_k\mathbf{H}_k\sum_{i=1}^{G}\mathbf{f}_i\mathbf{f}_i^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}$ and $\mathbf{a}_k = a_k\mathbf{H}_k\mathbf{f}_g$.

Upon replacing the objective function of Problem (3.10) by (3.14), the subproblem for the optimization of $\mathbf{e}$ is given by

$$\begin{aligned} \max_{\mathbf{e}} \ & \sum_{g=1}^{G} \min_{k\in\mathcal{K}_g} \left\{\text{const}_k + 2\text{Re}\left\{\mathbf{a}_k^{\mathrm{H}}\mathbf{e}\right\} - \mathbf{e}^{\mathrm{H}}\mathbf{A}_k\mathbf{e}\right\} \\ \text{s.t.} \ & \mathbf{e} \in \mathcal{S}_e. \end{aligned} \tag{3.15}$$

Also introducing auxiliary variables $\boldsymbol{\kappa} = [\kappa_1, ..., \kappa_G]^{\mathrm{T}}$, Problem (3.15) is equivalent to

$$\begin{aligned} \max_{\mathbf{e},\boldsymbol{\kappa}} \ & \sum_{g=1}^{G} \kappa_g \\ \text{s.t.} \ & \mathbf{e} \in \mathcal{S}_e, \\ & \text{const}_k + 2\text{Re}\left\{\mathbf{a}_k^{\mathrm{H}}\mathbf{e}\right\} - \mathbf{e}^{\mathrm{H}}\mathbf{A}_k\mathbf{e} \geq \kappa_g, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}. \end{aligned} \tag{3.16}$$

The above problem is still non-convex due to the non-convex unit-modulus set $\mathcal{S}_e$. To address this issue, it is replaced with a relaxed convex one as

$$\mathcal{S}_{e-relax} = \{\mathbf{e}^{\mathrm{H}}\mathrm{diag}(\mathbf{i}_m)\mathbf{e} \leq 1, \ \ \forall m = 1, \cdots, M, e_{M+1} = 1\},$$

where $\mathbf{i}_m \in \mathbb{R}^{(M+1)\times 1}$ is a selection vector whose m^{th} element is equal to one and all the other elements are equal to zero. Let us denote by $\widehat{\mathbf{e}}_1$ the optimal solution of the following relaxed version of the SOCP problem, i.e.,

$$\begin{aligned} \widehat{\mathbf{e}}_1 = \arg\max_{\mathbf{e}} \ & \sum_{g=1}^{G} \gamma_g \\ \text{s.t.} \ \ & \mathbf{e} \in \mathcal{S}_{e-relax}, \\ & \mathrm{const}_k + 2\mathrm{Re}\left\{\mathbf{a}_k^{\mathrm{H}}\mathbf{e}\right\} - \mathbf{e}^{\mathrm{H}}\mathbf{A}_k\mathbf{e} \geq \kappa_g, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}. \end{aligned} \tag{3.17}$$

Then, the locally optimal solution $\mathbf{e}$ in the n^{th} iteration is

$$\mathbf{e}^{t+1} = \begin{cases} \widehat{\mathbf{e}}_2, & \text{if } F\left(\mathbf{F}^{t+1}, \widehat{\mathbf{e}}_2 | \mathbf{F}^t, \mathbf{e}^t\right) \geq F\left(\mathbf{F}^{t+1}, \mathbf{e}^t | \mathbf{F}^t, \mathbf{e}^t\right), \\ \mathbf{e}^t, & \text{otherwise}, \end{cases} \tag{3.18}$$

where

$$\widehat{\mathbf{e}}_2 = \exp\left\{j\angle\left(\frac{\widehat{\mathbf{e}}_1}{[\widehat{\mathbf{e}}_1]_{M+1}}\right)\right\}, \tag{3.19}$$

and symbol $[\widehat{\mathbf{e}}_1]_m$ denotes the m^{th} element of the vector $\widehat{\mathbf{e}}_1$. Here the $\exp\{\cdot\}$ and the $\angle(\cdot)$ are both element-wise operations.

3.2.3 Algorithm Development

Based on the above analysis, Algorithm 3.1 summarizes the alternating update process between precoding matrix $\mathbf{F}$ and reflection coefficient vector $\mathbf{e}$ to maximize the sum rate of the whole system.

Algorithm 3.1: SOCP-based MM algorithm

Require: Initialize $\mathbf{F}^0$ and $\mathbf{e}^0$, and $t = 0$.

1: **repeat**
2: Calculate $\mathbf{F}^{t+1}$ by solving Problem (3.13) with given $\mathbf{e}^t$;
3: Calculate $\mathbf{e}^{t+1}$ by solving Problem (3.17) with given $\mathbf{F}^{t+1}$;
4: $t \leftarrow t + 1$;
5: **until** The value of function $F(\mathbf{F}, \mathbf{e})$ in (3.8) converges.

3.2.3.1 Complexity Analysis

This subsection analyzes the computational complexity of Algorithm 3.1, which mainly comes from optimizing $\mathbf{F}$ in the SOCP problem in (3.13) and optimizing $\mathbf{e}$ in the SOCP problem in (3.17).

According to [101], the complexity of solving an SOCP problem, with M_{socp} second order cone constraints where the dimension of each is N_{socp}, is $\mathcal{O}(N_{\text{socp}}M_{\text{socp}}^{3.5} + N_{\text{socp}}^3 M_{\text{socp}}^{2.5})$. Problem (3.13) contains one power constraint with dimension NG and K rate constraints with dimension NG. Therefore, the complexity of solving Problem (3.13) per iteration is $\mathcal{O}(NG + N^3G^3 + NGK^{3.5} + N^3G^3K^{2.5})$. Problem (3.17) has M constant modulus constraints with dimension one for sparse vector $\mathbf{i}_m$ and K rate constraints with dimension $M + 1$. Therefore, the complexity of solving Problem (3.17) per iteration is $\mathcal{O}(M^{3.5} + M^{2.5} + (M+1)K^{3.5} + (M+1)^3K^{2.5})$. Therefore, the approximate complexity of Algorithm 3.1 per iteration is $\mathcal{O}(N^3G^3K^{2.5} + M^{3.5} + MK^{3.5})$.

3.2.3.2 Convergence Analysis

The following theorem shows the convergence and solution properties of Algorithm 3.1.

Theorem 1. *The objective function value sequence* $\{F\left(\mathbf{F}^t, \mathbf{e}^t\right)\}$ *generated by Algorithm 3.1 is guaranteed to converge, and the optimal solution converges to a Karush-Kuhn-Tucker (KKT) point.*

Proof: Please refer to Appendix A.2. ■

3.3 Low-Complexity MM Method

As seen in Algorithm 3.1, two SOCP problems need to be solved in each iteration, which incurs a high computational complexity. This section aims to derive a low-complexity algorithm containing closed-form solutions.

Since $\min_{k\in\mathcal{K}_g}\left\{\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)\right\}$ in Problem (3.10) is non-differentiable, it is approximated as a smooth function by using the smooth log–sum–exp lower-bound [102]

$$\begin{aligned}\min_{k\in\mathcal{K}_g}\left\{\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)\right\} &\approx f_g\left(\mathbf{F},\mathbf{e}\right)\\ &= -\frac{1}{\mu_g}\log\Big(\sum_{k\in\mathcal{K}_g}\exp\left\{-\mu_g\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)\right\}\Big),\end{aligned} \tag{3.20}$$

where $\mu_g>0$ is a smoothing parameter which satisfies

$$\begin{aligned}f_g\left(\mathbf{F},\mathbf{e}\right) &\le \min_{k\in\mathcal{K}_g}\left\{\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)\right\}\\ &\le f_g\left(\mathbf{F},\mathbf{e}\right)+\frac{1}{\mu_g}\log\left(|\mathcal{K}_g|\right).\end{aligned} \tag{3.21}$$

Theorem 2. *$f_g\left(\mathbf{F},\mathbf{e}\right)$ is biconcave of $\mathbf{F}$ and $\mathbf{e}$.*

Proof: According to [79], if the Hessian matrix of a function is semi-negative definite, that function is concave. In particular, the Hessian matrix of the exp-sum-log function $f\left(x\right)=-\log\Big(\sum_{k\in\mathcal{K}_g}\exp\left\{-x\right\}\Big)$ is derived as

$$\nabla^2 f(x)=-\frac{1}{\left(\mathbf{1}\mathbf{z}^{\mathrm{T}}\right)^2}\left(\left(\mathbf{1}^{\mathrm{T}}\mathbf{z}\right)\mathrm{diag}(\mathbf{z})-\mathbf{z}\mathbf{z}^{\mathrm{T}}\right), \tag{3.22}$$

where $\mathbf{z}=(e^{x_1},\ldots,e^{x_N})$. Then for all $\mathbf{v}$, it has

$$\begin{aligned}\mathbf{v}^{\mathrm{T}}\nabla^2 f(x)\mathbf{v} &= -\frac{1}{\left(\mathbf{1}\mathbf{z}^{\mathrm{T}}\right)^2}\left(\left(\sum_{n=1}^{N}z_n\right)\left(\sum_{n=1}^{N}v_n^2 z_n\right)-\left(\sum_{n=1}^{N}v_n z_n\right)^2\right)\\ &= -\left(\mathbf{b}^{\mathrm{T}}\mathbf{b}\mathbf{a}^{\mathrm{T}}\mathbf{a}-\left(\mathbf{a}^{\mathrm{T}}\mathbf{b}\right)^2\right)\le 0,\end{aligned} \tag{3.23}$$

where the components of vectors $\mathbf{a}$ and $\mathbf{b}$ are $a_n = v_n\sqrt{z_n}$ and $b_n = \sqrt{z_n}$, respectively. The inequality follows from the Cauchy-Schwarz inequality. Then $\nabla^2 f(x) \preceq 0$, and the log-sum-exp function $f(x)$ is concave. Therefore, $-\frac{1}{\mu_g}\log\left(\sum_{k\in\mathcal{K}_g}\exp\left\{-\mu_g\widetilde{R}_k\right\}\right)$ is an increasing and concave function w.r.t. $\widetilde{R}_k$. Recall that $\widetilde{R}_k\left(\mathbf{F},\mathbf{e}|\mathbf{F}^t,\mathbf{e}^t\right)$ is biconcave of $\mathbf{F}$ and $\mathbf{e}$. Finally, according to the composition principle [79], $f_g(\mathbf{F},\mathbf{e})$ is biconcave of $\mathbf{F}$ and $\mathbf{e}$. The proof is complete. ■

Large μ_g leads to high accuracy of the approximation, but it also causes the problem to be nearly ill-conditioned. When μ_g is chosen appropriately, Problem (3.10) is approximated as

$$\begin{aligned}\max_{\mathbf{F},\mathbf{e}}\ & \sum_{g=1}^{G} f_g(\mathbf{F},\mathbf{e})\\ \text{s.t.}\ & \mathbf{F}\in\mathcal{S}_F, \mathbf{e}\in\mathcal{S}_e.\end{aligned} \tag{3.24}$$

This problem is still a biconvex problem of $\mathbf{F}$ and $\mathbf{e}$, which enables us to alternately update $\mathbf{F}$ and $\mathbf{e}$ by adopting the alternating optimization method.

3.3.1 Optimizing the Precoding Matrix

Given $\mathbf{e}$, the subproblem of Problem (3.24) for the optimization of $\mathbf{F}$ is

$$\begin{aligned}\max_{\mathbf{F}}\ & \sum_{g=1}^{G} f_g(\mathbf{F})\\ \text{s.t.}\ & \mathbf{F}\in\mathcal{S}_F.\end{aligned} \tag{3.25}$$

Even $f_g(\mathbf{F})$ is a concave and continuous function of precoding matrix $\mathbf{F}$, it is still very complex and difficult to be optimized directly. In this subsection, the surrogate function of $f_g(\mathbf{F})$ in the MM algorithm framework is given in the following theorem.

Theorem 3. *Since $f_g(\mathbf{F})$ is twice differentiable and concave, $f_g(\mathbf{F})$ is minorized at any*

fixed $\mathbf{F}^t$ *with a quadratic function* $\widetilde{f}_g(\mathbf{F}|\mathbf{F}^t)$ *satisfying Assumptions (A1)-(A4), as follows*

$$\widetilde{f}_g(\mathbf{F}|\mathbf{F}^t) = 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{U}_g^{\mathrm{H}}\mathbf{F}\right)\right\} + \alpha_g \mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right) + consF_g, \tag{3.26}$$

where

$$\mathbf{U}_g = \sum_{k\in\mathcal{K}_g} g_k(\mathbf{F}^t)(\mathbf{C}_k - \mathbf{B}_k^{\mathrm{H}}\mathbf{F}^t) - \alpha_g \mathbf{F}^t, \tag{3.27}$$

$$g_k(\mathbf{F}^t) = \frac{\exp\left\{-\mu_g \widetilde{R}_k\left(\mathbf{F}^t\right)\right\}}{\sum_{k\in\mathcal{K}_g}\exp\left\{-\mu_g \widetilde{R}_k\left(\mathbf{F}^t\right)\right\}}, k \in \mathcal{K}_g, \tag{3.28}$$

$$\alpha_g = -\max_{k\in\mathcal{K}_g}\left\{b_k \mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{e}\right\} - 2\mu_g \max_{k\in\mathcal{K}_g}\left\{tp_k\right\}, \tag{3.29}$$

$$tp_k = P_{\mathrm{T}} b_k^2 |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{e}|^2 + ||\mathbf{C}_k||_F^2 + 2\sqrt{P_{\mathrm{T}}}||\mathbf{B}_k\mathbf{C}_k||_F, \tag{3.30}$$

$$consF_g = f_g(\mathbf{F}^t) + \alpha_g \mathrm{Tr}\left((\mathbf{F}^t)^{\mathrm{H}}\mathbf{F}^t\right) - 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{D}_g^{\mathrm{H}}\mathbf{F}^t\right)\right\}. \tag{3.31}$$

Proof: Please refer to Appendix A.3. ■

Upon replacing the objective function of Problem (3.25) with (3.26), the following surrogate problem is obtained

$$\begin{aligned} \max_{\mathbf{F}} \quad & \sum_{g=1}^{G}\left(2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{U}_g^{\mathrm{H}}\mathbf{F}\right)\right\} + \alpha_g \mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right) + \mathrm{consF}_g\right) \\ \text{s.t.} \quad & \mathbf{F} \in \mathcal{S}_F. \end{aligned} \tag{3.32}$$

The optimal $\mathbf{F}^{t+1}$ could be obtained by introducing a Lagrange multiplier $\tau \geq 0$ associated with the power constraint, yielding the Lagrange function

$$\mathcal{L}(\mathbf{F},\tau) = 2\mathrm{Re}\left\{\mathrm{Tr}\left(\sum_{g=1}^{G}\mathbf{U}_g^{\mathrm{H}}\mathbf{F}\right)\right\} + \sum_{g=1}^{G}\alpha_g \mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right) + \sum_{g=1}^{G}\mathrm{consF}_g - \tau\left(\mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right) - P_{\mathrm{T}}\right). \tag{3.33}$$

By setting the first-order derivative of $\mathcal{L}(\mathbf{F},\tau)$ w.r.t. $\mathbf{F}^*$ to zero, it has

$$\frac{\partial \mathcal{L}(\mathbf{F})}{\partial \mathbf{F}^*} = \mathbf{0}.$$

Then the globally optimal solution of $\mathbf{F}$ in iteration t can be derived as

$$\mathbf{F}^{t+1} = \frac{1}{\tau - \sum_{g=1}^{G} \alpha_g} \sum_{g=1}^{G} \mathbf{U}_g. \tag{3.34}$$

By substituting (3.34) into the power constraint, one has

$$\frac{\mathrm{Tr}\left(\left(\sum_{g=1}^{G} \mathbf{U}_g\right)^{\mathrm{H}} \left(\sum_{g=1}^{G} \mathbf{U}_g\right)\right)}{(\tau - \sum_{g=1}^{G} \alpha_g)^2} \leq P_{\mathrm{T}}. \tag{3.35}$$

It is obvious that the left hand side of (3.35) is a decreasing function of τ.

- If the power constraint inequality (3.35) holds when $\tau = 0$, then

$$\mathbf{F}^{t+1} = \frac{-1}{\sum_{g=1}^{G} \alpha_g} \sum_{g=1}^{G} \mathbf{U}_g. \tag{3.36}$$

- Otherwise, there must exist a $\tau > 0$ that (3.35) holds with equality, then

$$\mathbf{F}^{t+1} = \sqrt{\frac{P_{\mathrm{T}}}{\mathrm{Tr}\left(\left(\sum_{g=1}^{G} \mathbf{U}_g\right)^{\mathrm{H}} \left(\sum_{g=1}^{G} \mathbf{U}_g\right)\right)}} \sum_{g=1}^{G} \mathbf{U}_g. \tag{3.37}$$

3.3.2 Optimizing the Reflection Coefficient Vector

Given $\mathbf{F}$, the subproblem of Problem (3.24) for the optimization of $\mathbf{e}$ is

$$\begin{aligned} \max_{\mathbf{e}} \ & \sum_{g=1}^{G} f_g(\mathbf{e}) \\ \text{s.t.} \ & \mathbf{e} \in \mathcal{S}_e. \end{aligned} \tag{3.38}$$

Upon adopting the MM algorithm framework, it firstly needs to find a minorizing function of $f_g(\mathbf{e})$ and denotes it as $\widehat{f}_g(\mathbf{e}|\mathbf{e}^t)$. Since $\mathcal{S}_e$ is a non-convex set, Assumption (A3) should be modified so as to claim stationarity convergence [103], [104]:

$$\widehat{f}_g^{'}(\mathbf{e}|\mathbf{e}^t;\mathbf{d})|_{\mathbf{e}=\mathbf{e}^t} = f_g^{'}(\mathbf{e}^t;\mathbf{d}), \forall \mathbf{d} \in \mathcal{T}_{\mathcal{S}_e}(\mathbf{e}^t),$$

where $\mathcal{T}_{\mathcal{S}_e}(\mathbf{e}^t)$ is the Boulingand tangent cone of $\mathcal{S}_e$ at $\mathbf{e}^t$. Therefore $\widehat{f}_g(\mathbf{e}|\mathbf{e}^t)$ is given in the following theorem.

Theorem 4. *Since $f_g(\mathbf{e})$ is twice differentiable and concave, $f_g(\mathbf{e})$ is minorized at any fixed $\mathbf{e}^t$ with a function $\widehat{f}_g(\mathbf{e}|\mathbf{e}^t)$ satisfying Assumptions (A1)-(A4), as follows*

$$\widetilde{f}_g(\mathbf{e}|\mathbf{e}^t) = 2\mathrm{Re}\left\{\mathbf{u}_g^{\mathrm{H}}\mathbf{e}\right\} + consE_g, \tag{3.39}$$

where

$$\mathbf{u}_g = \sum_{k\in\mathcal{K}_g} g_k(\mathbf{e}^t)(\mathbf{a}_k - \mathbf{A}_k^{\mathrm{H}}\mathbf{e}^t) - \beta_g\mathbf{e}^t, \tag{3.40}$$

$$g_k(\mathbf{e}^t) = \frac{\exp\left\{-\mu_g\widetilde{R}_k\left(\mathbf{e}^t\right)\right\}}{\sum_{k\in\mathcal{K}_g}\exp\left\{-\mu_g\widetilde{R}_k\left(\mathbf{e}^t\right)\right\}}, k \in \mathcal{K}_g, \tag{3.41}$$

$$\beta_g = -\mathrm{max}_{k\in\mathcal{K}_g}\left\{\lambda_{\max}(\mathbf{A}_k)\right\} - 2\mu_g\mathrm{max}_{k\in\mathcal{K}_g}\left\{tp2_k\right\}, \tag{3.42}$$

$$tp2_k = ||\mathbf{a}_k||_2^2 + (M+1)\lambda_{\max}(\mathbf{A}_k\mathbf{A}_k^{\mathrm{H}}) + 2||\mathbf{A}_k\mathbf{a}_k||_1, \tag{3.43}$$

$$consE_g = f_g(\mathbf{e}^t) + 2(M+1)\beta_g - 2\mathrm{Re}\left\{\mathbf{d}_g^{\mathrm{H}}\mathbf{e}^t\right\}. \tag{3.44}$$

Proof: Please refer to Appendix A.4. ■

Upon replacing the objective function of Problem (3.38) by (3.39), it obtains the following surrogate problem as

$$\max_{\mathbf{e}} \sum_{g=1}^{G}\left(2\mathrm{Re}\left\{\mathbf{u}_g^{\mathrm{H}}\mathbf{e}\right\} + \mathrm{consE}_g\right)$$

$$\text{s.t.} \quad \mathbf{e} \in \mathcal{S}_e. \tag{3.45}$$

Then, the globally optimal solution of $\mathbf{e}$ at the n^{th} iteration is

$$\mathbf{e}^{t+1} = \exp\left\{ j\angle\left(\left(\sum_{g=1}^{G} \mathbf{u}_g \right) / \left[\sum_{g=1}^{G} \mathbf{u}_g \right]_{M+1} \right) \right\}, \tag{3.46}$$

where $\exp\{j\angle(\cdot)\}$ is an element-wise operation.

3.3.3 Low-Complexity Algorithm Design

This section adopts alternating optimization algorithm to alternately optimize precoding matrix $\mathbf{F}$ and reflection coefficient vector $\mathbf{e}$. Note that the tightness of the lower bounds α_g in (3.29) and β_g in (3.42) affects the performance of the convergence speed. Here, SQUAREM [105] is adopted to accelerate the convergence speed of the proposed algorithm, which is summarized in Algorithm 3.2.

Let $\mathcal{M}_F(\cdot)$ denote the nonlinear fixed-point iteration map of the MM algorithm of $\mathbf{F}$ in (3.34), i.e., $\mathbf{F}^{t+1} = \mathcal{M}_F(\mathbf{F}^t)$, and $\mathcal{M}_e(\cdot)$ of $\mathbf{e}$ in (3.46), i.e., $\mathbf{e}^{t+1} = \mathcal{M}_e(\mathbf{e}^t)$. $\mathcal{P}_{\mathcal{S}}(\cdot)$ is project operation to force wayward points to satisfy their nonlinear constraints. For the power constraint in Problem (3.32), the projection can be done by using the function $\frac{(\cdot)}{||\cdot||_F}||\mathbf{F}_2||_F$ to the solution matrix, e.g., $\mathcal{P}_{\mathcal{S}}(\mathbf{X}) = \frac{(\mathbf{X})}{||\mathbf{X}||_F}||\mathbf{F}_2||_F$. For the unit-modulus constraints in Problem (3.45), it can be obtained by using function $\exp\{j\angle(\cdot)\}$ element-wise to the solution vector. Steps 10 to 13 and steps 21 to 24 are to maintain the ascent property of the proposed algorithm.

3.3.4 Complexity Analysis

The computational complexity of Algorithm 3.2 is composed of the nonlinear fixed-point iteration maps $\mathcal{M}_F(\cdot)$ and $\mathcal{M}_e(\cdot)$. In $\mathcal{M}_F(\cdot)$, the computational complexity of $\mathbf{U}_g$ in (3.30) mainly comes from $g_k(\mathbf{F}^t)$ in (3.28) and α_g in (3.29). Firstly, the computational complexity of $g_k(\mathbf{F}^t)$ is of order $\mathcal{O}(|\mathcal{K}_g|(2MNG + 3NG))$ since there are $|\mathcal{K}_g|$ $\widetilde{R}_k\left(\mathbf{F}^t\right)$

Algorithm 3.2: Low-complexity MM algorithm

Require: Initialize $\mathbf{F}^0$ and $\mathbf{e}^0$, and $t = 0$.

repeat
  Set $\mathbf{e} = \mathbf{e}^t$;
  $\mathbf{F}_1 = \mathcal{M}_F(\mathbf{F}^t)$;
  $\mathbf{F}_2 = \mathcal{M}_F(\mathbf{F}_1)$;
  $\mathbf{J}_1 = \mathbf{F}_1 - \mathbf{F}^t$;
  $\mathbf{J}_2 = \mathbf{F}_2 - \mathbf{F}_1 - \mathbf{J}_1$;
  $\omega = -\frac{||\mathbf{J}_1||_F}{||\mathbf{J}_2||_F}$;
  $\mathbf{F}^{t+1} = -\mathcal{P}_{\mathcal{S}}(\mathbf{F}^t - 2\omega\mathbf{J}_1 + \omega^2\mathbf{J}_2)$;
  while $F\left(\mathbf{F}^{t+1}\right) < F\left(\mathbf{F}^t\right)$ **do**
    $\omega = (\omega - 1)/2$;
    $\mathbf{F}^{t+1} = -\mathcal{P}_{\mathcal{S}}(\mathbf{F}^t - 2\omega\mathbf{J}_1 + \omega^2\mathbf{J}_2)$;
  end while
  Set $\mathbf{F} = \mathbf{F}^{t+1}$;
  $\mathbf{e}_1 = \mathcal{M}_e(\mathbf{e}^t)$;
  $\mathbf{e}_2 = \mathcal{M}_e(\mathbf{e}_1)$;
  $\mathbf{j}_1 = \mathbf{e}_1 - \mathbf{e}^t$;
  $\mathbf{j}_2 = \mathbf{e}_2 - \mathbf{e}_1 - \mathbf{j}_1$;
  $\omega = -\frac{||\mathbf{j}_1||_F}{||\mathbf{j}_2||_F}$;
  $\mathbf{e}^{t+1} = -\mathcal{P}_{\mathcal{S}}(\mathbf{e}^t - 2\omega\mathbf{j}_1 + \omega^2\mathbf{j}_2)$;
  while $F\left(\mathbf{e}^{t+1}\right) < F\left(\mathbf{e}^t\right)$ **do**
    $\omega = (\omega - 1)/2$;
    $\mathbf{e}^{t+1} = -\mathcal{P}_{\mathcal{S}}(\mathbf{e}^t - 2\omega\mathbf{j}_1 + \omega^2\mathbf{j}_2)$;
  end while
  $n \leftarrow n + 1$;
until The value of function $F(\mathbf{F}, \mathbf{e})$ in (3.8) converges.

in (3.9) of order $\mathcal{O}(2MNG + 3NG)$. Then each tp_k in (3.30) is of complexity $\mathcal{O}(4N^3 + 2N^2K - NK + 4MN)$ neglecting the lower-order terms, thus α_g is of order $\mathcal{O}(|\mathcal{K}_g|(4N^3 + 2N^2K + 4MN))$. Therefore, the approximate complexity of $\mathcal{M}_F(\cdot)$ is $\mathcal{O}(4N^3K + 2N^2K^2 + 2MNGK)$ neglecting the lower-order terms. In $\mathcal{M}_e(\cdot)$, the computational complexity of $g_k(\mathbf{e}^t)$ in (3.41) is the same as $g_k(\mathbf{F}^t)$, which is of complexity $\mathcal{O}(|\mathcal{K}_g|(2MNG + 3NG))$. Furthermore, the eigenvalue operations $\lambda_{\max}(\mathbf{A}_k)$ and $\lambda_{\max}(\mathbf{A}_k\mathbf{A}_k^{\mathrm{H}})$ of order $\mathcal{O}((M + 1)^3)$ contribute to the main complexity of calculating β_g in (3.42), which is of order $\mathcal{O}(|\mathcal{K}_g|(M+1)^3)$. Neglecting the lower-order terms, the approximate complexity of $\mathcal{M}_e(\cdot)$ is $\mathcal{O}(2MNGK + K(M+1)^3)$. Eventually, the approximate complexity of Algorithm 3.2 per iteration is $\mathcal{O}(4N^3K + 2N^2K^2 + 3MNGK + K(M+1)^3)$, neglecting the lower-order

terms.

The computational complexity of the proposed two algorithms are summarized and compared in Table 3-A. Comparing with Algorithm 3.1 based on SOCP, Algorithm 3.2 has a lower computational complexity and requires less CPU time, which will be shown in the following section.

3.3.5 Convergence Analysis

In each iteration, the MM algorithm is adopted to update each set of variables. The monotonicity of the MM algorithm has been proved in [93] and [106]. In the following, it claims the monotonicity of Algorithm 3.2. At the n^{th} iteration, with given $\mathbf{e}^t$, it has

$$f_g(\mathbf{F}^t, \mathbf{e}^t) = \widetilde{f}_g(\mathbf{F}^t, \mathbf{F}^t) \leq \widetilde{f}_g(\mathbf{F}^{t+1}, \mathbf{F}^t) \leq f_g(\mathbf{F}^{t+1}, \mathbf{e}^t),$$

where the first equality follows from (A1), the first inequality follows from (3.32), and the second one follows from (A2). Subsequently, with given $\mathbf{F}^{t+1}$, it is straightforward to have

$$f_g(\mathbf{F}^{t+1}, \mathbf{e}^t) = \widehat{f}_g(\mathbf{e}^t, \mathbf{e}^t) \leq \widehat{f}_g(\mathbf{e}^{t+1}, \mathbf{e}^t) \leq f_g(\mathbf{F}^{t+1}, \mathbf{e}^{t+1}).$$

Therefore, the objective function values $\{f_g(\mathbf{F}^{t+1}, \mathbf{e}^{t+1})\}$ generated during the procedure of the AO algorithm are monotonically increasing.

Let $\{\mathbf{F}^t\}$ be the sequence generated by the proposed algorithm. Since $\mathcal{S}_F$ is a convex set, every limit point of $\{\mathbf{F}^t\}$ is a d-stationary point of Problem (3.8), and the limit point $\mathbf{F}^\infty$ satisfies

$$f_g^{'}(\mathbf{F}^\infty; \mathbf{d}) \leq 0, \forall \mathbf{d} \text{ with } \mathbf{F}^\infty + \mathbf{d} \in \mathcal{S}_F.$$

The proof of converging to a d-stationary point can be found in [107].

Let $\{\mathbf{e}^t\}$ be the sequence generated by the proposed algorithm. Since $\mathcal{S}_e$ is a non-convex set, every limit point of $\{\mathbf{e}^t\}$ is a B-stationary point of Problem (3.8), and the

Table 3-A: Complexity analysis of the proposed MM algorithms

Algorithm	**Complexity**
SOCP-based MM algorithm	$\mathcal{O}(4N^3K + 2N^2K^2 + 3MNGK + K(M+1)^3)$
Low-complexity MM algorithm	$\mathcal{O}(N^3K^3 + NK^{4.5} + N^3K^{5.5} + MK^{3.5} + M^3K^{2.5})$

limit point $\mathbf{e}^\infty$ satisfies

$$f_g^{'}(\mathbf{e}^\infty; \mathbf{d}) \leq 0, \forall \mathbf{d} \in \mathcal{T}_{\mathcal{S}_e}(\mathbf{e}^\infty).$$

The proof of converging to a B-stationary point can be found in [103] and [104].

The property of the converged solution of Algorithm 3.2 is shown in the following Theorem.

Theorem 5. *The optimal solution converges to a KKT point of Problem (3.24).*

Proof: Please refer to Appendix A.5. ■

3.4 Simulation Results and Discussions

3.4.1 Simulation Setup

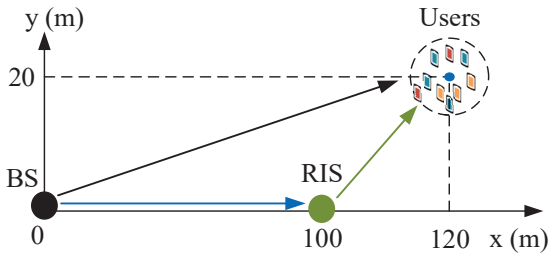


Figure 3.2: The simulated system setup.

In this section, extensive simulation results are provided to evaluate the performance of the proposed algorithms for an RIS-aided multigroup multicast MISO communication system. All experiments are performed on a PC with a 1.99 GHz i7-8550U CPU and 16 GB RAM. Each point in the following figures is obtained by averaging over 100 independent trials. The simulated model in Fig. 5.2 [2] is as follows: The BS locating at

[2] 2D scenario can be extended to 3D scenario.

(0 m, 0 m) employs a uniform linear array (ULA) with N antennas and the RIS locating at (100 m, 0 m) is equipped with a uniform planar array (UPA) with M reflecting elements, where the width of the UPA is fixed at 4 and the length is $M/4$. All users are randomly distributed in a circle centered at (120 m, 20 m) with radius 10 m.

The large-scale path loss is PL $= -30 - 10\alpha \log_{10}(d)$ dB, in which d is the link length in meters and the path loss exponents for the BS-RIS link, the RIS-user link, and the BS-user link are set as $\alpha_{\mathrm{BI}} = \alpha_{\mathrm{IU}} = 2$ and $\alpha_{\mathrm{BU}} = 4$, respectively [108]. The small-scale fading in $[\mathbf{H}_{\mathrm{dr}}, \{\mathbf{h}_{\mathrm{d},k}\}_{\forall k \in \mathcal{K}}]$ is assumed to follow Rayleigh distribution with zero-mean and unit variance due to the fact of the large lengths of the BS-RIS link and the BS-user link, while the small-scale fading in $\{\mathbf{h}_{\mathrm{r},k}\}_{\forall k \in \mathcal{K}}$ is assumed to be Rican fading with Ricean factor $\kappa_{\mathrm{IU}} = 10$. The Line-of-Sight (LoS) components are modeled as the product of the steering vectors of the transceivers and the non-LoS components are drawn from a Rayleigh distribution. Unless otherwise stated, the other parameters are set as: Transmission bandwidth of 10 MHz, noise power density of -174 dBm/Hz, convergence accuracy of $\epsilon = 10^{-6}$, smoothing parameter of $\mu_g = 100$ [102], $N = 4$, $N = 16$, $G = |\mathcal{K}_g| = 2$.

RIS-Alg. 1 is used to represent Algorithm 3.1 and **RIS-Alg. 2** to represent Algorithm 3.2. For comparison purposes, the performance of the scheme without RIS is showed, in which the precoding matrix is also obtained by the proposed two algorithms, denoted as **NRIS-Alg. 1** and **NRIS-Alg. 2**, respectively.

3.4.2 Baseline Schemes

Due to the hardware limitation, it is practically difficult to realize the continuous phase shifts at each reflection element considered in this work. Hence, two baseline schemes with 2 bit resolution are considered in the simulations to investigate the performance loss of using finite resolution reflection elements. Specifically, with optimal $\mathbf{e}^o$ generated

by Algorithm 3.1 or Algorithm 3.2, the m^{th} discrete phase shift can be obtained by

$$\theta_m^o = \arg \min_{\theta \in \mathcal{F}_\theta} |\exp\{j\angle\theta\} - e_m^o|,$$

where $\mathcal{F}_\theta = \{0, 2\pi/B, ..., 2\pi(B-1)/B\}$ and $B = 2^2$. Therefore, the two baseline schemes are called as **RIS-Alg. 1, 2 bit** and **RIS-Alg. 2, 2 bit.**

Besides, RIS is advocated as an energy-efficient device for assisting wireless communication. Hence, it is necessary to compare the performance of the RIS-based and the full-duplex Amplify-and-Forward (AF) relay-based multigroup multicast systems. To ensure a fair comparison with the proposed RIS-aided system, the **Relay** benchmark scheme, in which the relay is located at the same place of the RIS, has considered the same users' locations and channel realizations. Then, the sum rate maximization problem for the joint design of the precoder $\mathbf{F}$ and the relay beamforming $\mathbf{W}$ is given by

$$\begin{aligned} \max_{\mathbf{F},\mathbf{W}} \ & \sum_{g=1}^{G} \min_{k \in \mathcal{K}_g} R_k^{relay} \\ \text{s.t.} \ & ||\mathbf{F}||_F^2 \leq P_{\mathrm{T}} \\ & ||\mathbf{W}\mathbf{H}_{\mathrm{dr}}\mathbf{F}||_F^2 + ||\mathbf{W}||_F^2 \sigma_r^2 \leq P_{\mathrm{relay}}, \end{aligned} \tag{3.47}$$

where R_k^{relay} is given by

$$\log_2 \left(1 + \frac{|(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \mathbf{W}\mathbf{H}_{\mathrm{dr}})\mathbf{f}_g|^2}{\sum_{i \neq g}^{G} |(\mathbf{h}_{\mathrm{d},k}^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \mathbf{W}\mathbf{H}_{\mathrm{dr}})\mathbf{f}_i|^2 + ||\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \mathbf{W}||_2^2 \sigma_r^2 + \sigma_k^2} \right).$$

Here, P_{relay} is the maximum available transmit power at the relay, σ_r^2 is the noise power received by the relay, and the digital relay beamforming $\mathbf{W}$ is assumed to be a diagonal matrix.

The AO method is adopted to solve the above problem. Basically, the SCA method in [109] is extended to alternately update each variable in Problem (3.47).

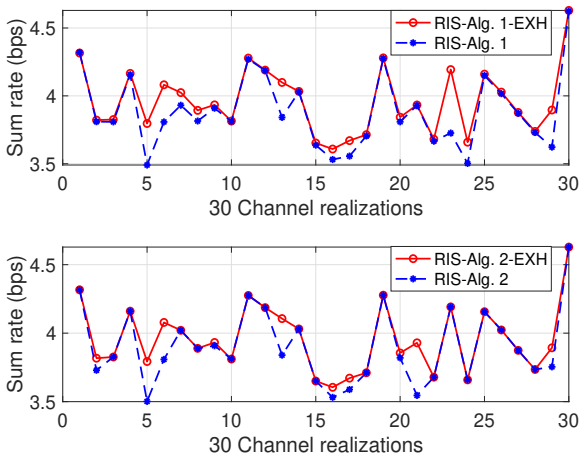
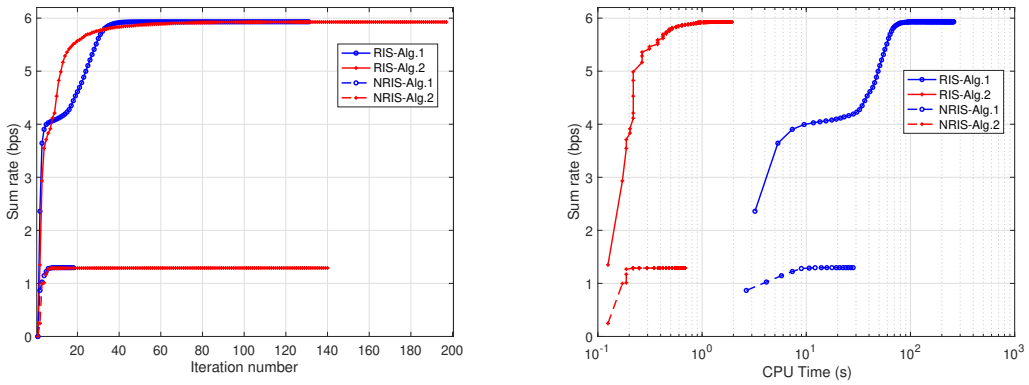


Figure 3.3: The performance comparison of different initialization, when $N = 4$, $M = 16$, $G = |\mathcal{K}_g| = 2$ and $P_{\mathrm{T}} = 15$ dBm.

3.4.3 Convergence of the Proposed Algorithms

Consider the fact of the nonconvexity of Problem (3.8), different initial points may result in different locally optimal solutions obtained by the the proposed algorithms. By testing 30 randomly channel realizations, Fig. 3.3 illustrates the impact of the initialization on the performance of the proposed algorithms. The initialization of RIS-Alg. 1 and RIS-Alg. 2 are: $\mathbf{F}$ is initialized by uniformly allocating maximum transmit power, $\mathbf{e}$ is initialized by setting each entry to 1. RIS-Alg. 1-EXH (RIS-Alg. 2-EXH) refers to the best initial point of 1000 random initial points for each channel realization. It can be seen that the sum rate of RIS-Alg. 1 (RIS-Alg. 2) is almost the same as that of RIS-Alg. 1-EXH (RIS-Alg. 2-EXH), implying that the simple uniform power allocation of $\mathbf{F}$ and all-one $\mathbf{e}$ is a good option for the initialization.

Fig. 3.4 investigates the convergence behaviour of various algorithms in terms of the iteration number and the CPU time when $P_{\mathrm{T}} = 20$ dBm. Fig. 3.4(a) compares convergence speed in terms of the number of iterations. Only a small number of iterations are sufficient for Algorithm 3.1 to converge for both RIS and NRIS schemes. The reason is that the lower bound of the original objective function in (3.9) used in Algorithm 3.1 is tighter than those in (3.26) and (3.39) used in Algorithm 3.2. Although Algorithm 3.2



(a) The sum rate versus iteration number

(b) The sum rate versus CPU time

Figure 3.4: The convergence behaviour of different algorithms, when $N = 4$, $M = 16$, $G = |\mathcal{K}_g| = 2$ and $P_{\mathrm{T}} = 20$ dBm.

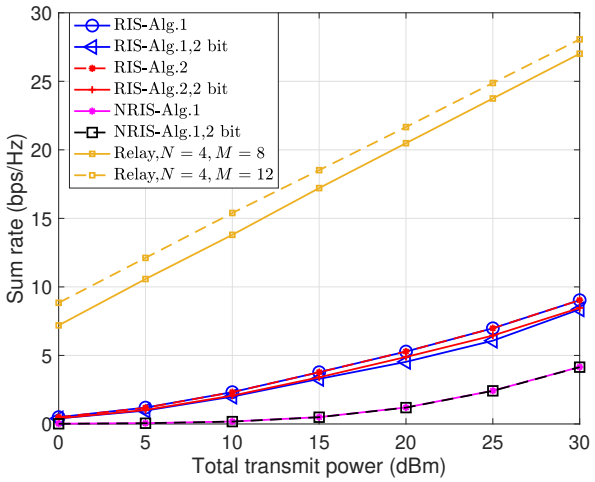
needs more iterations to converge, it has a fast convergence speed in terms of CPU time shown in Fig. 3.4(b). This is because in each iteration of Algorithm 3.2, there always exists closed-form solutions when designing precoding matrix and reflection coefficient vector. In addition, the optimal objective function values generated by both algorithms for RIS case and NRIS case are the same. Therefore, Algorithm 3.2 outperforms Algorithm 3.1 due to the fact that the former can generate the same gain with the latter while costing much less CPU running time

3.4.4 RIS vs AF Relay Performance Comparison

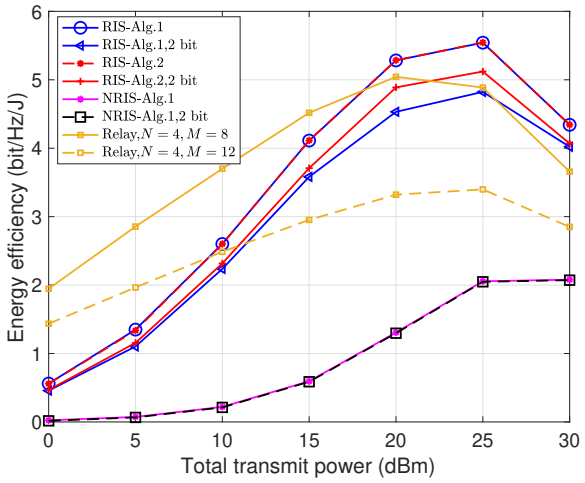
Fig. 3.5 shows the sum rate, the energy efficiency, and the corresponding CPU running time under different maximum transmit power. The energy efficiency (bit/Hz/J) is defined as the ratio of the sum rate to the power consumption, i.e.,

$$EE = \frac{Sum\ Rate}{Power}.$$

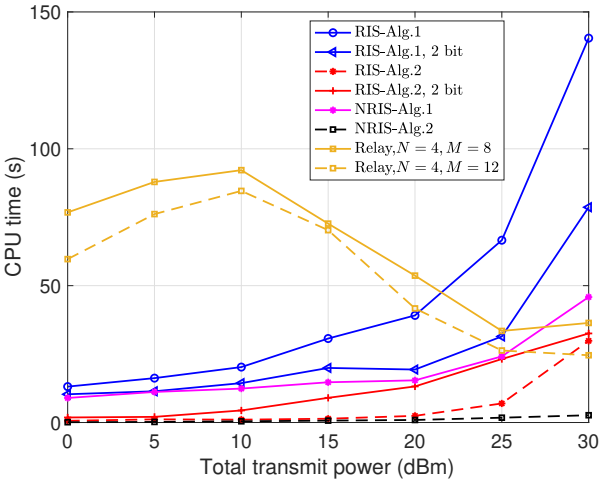
In the relay-aided system, it sets $P_{\mathrm{T}} = P_{\mathrm{relay}}$. The linear power consumption model is $Power = \eta(p_{\mathrm{T}} + p_{\mathrm{relay}}) + NP_t + 2MP_r$, where p_{T} and p_{relay} are the practical transmit power of the BS and the relay, respectively. Following [110], the reciprocal of the power amplifier efficiency is set as $\eta = 1.2$ and the circuit power consumption of the active



(a) Sum rate



(b) Energy efficiency



(c) CPU time

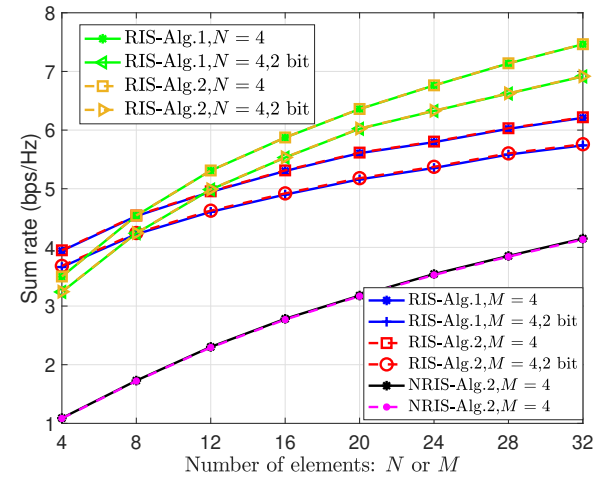
Figure 3.5: The sum rate, energy efficiency, and CPU time versus the total transmit power, when $N = 4$, $M = 16$ and $G = |\mathcal{K}_g| = 2$.

antennas at the BS and the relay as $P_t = P_r = 200$ mW. In the RIS-aided system, it adopts $Power = \eta(p_{\mathrm{T}} + p_{\mathrm{relay}}) + NP_t + MP_{RIS}$, where the circuit power consumption of the passive reflection elements is set as $P_{RIS} = 5$ mW [111].

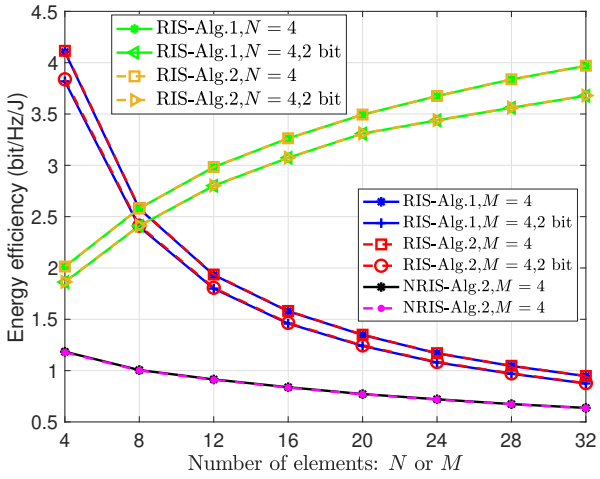
It can be seen in Fig. 3.5(a) that the RIS structure can obviously enhance the sum rate performance of the system without consuming additional transmit power, comparing with the system without the RIS structure. The performance loss of the '2 bit' phase shifter generated by the proposed two algorithms is much small compared with the continuous phase shifter cases. However, the relay-aided system outperforms the RIS-aided one, which is reasonable due the fact that the relay can amplify and forward the received signals by using the relay transmit power P_{relay}. The EE of the RIS-aided system shown in Fig. 3.5(b) is higher than the relay-aided one at high transmit power. The reason behind this is twofold. On the one hand, as P_{T} increases, the contribution of the relay transmit power P_{relay} to the system sum rate gain becomes less. On the other hand, the circuit power consumption of the relay is relatively high. Another observation from Fig. 3.5(b) is that the EE of the relay system decreases with the number of the active antennas deployed at the relay. From Fig. 3.5(c), it is observed that Algorithm 3.1 is time-consuming and the time required is unacceptable when P_{T} increases. In addition, the computational complexity of the joint optimization of the precoder and the relay beamforming is much higher than the RIS case when P_{T} is less than 20 dBm due to the fact that relay power constraint is complex. Finally, all the results obtained from Fig. 3.5 verify the performance gains of the RIS-aided system in terms of the EE and complexity.

3.4.5 RIS Performance Analysis

It is of practical significance to compare the communication performance of conventional large-scale antenna arrays deployed at the BS and large-scale passive elements deployed at the RIS, since RIS is regarded as an extension of massive MIMO antenna array. Fig. 3.6 illustrates the sum rate and the EE performance versus the numbers of antenna ele-



(a) Sum rate



(b) Energy efficiency

Figure 3.6: The sum rate versus the numbers of reflection elements at the RIS M or transmit antennas at the BS N, when $G = |\mathcal{K}_g| = 2$ and $P_{\mathrm{T}} = 20$ dBm.

ments at the BS and reflection elements at the RIS when $P_{\mathrm{T}} = 20$ dBm. It is observed from Fig. 3.6(a) that significant gains can be achieved by the RIS scheme over that without an RIS even when M is as small as 4, and also that the spectral efficiency performance gains achieved by increasing the number of reflection elements are much higher than those achieved by increasing the number of transmit antennas. In addition, in Fig. 3.6(b), it is more energy-efficient to deploy an RIS with passive elements than installing active large-scale antenna array with energy-consuming RF chains and power amplifiers.

The trend of EE decreasing with the number of transmit antennas comes from the fact that the circut energy consumption of more antennas outweighs the system sum rate gain introduced by deploying more antennas. These simulation results demonstrate that RIS technology is superior to traditional massive MIMO in terms of spectral efficiency and energy efficiency.

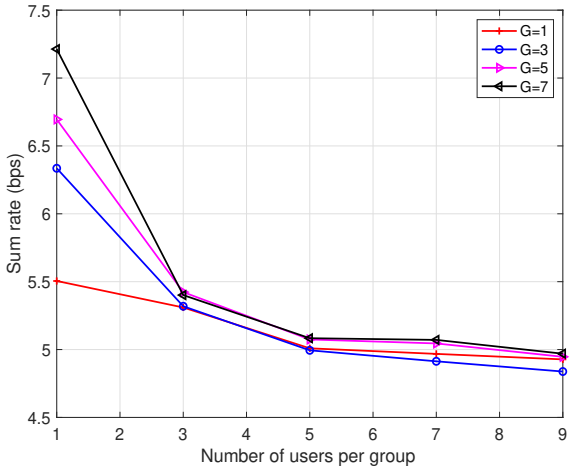


Figure 3.7: The sum rate versus the number of users per group, when $N = 4$, $M = 16$, and $P_{\mathrm{T}} = 20$ dBm.

The above simulation results show that Algorithm 3.2 requires less CPU time than Algorithm 3.1. Hence, Algorithm 3.2 is adopted to investigate the effect of an RIS on the performance of a multicast communication system. Fig. 3.7 illustrates the sum rate versus the number of users per group for various numbers of groups. It can be observed from this figure that the sum rate for all values of G decreases with the increase of the number of users per group. The reason is that the data rate for each group is limited by the user with the worst channel condition. With the increase of the number of users per group, the channel gain for the worst user becomes smaller.

Fig. 3.8 compares the effects of two improvements on the performance limit, namely, increasing the number of antennas at the BS and the number of reflection elements at the RIS, respectively. When $|\mathcal{K}_g| = 1$, the multicasting system reduces to a unitcasting system, in which the transmit antennas outperform the reflection elements in the aspect

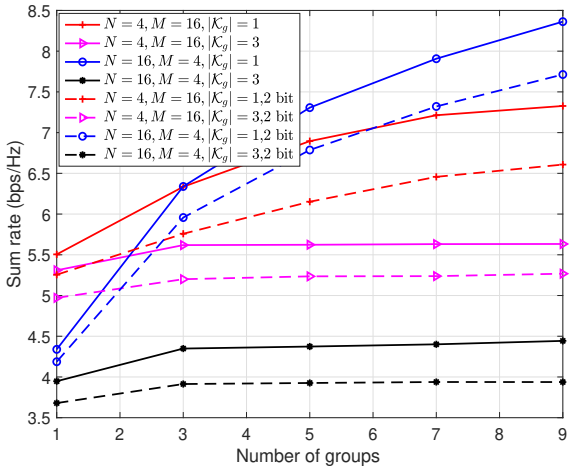


Figure 3.8: The sum rate versus the number of groups, when $P_{\mathrm{T}} = 20$ dBm.

of suppressing multi-user interference. While when $|\mathcal{K}_g| = 3$, the sum rate of the system increases slowly and tends to be stable with the increase of the number of multicasting groups for a given number of antenna/reflection elements. This is because the multicast transmission pays more gain to enhance the rate of the worst user in the multicast group for ensuring fairness, at the expense of sum rate improvement.

3.5 Summary

In this chapter, it has shown the performance benefits of introducing an RIS to the multi-group multicast systems. By carefully adjusting the reflection coefficients at the RIS, the signal reflected by the RIS can enhance the strength of the signal received by the user. It investigates the sum rate maximization problem by joint optimization of the precoding matrix at the BS and reflection coefficient vector at the RIS, while guaranteeing the transmit power constraint and the associated non-convex unit-modulus constraint at the RIS. Under the MM algorithm framework, it derives the concave lower bound of the original non-concave objective function, and then adopts alternating optimization method to update variables in an alternating manner. Furthermore, this work proposed a low-complexity algorithm under the MM algorithm framework in which there exists closed-form solutions at each iteration. The simulation results have demonstrated the

significant spectral and energy efficiency enhancement of the RIS in multigroup multicast systems and that the proposed algorithm converges rapidly in terms of CPU time.

Chapter 4

Robust Transmission in the Presence of Random Blockages

This chapter proposes a robust transmission strategy for application to RIS-aided mmWave communication systems, which accounts for the channel uncertainties caused by the presence of random blockages while ensuring the fairness among the users. Typical methods to handle the presence of channel uncertainties at the design stage are the outage constrained robust optimization and the worst-case robust optimization techniques [75]. However, both methods rely on the estimation of the instantaneous CSI. Furthermore, the worst-case robust optimization method is conservative and hence suboptimal due to the low probability of occurrence of the worst case. This chapter considers the design of robust beamforming schemes for application to mmWave systems without relying on the knowledge of instantaneous CSI. The proposed approach is based, on the other hand, on the knowledge of large-scale CSI and the blockage probability. The proposed approach is motivated by the results reported in [22], where the authors have shown that the time-scale at which mmWave signals are randomly disrupted by spatial blockages is of the order of a few 100 milliseconds (or more), as well as the results reported in [44], where the authors have shown that the blockage probability is determined by the trans-

mission distance and by some environment-specific parameters. Therefore, a mmWave link is disrupted by the same blockages for several physical layer resource blocks and the associated probability can be assumed to be known if the large-scale CSI is assumed to be known. Specifically, this work formulates a maximum outage probability minimization problem and solve it by using a stochastic optimization framework. The main contributions of this chapter can be summarized as follows:

- To the best of the knowledge, this is the first work that introduces a robust beamforming design for RIS-aided downlink multiuser mmWave systems that relies on the knowledge of large-scale CSI and blockage probability. Specifically, the considered optimization criterion is based on minimizing the maximum outage probability of all the users. In contrast to the sum outage probability minimization problem in [25], [48], the considered min-max outage probability problem ensures the desired QoS performance to the worst-case user. Because of the non-differentiable objective function of the considered problem, the Stochastic Gradient Descent (SGD) method adopted in [25], [48] cannot be directly applied. To circumvent this issue, two stochastic optimization frameworks are introduced for jointly optimizing the beamforming at the BS and at the RIS.

- First, it considers the single-user case and optimizes the beamforming schemes at the BS and RIS by minimizing the outage probability given the large-scale CSI and the blockage probability. Since the objective function of the considered problem is not formulated in a closed-form expression, it is approximated with the statistical expectation of a smooth function that is twice differentiable. The resulting expectation optimization problem is solved by adopting the SMM method. Specifically, an upper bound surrogate function of the original differentiable function is constructed for any new channel realization at each iteration. The constructed surrogate problem is shown to have a closed-form solution and to be computationally efficient. It is proved that the proposed SMM method is guaranteed to converge to the set of stationary points of the original expectation minimization problem.

- Then, it considers the multi-user case and a min-max outage probability optimization problem is formulated. To tackle the non-differentiability of the max objective function, it is replaced with the log-sum-exp upper bound. Then, the SSCA method is employed, which offers a greater flexibility than the SMM method in terms of selecting the surrogate function and results in closed-form expressions at each iteration. Also in this case, it is proved that the proposed SSCA method is guaranteed to converge to the set of stationary points of the original expectation minimization problem.

- It is demonstrated through numerical results that the proposed robust beamforming algorithm outperforms its non-robust counterpart and the robust beamforming algorithm for conventional systems in the absence of RISs. If the blockage probability is high, the proposed methods outperform the others in terms of maximum outage probability and minimum effective rate. Moreover, deploying multiple small-size RISs is shown to provide better performance than deploying a single large-size RIS in terms of improving the performance of the worst-case user.

4.1 System Model

4.1.1 Signal Model

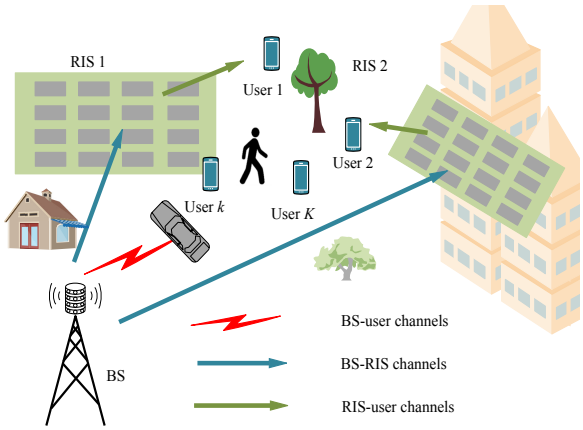


Figure 4.1: Multiple RIS-aided mmWave communication system.

As shown in Fig. 4.1 [1], an RIS-aided downlink mmWave communication system is considered. In order to ensure high QoS for the users in the presence of random blockages, U RISs, each of which has M reflecting elements, are deployed to assist the communication from the BS equipped with N active antennas to K single-antenna users (denoted by $\mathcal{K} \triangleq \{1, \ldots, K\}$). The RISs are assumed to be connected to controllers that exchange control information with the BS through dedicated channels [26], [27]. The baseband transmitted signal at the BS is $\mathbf{x} = \mathbf{F}\mathbf{s}$, where $\mathbf{s} \in \mathbb{C}^{K\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the Gaussian data symbol vector and $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_K] \in \mathbb{C}^{N\times K}$ denotes the full-digital beamforming matrix. The baseband transmit power is limited to the total transmit power P_{max}. Hence, $\mathbf{F}$ belongs to the set $\mathcal{S}_f = \{\mathbf{F} \mid ||\mathbf{F}||_F^2 \leq P_{max}\}$.

Let $\mathbf{h}_{\mathrm{b},k} \in \mathbb{C}^{N\times 1}$, $\mathbf{H}_u \in \mathbb{C}^{M\times N}$ and $\mathbf{h}_{u,k} \in \mathbb{C}^{M\times 1}$ denote the channels of the links from the BS to the k-th user, from the BS to the u-th RIS, and from the u-th RIS to the k-th user, respectively. Then, the received signal intended to the k-th user is expressed as [2]

$$y_k = \left(\mathbf{h}_{\mathrm{b},k}^{\mathrm{H}} + \sum_{u=1}^{U} \mathbf{h}_{u,k}^{\mathrm{H}} \mathbf{E}_u \mathbf{H}_u\right) \mathbf{x} + n_k, \tag{4.1}$$

where $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ is the AWGN, and $\mathbf{E}_u = \zeta \mathrm{diag}([e_{(u-1)M+1}, \ldots, e_{uM}])$ is the reflection coefficient matrix (also known as the RIS beamforming matrix) of the u-th RIS. The element $e_{(u-1)M+m}$ is the m-th unit modulus reflection coefficient at the u-th RIS, and $\zeta \in [0, 1]$ denotes the reflection efficiency. it is assumed, independently of the applied phase shift and of the angle of incidence, $\zeta = 1$ for simplicity, since it offers the best reflection performance for the RIS.

By defining the matrices $\mathbf{h}_k = [\mathbf{h}_{1,k}^{\mathrm{H}}, \ldots, \mathbf{h}_{U,k}^{\mathrm{H}}]^{\mathrm{H}}$ and $\mathbf{H} = [\mathbf{H}_1^{\mathrm{H}}, \ldots, \mathbf{H}_U^{\mathrm{H}}]^{\mathrm{H}}$, it obtains the equivalent channel $\mathbf{G}_k = \begin{bmatrix} \mathrm{diag}(\mathbf{h}_k^{\mathrm{H}})\mathbf{H} \\ \mathbf{h}_{\mathrm{b},k}^{\mathrm{H}} \end{bmatrix} \in \mathbb{C}^{(UM+1)\times N}$ between the BS and the

[1] 2D scenario can be extended to 3D scenario.

[2] For simplicity, the reflections of signals between the RISs is ignored, since they are typically weak in the mmWave frequency band. The impact of the reflected signals between RISs was recently addressed in [112].

k-th user. The corresponding equivalent reflection coefficient vector is given by $\mathbf{e} = [e_1, \ldots, e_{UM+1}]^{\mathrm{T}} \in \mathbb{C}^{(UM+1)\times 1}$ which belongs to the set $\mathcal{S}_e = \{\mathbf{e} \mid |e_m|^2 = 1, 1 \leq m \leq UM, e_{UM+1} = 1\}$. Then, (4.1) can be rewritten in a compact form as

$$y_k = \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\mathbf{F}\mathbf{s} + n_k. \tag{4.2}$$

Therefore, the corresponding achievable SINR is

$$\Gamma_k(\mathbf{F}, \mathbf{e}) = \frac{|\mathbf{e}^{\mathrm{H}}\mathbf{G}_k\mathbf{f}_k|^2}{\sum_{i\neq k}^{K} |\mathbf{e}^{\mathrm{H}}\mathbf{G}_k\mathbf{f}_i|^2 + \sigma_k^2}. \tag{4.3}$$

4.1.2 Channel Model

It is considered a typical Saleh-Valenzuela (SV) [113] channel model for application to mmWave systems. For simplicity, it ignores the randomness introduced by the presence of hardware impairments that may affect the performance of mmWave systems. In particular, it is assumed that a UPA is deployed at the BS and at the RIS. The steering vector of each UPA is denoted by $\mathbf{a}(\varphi, \phi)$, where $\varphi(\phi)$ denotes the azimuth (elevation) Angle of Departure (AoD) and Angle of Arrival (AoA) depending on whether a transmitter or a receiver is considered. It is assumed that there exist maximum $L_{\mathrm{b},k}$, $L_{u,k}$ and $L_{\mathrm{b},u}$ sparse scatterers on the links from the BS to the k-th user, from the u-th RIS to the k-th user, and from the BS to the u-th RIS, respectively. Also, it is assumed that each scatterer comprises I subpaths. In the far-field region, therefore, the mmWave channels can be expressed as

$$\mathbf{h}_{\mathrm{b},k} = g_0^{\mathrm{b},k}\mathbf{a}\left(\varphi_{\mathrm{b},k,0}^{\mathrm{AoD}}, \phi_{\mathrm{b},k,0}^{\mathrm{AoD}}\right) + \sqrt{\frac{1}{IL_{\mathrm{b},k}}}\sum_{l=1}^{L_{\mathrm{b},k}}\sum_{i=1}^{I} g_{l,i}^{\mathrm{b},k}\mathbf{a}\left(\varphi_{\mathrm{b},k,l,i}^{\mathrm{AoD}}, \phi_{\mathrm{b},k,l,i}^{\mathrm{AoD}}\right), \forall k, \tag{4.4}$$

$$\mathbf{h}_{u,k} = g_0^{u,k}\mathbf{a}\left(\varphi_{u,k,0}^{\mathrm{AoD}}, \phi_{u,k,0}^{\mathrm{AoD}}\right) + \sqrt{\frac{1}{IL_{u,k}}}\sum_{l=1}^{L_{u,k}}\sum_{i=1}^{I} g_{l,i}^{u,k}\mathbf{a}\left(\varphi_{u,k,l,i}^{\mathrm{AoD}}, \phi_{u,k,l,i}^{\mathrm{AoD}}\right), \forall k, \forall u, \tag{4.5}$$

$$\begin{aligned}\mathbf{H}_u = {} & g_0^{\mathrm{b},u}\mathbf{a}\left(\varphi_{u,0}^{\mathrm{AoA}}, \phi_{u,0}^{\mathrm{AoA}}\right)\mathbf{a}\left(\varphi_{\mathrm{b},0}^{\mathrm{AoD}}, \phi_{\mathrm{b},0}^{\mathrm{AoD}}\right)^{\mathrm{H}} \\ & + \sqrt{\frac{1}{IL_{\mathrm{b},u}}}\sum_{l=1}^{L_{\mathrm{b},u}}\sum_{i=1}^{I} g_{l,i}^{\mathrm{b},u}\mathbf{a}\left(\varphi_{u,l,i}^{\mathrm{AoA}}, \phi_{u,l,i}^{\mathrm{AoA}}\right)\mathbf{a}\left(\varphi_{\mathrm{b},l,i}^{\mathrm{AoD}}, \phi_{\mathrm{b},l,i}^{\mathrm{AoD}}\right)^{\mathrm{H}}, \forall u,\end{aligned} \tag{4.6}$$

where, by denoting an arbitrary element $q \in \{(\mathrm{b},k),(u,k),(\mathrm{b},u)\}_{\forall k,\forall u}$, $g_0^q \mathbf{a}\left(\varphi_{q,0}^{\mathrm{AoD}}, \phi_{q,0}^{\mathrm{AoD}}\right)$ is the LoS component whose fading coefficient has distribution $g_0^q \sim \mathcal{CN}(0, \zeta_0^q 10^{\frac{\mathrm{PL}}{10}})$, where $\zeta_0^q = \frac{\kappa}{1+\kappa}$ is the power that corresponds to the Rician factor κ, and PL is the distance-dependent path-loss. The remaining paths are the Non Line-of-Sight (NLoS) components whose fading coefficients have distribution $g_{l,i}^q \sim \mathcal{CN}(0, \zeta_l^q 10^{\frac{\mathrm{PL}}{10}})$ where $\zeta_l^q = \frac{1}{(L_q-1)I(1+\kappa)}$ is the corresponding power fraction.

It is assumed that the users' locations are quasi-static over milliseconds or even seconds. Therefore, the large-scale fading parameters, such as the distance-dependent path-loss, the number of clusters, the Rician factor, the cluster central angles, and the angular spreads, change relatively slowly and can be assumed to be known by the BS [114], [115]. However, the instantaneous CSI, which is given by $\{\mathbf{h}_{\mathrm{b},k}, \mathbf{h}_{u,k}, \mathbf{H}_u\}$, vary during the data transmission because of the rapidly varying small-scale fading coefficients $\{g_0^q, g_{l,i}^q\}$, AoDs and AoAs. In general, these parameters vary according to an ergodic stationary process. For example, the AoDs and AoAs can be generated according to a Gaussian distribution, whose mean value coincide with the central angles of the clusters and the variance coincides with the angular spread [114].

Besides the path-loss and the small-scale fading, the reliability of the communication links in the mmWave frequency band is determined by the presence of blockages [116]. In the context of RIS-aided communications, most existing contributions have considered the worst-case scenario where the BS-user links are completely blocked due to the presence of obstacles during the whole transmission, while the RIS-related links are not affected by the presence of blockages since the locations of the RISs can be appropriately chosen to ensure line-of-sight transmission. However, this assumption may not represent all possible deployment scenarios. Traditionally, the presence of blockages is incorporated in the shadowing model, along with the impact of reflections, scattering, and diffraction [43]. In contrast, it is adopted a recently proposed probabilistic model [48] to characterize the channel uncertainties caused by the presence of random blockages. The considered model is more realistic, since the impact of blockages and the

corresponding blockage probability depends on the transmission distance [43].

Specifically, let us introduce the blockage parameters $\omega_{k,l} \in \{0,1\}, 0 \leq l \leq L_{\mathrm{b},k}, \forall k \in \mathcal{K}$, which are random variables with a Bernoulli distribution. In particular, these random variables take the value one with probability p_k, which is referred to as the blockage probability. With the aid of these random variables, the presence of blockages can be taken into account in the formulation of the mmWave channels. Specifically, the BS-user links in (4.4) are modified as

$$\mathbf{h}_{\mathrm{b},k} = \omega_{k,0} g_0^{\mathrm{b},k} \mathbf{a}\left(\varphi_{\mathrm{b},k,0}^{\mathrm{AoD}}, \phi_{\mathrm{b},k,0}^{\mathrm{AoD}}\right) + \sqrt{\frac{1}{IL_{\mathrm{b},k}}} \sum_{l=1}^{L_{\mathrm{b},k}} \omega_{k,l} \sum_{i=1}^{I} g_{l,i}^{\mathrm{b},k} \mathbf{a}\left(\varphi_{\mathrm{b},k,l,i}^{\mathrm{AoD}}, \phi_{\mathrm{b},k,l,i}^{\mathrm{AoD}}\right), \forall k. \quad (4.7)$$

As far as the blockage probability is concerned, it is known that it usually depends on the transmission distance. For example, the authors of [43] have shown that the probability that a link is blocked, i.e., there is at least one object in between the transmitter and the receiver, can be formulated as $p_k(d_k) = \max(0, 1 - e^{-a_{out} d_k + b_{out}})$, where d_k is the transmission distance between the BS and the k-th user in the considered system model and a_{out} and b_{out} are environment-dependent parameters that can be obtained from theory or can be obtained from curve fitting from data [44], [Table II, [117]].

4.1.3 Problem Formulation

Since the RISs are not equipped with power amplifiers and with digital signal processing units, the acquisition of CSI is difficult to obtain. This is especially true if the BS and the RISs are optimized based on perfect instantaneous CSI, since a large training overhead would be needed [118]. Therefore, it is important to develop robust beamforming schemes that do not necessarily rely on the knowledge of the instantaneous CSI, but still account for the impact of the large-scale CSI, which is easier to acquire at a reduced overhead, and that are robust to the presence of random blockages. Motivated by this consideration, this chapter aims to design robust beamforming schemes for RIS-aided systems that depend only on the large-scale CSI and the blockage probability, but are independent of the short-term CSI, i.e., the fast fading. Also, this chapter aims to ensure that the

beamforming schemes provide fairness to the network users.

To this end, the min-max outage probability optimization problem is formulated as

$$\min_{\mathbf{F},\mathbf{e}} \ \max_{k\in\mathcal{K}} \Pr\{\Gamma_k(\mathbf{F},\mathbf{e}) \leq \gamma_k\} \tag{4.8a}$$

$$\text{s.t. } \mathbf{F} \in \mathcal{S}_f \tag{4.8b}$$

$$\mathbf{e} \in \mathcal{S}_e, \tag{4.8c}$$

where the outage probability $\Pr\{\Gamma_k(\mathbf{F},\mathbf{e}) \leq \gamma_k\}$ is the probability that the SINR $\Gamma_k(\mathbf{F},\mathbf{e})$ of the k-th user is less than the SINR reliability threshold γ_k for all possible realizations of the random channel $\mathbf{G} = [\mathbf{G}_1, \ldots, \mathbf{G}_K]$. Specifically, the probability in (4.8) is computed as a function of the small-scale fading coefficients, the AoDs and the AoAs of the subpaths of the scatterers. Notably, the outage probability in (4.8) depends on the transmission distance, the blockage probability, the number of clusters and their centers and angular spreads.

Compared with the sum outage probability minimization problem formulated in [25], the objective function in (4.8) ensures fairness among the users. However, due to the min-max formulation, the objective function is not smooth and differentiable, which makes the algorithms proposed in [25] not directly applicable to solve the problem in (4.8).

4.2 Single-User System

This section considers a single-user system model in order to obtain some design insights. By setting $K = 1$ and omitting the user index, the problem in (4.8) reduces to

$$\min_{\mathbf{f},\mathbf{e}} \ \Pr\{\Gamma(\mathbf{f},\mathbf{e}) \leq \gamma\} \tag{4.9a}$$

$$\text{s.t. } \mathbf{f} \in \mathcal{S}_f \tag{4.9b}$$

$$\mathbf{e} \in \mathcal{S}_e. \tag{4.9c}$$

4.2.1 Problem Reformulation

The probability $\Pr\{\Gamma(\mathbf{f},\mathbf{e}) \leq \gamma\}$ has no closed-form expression and thus the problem in (4.9) is prohibitively challenging to be solved. To tackle this issue, the probability function is reformulated in terms of an equivalent expectation function, i.e., $\Pr\{\Gamma(\mathbf{f},\mathbf{e}) \leq \gamma\} = \mathbb{E}_{\mathbf{G}}[\mathbb{I}_{\Gamma\leq\gamma}]$ where $\mathbb{I}_{\Gamma\leq\gamma}$ denotes the step function applied to the event $\Gamma \leq \gamma$. Thanks to this reformulation, several stochastic programming techniques can be used to solve the problem in (4.9). However, the step function is discontinuous, and the existing stochastic programming methods cannot be directly applied.

To circumvent this issue, the step function is approximated with the following smooth approximating function

$$u(x) = \frac{1}{1+e^{-\theta x}}, \tag{4.10}$$

where $x = \gamma - \Gamma$ and θ is a smooth parameter that controls the approximation error. Specifically, the larger θ is the closer to an ideal step function the function in (4.10) is.

By defining $f(\mathbf{f},\mathbf{e}|\mathbf{G}) = u\left(\gamma\sigma^2 - |\mathbf{e}^{\mathrm{H}}\mathbf{G}\mathbf{f}|^2\right)$, a convenient approximation of the problem formulated in (4.9) is

$$\min_{\mathbf{f}\in\mathcal{S}_f,\mathbf{e}\in\mathcal{S}_e} \; g(\mathbf{f},\mathbf{e}|\mathbf{G}) = \mathbb{E}\left[f(\mathbf{f},\mathbf{e}|\mathbf{G})\right]. \tag{4.11}$$

4.2.2 Stochastic Majorization-Minimization Method

A simple approach for solving the problem in (4.11) is the Sample Average Approximation (SAA) method. However, the SAA method is computationally prohibitive since it requires large-size memory storage due to the fact that the solution obtained at each iteration is calculated by averaging over a large number of channel realizations. To overcome these difficulties, it is adopted the widely used SMM [119] (also known as stochastic successive minimization [94]) method. Accordingly, an appropriately chosen upper bound approximation for the function $f(\mathbf{f},\mathbf{e}|\mathbf{G})$ is constructed at each iteration of the algorithm and for each channel realization. The solution is obtained as the average

over the channel realizations at each iteration.

The typical approach to apply this method consists of introducing an upper bound approximation function for $f(\mathbf{f}, \mathbf{e}|\mathbf{G})$ that satisfies Assumption A and B in Section 2.2.2 and then makes the corresponding surrogate problem easy to solve. In some cases, a closed-form solution may be found as well.

Since the variables $\mathbf{f}$ and $\mathbf{e}$ are highly coupled with each other, an AO method is adopted to update them. Based on Assumption A and B, the variables $\mathbf{f}$ and $\mathbf{e}$ are updated, at the t-th iteration of the algorithm, by solving the following two SMM sub-problems

$$\mathbf{f}^t = \arg\min_{\mathbf{f}\in\mathcal{S}_f} \ \frac{1}{n}\sum_{i=1}^{t} \hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}|\mathbf{G}^i\right), \tag{4.12}$$

and

$$\mathbf{e}^t = \arg\min_{\mathbf{e}\in\mathcal{S}_e} \ \frac{1}{n}\sum_{i=1}^{t} \hat{f}\left(\mathbf{e}, \mathbf{e}^{i-1}|\mathbf{G}^i\right), \tag{4.13}$$

where $\mathbf{G}^1, \mathbf{G}^2, ...$ are some independent samples of the random equivalent channel $\mathbf{G}$ [3]. Furthermore, $\hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}|\mathbf{G}^i\right)$ denotes the surrogate function associated to $\mathbf{f}$ when $\mathbf{e}$ is given, and $\hat{f}\left(\mathbf{e}, \mathbf{e}^{i-1}|\mathbf{G}^i\right)$ is the surrogate function that corresponds to $\mathbf{e}$ when $\mathbf{f}$ is given.

4.2.2.1 Optimizing f

First, $\hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}|\mathbf{G}^i\right)$ is constructed so as to fulfill the Assumptions A and B. This is given in the following lemma.

Lemma 2. *Given the twice differentiable function $f(\mathbf{f}|\mathbf{G}^i)$, this work considers the fol-*

[3] More precisely, it is assumed that the large-scale fading parameters are kept fixed, and that the samples of $\mathbf{G}$ are constituted by $\{\mathbf{h}_{\mathrm{b},k}, \mathbf{h}_{u,k}, \mathbf{H}_u\}$, which are obtained by generating the random variables $\{g_0^q, g_{l,i}^q\}$, AoDs and AoAs according to their distributions whose parameters are assumed to be known, as well as the Bernoulli random variables $\omega_{k,l}$ whose blockage probability p_k is assumed to be known.

lowing second-order upper bound approximation around any given $\mathbf{f}^{i-1}$

$$\hat{f}(\mathbf{f}, \mathbf{f}^{i-1}|\mathbf{G}^i) = 2\mathrm{Re}\left\{\mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \alpha_f^i||\mathbf{f}||_2^2 + const_f^i, \tag{4.14}$$

where

$$\mathbf{d}_f^i = \mathbf{m}_f^i - \alpha_f^i\mathbf{f}^{i-1}, \tag{4.15a}$$

$$\mathbf{m}_f^i = \frac{-\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}, \tag{4.15b}$$

$$x^i = \gamma\sigma^2 - |\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}|^2, \tag{4.15c}$$

$$\alpha_f^i = \frac{\theta^2}{2}P_{max}|\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}|^2, \tag{4.15d}$$

$$const_f^i = f(\mathbf{f}^{i-1}|\mathbf{G}^i) + \alpha_f^i||\mathbf{f}^{i-1}||_2^2 - 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}\mathbf{f}^{i-1}\right\}. \tag{4.15e}$$

Proof: Please refer to Appendix B.1. ■

By using (4.14) and ignoring the irrelevant constants, the subproblem in (4.12) for updating $\mathbf{f}$ is reformulated as

$$\min_{\mathbf{f}\in\mathcal{S}_f} \; 2\mathrm{Re}\left\{\frac{1}{t}\sum_{i=1}^{t}\mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \frac{1}{t}\sum_{i=1}^{t}\alpha_f^i||\mathbf{f}||_2^2. \tag{4.16}$$

The optimization problem in (4.16) is convex and can be solved by computing its Lagrange function given by

$$\mathcal{L}(\mathbf{f},\kappa) = 2\mathrm{Re}\left\{\frac{1}{t}\sum_{i=1}^{t}\mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \frac{1}{t}\sum_{i=1}^{t}\alpha_f^i||\mathbf{f}||_2^2 + \kappa\left(||\mathbf{f}||_2^2 - P_{max}\right), \tag{4.17}$$

where $\kappa \geq 0$ is the Lagrange multiplier associated with the power constraint. By setting $\partial\mathcal{L}(\mathbf{f})/\partial\mathbf{f}^* = \mathbf{0}$, the globally optimal solution of $\mathbf{f}$ at the t-th iteration is

$$\mathbf{f}^t = \frac{-1}{\kappa + \frac{1}{t}\sum_{i=1}^{t}\alpha_f^i}\frac{1}{t}\sum_{i=1}^{t}\mathbf{d}_f^i. \tag{4.18}$$

Also, (4.18) must satisfy the power constraint in (4.9b), which yields

$$\frac{||\frac{1}{n}\sum_{i=1}^{t}\mathbf{d}_f^i||_2^2}{(\kappa+\frac{1}{t}\sum_{i=1}^{t}\alpha_f^i)^2} \leq P_{max}. \tag{4.19}$$

Since the left hand side of (4.19) is a decreasing function of κ, the closed-form solution is obtained as

$$\mathbf{f}^t = \begin{cases} \frac{-1}{\sum_{i=1}^{t}\alpha_f^i}\sum_{i=1}^{t}\mathbf{d}_f^i, & \text{if } \frac{||\sum_{i=1}^{t}\mathbf{d}_f^i||_2^2}{(\sum_{i=1}^{t}\alpha_f^i)^2} \leq P_{max}, \\ -\sqrt{\frac{P_{max}}{||\sum_{i=1}^{t}\mathbf{d}_f^i||_2^2}}\sum_{i=1}^{t}\mathbf{d}_f^i, & \text{otherwise,} \end{cases} \tag{4.20}$$

where the first case in (4.20) is obtained by setting $\kappa = 0$, and the second case follows because there must exist a $\kappa > 0$ for which (4.19) holds with equality.

4.2.2.2 Optimizing e

Similar to the optimization of $\mathbf{f}$, it is first constructed a surrogate function for $\mathbf{e}$.

Lemma 3. *Given the twice differentiable function* $f\left(\mathbf{e}|\mathbf{G}^i\right)$*, this work considers the following second-order upper bound approximation around any feasible* $\mathbf{e}^{i-1}$

$$\hat{f}(\mathbf{e},\mathbf{e}^{i-1}|\mathbf{G}^i) = 2\mathrm{Re}\left\{\mathbf{d}_e^{i,\mathrm{H}}\mathbf{e}\right\} + const_e^i, \tag{4.21}$$

where

$$\mathbf{d}_e^i = \mathbf{m}_e^i - \alpha_e^i\mathbf{e}^{i-1}, \tag{4.22a}$$

$$\mathbf{m}_e^i = \frac{-\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2}\mathbf{G}^i\mathbf{f}^{i-1}\mathbf{f}^{i-1,\mathrm{H}}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}, \tag{4.22b}$$

$$\alpha_e^i = \frac{\theta^2}{2}(UM+1)|\mathbf{f}^{i-1,\mathrm{H}}\mathbf{G}^{i,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}|^2, \tag{4.22c}$$

$$const_e^i = f(\mathbf{e}^{i-1}|\mathbf{G}^i) + 2(UM+1)\alpha_e^i - 2\mathrm{Re}\left\{\mathbf{m}_e^{i,\mathrm{H}}\mathbf{e}^{i-1}\right\}. \tag{4.22d}$$

Proof: The proof of Lemma 3 is similar to that of Lemma 2 and it is hence omitted

for brevity. ■

By substituting (4.21) into the objective function of the subproblem in (4.13) and ignoring the irrelevant constants, it is obtained

$$\min_{\mathbf{e}\in\mathcal{S}_e} \; 2\mathrm{Re}\left\{\frac{1}{n}\sum_{i=1}^{t}\mathbf{d}_e^{i,\mathrm{H}}\mathbf{e}\right\}. \tag{4.23}$$

The globally optimal solution of the optimization problem in (4.23) is

$$\mathbf{e}^t = \exp\left\{\mathrm{j}\angle\left(\left(\sum_{i=1}^{t}\mathbf{d}_e^i\right) / \left[\sum_{i=1}^{t}\mathbf{d}_e^i\right]_{UM+1}\right)\right\}, \tag{4.24}$$

where $[\cdot]_m$ denotes the m-th element of the vector, $\mathrm{j} \triangleq \sqrt{-1}$ is the imaginary unit, $\angle(\cdot)$ denotes the angle of a complex number, and $\exp\{\mathrm{j}\angle(\cdot)\}$ is an element-wise operation.

4.2.3 Algorithm Development

By leveraging the SMM method, it has been obtained the closed-form solutions in (4.20) and (4.24) for $\mathbf{f}$ and $\mathbf{e}$, respectively. The closed-form solutions, at each iteration of the algorithm, greatly reduce the computational complexity. The whole numerical recipe is reported in Algorithm 4.1, which is referred to as the SMM-OutMin algorithm.

Algorithm 4.1: SMM-OutMin Algorithm

Require: Initialize $\mathbf{f}^0 \in \mathcal{S}_f$ and $\mathbf{e}^0 \in \mathcal{S}_e$. Set $t = 1$.

1: **repeat**
2: Obtain the sample channel $\mathbf{G}^t$.
3: Update $\mathbf{f}^t$ according to (4.20).
4: Update $\mathbf{e}^t$ according to (4.24).
5: $t = t + 1$.
6: **until** $||\mathbf{f}^t - \mathbf{f}^{t-1}||_2 \to 0$ and $||\mathbf{e}^t - \mathbf{e}^{t-1}||_2 \to 0$.

4.2.3.1 Convergence Analysis

The convergence of Algorithm 4.1 is analyzed in the following theorem.

Theorem 6. *Assume that Assumptions A and B are satisfied. Then, the sequence of the*

solutions obtained in each iteration of Algorithm 4.1 converges to the set of stationary points of the problem in (4.11) almost surely.

Proof: Please refer to Appendix B.2. ■

4.2.3.2 Complexity Analysis

The computational complexity for updating $\mathbf{f}^t$ and $\mathbf{e}^t$ at each iteration mainly depends on the computation of (4.20) and (4.24), respectively. In particular, due to the update rule in $\{\sum_{i=1}^{t}\alpha_f^i, \sum_{i=1}^{t}\mathbf{d}_f^i, \sum_{i=1}^{t}\mathbf{d}_e^i\}$, only $\{\alpha_e^t, \mathbf{d}_f^t, \mathbf{d}_e^t\}$ need to be calculated at the t-th iteration. Therefore, the approximate complexity of each iteration is $\mathcal{O}(4UMN + 12N)$.

4.3 Multiuser System

This section considers the general multiuser setup and solve the optimization problem in (4.8). The min-max problem in (4.8) is more challenging to tackle as compared with the problem in (4.9) due to the presence of the max function. To tackle the problem in (4.8), the SMM method applied to the single user case is extended.

4.3.1 Problem Reformulation

This work first approximate the probability function in the original formulation of the problem in (4.8) by still using the smooth function in (4.10). To this end, it is defined $f_k\left(\mathbf{F}, \mathbf{e}|\mathbf{G}\right) = u\left(\mathbf{e}^{\mathrm{H}}\mathbf{G}_k\mathbf{F}\boldsymbol{\Upsilon}_k\mathbf{F}^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\mathbf{e} + \gamma_k\sigma_k^2\right)$, where $\boldsymbol{\Upsilon}_k$ is a diagonal matrix whose diagonal entries are all equal to γ_k with the exception of the k-th diagonal element that is equal to -1. Therefore, an approximate expression for the objective function in (4.8) is $\max_{k\in\mathcal{K}}\mathbb{E}\left[f_k\left(\mathbf{F}, \mathbf{e}|\mathbf{G}\right)\right]$. However, the obtained objective function is still intractable since the maximization operation and the expectation operation make the functions $f_k, \forall k$ and the different channel states coupled, respectively. To circumvent these issues, this work uses Jensen's inequality

$$\max_{k\in\mathcal{K}}\mathbb{E}\left[f_k\left(\mathbf{F}, \mathbf{e}|\mathbf{G}\right)\right] \leq \mathbb{E}\left[\max_{k\in\mathcal{K}} f_k\left(\mathbf{F}, \mathbf{e}|\mathbf{G}\right)\right], \tag{4.25}$$

since the max function $\max_{k\in\mathcal{K}}\{x_1,\ldots,x_K\}$ is convex [79].

Furthermore, the non-differentiable max function, $\max_{k\in\mathcal{K}} f_k(\mathbf{F},\mathbf{e}|\mathbf{G})$, is approximated by adopting a smooth log–sum–exp upper-bound [102]

$$\max_{k\in\mathcal{K}} f_k(\mathbf{F},\mathbf{e}|\mathbf{G}) \approx F(\mathbf{F},\mathbf{e}|\mathbf{G}) = \mu \ln\Big(\sum_{k\in\mathcal{K}} \exp\left\{\frac{1}{\mu} f_k(\mathbf{F},\mathbf{e}|\mathbf{G})\right\}\Big), \tag{4.26}$$

where $\mu > 0$ is a smoothing parameter that fulfills the condition

$$\max_{k\in\mathcal{K}} f_k(\mathbf{F},\mathbf{e}|\mathbf{G}) \le F(\mathbf{F},\mathbf{e}|\mathbf{G}) \le \max_{k\in\mathcal{K}} f_k(\mathbf{F},\mathbf{e}|\mathbf{G}) + \frac{1}{\mu}\log(|\mathcal{K}|). \tag{4.27}$$

When μ is appropriately chosen, a smooth approximation for the problem in (4.8) is

$$\min_{\mathbf{F}\in\mathcal{S}_f,\mathbf{e}\in\mathcal{S}_e} \; G(\mathbf{F},\mathbf{e}|\mathbf{G}) = \mathbb{E}\left[F(\mathbf{F},\mathbf{e}|\mathbf{G})\right]. \tag{4.28}$$

4.3.2 Stochastic Successive Convex Approximation Method

Similar to the optimization problem in (4.11), the optimization problem in (4.28) may be solved by adopting the SMM method. However, the function $F(\mathbf{F},\mathbf{e}|\mathbf{G})$ in (4.26) is more complex and its second-order derivative, which is necessary to construct the upper bound surrogate function of $F(\mathbf{F},\mathbf{e}|\mathbf{G})$ as shown in Appendix B.1, is not easy to be calculated. Furthermore, the coefficient of the second-order term in the final upper bound surrogate function of $F(\mathbf{F},\mathbf{e}|\mathbf{G})$ (α_f^i in (4.14)) may not be very tight, which eventually results in a very slow convergence rate of the SMM algorithm. Therefore, it is adopted the SSCA method to overcome these issues with Assumption C in Section 2.2.3.

4.3.2.1 Optimizing F

By using (2.26) and the complex differential formula $\mathrm{d}F(\mathbf{X},\mathbf{X}^*) = \mathrm{Tr}((\frac{\partial F(\mathbf{X},\mathbf{X}^*)}{\partial \mathbf{X}})^{\mathrm{T}}\mathrm{d}\mathbf{X} + (\frac{\partial F(\mathbf{X},\mathbf{X}^*)}{\partial \mathbf{X}^*})^{\mathrm{T}}\mathrm{d}\mathbf{X}^*)$ (Eq. (3.4.55) in [120]), a surrogate function can be constructed for $\mathbf{F}$

around $\mathbf{F}^{i-1}$ when $\mathbf{e}$ is given. The surrogate function is given in

$$\begin{aligned}\hat{F}\left(\mathbf{F},\mathbf{F}^{i-1}|\mathbf{G}\right) &= F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)+\mathrm{Tr}\left(\left(\frac{\partial F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)}{\partial \mathbf{F}}\right)^{\mathrm{T}}\left(\mathbf{F}-\mathbf{F}^{i-1}\right)\right)\\ &\quad+\mathrm{Tr}\left(\left(\frac{\partial F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)}{\partial \mathbf{F}^{*}}\right)^{\mathrm{T}}\left(\mathbf{F}^{*}-\mathbf{F}^{*,i-1}\right)\right)+\frac{\tau^{i}}{2}||\mathbf{F}-\mathbf{F}^{i-1}||_{F}^{2}\\ &= F\left(\mathbf{F}^{i-1}|\mathbf{G}\right)+2\sum_{k\in\mathcal{K}} l_{k}^{i}\mathrm{Re}\left\{\mathrm{Tr}\left(\boldsymbol{\Upsilon}_{k}\mathbf{F}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}_{k}(\mathbf{F}-\mathbf{F}^{i-1})\right)\right\}\\ &\quad+\frac{\tau^{i}}{2}||\mathbf{F}-\mathbf{F}^{i-1}||_{F}^{2}\\ &= 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{P}_{f}^{i,\mathrm{H}}\mathbf{F}\right)\right\}+\frac{\tau^{i}}{2}||\mathbf{F}||_{F}^{2}+\mathrm{cons1}^{i},\end{aligned} \tag{4.29}$$

where the following parameters are introduced

$$\mathbf{P}_{f}^{i}=\mathbf{W}_{f}^{i}-\frac{\tau^{i}}{2}\mathbf{F}^{i-1}, \tag{4.30a}$$

$$\mathbf{W}_{f}^{i}=\sum_{k\in\mathcal{K}} l_{k}^{i}\mathbf{G}_{k}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{i}\mathbf{F}^{i-1}\boldsymbol{\Upsilon}_{k}, \tag{4.30b}$$

$$l_{k}^{i}=\frac{\exp\left\{\frac{1}{\mu}f_{k}\left(\mathbf{F}^{i-1},\mathbf{e}^{i-1}|\mathbf{G}^{i}\right)\right\}}{\sum_{k\in\mathcal{K}}\exp\left\{\frac{1}{\mu}f_{k}\left(\mathbf{F}^{i-1},\mathbf{e}^{i-1}|\mathbf{G}^{i}\right)\right\}}\frac{\theta e^{-\theta x_{k}^{i}}}{\left(1+e^{-\theta x_{k}^{i}}\right)^{2}}, \tag{4.30c}$$

$$x_{k}^{i}=\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{i}\mathbf{F}^{i-1}\boldsymbol{\Upsilon}_{k}\mathbf{F}^{i-1,\mathrm{H}}\mathbf{G}_{k}^{i,\mathrm{H}}\mathbf{e}^{i-1}+\gamma_{k}\sigma_{k}^{2}, \tag{4.30d}$$

$$\mathrm{cons1}^{i}=F\left(\mathbf{F}^{i-1}|\mathbf{G}^{i}\right)+\frac{\tau^{i}}{2}||\mathbf{F}^{i-1}||_{F}^{2}-2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{W}_{f}^{i,\mathrm{H}}\mathbf{F}^{i-1}\right)\right\}. \tag{4.30e}$$

By using (4.29), the optimization subproblem in (4.28) as a function of $\mathbf{F}$ is formulated, at the t-th iteration, as

$$\min_{\mathbf{F}\in\mathcal{S}_{F}}\ \frac{1}{t}\sum_{i=1}^{t}\hat{F}\left(\mathbf{F},\mathbf{F}^{i-1}|\mathbf{G}^{i}\right). \tag{4.31}$$

The obtained optimization problem in (4.31) can be solved by using the same methods as for the optimization problem in (4.16). Specifically, the global minimizer of the

optimization problem in (4.31) is

$$\widehat{\mathbf{F}}^t = \begin{cases} \frac{-2}{\sum_{i=1}^t \tau^i} \sum_{i=1}^t \mathbf{P}_f^i, & \text{if } \frac{4||\sum_{i=1}^t \mathbf{P}_f^i||_F^2}{(\sum_{i=1}^t \tau^i)^2} \leq P_{max}, \\ -\sqrt{\frac{P_{max}}{||\sum_{i=1}^t \mathbf{P}_f^i||_F^2}} \sum_{i=1}^t \mathbf{P}_f^i, & \text{otherwise.} \end{cases} \tag{4.32}$$

4.3.2.2 Optimizing e

Similarly, the optimization subproblem in (4.28) as a function of $\mathbf{e}$ when $\mathbf{F}$ is given, can be formulated, at the t-th iteration, as

$$\min_{\mathbf{e}\in\mathcal{S}_e} \ \frac{1}{t} \sum_{i=1}^{t} \hat{F}\left(\mathbf{e}, \mathbf{e}^{i-1} | \mathbf{G}^i\right), \tag{4.33}$$

where $\hat{F}\left(\mathbf{e}, \mathbf{e}^{i-1} | \mathbf{G}^i\right) = 2\text{Re}\left\{\mathbf{p}_e^{i,\text{H}} \mathbf{e}\right\} + \text{cons}2^i$, and

$$\mathbf{p}_e^i = \mathbf{w}_e^i - \frac{\tau^i}{2} \mathbf{e}^{i-1}, \tag{4.34a}$$

$$\mathbf{w}_e^i = \sum_{k\in\mathcal{K}} l_k \mathbf{G}_k \mathbf{F}^{i-1} \boldsymbol{\Upsilon}_k \mathbf{F}^{i-1,\text{H}} \mathbf{G}^{\text{H}} \mathbf{e}^{i-1}, \tag{4.34b}$$

$$\text{cons}2^i = F\left(\mathbf{e}^{i-1} | \mathbf{G}^i\right) + \tau^i (UM+1) - 2\text{Re}\left\{\mathbf{w}_e^{i,\text{H}} \mathbf{e}^{i-1}\right\}. \tag{4.34c}$$

Therefore, the minimizer of the optimization problem in (4.33) is

$$\widehat{\mathbf{e}}^t = \exp\left\{ \text{j}\angle \left(\left(\sum_{i=1}^t \mathbf{p}_e^i \right) / \left[\sum_{i=1}^t \mathbf{p}_e^i \right]_{UM+1} \right) \right\}. \tag{4.35}$$

4.3.3 Algorithm Development

The closed-form solutions for $\mathbf{F}$ in (4.32) and for $\mathbf{e}$ in (4.35) can greatly reduce the computational complexity. Algorithm 4.2 summarizes the proposed SSCA-based robust beamforming design for RIS-aided multiuser mmWave systems in which the BS-user links are subject to random blockages. The proposed algorithm is referred to as the SSCA-OutMin algorithm.

Algorithm 4.2: SSCA-OutMin Algorithm

Require: Initialize $\mathbf{F}^0 \in \mathcal{S}_f$ and $\mathbf{e}^0 \in \mathcal{S}_e$. Set $t = 0$.

1: **repeat**
2: $t = t + 1$.
3: Obtain the sample channel $\mathbf{G}^t$.
4: Calculate $\widehat{\mathbf{F}}^t$ according to (4.32).
5: Update $\mathbf{F}^t = \mathbf{F}^{t-1} + \xi_f^t \left(\widehat{\mathbf{F}}^t - \mathbf{F}^{t-1}\right)$.
6: Calculate $\widehat{\mathbf{e}}^t$ according to (4.35).
7: Update $\mathbf{e}^t = \mathbf{e}^{t-1} + \xi_e^t \left(\widehat{\mathbf{e}}^t - \mathbf{e}^{t-1}\right)$.
8: **until** $||\mathbf{F}^t - \mathbf{F}^{t-1}||_F^2 \to 0$ and $||\mathbf{e}^t - \mathbf{e}^{t-1}||_2 \to 0$.

4.3.3.1 Step-Size Selection

It is worth noting that the approximation in (4.29) has the same form as that in (4.14). However, τ^i in the SSCA method can be any positive number, and $\hat{F}\left(\mathbf{F}, \mathbf{F}^{i-1}|\mathbf{G}\right)$ might not be a global upper bound of $F\left(\mathbf{F}|\mathbf{G}\right)$. To account for this issue, the step sizes ξ_f^t and ξ_e^t in Algorithm 4.2 need to be carefully chosen to ensure convergence.

As an example, let us consider the choice of ξ_f^t to illustrate the update rule, which is a line-search (also called Armijo step-size) rule. Consider $\xi_f^0 > 0$ and $c_{1,f}, c_{2,f} \in (0,1)$. Let ξ_f^t be the largest element in $\{\xi_f^0 c_{2,f}^t\}_{t=0,1,\ldots}$ such that

$$F\left(\mathbf{F}^{t-1} + \xi_f^t \left(\widehat{\mathbf{F}}^t - \mathbf{F}^{t-1}\right)\right) \leq F(\mathbf{F}^{t-1}) + c_{1,f}\xi_f^t \mathrm{Tr}\left(\nabla_{\mathbf{F}} F(\mathbf{F}^{t-1})^{\mathrm{T}} \left(\widehat{\mathbf{F}}^t - \mathbf{F}^{t-1}\right)\right). \tag{4.36}$$

Theorem 7. *If $\{\xi_f^t\}_{t=1,2,\ldots}$ is chosen according to the line-search rule, then*

$$\lim_{n\to\infty} ||\widehat{\mathbf{F}}^t - \mathbf{F}^{t-1}|| = 0.$$

Proof: See Theorem 7 in [95]. ■

4.3.3.2 Convergence Analysis

The convergence of Algorithm 4.2 is given in the following theorem.

Theorem 8. *Assume that Assumptions B and C are satisfied. Then, every limit point of the iterations generated by Algorithm 4.2 is a stationary point of the optimization problem in (4.28) almost surely.*

Proof: Please refer to Appendix B.3. ■

4.3.3.3 Complexity Analysis

The computational complexity for updating $\mathbf{f}^t$ and $\mathbf{e}^t$ at each iteration mainly depends on the computation of (4.32) and (4.35), respectively. In particular, only $\{\mathbf{P}_f^t, \mathbf{p}_e^t\}$ needs to be calculated at the t-th iteration. Therefore, the approximate complexity of each iteration is given by $\mathcal{O}((K+2)2UMN + UMK + NK + (N+2)K^2 + 2N)$.

4.3.3.4 Initial Point

The optimization problem in (4.28) has, in general, multiple local minima due to the non-convex unit-modulus constraint and to $\mathbf{e} \in \mathcal{S}_e$. The accurate selection of the initial points in Algorithm 4.2 plays an important role for the convergence speed and the quality of the obtained local solution. To that end, $\mathbf{e}$ is first initialized to maximize the minimum equivalent total channel gain, resulting in the following optimization problem

$$\mathbf{e}^0 = \arg\max_{\mathbf{e}\in\mathcal{S}_e} \min_{k\in\mathcal{K}} ||\mathbf{e}^{\mathrm{H}}\mathbf{G}_k^0||_2^2. \tag{4.37}$$

The optimization problem in (4.37) can be efficiently solved by using the SDR method as follows

$$\max_{\mathbf{E}} \; t \tag{4.38a}$$

$$\text{s.t.} \; \mathrm{Tr}\{\mathbf{G}_k^0\mathbf{G}_k^{0,\mathrm{H}}\mathbf{E}\} \geq t, \forall k \in \mathcal{K} \tag{4.38b}$$

$$\mathbf{E} \succeq 0, \mathrm{rank}(\mathbf{E}) = 1, [\mathbf{E}]_{m,m} = 1, \forall m, \tag{4.38c}$$

where $\mathbf{E} = \mathbf{e}\mathbf{e}^{\mathrm{H}}$ and t is an auxiliary variable.

Furthermore, $\mathbf{F}$ is initialized by using the maximum-ratio transmission (MRT) solution as

$$\mathbf{F}^0 = P_{max} \frac{\mathbf{G}^0 \mathbf{e}^0}{||\mathbf{G}^0 \mathbf{e}^0||}. \tag{4.39}$$

4.4 Numerical Results and Discussion

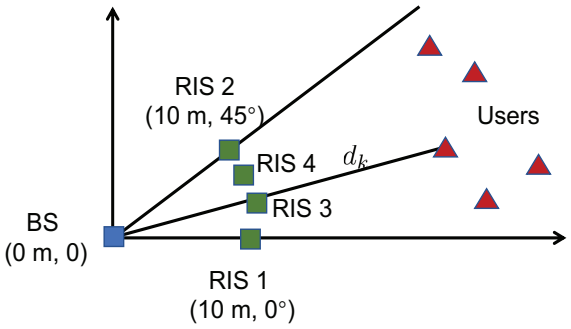


Figure 4.2: The simulated system setup.

4.4.1 Simulation Setup

This section numerically evaluates the performance of the proposed algorithms. All experiments are performed on a PC with a 1.99 GHz i7-8550U CPU and 16 GB RAM. It is adopted the polar coordinates to describe the simulated system setup as shown in Fig. 4.2 [4]. Specifically, the BS is located at (0 m, 0°), and the four RISs are deployed in the locations (10 m, 0°), (10 m, 45°), (10 m, 20°) and (10 m, 30°) which are close to the BS. The users are randomly placed in a region that is identified by the polar diameter $d_k \in$[50 m, 80 m] and the polar angle $\vartheta \in [0, 45°]$, where d_k is used to calculate the distance-dependent blockage probability. The large-scale fading, which corresponds to an Urban Micro (UMi)-street canyon scenario [115], is PL $= 32.4 + 20\log_{10}(f_c) + 10\alpha\log_{10}(D) + \xi$ in dB, where D is the link distance (in meters), α is the path loss exponent, and $\xi \sim \mathcal{CN}(0, \sigma_\xi^2)$ is the log-normal shadowing where σ_ξ^2 denotes the shadowing variance. The mmWave system operates at a carrier frequency $f_c = 28$ GHz and the bandwidth is 20

[4] 2D coordinate is considered here for convenience, the performance of which can be extended to 3D scenario.

MHz. Since the macro-scattering environment between the BS and the users is complex, only NLoS clusters are assumed to exist in the BS-user links, i.e., the Rician factor is $\kappa = 0$. The large-scale parameters of the NLoS links are $\alpha = 3.5$ and $\sigma_\xi = 8.2$ dB [115]. In practice, the RISs can be installed such that the BS-RIS links and the RIS-user links are blockage-free. Thus, the channels in (4.5) and (4.6) are assumed to have only the LoS cluster with a Rician factor $\kappa \to \infty$. The large-scale parameters of the LoS link are $\alpha = 2$ and $\sigma_\xi = 4$ dB according to [115]. Unless stated otherwise, it is assumed $L_{\mathrm{b},k} = L_{u,k} = L_{\mathrm{b},u} = 5$ and $I = 20$. The transmit power limit of the BS is $P_{max} = 30$ dBm and the noise power at each user is $\sigma_1^2 = \ldots = \sigma_K^2 = -94$ dBm. For simplicity, it is assumed an equal blockage probability, $p_{k,l} = p_{\mathrm{block}}, \forall k, l$, and an equal target SINR, $\gamma = \gamma_1 = \ldots = \gamma_K$, which yields the target rate $R_{\mathrm{targ}} = \log_2(1+\gamma)$. The smooth parameters are chosen to be $\theta = \frac{1}{\max_{\forall k \in \mathcal{K}} |x_k^0|}$ and $\mu = \frac{1}{100K}$.

To evaluate the performance of the proposed stochastic optimization algorithms, the following benchmark schemes are considered. 1) Perfect-Instantaneous: Perfect instantaneous CSI is assumed to be known, including the instantaneous channel gains, AoAs, AoDs, and blockage status of the links. This scheme is regarded as the performance upper bound. 2) NoRIS: In this case, no RIS is employed and the precoding at the BS is obtained by using the SMM or SSCA methods. This scheme is regarded as the performance lower bound. 3) No-robust: In this scheme, the beamforming schemes at the BS and at the RIS are designed by using the SMM or SSCA methods by taking into account the random small-scale parameters while assuming $p_{\mathrm{block}} = 0$. 4) Imperfect CSI: In this scheme, the beamforming schemes at the BS and at the RIS are designed by using the SMM or SSCA methods based on the imperfect knowledge of the central angles of the clusters. Specifically, it is assumed that the estimation error of the central angles of the clusters is 0.01 degrees, i.e., $\triangle\mathbb{E}\{\varphi(\phi)\} = 0.01$. 5) Quantization-1/2/3 bit: In this scheme, the optimal continuous-valued phase shifts of the RIS are first obtained by using the SMM algorithm and are then quantified with 1 bit or, 2 bit or 3 bit resolution. 6) SAA: In this scheme, 300 independent channel realizations are generated in advance,

and the solutions at each iteration of the algorithms are obtained as the average over the 300 channel samples. Also, the surrogate function used at each iteration is obtained by adopting the MM or SCA methods. To be specific, let us consider the beamforming design in the single-user case as an example. The beamforming designed by using the SAA-MM method is obtained by modifying the problems in (4.12) and (4.13) with the following updating rules

$$\mathbf{f}^t = \arg\min_{\mathbf{f}\in\mathcal{S}_f} \ \frac{1}{300}\sum_{i=1}^{300} \hat{f}\left(\mathbf{f}, \mathbf{f}^{t-1}|\mathbf{G}^i\right), \tag{4.40}$$

and

$$\mathbf{e}^t = \arg\min_{\mathbf{e}\in\mathcal{S}_e} \ \frac{1}{300}\sum_{i=1}^{300} \hat{f}\left(\mathbf{e}, \mathbf{e}^{t-1}|\mathbf{G}^i\right). \tag{4.41}$$

In order to demonstrate the robustness of the proposed algorithms, this work considers two performance metrics: the outage probability and the effective rate. In particular, the outage probability of each user is calculated by averaging over 1000 independent channel realizations. The corresponding effective rate of the k-th user is defined as $R_{\text{eff},k} \triangleq \mathbb{E}[\log_2(1+\Gamma_k(\mathbf{F},\mathbf{e}))]$ if $\Gamma_k(\mathbf{F},\mathbf{e}) \geq \gamma$ and $R_{\text{eff}} \triangleq 0$ otherwise.

4.4.2 Convergence

Table 4-A: Comparison of the CPU time

Algorithms	The CPU time (sec) per iteration	The CPU time (sec)
SMM	0.0025	1.8750
SSCA	0.0042	4.6719
SAA	0.3557	20.9844

Figure 4.3 illustrates the convergence behavior of the considered stochastic optimization algorithms. For comparison, the single-user case is considered in the presence of RIS 1 depicted in Fig. 4.2, and the other parameters are given in Fig. 4.3. In Fig. 4.3, the coordinate value on the y-axis is the objective value of the problems in (4.16) or (4.31), and it is not the actual outage probability of the original problem. It is observed that the

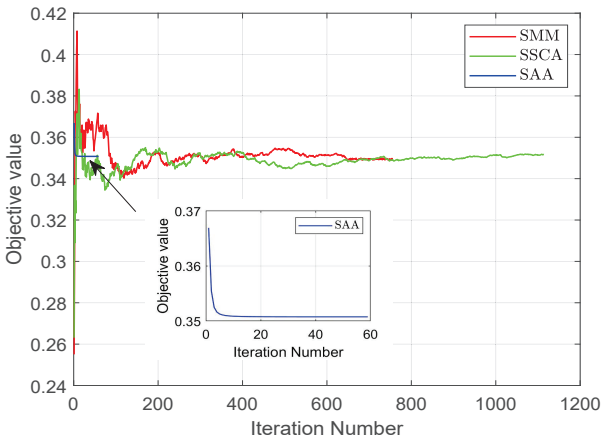
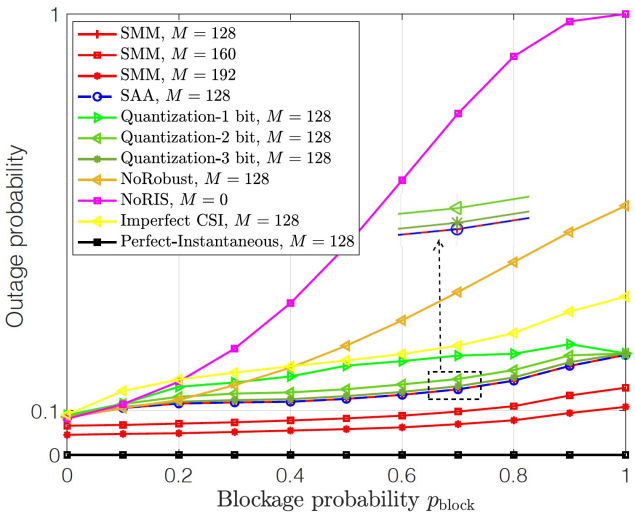


Figure 4.3: Convergence behavior of different algorithms, when $N = 8$, $M = 128$, $K = 1$, $U = 1$, and $R_{\text{targ}} = 0.1$ bps/Hz.

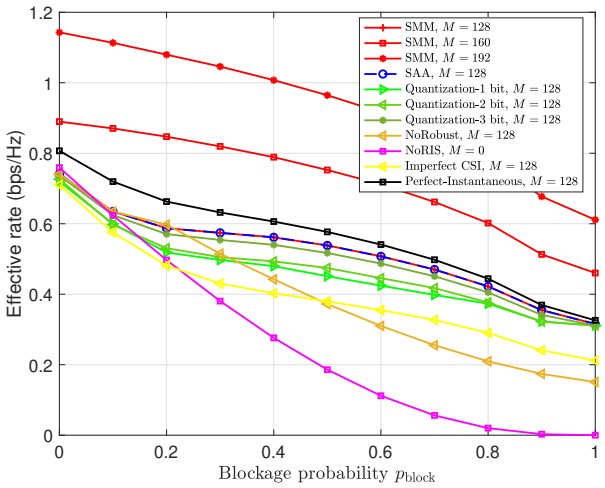
SMM and SSCA algorithms are characterized by an oscillatory convergence behavior, which depends on the random channel generations at each iteration. On the other hand, using 300 channel realizations for each iteration leads to the monotonic convergence behavior of the SAA algorithm when adopting a monotonically decreasing surrogate function for each channel realization. Although the SAA algorithm requires the least number of iterations to converge, it is much more computationally demanding than the other two algorithms. This observation is confirmed in Table 4-A, which compares the CPU time consumption of each iteration and the total CPU time consumption for the three considered algorithms. Theoretically, the computational complexity of each iteration of the SAA algorithm is 300 times higher than that of the SMM or SAA algorithms, because each parameter needs to be calculated 300 times for all channel realizations at each iteration of the SAA algorithm.

4.4.3 Single-User Case Study

This work considers a single-user system where the transmission of data is assisted by the RIS 1 in Fig. 4.2 and the target rate is $R_{\text{targ}} = 0.1$ bps/Hz. Figure 4.4 illustrates the performance of different algorithms as a function of the blockage probability. It is seen that the SMM-based beamforming scheme with $M = 128$ outperforms the NoRIS



(a) Outage probability

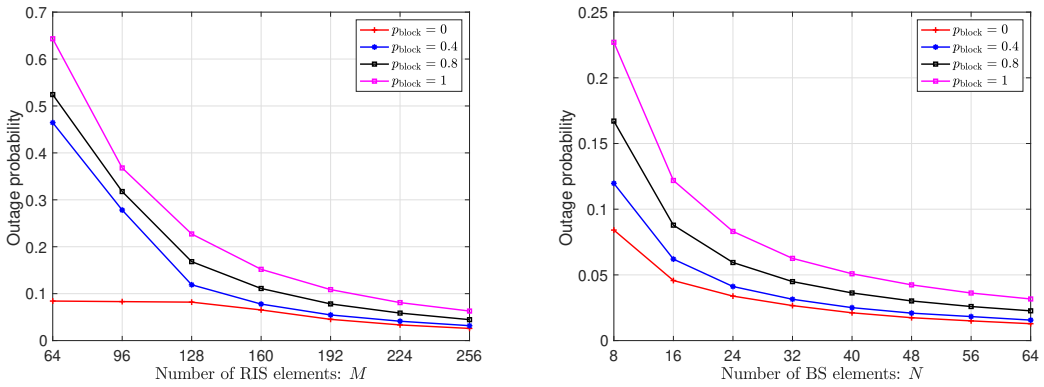


(b) Effective rate

Figure 4.4: Comparison of the outage probability and effective rate as a function of the blockage probability p_{block} for $N = 8$, $K = 1$, $U = 1$, and $R_{\text{targ}} = 0.1$ bps/Hz.

scheme when $p_{\text{block}} \geq 0.1$. If $p_{\text{block}} \leq 0.1$, on the other hand, the direct BS-user channel is much stronger than the cascaded BS-RIS-user channel, as the latter is subject to the double path loss law, which dominates the performance for long transmission distances at high frequency bands. When the blockage probability is small ($p_{\text{block}} \leq 0.1$), therefore, the BS tends to allocate the transmit power to the stronger direct path, thus reducing the contribution of the RIS to the system performance. If the number of reflecting elements at the RIS is increased to $M = 160$, the proposed SMM algorithm outperforms

the NoRIS system for any value of the blockage probability (i.e., $0 \leq p_{\text{block}} \leq 1$). The reason is that the RIS-aided channel provides a beamforming gain that compensates for the performance loss caused by the presence of blockages even if $p_{\text{block}} = 0$. When M is further increased to 192, the outage probability is compressed to about 0.05 during $p_{\text{block}} \in [0, 0.6]$. In addition, it is seen that the SMM-based beamforming scheme and the SAA-based beamforming scheme offer the same performance. As far as the impact of the phase quantization is concerned, it is seen that 1-bit resolution has a non-negligible negative impact on the system performance, while 3-bit resolution is sufficient to obtain performance very close to the continuous-valued phase shifts. As expected, the Perfect-Instantaneous scheme outperforms all the other schemes at the price of frequent channel estimations in each channel coherence block. Finally, the proposed robust designs outperform the NoRobust case and the imperfect CSI case, since the impact of random blockages is accounted for at the design stage.



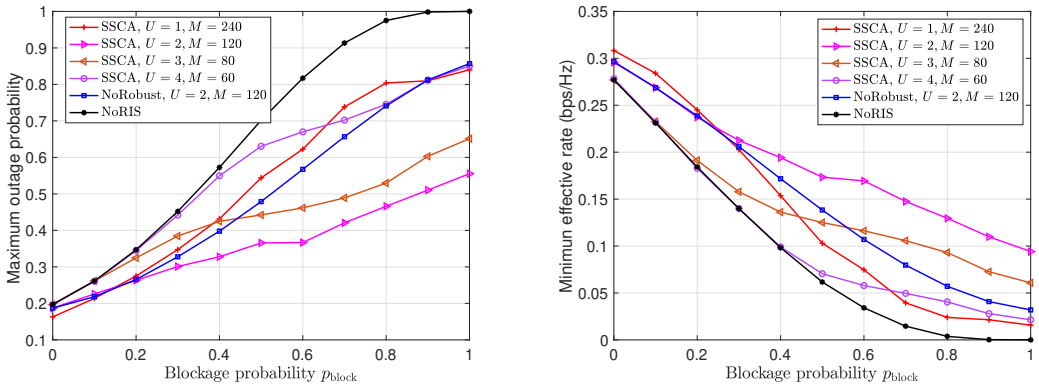
(a) Number of RIS elements (b) Number of BS antennas

Figure 4.5: Outage probability as a function of M with fixed $N = 8$ and as a function of N with fixed $M = 128$, when $K = 1$, $U = 1$, and $R_{\text{targ}} = 0.1$ bps/Hz.

Figure 4.5 shows the impact of the size of the RIS and the size of the antenna array at the BS on the outage probability. The SMM algorithm is considered. It can be observed from Fig. 4.5(a) that, when BS is equipped with $N = 8$ antennas, the RIS plays a significant role in guaranteeing the desired user's QoS and in improving the system robustness as the number of reflecting elements increases ($M : 64 \rightarrow 256$). Specifically, a

large-size RIS with $M \geq 224$ provides an outage probability smaller than 0.1, even if the direct channel from the BS to the user is blocked with unit probability. A similar trend is observed in Fig. 4.5(b) as the number of BS antennas is increased while the number of RIS elements is kept fixed and is equal to $M = 128$. The main difference is that the N antennas at the BS require power amplifiers, digital processing units, and multiple RF chains.

4.4.4 Multiuser Case Study



(a) Outage probability (b) Effective rate

Figure 4.6: Comparison of the maximum outage probability and minimum effective rate as a function of the blockage probability p_{block} for $N = 16$, $K = 3$, and $R_{\text{targ}} = 0.1$ bps/Hz.

This section analyzes a multiuser system with $K = 3$ users, and assume that the target rate is $R_{\text{targ}} = 0.1$ bps/Hz. Multiple RISs are distributed as shown in Fig. 4.2. For fairness, the total number of RIS elements is kept fixed, i.e., $UM = 240$. From Fig. 4.6, it comes to the conclusion that, compared with the single-RIS case, distributing the total number of RIS elements between two RISs significantly improves the system performance in terms of maximum outage probability and minimum effective rate. This is because a better spatial diversity gain is ensured in this case, while ensuring the each RIS has a sufficient number of reflecting elements to compensate for the path loss of the RIS-aided links. If the total number of reflecting elements is distributed among three or four RISs, the system performance is reduced. This is because each RIS cannot

compensate the distance-dependent path loss. In a multi-RIS scenario, therefore, the size of each RIS (i.e., the number of reflecting elements) is an optimization parameter that needs to be judiciously chosen. Furthermore, it is noted that the proposed robust designs significantly outperform the NoRobust and the NoRIS schemes, under similar setups.

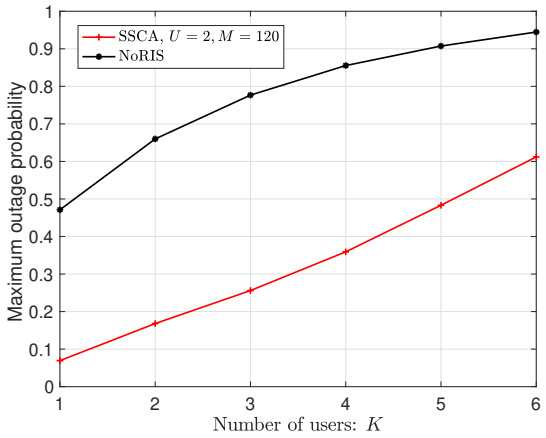


Figure 4.7: Comparison of the maximum outage probability as a function of the number of users K for $N = 16$, $p_{\text{block}} = 0.6$, and $R_{\text{targ}} = 0.1$ bps/Hz.

Finally, Fig. 4.7 illustrates the maximum outage probability as a function of the number of users. For a fair comparison, it is assumed the setup $N = 16, M = 120$ and $U = 2$. It is seen that the gain with respect to the NoRIS scheme is almost constant as the number of users K increases. Therefore, the proposed RIS-aided scheme can guarantee the desired QoS performance for the worst-case user even if the number of users increases.

4.5 Summary

This chapter has introduced schemes for improving the reliability of a mmWave system in the presence of random blockages by deploying multiple RISs and designing the corresponding robust beamforming. In order to reduce the system outage, this chapter has formulated and solved a maximum outage probability minimization problem which belongs to the family of stochastic optimization problems. More precisely, this chapter

has introduced robust beamforming schemes at the RIS that depend on the large-scale CSI and the blockage probability. The proposed schemes are obtained by solving complex stochastic optimization problems, for which closed-form solutions have been proposed at each iteration by leveraging the SMM and SSCA optimization methods. The proposed stochastic methods are proved to converge to the set of stationary points of the original stochastic problems. Selected numerical results have demonstrated the performance gains, in terms of outage probability and effective rate, that the proposed schemes can offer when applied to RIS-aided mmWave systems in the presence of random blockages. However, as shown in Fig. 4.4, the negative impact of imperfect CSI on outage probability is still a challenge. Therefore, in the next chapter, we will investigate robust beamforming designs to address the channel uncertainty caused by channel error.

Chapter 5

Robust Transmission under Channel Estimation Error

This chapter studies the robust transmission design based on the imperfect cascaded CBRUT. Specifically, this work aims to design a robust active and passive beamforming scheme to minimize the total transmit power under both the bounded CSI error model and the statistical CSI error model. Unfortunately, the robust beamforming algorithms developed in [33] and [34] are not applicable for the imperfect CBRUT case. Hence, the contributions of this work are summarized as follows:

- To the best of the knowledge, this is the first work to study the robust transmission design based on imperfect cascaded BS-RIS-user channels, which is more practical than the previous works in which imperfect RIS-user channels were considered. In addition, it considers the robust transmission design under two channel error models: the bounded CSI error model and the statistical CSI error model. However, both [33] and [34] only considered the bounded CSI error model.

- For the bounded CSI error, worst-case robust beamforming design problems are formulated that minimize the transmit power subject to unit modulus of the reflec-

tion beamforming and the worst-case QoS constraints with imperfect CBRUT. The worst-case robust design can guarantee that the achievable rate of each user is no less than its minimum rate requirement for all possible channel error realizations. To address this non-convex problem, S-procedure is firstly adopted to approximate the semi-infinite inequality constraints. Then, under the AO framework, the precoder is updated in an SOCP and the reflection beamforming is updated by using the penalty CCP.

- For the statistical CSI error model, the aim is to minimize the transmit power subject to unit-modulus constraints and the rate outage probability constraints. Here, the rate outage probability constraints represent the probability that the achievable rate of each user being below its minimum rate requirement needs to be less than a predetermined probability. By applying the BTI, the safe approximation of the rate outage probability is obtained to make the original problem tractable. Then, the precoder and the reflection beamforming are optimized by using the SDR and penalty CCP techniques respectively in an iterative manner.

- It is demonstrated through numerical results that the robust beamforming under the statistical CSI error model can achieve superior system performance in terms of the minimum transmit power, convergence speed and complexity, than that under the bounded CSI error model. In addition, it is observed that the level of the CBRUT error plays an important role in the RIS-aided systems. Specifically, when the CBRUT error is small, the total transmit power decreases with the number of the reflection elements due to the increased beamforming gain. However, when the CBRUT error is large, the transmit power increases with the number of the reflection elements due to the increased channel estimation error. Hence, whether to deploy the RIS in wireless communication systems depends on the level of the CBRUT error.

5.1 System Model

This section first introduces the system model of the RIS-aided MISO downlink communication system, and then discusses the channel uncertainty scenarios as well as the CSI error models.

5.1.1 Signal Transmission Model

As shown in Fig. 5.1, an RIS-aided MISO broadcast communication system is considered, which consists of one multi-antenna BS, K single-antenna users and one RIS. It is assumed that the BS is equipped with N active antennas, and transmits K Gaussian data symbols denoted by $\mathbf{s} = [s_1, \cdots, s_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ to all the users, where $\mathbb{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = \mathbf{I}$. RIS with M programmable phase shifters is deployed to enhance the system performance. Therefore, by defining the set of users as $\mathcal{K} = \{1, 2, ..., K\}$, the received baseband signal of users is given by

$$y_k = (\mathbf{h}_k^{\mathrm{H}} + \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}}\mathbf{E}\mathbf{H}_{\mathrm{dr}})\mathbf{F}\mathbf{s} + n_k, \forall k \in \mathcal{K}. \tag{5.1}$$

Here, $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_K] \in \mathbb{C}^{N\times K}$ is the precoder matrix, in which $\mathbf{f}_k$ is the precoding vector associated with user k. Then, the transmit power at the BS is $\mathbb{E}\{\mathrm{Tr}\left(\mathbf{F}\mathbf{s}\mathbf{s}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\right)\} = ||\mathbf{F}||_F^2$. n_k is the AWGN at user k, with zero mean and noise variance σ_k^2, i.e., $n_k \sim \mathcal{CN}(0, \sigma_k^2)$. The reflection beamforming of the RIS is a diagonal matrix $\mathbf{E} = \sqrt{\iota}\mathrm{diag}(e_1, \cdots, e_M) \in$

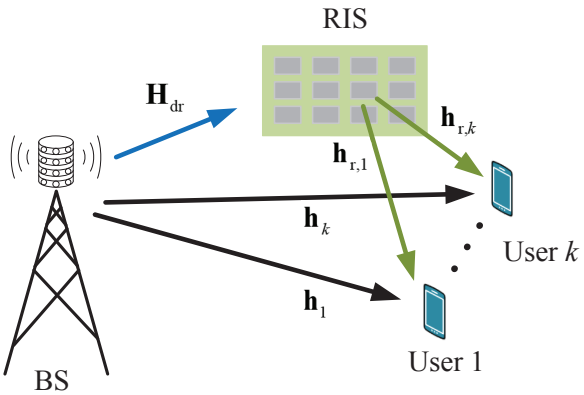


Figure 5.1: An RIS-aided multi-user communication system.

$\mathbb{C}^{M\times M}$, of which has unit-modulus phase shifts, i.e., $|e_m|^2 = 1$. $0 \leq \iota \leq 1$ indicates the reflection efficiency and the power loss of reflection operation usually comes from multiple reflections of signals. Here, it is assumed that only the first-order reflection on the RIS is considered and set $\iota = 1$. It is assumed that the phase shifts of the RIS are calculated by the BS and then fed back to the RIS controller through dedicated feedback channels [26], [27]. In addition, the channel vectors spanning from the BS to user k and from the RIS to user k are denoted by $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ and $\mathbf{h}_{\mathrm{r},k} \in \mathbb{C}^{M\times 1}$, respectively. The channel matrix between the BS and the RIS is represented by $\mathbf{H}_{\mathrm{dr}} \in \mathbb{C}^{M\times N}$.

Denote by $\mathbf{G}_k = \mathrm{diag}(\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}})\mathbf{H}_{\mathrm{dr}}$ the cascaded channel from the BS to user k via the RIS, by $\mathbf{e} = [e_1, \cdots, e_M]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ the vector containing diagonal elements of matrix $\mathbf{E}$, and by $\beta_k = ||(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k)\mathbf{F}_{-k}||_2^2 + \sigma_k^2$ the Interference-plus-Noises (INs) power of user k, where $\mathbf{F}_{-k} = [\mathbf{f}_1, \cdots, \mathbf{f}_{k-1}, \mathbf{f}_{k+1}, \cdots, \mathbf{f}_K]$. Then, the achievable data rate (bit/s/Hz) at user k is given by

$$\mathcal{R}_k(\mathbf{F}, \mathbf{e}) = \log_2\left(1 + \frac{1}{\beta_k}\left|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k\right|^2\right). \tag{5.2}$$

5.1.2 Two Scenarios and CSI Error Models

In the RIS-aided communication system, there are two types of channels: the direct channel $\mathbf{h}_k$, and the cascaded BS-RIS-user channel $\mathbf{G}_k$. The system performance of the RIS-aided communication system is highly affected by the accuracy of the Direct Channel State Information at the Transmitter (DCSIT) and the CBRUT. In the following, two scenarios of the channel uncertainties are first introduced and then two types of CSI error models.

1) Scenario 1: Partial Channel Uncertainty (PCU)

In RIS-aided communications, the CBRUT is much more challenging to obtain than the DCSIT due to the passive features of the RIS. Hence, in this scenario, it is assumed that the DCSIT is perfect, while the CBRUT is imperfect. The CBRUT can be repre-

sented as

$$\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k, \forall k \in \mathcal{K}, \tag{5.3}$$

where $\widehat{\mathbf{G}}_k$ is the estimated cascaded CSI known at the BS, $\triangle\mathbf{G}_k$ is the unknown CBRUT error.

2) Scenario 2: Full Channel Uncertainty (FCU)

In complex electromagnetic environment, the accurate DCSIT is also challenging to obtain. In this scenario, it is assumed both the DCSIT and the CBRUT are imperfect. In addition to the CBRUT error model in (5.3), the direct channel is expressed as

$$\mathbf{h}_k = \widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k, \forall k \in \mathcal{K}, \tag{5.4}$$

where $\widehat{\mathbf{h}}_k$ is the estimated DCSIT known at the BS and $\triangle\mathbf{h}_k$ is the unknown DCSIT error.

This work investigates two types of robust beamforming design for RIS-aided MISO communication systems depending on the CSI error models.

1) Error model 1: Bounded CSI error model

Specifically, one is the worst-case robust beamforming design subject to the bounded CSI error model, i.e.,

$$\|\triangle\mathbf{G}_k\|_F \leq \xi_{\mathrm{g},k}, \|\triangle\mathbf{h}_k\|_2 \leq \xi_{\mathrm{h},k}, \forall k \in \mathcal{K}, \tag{5.5}$$

where $\xi_{\mathrm{g},k}$ and $\xi_{\mathrm{h},k}$ are the radii of the uncertainty regions known at the BS. This CSI error model characterizes the channel quantization error which naturally belongs to a bounded region [121]. For example, in the Frequency Division Duplex (FDD) setting, the receiver estimates the downlink channel and then feeds the rate-limited quantified CSI back to the transmitter. Then, the acquired CSI is plagued by quantization errors.

2) Error model 2: Statistical CSI error model

The other is the outage-constrained robust beamforming design associated with the statistical CSI error model, in which each CSI error vector is assumed to follow the CSCG distribution, i.e.,

$$\mathrm{vec}(\triangle \mathbf{G}_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{g},k}), \mathbf{\Sigma}_{\mathrm{g},k} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.6a}$$

$$\triangle \mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mathrm{h},k}), \mathbf{\Sigma}_{\mathrm{h},k} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.6b}$$

where $\mathbf{\Sigma}_{\mathrm{g},k} \in \mathbb{C}^{MN \times MN}$ and $\mathbf{\Sigma}_{\mathrm{h},k} \in \mathbb{C}^{N \times N}$ are positive semidefinite error covariance matrices. In this case, the CSI imperfection is caused by the channel estimation error [122]. For example, in the Time Division Duplex (TDD) setting, noise and limited training will cause the uplink channel estimation error. The conventional MMSE method is generally adopted to estimate the cascaded channel, and thus the channel estimation generally follows the CSCG distribution.

In the following, the first type of robust beamforming design based on the bounded CSI error model is first considered. Then, the second one based on the statistical CSI error model is addressed.

5.2 Worst-Case Robust Beamforming Design

In this section, the worst-case robust beamforming design is considered under the bounded CSI error model. The aim is to minimize the total transmit power of the BS by the joint design of the precoder matrix $\mathbf{F}$ and reflection beamforming vector $\mathbf{e}$ under the unit-modulus constraints and the worst-case QoS constraints, i.e., ensuring the achievable rate of each user to be above a threshold for all possible channel error realizations. In order to solve the non-convex robust design problem with semi-infinite inequality constraints and coupled variables, an AO algorithm is proposed based on S-Procedure, SOCP and penalty CCP [123].

First, two useful lemmas about multiple complex valued uncertainties are formally introduced as follows, which will be used in the later derivations.

Lemma 4. *(General S-Procedure [124]) Define the quadratic functions of the variable* $\mathbf{x} \in \mathbb{C}^{n\times 1}$*:*

$$f_i(\mathbf{x}) = \mathbf{x}^{\mathrm{H}}\mathbf{W}_i\mathbf{x} + 2\mathrm{Re}\left\{\mathbf{w}_i^{\mathrm{H}}\mathbf{x}\right\} + w_i,\ i = 0, ..., P,$$

where $\mathbf{W}_i = \mathbf{W}_i^{\mathrm{H}}$*. The condition* $\{f_i(\mathbf{x}) \geq 0\}_{i=1}^{P} \Rightarrow f_0(\mathbf{x}) \geq 0$ *holds if and only if there exist* $\forall i, \varpi_i \geq 0$ *such that*

$$\begin{bmatrix} \mathbf{W}_0 & \mathbf{w}_0 \\ \mathbf{w}_0^{\mathrm{H}} & w_0 \end{bmatrix} - \sum_{i=1}^{P} \varpi_i \begin{bmatrix} \mathbf{W}_i & \mathbf{w}_i \\ \mathbf{w}_i^{\mathrm{H}} & w_i \end{bmatrix} \succeq \mathbf{0}.$$

Lemma 5. *(General sign-definiteness [125]) For a given set of matrices* $\mathbf{W} = \mathbf{W}^{\mathrm{H}}$*,* $\{\mathbf{Y}_i, \mathbf{Z}_i\}_{i=1}^{P}$*, the following Linear Matrix Inequality (LMI) satisfies*

$$\mathbf{W} \succeq \sum_{i=1}^{P} \left(\mathbf{Y}_i^{\mathrm{H}}\mathbf{X}_i\mathbf{Z}_i + \mathbf{Z}_i^{\mathrm{H}}\mathbf{X}_i^{\mathrm{H}}\mathbf{Y}_i\right), \forall i, ||\mathbf{X}_i||_F \leq \xi_i,$$

if and only if there exist real numbers $\forall i, \mu_i \geq 0$ *such that*

$$\begin{bmatrix} \mathbf{W} - \sum_{i=1}^{P} \mu_i \mathbf{Z}_i^{\mathrm{H}}\mathbf{Z}_i & -\xi_1\mathbf{Y}_1^{\mathrm{H}} & \cdots & -\xi_P\mathbf{Y}_P^{\mathrm{H}} \\ -\xi_1\mathbf{Y}_1 & \mu_1\mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\xi_P\mathbf{Y}_P & \mathbf{0} & \cdots & \mu_P\mathbf{I} \end{bmatrix} \succeq \mathbf{0}.$$

It is noted that Lemma 5 can be proved by applying Lemma 4 and the detailed proof is given in [126].

5.2.1 Scenario 1: Partial Channel Uncertainty

This subsection designs the robust beamforming for the RIS-aided communication system under Scenario 1 with perfect DCSIT and imperfect CBRUT. This problem is simpler

than the one with full channel uncertainty and the algorithm developed for Scenario 1 has lower complexity than that for Scenario 2. Mathematically, let $\mathcal{E}_k^{partial} \triangleq \{\forall \left\|\triangle \mathbf{G}_k\right\|_F \leq \xi_{\mathrm{g},k}\}$ and denote by $\mathcal{M} = \{1, 2, ..., M\}$ the set of reflection elements, the worst-case transmit power minimization problem is formulated as

$$\min_{\mathbf{F},\mathbf{e}} \ ||\mathbf{F}||_F^2 \tag{5.7a}$$

$$\text{s.t. } \mathcal{R}_k\left(\mathbf{F}, \mathbf{e}\right) \geq R_k, \mathcal{E}_k^{partial}, \forall k \in \mathcal{K} \tag{5.7b}$$

$$|e_m|^2 = 1, \forall m \in \mathcal{M}. \tag{5.7c}$$

Here, R_k is the target rate of user k. Constraints (5.7b) are the worst-case QoS requirements for the users, while constraints (5.7c) correspond to the unit-modulus requirements of the reflection elements at the RIS.

To start with, the non-convexity of constraints (5.7b) can be addressed by firstly treating the INs power $\boldsymbol{\beta} = [\beta_1, ..., \beta_K]^{\mathrm{T}}$ as auxiliary variables. Hence, constraints (5.7b) are reformulated as

$$\left|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k\right|^2 \geq \beta_k(2^{R_k} - 1), \mathcal{E}_k^{partial}, \forall k \in \mathcal{K}, \tag{5.8}$$

$$\left\|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{F}_{-k}\right\|_2^2 + \sigma_k^2 \leq \beta_k, \mathcal{E}_k^{partial}, \forall k \in \mathcal{K}. \tag{5.9}$$

Constraints (5.8) and (5.9) are termed as the worst-case useful signal power constraints and the worst-case INs power constraints, respectively.

Then, the non-convex semi-infinite inequality constraints (5.8) are handled by firstly approximating the non-convex parts and then dealing with the semi-infinite inequalities by using the S-Procedure. Specifically, the following lemma shows the linear approximation of the useful signal power in (5.8).

Lemma 6. *Substituting* $\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k$ *into the useful signal power in (5.8) and let* $\mathbf{f}_k^{(n)}$ *and* $\mathbf{e}^{(n)}$ *be the optimal solutions obtained at iteration* n*, then* $|[\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}(\widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k)]\mathbf{f}_k|^2$

is linearly approximated by its lower bound at $(\mathbf{f}_k^{(n)}, \mathbf{e}^{(n)})$ *as follows*

$$\mathrm{vec}^{\mathrm{T}}(\triangle\mathbf{G}_k)\mathbf{A}_k\mathrm{vec}(\triangle\mathbf{G}_k^*) + 2\mathrm{Re}\left\{\mathbf{a}_k^{\mathrm{T}}\mathrm{vec}(\triangle\mathbf{G}_k^*)\right\} + a_k, \tag{5.10}$$

where

$$\begin{aligned}
\mathbf{A}_k =& \mathbf{f}_k\mathbf{f}_k^{(n),\mathrm{H}} \otimes \mathbf{e}^*\mathbf{e}^{(n),\mathrm{T}} + \mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}} \otimes \mathbf{e}^{(n),*}\mathbf{e}^{\mathrm{T}} - (\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}} \otimes \mathbf{e}^{(n),*}\mathbf{e}^{(n),\mathrm{T}}),\\
\mathbf{a}_k =& \mathrm{vec}(\mathbf{e}\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}) + \mathrm{vec}(\mathbf{e}^{(n)}\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k\mathbf{f}_k^{(n),\mathrm{H}})\\
& - \mathrm{vec}(\mathbf{e}^{(n)}\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}),\\
a_k =& 2\mathrm{Re}\left\{\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}\left(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right)\right\}\\
& - \left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}\left(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}^{(n)}\right).
\end{aligned}$$

Proof: Please refer to Appendix C.1. ■

By replacing the useful signal power in (5.8) with its linear approximation (5.10), constraints (5.8) are reformulated as

$$\mathrm{vec}^{\mathrm{T}}(\triangle\mathbf{G}_k)\mathbf{A}_k\mathrm{vec}(\triangle\mathbf{G}_k^*) + 2\mathrm{Re}\left\{\mathbf{a}_k^{\mathrm{T}}\mathrm{vec}(\triangle\mathbf{G}_k^*)\right\} + a_k \geq \beta_k(2^{R_k} - 1), \mathcal{E}_k^{partial}, \forall k \in \mathcal{K}. \tag{5.11}$$

Lemma 4 is then used to tackle the CSI uncertainty in the above constraints. Specifically, constraint corresponding to each user in (5.11) can be recast by setting the parameters in Lemma 4 as follows

$$\begin{aligned}
& P = 1,\ \mathbf{W}_0 = \mathbf{A}_k,\ \mathbf{w}_0 = \mathbf{a}_k,\ w_0 = a_k - \beta_k(2^{R_k} - 1),\\
& \mathbf{x} = \mathrm{vec}(\triangle\mathbf{G}_k^*), \qquad \mathbf{W}_1 = -\mathbf{I},\ w_1 = \xi_k^2.
\end{aligned}$$

Then, (5.11) is transformed into the following equivalent LMIs as

$$\left[\begin{array}{cc} \varpi_{\mathrm{g},k}\mathbf{I}_{MN}+\mathbf{A}_k & \mathbf{a}_k \\ \mathbf{a}_k^{\mathrm{T}} & C_k^{partial} \end{array}\right] \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.12}$$

where $\boldsymbol{\varpi}_{\mathrm{g}} = [\varpi_{\mathrm{g},1}, ..., \varpi_{\mathrm{g},K}]^{\mathrm{T}} \geq 0$ are slack variables and $C_k^{partial} = a_k - \beta_k(2^{R_k}-1) - \varpi_{\mathrm{g},k}\xi_k^2$.

Next, the uncertainty in $\{\triangle\mathbf{G}_k\}_{\forall k\in\mathcal{K}}$ of (5.9) is considered. Specifically, Schur's complement Lemma [79] is firstly adopted to equivalently recast the INs power inequalities in (5.9) into matrix inequalities as follows

$$\left[\begin{array}{cc} \beta_k-\sigma_k^2 & \mathbf{t}_k^{\mathrm{H}} \\ \mathbf{t}_k & \mathbf{I} \end{array}\right] \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.13}$$

where $\mathbf{t}_k = ((\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k)\mathbf{F}_{-k})^{\mathrm{H}}$. By using $\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k$, (5.13) is then rewritten as

$$\left[\begin{array}{cc} \beta_k-\sigma_k^2 & \widehat{\mathbf{t}}_k^{\mathrm{H}} \\ \widehat{\mathbf{t}}_k & \mathbf{I} \end{array}\right] \succeq -\left[\begin{array}{c} \mathbf{0} \\ \mathbf{F}_{-k}^{\mathrm{H}} \end{array}\right] \triangle\mathbf{G}_k^{\mathrm{H}} \left[\begin{array}{cc} \mathbf{e} & \mathbf{0} \end{array}\right] - \left[\begin{array}{c} \mathbf{e}^{\mathrm{H}} \\ \mathbf{0} \end{array}\right] \triangle\mathbf{G}_k \left[\begin{array}{cc} \mathbf{0} & \mathbf{F}_{-k} \end{array}\right], \forall k \in \mathcal{K}, \tag{5.14}$$

where $\widehat{\mathbf{t}}_k = ((\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{F}_{-k})^{\mathrm{H}}$.

In order to use Lemma 5, the following parameters (It is noted that the subscript i in Lemma 5 has been ignored since $P = 1$.) are chosen for each constraint in (5.14) as

$$\mathbf{W} = \left[\begin{array}{cc} \beta_k-\sigma_k^2 & \widehat{\mathbf{t}}_k^{\mathrm{H}} \\ \widehat{\mathbf{t}}_k & \mathbf{I} \end{array}\right], \mathbf{Y} = -\left[\begin{array}{cc} \mathbf{0} & \mathbf{F}_{-k} \end{array}\right],$$
$$\mathbf{Z} = \left[\begin{array}{cc} \mathbf{e} & \mathbf{0} \end{array}\right], \qquad \mathbf{X} = \triangle\mathbf{G}_k^{\mathrm{H}}.$$

Then, the equivalent LMIs of the worst-case INs power constraints (5.9) are given by

$$\begin{bmatrix} \beta_k - \sigma_k^2 - \mu_{\mathrm{g},k} M & \widehat{\mathbf{t}}_k^{\mathrm{H}} & \mathbf{0}_{1\times N} \\ \widehat{\mathbf{t}}_k & \mathbf{I}_{(K-1)} & \xi_{\mathrm{g},k}\mathbf{F}_{-k}^{\mathrm{H}} \\ \mathbf{0}_{N\times 1} & \xi_{\mathrm{g},k}\mathbf{F}_{-k} & \mu_{\mathrm{g},k}\mathbf{I}_N \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.15}$$

where $\boldsymbol{\mu}_{\mathrm{g}} = [\mu_{\mathrm{g},1}, ..., \mu_{\mathrm{g},K}]^{\mathrm{T}} \geq 0$ are slack variables.

Based on the above discussions, Problem (5.7) is approximately rewritten as

$$\min_{\mathbf{F},\mathbf{e},\boldsymbol{\beta},\boldsymbol{\varpi}_{\mathrm{g}},\boldsymbol{\mu}_{\mathrm{g}}} \ ||\mathbf{F}||_F^2 \tag{5.16a}$$

$$\text{s.t. } (5.12), (5.15), (5.7c), \tag{5.16b}$$

$$\boldsymbol{\varpi}_{\mathrm{g}} \geq 0, \boldsymbol{\mu}_{\mathrm{g}} \geq 0. \tag{5.16c}$$

This problem is still non-convex and difficult to optimize $\mathbf{F}$ and $\mathbf{e}$ simultaneously since $\mathbf{F}$ and $\mathbf{e}$ are coupled in $\mathbf{A}_k$, $\mathbf{a}_k$ and $\widehat{\mathbf{t}}_k$. In the following, the AO method is adopted to optimize $\mathbf{F}$ and $\mathbf{e}$ sequentially in an iterative manner. In particular, the transmit power is minimized by first fixing the reflection beamforming $\mathbf{e}$ so that the problem reduces to a convex one with respect to $\mathbf{F}$. CVX tool [62] is adopted to solve the resulting convex problem. Precoder $\mathbf{F}$ is then fixed and the resulting non-convex problem of $\mathbf{e}$ is handled under the penalty CCP method. Specifically, for given $\mathbf{e}$, the subproblem of $\mathbf{F}$ is given by

$$\mathbf{F}^{(n+1)} = \arg \min_{\mathbf{F},\boldsymbol{\beta},\boldsymbol{\varpi}_{\mathrm{g}},\boldsymbol{\mu}_{\mathrm{g}}} \ ||\mathbf{F}||_F^2 \tag{5.17a}$$

$$\text{s.t. } (5.12), (5.15), (5.16c), \tag{5.17b}$$

where $\mathbf{F}^{(n+1)}$ is the optimal solution obtained in the $(n+1)$-th iteration. Problem (5.17) is a SDP and can be solved by the CVX tool.

Then, for given $\mathbf{F}$, the subproblem of $\mathbf{e}$ is a feasibility-check problem. According to [12], [33] and in order to improve the converged solution in the optimization of $\mathbf{e}$,

the useful signal power inequalities in (5.8) are modified by introducing slack variables $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_K]^{\mathrm{T}} \geq 0$ and recast as

$$\left|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k\right|^2 \geq \beta_k(2^{R_k} - 1) + \alpha_k, \forall k \in \mathcal{K}. \tag{5.18}$$

Subsequently, the LMIs (5.12) are modified as

$$\begin{bmatrix} \varpi_{\mathrm{g},k}\mathbf{I}_{MN} + \mathbf{A}_k & \mathbf{a}_k \\ \mathbf{a}_k^{\mathrm{T}} & C_k^{partial} - \alpha_k \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}. \tag{5.19}$$

In addition, it is noted that only the submatrix of $K \times K$ in the upper left corner of (5.15) depends on $\mathbf{e}$, so the dimension of the LMIs (5.15) can be reduced from $(K+N)\times(K+N)$ to $K \times K$ as

$$\begin{bmatrix} \beta_k - \sigma_k^2 - \mu_{\mathrm{g},k}M & \widehat{\mathbf{t}}_k^{\mathrm{H}} \\ \widehat{\mathbf{t}}_k & \mathbf{I}_{(K-1)} \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}. \tag{5.20}$$

Combining (5.19) and (5.20), the sub-problem of $\mathbf{e}$ can be formulated as

$$\max_{\boldsymbol{\alpha},\mathbf{e},\boldsymbol{\beta},\boldsymbol{\varpi}_{\mathrm{g}},\boldsymbol{\mu}_{\mathrm{g}}} \sum_{k=1}^{K} \alpha_k \tag{5.21a}$$

$$\text{s.t. } (5.19), (5.20), (5.7c), (5.16c), \tag{5.21b}$$

$$\boldsymbol{\alpha} \geq 0. \tag{5.21c}$$

Note that the solution of Problem (5.21) can yield a lower objective value compared with Problem (5.17), the explanation of which can be found in [12].

It is noted that the above problem is still non-convex due to the unit-modulus constraints. As in [33], the penalty CCP [123] is adopted here to deal with the non-convex constraints. Following the penalty CCP framework, the constraints (5.7c) are firstly equivalently rewritten as $1 \leq |e_m|^2 \leq 1, \forall m \in \mathcal{M}$. The non-convex parts of the resulting constraints are then linearized by $|e_m^{[i]}|^2 - 2\mathrm{Re}(e_m^* e_m^{[i]}) \leq -1, \forall m \in \mathcal{M}$, at fixed $e_m^{[i]}$. It

finally comes to the following convex subproblem of $\mathbf{e}$ as

$$\max_{\substack{\mathbf{e},\boldsymbol{\alpha},\mathbf{b},\\ \boldsymbol{\beta},\boldsymbol{\varpi}_{\mathrm{g}},\boldsymbol{\mu}_{\mathrm{g}}}} \sum_{k=1}^{K} \alpha_k - \lambda^{[i]} \sum_{m=1}^{2M} b_m \tag{5.22a}$$

$$\text{s.t. } (5.19),(5.20),(5.16\text{c}),(5.21\text{c}), \tag{5.22b}$$

$$|e_m^{[i]}|^2 - 2\mathrm{Re}(e_m^* e_m^{[i]}) \leq b_m - 1, \forall m \in \mathcal{M} \tag{5.22c}$$

$$|e_m|^2 \leq 1 + b_{M+m}, \forall m \in \mathcal{M} \tag{5.22d}$$

$$\mathbf{b} \geq 0, \tag{5.22e}$$

where $\mathbf{b} = [b_1, ..., b_{2M}]^{\mathrm{T}}$are slack variables imposed over the equivalent linear constraints of the unit-modulus constraints, and $||\mathbf{b}||_1$ is the penalty term in the objective function. $||\mathbf{b}||_1$ is scaled by the regularization factor $\lambda^{[i]}$ to control the feasibility of the constraints.

Problem (5.22) is an SDP and can be solved by the CVX tool. The steps of finding a feasible solution of $\mathbf{e}$ to Problem (5.21) is summarized in Algorithm 5.1. It remarks that: *a)* When χ is sufficiently low, constraints (5.7c) in the original Problem (5.21) is guaranteed by $||\mathbf{b}||_1 \leq \chi$; *b)* The maximum value λ_{max} is imposed to avoid a numerical problem, that is, a feasible solution satisfying $||\mathbf{b}||_1 \leq \chi$ may not be found when the iteration converges to the stopping criteria $||\mathbf{e}^{[i]} - \mathbf{e}^{[i-1]}||_1 \leq \nu$ with the increase of $\lambda^{[i]}$; *c)* Stopping criteria $||\mathbf{e}^{[i]} - \mathbf{e}^{[i-1]}||_1 \leq \nu$ controls the convergence of Algorithm 5.1; *d)* As mentioned in [123], a feasible solution to Problem (5.22) may not be feasible for Problem (5.21). Hence, the feasibility of Problem (5.21) is guaranteed by imposing a maximum number of iterations T_{max} and, in case it is reached, the iteration is restarted based on a new initial point.

Finally, under the AO framework, Problem (5.16) is solved by solving Problems (5.17) and (5.21) in an iterative manner. It is remarked that the fixed point $\mathbf{e}^{[i]}$ in constraint (5.22c) is updated iteratively in Algorithm 5.1, which is the same as $\lambda^{[i]}$. While fixed point $\mathbf{e}^{(n)}$ in constraint (5.19) is updated iteratively in the outer AO framework.

Algorithm 5.1: Penalty CCP optimization for reflection beamforming optimization

Require: Initialize $\mathbf{e}^{[0]}$, $\gamma^{[0]} > 1$, and set $t = 0$.
1: **repeat**
2: **if** $t < T_{max}$ **then**
3: Update $\mathbf{e}^{[i+1]}$ from Problem (5.22);
4: $\lambda^{[i+1]} = \min\{\gamma\lambda^{[i]}, \lambda_{max}\}$;
5: $t = t + 1$;
6: **else**
7: Initialize with a new random $\mathbf{e}^{[0]}$, set up $\gamma^{[0]} > 1$ again, and set $t = 0$.
8: **end if**
9: **until** $||\mathbf{b}||_1 \leq \chi$ and $||\mathbf{e}^{[i]} - \mathbf{e}^{[i-1]}||_1 \leq \nu$.
10: Output $\mathbf{e}^{(t+1)} = \mathbf{e}^{[i]}$.

5.2.2 Scenario 2: Full Channel Uncertainty

This subsection discusses the extension from the robust beamforming design in the previous subsection to the case where both the DCSIT and CBRUT are imperfect. By considering the full channel uncertainty in (5.3) and (5.4) and denoting $\mathcal{E}_k^{full} \triangleq \{\forall||\triangle\mathbf{h}_k||_2 \leq \xi_{\mathrm{h},k}, \forall||\triangle\mathbf{G}_k||_F \leq \xi_{\mathrm{g},k}\}$, constraints (5.7b) can be extended to

$$\mathcal{R}_k(\mathbf{F}, \mathbf{e}) \geq R_k, \mathcal{E}_k^{full}, \forall k \in \mathcal{K}, \tag{5.23}$$

which is then equivalent to

$$\left|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k\right|^2 \geq \beta_k(2^{R_k} - 1), \mathcal{E}_k^{full}, \forall k \in \mathcal{K}, \tag{5.24}$$

$$\left\|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{F}_{-k}\right\|_2^2 + \sigma_k^2 \leq \beta_k, \mathcal{E}_k^{full}, \forall k \in \mathcal{K}. \tag{5.25}$$

The above non-convex semi-infinite inequality constraints can be addressed in the same way as Scenario 1. In particular, the linear approximation of the useful signal power in (5.24) is given in the following lemma.

Lemma 7. *Let $\mathbf{f}_k^{(n)}$ and $\mathbf{e}^{(n)}$ be the optimal solutions obtained at iteration n, and by inserting $\mathbf{h}_k = \widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k$ and $\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k$ into the useful signal power in (5.24), then the resulting $|[(\widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k)^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}(\widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k)]\mathbf{f}_k|^2$ is lower bounded linearly at ($\mathbf{f}_k^{(n)}$,*

$\mathbf{e}^{(n)}$*) as follows*

$$\mathbf{x}_k^{\mathrm{H}}\widetilde{\mathbf{A}}_k\mathbf{x}_k + 2\mathrm{Re}\left\{\widetilde{\mathbf{a}}_k^{\mathrm{H}}\mathbf{x}_k\right\} + \widetilde{a}_k, \tag{5.26}$$

where

$$\begin{aligned}
\widetilde{\mathbf{A}}_k &= \mathbf{D}_k + \mathbf{D}_k^{\mathrm{H}} - \mathbf{Z}_k,\\
\mathbf{D}_k &= \begin{bmatrix} \mathbf{f}_k^{(n)} \\ \mathbf{f}_k^{(n)} \otimes \mathbf{e}^{(\mathrm{n}),*} \end{bmatrix} \begin{bmatrix} \mathbf{f}_k^{\mathrm{H}} & \mathbf{f}_k^{\mathrm{H}} \otimes \mathbf{e}^{\mathrm{T}} \end{bmatrix},\\
\mathbf{Z}_k &= \begin{bmatrix} \mathbf{f}_k^{(n)} \\ \mathbf{f}_k^{(n)} \otimes \mathbf{e}^{(\mathrm{n}),*} \end{bmatrix} \begin{bmatrix} \mathbf{f}_k^{(n),\mathrm{H}} & \mathbf{f}_k^{(n),\mathrm{H}} \otimes \mathbf{e}^{(n),\mathrm{T}} \end{bmatrix},\\
\widetilde{\mathbf{a}}_k &= \mathbf{d}_{1,k} + \mathbf{d}_{2,k} - \mathbf{z}_k,\\
\mathbf{d}_{1,k} &= \begin{bmatrix} \mathbf{f}_k\mathbf{f}_k^{(n),\mathrm{H}}\left(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}^{(n)}\right) \\ \mathrm{vec}^*(\mathbf{e}\left(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}) \end{bmatrix},\\
\mathbf{d}_{2,k} &= \begin{bmatrix} \mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}\left(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right) \\ \mathrm{vec}^*(\mathbf{e}^{(n)}\left(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k\mathbf{f}_k^{(n),\mathrm{H}}) \end{bmatrix},\\
\mathbf{z}_k &= \begin{bmatrix} \mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}\left(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}^{(n)}\right) \\ \mathrm{vec}^*(\mathbf{e}^{(n)}\left(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}) \end{bmatrix},\\
\widetilde{a}_k &= 2\mathrm{Re}\left\{d_k\right\} - z_k,\\
d_k &= \left(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}\left(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right),\\
z_k &= \left(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}\left(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}^{(n)}\right),\\
\mathbf{x}_k &= \begin{bmatrix} \triangle\mathbf{h}_k^{\mathrm{H}} & \mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G}_k^*) \end{bmatrix}^{\mathrm{H}}.
\end{aligned}$$

Proof: Please refer to Appendix C.2. ■

Based on Lemma 7, constraints (5.24) are equivalently rewritten as

$$\mathbf{x}_k^{\mathrm{H}}\widetilde{\mathbf{A}}_k\mathbf{x}_k + 2\mathrm{Re}\left\{\widetilde{\mathbf{a}}_k^{\mathrm{H}}\mathbf{x}_k\right\} + \widetilde{a}_k \geq \beta_k(2^{R_k}-1), \mathcal{E}_k^{full}, \forall k \in \mathcal{K}. \tag{5.27}$$

Before applying Lemma 4, it is beneficial to express $\mathcal{E}_k^{full}$ in terms of the following quadratic expressions as

$$\mathcal{E}_k^{full} \triangleq \begin{cases} \mathbf{x}_k^{\mathrm{H}} \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{x}_k - \xi_{\mathrm{h},k}^2 \leq 0, \\ \mathbf{x}_k^{\mathrm{H}} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{MN} \end{bmatrix} \mathbf{x}_k - \xi_{\mathrm{g},k}^2 \leq 0. \end{cases}$$

Therefore, after introducing $\boldsymbol{\varpi}_{\mathrm{h}} = [\varpi_{\mathrm{h},1}, ..., \varpi_{\mathrm{h},K}]^{\mathrm{T}} \geq 0$ and $\boldsymbol{\varpi}_{\mathrm{g}} = [\varpi_{\mathrm{g},1}, ..., \varpi_{\mathrm{g},K}]^{\mathrm{T}} \geq 0$ as slack variables, constraints (5.24) can be transformed by Lemma 4 into the following equivalent LMIs as

$$\begin{bmatrix} \widetilde{\mathbf{A}}_k + \begin{bmatrix} \varpi_{\mathrm{h},k}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \varpi_{\mathrm{g},k}\mathbf{I}_{MN} \end{bmatrix} & \widetilde{\mathbf{a}}_k \\ \widetilde{\mathbf{a}}_k^{\mathrm{H}} & C_k^{full} \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.28}$$

where $C_k^{full} = \widetilde{a}_k - \beta_k(2^{R_k} - 1) - \varpi_{\mathrm{h},k}\xi_{\mathrm{h},k}^2 - \varpi_{\mathrm{g},k}\xi_{\mathrm{g},k}^2$.

Next, by inserting $\mathbf{h}_k = \widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k$ and $\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k$ into the equivalent matrix inequality of the INs power in (5.13), it has

$$\begin{aligned} \mathbf{0} &\preceq \begin{bmatrix} \beta_k - \sigma_k^2 & \widetilde{\mathbf{t}}_k^{\mathrm{H}} \\ \widetilde{\mathbf{t}}_k & \mathbf{I} \end{bmatrix} + \begin{bmatrix} 0 & \left(\triangle\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\triangle\mathbf{G}_k\right)\mathbf{F}_{-k} \\ \mathbf{F}_{-k}^{\mathrm{H}}\left(\triangle\mathbf{h}_k + \triangle\mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right) & \mathbf{0} \end{bmatrix} \\ &\preceq \begin{bmatrix} \mathbf{0} \\ \mathbf{F}_{-k}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \triangle\mathbf{h}_{\mathrm{r},k} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \triangle\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{F}_{-k} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{F}_{-k}^{\mathrm{H}} \end{bmatrix} \triangle\mathbf{G}_k^{\mathrm{H}} \begin{bmatrix} \mathbf{e} & \mathbf{0} \end{bmatrix} \\ &\quad + \begin{bmatrix} \mathbf{e}^{\mathrm{H}} \\ \mathbf{0} \end{bmatrix} \triangle\mathbf{G}_k \begin{bmatrix} \mathbf{0} & \mathbf{F}_{-k} \end{bmatrix} + \begin{bmatrix} \beta_k - \sigma_k^2 & \widetilde{\mathbf{t}}_k^{\mathrm{H}} \\ \widetilde{\mathbf{t}}_k & \mathbf{I} \end{bmatrix}, \end{aligned} \tag{5.29}$$

where $\widetilde{\mathbf{t}}_k = ((\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{F}_{-k})^{\mathrm{H}}$.

Applying Lemma 5 and defining slack variables $\boldsymbol{\mu}_{\mathrm{g}} = [\mu_{\mathrm{g},1}, ..., \mu_{\mathrm{g},K}]^{\mathrm{T}} \geq 0$ and $\boldsymbol{\mu}_{\mathrm{h}} =$

$[\mu_{\mathrm{h},1}, ..., \mu_{\mathrm{h},K}]^{\mathrm{T}} \geq 0$, the equivalent LMIs of the worst-case INs power constraints (5.25) are given as

$$\begin{bmatrix} Temp_k & \widetilde{\mathbf{t}}_k^{\mathrm{H}} & \mathbf{0}_{1\times N} & \mathbf{0}_{1\times N} \\ \widetilde{\mathbf{t}}_k & \mathbf{I}_{(K-1)} & \xi_{\mathrm{g},k}\mathbf{F}_{-k}^{\mathrm{H}} & \xi_{\mathrm{h},k}\mathbf{F}_{-k}^{\mathrm{H}} \\ \mathbf{0}_{N\times 1} & \xi_{\mathrm{g},k}\mathbf{F}_{-k} & \mu_{\mathrm{g},k}\mathbf{I}_N & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times 1} & \xi_{\mathrm{h},k}\mathbf{F}_{-k} & \mathbf{0}_{N\times N} & \mu_{\mathrm{h},k}\mathbf{I}_N \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}, \tag{5.30}$$

where $Temp_k = \beta_k - \sigma_k^2 - \mu_{\mathrm{g},k}M - \mu_{\mathrm{h},k}$.

With (5.28) and (5.30), the worst-case robust beamforming design problem under full channel uncertainty can be formulated as

$$\min_{\substack{\mathbf{F},\mathbf{e},\boldsymbol{\beta},\boldsymbol{\varpi}_{\mathrm{g}},\\ \boldsymbol{\varpi}_{\mathrm{h}},\boldsymbol{\mu}_{\mathrm{g}}\boldsymbol{\mu}_{\mathrm{h}}}} \quad ||\mathbf{F}||_F^2 \tag{5.31a}$$

$$\text{s.t. } (5.28), (5.30), (5.7c), \tag{5.31b}$$

$$\boldsymbol{\varpi}_{\mathrm{g}} \geq 0, \boldsymbol{\varpi}_{\mathrm{h}} \geq 0, \boldsymbol{\mu}_{\mathrm{g}} \geq 0, \boldsymbol{\mu}_{\mathrm{h}} \geq 0. \tag{5.31c}$$

Problem (5.31) is again non-convex and has coupled variables, which can be solved similarly to Problem (5.16) and thus omitted for simplicity.

5.3 Outage Constrained Robust Beamforming Design

In general, the channel estimation error follows the Gaussian distribution [51]. Hence, it is unbounded. The above bounded channel model may not be able to characterize the practical channel error model. As a result, this section considers the statistical CSI error model. Specifically, by defining the maximum data rate outage probabilities $\rho_1, ..., \rho_K \in (0, 1]$, the transmit power minimization problem is formulated as

$$\min_{\mathbf{F},\mathbf{e}} \quad ||\mathbf{F}||_F^2 \tag{5.32a}$$

$$\text{s.t. } \Pr\{\mathcal{R}_k(\mathbf{F}, \mathbf{e}) \geq R_k\} \geq 1 - \rho_k, \forall k \in \mathcal{K} \tag{5.32b}$$

$$|e_m|^2 = 1, 1 \leq m \leq M. \tag{5.32c}$$

The rate outage constraints (5.32b) guarantee that the probability of each user that can successfully decode its message at a data rate of R_k is no less than $1 - \rho_k$.

The outage constrained robust beamforming design problem in (5.32) is computationally intractable due to the fact that the rate outage probability constraints (5.32b) have no simple closed-form expressions [127]. In order to solve Problem (5.32), a safe approximation based on Bernstein-type inequality is given in the following lemma.

Lemma 8. *(Bernstein-Type Inequality: Lemma 1 in [127]) Assume* $f(\mathbf{x}) = \mathbf{x}^{\mathrm{H}}\mathbf{U}\mathbf{x} + 2\mathrm{Re}\{\mathbf{u}^{\mathrm{H}}\mathbf{x}\} + u$, *where* $\mathbf{U} \in \mathbb{H}^{n\times n}$, $\mathbf{u} \in \mathbb{C}^{n\times 1}$, $u \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{C}^{n\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. *Then for any* $\rho \in [0, 1]$, *the following approximation holds:*

$$\Pr\{\mathbf{x}^{\mathrm{H}}\mathbf{U}\mathbf{x} + 2\mathrm{Re}\{\mathbf{u}^{\mathrm{H}}\mathbf{x}\} + u \geq 0\} \geq 1 - \rho \tag{5.33a}$$

$$\Rightarrow \mathrm{Tr}\,(\mathbf{U}) - \sqrt{2\ln(1/\rho)}x + \ln(\rho)\lambda_{\max}^{+}(-\mathbf{U}) + u \geq 0 \tag{5.33b}$$

$$\Rightarrow \begin{cases} \mathrm{Tr}\,(\mathbf{U}) - \sqrt{2\ln(1/\rho)}x + \ln(\rho)y + u \geq 0 \\ \sqrt{||\mathbf{U}||_F^2 + 2||\mathbf{u}||^2} \leq x \\ y\mathbf{I} + \mathbf{U} \succeq \mathbf{0}, y \geq 0, \end{cases} \tag{5.33c}$$

where $\lambda_{\max}^{+}(-\mathbf{U}) = \max(\lambda_{\max}(-\mathbf{U}), 0)$. x *and* y *are slack variables.*

Please refer to [127] for the proof of Lemma 8.

The following subsections first design the relatively simple robust beamforming under the partial channel uncertainty, and then extend it to the full channel uncertainty case.

5.3.1 Scenario 1: Partial Channel Uncertainty

Before the derivations, the rate outage probability of user k in (5.32b) is rewritten as

$$\Pr\left\{\log_2\left(1 + \frac{\left|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k\right|^2}{\left\|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{F}_{-k}\right\|_2^2 + \sigma_k^2}\right) \geq R_k\right\}$$

$$= \Pr\left\{\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\boldsymbol{\Phi}_k\left(\mathbf{h}_k + \mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right) - \sigma_k^2 \geq 0\right\}, \tag{5.34}$$

where $\boldsymbol{\Phi}_k = \mathbf{f}_k\mathbf{f}_k^{\mathrm{H}}/(2^{R_k}-1) - \mathbf{F}_{-k}\mathbf{F}_{-k}^{\mathrm{H}}$.

For the convenience of derivations, it is assumed that $\boldsymbol{\Sigma}_{\mathrm{g},k} = \varepsilon_{\mathrm{g},k}^2\mathbf{I}$, then the RCSIT error in (5.6) can be rewritten as $\mathrm{vec}(\triangle\mathbf{G}_k) = \varepsilon_{\mathrm{g},k}\mathbf{i}_{\mathrm{g},k}$ where $\mathbf{i}_{\mathrm{g},k} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$. Defining $\mathbf{E} = \mathbf{e}\mathbf{e}^{\mathrm{H}}$, the rate outage probability (5.34) is reformulated in

$$\begin{aligned}
&\Pr\left\{\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}(\widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k)\right)\boldsymbol{\Phi}_k\left(\mathbf{h}_k + (\widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k)^{\mathrm{H}}\mathbf{e}\right) - \sigma_k^2 \geq 0\right\} \\
&= \Pr\Big\{\mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G}_k)(\boldsymbol{\Phi}_k^{\mathrm{T}}\otimes\mathbf{E})\mathrm{vec}(\triangle\mathbf{G}_k) + 2\mathrm{Re}\{\mathrm{vec}^{\mathrm{H}}((\mathbf{e}\mathbf{h}_k^{\mathrm{H}} + \mathbf{E}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k)\mathrm{vec}(\triangle\mathbf{G}_k)\} \\
&\quad + (\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) - \sigma_k^2 \geq 0\Big\} \\
&= \Pr\Big\{\varepsilon_{\mathrm{g},k}^2\mathbf{i}_{\mathrm{g},k}^{\mathrm{H}}(\boldsymbol{\Phi}_k^{\mathrm{T}}\otimes\mathbf{E})\mathbf{i}_{\mathrm{g},k} + 2\mathrm{Re}\{\varepsilon_{\mathrm{g},k}\mathrm{vec}^{\mathrm{H}}((\mathbf{e}\mathbf{h}_k^{\mathrm{H}} + \mathbf{E}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k)\mathbf{i}_{\mathrm{g},k}\} \\
&\quad + (\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) - \sigma_k^2 \geq 0\Big\}.
\end{aligned} \tag{5.35}$$

Therefore, the rate outage constraints (5.32b) are given as

$$\Pr\left\{\mathbf{i}_{\mathrm{g},k}^{\mathrm{H}}\mathbf{U}_k\mathbf{i}_{\mathrm{g},k} + 2\mathrm{Re}\{\mathbf{u}_k^{\mathrm{H}}\mathbf{i}_{\mathrm{g},k}\} + u_k \geq 0\right\} \geq 1-\rho_k, \forall k \in \mathcal{K}, \tag{5.36}$$

where

$$\mathbf{U}_k = \varepsilon_{\mathrm{g},k}^2(\boldsymbol{\Phi}_k^{\mathrm{T}}\otimes\mathbf{E}), \tag{5.37a}$$

$$\mathbf{u}_k = \varepsilon_{\mathrm{g},k}\mathrm{vec}((\mathbf{e}\mathbf{h}_k^{\mathrm{H}} + \mathbf{E}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k^{\mathrm{H}}), \tag{5.37b}$$

$$u_k = (\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) - \sigma_k^2. \tag{5.37c}$$

Applying Lemma 8 and introducing auxiliary variables $\mathbf{x} = [x_1, ..., x_K]^{\mathrm{T}}$ and $\mathbf{y} = [y_1, ..., y_K]^{\mathrm{T}}$, rate outage constraint of user k in (5.36) is transformed into the deterministic form as

$$\mathrm{Tr}\left(\mathbf{U}_k\right) - \sqrt{2\ln(1/\rho_k)}x_k + \ln(\rho_k)y_k + u_k \geq 0, \tag{5.38a}$$

$$\sqrt{||\mathbf{U}_k||_F^2 + 2||\mathbf{u}_k||^2} \leq x_k, \tag{5.38b}$$

$$y_k\mathbf{I} + \mathbf{U}_k \succeq \mathbf{0}, y_k \geq 0. \tag{5.38c}$$

(5.38) can be further simplified by some mathematical transformations as follows

$$\mathrm{Tr}\left(\mathbf{U}_k\right) = \varepsilon_{\mathrm{g},k}^2 \mathrm{Tr}\left(\mathbf{\Phi}_k^{\mathrm{T}} \otimes \mathbf{E}\right) = \varepsilon_{\mathrm{g},k}^2 \mathrm{Tr}\left(\mathbf{\Phi}_k\right) \mathrm{Tr}\left(\mathbf{E}\right) = \varepsilon_{\mathrm{g},k}^2 M \mathrm{Tr}\left(\mathbf{\Phi}_k\right), \tag{5.39a}$$

$$||\mathbf{U}_k||_F^2 = \varepsilon_{\mathrm{g},k}^4 ||(\mathbf{\Phi}_k^{\mathrm{T}} \otimes \mathbf{E})||_F^2 = \varepsilon_{\mathrm{g},k}^4 ||\mathbf{\Phi}_k||_F^2 ||\mathbf{E}||_F^2 = \varepsilon_{\mathrm{g},k}^4 M^2 ||\mathbf{\Phi}_k||_F^2, \tag{5.39b}$$

$$||\mathbf{u}_k||^2 = \varepsilon_{\mathrm{g},k}^2 ||\mathrm{vec}((\mathbf{e}\mathbf{h}_k^{\mathrm{H}} + \mathbf{E}\widehat{\mathbf{G}}_k)\mathbf{\Phi}_k^{\mathrm{H}})||^2 = \varepsilon_{\mathrm{g},k}^2 M || \left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\right) \mathbf{\Phi}_k ||_2^2, \tag{5.39c}$$

$$\lambda(\mathbf{U}_k) = \lambda(\varepsilon_{\mathrm{g},k}^2(\mathbf{\Phi}_k^{\mathrm{T}} \otimes \mathbf{E})) = \varepsilon_{\mathrm{g},k}^2 \lambda(\mathbf{\Phi}_k^{\mathrm{T}} \otimes \mathbf{E}) = \varepsilon_{\mathrm{g},k}^2 \lambda(\mathbf{\Phi}_k)\lambda(\mathbf{E}) = \varepsilon_{\mathrm{g},k}^2 M \lambda(\mathbf{\Phi}_k). \tag{5.39d}$$

Operation $\lambda(\mathbf{X})$ means the eigenvalues of $\mathbf{X}$. (5.39a) and (5.39b) are from [P76 in [120]], (5.39d) is from [P421 in [120]].

Therefore, according to Lemma 8 and equation (5.39), the approximation problem of Problem (5.32) can be given as

$$\min_{\mathbf{F},\mathbf{e},\mathbf{x},\mathbf{y}} \ ||\mathbf{F}||_F^2 \tag{5.40a}$$

$$\text{s.t. } \varepsilon_{\mathrm{g},k}^2 M \mathrm{Tr}\left(\mathbf{\Phi}_k\right) - \sqrt{2\ln(1/\rho_k)} x_k - \ln(1/\rho_k) y_k + u_k \geq 0, \forall k \in \mathcal{K} \tag{5.40b}$$

$$\left\| \begin{array}{c} \varepsilon_{\mathrm{g},k}^2 M \mathrm{vec}(\mathbf{\Phi}_k) \\ \sqrt{2M}\varepsilon_{\mathrm{g},k}\mathbf{\Phi}_k \left(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right) \end{array} \right\| \leq x_k, \forall k \in \mathcal{K} \tag{5.40c}$$

$$y_k\mathbf{I} + \varepsilon_{\mathrm{g},k}^2 M \mathbf{\Phi}_k \succeq \mathbf{0}, y_k \geq 0, \forall k \in \mathcal{K} \tag{5.40d}$$

$$|e_m|^2 = 1, \forall m \in \mathcal{M}. \tag{5.40e}$$

Problem (5.40) is still difficult to solve because constraints (5.40c) are non-convex and have coupled variables $\mathbf{F}$ and $\mathbf{e}$. The AO method is used to update $\mathbf{F}$ and $\mathbf{e}$ in an iterative manner. More specifically, by first fixing $\mathbf{e}$, the non-convex problem in $\mathbf{F}$ at hand is relaxed by employing the SDR technique [19] and solved by CVX. $\mathbf{F}$ is then fixed and the resulting non-convex problem of $\mathbf{e}$ is also handled under the penalty CCP

method.

For fixed $\mathbf{e}$, let $\mathbf{\Phi}_k = \mathbf{\Gamma}_k/(2^{R_k}-1) - \sum_{i=1,i\neq k}^{K}\mathbf{\Gamma}_i$ where $\mathbf{\Gamma}_k = \mathbf{f}_k\mathbf{f}_k^{\mathrm{H}}$, Problem (5.40) corresponding to $\mathbf{F}$ is rewritten as

$$\min_{\mathbf{\Gamma},\mathbf{x},\mathbf{y}} \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{\Gamma}_k\right) \tag{5.41a}$$

$$\text{s.t. } \varepsilon_{\mathrm{g},k}^2 M \mathrm{Tr}\left(\mathbf{\Phi}_k\right) - \sqrt{2\ln(1/\rho_k)}x_k - \ln(1/\rho_k)y_k + u_k \geq 0, \forall k \in \mathcal{K} \tag{5.41b}$$

$$\left\| \begin{array}{c} \varepsilon_{\mathrm{g},k}^2 M \mathrm{vec}(\mathbf{\Phi}_k) \\ \sqrt{2M}\varepsilon_{\mathrm{g},k}\mathbf{\Phi}_k\left(\mathbf{h}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right) \end{array} \right\| \leq x_k, \forall k \in \mathcal{K} \tag{5.41c}$$

$$y_k\mathbf{I} + \varepsilon_{\mathrm{g},k}^2 M\mathbf{\Phi}_k \succeq \mathbf{0}, y_k \geq 0, \forall k \in \mathcal{K} \tag{5.41d}$$

$$\mathbf{\Gamma}_k \succeq \mathbf{0}, \forall k \in \mathcal{K} \tag{5.41e}$$

$$\mathrm{rank}(\mathbf{\Gamma}_k) = 1, \forall k \in \mathcal{K}, \tag{5.41f}$$

where $\mathbf{\Gamma} = [\mathbf{\Gamma}_1, ..., \mathbf{\Gamma}_K]$. Problem (5.41) can be solved by adopting the SDR technique, i.e., removing $\mathrm{rank}(\mathbf{\Gamma}_k) = 1, \forall k \in \mathcal{K}$ from the problem formulation, the resulting convex SDP problem is then efficiently solved by the CVX tools. The following theorem reveals the tightness of the SDR.

Theorem 9. *If the relaxed version of Problem (5.41) is feasible, then there always exists a feasible solution, denoted as* $\mathbf{\Gamma}^\star = [\mathbf{\Gamma}_1^\star, ..., \mathbf{\Gamma}_K^\star]$, *satisfying* $\mathrm{rank}(\mathbf{\Gamma}_k^\star) = 1, \forall k \in \mathcal{K}$.

Proof: Please refer to Appendix C.3. ■

Remark 1: Numerical results show that, the optimal $\mathbf{\Gamma}_k^\star$ is usually of rank one before the construction of the rank-1 solution mentioned in Appendix C.3. The optimal $\mathbf{f}_k$ can be obtained from $\mathbf{\Gamma}_k^\star$ by using eigenvalue decomposition.

This work now considers the subproblem of $\mathbf{e}$ with fixed $\mathbf{F}$. With the same purpose of (5.18), slack variables $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_K]^{\mathrm{T}}$ is introduced to the rate outage probability

in (5.34), which leads to

$$\Pr\left\{\left(\mathbf{h}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\boldsymbol{\Phi}_k\left(\mathbf{h}_k+\mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right)-\sigma_k^2-\alpha_k\geq 0\right\}. \tag{5.42}$$

Then, (5.37c) is also modified as follows

$$u_k^{\mathbf{e}}=(\mathbf{h}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k(\mathbf{h}_k+\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e})-\sigma_k^2-\alpha_k. \tag{5.43}$$

It is noted that (5.43) is non-concave in $\mathbf{e}$ due to the fact that $\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\mathbf{f}_k\mathbf{f}_k^{\mathrm{H}}\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}/(2^{R_k}-1)$ in $\widehat{\mathbf{G}}_k\boldsymbol{\Phi}_k\widehat{\mathbf{G}}_k^{\mathrm{H}}$ is convex. By using the first-order Taylor inequality given in Appendix C.1, $\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\mathbf{f}_k\mathbf{f}_k^{\mathrm{H}}\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}/(2^{R_k}-1)$ can be lower bounded linearly by $u_{\mathrm{linear},k}^{\mathbf{e}}=(2\mathrm{Re}\{\mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\mathbf{f}_k\mathbf{f}_k^{\mathrm{H}}\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\}-\mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k\mathbf{f}_k\mathbf{f}_k^{\mathrm{H}}\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}^{(n)})/(2^{R_k}-1)$. This work then constructs an equivalent concave version of (5.43), which is given as

$$\begin{aligned}u_k^{\mathbf{e}}=&u_{\mathrm{linear},k}^{\mathbf{e}}-\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\mathbf{F}_{-k}\mathbf{F}_{-k}^{\mathrm{H}}\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}+2\mathrm{Re}\{\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\boldsymbol{\Phi}_k\mathbf{h}_k\}+\mathbf{h}_k^{\mathrm{H}}\boldsymbol{\Phi}_k\mathbf{h}_k-\sigma_k^2-\alpha_k\\&+M\mathrm{const}\mathrm{E}_k.\end{aligned} \tag{5.44}$$

In addition, constraints (5.40d) are independent of $\mathbf{e}$ and transformed from $\lambda_{\max}^{+}(-\mathbf{U})$ in Lemma 8, which then can have $y_k=\max(\lambda_{\max}(-\varepsilon_{\mathrm{g},k}^2M\boldsymbol{\Phi}_k),0),\forall k\in\mathcal{K}$. With $\boldsymbol{\alpha}$ and (5.44), the subproblem of (5.40) corresponding to $\mathbf{e}$ is formulated as

$$\max_{\mathbf{e},\boldsymbol{\alpha},\mathbf{x},\mathbf{y}}\ \sum_{k=1}^{K}\alpha_k \tag{5.45a}$$

$$\text{s.t. }\ \varepsilon_{\mathrm{g},k}^2M\mathrm{Tr}\left(\boldsymbol{\Phi}_k\right)-\sqrt{2\ln(1/\rho_k)}x_k-\ln(1/\rho_k)y_k+u_k^{\mathbf{e}}\geq 0,\forall k\in\mathcal{K} \tag{5.45b}$$

$$\left\|\begin{array}{c}\varepsilon_{\mathrm{g},k}^2M||\boldsymbol{\Phi}_k||_F\\ \sqrt{2M}\varepsilon_{\mathrm{g},k}\boldsymbol{\Phi}_k\left(\mathbf{h}_k+\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right)\end{array}\right\|\leq x_k,\forall k\in\mathcal{K} \tag{5.45c}$$

$$\boldsymbol{\alpha}\geq 0 \tag{5.45d}$$

$$|e_m|^2=1,\forall m\in\mathcal{M}. \tag{5.45e}$$

The non-convex constraints (5.45e) in Problem (5.45) is solved by using the same tech-

niques as those used for solving Problem (5.21), then the resulting approximation problem for Problem (5.45) can be solved by using Algorithm 5.1.

5.3.2 Scenario 2: Full Channel Uncertainty

This subsection introduces the extension of the outage constrained robust beamforming design from the partial channel uncertainty to the case where all the channels are imperfect at the BS. By considering the full statistical CSI error in (5.6), (5.34) is then formulated as

$$\Pr\left\{\left(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\right)\boldsymbol{\Phi}_k\left(\widehat{\mathbf{h}}_k+\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right)+2\mathrm{Re}\left\{\left(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k\right)\boldsymbol{\Phi}_k\left(\triangle\mathbf{h}_k+\triangle\mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right)\right\}-\sigma_k^2 + \left(\triangle\mathbf{h}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\triangle\,\mathbf{G}_k\right)\boldsymbol{\Phi}_k\left(\triangle\mathbf{h}_k+\triangle\mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right)\geq 0\right\}. \quad (5.46)$$

Assuming that $\boldsymbol{\Sigma}_{\mathrm{h},k}=\varepsilon_{\mathrm{h},k}^2\mathbf{I}$, then the DCSIT can be expressed as $\triangle\mathbf{h}_k=\varepsilon_{\mathrm{h},k}\mathbf{i}_{\mathrm{h},k}$ where $\mathbf{i}_{\mathrm{h},k}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$. The second term inside (5.46) is rewritten as

$$\begin{aligned}
&2\mathrm{Re}\left\{(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k\triangle\,\mathbf{h}_k+\mathrm{vec}^{\mathrm{T}}(\mathbf{e}(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k)\mathrm{vec}(\triangle\mathbf{G}_k^*)\right\}\\
&=2\mathrm{Re}\left\{\varepsilon_{\mathrm{h},k}(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k\mathbf{i}_{\mathrm{h},k}+\varepsilon_{\mathrm{g},k}\mathrm{vec}^{\mathrm{T}}(\mathbf{e}(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k)\mathbf{i}_{\mathrm{g},k}^*\right\}\\
&=2\mathrm{Re}\left\{\widetilde{\mathbf{u}}_k^{\mathrm{H}}\widetilde{\mathbf{i}}_k\right\},
\end{aligned}$$

where $\widetilde{\mathbf{i}}_k=\left[\begin{array}{cc}\mathbf{i}_{\mathrm{h},k}^{\mathrm{H}} & \mathbf{i}_{\mathrm{g},k}^{\mathrm{T}}\end{array}\right]^{\mathrm{H}}$ and

$$\widetilde{\mathbf{u}}_k=\left[\begin{array}{c}\varepsilon_{\mathrm{h},k}\boldsymbol{\Phi}_k(\widehat{\mathbf{h}}_k+\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e})\\ \varepsilon_{\mathrm{g},k}\mathrm{vec}^*(\mathbf{e}(\widehat{\mathbf{h}}_k^{\mathrm{H}}+\mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k)\end{array}\right].$$

The fourth term on the left hand side of (5.46) is rewritten as

$$\begin{aligned}
&\triangle\mathbf{h}_k^{\mathrm{H}}\boldsymbol{\Phi}_k\triangle\,\mathbf{h}_k+2\mathrm{Re}\left\{\mathbf{e}^{\mathrm{H}}\triangle\,\mathbf{G}_k\boldsymbol{\Phi}_k\triangle\,\mathbf{h}_k\right\}+\mathbf{e}^{\mathrm{H}}\triangle\,\mathbf{G}_k\boldsymbol{\Phi}_k\triangle\mathbf{G}_k^{\mathrm{H}}\mathbf{e}\\
&=\varepsilon_{\mathrm{h},k}^2\mathbf{i}_{\mathrm{h},k}^{\mathrm{H}}\boldsymbol{\Phi}_k\mathbf{i}_{\mathrm{h},k}+2\mathrm{Re}\left\{\triangle\mathbf{h}_k^{\mathrm{H}}(\boldsymbol{\Phi}_k\otimes\mathbf{e}^{\mathrm{T}})\mathrm{vec}(\triangle\mathbf{G}_k^*)\right\}+\mathrm{vec}^{\mathrm{T}}(\triangle\mathbf{G}_k)(\boldsymbol{\Phi}_k\otimes\mathbf{E}^{\mathrm{T}})\mathrm{vec}(\triangle\mathbf{G}_k^*)\\
&=\varepsilon_{\mathrm{h},k}^2\mathbf{i}_{\mathrm{h},k}^{\mathrm{H}}\boldsymbol{\Phi}_k\mathbf{i}_{\mathrm{h},k}+2\mathrm{Re}\left\{\varepsilon_{\mathrm{h},k}\varepsilon_{\mathrm{g},k}\mathbf{i}_{\mathrm{h},k}^{\mathrm{H}}(\boldsymbol{\Phi}_k\otimes\mathbf{e}^{\mathrm{T}})\mathbf{i}_{\mathrm{g},k}^*\right\}+\varepsilon_{\mathrm{g},k}^2\mathbf{i}_{\mathrm{g},k}^{\mathrm{T}}(\boldsymbol{\Phi}_k\otimes\mathbf{E}^{\mathrm{T}})\mathbf{i}_{\mathrm{g},k}^*
\end{aligned}$$

$$= \widetilde{\mathbf{i}}_k^{\mathrm{H}} \widetilde{\mathbf{U}}_k \widetilde{\mathbf{i}}_k,$$

where

$$\widetilde{\mathbf{U}}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{h},k}^{1/2} \boldsymbol{\Phi}_k \boldsymbol{\Sigma}_{\mathrm{h},k}^{1/2} & \varepsilon_{\mathrm{h},k}\varepsilon_{\mathrm{g},k}(\boldsymbol{\Phi}_k \otimes \mathbf{e}^{\mathrm{T}}) \\ \varepsilon_{\mathrm{h},k}\varepsilon_{\mathrm{g},k}(\boldsymbol{\Phi}_k \otimes \mathbf{e}^{*}) & \varepsilon_{\mathrm{g},k}^2(\boldsymbol{\Phi}_k \otimes \mathbf{E}^{\mathrm{T}}) \end{bmatrix}.$$

Denote $\widetilde{u}_k = (\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) - \sigma_k^2$, the rate outage constraint (5.46) is then equivalent to

$$\Pr\left\{\widetilde{\mathbf{i}}_k^{\mathrm{H}}\widetilde{\mathbf{U}}_k\widetilde{\mathbf{i}}_k + 2\mathrm{Re}\left\{\widetilde{\mathbf{u}}_k^{\mathrm{H}}\widetilde{\mathbf{i}}_k\right\} + \widetilde{u}_k \geq 0\right\} \geq 1 - \rho_k. \tag{5.47}$$

Combining Lemma 8 and new auxiliary variables $\widetilde{\mathbf{x}} = [\widetilde{x}_1, ..., \widetilde{x}_K]^{\mathrm{T}}$ and $\widetilde{\mathbf{y}} = [\widetilde{y}_1, ..., \widetilde{y}_K]^{\mathrm{T}}$, the approximation of the data rate outage constraint of user k in (5.47) is given by

$$\mathrm{Tr}\left(\widetilde{\mathbf{U}}_k\right) - \sqrt{2\ln(1/\rho_k)}\widetilde{x}_k + \ln(\rho_k)\widetilde{y}_k + \widetilde{u}_k \geq 0, \tag{5.48a}$$

$$\sqrt{||\widetilde{\mathbf{U}}_k||_F^2 + 2||\widetilde{\mathbf{u}}_k||^2} \leq \widetilde{x}_k, \tag{5.48b}$$

$$\widetilde{y}_k\mathbf{I} + \widetilde{\mathbf{U}}_k \succeq \mathbf{0}, \widetilde{y}_k \geq 0. \tag{5.48c}$$

Some terms in (5.48) are simplified as follows:

$$\mathrm{Tr}\left(\widetilde{\mathbf{U}}_k\right) = \mathrm{Tr}\left(\begin{bmatrix} \varepsilon_{\mathrm{h},k}\boldsymbol{\Phi}_k^{1/2} \\ \varepsilon_{\mathrm{g},k}(\boldsymbol{\Phi}_k^{1/2} \otimes \mathbf{e}^{*}) \end{bmatrix} \begin{bmatrix} \varepsilon_{\mathrm{h},k}\boldsymbol{\Phi}_k^{1/2} & \varepsilon_{\mathrm{g},k}(\boldsymbol{\Phi}_k^{1/2} \otimes \mathbf{e}^{\mathrm{T}}) \end{bmatrix}\right)$$

$$= (\varepsilon_{\mathrm{h},k}^2 + \varepsilon_{\mathrm{g},k}^2 M)\mathrm{Tr}\left(\boldsymbol{\Phi}_k\right), \tag{5.49a}$$

$$||\widetilde{\mathbf{U}}_k||_F^2 = (\varepsilon_{\mathrm{h},k}^2 + \varepsilon_{\mathrm{g},k}^2 M)^2||\boldsymbol{\Phi}_k||_F^2, \tag{5.49b}$$

$$||\widetilde{\mathbf{u}}_k||^2 = (\varepsilon_{\mathrm{h},k}^2 + \varepsilon_{\mathrm{g},k}^2 M)||(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k)\boldsymbol{\Phi}_k||_2^2, \tag{5.49c}$$

$$\widetilde{y}_k\mathbf{I} + \widetilde{\mathbf{U}}_k \succeq \mathbf{0} \Longrightarrow \widetilde{y}_k\mathbf{I} + (\varepsilon_{\mathrm{h},k}^2 + \varepsilon_{\mathrm{g},k}^2 M)\boldsymbol{\Phi}_k \succeq \mathbf{0}. \tag{5.49d}$$

The derivations of (5.49) are similar to (5.39).

Based on the above results, Problem (5.32) with imperfect DCSIT and imperfect CBRUT is given by

$$\min_{\mathbf{F},\mathbf{e},\widetilde{\mathbf{x}},\widetilde{\mathbf{y}}} \ ||\mathbf{F}||_F^2 \tag{5.50a}$$

$$\text{s.t.} \ (\varepsilon_{\mathrm{h},k}^2+\varepsilon_{\mathrm{g},k}^2 M)\mathrm{Tr}\left(\mathbf{\Phi}_k\right)-\sqrt{2\ln(1/\rho_k)}x_k-\ln(1/\rho_k)y_k+\widetilde{u}_k\geq 0,\forall k\in\mathcal{K} \tag{5.50b}$$

$$\left\| \begin{array}{c} (\varepsilon_{\mathrm{h},k}^2+\varepsilon_{\mathrm{g},k}^2 M)\mathrm{vec}(\mathbf{\Phi}_k) \\ \sqrt{2(\varepsilon_{\mathrm{h},k}^2+\varepsilon_{\mathrm{g},k}^2 M)}\mathbf{\Phi}_k\left(\widehat{\mathbf{h}}_k+\widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}\right) \end{array} \right\| \leq \widetilde{x}_k,\forall k\in\mathcal{K} \tag{5.50c}$$

$$\widetilde{y}_k\mathbf{I}+(\varepsilon_{\mathrm{h},k}^2+\varepsilon_{\mathrm{g},k}^2 M)\mathbf{\Phi}_k\succeq\mathbf{0},\widetilde{y}_k\geq 0,\forall k\in\mathcal{K} \tag{5.50d}$$

$$|e_m|^2=1,\forall m\in\mathcal{M}. \tag{5.50e}$$

Comparing Problem (5.50) with Problem (5.40), it is found that the former can be obtained from the latter by replacing $\varepsilon_{\mathrm{g},k}^2 M$ with $\varepsilon_{\mathrm{h},k}^2+\varepsilon_{\mathrm{g},k}^2 M$ and $\mathbf{h}_k$ with $\widehat{\mathbf{h}}_k$. Therefore, Problem (5.50) can be solved by using the same techniques as those used to solve Problem (5.40). In addition, when M is large, the impact of imperfect CBRUT dominates the performance of the system, which will be illustrated in the numerical results later. Thus, it is significant to investigate the robust beamforming in an RIS-aided system in which there are a large number of reflection elements with high channel estimation error.

5.4 Computational Complexity

This section analyzes the computational complexity of the proposed robust transmission design methods. Since all the resulting convex problems involving LMI, Second-Order Cone (SOC) constraints and linear constraints that can be solved by a standard interior point method [101], it can compare the computational complexity of different methods in terms of their worst-case runtime, the general expression (the complexity of the linear

constraints are ignored) of which is given by

$$\mathcal{O}((\sum_{j=1}^{J} b_j + 2I)^{1/2} n(n^2 + \underbrace{n\sum_{j=1}^{J} b_j^2 + \sum_{j=1}^{J} b^3}_{\text{due to LMI}} + \underbrace{n\sum_{i=1}^{I} a_i^2}_{\text{due to SOC}})),$$

where n is the number of variables, J is the number of LMIs of size b_j, and I is the number of SOC of size a_i. Based on the above expression, the computational complexity per iteration of the proposed methods is provided as follows:

1) *PCU-bounded method* denotes the worst-case beamforming design method under Scenario 1. The approximate complexity of Problem (5.17) is $o_{\mathbf{F}} = \mathcal{O}([K(MN + K + N + 1)]^{1/2} n_1 [n_1^2 + n_1 K((MN+1)^2 + (K+N)^2) + K((MN+1)^3 + (K+N)^3)])$ where $n_1 = NK$, and that of Problem (5.22) is $o_{\mathbf{e}} = \mathcal{O}([K(MN+1+K) + 2M]^{1/2} n_2 [n_2^2 + n_2 K((MN+1)^2 + K^2) + K((MN+1)^3 + K^3) + n_2 M])$ where $n_2 = M$. Finally, the approximate complexity of PCU-bounded method per iteration is $o_{\mathbf{F}} + o_{\mathbf{e}}$.

2) *FCU-bounded method* denotes the worst-case beamforming design method under Scenario 2. The approximate complexity of Problem (5.31) is $o_{\mathbf{F}} + o_{\mathbf{e}}$, where $o_{\mathbf{F}} = \mathcal{O}([K(MN + 3N + K + 1)]^{1/2} n_1 [n_1^2 + n_1 K((MN+N+1)^2 + (K+2N)^2) + K((MN + N + 1)^3 + (K+2N)^2)])$ with $n_1 = NK$, and $o_{\mathbf{e}} = \mathcal{O}([K(MN+1+K) + 2M]^{1/2} n_2 [n_2^2 + n_2 K((MN+1)^2 + K^2) + K((MN+1)^3 + K^3) + n_2 M])$ with $n_2 = M$.

3) *PCU-statistic method* denotes the outage constrained beamforming design method under Scenario 1. The approximate complexity of Problem (5.41) is $o_{\mathbf{F}} = \mathcal{O}([2K(N + 1)]^{1/2} n_1 [n_1^2 + 2n_1 KN^2 + 2KN^3 + nKN^2(N+1)^2])$ where $n_1 = NK$, and that of Problem (5.45) is $o_{\mathbf{e}} = \mathcal{O}([4K + 2M]^{1/2} n_2 [n_2^2 + n_2(K(M^2 + (N+1)^2) + M)])$ where $n_2 = M$. Finally, the approximate complexity of PCU-statistic method per iteration is $o_{\mathbf{F}} + o_{\mathbf{e}}$.

4) *FCU-statistic method* denotes the outage constrained beamforming design method under Scenario 2. Here, the approximate complexity per iteration is the same with the PCU-statistic method since they only have some different coefficients.

5.5 Numerical Results and Discussions

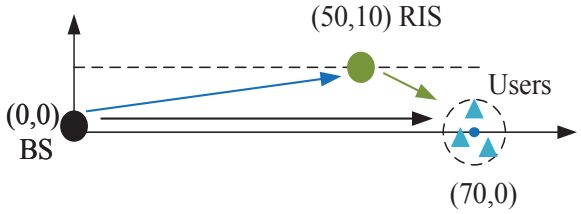


Figure 5.2: The simulated system setup.

Table 5-A: System parameters

Path loss exponents of BS-user link	$\alpha_{\mathrm{BU}} = 4$
Path loss exponents of BS-RIS link	$\alpha_{\mathrm{BI}} = 2.2$
Path loss exponents of RIS-user link	$\alpha_{\mathrm{IU}} = 2$
Noise power	$\sigma_1^2 = ... = \sigma_K^2 = -80$ dBm
Convergence tolerance	10^{-4}
Maximun outage probabilities	$\rho_1 = ... = \rho_K = \rho = 0.05$

This section provides numerical results to evaluate the performance of the proposed algorithms. The simulated system setup of the considered network is shown in Fig. 5.2 [1], in which it is assumed that the BS is located at (0 m, 0 m) and the RIS is placed at (50 m, 10 m). K users are randomly and uniformly distributed in a circle centered at (70 m, 0 m) with radius of 5 m. The channel models, i.e., $\{\mathbf{h}_k, \mathbf{G}_k\}_{\forall k \in \mathcal{K}}$, are assumed to include large-scale fading and small-scale fading. The large-scale fading model is expressed as $\mathrm{PL} = -\mathrm{PL}_0 - 10\alpha \log_{10}(d)$ dB, where α is the path loss exponent and d is the link distance in meters. PL_0 denotes the pathloss at the distance of 1 meter, which is set as 40 dB based on the 3GPP UMi model [128] with 3.5 GHz carrier frequency (i.e., carrier frequency of 5G in China). The small-scale fading in $\{\mathbf{h}_k, \mathbf{G}_k\}_{\forall k \in \mathcal{K}}$ is assumed to be Rayleigh fading distribution. For the statistical CSI error model, the variance of $\mathrm{vec}(\triangle \mathbf{G}_k)$ and $\triangle \mathbf{h}_k$ are defined as $\varepsilon_{\mathrm{g},k}^2 = \delta_{\mathrm{g}}^2 ||\mathrm{vec}(\widehat{\mathbf{G}}_k)||_2^2$ and $\varepsilon_{\mathrm{h},k}^2 = \delta_{\mathrm{h}}^2 ||\widehat{\mathbf{h}}_k||_2^2$, respectively. $\delta_{\mathrm{g}} \in [0, 1)$ and $\delta_{\mathrm{h}} \in [0, 1)$ measure the relative amount of CSI uncertainties. For the bounded CSI error model, the radii of the uncertainty regions are set as

$$\xi_{\mathrm{g},k} = \sqrt{\frac{\varepsilon_{\mathrm{g},k}^2}{2} F_{2MN}^{-1}(1-\rho)},$$

[1] 2D scenario can be extended to 3D scenario.

and

$$\xi_{\mathrm{h},k} = \sqrt{\frac{\varepsilon_{\mathrm{h},k}^2}{2} F_{2N}^{-1}(1-\rho)},$$

where $F_{2MN}^{-1}(\cdot)$ and $F_{2N}^{-1}(\cdot)$ denote the inverse Cumulative Distribution Function (CDF) of the Chi-square distribution with degrees of freedom equal to $2MN$ and $2N$, respectively. According to [127], the above bounded CSI error model provides a fair comparison between the performance of the worst-case robust design and the outage constrained robust design. In addition, the target rates of all users are assumed to be the same, i.e., $R_1 = ... = R_K = R$ and the fixed simulation settings for the simulations are given in Table 5-A.

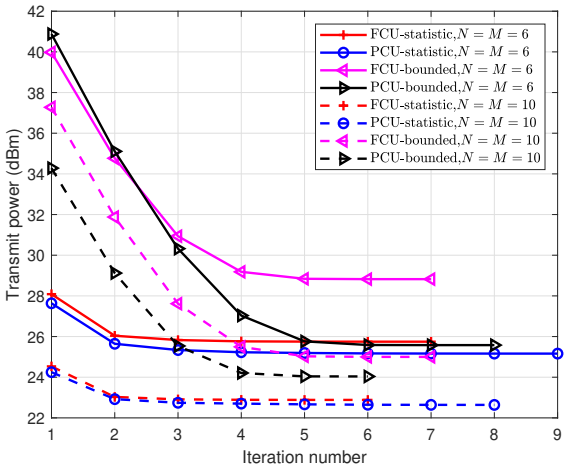


Figure 5.3: Transmit power versus the number of iteration of different algorithms, when $K = 3$ and $\{\delta_{\mathrm{g}}, \delta_{\mathrm{h}}\} = \{0.01, 0.02\}$.

Fig. 5.3 illustrates the convergence behavior of the proposed four algorithms. Here, the minimum rate is set as $R = 2$ bit/s/Hz, and the channel uncertainty levels are chosen as $\{\delta_{\mathrm{g}}, \delta_{\mathrm{h}}\} = \{0.01, 0.02\}$. It is observed that all algorithms converge rapidly and 10 iterations are sufficient for the algorithms to converge. It also shows that the convergence speed increases with the number of antennas. In addition, the algorithms under the statistical error model converge faster than those under the bounded error model.

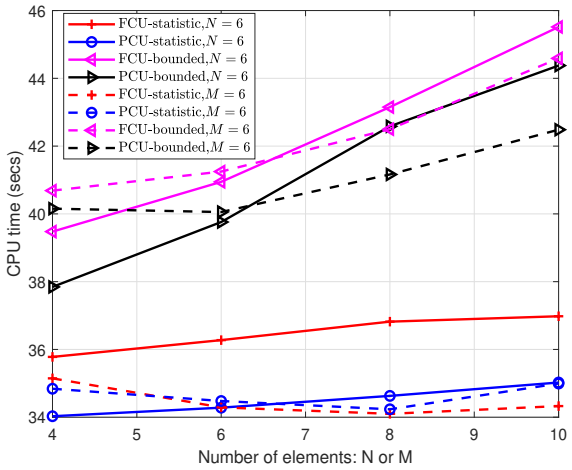


Figure 5.4: Average CPU time versus the number of antenna elements at the RIS M and at the BS N, when $K = 2$ and $\{\delta_{\rm g}, \delta_{\rm h}\} = \{0.01, 0.02\}$.

Fig. 5.4 compares the average CPU running time of the proposed algorithms versus the numbers of antenna elements at the BS and/or reflection elements at the RIS. The results are obtained by using a computer with a 1.99 GHz i7-8550U CPU and 16 GB RAM. Here, it is set $K = 2$, $R = 2$ bit/s/Hz, and $\{\delta_{\rm g}, \delta_{\rm h}\} = \{0.01, 0.02\}$. Firstly, it is observed that the robust algorithms under the statistical CSI error model require much less CPU running time than those under the bounded CSI error model. This is due to the fact that there are some large-dimensional LMIs that increase the computational complexity of the worst-case algorithms. Secondly, the FCU-bounded algorithm requires more CPU time than the PCU-bounded algorithm because the DCSIT error $\triangle \mathbf{h}_k$ increases the dimension of the LMIs. Thirdly, when $M = 6$, the CPU running time of the outage constrained algorithm under two scenarios is similar due to the fact that no additional complexity is introduced by considering the additional DCSIT error. Finally, under bounded CSI error model, the gap of CPU time between FCU and PCU cases (see solid pink and black curves) reduces when M becomes large due to the fact that the role of M becomes weak in the complexity, While larger N has more impact in the complexity such that the gap between FCU and PCU cases (see dash pink and black curves) increases with N.

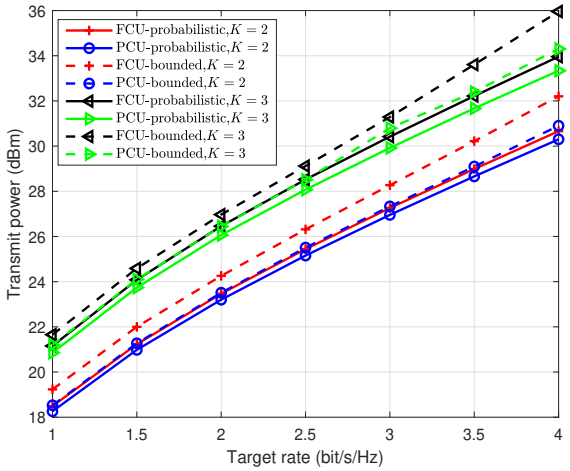
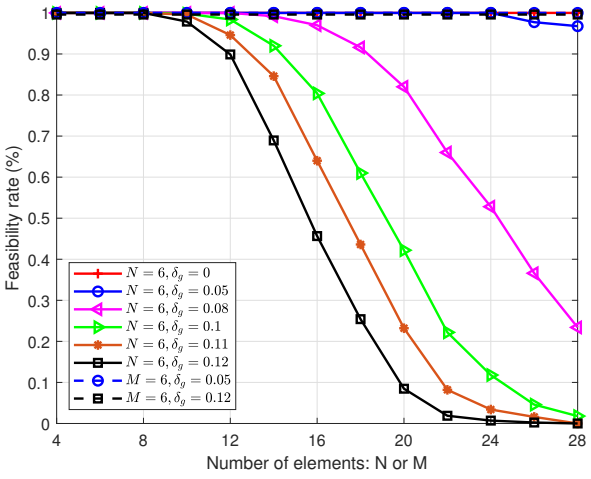


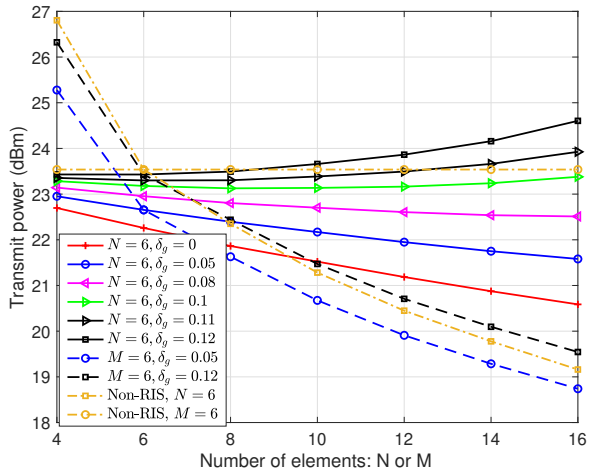
Figure 5.5: Transmit power versus the target rate R under $N = M = 6$ and $\{\delta_{\mathrm{g}}, \delta_{\mathrm{h}}\} = \{0.01, 0.02\}$.

Fig. 5.5 shows the minimum transmit power of the RIS-aided communication system versus the target rate requirements of users under various CSI error models. Some system parameters are set as $N = M = 6$, $K = \{2, 3\}$, $\{\delta_{\mathrm{g}}, \delta_{\mathrm{h}}\} = \{0.01, 0.02\}$. It is seen that the minimum transmit power increases with the target rate for both channel uncertainty scenarios and both CSI error models. In addition, it is also observed that the minimum transmit power of the worst-case robust design algorithms is larger than that of the outage constrained robust design algorithms. This is due to the fact that the worst-case optimization is the most conservative robust design, which requires more transmit power with the aim of ensuring that the achievable rate of each user meets the target rate requirement for the worst-case CSI error realization. Finally, the FCU scenario requires more transmit power to achieve the target rate than the PCU scenario, since FCU scenario needs more transmit power to combact the transmission loss caused by both the direct channel and reflection channel error.

In the following, the impact of the accuracy of the CSI on the system performance is studied. The outage constrained robust beamforming design algorithms is adopted since the computational complexity of the worst-case robust beamforming design algorithms is unacceptable at large numbers of antennas.



(a) Feasibility rate



(b) Transmit power

Figure 5.6: Feasibility rate and transmit power versus the number of antenna elements under the PCU scenario, when $K = 2$.

Fig. 5.6 shows the feasibility rate and the minimum transmit power versus N or M when only the CBRUT is imperfect, i.e., $\delta_{\mathrm{h}} = 0$. It is assumed that there are $K = 2$ users with $R = 2$ bit/s/Hz. The feasibility rate is defined as the ratio of the number of feasible channel realizations to the total number of channel realizations, where the feasible channel realization means that there exists a feasible solution to the outage constrained problem in (5.32) with this channel realization. An interesting phenomenon can be observed from Fig. 5.6(a). When fixing the number of transmit antennas N,

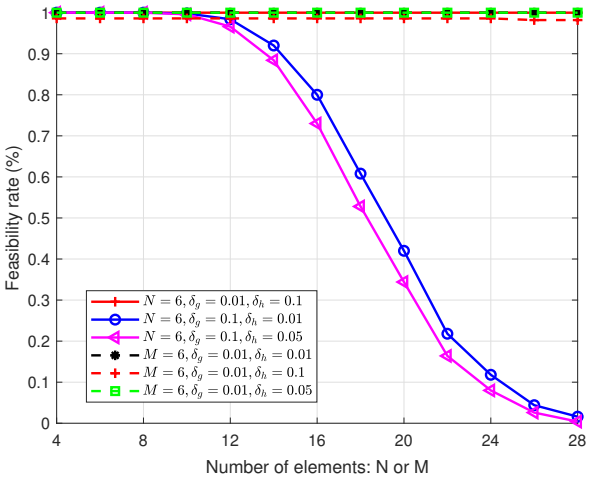
the feasibility rate decreases rapidly with the number of phase shifters at a high level of channel uncertainty ($\delta_{\mathrm{g}} \geq 0.08$). By contrast, when fixing the number of phase shifters M the feasibility rate keeps stable for different numbers of antennas even at a high level of channel uncertainty.

Based on the observations of Fig. 5.6(a), the minimum transmit power consumption of different channel uncertainty levels is further examined in Fig. 5.6(b) with a benchmark scheme without RIS. Fig. 5.6(b) is generated based on the channel realizations for which the feasible solutions can be obtained at $N = 16$ or $M = 16$.

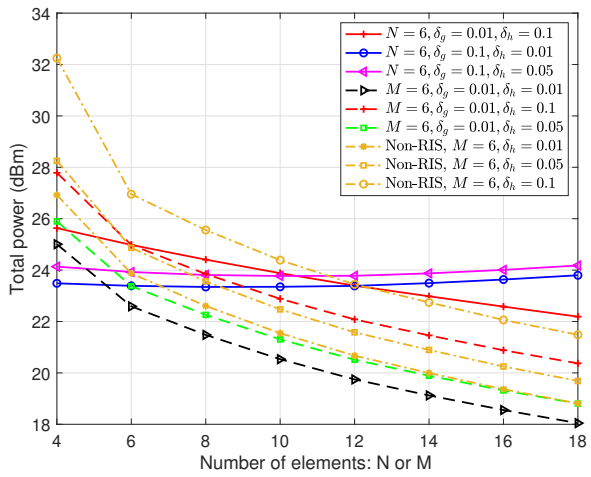
This work first studies the case with fixed number of transmit antennas $N = 6$. In Fig. 5.6(b), the case with $\delta_{\mathrm{g}} = 0$ can be regarded as the perfect CBRUT case, and its minimum transmit power decreases with the number of the reflection elements. This trend is consistent with that of Fig. 4 in [12]. The minimum transmit power consumption values under small values of δ_{g}, e.g., $\delta_{\mathrm{g}} = \{0.05, 0.08\}$, also decrease with the number of the reflection elements, and are higher than that of the perfect CBRUT case. The reason is that the BS needs to consume more power to compensate for the rate loss caused by the CBRUT error. However, when δ_{g} increases to 0.1 or larger, transmit power consumption starts to increase with the number of reflection elements. The reason is that increasing the number of reflection elements cannot only reduce the transmit power due to its increased beamforming gain, but also increase the channel estimation error that more transmit power is required to compensate for the channel errors. Therefore, when the CBRUT error is small, the benefits brought by the increase of M, outweighs its drawbacks, and vice versa. As a result, the number of RIS reflection elements should be carefully chosen, and the accuracy of the CBRUT estimation is crucial to reap the benefits offered by the RIS.

On the other hand, for the case with a fixed number of reflection elements, the transmit power consumption values decrease with the number of antennas at the BS even when the CBRUT error is high as $\delta_{\mathrm{g}} = 0.12$. The reason is that when the number of antennas is large, more degrees of freedom can be exploited to optimize the active

beamforming vector at the BS to compensate for the channel estimation error. Finally, compared with the system without RIS, the RIS may lose its performance gain advantage under high CBRUT error.



(a) Feasibility rate



(b) Transmit power

Figure 5.7: Feasibility rate and transmit power versus the number of antenna elements under the FCU scenario, when $K = 2$.

Fig. 5.7 shows the feasibility rate and the minimum transmit power versus M or N when both the DCSIT and the CBRUT are imperfect. The simulation parameters are the same as those in Fig. 5.6. Fig. 5.7(a) shows that when δ_{g} is low, the feasibility rates achieved by various cases are always high. In addition, from Fig. 5.7(b), it is

found that the increase of the number of antennas at the BS is effective in reducing the transmit power consumption, which is not affected by the DCSIT error δ_{h} (see curves $M = 6, \delta_{\mathrm{g}} = 0.01, \delta_{\mathrm{h}} = \{0.01, 0.05, 0.1\}$).

5.6 Summary

This chapter investigated robust beamforming designs under imperfect CBRUT for the RIS-aided MU-MISO system. The aim was to minimize the transmit power subject to the worst-case rate constraints under the bounded CSI error model and the rate outage probability constraints under the statistical CSI error model. The CSI uncertainties under the bounded CSI error model were addressed applying the S-procedure, and those under the statistical CSI error model were tackled by using the Bernstein-Type Inequality. The reformulated problems were efficiently solved under the AO framework. It is shown that the performance in terms of the minimum achievable transmit power, convergence and complexity under the statistical CSI error model is higher than that under the bounded CSI error model. The number of elements on the RIS may have a negative impact on system performance when the CBRUT error is large. This conclusion provides an engineering insight for the careful selection of the size of the RIS. In the end, this work provides a framework of robust transmission design in a simple single-cell multiuser scenario. The more complicated scenarios, such as the RISs-assisted full-duplex communication systems and RIS-aided physical layer security systems, will be studied as the future work. Furthermore, the robustness of the RIS in mmWave system under a geometric channel model is also worth studying.

Chapter 6

Robust Transmission in the Presence of Eavesdropper

This chapter studies the RIS-aided secrecy communication under the active attacks and passive eavesdropping. The contributions of this chapter are summarized as follows:

- This chapter proposes an RIS-aided two-phase secrecy communication scheme for a scenario where the ED has a similar channel direction as a LU in order to acquire high-quality eavesdropping information. In particular, in the multicasting phase, the BS transmits signals to the LU with low transmission power to reduce the information leakage to the ED. In the user cooperation phase, other LUs forward the received signals to the attacked LU with the assistance of RIS by using the energy harvested in the previous phase. In addition, two models of ED are considered in this work, i.e. active attack and passive eavesdropping.

- In the presence of statistical CSI error under the active attack, this work develops an outage constrained beamforming design that maximizes the secrecy rate subject to the unit-modulus constraint, the energy harvesting constraint and the secrecy rate outage probability constraint. Here, the outage probability constraint guar-

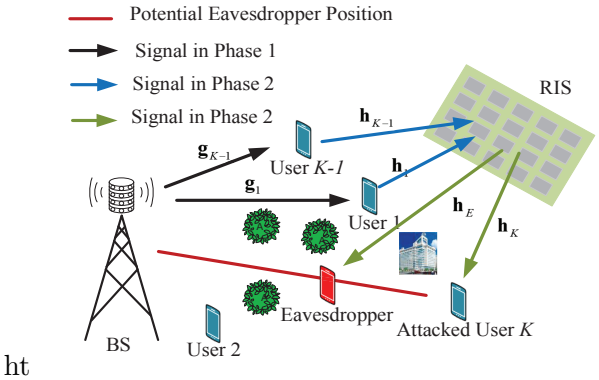
antees the maximum secrecy rate of the system for secure communication under a predetermined probability. By resorting to the BTI and convex approximations, the non-convexity of constraints is addressed. Then, the active precoders and the passive reflecting beamforming are updated by using the proposed SDP and penalty CCP technique respectively in an iterative manner.

- For the passive ED case with only partial CSI, it maximizes an average secrecy rate subject to the unit-modulus constraint of the reflecting beamforming and the energy harvesting constraint. To address the numerical integration in the objective function, an angular secrecy model, which is analytically non-convex, is proposed. A low-complexity algorithm is proposed based on the MM-based AO framework, where the active beamforming vectors are updated by solving a convex optimization problem and the reflecting beamforming vectors are updated in a closed-form solution which is globally optimal.

- The numerical results demonstrate that the level of the cascaded CSI error plays a vital role in the RIS-aided secure communication systems. In particular, at low error of cascaded CSI, the secrecy rate increases with the number of elements at the RIS due to the increased beamforming gain. However, at large level of cascaded CSI error, the secrecy rate decreases with the number of elements at the RIS due to the increased channel estimation error. Hence, whether to enable the RIS for enhancing the security capacity in the communication systems depends on the level of the cascaded CSI error. In addition, the RIS can enhance the average secrecy rate under the passive eavesdropping.

6.1 System Model

As shown in Fig. 6.1, a BS with N transmit antennas communicates with K single-antenna LUs in the presence of a single-antenna ED. An RIS with M reflecting elements is introduced to aid the secure communication.



ht

Figure 6.1: Two-phase communication system

6.1.1 Channel Model

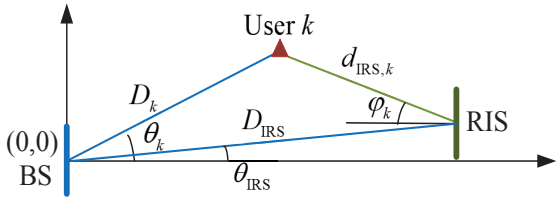


Figure 6.2: Coordinates of communication nodes in the system

Rician wiretap channels are considered for the analysis of angle aware two-phase security communication. Define the set of all LUs as $\mathcal{K} = \{1, 2, ..., K\}$, and denote set $\mathcal{K}_{-K} = \mathcal{K}/\{K\}$ and set $\mathcal{K}_{+E} = \mathcal{K} \cup \{E\}$. By denoting $\{D_i, \theta_i\}_{\forall i \in \mathcal{K}_{+E}}$ as the distances and the azimuth angles respectively from the BS to the LUs and the ED, as shown in Fig. 6.1, Rician channel is used to model the corresponding channels $\{\mathbf{g}_i \in \mathbb{C}^{N\times 1}\}_{\forall i \in \mathcal{K}_{+E}}$ [129]:

$$\mathbf{g}_i = \sqrt{\varrho_0 \left(\frac{D_i}{d_0}\right)^{-\alpha_{\mathrm{BS}}}} \left(\sqrt{\frac{K_{\mathrm{BS}}}{1+K_{\mathrm{BS}}}} \mathbf{g}_i^{\mathrm{LOS}} + \sqrt{\frac{1}{1+K_{\mathrm{BS}}}} \mathbf{g}_i^{\mathrm{NLOS}} \right), \forall i \in \mathcal{K}_{+E}, \tag{6.1}$$

where ϱ_0 is the pathloss at the reference distance of d_0, α_{BS} and K_{BS} are the pathloss exponent and the Rician factor of the BS-related links, respectively. It is assumed that the BS is equipped with a ULA. Then, the LoS component is given by $\mathbf{g}_i^{\mathrm{LOS}} = \left[1, e^{-\mathrm{j}\pi \sin\theta_i}, \cdots, e^{-\mathrm{j}(N-1)\pi \sin\theta_i}\right]$, and the non-LoS component is drawn from a Rayleigh fading, i.e., $\mathbf{g}_i^{\mathrm{NLOS}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$.

Furthermore, by denoting $\{D_{\mathrm{RIS}}, \theta_{\mathrm{RIS}}\}$ as the distance and the azimuth angle from the BS to the RIS, it is straightforward to obtain the distances $\{d_{\mathrm{RIS},i}\}_{\forall i \in \mathcal{K}_{+E}}$ and the azimuth angles $\{\varphi_i\}_{\forall i \in \mathcal{K}_{+E}}$ from the RIS to the LUs and the ED as shown in Fig. 6.2:

$$
\begin{aligned}
d_{\mathrm{RIS},i} &= \left[(D_{\mathrm{RIS}} \cos\theta_{\mathrm{RIS}} - D_i \cos\theta_i)^2 + (D_{\mathrm{RIS}} \sin\theta_{\mathrm{RIS}} - D_i \sin\theta_i)^2 \right]^{-1/2}, \\
\sin\varphi_i &= \frac{1}{d_{\mathrm{RIS},i}} (D_i \sin\theta_i - D_{\mathrm{RIS}} \sin\theta_{\mathrm{RIS}}), \\
\cos\varphi_i &= \frac{1}{d_{\mathrm{RIS},i}} (D_{\mathrm{RIS}} \cos\theta_{\mathrm{RIS}} - D_i \cos\theta_i).
\end{aligned}
$$

The corresponding channels $\{\mathbf{h}_i \in \mathbb{C}^{N\times 1}\}_{\forall i \in \mathcal{K}_{+E}}$ are given by

$$
\mathbf{h}_i = \sqrt{\varrho_0 \left(\frac{d_{\mathrm{RIS},i}}{d_0} \right)^{-\alpha_{\mathrm{RIS}}}} \left(\sqrt{\frac{K_{\mathrm{RIS}}}{1+K_{\mathrm{RIS}}}} \mathbf{h}_i^{\mathrm{LOS}} + \sqrt{\frac{1}{1+K_{\mathrm{RIS}}}} \mathbf{h}_i^{\mathrm{NLOS}} \right), \forall i \in \mathcal{K}_{+E}, \quad (6.2)
$$

where α_{RIS} and K_{RIS} are the pathloss exponent and the Rician factor of the RIS-related links, respectively. $\mathbf{h}_i^{\mathrm{NLOS}}$ is the non-LoS component, and its distribution is the same as that of $\mathbf{g}_i^{\mathrm{NLOS}}$. By assuming an $M = M_{\mathrm{x}} \times M_{\mathrm{y}}$ UPA deployed at the RIS with M_{x} and M_{y} being the number of reflecting elements in x-axis and y-axis, respectively, the LoS component is then written as

$$
\mathbf{h}_i^{\mathrm{LOS}} = \mathbf{h}_i^{\mathrm{x}} \otimes \mathbf{h}_i^{\mathrm{y}}, \quad (6.3)
$$

where $\mathbf{h}_i^{\mathrm{x}} = [1, \cdots, e^{-\mathrm{j}\pi(M_{\mathrm{x}}-1)\cos\varphi_i \cos\phi \sin\theta_i}]^{\mathrm{T}}$, $\mathbf{h}_i^{\mathrm{y}} = [1, \cdots, e^{-\mathrm{j}\pi(M_{\mathrm{x}}-1)\sin\varphi_i \cos\phi \sin\theta_i}]^{\mathrm{T}}$, and ϕ is the elevation angle observed at the RIS side.

6.1.2 Signal Transmission

As shown in Fig. 6.1, it is assumed that the ED hides at the line connecting the BS to one of the users, denoted as user K, to achieve high success rate of attack. In this situation, the signal received by the ED is highly correlated with user K [38], [39], which leads to $\theta_E \approx \theta_K$, $\mathbf{g}_E^{\mathrm{LOS}} \approx \mathbf{g}_K^{\mathrm{LOS}}$ and $D_E \in (0, D_K)$. When the Rician factor K_{BS} is sufficiently large, the channel $\mathbf{g}_E$ is approximately equal to the channel $\mathbf{g}_K$.

In particular, in the first phase, the BS multicasts the common signal to all users

except user K [1]. In the second phase, the helping users ($\forall k \in \mathcal{K}_{-K}$) forward the decoded common signal to user K via the RIS. In this work, in order to implement the AAUC scheme without consuming extra energy, the LUs adopt the hybrid information and energy harvesting receiving mode which splits the received signal into two power streams with power splitting ratios t_k and $1 - t_k$. The former is used for decoding the signal and the latter is for energy harvesting.

6.1.2.1 Multicasting Phase

In this phase, the BS multicasts the signal s to the helping LUs through beamforming vector $\mathbf{f} \in \mathbb{C}^{N\times 1}$ which is limited to the maximum transmit power $P_{\max}$, i.e., $||\mathbf{f}||_2^2 \leq P_{\max}$. Since $\mathbf{g}_E \approx \mathbf{g}_K$, the beamforming $\mathbf{f}$ needs to satisfy $|\mathbf{g}_K^{\mathrm{H}}\mathbf{f}| = 0$ to ensure that $|\mathbf{g}_E^{\mathrm{H}}\mathbf{f}| \approx 0$. Let $\mathbf{Q} \in \mathbb{C}^{N\times(N-1)}$ be the orthogonal matrix which spans the null space of $\mathbf{g}_K$ by using the QR decomposition, i.e., $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{I}$. Then, design $\mathbf{f} = \mathbf{Q}\mathbf{z}$, where $\mathbf{z} \in \mathbb{C}^{(N-1)\times 1}$ is a newly introduced variable. Therefore, the signal received by LU k is given by $\mathbf{g}_k^{\mathrm{H}}\mathbf{Q}\mathbf{z} + n_k$, where n_k is the received noise with the noise power of σ_k^2. By adopting the hybrid receiving mode and let $\mathbf{t} = [t_1, ..., t_{K-1}]^{\mathrm{T}}$ where t_k is the power splitting ratio of LU k, the achievable rate at LU $k \neq K$ is

$$R_k\left(\mathbf{z}, t_k\right) = \frac{1}{2}\log_2\left(1 + \frac{t_k}{\sigma_k^2}\left|\mathbf{g}_k^{\mathrm{H}}\mathbf{Q}\mathbf{z}\right|^2\right), \tag{6.4}$$

where the factor $1/2$ is due to the assumption that the total time duration is evenly distributed to two transmission phases. The harvested power at LU $k \neq K$ is

$$\left(1 - t_k\right)\left|\mathbf{g}_k^{\mathrm{H}}\mathbf{Q}\mathbf{z}\right|^2. \tag{6.5}$$

[1]The RIS is assumed to be turned off in the first phase to avoid reflecting useful signals to the ED. Please note that this assumption is practical due to the fact that the BS has no responsibility of designing secure reflecting beaming for the RIS, which reduces the computational complexity and hardware cost at the BS.

6.1.2.2 User Cooperation Phase

In this phase, the helping LUs ($\forall k \in \mathcal{K}_{-K}$) forward the signal s to LU K through a beamforming vector $\mathbf{w} \in \mathbb{C}^{(K-1)\times 1} = [w_1, ..., w_{K-1}]^{\mathrm{T}}$ by using the power harvested in the multicasting phase. Since LU K is randomly selected by the ED and assume that there are many obstacles in the communication environment, such as indoor applications, the direct links between the helping LUs and the LU K may be blocked. To address this issue, an RIS can be installed on the building with a certain height, and thus the RIS is capable of reflecting the signals forwarded by the helping LUs to LU K. Denote by $\mathbf{e}$ the reflection coefficient vector of the RIS, where $|e_m|^2 = 1, \forall m = 1, \cdots, M$. Then, the signal received by LU K is given by

$$\begin{aligned} y_K &= \mathbf{h}_K^{\mathrm{H}} \mathrm{diag}(\mathbf{e}^*) \mathbf{H}_{\mathrm{RIS}} \mathbf{w} s + n_K \\ &= \mathbf{e}^{\mathrm{H}} \mathbf{H}_K \mathbf{w} s + n_K, \end{aligned}$$

where $\mathbf{H}_{\mathrm{RIS}} = [\mathbf{h}_1, ..., \mathbf{h}_{K-1}]$, $\mathbf{H}_K = [\mathbf{h}_K^* \odot \mathbf{h}_1, ..., \mathbf{h}_K^* \odot \mathbf{h}_{K-1}]$ is the cascaded LU-RIS-LU (LRL) channel, and $n_K \sim \mathcal{CN}(\mathbf{0}, \sigma_K^2)$ is the noise. The corresponding achievable rate is

$$R_K(\mathbf{w}, \mathbf{e}) = \frac{1}{2} \log_2 \left(1 + \frac{1}{\sigma_K^2} \left| \mathbf{e}^{\mathrm{H}} \mathbf{H}_K \mathbf{w} \right|^2 \right). \tag{6.6}$$

On the other hand, the signal received by the ED is $y_E = \mathbf{e}^{\mathrm{H}} \mathbf{H}_E \mathbf{w} s + n_E$, where $\mathbf{H}_E = [\mathbf{h}_E^* \odot \mathbf{h}_1, ..., \mathbf{h}_E^* \odot \mathbf{h}_{K-1}]$ is the cascaded LU-RIS-ED (LRE) channel, and $n_E \sim \mathcal{CN}(\mathbf{0}, \sigma_E^2)$ is the received noise at the ED.

The corresponding eavesdropping rate is

$$R_E(\mathbf{w}, \mathbf{e}) = \frac{1}{2} \log_2 \left(1 + \frac{1}{\sigma_E^2} \left| \mathbf{e}^{\mathrm{H}} \mathbf{H}_E \mathbf{w} \right|^2 \right). \tag{6.7}$$

Finally, the secrecy rate of this system under the AAUC scheme can be given by [39]:

$$\left[\min_{\forall k \in \mathcal{K}} R_k - R_E \right]^+. \tag{6.8}$$

The following two sections consider the system design for two ED models: the active eavesdropper model and the passive eavesdropper model.

6.2 ED Model I: Active Eavesdropper Model

This section considers the active attack case, in which the ED pretends to be an LU sending pilot signals to the transmitters (including the BS and the helping LUs) during the channel estimation procedure [38], [39]. It is reasonable to assume that the BS is capable of addressing this attack by using the multi-antenna technique, so as to obtain perfect CSI of the system. Nevertheless, the single-antenna helping LUs only have the imperfect CSI of LU K and the ED due to their limited anti-interference ability.

6.2.1 Channel Uncertainties

Based on the above assumption, the cascaded channels can be modeled as

$$\mathbf{H}_K = \widehat{\mathbf{H}}_K + \triangle_K,\ \mathbf{H}_E = \widehat{\mathbf{H}}_E + \triangle_E, \tag{6.9}$$

where $\widehat{\mathbf{H}}_K$ and $\widehat{\mathbf{H}}_E$ are the estimated cascaded channels, $\triangle_K = [\triangle_1^K \cdots \triangle_{K-1}^K]$ and $\triangle_E = [\triangle_1^E \cdots \triangle_{K-1}^E]$ are the unknown cascaded channel errors. $\triangle_k^K$ and $\triangle_k^E$ are the unknown cascaded LRL and LRE channel error vectors at LU k, respectively.

According to [75], the robust beamforming under the statistical CSI error model outperforms the bounded CSI error model in terms of the minimum transmit power, convergence speed and computational complexity. In addition, the statistical channel error model is more suitable to model the channel estimation error when the channel estimation is based on the Minimum Mean Sum Error (MMSE) method. Hence, the statistical model is adopted to characterize the cascaded CSI imperfection [75], i.e., each CSI error vector is assumed to follow the CSCG distribution, i.e.,

$$\triangle_k^K \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k^K), \mathbf{\Sigma}_k^K \succeq \mathbf{0}, \forall k \in \mathcal{K}_{-K}, \tag{6.10a}$$

$$\triangle_k^E \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k^E), \mathbf{\Sigma}_k^E \succeq \mathbf{0}, \forall k \in \mathcal{K}_{-K}, \tag{6.10b}$$

where $\mathbf{\Sigma}_k^K \in \mathbb{C}^{M\times M}$ and $\mathbf{\Sigma}_k^E \in \mathbb{C}^{M\times M}$ are positive semidefinite error covariance matrices. Note that the CSI error vectors of different LUs are independent with each other. Therefore, it has

$$\text{vec}(\triangle_K) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_K),\ \text{vec}(\triangle_E) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_E), \tag{6.11}$$

where $\mathbf{\Sigma}_K$ and $\mathbf{\Sigma}_E$ are block diagonal matrices, i.e., $\mathbf{\Sigma}_K = \text{diag}(\mathbf{\Sigma}_1^K, ..., \mathbf{\Sigma}_{K-1}^K)$ and $\mathbf{\Sigma}_E = \text{diag}(\mathbf{\Sigma}_1^E, ..., \mathbf{\Sigma}_{K-1}^E)$.

6.2.2 Outage Constrained Beamforming Design

Under the statistical CSI error model, a probabilistic robust algorithm is developed for the secrecy rate maximization problem, which is formulated as

$$\max_{R_{\text{sec}},\mathbf{z},\mathbf{w},\mathbf{e},\mathbf{t}} R_{\text{sec}} \tag{6.12a}$$

$$\text{s.t.Pr}\left\{\min_{\forall k\in\mathcal{K}} R_k - R_E \geq R_{\text{sec}}\right\} \geq 1-\rho \tag{6.12b}$$

$$||\mathbf{z}||_2^2 \leq P_{\max} \tag{6.12c}$$

$$|e_m|^2 = 1, 1 \leq m \leq M \tag{6.12d}$$

$$0 \leq \mathbf{t} \leq 1 \tag{6.12e}$$

$$|w_k|^2 \leq (1-t_k)\left|\mathbf{g}_k^{\text{H}}\mathbf{Q}\mathbf{z}\right|^2, \forall k \in \mathcal{K}_{-K}, \tag{6.12f}$$

where $\rho \in (0, 1]$ is the secrecy rate outage probability.

Problem (6.12) is difficult to solve due to the computationally intractable rate outage probability constraint (6.12b), the non-convex unit-modulus constraint (6.12d), and the non-convex power constraint (6.12f).

Firstly, constraint (6.12b) is replaced with the development of a safe approximation consisting of three steps in the following.

Step 1: Decouple the Probabilistic Constraint: First of all, based on the independence between $\{\mathbf{g}_k\}_{\forall k\in\mathcal{K}_{-K}}$ and $\mathbf{H}_K$, it comes to

$$(6.12b) \Leftrightarrow \prod_{k=1}^{K} \Pr\left\{R_k - R_E \geq R_{\text{sec}}\right\} \geq 1-\rho \tag{6.13}$$

$$\Leftarrow \Pr\left\{R_k - R_E \geq R_{\text{sec}}\right\} \geq 1-\bar{\rho}, \forall k \in \mathcal{K}_K, \tag{6.14}$$

where $\bar{\rho} = 1-(1-\rho)^{1/K}$.

Step 2: Convenient Approximations: To address the non-concavity of $R_k - R_E, \forall k \in \mathcal{K}_K$, it needs to construct a sequence of surrogate functions of $\{R_i\}_{\forall i\in\mathcal{K}_{+E}}$. More specifically, the following lemmas is needed.

Lemma 9. *[79] The quadratic-over-linear function $\frac{x^2}{y}$ is jointly convex in (x,y), and lower bounded by its linear first-order Taylor approximation $\frac{2x^t}{y^t}x - (\frac{x^t}{y^t})^2 y$ at fixed point (x^t, y^t).*

By substituting x with $\mathbf{g}_k^{\text{H}}\mathbf{Q}\mathbf{z}$ and y with $1/t_k$, Lemma 9 is utilized to obtain a concave lower bound of rate $R_k(\mathbf{z}, t_k)$ for $\forall k \in \mathcal{K}_{-K}$. The lower bound is given by

$$\widetilde{R}_k(\mathbf{z}, t_k|\mathbf{z}^t, t_k^t) = \frac{1}{2}\log_2\left(1 - \frac{t_k^{t,2}}{\sigma_k^2 t_k}\left|\mathbf{g}_k^{\text{H}}\mathbf{Q}\mathbf{z}^t\right|^2 + 2t_k^t \text{Re}\left\{\frac{1}{\sigma_k^2}\mathbf{z}^{t,\text{H}}\mathbf{Q}^{\text{H}}\mathbf{g}_k\mathbf{g}_k^{\text{H}}\mathbf{Q}\mathbf{z}\right\}\right) \tag{6.15}$$

for any feasible solution $\{\mathbf{z}^t, t_k^t\}$.

Lemma 10. *The upper bound of rate $R_E(\mathbf{w}, \mathbf{e})$ is given by*

$$\widetilde{R}_E(\mathbf{w}, \mathbf{e}, a_E) = \frac{a_E\left|\mathbf{e}^{\text{H}}\mathbf{H}_E\mathbf{w}\right|^2/\sigma_E^2 + a_E - \ln a_E - 1}{2\ln 2},$$

where a_E is the auxiliary variable.

Proof: Please refer to Appendix D.1. ■

Lemma 11. *The lower bound of rate* $R_K(\mathbf{w}, \mathbf{e})$ *is given by*

$$\begin{aligned}\widetilde{R}_K(\mathbf{w}, \mathbf{e}, a_K, v) =& \frac{1}{2\ln 2}\left(-a_K|v|^2|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}|^2 - \sigma_K^2 a_K|v|^2 + 2a_K Re\left\{v\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right\}\right.\\ &\left.-a_K + \ln a_K + 1\right),\end{aligned}$$

where a_K *and* v *are the auxiliary variables.*

Proof: Please refer to Appendix D.2. ■

For the convenience of derivations, it is assumed that $\mathbf{\Sigma}_k^K = \varepsilon_{K,k}^2\mathbf{I}_M$ and $\mathbf{\Sigma}_k^E = \varepsilon_{E,k}^2\mathbf{I}_M$, then $\mathbf{\Sigma}_K = \mathbf{\Lambda}_K \otimes \mathbf{I}_M$ where $\mathbf{\Lambda}_K = \mathrm{diag}(\varepsilon_{K,1}^2, ..., \varepsilon_{K,K-1}^2)$, and $\mathbf{\Sigma}_E = \mathbf{\Lambda}_E \otimes \mathbf{I}_M$ where $\mathbf{\Lambda}_E = \mathrm{diag}(\varepsilon_{E,1}^2, ..., \varepsilon_{E,K-1}^2)$. Furthermore, the error vectors in (6.11) can be rewritten as $\mathrm{vec}(\triangle_K) = \mathbf{\Sigma}_K^{\frac{1}{2}}\mathbf{i}_K$ where $\mathbf{i}_K \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{M(K-1)})$, and $\mathrm{vec}(\triangle_E) = \mathbf{\Sigma}_E^{\frac{1}{2}}\mathbf{i}_E$ where $\mathbf{i}_E \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{M(K-1)})$. Define $\mathbf{E} = \mathbf{e}\mathbf{e}^{\mathrm{H}}$ and $\mathbf{W} = \mathbf{w}\mathbf{w}^{\mathrm{H}}$. Combining (6.15) with Lemma 10, Left Hand Side (LHS) of (6.14) corresponding to the users in $\mathcal{K}_{-K}$ can be replaced by its lower bound:

$$\begin{aligned}&\Pr\left\{R_k - R_E \geq R_{\mathrm{sec}}\right\}\\ \geq&\Pr\left\{\widetilde{R}_k - \widetilde{R}_E \geq R_{\mathrm{sec}}\right\}\\ =&\Pr\left\{a_E\mathrm{Tr}\left(\mathbf{E}(\widehat{\mathbf{H}}_E + \triangle_E)\mathbf{W}(\widehat{\mathbf{H}}_E^{\mathrm{H}} + \triangle_E^{\mathrm{H}})\right) - [2\ln 2(\widetilde{R}_k - R_{\mathrm{sec}}) - a_E + \ln a_E + 1]\sigma_E^2 \leq 0\right\}\\ =&\Pr\left\{\mathbf{i}_E^{\mathrm{H}}\mathbf{U}_E\mathbf{i}_E + 2\mathrm{Re}\left\{\mathbf{u}_E^{\mathrm{H}}\mathbf{i}_E\right\} + u_k \leq 0\right\},\end{aligned} \tag{6.16}$$

where

$$\mathbf{U}_E = a_E\mathbf{\Sigma}_E^{\frac{1}{2}}(\mathbf{W}^{\mathrm{T}} \otimes \mathbf{E})\mathbf{\Sigma}_E^{\frac{1}{2}}, \tag{6.17a}$$

$$\mathbf{u}_E = a_E\mathbf{\Sigma}_E^{\frac{1}{2}}\mathrm{vec}(\mathbf{E}\widehat{\mathbf{H}}_E\mathbf{W}), \tag{6.17b}$$

$$u_k = a_E\mathrm{Tr}\left(\mathbf{E}\widehat{\mathbf{H}}_E\mathbf{W}\widehat{\mathbf{H}}_E^{\mathrm{H}}\right) - [(\widetilde{R}_k - R_{\mathrm{sec}})2\ln 2 - a_E + \ln a_E + 1]\sigma_E^2. \tag{6.17c}$$

Combining Lemma 10 with Lemma 11, the LHS of (6.14) corresponding to LU K can

be replaced by its lower bound:

$$\begin{aligned}
&\Pr\{R_K - R_E \geq R_{\text{sec}}\} \\
\geq&\Pr\left\{\widetilde{R}_K - \widetilde{R}_E \geq R_{\text{sec}}\right\} \\
=&\Pr\Big\{a_K|v|^2\text{Tr}\left(\mathbf{E}(\widehat{\mathbf{H}}_K + \triangle_K)\mathbf{W}(\widehat{\mathbf{H}}_K^{\text{H}} + \triangle_K^{\text{H}})\right) - 2a_K\text{Re}\left\{v\mathbf{e}^{\text{H}}(\widehat{\mathbf{H}}_K + \triangle_K)\mathbf{w}\right\} + \\
&\frac{a_E}{\sigma_E^2}\text{Tr}\left(\mathbf{E}(\widehat{\mathbf{H}}_E + \triangle_E)\mathbf{W}(\widehat{\mathbf{H}}_E^{\text{H}} + \triangle_E^{\text{H}})\right) - c \leq 0\Big\} \\
=&\Pr\left\{\mathbf{i}^{\text{H}}\mathbf{U}_K\mathbf{i} + 2\text{Re}\left\{\mathbf{u}_K^{\text{H}}\mathbf{i}\right\} + u_K \leq 0\right\},
\end{aligned} \tag{6.18}$$

where

$$\mathbf{i} = [\ \mathbf{i}_K^{\text{H}},\ \ \mathbf{i}_E^{\text{H}}\ ]^{\text{H}}, \tag{6.19a}$$

$$\mathbf{U}_K = \text{diag}\Big\{a_K|v|^2\boldsymbol{\Sigma}_K^{\frac{1}{2}}(\mathbf{W}^{\text{T}} \otimes \mathbf{E})\boldsymbol{\Sigma}_K^{\frac{1}{2}},\ \frac{a_E}{\sigma_E^2}\boldsymbol{\Sigma}_E^{\frac{1}{2}}(\mathbf{W}^{\text{T}} \otimes \mathbf{E})\boldsymbol{\Sigma}_E^{\frac{1}{2}}\Big\}, \tag{6.19b}$$

$$\mathbf{u}_K = [a_K|v|^2\text{vec}^{\text{H}}(\mathbf{E}\widehat{\mathbf{H}}_K\mathbf{W})\boldsymbol{\Sigma}_K^{\frac{1}{2}} - a_K v\text{vec}^{\text{H}}(\mathbf{e}\mathbf{w}^{\text{H}})\boldsymbol{\Sigma}_K^{\frac{1}{2}},\ \frac{a_E}{\sigma_E^2}\text{vec}^{\text{H}}(\mathbf{E}\widehat{\mathbf{H}}_E\mathbf{W})\boldsymbol{\Sigma}_E^{\frac{1}{2}}]^{\text{H}}, \tag{6.19c}$$

$$u_K = a_K|v|^2\text{Tr}\left(\mathbf{E}\widehat{\mathbf{H}}_K\mathbf{W}\widehat{\mathbf{H}}_K^{\text{H}}\right) + \frac{a_E}{\sigma_E^2}\text{Tr}\left(\mathbf{E}\widehat{\mathbf{H}}_E\mathbf{W}\widehat{\mathbf{H}}_E^{\text{H}}\right)\ - 2a_K\text{Re}\left\{v\mathbf{e}^{\text{H}}\widehat{\mathbf{H}}_K\mathbf{w}\right\} - c, \tag{6.19d}$$

$$c = \ln a_E + \ln a_K - a_E - a_K - 2R_{\text{sec}}\ln 2\ - \sigma_K^2 a_K|v|^2 + 2. \tag{6.19e}$$

Now, by substituting (6.16) and (6.18) into (6.14), (6.14) can be approximated as

$$\Pr\left\{\mathbf{i}_E^{\text{H}}\mathbf{U}_E\mathbf{i}_E + 2\text{Re}\left\{\mathbf{u}_E^{\text{H}}\mathbf{i}_E\right\} + u_k \leq 0\right\} \geq 1 - \bar{\rho}, \forall k \in \mathcal{K}_{-K}, \tag{6.20a}$$

$$\Pr\left\{\mathbf{i}^{\text{H}}\mathbf{U}_K\mathbf{i} + 2\text{Re}\left\{\mathbf{u}_K^{\text{H}}\mathbf{i}\right\} + u_K \leq 0\right\} \geq 1 - \bar{\rho}. \tag{6.20b}$$

Step 3: A BTI-Based Safe Approximation: The outage probabilities in (6.20) are characterized by quadratic inequalities, which can be safely approximated by using the following lemma.

Lemma 12. *(Bernstein-Type Inequality) [130] Assume* $f(\mathbf{x}) = \mathbf{x}^{\mathrm{H}}\mathbf{U}\mathbf{x} + 2\mathrm{Re}\{\mathbf{u}^{\mathrm{H}}\mathbf{x}\} + u$, *where* $\mathbf{U} \in \mathbb{H}^{n\times n}$, $\mathbf{u} \in \mathbb{C}^{n\times 1}$, $u \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{C}^{n\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. *Then for any* $\rho \in [0, 1]$, *the following approximation holds:*

$$\begin{aligned}
&\Pr\{\mathbf{x}^{\mathrm{H}}\mathbf{U}\mathbf{x} + 2\mathrm{Re}\{\mathbf{u}^{\mathrm{H}}\mathbf{x}\} + u \le 0\} \ge 1 - \rho \\
\Rightarrow &\mathrm{Tr}\,(\mathbf{U}) + \sqrt{2\ln(1/\rho)}x - \ln(\rho)\lambda_{\max}^{+}(\mathbf{U}) + u \le 0 \\
\Rightarrow &\begin{cases} \mathrm{Tr}\,(\mathbf{U}) + \sqrt{2\ln(1/\rho)}x - \ln(\rho)y + u \le 0 \\ \sqrt{||\mathbf{U}||_F^2 + 2||\mathbf{u}||_2^2} \le x \\ y\mathbf{I} - \mathbf{U} \succeq \mathbf{0}, y \ge 0, \end{cases}
\end{aligned} \tag{6.21}$$

where $\lambda_{\max}^{+}(\mathbf{U}) = \max(\lambda_{\max}(\mathbf{U}), 0)$. x *and* y *are slack variables.*

Before using Lemma 12, the following simplified derivations is needed for LU k, $\forall k \in \mathcal{K}_{-K}$, i.e.,

$$\begin{aligned}
\mathrm{Tr}\,(\mathbf{U}_E) &= \mathrm{Tr}\left(a_E\boldsymbol{\Sigma}_E^{\frac{1}{2}}(\mathbf{W}^{\mathrm{T}} \otimes \mathbf{E})\boldsymbol{\Sigma}_E^{\frac{1}{2}}\right) \\
&= \mathrm{Tr}\left(a_E(\mathbf{W}^{\mathrm{T}} \otimes \mathbf{E})(\boldsymbol{\Lambda}_E \otimes \mathbf{I}_M)\right) \\
&= a_E M\mathrm{Tr}\,(\boldsymbol{\Lambda}_E\mathbf{W}), \qquad (6.22a) \\
||\mathbf{U}_E||_F^2 &= a_E^2M^2||\boldsymbol{\Lambda}_E\mathbf{W}||_F^2, \qquad (6.22b) \\
||\mathbf{u}_E||_2^2 &= a_E^2\mathrm{vec}^{\mathrm{H}}(\mathbf{E}\widehat{\mathbf{H}}_E\mathbf{W})(\boldsymbol{\Lambda}_E \otimes \mathbf{I}_M)\mathrm{vec}(\mathbf{E}\widehat{\mathbf{H}}_E\mathbf{W}) \\
&= a_E^2M||\boldsymbol{\Lambda}_E^{\frac{1}{2}}\mathbf{W}\widehat{\mathbf{H}}_E^{\mathrm{H}}\mathbf{e}||_2^2, \qquad (6.22c) \\
\lambda_{\max}(\mathbf{U}_E) &= \lambda_{\max}(a_E\boldsymbol{\Sigma}_E^{\frac{1}{2}}(\mathbf{W}^{\mathrm{T}} \otimes \mathbf{E})\boldsymbol{\Sigma}_E^{\frac{1}{2}}) \\
&= \lambda_{\max}(a_E(\boldsymbol{\Lambda}_E\mathbf{W}^{\mathrm{T}} \otimes \mathbf{E})) \\
&= a_E\lambda_{\max}(\boldsymbol{\Lambda}_E\mathbf{W})\lambda(\mathbf{E}) = a_EM\mathbf{w}^{\mathrm{H}}\boldsymbol{\Lambda}_E\mathbf{w}. \qquad (6.22d)
\end{aligned}$$

By substituting (6.22) into (6.21) and introducing slack variables $\{x_E, y_E\}$, the con-

straints for $\forall k \in \mathcal{K}_{-K}$ in (6.20a) are transformed into the following deterministic form:

$$\mathrm{BTI}_1 \triangleq \begin{cases} \mathrm{Tr}\left(\mathbf{U}_E\right) + \sqrt{2\ln(1/\bar{\rho})}x_E - \ln(\bar{\rho})y_E \\ \qquad +u_k \leq 0, \forall k \in \mathcal{K}_{-K} \\ \left\| \begin{array}{c} a_E M \mathrm{vec}(\mathbf{\Lambda}_E \mathbf{W}) \\ \sqrt{2M} a_E \mathbf{\Lambda}_E^{\frac{1}{2}} \mathbf{W} \widehat{\mathbf{H}}_E^{\mathrm{H}} \mathbf{e} \end{array} \right\| \leq x_E \\ y_E \geq a_E M \mathbf{w}^{\mathrm{H}} \mathbf{\Lambda}_E \mathbf{w}. \end{cases} \tag{6.23}$$

On the other hand, the simplified derivations for LU K are given by

$$\mathrm{Tr}\left(\mathbf{U}_K\right) = a_K |v|^2 M \mathrm{Tr}\left(\mathbf{\Lambda}_K \mathbf{W}\right) + \frac{a_E}{\sigma_E^2} M \mathrm{Tr}\left(\mathbf{\Lambda}_E \mathbf{W}\right), \tag{6.24a}$$

$$||\mathbf{U}_K||_F^2 = a_K^2 |v|^4 M^2 ||\mathbf{\Lambda}_K \mathbf{W}||_F^2 + \frac{a_E^2}{\sigma_E^4} M^2 ||\mathbf{\Lambda}_E \mathbf{W}||_F^2, \tag{6.24b}$$

$$||\mathbf{u}_K||^2 = M ||\mathbf{\Lambda}_K^{\frac{1}{2}} \left(a_K |v|^2 \mathbf{W} \widehat{\mathbf{H}}_K^{\mathrm{H}} \mathbf{e} - a_K v \mathbf{w} \right) ||_2^2 + \frac{a_E^2}{\sigma_E^4} M ||\mathbf{e}^{\mathrm{H}} \widehat{\mathbf{H}}_E \mathbf{W} \mathbf{\Lambda}_E^{\frac{1}{2}}||_2^2, \tag{6.24c}$$

$$\lambda_{\max}(\mathbf{U}_K) = \max \left\{ a_K |v|^2 M \mathbf{w}^{\mathrm{H}} \mathbf{\Lambda}_K \mathbf{w}, \ \frac{a_E}{\sigma_E^2} M \mathbf{w}^{\mathrm{H}} \mathbf{\Lambda}_E \mathbf{w} \right\}.$$

By substituting the above equations into (6.21) and introducing slack variables $\{x_K, y_K\}$, the constraint for LU K in (6.20b) is transformed into the following deterministic form:

$$\mathrm{BTI}_2 \triangleq \begin{cases} \mathrm{Tr}\left(\mathbf{U}_K\right) + \sqrt{2\ln(1/\bar{\rho})}x_K - \ln(\bar{\rho})y_K + u_K \leq 0 \\ \left\| \begin{array}{c} a_K |v|^2 M \mathrm{vec}(\mathbf{\Lambda}_K \mathbf{W}) \\ a_E M \mathrm{vec}(\mathbf{\Lambda}_E \mathbf{W})/\sigma_E^2 \\ \sqrt{2M} \mathbf{\Lambda}_K^{\frac{1}{2}} \left(a_K |v|^2 \mathbf{W} \widehat{\mathbf{H}}_K^{\mathrm{H}} \mathbf{e} - a_K v \mathbf{w} \right) \\ \sqrt{2M} a_E \mathbf{\Lambda}_E^{\frac{1}{2}} \mathbf{W} \widehat{\mathbf{H}}_E^{\mathrm{H}} \mathbf{e}/\sigma_E^2 \end{array} \right\| \leq x_K \\ y_K \geq \lambda_{\max}(\mathbf{U}_K), y_K \geq 0. \end{cases} \tag{6.25}$$

Then, to handle the non-convex power constraint (6.12f), the right hand side of (6.12f) is replaced with its linear lower bound

$$\Xi(\mathbf{z}, t_k) = 2(1 - t_k^t)\mathrm{Re}\left\{ \mathbf{z}^{t,\mathrm{H}} \mathbf{Q}^{\mathrm{H}} \mathbf{g}_k \mathbf{g}_k^{\mathrm{H}} \mathbf{Q} \mathbf{z} \right\} - \frac{(1 - t_k^t)^2 \left| \mathbf{g}_k^{\mathrm{H}} \mathbf{Q} \mathbf{z}^t \right|^2}{(1 - t_k)} \tag{6.26}$$

at feasible point $\{\mathbf{z}^t, t_k^t\}$ by adopting the same first-order Taylor approximation used in Lemma 9.

Therefore, based on (6.23), (6.25) and (6.26) and defining $\mathbf{x} = [x_E, x_K]^{\mathrm{T}}$and $\mathbf{y} = [y_E, y_K]^{\mathrm{T}}$, the Problem (6.12) can be approximated as

$$\max_{R_{\mathrm{sec}},\mathbf{z},\mathbf{w},\mathbf{e},\mathbf{t},a_K,a_E,v,\mathbf{x},\mathbf{y}} R_{\mathrm{sec}} \tag{6.27a}$$

$$\text{s.t.}(6.23), (6.25), (6.12c) - (6.12e), \tag{6.27b}$$

$$|w_k|^2 \leq \Xi(\mathbf{z}, t_k), \forall k \in \mathcal{K}_{-K}. \tag{6.27c}$$

For given $\{\mathbf{e}, a_K, a_E, v\}$, a new variable is introduced as $\mathbf{W} = \mathbf{w}\mathbf{w}^{\mathrm{H}}$ with $\mathrm{rank}(\mathbf{W}) = 1$. However, different from the general SDP, $\mathbf{w}$ and $\mathbf{W}$, coexist in (6.24c). Therefore, the SDR technique is not applicable here. In order to handle this problem, it is assumed $\mathbf{w}$ and $\mathbf{W}$ are two different variables. If $\mathrm{Tr}\,(\mathbf{W}) = \lambda_{\max}(\mathbf{W})$, then $\mathrm{rank}(\mathbf{W}) = 1$. If the obtained $\mathbf{W}$ is not rank one, then $\mathrm{Tr}\,(\mathbf{W}) - \lambda_{\max}(\mathbf{W}) > 0$. Therefore, $\mathrm{Tr}\,(\mathbf{W}) - \lambda_{\max}(\mathbf{W})$ is constrained to be less than a very small real positive number threshold ε to guarantee the rank-1 condition of $\mathbf{W}$, yielding the surrogate constraint of rank-1 constraint as

$$\mathrm{Tr}\,(\mathbf{W}) - \lambda_{\max}(\mathbf{W}) \leq \varepsilon. \tag{6.28}$$

When $\mathrm{rank}(\mathbf{W}) \approx 1$, the relationship between $\mathbf{w}$ and $\mathbf{W}$ is given by the following constraint:

$$-\varepsilon \leq ||\mathbf{w}||^2 - \mathrm{Tr}\,(\mathbf{W}) \leq \varepsilon. \tag{6.29}$$

As for constraint (6.28), since $\lambda_{\max}(\mathbf{W})$ is a convex function of $\mathbf{W}$ [79], the left hand side of (6.28) is concave, which is the difference between a linear function and a convex function. Hence, it needs to construct a convex approximation of constraint (6.28). To address this issue, the following lemma is introduced.

Lemma 13. *Denote by* $\mathbf{v}_{max}$ *the eigenvector corresponding to the maximum eigenvalue of a matrix* $\mathbf{V}$*, it has*

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{v}_{max}\mathbf{v}_{max}^{\mathrm{H}}(\mathbf{Z}-\mathbf{V})\right) &= \mathbf{v}_{max}^{\mathrm{H}}\mathbf{Z}\mathbf{v}_{max} - \mathbf{v}_{max}^{\mathrm{H}}\mathbf{V}\mathbf{v}_{max} \\ &= \mathbf{v}_{max}^{\mathrm{H}}\mathbf{Z}\mathbf{v}_{max} - \lambda_{\max}(\mathbf{V}) \\ &\leq \lambda_{\max}(\mathbf{Z}) - \lambda_{\max}(\mathbf{V})\end{aligned}$$

for any Hermitian matrix $\mathbf{Z}$*.* ■

Let $\mathbf{d}_{\mathbf{W}}^{t}$ be the eigenvector corresponding to the maximum eigenvalue of the feasible point $\mathbf{W}^{t}$. Then, by using Lemma 13, the surrogate convex constraint of (6.28) is given by

$$\mathrm{Tr}\left(\mathbf{W}\right) - \lambda_{\max}(\mathbf{W}^{t}) - \mathrm{Tr}\left(\mathbf{d}_{\mathbf{W}}^{t}\mathbf{d}_{\mathbf{W}}^{t,\mathrm{H}}(\mathbf{W}-\mathbf{W}^{t})\right) \leq \varepsilon. \tag{6.30}$$

Now, constraint (6.29) is considered. By applying the first-order Taylor approximation to $||\mathbf{w}||^2$, it obtains the following convex approximation of the constraint in (6.29) as

$$||\mathbf{w}||^2 - \mathrm{Tr}\left(\mathbf{W}\right) \leq \varepsilon, \tag{6.31a}$$

$$2\mathrm{Re}\left\{\mathbf{w}^{t,\mathrm{H}}\mathbf{w}\right\} - ||\mathbf{w}^{t}||^2 - \mathrm{Tr}\left(\mathbf{W}\right) \geq -\varepsilon. \tag{6.31b}$$

Finally, the subproblem w.r.t., $\{\mathbf{z}, \mathbf{w}, \mathbf{W}, \mathbf{t}\}$ is formulated as

$$\max_{R_{\mathrm{sec}},\mathbf{z},\mathbf{w},\mathbf{W},\mathbf{t},\mathbf{x},\mathbf{y}} R_{\mathrm{sec}} \tag{6.32a}$$

$$\text{s.t.} (6.23), (6.25), (6.12c), (6.12e), \tag{6.32b}$$

$$(6.27c), (6.30), (6.31) \tag{6.32c}$$

$$\mathbf{W} \succeq \mathbf{0}. \tag{6.32d}$$

Problem (6.32) is an SDP and can be solved by the CVX tool [62].

For given $\{\mathbf{w}, a_K, a_E, v\}$, Problem (6.27) with optimization variable $\mathbf{e}$ can be solved by applying the penalty CCP [33], [75], [123] to relax the unit-modulus constraint (6.12d). Comparing with the SDR technique, the penalty CCP method is capable of finding a feasible solution to meet constraint (6.12d). In particular, the constraint of (6.12d) can be relaxed by

$$|e_m^{[j]}|^2 - 2\mathrm{Re}(e_m^{\mathrm{H}} e_m^{[j]}) \leq b_m - 1, 1 \leq m \leq M, \tag{6.33a}$$

$$|e_m|^2 \leq 1 + b_{M+m}, 1 \leq m \leq M, \tag{6.33b}$$

where $e_m^{[j]}$ is any feasible solution and $\mathbf{b} = [b_1, ..., b_{2M}]^{\mathrm{T}}$ are slack vector variables. The proof of (6.33) can be found in [33], [75]. Following the penalty CCP framework, the subproblem for optimizing $\mathbf{e}$ is formulated as

$$\max_{R_{\mathrm{sec}}, \mathbf{e}, \mathbf{x}, \mathbf{y}} R_{\mathrm{sec}} - \lambda^{[j]} ||\mathbf{b}||_1 \tag{6.34a}$$

$$\text{s.t.} (6.23), (6.25), (6.33). \tag{6.34b}$$

Problem (6.34) is an SDP and can be solved by the CVX tool. The algorithm for finding a feasible solution of $\mathbf{e}$ is summarized in Algorithm 6.1.

Algorithm 6.1: Penalty CCP optimization for reflection beamforming optimization

Require: Initialize $\mathbf{e}^{[0]}$, $\gamma^{[0]} > 1$, and set $j = 0$.

repeat
  if $j < J_{max}$ **then**
    Update $\mathbf{e}^{[j+1]}$ by solving Problem (6.34);
    Update $\lambda^{[j+1]} = \min\{\gamma\lambda^{[j]}, \lambda_{max}\}$;
    $j = j + 1$;
  else
    Initialize with a new random $\mathbf{e}^{[0]}$, set $\gamma^{[0]} > 1$ again, and set $j = 0$.
  end if
until $||\mathbf{b}||_1 \leq \chi$ and $||\mathbf{e}^{[j]} - \mathbf{e}^{[j-1]}||_1 \leq \nu$.
Output $\mathbf{e}^{(n+1)} = \mathbf{e}^{[j]}$.

In addition, Problem (6.27) is convex w.r.t. $\{R_{\text{sec}}, v, \mathbf{x}, \mathbf{y}\}$ for given $\{\mathbf{z}, \mathbf{w}, \mathbf{e}, \mathbf{t}, a_K, a_E, \}$, and convex w.r.t. $\{R_{\text{sec}}, a_K, a_E, \mathbf{x}, \mathbf{y}\}$ for given $\{\mathbf{z}, \mathbf{w}, \mathbf{e}, \mathbf{t}, v, \}$. Finally, Problem (6.27) is addressed under the AO framework containing four subproblems. The convergence of the AO framework can be guaranteed due to the fact that each subproblem can obtain a non-decreasing sequence of objective function values.

6.3 ED Model II: Passive Eavesdropper Model

This section focuses on the transmission design for the passive attack, which is more practical and more challenging to address, since the passive ED can hide itself and its CSI is not known [38], [39]. The authors in [129] proposed to exploit the angular information of the ED to combat its passive attack, which is also applicable here. In this section, the cascaded LRL channel $\mathbf{H}_K$ and the channel $\mathbf{H}_{\text{RIS}}$ are assumed to be perfect, which is reasonable due to the fact that the pilot information for channel estimation for LUs is known at the BS.

6.3.1 Average Eavesdropping Rate Maximization

The signal received by the ED is formulated as

$$y_E = \mathbf{h}_E^{\text{H}} \text{diag}(\mathbf{e}^*) \mathbf{H}_{\text{RIS}} \mathbf{w} s + \sigma_E^2.$$

Since the ED is passive, it can only detect the activity of the ED on the line segment between the BS and LU K without knowing its exact location. This detection of a passive attack is based on spectrum sensing [131]. Hence, the average eavesdropping rate is considered which can be computed as follows [129], [132], [133]:

$$R_E^{av}(\mathbf{w}, \mathbf{e}) = \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\mathbf{h}_E\}} \left[\frac{1}{2} \log_2 \left(1 + \frac{1}{\sigma_E^2} \left| \mathbf{h}_E^{\text{H}} \text{diag}(\mathbf{e}^*) \mathbf{H}_{\text{RIS}} \mathbf{w} \right|^2 \right) \right] \text{d}_{D_E}. \quad (6.35)$$

With (6.35), the following optimization problem is formulated:

$$\max_{\mathbf{z},\mathbf{w},\mathbf{e},\mathbf{t}} \left\{ \min_{\forall k \in \mathcal{K}} R_k - R_E^{av}\left(\mathbf{w},\mathbf{e}\right) \right\} \tag{6.36a}$$

$$\text{s.t.} (6.12c) - (6.12f). \tag{6.36b}$$

The main challenge to solve Problem (6.36) is from the average eavesdropping rate containing the integration over D_E and the expectation over $\mathbf{h}_E$. To address this issue, Jensen's inequality is used to construct an upper bound of $R_E^{av}\left(\mathbf{w},\mathbf{e}\right)$ given by

$$\begin{aligned} R_E^{up}\left(\mathbf{w},\mathbf{e}\right) &= \frac{1}{2}\log_2\left(1+\frac{\int_0^{D_K}\mathbb{E}_{\{\mathbf{h}_E\}}\left[\left|\mathbf{h}_E^{\mathrm{H}}\mathrm{diag}(\mathbf{e}^*)\mathbf{H}_{\mathrm{RIS}}\mathbf{w}\right|^2\right]\mathrm{d}_{D_E}}{\sigma_E^2 D_K}\right) \\ &= \frac{1}{2}\log_2\left(1+\frac{1}{\sigma_E^2}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{RIS}}^{\mathrm{H}}\mathrm{diag}(\mathbf{e})\mathbf{R}_E\mathrm{diag}(\mathbf{e}^*)\mathbf{H}_{\mathrm{RIS}}\mathbf{w}\right), \end{aligned} \tag{6.37}$$

where $\mathbf{R}_E = \frac{1}{D_K}\int_0^{D_K}\mathbb{E}_{\{\mathbf{h}_E\}}[\mathbf{h}_E\mathbf{h}_E^{\mathrm{H}}]\mathrm{d}_{D_E}$ that can be computed via one-dimension integration.

According to (6.2), define

$$\overline{\mathbf{h}}_E = \sqrt{\varrho_0\left(\frac{d_{\mathrm{RIS},E}}{d_0}\right)^{-\alpha_{\mathrm{RIS}}}\frac{K_{\mathrm{RIS}}}{1+K_{\mathrm{RIS}}}}\mathbf{h}_E^{\mathrm{LOS}}, \tag{6.38a}$$

$$\mathbf{R}_E = \varrho_0\left(\frac{d_{\mathrm{RIS},E}}{d_0}\right)^{-\alpha_{\mathrm{RIS}}}\frac{1}{1+K_{\mathrm{RIS}}}\mathbf{I}_M, \tag{6.38b}$$

where $\overline{\mathbf{h}}_E$ describes the LoS component and is the mean of channel $\mathbf{h}_E$. Moreover, $\mathbf{R}_E$ is a positive semi-definite covariance matrix representing the spatial correlation characteristics of the non-LoS component. Therefore, it comes to $\mathbf{h}_E \sim \mathcal{CN}(\overline{\mathbf{h}}_E, \mathbf{R}_E)$ [134], and further obtains

$$\begin{aligned} \mathbb{E}_{\{\mathbf{h}_E\}}[\mathbf{h}_E\mathbf{h}_E^{\mathrm{H}}] &= \left[\mathbf{R}_E + \overline{\mathbf{h}}_E\overline{\mathbf{h}}_E^{\mathrm{H}}\right] \\ &= \varrho_0\left(\frac{d_{\mathrm{RIS},E}}{d_0}\right)^{-\alpha_{\mathrm{RIS}}}\left[\frac{1}{1+K_{\mathrm{RIS}}}\mathbf{I}_M + \frac{K_{\mathrm{RIS}}}{1+K_{\mathrm{RIS}}}\mathbf{h}_E^{\mathrm{LOS}}(\mathbf{h}_E^{\mathrm{LOS}})^{\mathrm{H}}\right]. \end{aligned}$$

6.3.2 Proposed Algorithm

By replacing $R_E^{av}(\mathbf{w},\mathbf{e})$ with $R_E^{up}(\mathbf{w},\mathbf{e})$ in the objective function of Problem (6.36), it has

$$\max_{\mathbf{z},\mathbf{w},\mathbf{e},\mathbf{t}} \left\{ \min_{\forall k \in \mathcal{K}} R_k - R_E^{up} \right\} \tag{6.39a}$$

$$\text{s.t.} (6.12c) - (6.12f). \tag{6.39b}$$

Problem (6.39) is still difficult to solve due to the non-convex constraints and objective function, as well as the coupled variables $\mathbf{w}$ and $\mathbf{e}$. Hence, an MM-based AO method is proposed to update $\mathbf{w}$ and $\mathbf{e}$ iteratively. More specifically, by first fixing $\mathbf{e}$, the non-concave objective function w.r.t., $\{\mathbf{z},\mathbf{w},\mathbf{t}\}$ is replaced by its customized concave surrogate function and then solved by the CVX. $\{\mathbf{z},\mathbf{w},\mathbf{t}\}$ are then fixed and the closed-form solution of $\mathbf{e}$ can be found by constructing an easy-to-solve surrogate objective function w.r.t $\mathbf{e}$.

The surrogate functions of $R_k(\mathbf{z},t_k)$ for $\forall k \in \mathcal{K}_{-K}$ are given by $\widehat{R}_k(\mathbf{z},t_k|\mathbf{z}^t,t_k^t) = \widetilde{R}_k(\mathbf{z},t_k|\mathbf{z}^t,t_k^t)$ given in (6.15), and those of $R_K(\mathbf{w},\mathbf{e})$ and $R_E(\mathbf{w},\mathbf{e})$ are given in the following lemma by using the first-order Taylor approximation.

Lemma 14. *Let $\{\mathbf{w}^t,\mathbf{e}^t\}$ be any feasible solution, then $R_K(\mathbf{w},\mathbf{e})$ is lower bounded by a concave surrogate function $\widehat{R}_K\left(\mathbf{w},\mathbf{e}|\mathbf{w}^t,\mathbf{e}^t\right)$ defined by*

$$\widehat{R}_K(\mathbf{w},\mathbf{e}|\mathbf{w}^t,\mathbf{e}^t) = \frac{1}{2}\log_2\left(1 - \frac{q_K^t}{\sigma_K^2} + 2\mathrm{Re}\left\{\frac{q_K}{\sigma_K^2}\right\}\right), \tag{6.40}$$

where $q_K^t = \left|\mathbf{e}^{t,\mathrm{H}}\mathbf{H}_K\mathbf{w}^t\right|^2$ and $q_K = \mathbf{e}^{t,\mathrm{H}}\mathbf{H}_K\mathbf{w}^t\mathbf{w}^{\mathrm{H}}\mathbf{H}_K^{\mathrm{H}}\mathbf{e}$.

Meanwhile, $R_E(\mathbf{w},\mathbf{e})$ is upper bounded by a convex surrogate function $\widehat{R}_E\left(\mathbf{w},\mathbf{e}|\mathbf{w}^t,\mathbf{e}^t\right)$ given by

$$\widehat{R}_E^{up}\left(\mathbf{w},\mathbf{e}|\mathbf{w}^t,\mathbf{e}^t\right) = \frac{1}{2}\log_2\left(1 + \frac{q_E^t}{\sigma_E^2}\right) + \frac{q_E - q_E^t}{2(\sigma_E^2 + q_E^t)\ln 2}, \tag{6.41}$$

where $q_E^t = \mathbf{w}^{t,\mathrm{H}}\mathbf{H}_{\mathrm{RIS}}^{\mathrm{H}}\mathrm{diag}(\mathbf{e}^t)\mathbf{R}_E\mathrm{diag}(\mathbf{e}^{t,*})\mathbf{H}_{\mathrm{RIS}}\mathbf{w}^t$ *and* $q_E = \mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{RIS}}^{\mathrm{H}}\mathrm{diag}(\mathbf{e})\mathbf{R}_E\mathrm{diag}(\mathbf{e}^*)\mathbf{H}_{\mathrm{RIS}}\mathbf{w}$.

Furthermore, the following proposition is given.

The functions $\{\widehat{R}_k, \widehat{R}_K, \widehat{R}_E^{up}\}$ preserve the first-order property of functions $\{R_k, R_K, R_E\}$, respectively. Let's take $\widehat{R}_K$ and R_K as an example

$$\nabla_{\mathbf{w}}\widehat{R}_K(\mathbf{w}, \mathbf{e}^t|\mathbf{w}^t, \mathbf{e}^t)|_{\mathbf{w}=\mathbf{w}^t} = \nabla_{\mathbf{w}}R_K\left(\mathbf{w}, \mathbf{e}^t\right)|_{\mathbf{w}=\mathbf{w}^t},$$
$$\nabla_{\mathbf{e}}\widehat{R}_K(\mathbf{w}^t, \mathbf{e}|\mathbf{w}^t, \mathbf{e}^t)|_{\mathbf{e}=\mathbf{e}^t} = \nabla_{\mathbf{e}}R_K\left(\mathbf{w}^t, \mathbf{e}\right)|_{\mathbf{e}=\mathbf{e}^t}.$$

Proof: See Appendix B in [129]. ■

Giving $\mathbf{e}$, by using (6.15), (6.40), (6.41) and (6.27c), the subproblem of optimizing $\{\mathbf{z}, \mathbf{w}, \mathbf{t}\}$ is formulated as

$$\max_{\mathbf{z},\mathbf{w},\mathbf{t}} \left\{ \min_{\forall k \in \mathcal{K}} \widehat{R}_k - \widehat{R}_E^{up} \right\} \tag{6.42a}$$
$$\text{s.t.}(6.12c), (6.12e), (6.27c). \tag{6.42b}$$

Introducing auxiliary variable r, Problem (6.42) can be transformed into

$$\max_{\mathbf{z},\mathbf{w},\mathbf{t},r} \left\{ r - \widehat{R}_E^{up} \right\} \tag{6.43a}$$
$$\text{s.t.}(6.12c), (6.12e), (6.27c) \tag{6.43b}$$
$$\widehat{R}_k \geq r, \forall k \in \mathcal{K}, \tag{6.43c}$$

which is convex and can be solved by using CVX.

Giving $\{\mathbf{z}, \mathbf{w}, \mathbf{t}\}$ and combining (6.15), (6.40) and (6.41), the subproblem w.r.t., $\mathbf{e}$ is formulated as

$$\max_{\mathbf{e}} \left\{ \min_{\forall k \in \mathcal{K}} \widehat{R}_k - \widehat{R}_E^{up} \right\}, \text{s.t.}(6.12d). \tag{6.44}$$

Problem (6.44) can be solved by transforming it into a SOCP under the penalty CCP method mentioned in Section 6.2.2. However, the penalty CCP method needs to solve a series of SOCP problems which incurs a high computational complexity. In the following, the aim is to derive a low-complexity algorithm containing the closed-form solution of $\mathbf{e}$.

Let $\mathcal{R} = \min_{k=1}^{K-1} \{R_k(\mathbf{z}, t_k)\}$ which is a constant. Then, Problem (6.39) is reformulated as

$$\max_{\mathbf{e}} \left\{\min\{\mathcal{R}, R_K(\mathbf{e})\} - R_E^{up}(\mathbf{e})\right\}, \text{s.t.} (6.12d) \tag{6.45}$$

for the optimization of $\mathbf{e}$.

Before solving Problem (6.45), the following two subproblems are first considered:

$$\mathcal{P}1: \quad \min_{\mathbf{e}} R_E^{up}(\mathbf{e}), \text{s.t.} (6.12d), R_K(\mathbf{e}) \geq \mathcal{R}. \tag{6.46}$$

$$\mathcal{P}2: \quad \max_{\mathbf{e}} R_K(\mathbf{e}) - R_E^{up}(\mathbf{e}), \text{s.t.} (6.12d), R_K(\mathbf{e}) \leq \mathcal{R}. \tag{6.47}$$

Denote the solutions to $\mathcal{P}1$ and $\mathcal{P}2$ as $\mathbf{e}_1^{\#}$ and $\mathbf{e}_2^{\#}$, respectively. In addition, let us denote the objective function value of Problem (6.45) as $obj(\mathbf{e})$, which is a function of $\mathbf{e}$. If $obj(\mathbf{e}_1^{\#}) \geq obj(\mathbf{e}_2^{\#})$, then the optimal solution of Problem (6.45) is given by $\mathbf{e}_1^{\#}$. Otherwise, the optimal solution is $\mathbf{e}_2^{\#}$.

The following lemma shows the solutions of Problem $\mathcal{P}1$.

Lemma 15. *The optimal solution of* $\mathcal{P}1$ *is given by*

$$\mathbf{e}_1^{\#} = \exp\{\mathrm{j} \arg((\lambda_{\max}(\mathbf{A}_E)\mathbf{I} - \mathbf{A}_E + \varrho_1^{opt}\mathbf{A}_K)\mathbf{e}^t)\}, \tag{6.48}$$

where $\mathbf{A}_E = (\mathbf{H}_{\mathrm{RIS}}\mathbf{w}\mathbf{w}^{\mathrm{H}}\mathbf{H}_{\mathrm{RIS}}^{\mathrm{H}}) \odot (\mathbf{R}_E^{\mathrm{T}}/\sigma_E^2)$, $\mathbf{A}_K = \mathbf{H}_K\mathbf{w}\mathbf{w}^{\mathrm{H}}\mathbf{H}_K^{\mathrm{H}}/\sigma_K^2$ *and* ϱ_1^{opt} *is the price introduced by the price mechanism [26].*

The optimal solution of $\mathcal{P}2$ is given by

$$\mathbf{e}_2^{\#} = \exp\{\mathrm{j}\arg(\mathbf{c} + \varrho_2^{opt}(\lambda_{\max}(\mathbf{A}_K)\mathbf{I} - \mathbf{A}_K)\mathbf{e}^t)\}, \tag{6.49}$$

where ϱ_2^{opt} is the price and

$$\mathbf{c} = \frac{1 + d_K^t}{(1 + d_E^t)^2}(\lambda_{\max}(\mathbf{A}_E)\mathbf{I} - \mathbf{A}_E)\mathbf{e}^t + \frac{\mathbf{A}_K\mathbf{e}^t}{1 + d_E^t},$$
$$d_K^t = \mathbf{e}^{t,\mathrm{H}}\mathbf{A}_K\mathbf{e}^t, \qquad d_E^t = \mathbf{e}^{t,\mathrm{H}}\mathbf{A}_E\mathbf{e}^t.$$

Proof: Please refer to Appendix D.3. ■

Since $\mathbf{e}_1^{\#}$ and $\mathbf{e}_2^{\#}$ are the globally optimal solutions of $\mathcal{P}1$ and $\mathcal{P}2$, respectively, hence $\mathbf{e}^{\#}$ is the globally optimal solution of Problem (6.45). The optimal price parameter can be obtained by using the bisection search method detailed in [26].

According to Proposition 6.3.2 and Theorem 1 in [135], the sequence $\{\mathbf{z}^t, \mathbf{w}^t, \mathbf{t}^t, \mathbf{e}^t\}_{n=1,2,3,\ldots}$ obtained in each iteration is guaranteed to converge to the set of stationary points of Problem (6.39). The computational complexity of Problem (6.43) mainly comes from SOC constraints and is given by $\mathcal{O}(\sqrt{2I}(n^3+n^2\sum_{i=1}^{I} a_i^2))$, where I is the number of SOC of size a_i and n is the number of variables [75]. Thus, the computational complexity of Problem (6.43) is $\mathcal{O}(\sqrt{2N}(n^3 + n^2((N-1)^2 + K - 1))$, where $n = N + 2K - 3$. The computational complexity of (6.48) or (6.49) is mainly comes from the eigenvalue operation whose complexity is $\mathcal{O}(N^3)$. Therefore, the total complexity of solving Problem (6.39) at each iteration is given by $\mathcal{O}(\sqrt{2N}(n^3 + n^2((N-1)^2 + K - 1) + N^3)$.

6.4 Numerical Results and Discussions

This section illustrates the performance of the proposed schemes in terms of the secrecy rate and the proposed algorithms in terms of the feasibility rate and complexity. The results are obtained by using a computer with a 1.99 GHz i7-8550U CPU and 16 GB

RAM. Polar coordinate system is used to describe the simulated system setup: The BS is located at (0 m, 0 m) and the RIS is placed at (50 m, 0 m) with elevation angle $\phi = \frac{2\pi}{3}$; K LUs are randomly and uniformly distributed in an area with $D_k \sim \mathcal{U}(20\,\text{m}, 40\,\text{m})$ and $\theta_k \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ for $\forall k \in \mathcal{K}$, where $\mathcal{U}$ is the uniform distribution. The ED is located at (D_E, θ_K) with $D_E \in (0, D_K)$. The pathloss at the distance of 1 m is -30 dB, the pathloss exponents are set to $\alpha_{\text{BS}} = \alpha_{\text{RIS}} = 2.2$ and the Rician factor is 5. The transmit power budget at the BS is $P_{\max} = 30$ dBm and the noise powers are $\{\sigma_i^2 = -90\ \text{dBm}\}_{\forall i \in \mathcal{K}_{+E}}$. For the statistical CSI error model, the variances of $\{\triangle_i^K, \triangle_i^E\}_{\forall i \in \mathcal{K}_{-K}}$ are defined as $\{\varepsilon_{K,i}^2 = \delta_K^2 ||\mathbf{h}_K^* \odot \mathbf{h}_i||_2^2, \varepsilon_{E,i}^2 = \delta_E^2 ||\mathbf{h}_E^* \odot \mathbf{h}_i||_2^2\}_{\forall i \in \mathcal{K}_{-K}}$, where $\delta_K \in [0, 1)$ and $\delta_E \in [0, 1)$ measure the relative amount of CSI uncertainties. In addition, the outage probability of secrecy rate is $\rho = 0.05$.

6.4.1 Robust Secrecy Rate in ED Model I

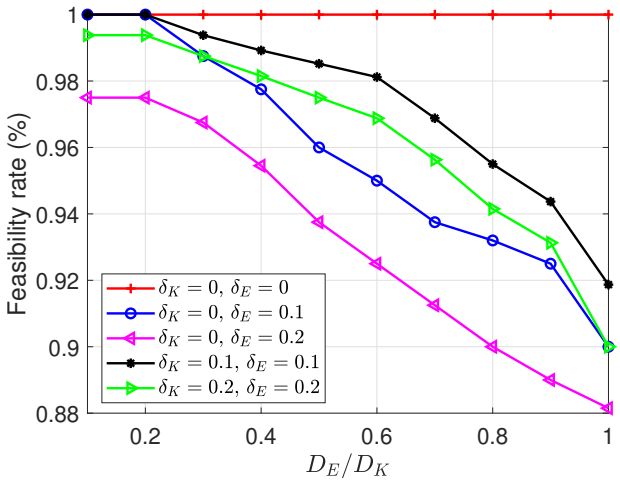
In order to verify the performance of the proposed outage constrained beamforming in the AAUC, the case of $N = 8$ and $K = 5$ is simulated. For comparison, the "Non-robust" is also considered as the benchmark scheme, in which the estimated cascaded LRL and LRE channels are naively regarded as perfect channels, resulting in the following problem

$$\max_{\mathbf{z},\mathbf{w},\mathbf{e},\mathbf{t}} \left\{ \min_{\forall k \in \mathcal{K}} R_k - R_E \right\} \tag{6.51a}$$

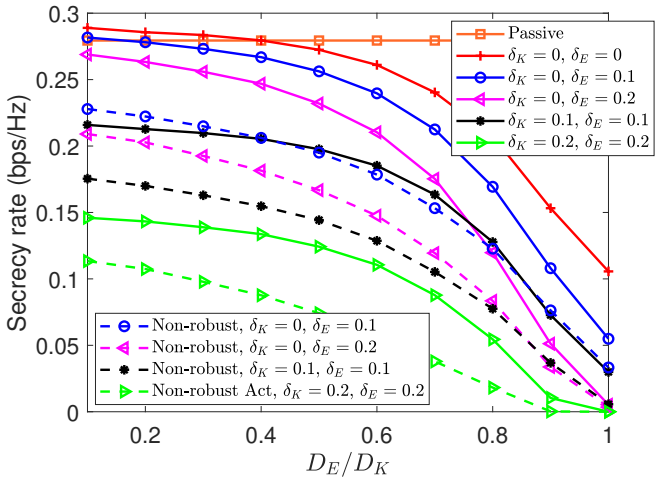
$$\text{s.t.} (6.12c) - (6.12f), \tag{6.51b}$$

where $R_K(\mathbf{e}) = \frac{1}{2}\log_2(1 + |\mathbf{e}^{\text{H}}\widehat{\mathbf{H}}_K\mathbf{w}|^2/\sigma_K^2)$ and $R_E(\mathbf{e}) = \frac{1}{2}\log_2(1 + |\mathbf{e}^{\text{H}}\widehat{\mathbf{H}}_E\mathbf{w}|^2/\sigma_E^2)$. Problem (6.51) can be solved by using the proposed low-complexity algorithm used to solve Problem (6.36).

Fig. 6.3 investigates the feasibility rate and the maximum secrecy rate versus the distance of the ED, in which the coordinate of X-axis is set to the ratio of D_E/D_K. The feasibility rate is defined as the ratio of the number of channel realizations that have



(a) Feasibility rate



(b) Secrecy rate

Figure 6.3: Performance versus D_E/D_K under $N = 8$, $M = 32$ and $K = 5$.

a feasible solution to the outage constrained problem of (6.12) to the total number of channel realizations. It is observed from Fig. 6.3(a) that the closer the ED is to LU K, the lower the feasibility rate will be, which means that the location of the eavesdropper imposes a great threat to the security system. From Fig. 6.3(b), it can be seen that the secrecy rate drops fast when the ED approaches LU K, and this secrecy rate reduces to zero when the channel error is large. At this situation, the whole system is no longer suitable for secure communication.

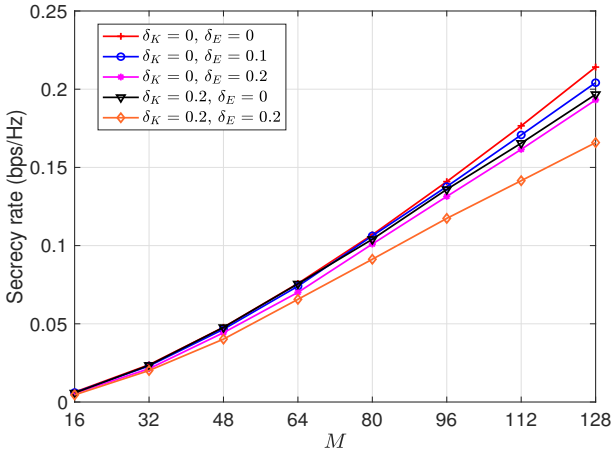


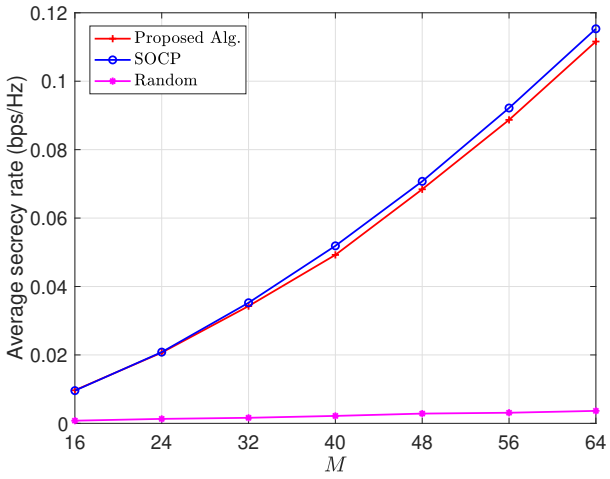
Figure 6.4: Secrecy rate versus M under N = 8 and K = 5.

Next, the performance versus the size of the RIS, i.e., M, is verified in Fig. 6.4. Assume that the ED is located at $D_E/D_K = 0.5$. In Fig. 6.4, the case of $\delta_K = \delta_E = 0$ is regarded as the perfect cascaded CSI case, and its maximum secrecy rate increases with M, which is consistent with that of Fig. 6 in [29]. The performance of $\delta_K = \delta_E = 0$ can be used as the performance upper bound of the proposed outage constrained beamforming method. Furthermore, the maximum secrecy rate is obviously degraded with enlarged channel uncertainty levels. In addition, it is observed that black line of $\{\delta_K = 0.2, \delta_E = 0\}$ outperforms the pink line of $\{\delta_K = 0, \delta_E = 0.2\}$, which means that the negative impact of cascaded LRL channel error on secrecy rate is higher than that of the cascaded LRE channel error.

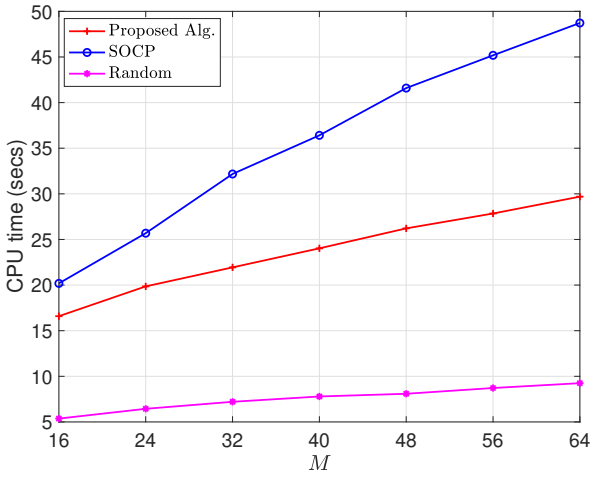
6.4.2 Average Secrecy Rate in ED Model II

This subsection evaluates the performance of the proposed scheme under the passive attack. In order to evaluate the performance of the proposed low-complexity algorithm, two benchmark algorithms are considered and given by: 1) The "SOCP" scheme, in which the CVX tool is used to solve the SOCP version of Problem (6.44). 2) The "Random" scheme, in which the phases of the reflecting elements are randomly generated.

Fig. 6.5 illustrates the performance in terms of the average secrecy rate and the



(a) Average secrecy rate

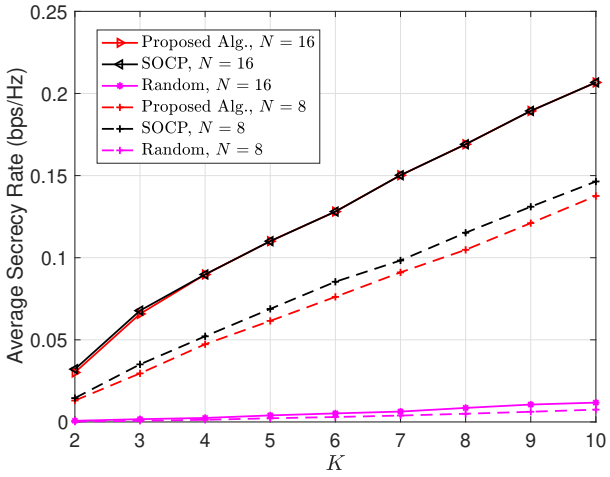


(b) Average CPU time

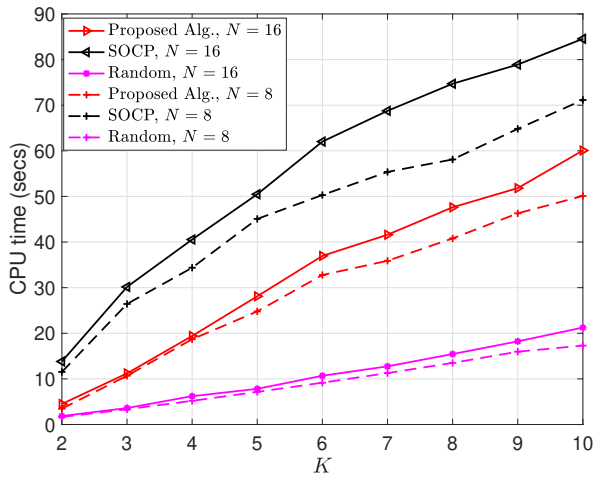
Figure 6.5: Performance versus M under $N = 8$ and $K = 5$.

computational complexity under the system setting of $N = 8$ and $K = 5$. It is observed from Fig. 6.5(a) that the proposed algorithm with semi-closed-form solution is almost the same as that of the SOCP-based algorithm with global optimal solution, and both of them outperform the scheme with random reflecting phases. Moreover, increasing the number of reflecting elements at the RIS can significantly enhance the average secrecy rate of the system. To evaluate the computational complexity of the algorithms, Fig. 6.5(b) describes the CPU time consumption required for these three algorithms. It can be seen that the proposed algorithm with closed-form solution requires much less CPU

running time than the SOCP-based scheme. In addition, the CPU running time of the SCOP-based algorithm is scaled with M, but the proposed algorithm is not sensitive to M, due to the fact that the computational complexity of the SOCP depends on M, while that of the closed-form solution does not.



(a) Average secrecy rate



(b) Average CPU time

Figure 6.6: Performance versus K under $M = 64$.

Finally, Fig. 6.6 illustrates the performance versus the number of users when $M = 64$. It can be obtained the same conclusion as above that the proposed algorithm has the same performance of the SOCP-based algorithm but consumes less CPU running time.

6.5 Summary

This chapter has proposed a two-phase RIS-aided communication system to realize the secure communication under the active attacks and passive eavesdropping. In order to address the cascaded channel error caused by the active attacks, the secrecy rate is maximized subject to secrecy rate outage probability constraints, which has been tackled by using the BTI. For the case of the partial CSI of the ED, average secrecy rate maximization problem was considered, which is addressed by the proposed low-complexity algorithm. It was shown that the negative effect of the ED's channel error is greater than that of the LU. In addition, the number of elements on the RIS has a negative impact on system performance when the channel error is large. This conclusion provides an engineering insight for the careful selection of the number of elements at the RIS.

Chapter 7

Conclusion

The purpose of this thesis was to develop optimization approaches for transmit beamforming designs of the RIS-aided wireless communication systems under different CSI challenges. Specifically, four CSI challenges are: worst channel condition in multicast transmissions, CSI uncertainty caused by the presence of random blockages, CSI uncertainty caused by the cascaded channel error and CSI uncertainty caused by the presence of eavesdropper. For each practical CSI issue, this thesis proposes effective or low-complexity beamforming algorithms from two aspects, and verifies the excellent performance of the designed beamforming algorithms in terms of computational complexity, spectrum efficiency, energy efficiency and robustness from theory and simulation.

Chapter 3 started by considering the problem of minimum rate maximization in downlink multigroup multicast communication systems under perfect CSI. To address this problem, two efficient algorithms were developed based on MM framework. The low-complexity algorithm with closed-form solutions was proved to have the same performance as the general SOCP-based algorithm through simulations. Moreover, the gains in terms of the spectral and energy efficiency achieved by employing RIS were demonstrated through simulations for a variety of system parameters.

Chapter 4 addressed the problem of maximum outage probability minimization in the

mmWave systems under CSI uncertainty caused by the presence of random blockages. It started by approximating the outage probability objective function by a continuous function. Then, robust stochastic optimization algorithms were developed based on the SMM and SSCA methods. The convergence of the developed algorithms to the set of stationary points of the original stochastic problems was established. Simulation results showed that the proposed robust beamforming for RIS-aided systems can effectively compensate for the performance loss caused by the presence of random blockages, especially when the blockage probability is high.

Chapter 5 developed robust beamforming designs to address the CSI uncertainty caused by channel estimation error. It started by characterising the cascaded BS-RIS-user channel error with two models: the bounded CSI error model and the statistical CSI error model. Then, the non-convex worst-case rate constraints under the bounded CSI error model were approximated by using the S-procedure and the non-convex rate outage probability constraints under the statistical CSI error model were addressed by utilizing the BTI. The significant performance gains achieved of robust transmission were demonstrated through simulations, and the negative impact of the CBRUT error on the system performance was greater than that of the direct CSI error.

In Chapter 6, the robust secrecy transmission was investigated to the physical layer security systems. It started by addressing the robust transmission designs under the CSI uncertainty caused by the cascaded channel estimation error when the eavesdropper launched an active attack. Then, a robust beamforming method was developed based on the average secrecy rate maximization with only the knowledge of large-scale eavesdropper CSI when the eavesdropper conducted passive eavesdropping. Numerical results demonstrated that the negative effect of the eavesdropper's channel error is larger than that of the legitimate user.

7.1 Future Work

In conclusion of this these, some future research directions are proposed.

- **Channel Estimation:** The performance gain provided by the RIS relies heavily on the accuracy of the CSI. The channel estimation methods of the RIS-aided system have been widely investigated. However, the pilot overhead of the channel estimation methods of the sub-6G systems in the existing literature is prohibitively high and scales with the number of RIS elements. While, the estimation methods for the sparse channel of the mmWave system in the existing literature is computational complex or has error propagation. Therefore, the channel estimation method with low pilot overhead, low computational complexity and low estimation error is still worth exploring.

- **Double/multi-RIS:** Most of the existing works have considered quasi passive beamforming designs and channel estimation schemes in systems with a single RIS. In some scenarios, however, it may be convenient to enable the transmission of signals through reflections from multiple RISs in order to route the signals and bypass the blocking objects in a smart manner, directly at the electromagnetic level (electromagnetic routing). Thus, double/multiple RISs may be utilized to realize a blockage-free communication network via multiple signal reflections. In addition, existing research works on double/multi-RIS aided systems often ignore the impact of the secondary reflections among the RISs, which may be a reasonable approximation if the RISs are in the far-field of each other but it may not hold anymore if the RISs are closely located. Finally, the analysis and design of multi-RIS communications at high frequency bands is an open research issue as well.

- **Multifunction RISs:** In the existing literature, most of the research works concentrate on RISs that operate as anomalous reflectors or as reflecting lenses. An RIS, however, can realize multiple signal transformations depending on how the scattering matrix (or equivalently the surface impedance) is designed. Recently,

notably, a few research attempts have been made to design RISs that operate as anomalous refracting mirrors or as anomalous refracting lenses as well RISs that can simultaneously realize reflections and re-fractions in order to guarantee omni-coverage performance. Multifunction RISs are an emerging research topic, and the corresponding modeling, performance evaluation, and optimization are still at its infancy.

Appendix A

Appendix of Chapter 3

A.1 The proof of Lemma 1

It performs some equivalent transformations of the rate expression (3.6) to show its hidden convexity, as follows

$$\begin{aligned} R_k\left(\mathbf{F},\mathbf{e}\right) &= \log_2\left(1+\frac{|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g|^2}{\sum_{i\neq g}^{G}|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2+\sigma_k^2}\right) \\ &= \log_2\left(1+r_{k,-g}^{-1}|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g|^2\right) \\ &= -\log_2\left(1-\left(r_{k,-g}+|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g|^2\right)^{-1}|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g|^2\right) \\ &= -\log_2\left(1-r_k^{-1}|t_k|^2\right), \end{aligned} \tag{A.1}$$

where $t_k=\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g$, $r_k=r_{k,-g}+|t_k|^2$, and $r_{k,-g}=\sum_{i\neq g}^{G}|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2+\sigma_k^2$.

Denoting $R_k(t_k,r_k)$ as the last equation expression of $R_k\left(\mathbf{F},\mathbf{e}\right)$ in (A.1), $R_k(t_k,r_k)$ is jointly convex in $\{t_k,r_k\}$ [136], thus its lower bound surrogate function could be obtained by the first-order approximation, e.g.,

$R_k\left(t_k,r_k\right)$

$$
\begin{aligned}
&\geq R_k\left(t_k^t, r_k^t\right) + \frac{\partial R_k}{\partial t_k}|_{t_k=t_k^t}\left(t_k - t_k^t\right) + \frac{\partial R_k}{\partial t_k^*}|_{t_k^*=t_g^{t,*}}\left(t_k^* - t_k^{t,*}\right) + \frac{\partial R_k}{\partial r_k}|_{r_k=r_k^t}\left(r_k - r_k^t\right) \\
&= R_k\left(t_k^t, r_k^t\right) + 2\mathrm{Re}\left\{\frac{t_k^{t,*}(t_k - t_k^t)}{r_k^t - |t_k^t|^2}\right\} - \frac{|t_k^t|^2(r_k - r_k^t)}{r_k^t(r_k^t - |t_k^t|^2)} \\
&= R_k\left(t_k^t, r_k^t\right) + 2\mathrm{Re}\left\{\frac{t_k^{t,*}}{r_k^t - |t_k^t|^2}t_k\right\} - \frac{|t_k^t|^2}{r_k^t(r_k^t - |t_k^t|^2)}r_k - \frac{|t_k^t|^2}{r_k^t - |t_k^t|^2}.
\end{aligned} \tag{A.2}
$$

Undo $t_k = \mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g$, $t_k^t = (\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g^t$, $r_k = \sum_{i=1}^G |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2 + \sigma_k^2$, and $r_k^t = \sum_{i=1}^G |(\mathbf{e}^t)^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i^t|^2 + \sigma_k^2$, and substitute them into the right hand side of the last equation in (A.2), it has

$$
\begin{aligned}
R_k(\mathbf{F}, \mathbf{e}) &\geq R_k\left(\mathbf{F}^t, \mathbf{e}^t\right) + 2\mathrm{Re}\left\{a_k\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g\right\} - \frac{|t_k^t|^2}{r_k^t - |t_k^t|^2} - b_k\sum_{i=1}^G |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2 - b_k\sigma_k^2 \\
&= \mathrm{const}_k + 2\mathrm{Re}\left\{a_k\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_g\right\} - b_k\sum_{i=1}^G |\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{f}_i|^2 \\
&= \widetilde{R}_k(\mathbf{F}, \mathbf{e}).
\end{aligned} \tag{A.3}
$$

Hence, the proof is completed.

A.2 The proof of Theorem 1

The monotonic property of the objective function value sequence $\{F(\mathbf{F}^t, \mathbf{e}^t)\}$ of Algorithm 3.1 can be guaranteed by (3.18). In addition, the sequence $\{\mathbf{F}^t, \mathbf{e}^t\}$ generated at each iteration of Algorithm 3.1 converges to a stable point as $t \to \infty$ because $\mathbf{F}^t$ and $\mathbf{e}^t$ are bounded in their feasible sets $\mathcal{S}_F$ and $\mathcal{S}_e$, respectively [137]. Denote by $\{\mathbf{F}^o, \mathbf{e}^o\}$ the converged solution. In the following, it is proved that $\{\mathbf{F}^o, \mathbf{e}^o\}$ is the KKT point based on the fact that all the locally optimal solutions (including the globally optimal solution) of a nonconvex optimization problem should satisfy the KKT optimality conditions [79].

Firstly, the Lagrangian of Problem (3.13) is given by

$$\mathcal{L}(\mathbf{F},\boldsymbol{\gamma},\boldsymbol{\lambda}^{(1)},\lambda^{(2)})$$

$$= \sum_{g=1}^{G} \gamma_g - \sum_{g=1}^{G} \sum_{k \in \mathcal{K}_g} \lambda_k^{(1)} (\gamma_g - \widetilde{R}_k \left(\mathbf{F}, \mathbf{e}^o | \mathbf{F}^o, \mathbf{e}^o\right)) - \lambda^{(2)} (\mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}} \mathbf{F}\right) - P_{\mathrm{T}})$$

where $\boldsymbol{\lambda}^{(1)} = [\lambda_1^{(1)}, ..., \lambda_K^{(1)}]$ and $\lambda^{(2)}$ are the dual variables. Since $\mathbf{F}^o$ is the globally optimal solution of Problem (3.13), there must exist a $\boldsymbol{\lambda}^{(1),o}$ and $\lambda^{(2),o}$ satisfying the following partial KKT conditions:

$$\sum_{g=1}^{G} \sum_{k \in \mathcal{K}_g} \lambda_k^{(1),o} \nabla_{\mathbf{F}^*} \widetilde{R}_k \left(\mathbf{F}, \mathbf{e}^o | \mathbf{F}^o, \mathbf{e}^o\right) |_{\mathbf{F}=\mathbf{F}^o} - \lambda^{(2),o} \mathbf{F}^o = \mathbf{0}, \tag{A.4}$$

$$\lambda_k^{(1),o} (\gamma_g - \widetilde{R}_k \left(\mathbf{F}^o, \mathbf{e}^o | \mathbf{F}^o, \mathbf{e}^o\right)) = 0, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}, \tag{A.5}$$

$$\lambda^{(2),o} (\mathrm{Tr}\left(\mathbf{F}^{\mathrm{H},o} \mathbf{F}^o\right) - P_{\mathrm{T}}) = 0. \tag{A.6}$$

According to the Assumptions (A1) and (A3), it has

$$\widetilde{R}_k \left(\mathbf{F}^o, \mathbf{e}^o | \mathbf{F}^o, \mathbf{e}^o\right) = R_k \left(\mathbf{F}^o, \mathbf{e}^o\right), \tag{A.7}$$

$$\nabla_{\mathbf{F}^*} \widetilde{R}_k \left(\mathbf{F}, \mathbf{e}^o | \mathbf{F}^o, \mathbf{e}^o\right) |_{\mathbf{F}=\mathbf{F}^o} = \nabla_{\mathbf{F}^*} R_k \left(\mathbf{F}, \mathbf{e}^o\right) |_{\mathbf{F}=\mathbf{F}^o}. \tag{A.8}$$

By substituting (A.8) and (A.7) into (A.4) and (A.5) respectively, it comes to

$$\sum_{g=1}^{G} \sum_{k \in \mathcal{K}_g} \lambda_k^{(1),o} \nabla_{\mathbf{F}^*} R_k \left(\mathbf{F}, \mathbf{e}^o\right) |_{\mathbf{F}=\mathbf{F}^o} - \lambda^{(2),o} \mathbf{F}^o = \mathbf{0}, \tag{A.9}$$

$$\lambda_k^{(1),o} (\gamma_g - R_k \left(\mathbf{F}^o, \mathbf{e}^o\right)) = 0, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}. \tag{A.10}$$

Then, $\mathbf{e}^o$ is the locally optimal solution of Problem (3.16) and satisfies the following KKT conditions:

$$\sum_{g=1}^{G} \sum_{k \in \mathcal{K}_g} \xi_k^{(1),o} \nabla_{\mathbf{e}^*} \widetilde{R}_k \left(\mathbf{F}^o, \mathbf{e} | \mathbf{F}^o, \mathbf{e}^o\right) |_{\mathbf{e}=\mathbf{e}^o} -$$

$$\sum_{m=1}^{M} \xi_m^{(2),o} (\nabla_{\mathbf{e}^*} |e_m|)|_{\mathbf{e}=\mathbf{e}^o} - \xi_{M+1}^{(2),o} (\nabla_{\mathbf{e}^*} e_{M+1})|_{\mathbf{e}=\mathbf{e}^o} = \mathbf{0}, \tag{A.11}$$

$$\xi_k^{(1),o} (\kappa_g - \widetilde{R}_k \left(\mathbf{F}^o, \mathbf{e}^o | \mathbf{F}^o, \mathbf{e}^o\right)) = 0, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}, \tag{A.12}$$

$$\xi_m^{(2),o}(|e_m^o| - 1) = 0, 1 \leq m \leq M, \xi_{M+1}^{(2),o}(e_{M+1}^o - 1) = 0, \tag{A.13}$$

where $\boldsymbol{\xi}^{(1),o} = [\xi_1^{(1),o}, ..., \xi_K^{(1),o}]$ and $\xi^{(2),o}$ are the optimal Lagrange multipliers.

Furthermore, it can be readily checked that

$$\nabla_{\mathbf{e}^*} \widetilde{R}_k \left(\mathbf{F}^o, \mathbf{e} | \mathbf{F}^o, \mathbf{e}^o\right)|_{\mathbf{e}=\mathbf{e}^o} = \nabla_{\mathbf{e}^*} R_k \left(\mathbf{F}^o, \mathbf{e}\right)|_{\mathbf{e}=\mathbf{e}^o}. \tag{A.14}$$

By substituting (A.14) into (A.11), it comes to

$$\sum_{g=1}^{G} \sum_{k \in \mathcal{K}_g} \xi_k^{(1),o} \nabla_{\mathbf{e}^*} R_k \left(\mathbf{F}^o, \mathbf{e}\right)|_{\mathbf{e}=\mathbf{e}^o} - \xi_{M+1}^{(2),o} (\nabla_{\mathbf{e}^*} e_{M+1})|_{\mathbf{e}=\mathbf{e}^o} - \sum_{m=1}^{M} \xi_m^{(2),o} (\nabla_{\mathbf{e}^*} |e_m|)|_{\mathbf{e}=\mathbf{e}^o} = \mathbf{0}, \tag{A.15}$$

Now, it moves to Problem (3.8). The general equivalent problem of the max-min Problem (3.8) is given by

$$\begin{aligned} \max_{\mathbf{F},\mathbf{e},\mathbf{r}} \ & \sum_{g=1}^{G} r_g \\ \text{s.t.} \ & \mathbf{F} \in \mathcal{S}_F, \mathbf{e} \in \mathcal{S}_e \\ & R_k \left(\mathbf{F}, \mathbf{e}\right) \geq r_g, \forall k \in \mathcal{K}_g, \forall g \in \mathcal{G}. \end{aligned} \tag{A.16}$$

where $\mathbf{r} = [r_1, ..., r_G]^{\mathrm{T}}$ are auxiliary variables. It can be readily verified that the set of equations (A.9), (A.15), (A.10), (A.6), and (A.13) constitute exactly the KKT conditions of Problem (A.16).

Hence, the proof is completed.

A.3 The proof of Theorem 3

Since $f_g(\mathbf{F})$ is twice differentiable and concave, a quadratic surrogate function is proposed to minorize $f_g(\mathbf{F})$, as follows

$$f_g(\mathbf{F}) \geq f_g(\mathbf{F}^t) + 2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{D}_g^{\mathrm{H}}(\mathbf{F}-\mathbf{F}^t)\right]\right\} + \mathrm{Tr}\left((\mathbf{F}-\mathbf{F}^t)^{\mathrm{H}}\mathbf{M}_g(\mathbf{F}-\mathbf{F}^t)\right) \tag{A.17}$$

where matrices $\mathbf{D}_g \in \mathbb{C}^{N\times N}$ and $\mathbf{M}_g \in \mathbb{C}^{N\times N}$ are determined to satisfy Assumptions (A1)-(A4).

Note that Assumptions (A1) and (A4) are already satisfied. Then it comes to prove that Assumption (A3) also holds. Let $\widetilde{\mathbf{F}}$ be a matrix belonging to $\mathcal{S}_F$. The directional derivative of the right hand side of (A.17) at $\mathbf{F}^t$ with direction $\widetilde{\mathbf{F}}-\mathbf{F}^t$ is given by:

$$2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{D}_g^{\mathrm{H}}(\widetilde{\mathbf{F}}-\mathbf{F}^t)\right]\right\}. \tag{A.18}$$

The directional derivative of $f_g(\mathbf{F})$ is

$$2\mathrm{Re}\left\{\mathrm{Tr}\left[\sum_{k\in\mathcal{K}_g} g_k(\mathbf{F}^t)(\mathbf{C}_k^{\mathrm{H}} - (\mathbf{F}^t)^{\mathrm{H}}\mathbf{B}_k)(\widetilde{\mathbf{F}}-\mathbf{F}^t)\right]\right\}, \tag{A.19}$$

where $g_k(\mathbf{F}^t)$ is defined in (3.28).

In order to satisfy Assumption (A3), the two directional derivatives (A.18) and (A.19) must be equal, which means

$$\mathbf{D}_g = \sum_{k\in\mathcal{K}_g} g_k(\mathbf{F}^t)(\mathbf{C}_k - \mathbf{B}_k^{\mathrm{H}}\mathbf{F}^t). \tag{A.20}$$

Now it proceeds to prove that Assumption (A2) also holds. If surrogate function $\widetilde{f}_g(\mathbf{F}|\mathbf{F}^t)$ is a lower bound for each linear cut in any direction, Assumption (A2) could

be satisfied. Let $\mathbf{F} = \mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t), \forall \gamma \in [0, 1]$. Then, it suffices to show

$$f_g(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t)) \geq f_g(\mathbf{F}^t) + 2\gamma \mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{D}_g^{\mathrm{H}}(\widetilde{\mathbf{F}} - \mathbf{F}^t)\right]\right\} + \gamma^2 \mathrm{Tr}\left[(\widetilde{\mathbf{F}} - \mathbf{F}^t)^{\mathrm{H}} \mathbf{M}_g (\widetilde{\mathbf{F}} - \mathbf{F}^t)\right], \tag{A.21}$$

Let us define $L_g(\gamma) = f_g(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))$,and $l_k(\gamma) = \widetilde{R}_k(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))$. Now, a sufficient condition for (A.21) to hold is that the second derivative of the right hand side of (A.21) is lower than or equal to the second derivative of the left hand side of (A.21) for $\forall \gamma \in [0, 1]$ and $\forall \widetilde{\mathbf{F}}, \forall \mathbf{F}^t \in \mathcal{S}_F$, which is formulated as follows

$$\frac{\partial^2 L_g(\gamma)}{\partial \gamma^2} \geq 2\mathrm{Tr}\left[(\widetilde{\mathbf{F}} - \mathbf{F}^t)^{\mathrm{H}} \mathbf{M}_g (\widetilde{\mathbf{F}} - \mathbf{F}^t)\right]. \tag{A.22}$$

In order to calculate the left hand side of (A.22), the first-order derivative is first calculated, as follows

$$\frac{\partial L_g(\gamma)}{\partial \gamma} = \sum_{k \in \mathcal{K}_g} g_k(\gamma) \nabla_\gamma l_k(\gamma), \tag{A.23}$$

where

$$\begin{aligned}
g_k(\gamma) &= \frac{\exp\{-\mu_g l_k(\gamma)\}}{\sum_{k \in \mathcal{K}_g} \exp\{-\mu_g l_k(\gamma)\}}, k \in \mathcal{K}_g, \\
\nabla_\gamma l_k(\gamma) &= 2\mathrm{Re}\Big\{\mathrm{Tr}\left[\mathbf{C}_k^{\mathrm{H}}(\widetilde{\mathbf{F}} - \mathbf{F}^t)\right] - \mathrm{Tr}\left((\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))^{\mathrm{H}} \mathbf{B}_k (\widetilde{\mathbf{F}} - \mathbf{F}^t)\right)\Big\} \\
&= 2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{Q}_k^{\mathrm{H}}(\widetilde{\mathbf{F}} - \mathbf{F}^t)\right]\right\} \\
&= 2\mathrm{Re}\left\{\mathbf{q}_k^{\mathrm{H}} \mathbf{f}\right\}, \\
\mathbf{Q}_k^{\mathrm{H}} &= \mathbf{C}_k^{\mathrm{H}} - (\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))^{\mathrm{H}} \mathbf{B}_k, \\
\mathbf{q}_k &= \mathrm{vec}(\mathbf{Q}_k), \\
\mathbf{f} &= \mathrm{vec}(\widetilde{\mathbf{F}} - \mathbf{F}^t),
\end{aligned}$$

Then, the second-order derivative is derived as

$$\frac{\partial^2 L_g(\gamma)}{\partial \gamma^2} = \sum_{k\in\mathcal{K}_g} \left(g_k(\gamma)\nabla_\gamma^2 l_k(\gamma) - \mu_g g_k(\gamma)\nabla_\gamma l_k(\gamma) \left(\nabla_\gamma l_k(\gamma)\right)^{\mathrm{T}} \right) + \mu_g \left(\sum_{k\in\mathcal{K}_g} g_k(\gamma)\nabla_\gamma l_k(\gamma) \right) \left(\sum_{k\in\mathcal{K}_g} g_k(\gamma)\nabla_\gamma l_k(\gamma) \right)^{\mathrm{T}}, \tag{A.24}$$

where

$$\begin{aligned} \nabla_\gamma^2 l_k(\gamma) &= -2\mathrm{Tr}\left[(\widetilde{\mathbf{F}} - \mathbf{F}^t)^{\mathrm{H}} \mathbf{B}_k (\widetilde{\mathbf{F}} - \mathbf{F}^t) \right] \\ &= -2\mathbf{f}^{\mathrm{H}} (\mathbf{I} \otimes \mathbf{B}_k)\mathbf{f}. \end{aligned}$$

This work reformulates $\frac{\partial^2 L_g(\gamma)}{\partial \gamma^2}$ in (A.24) into a quadratic form of $\mathbf{f}$, as follows

$$\frac{\partial^2 L_g(\gamma)}{\partial \gamma^2} = \begin{bmatrix} \mathbf{f}^{\mathrm{H}} & \mathbf{f}^{\mathrm{T}} \end{bmatrix} \mathbf{\Phi} \begin{bmatrix} \mathbf{f} \\ \mathbf{f}^* \end{bmatrix},$$

where $\mathbf{\Phi}$ is given in

$$\mathbf{\Phi}_g = \sum_{k\in\mathcal{K}_g} \left(g_k(\gamma) \begin{bmatrix} -\mathbf{I}\otimes\mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}\otimes\mathbf{B}_k^{\mathrm{T}} \end{bmatrix} - \mu_g g_k(\gamma) \begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_k^* \end{bmatrix} \begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_k^* \end{bmatrix}^{\mathrm{H}} \right) + \mu_g \begin{bmatrix} \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k \\ \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^* \end{bmatrix} \begin{bmatrix} \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k \\ \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^* \end{bmatrix}^{\mathrm{H}}. \tag{A.25}$$

The right hand side of (A.22) is also manipulated into a quadratic form of $\mathbf{f}$ by using vectorization operation $\mathrm{Tr}[\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{C}] = \mathrm{vec}^{\mathrm{T}}(\mathbf{A})(\mathbf{I}\otimes\mathbf{B})\mathrm{vec}(\mathbf{C})$ [138], as follows

$$2\mathrm{Tr}\left[(\widetilde{\mathbf{F}} - \mathbf{F}^t)^{\mathrm{H}} \mathbf{M}_g (\widetilde{\mathbf{F}} - \mathbf{F}^t) \right] = \begin{bmatrix} \mathbf{f}^{\mathrm{H}} & \mathbf{f}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{I}\otimes\mathbf{M}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\otimes\mathbf{M}_g^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{f}^* \end{bmatrix}.$$

Then, (A.22) is equivalent to

$$\left[\begin{array}{cc} \mathbf{f}^{\mathrm{H}} & \mathbf{f}^{\mathrm{T}} \end{array}\right] \mathbf{\Phi}_g \left[\begin{array}{c} \mathbf{f} \\ \mathbf{f}^* \end{array}\right] \geq \left[\begin{array}{cc} \mathbf{f}^{\mathrm{H}} & \mathbf{f}^{\mathrm{T}} \end{array}\right] \left[\begin{array}{cc} \mathbf{I} \otimes \mathbf{M}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{M}_g^{\mathrm{T}} \end{array}\right] \left[\begin{array}{c} \mathbf{f} \\ \mathbf{f}^* \end{array}\right],$$

where it needs to find an $\mathbf{M}_g$ that satisfies

$$\mathbf{\Phi}_g \succeq \left[\begin{array}{cc} \mathbf{I} \otimes \mathbf{M}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{M}_g^{\mathrm{T}} \end{array}\right].$$

For convenience, $\mathbf{M}_g = \alpha_g \mathbf{I} = \lambda_{\min}\left(\mathbf{\Phi}_g\right)\mathbf{I}$ is chosen. Finally, (A.17) is equivalent to

$$\begin{aligned} f_g(\mathbf{F}) &\geq f_g(\mathbf{F}^t) + 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{D}_g^{\mathrm{H}}(\mathbf{F}-\mathbf{F}^t)\right)\right\} + \alpha_g \mathrm{Tr}\left[(\mathbf{F}-\mathbf{F}^t)^{\mathrm{H}}(\mathbf{F}-\mathbf{F}^t)\right] \\ &= 2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{U}_g^{\mathrm{H}}\mathbf{F}\right]\right\} + \alpha_g \mathrm{Tr}\left(\mathbf{F}^{\mathrm{H}}\mathbf{F}\right) + \mathrm{consF}_g \end{aligned} \tag{A.26}$$

where $\mathbf{U}_g$ and consF_g are given in (3.27) and (3.31), respectively. α_g in (3.29) is difficult to obtain for the complex expression of $\mathbf{\Phi}_g$. In the following, it proceeds to obtain the value of α_g.

The following inequalities and equalities will be used later:

(B1): [138] $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices: $\lambda_{\min}(\mathbf{A}) + \lambda_{\min}(\mathbf{B}) \leq \lambda_{\min}(\mathbf{A}+\mathbf{B})$.

(B2): [138] $\mathbf{A}$ is rank one: $\lambda_{\max}(\mathbf{A}) = \mathrm{Tr}\left[\mathbf{A}\right], \lambda_{\min}(\mathbf{A}) = 0$.

(B3): (Theorem 30 in **book-Matrix**) a_k and b_k are positive: $\sum_{k=1}^{K} a_k b_k \leq \max_{k=1}^{K}\{b_k\}$, if $\sum_{k=1}^{K} a_k = 1$.

(B4): [138] $\mathbf{A}$ is positive semidifinite with maximum eigenvalue $\lambda_{\max}(\mathbf{A})$ and $\mathbf{B}$ is positive semidifinite: $\mathrm{Tr}\left[\mathbf{AB}\right] \leq \lambda_{\max}(\mathbf{A})\mathrm{Tr}\left[\mathbf{B}\right]$.

$\mathbf{\Phi}_g$ is complex and cannot be determined by a constant, thus Assumptions (A1)-(A4)

are used to find its lower bound shown in (A.27).

$$\begin{aligned}
\lambda_{\min}\left(\mathbf{\Phi}_g\right) &\overset{(B1)}{\geq} -\sum_{k\in\mathcal{K}_g} g_k(\gamma)\lambda_{\max}\left(\begin{bmatrix} \mathbf{I}\otimes\mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\otimes\mathbf{B}_k^{\mathrm{T}} \end{bmatrix}\right) \\
&\quad -\mu_g\sum_{k\in\mathcal{K}_g} g_k(\gamma)\lambda_{\max}\left(\begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_k^* \end{bmatrix}\begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_k^* \end{bmatrix}^{\mathrm{H}}\right) \\
&\quad +\lambda_{\min}\left(\mu_g\begin{bmatrix} \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k \\ \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^* \end{bmatrix}\begin{bmatrix} \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k \\ \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^* \end{bmatrix}^{\mathrm{H}}\right) \\
&\overset{(B2)}{=} -\sum_{k\in\mathcal{K}_g} g_k(\gamma)\lambda_{\max}(\mathbf{B}_k) - 2\mu_g\sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^{\mathrm{H}}\mathbf{q}_k \\
&\overset{(B2)}{=} -\sum_{k\in\mathcal{K}_g} b_k g_k(\gamma)\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{e} - 2\mu_g\sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^{\mathrm{H}}\mathbf{q}_k \\
&\overset{(B3)}{\geq} -\mathrm{max}_{k\in\mathcal{K}_g}\left\{b_k\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{e}\right\} - 2\mu_g\mathrm{max}_{k\in\mathcal{K}_g}\left\{||\mathbf{q}_k||_2^2\right\} \\
&= -\mathrm{max}_{k\in\mathcal{K}_g}\left\{b_k\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{e}\right\} - 2\mu_g\mathrm{max}_{k\in\mathcal{K}_g}\left\{||\mathbf{Q}_k||_F^2\right\}. \qquad \text{(A.27)}
\end{aligned}$$

Recall that $\mathbf{F} = \mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t), \forall\gamma \in [0,1]$, therefore $||\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t)||_F^2 \leq P_{\mathrm{T}}$. By using (A4), the last term in the right hand side of the last equation of (A.27) satisfies inequality (A.28) as

$$\begin{aligned}
&||\mathbf{Q}_k||_F^2 \\
&= ||\mathbf{C}_k - \mathbf{B}_k^{\mathrm{H}}(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))||_F^2 \\
&= ||(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))^{\mathrm{H}}\mathbf{B}_k||_F^2 + ||\mathbf{C}_k||_F^2 - 2\mathrm{Re}\left\{\mathrm{Tr}\left(\mathbf{C}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))\right)\right\} \\
&\overset{(B4)}{\leq} \lambda_{\max}(\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}})||\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t)||_F^2 + ||\mathbf{C}_k||_F^2 - 2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{C}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}(\mathbf{F}^t + \gamma(\widetilde{\mathbf{F}} - \mathbf{F}^t))\right]\right\} \\
&\leq P_{\mathrm{T}}\lambda_{\max}(\mathbf{B}_k\mathbf{B}_k^{\mathrm{H}}) + ||\mathbf{C}_k||_F^2 + 2\sqrt{P_{\mathrm{T}}}||\mathbf{B}_k\mathbf{C}_k||_F \\
&= P_{\mathrm{T}}b_k^2|\mathbf{e}^{\mathrm{H}}\mathbf{H}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{e}|^2 + ||\mathbf{C}_k||_F^2 + 2\sqrt{P_{\mathrm{T}}}||\mathbf{B}_k\mathbf{C}_k||_F. \qquad \text{(A.28)}
\end{aligned}$$

The third term in the right hand side of the last inequality of (A.28) is the optimal

objective value of the following Problem (A.29) which has a closed-form solution.

$$\begin{aligned} \min_{\mathbf{X}} \ & 2\mathrm{Re}\left\{\mathrm{Tr}\left[\mathbf{C}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\mathbf{X}\right]\right\} \\ \text{s.t.} \ & \mathrm{Tr}\left[\mathbf{X}^{\mathrm{H}}\mathbf{X}\right] \leq P_{\mathrm{T}}. \end{aligned} \tag{A.29}$$

Finally, combining (A.27) with (A.28), it comes to (3.29). Hence, the proof is completed.

A.4 The proof of Theorem 4

Since $f_g(\mathbf{e})$ is twice differentiable and concave, $f_g(\mathbf{e})$ is minorized at $\mathbf{e}^t$ with a quadratic function, as follows

$$f_g(\mathbf{e}) \geq f_g(\mathbf{e}^t) + 2\mathrm{Re}\left\{\mathbf{d}_g^{\mathrm{H}}(\mathbf{e}-\mathbf{e}^t)\right\} + (\mathbf{e}-\mathbf{e}^t)^{\mathrm{H}}\mathbf{N}_g(\mathbf{e}-\mathbf{e}^t), \tag{A.30}$$

where vectors $\mathbf{d}_g \in \mathbb{C}^{M\times 1}$ and matrices $\mathbf{N}_g \in \mathbb{C}^{M\times M}$ are determined to satisfy Assumptions (A1)-(A4).

Obviously, (A1) and (A4) are already satisfied. In order to satisfy Assumption (A3), the directional derivatives of $f_g(\mathbf{e})$ and the right hand side of (A.30) must be equal, yielding

$$\mathbf{d}_g = \sum_{k\in\mathcal{K}_g} g_k(\mathbf{e}^t)(\mathbf{a}_k - \mathbf{A}_k^{\mathrm{H}}\mathbf{e}^t), \tag{A.31}$$

where $g_k(\mathbf{e}^t)$ is defined in (3.41).

Let $\mathbf{e} = \mathbf{e}^t + \gamma(\widetilde{\mathbf{e}} - \mathbf{e}^t), \forall\gamma \in [0,1]$. In order to satisfy Assumption (A2), it suffices to show

$$f_g(\mathbf{e}^t + \gamma(\widetilde{\mathbf{e}} - \mathbf{e}^t)) \geq f_g(\mathbf{e}^t) + 2\gamma\mathrm{Re}\left\{\mathbf{d}_g^{\mathrm{H}}(\widetilde{\mathbf{e}} - \mathbf{e}^t)\right\} + \gamma^2(\widetilde{\mathbf{e}} - \mathbf{e}^t)^{\mathrm{H}}\mathbf{N}_g(\widetilde{\mathbf{e}} - \mathbf{e}^t). \tag{A.32}$$

Then, it needs to calculate the second-order derivatives of the left hand side and the

right hand side of (A.32), and make the latter one lower than or equal to the former for $\forall\gamma \in [0,1]$ and $\forall\widetilde{\mathbf{e}}, \forall\mathbf{e}^t \in \mathcal{S}_e$.

The second-order derivative of the left hand side of (A.32) is given by

$$\frac{\partial^2 L_g(\gamma)}{\partial\gamma^2} = \begin{bmatrix} \mathbf{t}^{\mathrm{H}} & \mathbf{t}^{\mathrm{T}} \end{bmatrix} \mathbf{\Psi}_g \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}, \tag{A.33}$$

with $\mathbf{t} = \widetilde{\mathbf{e}} - \mathbf{e}^t$. $\mathbf{\Psi}_g$ is shown in (A.34) where

$$\begin{aligned} \mathbf{\Psi}_g = \sum_{k\in\mathcal{K}_g} \left(g_k(\gamma) \begin{bmatrix} -\mathbf{A}_k & 0 \\ \mathbf{0} & -\mathbf{A}_k^{\mathrm{T}} \end{bmatrix} - \mu_g g_k(\gamma) \begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_k^* \end{bmatrix} \begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_k^* \end{bmatrix}^{\mathrm{H}} \right) \\ + \mu_g \begin{bmatrix} \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k \\ \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^* \end{bmatrix} \begin{bmatrix} \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k \\ \sum_{k\in\mathcal{K}_g} g_k(\gamma)\mathbf{q}_k^* \end{bmatrix}^{\mathrm{H}}, \end{aligned} \tag{A.34}$$

$$\mathbf{q}_k = \mathbf{a}_k - \mathbf{A}_k^{\mathrm{H}}(\mathbf{e}^t + \gamma(\widetilde{\mathbf{e}} - \mathbf{e}^t)) \tag{A.35}$$

$$g_k(\gamma) = \frac{\exp\{-\mu_g l_k(\gamma)\}}{\sum_{k\in\mathcal{K}_g}\exp\{-\mu_g l_k(\gamma)\}}, k \in \mathcal{K}_g \tag{A.36}$$

The second-order derivative of the right hand side of (A.32) is

$$2(\widetilde{\mathbf{e}} - \mathbf{e}^t)^{\mathrm{H}}\mathbf{N}_g(\widetilde{\mathbf{e}} - \mathbf{e}^t) = \begin{bmatrix} \mathbf{t}^{\mathrm{H}} & \mathbf{t}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{I}\otimes\mathbf{N}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\otimes\mathbf{N}_g^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}. \tag{A.37}$$

Combining (A.33) with (A.37), $\mathbf{N}_g$ must satisfy

$$\mathbf{\Psi}_g \succeq \begin{bmatrix} \mathbf{I}\otimes\mathbf{N}_g & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\otimes\mathbf{N}_g^{\mathrm{T}} \end{bmatrix}.$$

For simplicity, $\mathbf{N}_g = \beta_g\mathbf{I} = \lambda_{\min}(\mathbf{\Psi}_g)\mathbf{I}$ is chosen. Eventually, (A.30) is equivalent to

$$f_g(\mathbf{e}) \geq f_g(\mathbf{e}^t) + 2\mathrm{Re}\left\{\mathbf{d}_g^{\mathrm{H}}(\mathbf{e} - \mathbf{e}^t)\right\} + \beta_g(\mathbf{e} - \mathbf{e}^t)^{\mathrm{H}}(\mathbf{e} - \mathbf{e}^t)$$

$$= 2\mathrm{Re}\left\{\mathbf{u}_g^{\mathrm{H}}\mathbf{e}\right\} + \mathrm{consE}_g, \tag{A.38}$$

where $\mathbf{u}_g$, β_g, and consE_g are given in (3.40), (3.42), and (3.44), respectively. The last equation of (A.38) is from the unit-modulus constraints, i.e., $\mathbf{e}^{\mathrm{H}}\mathbf{e} = (\mathbf{e}^t)^{\mathrm{H}}\mathbf{e}^t = M + 1$. The method to get the value of β_g is similar as α_g, so it is omitted here. Hence, the proof is completed.

A.5 The proof of Theorem 5

Let us denote the converged solution of Problem (3.24) by $\{\mathbf{F}^o, \mathbf{e}^o\}$. In the following, $\{\mathbf{F}^o, \mathbf{e}^o\}$ is proved to satisfy the KKT conditions of Problem (3.24).

Firstly, since $\mathbf{F}^o$ is the globally optimal solution of Problem (3.32), the KKT conditions of the Lagrangian in (3.33) of Problem (3.32) is given by

$$\sum_{g=1}^{G} \nabla_{\mathbf{F}^*} \widetilde{f}_g(\mathbf{F}|\mathbf{F}^t)|_{\mathbf{F}=\mathbf{F}^o} - \tau^o\mathbf{F}^o = \mathbf{0}, \tag{A.39}$$

$$\tau^o(\mathrm{Tr}\left(\mathbf{F}^{\mathrm{H},o}\mathbf{F}^o\right) - P_{\mathrm{T}}) = 0, \tag{A.40}$$

where τ^o is the optimal Lagrange multiplier. According to the Assumption (A3), it has

$$\nabla_{\mathbf{F}^*} \widetilde{f}_g(\mathbf{F}|\mathbf{F}^t)|_{\mathbf{F}=\mathbf{F}^o} = \nabla_{\mathbf{F}^*} f_g\left(\mathbf{F}, \mathbf{e}^o\right)|_{\mathbf{F}=\mathbf{F}^o}. \tag{A.41}$$

By substituting (A.41) into (A.39), it comes to

$$\sum_{g=1}^{G} \nabla_{\mathbf{F}^*} f_g\left(\mathbf{F}, \mathbf{e}^o\right)|_{\mathbf{F}=\mathbf{F}^o} - \tau^o\mathbf{F}^o = \mathbf{0}, \tag{A.42}$$

Then, since $\mathbf{e}^o$ is the locally optimal solution of Problem (3.45), it is readily to obtain

the following KKT conditions:

$$\sum_{g=1}^{G} \nabla_{\mathbf{e}^*} f_g\left(\mathbf{F}^o, \mathbf{e}\right)|_{\mathbf{e}=\mathbf{e}^o} - \sum_{m=1}^{M} \tau_m^{(2),o}(\nabla_{\mathbf{e}^*}|e_m|)|_{\mathbf{e}=\mathbf{e}^o} - \tau_{M+1}^{(2),o}(\nabla_{\mathbf{e}^*} e_{M+1})|_{\mathbf{e}=\mathbf{e}^o} = \mathbf{0}, \quad \text{(A.43)}$$

$$\tau_m^{(2),o}(|e_m^o| - 1) = 0, 1 \leq m \leq M, \tau_{M+1}^{(2),o}(e_{M+1}^o - 1) = 0, \quad \text{(A.44)}$$

where $\boldsymbol{\tau}^{(2),o} = [\tau_1^{(2),o}, ..., \tau_{M+1}^{(2),o}]$ are the optimal Lagrange multipliers.

Then, the set of equations (A.42), (A.40), (A.43), and (A.44) constitute exactly the KKT conditions of Problem (3.24).

Hence, the proof is completed.

Appendix B

Appendix of Chapter 4

B.1 The proof of Lemma 2

In this appendix, $\mathbf{G}^i$ is omitted for simplicity, i.e., $f\left(\mathbf{f}|\mathbf{G}^i\right)$ is replaced by $f\left(\mathbf{f}\right)$. Since $f\left(\mathbf{f}\right)$ is twice differentiable, a second-order approximation is proposed to upper bound $f\left(\mathbf{f}\right)$ at any fixed point $\mathbf{f}^{i-1}$

$$\begin{aligned} f\left(\mathbf{f}\right) \leq & \hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}\right) \\ = & f\left(\mathbf{f}^{i-1}\right) + 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\mathbf{f}-\mathbf{f}^{i-1})\right\} + (\mathbf{f}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{M}_f^i(\mathbf{f}-\mathbf{f}^{i-1}), \end{aligned} \tag{B.1}$$

where $\mathbf{m}_f^i$ and $\mathbf{M}_f^i$ are to be designed to satisfy Assumption A.

Assumptions (A1) and (A4) are readily satisfied. Assumption (A3) is a consistency condition for the first-order directional derivative. Given $\tilde{\mathbf{f}} \in \mathcal{S}_f$, the directional derivative of $\hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}\right)$ at $\mathbf{f}^{i-1}$ with direction $\tilde{\mathbf{f}} - \mathbf{f}^{i-1}$ is (from Eq. (3.4.17) in [120])

$$\begin{aligned} \hat{f}'(\mathbf{f}^{i-1}; \tilde{\mathbf{f}} - \mathbf{f}^{i-1}) = & \left(\left.\frac{\partial \hat{f}}{\partial \mathbf{f}}\right|_{\mathbf{f}=\mathbf{f}^{i-1}}\right)^{\mathrm{T}} (\tilde{\mathbf{f}} - \mathbf{f}^{i-1}) + \left(\left.\frac{\partial \hat{f}}{\partial \mathbf{f}^*}\right|_{\mathbf{f}^*=\mathbf{f}^{*,i-1}}\right)^{\mathrm{T}} (\tilde{\mathbf{f}} - \mathbf{f}^{i-1})^* \\ = & 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\tilde{\mathbf{f}} - \mathbf{f}^{i-1})\right\}. \end{aligned} \tag{B.2}$$

The corresponding directional derivative of $f(\mathbf{f})$ is (from Eq. (3.4.17) in [120])

$$\begin{aligned} f'(\mathbf{f}^{i-1};\tilde{\mathbf{f}}-\mathbf{f}^{i-1}) &= \left(\left.\frac{\partial f}{\partial \mathbf{f}}\right|_{\mathbf{f}=\mathbf{f}^{i-1}}\right)^{\mathrm{T}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}) + \left(\left.\frac{\partial f}{\partial \mathbf{f}^*}\right|_{\mathbf{f}^*=\mathbf{f}^{*,i-1}}\right)^{\mathrm{T}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^* \\ &= \frac{-\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2} 2\mathrm{Re}\left\{\mathbf{f}^{i-1,\mathrm{H}}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}\left(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}\right)\right\}, \end{aligned} \tag{B.3}$$

where x^i is given in (4.15c).

Assumption (A3) is satisfied only when (B.2) and (B.3) are equal, yielding

$$\mathbf{m}_f^i = -\frac{\theta e^{-\theta x^i}}{\left(1+e^{-\theta x^i}\right)^2}\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{f}^{i-1}. \tag{B.4}$$

In order for Assumption (A2) to hold, it is sufficient to show that $\hat{f}(\mathbf{f},\mathbf{f}^{i-1})$ is an upper bound for each linear cut in any direction. In particular, defining $\mathbf{f}=\mathbf{f}^{i-1}+\xi(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}), \forall \xi\in[0,1]$, it is needed to show

$$f\left(\mathbf{f}^{i-1}+\xi(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})\right) \le f\left(\mathbf{f}^{i-1}\right) + 2\xi\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})\right\} + \xi^2(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{M}_f^i(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}). \tag{B.5}$$

Let us introduce the functions $L(\xi) = f(\mathbf{f}^{i-1}+\xi(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}))$ and $l(\xi) = \gamma\sigma^2 - |\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i(\mathbf{f}^{i-1}+\xi(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1}))|^2$. The inequality in (B.5) is fulfilled if Assumption (A3) holds, which is ensured by using (B.4), and if the second-order derivative of $L(\xi)$ is no greater than the second-order derivative in the right hand side of (B.5) for any value of ξ.

The corresponding sufficient condition can be formulated as

$$\frac{\partial^2 L(\xi)}{\partial \xi^2} \le 2(\tilde{\mathbf{f}}-\mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{M}_f^i(\tilde{\mathbf{f}}-\mathbf{f}^{i-1}). \tag{B.6}$$

The next step is to compute $\partial^2 L(\xi)/\partial\xi^2$. To this end, the first-order derivative of

$L(\xi)$ is first calculated, as follows

$$\frac{\partial L(\xi)}{\partial \xi} = g(\xi)\frac{\partial l(\xi)}{\partial \xi}, \tag{B.7}$$

where $g(\xi) = \frac{\theta e^{-\theta l(\xi)}}{(1+e^{-\theta l(\xi)})^2}$, $\frac{\partial l(\xi)}{\partial \xi} = -2\mathrm{Re}\{\mathbf{q}^{\mathrm{H}}(\widetilde{\mathbf{f}} - \mathbf{f}^{i-1})\}$, and $\mathbf{q} = \mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^{i}(\mathbf{f}^{i-1} + \xi(\widetilde{\mathbf{f}} - \mathbf{f}^{i-1}))$.

Then, the second-order derivative can be formulated as

$$\frac{\partial^2 L(\xi)}{\partial \xi^2} = g(\xi)\frac{\partial^2 l(\xi)}{\partial \xi^2} - \theta g(\xi)\left(\frac{\partial l(\xi)}{\partial \xi}\right)^2 + 2\left(1 + e^{-\theta l(\xi)}\right)\left(g(\xi)\frac{\partial l(\xi)}{\partial \xi}\right)^2, \tag{B.8}$$

where $\frac{\partial^2 l(\xi)}{\partial \xi^2} = -2\mathrm{Re}\{(\widetilde{\mathbf{f}} - \mathbf{f}^{i-1})^{\mathrm{H}}\mathbf{\Theta}(\widetilde{\mathbf{f}} - \mathbf{f}^{i-1})\}$ and $\mathbf{\Theta} = \xi\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^{i}$.

Equation (B.8) can be rewritten as a quadratic form of $\mathbf{t} = \widetilde{\mathbf{f}} - \mathbf{f}^{i-1}$, as follows

$$\frac{\partial^2 L(\xi)}{\partial \xi^2} = \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}^{\mathrm{H}} \mathbf{\Phi} \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}, \tag{B.9}$$

where

$$\mathbf{\Phi} = g(\xi)\left(2\left(1 + e^{-\theta l(\xi)}\right)g(\xi) - \theta\right)\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^* \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^* \end{bmatrix}^{\mathrm{H}} - g(\xi)\mathbf{I}_2 \otimes \mathbf{\Theta}. \tag{B.10}$$

Furthermore, (B.6) can be reformulated in a form similar to (B.9), as follows

$$\begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} \mathbf{I} \otimes \mathbf{M}_f^i & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{M}_f^{i,\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}. \tag{B.11}$$

Combining (B.9) and (B.11), the sufficient condition in (B.6) is equivalent to

$$\begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}^{\mathrm{H}} \mathbf{\Phi} \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix} \leq \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} \mathbf{I} \otimes \mathbf{M}_f^i & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{M}_f^{i,\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{t}^* \end{bmatrix},$$

which is satisfied when $\mathbf{M}_f^i$ is chosen so as to fulfill the condition

$$\mathbf{\Phi} \preceq \begin{bmatrix} \mathbf{I} \otimes \mathbf{M}_f^i & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \otimes \mathbf{M}_f^{i,\mathrm{T}} \end{bmatrix}.$$

A convenient choice that fulfills this condition is $\mathbf{M}_f^i = \alpha_f^i \mathbf{I} = \lambda_{\max}(\mathbf{\Phi})\,\mathbf{I}$. Then, $\hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}\right)$ in (B.1) can be formulated as follows

$$\begin{aligned} \hat{f}\left(\mathbf{f}, \mathbf{f}^{i-1}\right) &= f\left(\mathbf{f}^{i-1}\right) + 2\mathrm{Re}\left\{\mathbf{m}_f^{i,\mathrm{H}}(\mathbf{f} - \mathbf{f}^{i-1})\right\} + \alpha_f^i ||\mathbf{f} - \mathbf{f}^{i-1}||_2^2 \\ &= 2\mathrm{Re}\left\{\mathbf{d}_f^{i,\mathrm{H}}\mathbf{f}\right\} + \alpha_f^i ||\mathbf{f}||_2^2 + \mathrm{const}_f^i, \end{aligned}$$

where $\mathbf{d}_f^{i,\mathrm{H}}$, α_f^i and const_f^i are defined in Lemma 2. The deterministic expression of $\lambda_{\max}(\mathbf{\Phi})$ is difficult to obtain, therefore the upper bound is derived as follows

$$\begin{aligned} \lambda_{\max}(\mathbf{\Phi}) &\overset{(\mathrm{p1})}{\leq} 2\left(1 + e^{-\theta l(\xi)}\right) g^2(\xi) \lambda_{\max}\left(\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^* \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^* \end{bmatrix}^{\mathrm{H}}\right) \\ &\quad - g(\xi)\lambda_{\min}(\mathbf{I}_2 \otimes \mathbf{\Theta}) - \theta g(\xi)\lambda_{\min}\left(\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^* \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ \mathbf{q}^* \end{bmatrix}^{\mathrm{H}}\right) \\ &\overset{(\mathrm{p2})}{=} 4\left(1 + e^{-\theta l(\xi)}\right) g^2(\xi) ||\mathbf{q}||_2^2 \\ &\overset{(\mathrm{p3})}{<} \frac{\theta^2}{2} ||\mathbf{q}||_2^2 \\ &\overset{(\mathrm{p4})}{\leq} \frac{\theta^2}{2} \lambda_{\max}(\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i) \cdot ||\mathbf{f}^{i-1} + \gamma(\widetilde{\mathbf{f}} - \mathbf{f}^{i-1})||_2^2 \\ &\overset{(\mathrm{p5})}{\leq} \frac{\theta^2}{2} P_{max} |\mathbf{e}^{i-1,\mathrm{H}}\mathbf{G}^i\mathbf{G}^{i,\mathrm{H}}\mathbf{e}^{i-1}|^2. \end{aligned}$$

where the inequalities are obtained by departing from (B.10) and by invoking the following properties.

(p1): If $\mathbf{A}$ and $\mathbf{B}$ are Hermitian matrices, then $\lambda_{\max}(\mathbf{A}) + \lambda_{\max}(\mathbf{B}) \geq \lambda_{\max}(\mathbf{A} + \mathbf{B})$ [138].

(p2): If $\mathbf{A}$ is rank one, then $\lambda_{\max}(\mathbf{A}) = \mathrm{Tr}[\mathbf{A}], \lambda_{\min}(\mathbf{A}) = 0$ [138].

(p3): $\left(1+e^{-\theta l(\xi)}\right) g^2(\xi) \leq \theta^2/8$, where the equality holds when $l(\xi)=0$.

(p4): If $\mathbf{A}$ is positive semidefinite with maximum eigenvalue $\lambda_{\max}(\mathbf{A})$ and $\mathbf{B}$ is positive semidefinite, then $\mathrm{Tr}\left[\mathbf{AB}\right] \leq \lambda_{\max}(\mathbf{A})\mathrm{Tr}\left[\mathbf{B}\right]$ [138].

(p5): The power constraint $||\mathbf{f}^{i-1}+\gamma(\widetilde{\mathbf{f}}-\mathbf{f}^{i-1})||_2^2 \leq P_{max}$ needs to be fulfilled.

Hence, the proof is completed.

B.2 The proof of Theorem 6

Define the random functions

$$g^t(\mathbf{x}) = \frac{1}{t}\sum_{i=1}^{t} f\left(\mathbf{x}|\mathbf{G}^i\right), \tag{B.12}$$

$$\hat{g}^t(\mathbf{x}) = \frac{1}{t}\sum_{i=1}^{t} \hat{f}\left(\mathbf{x},\mathbf{x}^{i-1}|\mathbf{G}^i\right). \tag{B.13}$$

To prove the convergence of Algorithm 1, the following lemmas are used.

Lemma 16. *Assume that Assumptions A and B are satisfied and define a limit point* $\bar{\mathbf{x}}$ *of the subsequence* $\{\mathbf{x}^{t_j}\}_{j=1}^{\infty}$. *Then, there exists uniformly continuous functions* $g(\mathbf{x})$ *and* $\hat{g}(\mathbf{x})$ *such that*

$$g(\mathbf{x}) = \lim_{n\to\infty} g^t(\mathbf{x}) = \mathbb{E}\left[f\left(\mathbf{x}|\mathbf{G}\right)\right], \forall \mathbf{x} \in \mathcal{S}_x, \tag{B.14}$$

$$g(\bar{\mathbf{x}}) = \lim_{j\to\infty} g^{t_j}(\mathbf{x}^{t_j}), \tag{B.15}$$

$$\hat{g}(\mathbf{x}) = \lim_{j\to\infty} \hat{g}^{t_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x, \tag{B.16}$$

$$\hat{g}(\bar{\mathbf{x}}) = \lim_{j\to\infty} \hat{g}^{t_j}(\mathbf{x}^{t_j}). \tag{B.17}$$

Proof: First, $f\left(\mathbf{x},\mathbf{G}\right)$ is bounded for $\forall \mathbf{x} \in \mathcal{S}_x$ and for all channel realizations due to the Assumption (B2). Therefore, (B.14) holds by using the strong law of large numbers [139]. Also, the families of functions $\{g^{t_j}(\mathbf{x})\}$ are equicontinuous and bounded over the

compact set $\mathcal{S}_x$ due to the Assumption (B2) and the use of the mean value theorem. Thus, by restricting to a subsequence, it has (B.15). Furthermore, the families of functions $\{\hat{g}^t(\mathbf{x})\}$ are also equicontinuous and bounded over the compact set $\mathcal{S}_x$ due to the Assumption (B2) and because $||\nabla_{\mathbf{x}}\hat{f}(\mathbf{x}, \mathbf{x}^{i-1}, \mathbf{G})||$ is bounded. Hence, the Arzelà-Ascoli theorem [140] implies that, by restricting to a subsequence, there exists a uniformly continuous function $\hat{g}(\mathbf{x})$ such that (B.16) and (B.17) hold. ■

In addition, the update rule of Algorithm 1 leads the following lemma.

Lemma 17. $\lim_{n\to\infty} |\hat{g}^t(\mathbf{x}^t) - g^t(\mathbf{x}^t)| = 0$, *almost surely.*

Proof: The proof of Lemma 17 is the same as the proof of (Lemma 1 in [94]) and is hence omitted for conciseness. ■

Assumption (A2) implies that $\hat{g}^{t_j}(\mathbf{x}) \geq g^{t_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x$, which combined with (B.14) and (B.16) leads to

$$\hat{g}(\mathbf{x}) \geq g(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x. \tag{B.18}$$

Moreover, combining Lemma 17 with (B.15) and (B.17) it has

$$\hat{g}(\bar{\mathbf{x}}) = g(\bar{\mathbf{x}}). \tag{B.19}$$

Then, (B.18) and (B.19) imply that $\bar{\mathbf{x}}$ is a minimizer of the function $\hat{g}(\mathbf{x}) - g(\mathbf{x})$. Hence the first-order optimality condition is satisfied

$$\nabla\hat{g}(\bar{\mathbf{x}}) - \nabla g(\bar{\mathbf{x}}) = 0. \tag{B.20}$$

Due to the fact that $\bar{\mathbf{x}}$ is the limit point of the problem in (4.12) or the problem in (4.13), it has $\hat{g}(\bar{\mathbf{x}}) \leq \hat{g}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x$, which implies

$$\langle \nabla\hat{g}(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x. \tag{B.21}$$

Combining (B.21) with (B.20), it obtains

$$\langle \nabla g(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x, \tag{B.22}$$

which means that the directional derivative of the objective function $g(\mathbf{x})$ is non-negative for every feasible direction at $\bar{\mathbf{x}}$. Recalling that $\mathbf{x} \in \{\mathbf{f}, \mathbf{e}\}$ and defining the limit points $\{\bar{\mathbf{f}}, \bar{\mathbf{e}}\}$, (B.22) is equivalent to

$$\begin{cases} \left\langle \nabla g(\bar{\mathbf{f}}), \mathbf{f} - \bar{\mathbf{f}} \right\rangle \geq 0, \forall \mathbf{f} \in \mathcal{S}_f, \\ \langle \nabla g(\bar{\mathbf{e}}), \mathbf{e} - \bar{\mathbf{e}} \rangle \geq 0, \forall \mathbf{e} \in \mathcal{S}_e. \end{cases}$$

Therefore, according to [95], $\{\bar{\mathbf{f}}, \bar{\mathbf{e}}\}$ is a stationary point of the problem in (4.11) due to the regularity of $g(\cdot)$.

B.3 The proof of Theorem 8

Define the random functions

$$G^t(\mathbf{x}) = \frac{1}{t} \sum_{i=1}^{t} F\left(\mathbf{x} | \mathbf{G}^i\right), \tag{B.23}$$

$$\hat{G}^t(\mathbf{x}) = \frac{1}{t} \sum_{i=1}^{t} \hat{F}\left(\mathbf{x}, \mathbf{x}^{i-1} | \mathbf{G}^i\right). \tag{B.24}$$

To analyze the convergence, it needs the following lemmas.

Lemma 18. *Assume that Assumptions B and C are satisfied and define a limit point* $\bar{\mathbf{x}}$ *for the subsequence* $\{\mathbf{x}^{t_j}\}_{j=1}^{\infty}$. *Then, there exists uniformly continuous functions* $G(\mathbf{x})$ *and* $\hat{G}(\mathbf{x})$ *such that*

$$G(\mathbf{x}) = \lim_{n \to \infty} G^t(\mathbf{x}) = \mathbb{E}\left[F\left(\mathbf{x} | \mathbf{G}\right)\right], \forall \mathbf{x} \in \mathcal{S}_x, \tag{B.25}$$

$$G(\bar{\mathbf{x}}) = \lim_{j \to \infty} G^{t_j}(\mathbf{x}^{t_j}), \tag{B.26}$$

$$\hat{G}(\mathbf{x}) = \lim_{j\to\infty} \hat{G}^{t_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x, \tag{B.27}$$

$$\hat{G}(\bar{\mathbf{x}}) = \lim_{j\to\infty} \hat{G}^{t_j}(\mathbf{x}^{t_j}). \tag{B.28}$$

Proof: The proof of Lemma 18 is the same as the proof of Lemma 16 and is omitted for brevity. ■

Furthermore, $\mathbf{x}^{t_j}$ is the minimizer of $\hat{G}^{t_j}(\mathbf{x})$, which implies

$$\hat{G}^{t_j}(\mathbf{x}^{t_j}) \leq \hat{G}^{t_j}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x. \tag{B.29}$$

Assuming $j \to \infty$, and combining (B.27) and (B.28), it obtains $\hat{G}(\bar{\mathbf{x}}) \leq \hat{G}(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}_x$, which implies that its first-order optimality condition is satisfied

$$\left\langle \nabla \hat{G}(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \right\rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x. \tag{B.30}$$

By combining (B.30) and Assumption (C3), it finally obtains

$$\langle \nabla G(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{S}_x. \tag{B.31}$$

Since $\mathbf{x} \in \{\mathbf{F}, \mathbf{e}\}$, define the limit points $\{\bar{\mathbf{F}}, \bar{\mathbf{e}}\}$ and (B.31) is then equivalent to

$$\begin{cases} \left\langle \nabla G(\bar{\mathbf{F}}), \mathbf{F} - \bar{\mathbf{F}} \right\rangle \geq 0, \forall \mathbf{F} \in \mathcal{S}_f, \\ \langle \nabla G(\bar{\mathbf{e}}), \mathbf{e} - \bar{\mathbf{e}} \rangle \geq 0, \forall \mathbf{e} \in \mathcal{S}_e. \end{cases}$$

Therefore, according to [95], $\{\bar{\mathbf{F}}, \bar{\mathbf{e}}\}$ is a stationary point of the problem in (4.28) due to the regularity of $G(\cdot)$.

Appendix C

Appendix of Chapter 5

C.1 The proof of Lemma 6

Let x be a complex scalar variable, it has the first-order Taylor inequality

$$|x|^2 \geq 2\mathrm{Re}\left\{x^{*,(n)}x\right\} - x^{*,(n)}x^{(n)}, \tag{C.1}$$

for any fixed point $x^{(n)}$. By replacing x and $x^{(n)}$ in (C.1) with $(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k)\mathbf{f}_k$ and $(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\mathbf{G}_k)\mathbf{f}_k^{(n)}$, respectively, it has

$$\begin{aligned}\left|\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k\right|^2 \geq & 2\mathrm{Re}\left\{\left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}\left(\mathbf{h}_k + \mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right)\right\} \\ & - \left(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\mathbf{G}_k\right)\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}\left(\mathbf{h}_k + \mathbf{G}_k^{\mathrm{H}}\mathbf{e}^{(n)}\right).\end{aligned} \tag{C.2}$$

By plugging $\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k$ into the right hand side of (C.2) and expanding it by using mathematical transformations, i.e., $\mathrm{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{B}) = \mathrm{vec}^{\mathrm{H}}(\mathbf{A})\mathrm{vec}(\mathbf{B})$ and $\mathrm{Tr}(\mathbf{ABCD}) = (\mathrm{vec}^{\mathrm{T}}(\mathbf{D}))^{\mathrm{T}}(\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [120], (5.10) can be obtained.

Hence, the proof is completed.

C.2 The proof of Lemma 7

The lower bound of (5.26) can also be derived from (C.2) under the full channel uncertainty. In particular, $\mathbf{h}_k = \widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k$ and $\mathbf{G}_k = \widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k$ are inserted into the first term on the right hand side of (C.2), and then get (C.3) at the top of the next page.

$$
\begin{aligned}
&\left[(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \triangle\mathbf{h}_k^{\mathrm{H}}) + \mathbf{e}^{(n),\mathrm{H}}(\widehat{\mathbf{G}}_k + \triangle\mathbf{G}_k)\right]\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}\left[(\widehat{\mathbf{h}}_k + \triangle\mathbf{h}_k) + (\widehat{\mathbf{G}}_k^{\mathrm{H}} + \triangle\mathbf{G}_k^{\mathrm{H}})\mathbf{e}\right]\\
=&(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) + (\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}(\triangle\mathbf{h}_k + \triangle\mathbf{G}_k^{\mathrm{H}}\mathbf{e})\\
&+ (\triangle\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}} \triangle \mathbf{G}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) + (\triangle\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}} \triangle \mathbf{G}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}(\triangle\mathbf{h}_k + \triangle\mathbf{G}_k^{\mathrm{H}}\mathbf{e})\\
=&(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) + (\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}} \triangle \mathbf{h}_k\\
&+ \mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G}_k)\mathrm{vec}(\mathbf{e}(\widehat{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\widehat{\mathbf{G}}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}) + \triangle\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}}(\widehat{\mathbf{h}}_k + \widehat{\mathbf{G}}_k^{\mathrm{H}}\mathbf{e}) + \triangle\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_k^{(n)}\mathbf{f}_k^{\mathrm{H}} \triangle \mathbf{h}_k\\
&+ \mathrm{vec}^{\mathrm{H}}(\mathbf{e}^{(n)}(\widehat{\mathbf{h}}_k + \mathbf{e}^{\mathrm{H}}\widehat{\mathbf{G}}_k^{\mathrm{H}})\mathbf{f}_k\mathbf{f}_k^{(n),\mathrm{H}})\mathrm{vec}(\triangle\mathbf{G}_k) \ + \mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G}_k)(\mathbf{f}_k^{*}\mathbf{f}_k^{(n),\mathrm{T}} \otimes \mathbf{e})\triangle\mathbf{h}_k^{*}\\
&+ \triangle\mathbf{h}_k^{\mathrm{T}}(\mathbf{f}_k^{*}\mathbf{f}_k^{(n),\mathrm{T}} \otimes \mathbf{e}^{(n),\mathrm{H}})\mathrm{vec}(\triangle\mathbf{G}_k) + \mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G}_k)(\mathbf{f}_k^{*}\mathbf{f}_k^{(n),\mathrm{T}} \otimes \mathbf{e}\mathbf{e}^{(n),\mathrm{H}})\mathrm{vec}(\triangle\mathbf{G}_k)\\
=&\widetilde{\mathbf{i}}_k^{\mathrm{H}}\mathbf{D}_k\widetilde{\mathbf{i}}_k + \mathbf{d}_{1,k}^{\mathrm{H}}\widetilde{\mathbf{i}}_k + \widetilde{\mathbf{i}}_k^{\mathrm{H}}\mathbf{d}_{2,k} + d_k. \qquad \text{(C.3)}
\end{aligned}
$$

With the similar mathematical transformations, the remaining two terms on the right hand side of (C.2) under the full channel uncertainty can be expressed as

$$
\begin{aligned}
&(\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{\mathrm{H}}\mathbf{G}_k)\mathbf{f}_k\mathbf{f}_k^{(n),\mathrm{H}}(\mathbf{h}_k + \mathbf{G}_k^{\mathrm{H}}\mathbf{e}^{(n)})\\
=&\widetilde{\mathbf{i}}_k^{\mathrm{H}}\mathbf{D}_k^{\mathrm{H}}\widetilde{\mathbf{i}}_k + \mathbf{d}_{2,k}^{\mathrm{H}}\widetilde{\mathbf{i}}_k + \widetilde{\mathbf{i}}_k^{\mathrm{H}}\mathbf{d}_{1,k} + d_k^{*} + (\mathbf{h}_k^{\mathrm{H}} + \mathbf{e}^{(n),\mathrm{H}}\mathbf{G}_k)\mathbf{f}_k^{(n)}\mathbf{f}_k^{(n),\mathrm{H}}(\mathbf{h}_k + \mathbf{G}_k^{\mathrm{H}}\mathbf{e}^{(n)})\\
=&\widetilde{\mathbf{i}}_k^{\mathrm{H}}\mathbf{Z}_k\widetilde{\mathbf{i}}_k + \mathbf{z}_k^{\mathrm{H}}\widetilde{\mathbf{i}}_k + \widetilde{\mathbf{i}}_k^{\mathrm{H}}\mathbf{z}_k + z_k. \qquad \text{(C.4)}
\end{aligned}
$$

Hence, the proof is completed.

C.3 The proof of Theorem 9

Denote by $\widehat{\mathbf{\Gamma}}^{\star} = [\widehat{\mathbf{\Gamma}}_1^{\star}, ..., \widehat{\mathbf{\Gamma}}_K^{\star}]$ the optimal solution of the relaxed version of Problem (5.41) and define the projection matrices as $\mathbf{P}_k = \widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}}\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\mathrm{H}}\widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}}/||\widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}}\widehat{\mathbf{h}}_k||^2, \forall k \in \mathcal{K}$,

where $\widehat{\mathbf{h}}_k = \left(\mathbf{h}_k + \mathbf{G}_k^{\mathrm{H}}\mathbf{e}\right)$. Then, a rank-one solution is constructed as $\widetilde{\mathbf{\Gamma}}^\star = [\widetilde{\mathbf{\Gamma}}_1^\star, ..., \widetilde{\mathbf{\Gamma}}_K^\star]$, each sub-matrix of which is given by

$$\widetilde{\mathbf{\Gamma}}_k^\star = \widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}} \mathbf{P}_k \widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}}. \tag{C.5}$$

Firstly, it checks the objective value of Problem (5.41) with solution $\widetilde{\mathbf{\Gamma}}^\star$:

$$\begin{aligned}\sum_{k=1}^{K} \mathrm{Tr}\left(\widetilde{\mathbf{\Gamma}}_k^\star\right) - \sum_{k=1}^{K} \mathrm{Tr}\left(\widehat{\mathbf{\Gamma}}_k^\star\right) &= \sum_{k=1}^{K} \mathrm{Tr}\left(\widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}} (\mathbf{P}_k - \mathbf{I}) \widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}}\right) \\ &\leq 0,\end{aligned} \tag{C.6}$$

which means that the objective value achieved by using the solution $\widetilde{\mathbf{\Gamma}}^\star$ is no more than that generated from the optimal solution $\widehat{\mathbf{\Gamma}}^\star$.

Then, since it is computationally intractable to check whether the constructed solution satisfies the constraints (5.41b)-(5.41e) directly, it instead considers the constraint (5.32b) in the original Problem (5.32). Specifically, from (5.34), it comes to

$$\frac{\widehat{\mathbf{h}}_k^{\mathrm{H}} \widetilde{\mathbf{\Gamma}}_k^\star \widehat{\mathbf{h}}_k}{(2^{R_k} - 1)} = \frac{|\widehat{\mathbf{h}}_k^{\mathrm{H}} \widehat{\mathbf{\Gamma}}_k^\star \widehat{\mathbf{h}}_k|^2}{||\widehat{\mathbf{\Gamma}}_k^{\star\frac{1}{2}} \widehat{\mathbf{h}}_k||^2 (2^{R_k} - 1)} = \frac{\widehat{\mathbf{h}}_k^{\mathrm{H}} \widehat{\mathbf{\Gamma}}_k^\star \widehat{\mathbf{h}}_k}{(2^{R_k} - 1)}, \tag{C.7}$$

as well as

$$\begin{aligned}\widehat{\mathbf{h}}_k^{\mathrm{H}} \widetilde{\mathbf{\Gamma}}_i^\star \widehat{\mathbf{h}}_k &= \widehat{\mathbf{h}}_i^{\mathrm{H}} \widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}} \frac{\widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}} \widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}}}{||\widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}} \widehat{\mathbf{h}}_i||^2} \widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}} \widehat{\mathbf{h}}_i \\ &\leq \lambda_{max}\left(\widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}} \widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \widehat{\mathbf{\Gamma}}_i^{\star\frac{1}{2}}\right) = \widehat{\mathbf{h}}_k^{\mathrm{H}} \widehat{\mathbf{\Gamma}}_i^\star \widehat{\mathbf{h}}_k.\end{aligned} \tag{C.8}$$

Combining (C.7) with (C.8), it has

$$\widehat{\mathbf{h}}_k^{\mathrm{H}} [\widetilde{\mathbf{\Gamma}}_k^\star / (2^{R_k} - 1) - \sum_{i \neq k}^{K} \widetilde{\mathbf{\Gamma}}_k^\star] \widehat{\mathbf{h}}_k \geq \widehat{\mathbf{h}}_k^{\mathrm{H}} [\widehat{\mathbf{\Gamma}}_k^\star / (2^{R_k} - 1) - \sum_{i \neq k}^{K} \widehat{\mathbf{\Gamma}}_i^\star] \widehat{\mathbf{h}}_k, \tag{C.9}$$

which implies that the constructed solution $\widetilde{\mathbf{\Gamma}}^\star$ satisfies constraint (5.32b) and then

satisfies constraints (5.41b)-(5.41e).

With (C.6) and (C.9), it comes to the conclusion that $\widetilde{\mathbf{\Gamma}}^{\star}$ is also a feasible solution of the relaxed version of Problem (5.41) with rank one.

Hence, the proof is completed.

Appendix D

Appendix of Chapter 6

D.1 The proof of Lemma 10

To begin, the following lemma is introduced.

Lemma 19. *Let $x \geq 0$ be a positive real number, and consider the function $g_1(a, x) = -ax + \ln a + 1$, then*

$$\ln x^{-1} = \max_{a \geq 0} g_1(a, x).$$

By applying Lemma 19, an upper bound of rate $R_E(\mathbf{w}, \mathbf{e})$ can be constructed as

$$\begin{aligned} R_E(\mathbf{w}, \mathbf{e}) &= \frac{-\ln(1 + \left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_E\mathbf{w}\right|^2 / \sigma_E^2)^{-1}}{2\ln 2} \\ &\overset{\text{(a)}}{=} \frac{-\max_{a_E \geq 0} g_1(a_E, 1 + \left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_E\mathbf{w}\right|^2 / \sigma_E^2)}{2\ln 2} \\ &= \frac{\min_{a_E \geq 0} -g_1(a_E, 1 + \left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_E\mathbf{w}\right|^2 / \sigma_E^2)}{2\ln 2} \\ &\leq \frac{-g_1(a_E, 1 + \left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_E\mathbf{w}\right|^2 / \sigma_E^2)}{2\ln 2}, \text{for any } a_E \geq 0 \\ &= \frac{a_E \left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_E\mathbf{w}\right|^2 / \sigma_E^2 + a_E - \ln a_E - 1}{2\ln 2} \\ &= \widetilde{R}_E(\mathbf{w}, \mathbf{e}, a_E), \end{aligned} \tag{D.1}$$

where the equality (a) is from Lemma 19.

Hence, the proof is completed.

D.2 The proof of Lemma 11

To prove Lemma 11, the following lemma is needed.

Lemma 20. *Let v be a complex number, and consider the function $g_2(v,x) = (|x|^2 + \sigma^2)|v|^2 - 2Re\{vx\} + 1$, then*

$$\frac{\sigma^2}{|x|^2+\sigma^2} = \min_v g_2(a,x).$$

By applying Lemma 20, a lower bound of rate $R_K(\mathbf{w},\mathbf{e})$ can be constructed as

$$\begin{aligned}
R_K(\mathbf{w},\mathbf{e}) &= \frac{\ln\left(1-\frac{\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2}{\sigma_K^2+\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2}\right)^{-1}}{2\ln 2} \\
&\overset{(a)}{=} \frac{\max_{a_K\geq 0} g(a_K, 1-\frac{\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2}{\sigma_K^2+\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2})}{2\ln 2} \\
&\geq \frac{g(a_K, 1-\frac{\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2}{\sigma_K^2+\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2})}{2\ln 2}, \text{for any } a_K \geq 0 \\
&= \frac{-a_K\left(\frac{\sigma_K^2}{\sigma_K^2+\left|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right|^2}\right)+\ln a_K+1}{2\ln 2} \\
&\overset{(b)}{=} \frac{-a_K\left(\min_v g_2(v,\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w})\right)+\ln a_K+1}{2\ln 2} \\
&= \frac{a_K\left(\max_v -g_2(v,\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w})\right)+\ln a_K+1}{2\ln 2} \\
&\geq \frac{a_K\left(-g_2(v,\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w})\right)+\ln a_K+1}{2\ln 2}, \text{for any } v \geq 0 \\
&= \frac{1}{2\ln 2}(-a_K|v|^2|\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}|^2 - \sigma_K^2 a_K|v|^2 + 2a_K\mathrm{Re}\left\{v\mathbf{e}^{\mathrm{H}}\mathbf{H}_K\mathbf{w}\right\} - a_K + \ln a_K + 1) \\
&= \widetilde{R}_K(\mathbf{w},\mathbf{e},a_K,v) \qquad \text{(D.2)}
\end{aligned}$$

where Equality (a) is from Lemma 19, and Equality (b) is from Lemma 20.

Hence, the proof is completed.

D.3 The proof of Lemma 15

To begin with, $\mathcal{P}1$ is solved: $\mathcal{P}1$ is equivalent to

$$\min_{\mathbf{e}} \ \mathbf{e}^{\mathrm{H}}\mathbf{A}_E\mathbf{e} \tag{D.3a}$$

$$\text{s.t.}(6.12d), \tag{D.3b}$$

$$\mathbf{e}^{\mathrm{H}}\mathbf{A}_K\mathbf{e} \geq e^{2\mathcal{R}} - 1. \tag{D.3c}$$

Step 1: Construct a surrogate problem: Under the MM algorithm framework [93], it comes to the following lemma.

Lemma 21. *[27], [141] Given* $\mathbf{A} \succeq \mathbf{A}_0$ *and* $\mathbf{x}$*, then quadratic function* $\mathbf{x}^{\mathrm{H}}\mathbf{A}_0\mathbf{x}$ *is majorized by* $\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} - 2\mathrm{Re}\{\mathbf{x}^{t,\mathrm{H}}(\mathbf{A}-\mathbf{A}_0)\mathbf{x}\} + \mathbf{x}^{t,\mathrm{H}}(\mathbf{A}-\mathbf{A}_0)\mathbf{x}^t$ *at* $\mathbf{x}^t$*.*

By adopting Lemma 21 and setting $\mathbf{A} = \lambda_{\max}(\mathbf{A}_E)\mathbf{I}$ for simplicity, the quadratic objective function in (D.3a) is majorized by

$$2\lambda_{\max}(\mathbf{A}_E)M - 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}(\lambda_{\max}(\mathbf{A}_E)\mathbf{I} - \mathbf{A}_E)\mathbf{e}\} - \mathbf{e}^{t,\mathrm{H}}\mathbf{A}_E\mathbf{e}^t \tag{D.4}$$

at feasible point $\mathbf{e}^t$. To deal with the non-convex constraint (D.3c), $\mathbf{e}^{\mathrm{H}}\mathbf{A}_K\mathbf{e}$ is replaced with its linear lower bound, resulting in the following equivalent constraint

$$(D.3c) \Rightarrow 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}\mathbf{A}_K\mathbf{e}\} \geq e^{2\mathcal{R}} - 1 + \mathbf{e}^{t,\mathrm{H}}\mathbf{A}_K\mathbf{e}^t. \tag{D.5}$$

Step 2: Closed-form solution :

By omitting the constant, Problem (D.3) then becomes

$$\max_{\mathbf{e}} \ 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}(\lambda_{\max}(\mathbf{A}_E)\mathbf{I} - \mathbf{A}_E)\mathbf{e}\}, \tag{D.6a}$$

$$\text{s.t. } (6.12d), (D.5). \tag{D.6b}$$

According to [26], a price mechanism is introduced for solving Problem (D.6), i.e.,

$$\begin{aligned}&\max_{\mathbf{e}} \ 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}(\lambda_{\max}(\mathbf{A}_E)\mathbf{I}-\mathbf{A}_E)\mathbf{e}\}+\varrho_1 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}\mathbf{A}_K\mathbf{e}\}\\ &\text{s.t. (6.12d).}\end{aligned}$$

where ϱ_1 is a non-negative price. Then, the globally optimal solution is given by

$$\mathbf{e}_1^{\#}(\varrho_1^{opt})=\exp\{\mathrm{j}\arg((\lambda_{\max}(\mathbf{A}_E)\mathbf{I}-\mathbf{A}_E+\varrho_1\mathbf{A}_K)\mathbf{e}^t)\}.$$

The optimal ϱ_1^{opt} is determined by using the bisection search method, the detailed information about which can be found in [26].

Then, $\mathcal{P}2$ is solved: $\mathcal{P}2$ is equivalent to

$$\max_{\mathbf{e}} \ \frac{1+\mathbf{e}^{\mathrm{H}}\mathbf{A}_K\mathbf{e}}{1+\mathbf{e}^{\mathrm{H}}\mathbf{A}_E\mathbf{e}}, \tag{D.7a}$$

$$\text{s.t. } (6.12d), \tag{D.7b}$$

$$\mathbf{e}^{\mathrm{H}}\mathbf{A}_K\mathbf{e}\le e^{2\mathcal{R}}-1. \tag{D.7c}$$

Step 1: Construct a surrogate problem: Under the MM algorithm framework, a linear lower bound of the objective function in (D.7a) is constructed as

$$\begin{aligned}\frac{1+\mathbf{e}^{\mathrm{H}}\mathbf{A}_K\mathbf{e}}{1+\mathbf{e}^{\mathrm{H}}\mathbf{A}_E\mathbf{e}} &\overset{\text{(a)}}{\ge} \frac{2\mathrm{Re}\{1+d_K\}}{1+d_E^t}-\frac{1+d_K^t}{(1+d_E^t)^2}\left(1+\mathbf{e}^{\mathrm{H}}\mathbf{A}_E\mathbf{e}\right)\\ &\overset{\text{(b)}}{\ge} \frac{2\mathrm{Re}\{1+d_K\}}{1+d_E^t}-\frac{1+d_K^t}{(1+d_E^t)^2}-\frac{1+d_K^t}{(1+d_E^t)^2}(2\lambda_{\max}(\mathbf{A}_E)M\\ &\quad -2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}(\lambda_{\max}(\mathbf{A}_E)\mathbf{I}-\mathbf{A}_E)\mathbf{e}\}-d_E^t)\\ &= 2\mathrm{Re}\{\mathbf{c}^{\mathrm{H}}\mathbf{e}\}+\text{const},\end{aligned}$$

where $d_K=\mathbf{e}^{t,\mathrm{H}}\mathbf{A}_K\mathbf{e}$. $\{d_K^t, d_E^t, \mathbf{c}\}$ are defined in Lemma 15. Inequality (a) is due to

Lemma 9, and inequality (b) is from Lemma 21. By using Lemma 21 again, the convex constraint (D.7c) can be replaced by an easy-to-solve form as

$$(D.7c) \Rightarrow 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}(\lambda_{\max}(\mathbf{A}_K)\mathbf{I}-\mathbf{A}_K)\mathbf{e}\} \geq -2\lambda_{\max}(\mathbf{A}_K)M - e^{2\mathcal{R}} + 1 - \mathbf{e}^{t,\mathrm{H}}\mathbf{A}_K\mathbf{e}^t. \tag{D.8}$$

Step 2: Closed-form solution: By omitting the constant, Problem (D.7) is then equivalent to

$$\max_{\mathbf{e}} \ 2\mathrm{Re}\{\mathbf{c}^{\mathrm{H}}\mathbf{e}\}, \tag{D.9a}$$

$$\text{s.t. } (6.12d), (D.8). \tag{D.9b}$$

By using the price mechanism, Problem (D.7) is reformulated as

$$\max_{\mathbf{e}} \ 2\mathrm{Re}\{\mathbf{c}^{\mathrm{H}}\mathbf{e}\} + \varrho_2 2\mathrm{Re}\{\mathbf{e}^{t,\mathrm{H}}(\lambda_{\max}(\mathbf{A}_K)\mathbf{I}-\mathbf{A}_K)\mathbf{e}\}$$

$$\text{s.t.} (6.12d).$$

where ϱ_2 is a non-negative price. The globally optimal solution is given by $\mathbf{e}_2^{\#}(\varrho_2^{opt}) = \exp\{\mathrm{j}\arg(\mathbf{c} + \varrho_2(\lambda_{\max}(\mathbf{A}_K)\mathbf{I}-\mathbf{A}_K)\mathbf{e}^t)\}$ where the optimal ϱ_2^{opt} is determined by using the bisection search method.

Hence, the proof is completed.

Bibliography

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, 2020.

[2] X. You, C.-X. Wang, J. Huang, *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, 2020.

[3] Z. Zhang, Y. Xiao, Z. Ma, *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, 2019.

[4] J. G. Andrews, S. Buzzi, W. Choi, *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[5] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, 2019.

[6] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, 2020.

[7] C. Pan, H. Ren, K. Wang, *et al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, 2021.

[8] X. Yuan, Y.-J. A. Zhang, Y. Shi, W. Yan, and H. Liu, "Reconfigurable-intelligent-surface empowered wireless communications: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 136–143, 2021.

[9] C. Huang, S. Hu, G. C. Alexandropoulos, *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, 2020.

[10] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light Sci. Appl.*, vol. 3, no. 10, e218–e218, 2014.

[11] X. Tan, Z. Sun, J. M. Jornet, and D. Pados, "Increasing indoor spectrum sharing capacity using smart reflect-array," in *2016 IEEE Intern. Conf. Commun. (ICC)*, 2016, pp. 1–6.

[12] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.

[13] N. Golrezaei, A. F. Molisch, A. G. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, 2013.

[14] N. D. Sidiropoulos, T. N. Davidson, and Z. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, 2006.

[15] E. Karipidis, N. D. Sidiropoulos, and Z. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, 2008.

[16] L. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, 2014.

[17] Z. Xiang, M. Tao, and X. Wang, "Massive MIMO multicasting in noncooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1180–1193, 2014.

[18] M. Sadeghi, L. Sanguinetti, R. Couillet, and C. Yuen, "Reducing the computational complexity of multicasting in large-scale antenna aystems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2963–2975, 2017.

[19] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[20] V. Raghavan *et al.*, "Spatio-temporal impact of hand and body blockage for millimeter-wave user equipment design at 28 GHz," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 46–52, 2018.

[21] V. Raghavan, M.-L. Chi, M. A. Tassoudji, O. H. Koymen, and J. Li, "Antenna placement and performance tradeoffs with hand blockage in millimeter wave systems," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 3082–3096, 2019.

[22] V. Raghavan *et al.*, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microwave Theory Tech.*, vol. 67, no. 7, pp. 3010–3024, 2019.

[23] B. Di, H. Zhang, L. Li, L. Song, Y. Li, and Z. Han, "Practical hybrid beamforming with finite-resolution phase shifters for reconfigurable intelligent surface based multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4565–4570, 2020.

[24] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Tech.*, vol. 69, no. 12, pp. 14 960–14 973, 2020.

[25] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. Di Renzo, "Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1057–1061, 2021.

[26] C. Pan, H. Ren, K. Wang, *et al.*, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, 2020.

[27] C. Pan, H. Ren, K. Wang, *et al.*, "Multicell MIMO communications relying on intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.

[28] T. Bai, C. Pan, Y. Deng, *et al.*, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, 2020.

[29] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent Reflecting Surface Aided Multigroup Multicast MISO Communication Systems," *IEEE Trans. Signal Process.*, vol. 68, pp. 3236–3251, 2020.

[30] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," pp. 1–6, 2019.

[31] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, 2019.

[32] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, 2020.

[33] G. Zhou, C. Pan, H. Ren, *et al.*, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, 2020.

[34] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," vol. 38, no. 11, pp. 2637–2652, 2020.

[35] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-sided secure MIMO wireless communications via intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7851–7866, 2020.

[36] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?," vol. 9, no. 6, pp. 778–782, 2020.

[37] S. Hong, C. Pan, H. Ren, K. Wang, K. K. Chai, and A. Nallanathan, "Robust transmission design for intelligent reflecting surface aided secure communication systems with imperfect cascaded CSI," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2487–2501, 2021.

[38] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, 2015.

[39] Y. Wu, A. Khisti, C. Xiao, G. Caire, K. Wong, and X. Gao, "A survey of physical layer security techniques for 5G wireless networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 679–695, 2018.

[40] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces: (invited paper)," pp. 735–740, 2019.

[41] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, 2020.

[42] X. Guan, Q. Wu, and R. Zhang, "Joint power control and passive beamforming in IRS-assisted spectrum sharing," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1553–1557, 2020.

[43] T. Bai, R. Vaze, and R. W. Heath, "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, 2014.

[44] M. R. Akdeniz, Y. Liu, M. K. Samimi, *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[45] T. Nishio *et al.*, "Proactive received power prediction using machine learning and depth images for mmWave networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2413–2427, 2019.

[46] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, 2020.

[47] D. Kumar, J. Kaleva, and A. Tolli, "Rate and reliability trade-off for mmWave communication via multi-point connectivity," in *IEEE GLOBECOM*, 2019, pp. 1–6.

[48] H. Iimori *et al.*, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, 2020.

[49] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, 2021.

[50] Z. Zhou, N. Ge, Z. Wang, and L. Hanzo, "Joint transmit precoding and reconfigurable intelligent surface phase adjustment: A decomposition-aided channel estimation approach," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1228–1243, 2021.

[51] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, 2020.

[52] P. Wang, J. Fang, H. Duan, and H. Li, "Compressed channel estimation and joint beamforming for intelligent reflecting surface-assisted millimeter wave systems," *IEEE Signal Processing Lett.*, vol. 27, pp. 905–909, 2020.

[53] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, "Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems," 2019. [Online]. Available: `https://arxiv.org/abs/1912.03619`.

[54] W. Shi, J. Li, G. Xia, *et al.*, "Secure multigroup multicast communication systems via intelligent reflecting surface," *China Communications*, vol. 18, no. 3, pp. 39–51, 2021.

[55] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82 599–82 612, 2019.

[56] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090–1094, 2020.

[57] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, 2019.

[58] Z. Chu, W. Hao, P. Xiao, and J. Shi, "Intelligent reflecting surface aided multi-antenna secure transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 108–112, 2020.

[59] G. Yang, X. Xu, Y.-C. Liang, and M. Di Renzo, "Reconfigurable intelligent surface-assisted non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3137–3151, 2021.

[60] G. Zhou, C. Pan, H. Ren, K. Wang, and Z. Peng, "Secure wireless communication in RIS-aided MISO system with hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1309–1313, 2021.

[61] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 234–238, 2021.

[62] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," *Version 2.1. [Online] http://cvxr.com/cvx, Dec. 2018.*,

[63] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[64] L. Dong and H.-M. Wang, "Secure MIMO transmission via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 787–790, 2020.

[65] Z. Peng, Z. Zhang, C. Pan, L. Li, and A. L. Swindlehurst, "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol. 69, pp. 837–851, 2021.

[66] P. Wang, J. Fang, L. Dai, and H. Li, "Joint transceiver and large intelligent surface design for massive MIMO mmWave systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052–1064, 2021.

[67] H.-M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's CSI," *IEEE Signal Process. Lett.*, vol. 27, pp. 1300–1304, 2020.

[68] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, 2020.

[69] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1735–1748, 2020.

[70] Y. Omid, S. M. Shahabi, C. Pan, Y. Deng, and A. Nallanathan, "Low-complexity robust beamforming design for IRS-aided MISO systems with imperfect channels," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1697–1701, 2021.

[71] M. Fu, Y. Zhou, and Y. Shi, "Intelligent reflecting surface for downlink non-orthogonal multiple access networks," in *Proc. IEEE Global Commun. Conf. Workshops (GC Wkshps)*, 2019, pp. 1–6.

[72] X. Hu, C. Masouros, and K.-K. Wong, "Reconfigurable intelligent surface aided mobile edge computing: From optimization-based to location-only learning-based solutions," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3709–3725, 2021.

[73] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA reconfigurable intelligent surface-aided downlink communication networks," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 664–674, 2021.

[74] H. Yang, X. Yuan, J. Fang, and Y.-C. Liang, "Reconfigurable intelligent surface aided constant-envelope wireless power transfer," *IEEE Trans. Signal Process.*, vol. 69, pp. 1347–1361, 2021.

[75] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 5092–5106, 2020.

[76] Y. Chen, Y. Wang, and L. Jiao, "Robust transmission for reconfigurable intelligent surface aided millimeter wave vehicular communications with statistical CSI," *early access in IEEE Trans. Wireless Commun.*, pp. 1–1, 2021. DOI: 10.1109/TWC.2021.3100492.

[77] J. Ye, S. Guo, and M.-S. Alouini, "Joint reflecting and precoding designs for SER minimization in reconfigurable intelligent surfaces assisted MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5561–5574, 2020.

[78] X. Hu, C. Zhong, and Z. Zhang, "Angle-domain intelligent reflecting surface systems: Design and analysis," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4202–4215, 2021.

[79] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ. Press, 2004.

[80] M. Shao, Q. Li, and W.-K. Ma, "Minimum symbol-error probability symbol-level precoding with intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1601–1605, 2020.

[81] S. Wang, Q. Li, and M. Shao, "One-bit symbol-level precoding for MU-MISO downlink with intelligent reflecting surface," *IEEE Signal Process. Lett.*, vol. 27, pp. 1784–1788, 2020.

[82] N. S. Perović, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, "On the maximum achievable sum-rate of the RIS-aided MIMO broadcast channel," 2021. [Online]. Available: https://arxiv.org/abs/2110.01700.

[83] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[84] K. Zhi, C. Pan, H. Ren, and K. Wang, "Ergodic rate analysis of reconfigurable intelligent surface-aided massive MIMO systems with ZF detectors," *early access in IEEE Commun. Lett.*, 2021.

[85] A. Papazafeiropoulos, C. Pan, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Intelligent reflecting surface-assisted MU-MISO systems with imperfect hardware: Channel estimation and beamforming design," *early access in IEEE Trans. Wireless Commun.*, pp. 1–1, 2021. DOI: 10.1109/TWC.2021.3109391.

[86] K. Zhi, C. Pan, H. Ren, and K. Wang, "Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI," 2020. [Online]. Available: https://arxiv.org/abs/2010.13525.

[87] K. Zhi, C. Pan, H. Ren, and K. Wang, "Statistical CSI-based design for reconfigurable intelligent surface-aided massive MIMO systems with direct links," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 1128–1132, 2021.

[88] K. Zhi, C. Pan, H. Ren, *et al.*, "Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI," 2021. [Online]. Available: https://arxiv.org/abs/2108.07622.

[89] J. Dai, Y. Wang, C. Pan, K. Zhi, H. Ren, and K. Wang, "Reconfigurable intelligent surface aided massive MIMO systems with low-resolution DACs," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3124–3128, 2021.

[90] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 745–749, 2020.

[91] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, 2021.

[92] Hunter, D. R., and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[93] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2017.

[94] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Springer Verlag New York*, pp. 515–545, 2016.

[95] M. Razaviyayn, *Successive convex approximation: Analysis and applications.* Ph.D. dissertation, Univ. Minnesota, Minneapolis, MN, USA, 2014.

[96] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications.* Cambridge University Press, 2011, ISBN: 9781139498043.

[97] G. Scutari and Y. Sun, "Parallel and distributed successive convex approximation methods for big-data optimization," *Multi-Agent Optim.: Cetraro Italy 2014, Berlin, Germany: Springer*, pp. 141–308, 2018.

[98] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, 2020.

[99] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: A survey and extensions," *Math. Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007.

[100] "The mosek optimization toolbox for MATLAB manual," *Version 7.1 (revision 28). [Online] http://mosek.com, accessed on: Mar. 20, 2015.*,

[101] A. Ben-Tal and A. Nemirovski, *(Lectures on modern convex optimization: Analysis, algorithms, and engineering applications). Philadelphia, PA, USA: SIAM.* MPSSIAM Ser. Optim., 2001.

[102] S. Xu, "Smoothing method for minimax problems," *Comput. Optim. Appl.*, vol. 20, no. 3, pp. 267–279, 2001.

[103] J. Pang, "Partially B-regular optimization and equilibrium problems," *Math. Oper. Res.*, vol. 32, no. 3, pp. 687–699, 2007.

[104] J. Pang, M. Razaviyayn, and A. Alvarado, "Computing B-stationary points of nonsmooth DC programs," *Math. Oper. Res.*, vol. 42, no. 1, pp. 95–118, 2017.

[105] R. Varadhan and C. Roland, "Simple and globally convergent methods for accelerating the convergence of any EM algorithm," *Scand. J. Statist.*, vol. 35, no. 2, pp. 335–353, 2008.

[106] M. W. Jacobson and J. A. Fessler, "An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2411–2422, 2007.

[107] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[108] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 581–585, 2020.

[109] O. Tervo, L. Tran, H. Pennanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multicell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, 2018.

[110] Shuguang Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, 2005.

[111] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2020.

[112] T. Nguyen, D. Nguyen, M. Di Renzo, and R. Zhang, "Leveraging secondary reflections and mitigating interference in multi-IRS/RIS aided wireless network," 2021. [Online]. Available: `https://doi.org/10.36227/techrxiv.17140004.v1`.

[113] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, 1987.

[114] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Wireless Communications. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[115] *Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 16)*. document 3GPP TR 38.901 V16.1.0, 2019.

[116] M. Di Renzo, "Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, 2015.

[117] M. Di Renzo, W. Lu, and P. Guan, "The intensity matching approach: A tractable stochastic geometry approximation to system-level analysis of cellular networks," vol. 15, no. 9, pp. 5963–5983, 2016.

[118] C. Pan, G. Zhou, K. Zhi, *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," 2021. [Online]. Available: `https://arxiv.org/abs/2112.05989`.

[119] J. Mairal, "Stochastic majorization-minimization algorithms for largescale optimization," *Advances in Neural Information Processing Systems*, pp. 2283–2291, 2013.

[120] X.-D. Zhang, *Matrix analysis and applications*. Cambridge Univ. Press, 2017.

[121] M. Botros and T. N. Davidson, "Convex conic formulations of robust downlink precoder designs with quality of service constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 714–724, 2007.

[122] J. Zhang, M. Kountouris, J. G. Andrews, and R. W. Heath, "Multimode transmission for the MIMO broadcast channel with imperfect channel state information," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 803–814, 2011.

[123] T. Lipp and S. Boyd, "Variations and extension of the convex-concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, 2016. [Online]. Available: `https://doi.org/10.1007/s11081-015-9294-x`.

[124] S. Boyd, L. G. El, E. Ferron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. Philadelphia, PA: SIAM, 1994.

[125] I. R. Petersen, "A stabilization algorithm for a class of uncertain linear systems," *Syst. Contr. Lett.*, no. 8, pp. 351–357, 1987.

[126] E. A. Gharavol and E. G. Larsson, "The sign-definiteness lemma and its applications to robust transceiver optimization for multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 238–252, 2013.

[127] K. Wang, A. M. So, T. Chang, W. Ma, and C. Chi, "Outage constrained robust transmit optimization for multiuser miso downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690–5705, 2014.

[128] 3GPP, "Technical specification group radio access network; study on 3D channel model for LTE (release 12)," *TR 36.873 V12.7.0*, 2017.

[129] S. Wang, M. Wen, M. Xia, R. Wang, Q. Hao, and Y.-C. Wu, "Angle aware user cooperation for secure massive MIMO in rician fading channel," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2182–2196, 2020.

[130] I. Bechar, "A bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables," 2009. [Online]. Available: `https://arxiv.org/abs/0909.3595`.

[131] A. Chaman, J. Wang, J. Sun, H. Hassanieh, and R. R. Choudhury, "Ghostbuster: Detecting the presence of hidden eavesdroppers," *Proc. ACM Mobico, New Delhi, Indiam*, pp. 337–351,

[132] A. Mukherjee, "Physical-layer security in the internet of things: Sensing and communication confidentiality under resource constraints," *Proc. IEEE*, vol. 103, no. 10, pp. 1747–1761, 2015.

[133] L. Mucchi, L. Ronga, X. Zhou, K. Huang, Y. Chen, and R. Wang, "A new metric for measuring the security of an environment: The secrecy pressure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3416–3430, 2017.

[134] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Massive MIMO with spatially correlated rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, 2019.

[135] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operation Research*, vol. 26, no. 4, 1978.

[136] Z. Wang, P. Babu, and D. P. Palomar, "Design of PAR-constrained sequences for MIMO channel estimation via majorization-minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6132–6144, 2016.

[137] A. A. Nasir, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Secrecy rate beamforming for multicell networks with information and energy harvesting," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 677–689, 2017.

[138] P. Maher and H. Luthepohl, "Handbook of matrices," *The Math. Gaz.*, vol. 83, no. 498, p. 557, 1999.

[139] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*. Birkhuser, Boston, 1996.

[140] N. Dunford and J. Schwartz, *Linear operators. Part 1: General theory*. Interscience Publications, New York, 1958.

[141] J. Song, P. Babu, and D. P. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, 2015.