Fast Algorithm for Joint Unicast and Multicast Beamforming in Large-Scale Systems

by

Shadi Mohammadi

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Masters of Applied Science in Electrical and Computer Engineering

The Faculty of Engineering and Applied Science

Department of Electrical and Computer Engineering

University of Ontario Institute of Technology(Ontario Tech University)

Oshawa, Ontario, Canada

September 2021

THESIS EXAMINATION INFORMATION

Submitted by: **Shadi Mohammadi**

Masters of Applied Science in **Electrical and Computer Engineering**

Thesis title: Fast Algorithm for Joint Unicast and Multicast Beamforming in Large-Scale Systems

An oral defense of this thesis took place on September 24, 2021 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee :	Dr. Ramiro Liscano
Research Supervisor :	Dr. Min Dong
Research Co-supervisor :	Dr. Shahram ShanbazPanahi
Examining Committee Member :	Dr. Ying Wang
Thesis Examiner:	Dr. Mehran Ebrahimi, Associate Professor, Faculty of Science, Ontario Tech University

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

Abstract

We consider a joint unicast and multi-group multicast beamforming problem for massive multiple-input multiple-output (MIMO) system with a large number of unicast users. We propose an alternating direction method of multipliers (ADMM)-based fast algorithm that efficiently obtains the beamforming solutions for unicast and multicast users to minimize the transmit power subject to quality-of-service constraints. Utilizing the optimal multicast beamforming structure obtained recently, we separate the original problem into unicast and multicast subproblems to be solved using the alternating optimization technique. We solved the unicast subproblem in closed-form by exploring the unicast beamforming structure, which reduces the computational complexity, substantially. For the multicast subproblem, we apply the successive convex approximation (SCA) method to solve it iteratively. Each SCA subproblem is then reformulated to the ADMM form, providing the closed-form update for the multicast subproblem. Simulation results show that the proposed algorithm achieves a near-optimal performance with low complexity for large-scale systems.

Keywords: unicast beamforming; multicast beamforming; optimal structure; large-scale optimization; alternating direction method of multipliers (ADMM)

Author's Declaration

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

Shadi Mohammadi

Statement of Contributions

Part of this thesis has been published as :

S.Mohammadi, M.Dong, S.ShahbazPanahi, "Fast Algorithm for Joint Unicast and Multicast Beamforming in Large-Scale Systems," in Proceeding of *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) , Lucca, Italy, Sep 27–30, 2021.*

I developed the research design and performed all the simulations and also the majority of witting the manuscript under the supervision of Prof. Min Dong and Prof. Shahram ShahbazPanahi. I have used standard referencing practices to acknowledge ideas, research techniques, or other materials that belong to others.

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisors Prof. Min Dong and Prof. Shahram ShahbazPanahi for giving me the opportunity to do research and their invaluable support, motivation, enthusiasm, and immense knowledge.

I would also like to thank all my group members and friends who helped me a lot in both academia and integrating myself into new culture and country, especially Niloofar, Sara, Yong, Ololade and Rosa.

I would also like to give special thanks to Mehran for being always there for me and his continuous support and understanding when undertaking my research.

Finally, I'm extremely grateful to my family for their unconditional love and support and sacrifices that they made to give me the opportunity of high quality education and prepare me for the future.

Contents

List of Figures

List of Tables

List of Abbreviations

1G	first generation
3G	3rd generation
4G	4th generation
5G	5th generation
ADMM	alternating direction method of multipliers
AO	alternating optimization
AR	augmented reality
BS	base station
CDF	cumulative distribution function
CSI	channel state information
DoF	degrees of freedom
EE	energy efficiency
FDM	frequency division multiplexing
IoT	internet of things

IoV	internet of vehicles
JMU-INA	joint multicast and unicast interference nulling and alignment
KKT	Karush-Kuhn-Tucker
LDM	layered division multiplexing
M2M	machine to machine
MIMO	multiple-input multiple-output
MBF-ADMM	multicast beamforming structure with ADMM
MBF-SCA	multicast beamforming structure with SCA
MBF-SDR	multicast beamforming structure with SDR
MMF	max-min fair
MMSE	minimum mean square error
MRT	maximum ratio transmission
MU-MIMO	multi-user MIMO
NOUM	non-orthogonal unicast and multicast
NP-hard	non-deterministic polynomial-time hard
QCQP	quadratically constrained quadratic program
QoS	quality of service
RS	rate splitting

SCA	successive convex approximation
SDR	semi-definite relaxation
SINR	signal-to-interference-plus-noise ratio
SISO	single-input-single-output
SNR	signal-to-noise ratio
SU-MIMO	single-user MIMO
TDM	time division multiplexing
UMBF-ADMM	joint unicast and multicast beamforming structure with ADMM
UMBF-SCA	joint unicast and multicast beamforming structure with SCA
WSR	weighted sum-rate
ZF	zero-forcing

Chapter 1

Introduction

1.1 Overview

Mobile communication technology has evolved remarkably over the past few decades. The first generation of wireless cellular networks (1G) was introduced around 40 years ago for making phone calls. Moving from 1G to 3G, data services have been introduced and the wireless networks were able to support multimedia services. In particular, in 3G networks, the users were able to make video calls, play online games, browse the web and watch TV online. As the demand of higher data rate grows quickly, 4th generation (4G) technology offered higher data rate and capacity with a better quality of service (QoS) for managing more data traffic. However, in recent years, due to the emerging technologies and the growth in the number of mobile users, 4G networks has been stretched to their limits to manage the high demand placed upon them. The advent of new technologies such as augmented reality (AR) , internet of things (IoT), internet of vehicles (IoV), and machine to machine (M2M) communications led to the rapid increase in wireless data usage and connectivity. Handling this massive wireless data traffic was challenging with the capabilities of previous wireless generation systems.

5th generation (5G) networks have been developed to enable ultra high data rates, coverage and connectivity, while allowing ultra low latency and energy consumption [1]. To accommodate the ever-increasing growth in wireless traffic in 5G networks, many wireless technologies have been developed. Massive multiple-input multiple-output (MIMO) systems have played a key role in the development of 5G networks and beyond. Massive MIMO, as an ultimate and most powerful form of multi–user MIMO (MU-MIMO) technology, uses a large number of antennas at a base station (BS) to improve throughput and spectral efficiency [2]. The general MU-MIMO has been around for decades, but using large number of antennas for massive MIMO transceiver design is relatively new [3]. The scalability feature is one of the important advantages of massive MIMO over traditional MIMO networks [4]. Additional benefits of massive MIMO technology include low latency, improved robustness, and enhanced security.

The increasing demand for group-oriented services such as multicast and broadcast has driven the multicast bemaforming as a vital technology in 5G networks and beyond. Multi-antenna multicast beamforming is a transmission technology for delivering common messages to multiple users simultaneously to support high-speed content distribution in wireless systems. There are different types of multicast beamforming such as single-group, multi-group ,and multi-cell multicast beamforming. Unicast beamforming is the first scenario introduced in the literature to transmit the private data streams, each addressed to a specific user [5]. Due to the rise of ever-increasing demands for delivering common data to different users in the emerging wireless services and applications, multicast beamforming has been developed as a solution to grant multiple users in a group simultaneous accesses to the common multimedia content.

The problem of a single-group multicasting was initially presented in [6], where both max-min fair (MMF) and quality of service (QoS) problem have been considered. This problem later has been extended to multi-group multicast [7–9] , and multi-cell networks [10, 11]. It has also been considered in different networks such as relay networks [12–15], cognitive radio networks [16–19] and cloud-radio networks [20].

Although the solution to unicast beamforming problems is well understood [5], solving multi-group multicast beamforming problems is challenging, as this family of problems is non-deterministic polynomial-time hard (NP-hard). Existing works developed numerical algorithms or signal processing methods to obtain suboptimal solutions. For traditional transmit antenna systems, semi-definite relaxation (SDR) has been the popular conventional method to solve this family of problems [6, 7, 11]. For large-scale systems, especially massive MIMO systems, successive convex approximation (SCA) has been adopted to reduce the computational complexity [21]. Other reduced-complexity algorithms were proposed for further improvement [22,23]. The optimal multi-group multicast beamforming structure has been recently obtained in [23], in which the inherent low-dimensional structure has been identified. This optimal structure leads to a substantial reduction of the computational complexity, independent of the number of antennas, for the massive MIMO systems.

Besides pure unicast and multicast data services, in practical systems, users requesting private and common data are often present at the same time. This requires effective beamforming design to support both type of transmissions, and has motivated studies on the joint unicast and multicast beamforming problems.

1.2 Motivation and Objective

In practical downlink transmission scenarios, data traffic is typically a mix of both private data for individuals and common data for a group of users. Effective transmission design to support both type of data is critical, especially in large-scale systems. Traditionally, the joint transmission of the private and common data has been managed by allocating different time/frequency recourses. With the massive MIMO technology, spatial multiplexing unicast and broadcast data streams via joint beamforming was studied and compared with the orthogonal resource approach in [24, 25]. In [26], spectral efficiency was analyzed for a joint unicast and multi-group multicast massive MIMO system using the maximum ratio transmission (MRT) and zero-forcing (ZF) beamforming schemes under the effect of uplink channel estimation error. However, to our best knowledge, there is no existing work on developing efficient algorithms for jointly optimal unicast and multicast beamforming design for large-scale systems.

Motivated by the above, in this thesis, we consider the problem of joint unicast and multi-group multicast downlink transmission design that is computational efficient for large-scale massive MIMO systems. Note that the joint unicast and multicast beamforming can be treated as a pure multi-group multicast beamforming scenario, where unicast users are viewed as additional multicast groups of size one. However, instead of a pure multicasting problem, we will explore both unicast and multicast beamforming features to significantly improve the computational efficiency in searching a solution.

1.3 Thesis Contribution

We consider a joint unicast and multi-group multicast transmission in cell of a massive MIMO system, where there may be a large number of unicast users. We propose a low-complexity high-performing algorithm for the joint downlink unicast and multi-group multicast beamforming problem, especially suitable for massive MIMO systems. The main contributions and research findings of this thesis are summarized below:

- Utilizing the optimal beamforming structure for multi-group multicast beamforming obtained in [27, 28], we transform the problem into a weight optimization, and then we separate this weight optimization problem into two subproblems for unicast and multicast transmissions to be solved alternatingly using the alternating optimization (AO) approach. By adopting the optimal structure, each subproblem dimension is of a much smaller size, independent of the number of antennas.

- For the unicast subproblem, we show that it resembles the pure unicast beamforming power minimization problem but with extra convex constraint. We obtain the optimal solution for the unicast subproblem in a simple closed-from. This solution provides a key step in reducing the overall computational complexity of the solution.

- For the multicast subproblem, which is a non-convex and NP-hard problem, we apply the SCA method to solve it iteratively. For each SCA iteration, we propose a fast algorithm based on the alternating direction method of multipliers (ADMM), to solve each convex approximation problem with convergence guarantee. In the proposed ADMM-based algorithm, we decompose the optimization problem into smaller-size subproblems and obtain closed-form or semi closed-form solutions to each subproblems.

- As the proposed algorithm solves the unicast subproblem separately, in a closed-form, the computational complexity grows only mildly with the number of unicast users.

- Furthermore, the proposed ADMM-based low complexity algorithm for multicast subproblem provides either closed-form or semi-closed form updates in each iteration, which incur a very low computational complexity.

- Simulation results demonstrate that the proposed algorithm has the near-optimal performance, while being scalable in both number of antennas and number of unicast users.

The research work from this thesis has resulted in the following publications that has been either accepted or in preparation:

1. S.Mohammadi, M.Dong, S.ShahbazPanahi, "Fast Algorithm for Joint Unicast and Multicast Beamforming in Large-Scale Systems," in Proceeding of *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) , Lucca, Italy, Sep 27–30, 2021*

2. S.Mohammadi, M.Dong, S.ShahbazPanahi, "ADMM-based Fast Algorithm for Joint Unicast and Multicast Beamforming in Large-Scale Systems," to be submitted to *IEEE Transactions on Wireless Communications, August 2021*

1.4 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2, a literature review on beamforming techniques, and benefits and challenges of massive MIMO technology is presented. In Chapter 3, the ADMM-based fast algorithm for the joint unicast and multicast beamforming is developed . In Chapter 4, the simulation results on performance and convergence analysis are given and discussed. The conclusion and future work of this thesis is written in Chapter 5.

1.5 Notation

The main notations used in this thesis are summarized below. Hermitian, transpose, and conjugate are denoted as $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$, respectively. The Euclidean norm of a vector is denoted by $\|\cdot\|$. The notation $\mathbf{a} \succcurlyeq \mathbf{0}$ means element-wise non-negative, $\mathbf{A} \succcurlyeq 0$ indicates matrix $\mathbf{A}$ being positive semi-definite, and $[\mathbf{A}]_{ij}$ indicates the (i, j)th element in $\mathbf{A}$. The real part of x is denoted by $\mathfrak{Re}\{x\}$, and $E(x)$ denotes the expectation of x. The abbreviation i.i.d. stands for independent and identically distributed, and $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ means $\mathbf{x}$ is a complex Gaussian random vector with zero mean and covariance $\mathbf{I}$.

Chapter 2

Literature Review

2.1 Multi-antenna Downlink Beamforming

Transmit beamforming is a powerful technique for directional signal transmission from a multi-antenna base station to one or multiple users [29]. Beamforming techniques can be divided into unicast, multicast, and the joint unicast-multicast beamfoming. In this chapter, we give an overview of multi-antenna downlink beamforming techniques and the challenges we face in developing beamforming solutions in massive MIMO systems.

2.1.1 Beamforming

Beamforming is a signal processing technique for directional signal transmission in a way that increases the signal power at the intended user, and reduces interference with nearby users along the route. It is a powerful approach to transmit or receive signals of interest to achieve spatial diversity and throughput improvements, and there is a very rich literature on beamforming design in various systems or scenarios. This technology also can boost the performance of multicasting services [30]. Depending upon the channel settings, we can implement beamforming technology in several different ways in wireless networks. Beamforming improves the spectrum efficiency and data rate in massive MIMO systems. In summary, beamforming can bring better perfomance and reduced interference to the system and has been studied widely in several different communication networks [5–11,21,31–37]. The focus of this thesis is on joint unicast and multicast beamforming. In order to understand this joint unicast and multicast scenario, in following, we will first survey the pure unicast and multicast beamforming. Then, we will review the several techniques for combining unicast and multicast technologies.

2.1.2 Unicast Beamforming

Unicast beamforming is the most common scenario for data transmission in wireless systems. It is used, for dedicated data transmission to each individual user, which has been studied extensively in the literature. In this case, base station sends one private data stream to only one user. Traditional unicast beamforming has been studied in several works in different research scopes e.g. exploring QoS and MMF problem [5,31–33]. In particular, for downlink multi-user unicast beamforming, a simple closed-form solution for unicast power minimization problem has been obtained in [33]. This solution has a simple and intuitive structure with only one design parameter per user. In this thesis, we will explore this feature of unicast beamforming to reduce the overall computational complexity of a joint unicast and multicast problem.

With the rise of new wireless technologies and applications, including video sharing and mass advertisement, the demand for delivering common data to different users has been increased. Therefore, multicast beamforming has been developed as a solution for transmitting common data to multiple users.

2.1.3 Multicast Beamforming

For data multicasting, each group of users receive the common data stream, simultaneously. The initial problem of a single-group multicasting was presented in [6]. The problem then has been extended to the multi-group multicasting [7–9]. Later, the problem has been extended to multi-cell [10, 11], where the inter-cell interference has been considered. The multicast beaforming problem has been studied in other scenarios like relay systems [12–15] and cognitive radio networks [16–19].

Among several different objective designs, QoS and MMF problems are the most commonly used problem formulations for the multicast beamforming transmission. In QoS problem, beamforming vectors are designed to minimize the transmit power, while satisfying QoS constraints. Several studies have been exploring the QoS problem for the multicast beamforming [7, 38–40]. Although in the single-group multicast beamforming, the QoS problem is always feasible, the multi-group multicast scenario needs to satisfy a condition on the signal-to-interference-plus-noise ratio (SINR) target to make the QoS problem feasible [35]. The MMF problem formulation is designed for maximizing the minimum SINR among all users subject to a constraint that the transmit power is less than a desired value. Based on the inverse relation between the QoS and the MMF problems [7], the solution to the MMF problem can be obtained by solving the QoS problem iteratively [21, 23, 40].

Multicast beamforming problems have been proved to be NP-hard [6]. Numerical algorithms have been developed in the literature to obtain the suboptimal solution for traditional antenna systems with only a few number of antennas at the BS. SDR is a popular technique to solve these multicasting problems by relaxing them to semi-definite problems (SDP) [6, 7, 34, 35]. However, as the number of antennas increases, in addition to deterioration of performance,the SDR-based methods suffer from high computational complexity. This behaviour is highly restraining especially in large-scale antenna array systems like massive MIMO networks.

SCA is an iterative method which has been proposed by [36] to solve a single-group multicasting problem. This work was later extended to multi-group [21] and multi-cell [37] systems, and the results show that the SCA-based methods have better performance than the SDR-based algorithms in large-scale systems. Despite the fact that the SCA-based method has a near-optimal performance, it still has a high computational complexity for massive MIMO systems, which grows fast with the number of antennas. Therefore, there is an increasing demand for efficient algorithms to solve the multicast beamforming problems in large-scale antenna array systems.

Several studies have been conducted on the multicast beamforming design to improve computational complexity in massive MIMO for multi-group [22, 23], and multi-cell networks [41, 42]. These studies include combining MRT and ZF precoding schemes with the SCA method or applying ADMM to reduce complexity. Despite these studies, still there was no fundamental understanding of the beamforming structure for the multicasting problem and in the practical systems, the existing algorithms faced challenges both in performance and computational complexity for large-scale systems.

To address these issues, recently, the optimal beamforming solution structure for the multi-group multicasting problem has been obtained by [27], where the optimal beamforming structure for both QoS and MMF problems have been derived. By extending the uplink-downlink duality to the multicasting problem, it has been shown that the multi-group multicast beamformer has an intuitive weighted minimum mean square error (MMSE) structure based on a group-channel direction.

Additionally, it has been shown that the optimal multicast beamformer has an inherent low-dimensional structure that is independent of the number of transmit antennas. For finding the optimal multi-group multicast solution, the problem has been converted to the weight optimization problem of a smaller size. Based on this structure, efficient algorithms have been proposed by [27] to obtain the weights. Exploiting the small size of the problem, both SDR and SCA methods have been applied to obtain the weights and therefore the optimal solution of the multi-group multicast beamforming problem. This optimal structure has been used by other studies to develop fast first-order algorithm for multi-group multicast beamforming with the QoS [43] and the MMF problem formulation [44].

2.1.4 Joint Unicast and Multicast Beamforming

A practical system should have the capability of combining unicast and multicast beamforming techniques to transmit both private and common data streams, at the same time. Traditionally, time division multiplexing (TDM)/ frequency division multiplexing (FDM) were used for the joint transmission of the private and common data by allocating different time/frequency recourses to the unicast and multicast services. However, one of the main issues with TDM/FDM is the low spectrum efficiency caused by the fact that we can't reuse time/frequency. To resolve this problem, spatial multiplexing has been developed in [45], and [25], where the unicast and multicast data streams are transmitted with different beamformers to separate different messages in the spatial domain.

Later, layered division multiplexing (LDM) [46] has been developed by [47] for the joint unicast and multicast transmission, as a form of non-orthogonal technology. LDM was accepted as a baseline physical layer in digital TV standard ATSC 3.0 [48]. A layered transmission structure is supported by an LDM system to transmit multiple signals in a different power levels and robustness. It has been shown that LDM can achieve higher spectral efficiency than TDM/FDM [49]. From an information-theoretic perspective, the concept of LDM has been described as a form of superposition coding [50]. Such non-orthogonal unicast and multicast (NOUM) transmission has been studied with different optimization targets, .e.g., maximizing the weighted sum-rate (WSR) [51, 52], the energy efficiency (EE) [53, 54], and minimizing the transmit power [55].

All the aforementioned studies are mainly suited to the under-loaded regimes, where the number of data streams are smaller than number of antennas. To further address the need of considering systems with fewer antennas than users, the rate splitting (RS) technique has been recently introduced in [56] as a promising multi-user non-orthogonal transmission technology for MIMO networks. In RS technique, each data stream is split into a private and a common part. The common part can be recovered by all the users, while the private part is only decoded by the intended user. Recently, the RS technique has been investigated for the joint unicast and multicast transmission in [57], where WSR and EE problems are studied. In that respect, the unicast messages are split into private and common parts and the common parts can be encoded along with the multicast messages into super-common stream decoded by all the users. Also, rate splitting method has been proposed in overloaded systems for pure multi-group multicast transmission [58–60], as well as pure unicast problem [61].

Besides the above mentioned works, the design and implementation of HybridCast has been presented in [24]. HybridCast is a MU-MIMO system that leverages unused degrees of freedom (DoF) and available link margin to enable concurrent unicast and multicast transmission. Moreover, [24] propose a precoding scheme, called joint multicast and unicast interference nulling and alignment (JMU-INA) to ensure that the multicast rate, will not be affected by the concurrent unicast streams. However, [24] neither studies QoS problem nor considers the multi-group multicast transmission.

2.2 Massive MIMO Technology

Massive MIMO has been introduced few years ago and become a key technology in 5G and beyond. Traditionally, single-input-single-output (SISO) systems were mostly used, which could not support large number of users with high reliability. Due to the rapid growth of mobile devices and development of new wireless technologies, the demand for wireless throughput has been increased in recent years. To accommodate the ever-increasing user demand, various new MIMO technology were developed. Different types of MIMO systems are single-user MIMO (SU-MIMO) [62,63], MU-MIMO [64–66] of which massive MIMO is developing as its most powerful form [2, 67, 68]. Massive MIMO is an extension of multi-user MIMO which brings together dozens or hundreds of antennas at the base station simultaneously and serves tens of users in the same time/frequency resource [3]. With extra antennas the radiated beam can be spatially focused towards the intended user and reduces the interference from nearby users. As the number of antennas increases, the spectral efficiency and throughput of the system improves.

2.2.1 Benefits of Massive MIMO

In massive MIMO systems, the capacity is expected to be increased ten times or more, and at the same time the network improves the radiated energy efficiency on the order of a hundred times [2]. Indeed, with a large number of antennas, energy can be focused with more sharpness into smaller region leading to this dramatic improvement in energy efficiency. Furthermore, massive MIMO boosts the robustness against both unintended man-made interference and internal jamming [69]. Also, a large number of antennas results in robustness of a system to individual antenna failures . The high spectral efficiency and reliability in massive MIMO networks are the results of high multiplexing gain and diversity gain, respectively [68]. Additional benefits of massive MIMO technology are high data rate, user tracking, low power consumption, less fading, low latency, and enhanced security [70]. Ultimately, due to the aforementioned benefits ,the network can offer a better overall user experience in massive MIMO systems. In literature, there are several works exploring the benefits of massive MIMO systems, including the spectral efficiency [71, 72], the energy efficiency [72], and the system capacity [73].

2.2.2 Challenges of Massive MIMO

Although beamforming technique increases throughput and reduces interference, but it adds to the computational complexity of a system. Additionally, This computational complexity increases with the number of antennas which could be a problem in large-scale systems like massive MIMO networks. As the wireless networks are moving towards large-scale (in both the number of transmit antennas and users), developing scalable transmit beamforming solutions in such scenarios for large-scale systems is of a critical importance. Therefore, finding a low complex and efficient beamfomers for massive MIMO systems has become an essential area of research.

There has been several studies working on reducing the computational complexity of the multicast beamforming design algorithms in massive MIMO networks for multi-group [22, 23], and multi-cell systems [41, 42]. The optimal multi-group multicast beamforming structure has been recently obtained in [27], in which the inherent low-dimensional structure has been developed. In this thesis, we utilize this optimal structure to develop fast algorithm for joint unicast and multicast problem in massive MIMO network.

ADMM [74] is a robust and fast numerical method for solving large-scale problems. It has been widely considered in recent years in many applications [23, 75–80]. The ADMM method has been increasingly attracting attention in the beamforming design problems for large-scale systems, due to it's benefits in reducing the problem size as well as computational complexity. This method has been studied in different works to develop a fast algorithm for multi-group multicast beamforming problem under a perfect channel state information (CSI) [23, 80] or the robust design [78, 79].

Massive MIMO systems with the capability of joint unicast and multicast beamforming could bring us several advantages particularly where both uincast and multicast data traffics are present. There are several researches in the literature studying the combination of joint unicast-multicast beamforming and massive MIMO networks for different objectives [24–26, 52, 81, 82]. However, none of these studies explores the beamfomring design QoS problem while accounting for computational complexity in large-scale systems.

In [24], the HybridCast has been proposed as a MU-MIMO system that leverages unused DoF and available link margin for transmitting unicast and multicast data streams at the same time. Spatial multiplexing has been developed in [25] by separating different messages in spatial domain for the transmission of unicast and multicat data streams in massive MIMO system. Authors in [26], analysed spectral efficiency for the problem of joint unicast and multi-group multicast massive MIMO system using MRT and ZF beamforming schemes while considering the effect of uplink channel estimation error. In [52], authors have investigated non-orthogonal unicast and multicast transmission for massive MIMO to maximize a weighted sum of the achievable ergodic unicast rate and multicast rate.

A hybrid unicast/multicast transmission systems has been proposed by [81] in which the multicast groups adopt asymptotic multicast beamforming while the unicast group adopts MU-MIMO linear precoding. A first-order algorithm was developed in [82] to solve the EE maximization problem for the LDM-based nonorthogonal multicast and unicast transmission in a cell-free massive MIMO system.

Along with the computational complexity, there are other challenges that need to be addressed in massive MIMO networks, including pilot contamination, channel estimation, signal detection, user scheduling, and hardware impairment. In literature, researchers addressed the aforementioned challenges in several works such as mitigating pilot contamination in [83, 84], developing algorithm for channel estimation in [85, 86] and finding an optimal user scheduling algorithm in [87].

Chapter 3

ADMM-based Fast Algorithm for Joint Unicast-Multicast Beamforming

In this chapter, we consider the QoS problem for the joint unicast and multicast beamforming scenario. Utilizing the optimal multicast beamforming structure recently obtained, we propose a low-complexity ADMM-based algorithm that efficiently obtains the beamforming solutions for both unicast and multicast users. In the following, first we introduce our system model and formulate the optimization problem. Then, we briefly describe the optimal multicast beamforming structure and later we use this structure to develop our proposed algorithm for joint unicast-multicast beamforming problem.

3.1 System Model

We consider a joint downlink unicast and multi-group multicast transmission scenario, as shown in Fig.3.1. The base station (BS) is equipped with N antennas and simultaneously serves K_{u} unicast users and G multicast groups. Each multicast group has

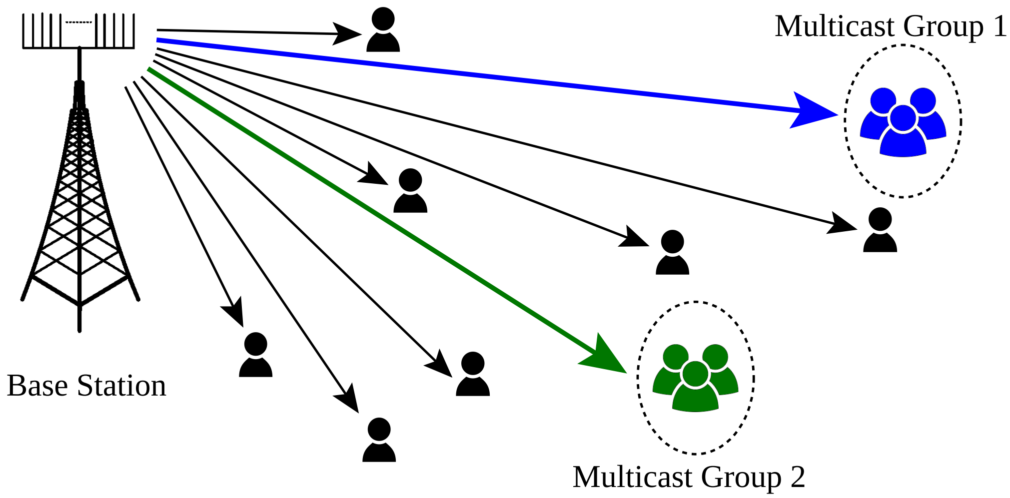

Figure 3.1: An example of mixed unicast and multicast scenario.

$K_{\mathrm{m}} \geq 2$ users[1]. All users are assumed to have a single antenna. A common message is sent to all users in a multicast group, and messages intended for all multicast groups and unicast users are independent to each other. Let $\mathcal{K}_{\mathrm{u}} \triangleq \{1, ..., K_{\mathrm{u}}\}$ denote the index set of all unicast user indexes. Let $\mathcal{G} \triangleq \{1, ..., G\}$ and $\mathcal{K}_{\mathrm{m}} \triangleq \{1, ..., K_{\mathrm{m}}\}$ denote the set of multicast group indexes and the set of user indexes in each group, respectively. Note that a user in the system is either a unicast user or a multicast user, and a multicast user can only be associated with one multicast group.

Let $\mathbf{g}_j$ denote the $N \times\ 1$ channel vector between the BS and unicast user $j \in \mathcal{K}_{\mathrm{u}}$, and $\mathbf{u}_j$ the $N \times\ 1$ unicast beamforming vector for user j. Let $\mathbf{h}_{ik}$ denote the $N \times 1$ channel vector between the BS and multicast user $k \in \mathcal{K}_{\mathrm{m}}$ in group $i \in \mathcal{G}$, and $\mathbf{w}_i$ the $N \times 1$ multicast beamforming vector for group i. The signals received at unicast user

[1]For notation simplicity, we assume that all multicast groups have the same number of users. The model can be straightforwardly extended to the general scenario, where different multicast groups may have different number of users.

$j \in \mathcal{K}_\mathrm{u}$ is given by

$$y_j^\mathrm{u} = \mathbf{u}_j^H \mathbf{g}_j s_j^\mathrm{u} + \sum_{j' \in \mathcal{K}_\mathrm{u}^{-j}} \mathbf{u}_{j'}^H \mathbf{g}_j s_{j'}^\mathrm{u} + \sum_{i \in \mathcal{G}} \mathbf{w}_i^H \mathbf{g}_j s_i^\mathrm{m} + n_j \tag{3.1}$$

where $\mathcal{K}_\mathrm{u}^{-j} \triangleq \mathcal{K}_\mathrm{u} \backslash \{j\}$, s_j^u is the data symbol intended for user j with $\mathbb{E}[|s_j^\mathrm{u}|^2] = 1$, and n_j is the additive white Gaussian noises at this user's receiver with zero mean and variance σ^2. The second and third terms in (3.1) correspond to the interference from other unicast users and from all multicast users, respectively. Similarly, the signal received at multicast user $k \in \mathcal{K}_\mathrm{m}$ in group $i \in \mathcal{G}$ is given by

$$y_{ik}^\mathrm{m} = \mathbf{w}_i^H \mathbf{h}_{ik} s_i^\mathrm{m} + \sum_{l \in \mathcal{G}^{-i}} \mathbf{w}_l^H \mathbf{h}_{ik} s_l^\mathrm{m} + \sum_{j \in \mathcal{K}_\mathrm{u}} \mathbf{u}_j^H \mathbf{h}_{ik} s_j^\mathrm{u} + n_{ik} \tag{3.2}$$

where $\mathcal{G}^{-i} \triangleq \mathcal{G} \backslash \{i\}$, s_i^m is the data symbol intended for user k in group i with $\mathbb{E}[|s_i^\mathrm{m}|^2] = 1$, and n_{ik} is the additive white Gaussian noise at this user's receiver with zero mean and variance σ^2. The received SINRs at unicast user j and multicast user k in group i are respectively given by

$$\mathrm{SINR}_j^\mathrm{u} = \frac{|\mathbf{u}_j^H \mathbf{g}_j|^2}{\sum_{j' \in \mathcal{K}_\mathrm{u}^{-j}} |\mathbf{u}_{j'}^H \mathbf{g}_j|^2 + \sum_{i \in \mathcal{G}} |\mathbf{w}_i^H \mathbf{g}_j|^2 + \sigma^2}, \tag{3.3}$$

$$\mathrm{SINR}_{ik}^\mathrm{m} = \frac{|\mathbf{w}_i^H \mathbf{h}_{ik}|^2}{\sum_{l \in \mathcal{G}^{-i}} |\mathbf{w}_l^H \mathbf{h}_{ik}|^2 + \sum_{j \in \mathcal{K}_u} |\mathbf{u}_j^H \mathbf{h}_{ik}|^2 + \sigma^2}. \tag{3.4}$$

The total transmit power at the BS is

$$P_\mathrm{tot} = \sum_{j \in \mathcal{K}_\mathrm{u}} \|\mathbf{u}_j\|^2 + \sum_{i \in \mathcal{G}} \|\mathbf{w}_i\|^2. \tag{3.5}$$

In this thesis, we focus on the QoS problem of jointly optimizing the unicast and multicast beamforming vectors $\{\mathbf{u}_j\}$ and $\{\mathbf{w}_i\}$ to minimize the total transmit power at the BS, while meeting the minimum received SINR target at each user. The problem is formulated as

$$\begin{aligned}\mathcal{P}_{\rm o}: \min_{\{\mathbf{u}_j\},\{\mathbf{w}_i\}} \ & \sum_{j\in\mathcal{K}_{\rm u}} \|\mathbf{u}_j\|^2 + \sum_{i\in\mathcal{G}} \|\mathbf{w}_i\|^2 \\ \text{s.t.} \quad & \text{SINR}^{\rm u}_j \geq \gamma^{\rm u}_j, \quad j\in\mathcal{K}_{\rm u} & \text{(3.6a)} \\ & \text{SINR}^{\rm m}_{ik} \geq \gamma^{\rm m}_{ik},\ k\in\mathcal{K}_{\rm m}, i\in\mathcal{G} & \text{(3.6b)}\end{aligned}$$

where $\gamma^{\rm u}_j$ and $\gamma^{\rm m}_{ik}$ are the SINR targets for unicast user j and multicast user k in group i, respectively. Since the multi-group multicast beamforming problem is NP hard [6], $\mathcal{P}_{\rm o}$ is an NP-hard problem. Note that $\mathcal{P}_{\rm o}$ can be treated as a pure multi-group multicast beamforming problem, where $K_{\rm u}$ unicast users are viewed as additional $K_{\rm u}$ multicast groups of size one. However, instead of considering $\mathcal{P}_{\rm o}$ as a pure multicasting problem, we will explore both unicast and multicast beamforming structures to compute the solution with significantly improved computational efficiency. The optimal multicast beamforming structure for multi-group multicast beamforming has been obtained in [27]. In what follows, we first briefly describe this optimal structure. Then, exploring both unicast and multicast beamforming features, we develop a fast algorithm to solve $\mathcal{P}_{\rm o}$.

3.2 Multi-group Multicast Beamforming Structure

The optimal structure of multi-group multicast beamforming for the QoS problem is obtained [27]. In this section, we briefly describe this structure. Consider a general multi-group multicast scenario with $\widetilde{G}$ multicasting groups, each group i having $\widetilde{K}_i$ users.[2] The set of group indexes and the set of user indexes in each group are

[2]Here, we consider a general multicast setting with $\widetilde{K}_i \geq 1$ for each group i , where a unicast user can be considered as a special case where $\widetilde{K}_i = 1$.

respectively denoted by $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{K}}_i$. The SINR at user k in group i is given by

$$\text{SINR}_{ik} = \frac{|\mathbf{w}_i^H \mathbf{h}_{ik}|^2}{\sum_{l \in \widetilde{\mathcal{G}}-i} |\mathbf{w}_l^H \mathbf{h}_{ik}|^2 + \sigma^2}, \quad k \in \widetilde{\mathcal{K}}_i, i \in \widetilde{\mathcal{G}}. \tag{3.7}$$

The QoS problem for power minimization while meeting the SINR targets is given by

$$\min_{\{\mathbf{w}_i\}} \sum_{i \in \mathcal{G}} \|\mathbf{w}_i\|^2 \tag{3.8a}$$

$$\text{SINR}_{ik} \geq \tilde{\gamma}_{ik}, \; k \in \widetilde{\mathcal{K}}_i, i \in \widetilde{\mathcal{G}} \tag{3.8b}$$

where $\tilde{\gamma}_{ik}$ is the SINR target for user k in group i. Although the optimization problem in (3.8) is an NP-hard problem, it is shown in [27] that the optimal solution to problem (3.8) is a weighted MMSE filter, given by

$$\mathbf{w}_i = \widetilde{\mathbf{R}}^{-1}(\boldsymbol{\lambda}) \mathbf{H}_i \mathbf{a}_i, \quad i \in \widetilde{\mathcal{G}} \tag{3.9}$$

where $\mathbf{H}_i \triangleq [\mathbf{h}_{i1}, ..., \mathbf{h}_{i\widetilde{K}_i}]$ is the channel matrix for group i, $\mathbf{a}_i$ is a $\widetilde{K}_i \times 1$ weight vector for group $i \in \widetilde{\mathcal{G}}$, and $\widetilde{\mathbf{R}}(\boldsymbol{\lambda}) \triangleq \mathbf{I} + \sum_{i \in \widetilde{\mathcal{G}}} \sum_{k \in \widetilde{\mathcal{K}}_i} \lambda_{ik} \tilde{\gamma}_{ik} \mathbf{h}_{ik} \mathbf{h}_{ik}^H$ is the noise plus weighted channel covariance matrix, in which λ_{ik} is the Lagrange multiplier associated with the SINR constraint in (3.8b), and $\boldsymbol{\lambda} \triangleq [\boldsymbol{\lambda}_1^T, ..., \boldsymbol{\lambda}_{\widetilde{G}}^T]^T$ with $\boldsymbol{\lambda}_i \triangleq [\lambda_{i1}, ..., \lambda_{i\widetilde{K}_i}]^T$.

The solution $\mathbf{w}_i$ in (3.9) is a semi-closed-form solution, where $\boldsymbol{\lambda}$ and $\{\mathbf{a}_i\}$ need to be numerically determined. The Lagrange multiplier vector $\boldsymbol{\lambda}$ can be approximately computed by the simple fixed-point iterative method proposed in [27], which is shown to be asymptotically optimal as $N \to \infty$. It is summarized below:

1. Initialize $\boldsymbol{\lambda}^{(0)}$; Set $n = 0$.

2. Compute $\lambda_{ik}^{(n+1)}$ for each $k \in \mathcal{K}$, $i \in \mathcal{G}$ as

$$\lambda_{ik}^{(n+1)} = \frac{1}{(1+\tilde{\gamma}_{ik})\mathbf{h}_{ik}^H \widetilde{\mathbf{R}}^{-1}(\boldsymbol{\lambda}^{(n)})\mathbf{h}_{ik}}.$$

3. Set $n = n+1$; Repeat Steps 2-3 until convergence.

Given $\boldsymbol{\lambda}$ and $\mathbf{w}_i$ in (3.9), $\mathcal{P}_\text{o}$ is then transformed into a weight optimization problem with respect to (w.r.t) weight vectors $\{\mathbf{a}_i\}$. The transformed problem can be solved via existing numerical methods, such as the SCA method used in [27]. There are total $\sum_{i\in\widetilde{\mathcal{G}}} \widetilde{K}_i$ variables in $\{\mathbf{a}_i\}$. Compared with the original problem $\mathcal{P}_\text{o}$ with $\widetilde{G}N$ variables, the transformed weight optimization problem is of a much smaller dimension than $\mathcal{P}_\text{o}$ for massive MIMO systems with $\widetilde{K}_i \ll N$. Utilizing this inherent low-dimensional structure of the optimal solution $\mathbf{w}_i$ in (3.9) can lead to significant computational saving in obtaining the multicast beamforming solution in massive MIMO systems. Next, based on this optimal beamforming structure, we develop a new approach to solve the joint unicast and multicast problem.

3.3 Algorithm for Joint Unicast-Multicast Beamforming

It is common that a system may contain many unicast users and only a few multicast groups for data transmission. In such scenario, directly treating the beamforming problem as a multi-group multicast beamforming and using the method in [27] to solve may not be efficient. In this section, we propose a fast algorithm for this mixed unicast and multicast scenario.

As discussed in the previous section, treating all the unicast users as single-user groups, we can view our problem as a multi-group multicast beamforming problem, where the optimal solution to $\mathcal{P}_{\mathrm{o}}$ is given by (3.9). Specifically, at the optimum, the unicast and multicast beamforming vectors for the joint problem have the following structures:

$$\text{Unicast}: \quad \mathbf{u}_j = b_j \mathbf{R}^{-1}(\boldsymbol{\lambda}, \boldsymbol{\nu})\mathbf{g}_j, \quad j \in \mathcal{K}_{\mathrm{u}} \tag{3.10}$$

$$\text{Multicast}: \quad \mathbf{w}_i = \mathbf{R}^{-1}(\boldsymbol{\lambda}, \boldsymbol{\nu})\mathbf{H}_i \mathbf{a}_i, \quad i \in \mathcal{G} \tag{3.11}$$

where

$$\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq \mathbf{I} + \sum_{i \in \mathcal{G}} \sum_{k \in \mathcal{K}_{\mathrm{m}}} \lambda_{ik} \gamma_{ik}^{\mathrm{m}} \mathbf{h}_{ik} \mathbf{h}_{ik}^H + \sum_{j \in \mathcal{K}_{\mathrm{u}}} \nu_j \gamma_j^{\mathrm{u}} \mathbf{g}_j \mathbf{g}_j^H, \tag{3.12}$$

with ν_j and λ_{ik} being the Lagrange multipliers associated with the SINR constraints for unicast users in (3.6a) and multicast users in (3.6b), respectively, and $\boldsymbol{\nu} \triangleq [\nu_1, \ldots, \nu_{K_u}]^T$; also, b_j denotes the weight for unicast user j. Since the beamforming solution $\mathbf{u}_j$ is unique up to a phase rotation, without loss of optimality, we assume b_j is real.

The Lagrange multiplier vectors $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ for the solutions in (3.10) and (3.11) can be computed using an approximation method similar to the one shown in Section 3.2. Specifically, we obtain λ_{ik} and ν_j by solving the following fixed-point equations:

$$\begin{aligned} \lambda_{ik} &= \frac{1}{(1+\gamma_{ik}^{\mathrm{m}})\mathbf{h}_{ik}^H \mathbf{R}^{-1}(\boldsymbol{\lambda}, \boldsymbol{\nu})\mathbf{h}_{ik}}, \quad k \in \mathcal{K}_{\mathrm{m}}, i \in \mathcal{G}, \\ \nu_j &= \frac{1}{(1+\gamma_j^{\mathrm{u}})\mathbf{g}_j^H \mathbf{R}^{-1}(\boldsymbol{\lambda}, \boldsymbol{\nu})\mathbf{g}_j}, \quad j \in \mathcal{K}_{\mathrm{u}}. \end{aligned} \tag{3.13}$$

The fixed-point iterative method for computing $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is summarized below.

S1) Initialize $\boldsymbol{\lambda}^{(0)}$ and $\boldsymbol{\nu}^{(0)}$. Set $n = 0$.

S2) Compute $\lambda_{ik}^{(n+1)}$ for each $k \in \mathcal{K}_{\mathrm{m}}$, $i \in \mathcal{G}$:

$$\lambda_{ik}^{(n+1)} = \frac{1}{(1+\gamma_{ik}^{\mathrm{m}})\mathbf{h}_{ik}^H \mathbf{R}^{-1}(\boldsymbol{\lambda}^{(n)}, \boldsymbol{\nu}^{(n)})\mathbf{h}_{ik}}.$$

S3) Compute $\nu_j^{(n)}$ for each $j \in \mathcal{K}_{\mathrm{u}}$:

$$\nu_j^{(n+1)} = \frac{1}{(1+\gamma_j^{\mathrm{u}})\mathbf{g}_j^H \mathbf{R}^{-1}(\boldsymbol{\lambda}^{(n)}, \boldsymbol{\nu}^{(n)})\mathbf{g}_j}.$$

S4) Set $n \leftarrow n+1$; Repeat S1)-S4) until convergence.

Define $\mathbf{b} \triangleq [b_1, \ldots, b_{K_{\mathrm{u}}}]^T$. Once $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ are obtained, using $\mathbf{u}_j$ in (3.10) and $\mathbf{w}_i$ in (3.11), we transform $\mathcal{P}_{\mathrm{o}}$ to a joint weight optimization problem w.r.t. weight vectors $\mathbf{b}$ and $\{\mathbf{a}_i\}$ for the respective unicast and multicast users, given as follows:

$$\mathcal{P}_{\mathrm{o}}' : \min_{\mathbf{b}, \{\mathbf{a}_i\}} \sum_{i \in \mathcal{G}} \|\widetilde{\mathbf{H}}_i \mathbf{a}_i\|^2 + \sum_{j \in \mathcal{K}_{\mathrm{u}}} b_j^2 \|\tilde{\mathbf{g}}_j\|^2 \tag{3.14a}$$

$$\text{s.t.} \quad \frac{b_j^2 |\tilde{\mathbf{g}}_j^H \mathbf{g}_j|^2}{\sum_{j' \in \mathcal{K}_{\mathrm{u}}^{-j}} b_{j'}^2 |\tilde{\mathbf{g}}_{j'}^H \mathbf{g}_j|^2 + \sum_{i \in \mathcal{G}} |\mathbf{a}_i^H \widetilde{\mathbf{H}}_i^H \mathbf{g}_j|^2 + \sigma^2} \geq \gamma_j^{\mathrm{u}}, \ j \in \mathcal{K}_{\mathrm{u}}, \tag{3.14b}$$

$$\frac{|\mathbf{a}_i^H \widetilde{\mathbf{H}}_i^H \mathbf{h}_{ik}|^2}{\sum_{l \in \mathcal{G}^{-i}} |\mathbf{a}_l^H \widetilde{\mathbf{H}}_l^H \mathbf{h}_{ik}|^2 + \sum_{j \in \mathcal{K}_{\mathrm{u}}} b_j^2 |\tilde{\mathbf{g}}_j^H \mathbf{h}_{ik}|^2 + \sigma^2} \geq \gamma_{ik}^{\mathrm{m}}, \ k \in \mathcal{K}_{\mathrm{m}}, i \in \mathcal{G} \tag{3.14c}$$

where $\tilde{\mathbf{g}}_j \triangleq \mathbf{R}^{-1}(\boldsymbol{\lambda}, \boldsymbol{\nu})\mathbf{g}_j$, $\widetilde{\mathbf{H}}_i \triangleq \mathbf{R}^{-1}(\boldsymbol{\lambda}, \boldsymbol{\nu})\mathbf{H}_i$.

Note that using the optimal beamforming structures reduces the problem size significantly. Problem $\mathcal{P}_{\mathrm{o}}'$ has $G(K_{\mathrm{m}}) + K_{\mathrm{u}}$ variables and $G(K_{\mathrm{m}}) + K_{\mathrm{u}}$ constraints, which no longer depends on N. However, problem $\mathcal{P}_{\mathrm{o}}'$ is still NP-hard and requires suitable numerical methods to solve. For relatively large K_{u}, the computational complexity for directly solving $\mathcal{P}_{\mathrm{o}}'$ via the SCA method is still high. Instead, we note that the optimal solution of a pure multi-user unicast beamforming problem is known in the literature, which can be computed efficiently [5]. Thus, we utilize the unicast beamforming solution structure to improve the computational efficiency for $\mathcal{P}'_{\mathrm{o}}$. Specifically, we propose an alternating optimization method (AO) to solve the weight vectors $\mathbf{b}$ and $\{\mathbf{a}_i\}$ alternatingly, *i.e.,* we solve $\mathcal{P}_o$ for $\{\mathbf{a}_i\}$ while fixing $\mathbf{b}$, and vice versa. Following this, in the next two subsections, we propose different methods to obtain the weight vectors $\mathbf{b}$ and $\{\mathbf{a}_i\}$ for the unicast and multicast users, respectively.

3.3.1 Obtaining Weight Vector For Unicast Users

With fixed weight vectors $\{\mathbf{a}_i\}$ for multicast users, $\mathcal{P}'_{\mathrm{o}}$ is equivalent to the following weight optimization problem for the unicast users:

$$\mathcal{P}_{\mathrm{u}}: \min_{\mathbf{b}} \sum_{j\in\mathcal{K}_{\mathrm{u}}} b_j^2\|\tilde{\mathbf{g}}_j\|^2 \tag{3.15a}$$

$$\text{s.t.} \quad \frac{b_j^2|\tilde{\mathbf{g}}_j^H\mathbf{g}_j|^2}{\sum_{j'\in\mathcal{K}_{\mathrm{u}}^{-j}} b_{j'}^2|\tilde{\mathbf{g}}_{j'}^H\mathbf{g}_j|^2 + I_{\mathrm{u},j}^{\mathrm{m}} + \sigma^2} \geq \gamma_j^{\mathrm{u}},\ j\in\mathcal{K}_{\mathrm{u}} \tag{3.15b}$$

$$\frac{e_{ik}^{\mathrm{m}}}{\sum_{j\in\mathcal{K}_{\mathrm{u}}} b_j^2|\tilde{\mathbf{g}}_j^H\mathbf{h}_{ik}|^2 + I_{\mathrm{m},ik}^{\mathrm{m}} + \sigma^2} \geq \gamma_{ik}^{\mathrm{m}},\ k\in\mathcal{K}_{\mathrm{m}}, i\in\mathcal{G} \tag{3.15c}$$

where $e_{ik}^{\mathrm{m}} \triangleq |\mathbf{a}_i^H\widetilde{\mathbf{H}}_i^H\mathbf{h}_{ik}|^2$ is the received signal power at multicast user k in group i, $I_{\mathrm{u},j}^{\mathrm{m}} \triangleq \sum_{i\in\mathcal{G}} |\mathbf{a}_i^H\widetilde{\mathbf{H}}_i^H\mathbf{g}_j|^2$ represents the interference from multicast users to the desired unicast user, and $I_{\mathrm{m},ik}^{\mathrm{m}} \triangleq \sum_{l\in\mathcal{G}^{-i}} |\mathbf{a}_l^H\widetilde{\mathbf{H}}_l^H\mathbf{h}_{ik}|^2$ denotes the interference form other multicast users to the desired multicast user.

We note that problem $\mathcal{P}_{\mathrm{u}}$ is similar to the power minimization problem for downlink multi-user unicast beamforming, under the optimized beamforming vectors [5]. The difference of $\mathcal{P}_{\mathrm{u}}$ from the classical unicast beamforming problem is that $\mathcal{P}_{\mathrm{u}}$ has an extra constraint in (3.15c), which is convex w.r.t. $\mathbf{b}$. We express the corresponding classical power minimization problem without the constraint in (3.15c) as

$$\mathcal{P}'_{\mathrm{u}}: \quad \min_{\mathbf{b}} \sum_{j\in\mathcal{K}_{\mathrm{u}}} b_j^2 \|\tilde{\mathbf{g}}_j\|^2, \quad \text{s.t.} \quad (3.15\text{b}).$$

Below, we show that compared to $\mathcal{P}'_{\mathrm{u}}$, the constraint in (3.15c) serves only as a feasibility check for $\mathcal{P}_{\mathrm{u}}$.

Proposition 1. *Assume $\mathcal{P}_u$ is feasible. Then $\mathcal{P}'_u$ and $\mathcal{P}_u$ are equivalent, i.e., the optimal solution $\mathbf{b}^\star$ to $\mathcal{P}'_u$ is also optimal to $\mathcal{P}_u$, and vise versa.*

Proof. First, we show that at the optimum of $\mathcal{P}_{\mathrm{u}}$, the SINR constraints in (3.15b) must be satisfied with equality. We prove this by contradiction. Assume that $\mathbf{b}^\star$ is the optimal solution to $\mathcal{P}_{\mathrm{u}}$, and at the optimum, not all SINR constraints in (3.15b) are attained with equality. This means that there exists some $j^o \in \mathcal{K}_{\mathrm{u}}$, such that

$$\frac{b_{j^o}^{\star 2} |\tilde{\mathbf{g}}_{j^o}^H \mathbf{g}_{j^o}|^2}{\sum_{j'\in\mathcal{K}_{\mathrm{u}}^{-j}} b_{j'}^{\star 2} |\tilde{\mathbf{g}}_{j'}^H \mathbf{g}_j|^2 + I_{\mathrm{u},j^o}^{\mathrm{m}} + \sigma^2} > \gamma_{j^o}^{\mathrm{u}}. \tag{3.16}$$

We can find $0 < \alpha < 1$, such that replacing $b_{j^o}^\star$ in (3.16) by $\tilde{b}_{j^o} = \alpha b_{j^o}^\star$, the SINR constraint is satisfied with equality as

$$\frac{\tilde{b}_{j^o}^{2} |\tilde{\mathbf{g}}_{j^o}^H \mathbf{g}_{j^o}|^2}{\sum_{j'\in\mathcal{K}_{\mathrm{u}}^{-j}} \tilde{b}_{j'}^{2} |\tilde{\mathbf{g}}_{j'}^H \mathbf{g}_j|^2 + I_{\mathrm{u},j^o}^{\mathrm{m}} + \sigma^2} = \gamma_{j^o}^{\mathrm{u}}. \tag{3.17}$$

Since $\mathbf{b}^\star$ satisfies the constraint in (3.15c), and $\tilde{b}_{j^o}^2 < b_{j^o}^{\star 2}$, the constraint in (3.15c) still hold for $\tilde{b}_{j^o}$. However, the objective value in $\mathcal{P}_{\mathrm{u}}$ is reduced with $b_{j^o}^\star$ replaced by $\tilde{b}_{j^o}$, contradicting our assumption that $\mathbf{b}^\star$ is the optimal solution for $\mathcal{P}_{\mathrm{u}}$. Thus, at the optimum of $\mathcal{P}_{\mathrm{u}}$, the constraints in (3.15b) are satisfied with equality. Solving these K_{u} equations in (3.15b) for K_{u} variables $\{b_j^2\}$, we have the unique solution $\{b_j^{\star 2}\}$. If $\mathcal{P}_{\mathrm{u}}$ is feasible, then $\mathbf{b}^\star$ satisfies the constraint in (3.15c) and is optimal to $\mathcal{P}_{\mathrm{u}}$ is unique (up to a phase rotation).

Assume $\mathcal{P}'_{\mathrm{u}}$ is feasible. Recall the power minimization problem for downlink multi-user unicast beamforming. Under the optimal beamforming vector, which is similar to (3.10), the resulting power minimization problem is in the same form as $\mathcal{P}'_{\mathrm{u}}$ [5]. It is known that at the optimum, the SINR constraints in (3.15b) are satisfied with equality, *i.e.,* the optimal $\mathbf{b}'^{\star}$ to $\mathcal{P}'_{\mathrm{u}}$ is unique solution to (3.15b) with equality (up to a phase rotation). Since the optimal solution to $\mathcal{P}_{\mathrm{u}}$ is also the unique solution to (3.15b) with equality, we conclude that if $\mathbf{b}'^{\star}$ satisfies (3.15c), then it is optimal to $\mathcal{P}_{\mathrm{u}}$. Otherwise, no feasible solution exists, and $\mathcal{P}_{\mathrm{u}}$ is infeasible.

Conversely, assume $\mathcal{P}_{\mathrm{u}}$ being feasible, and $\mathbf{b}^{\star}$ is optimal to $\mathcal{P}_{\mathrm{u}}$. From the above, $\mathbf{b}^{\star}$ satisfies the constraint in (3.15b) with equality and is unique. It immediately follows that $\mathbf{b}^{\star}$ is optimal to $\mathcal{P}'_{\mathrm{u}}$. $\square$

From Proposition 1, instead of solving $\mathcal{P}_{\mathrm{u}}$, we can first solve $\mathcal{P}'_{\mathrm{u}}$. As discussed earlier and also indicated in the proof of Proposition 1, $\mathcal{P}'_{\mathrm{u}}$ is the same as the power minimization problem for downlink multi-user unicast beamforming. At the optimum, the constraint in (3.15b) is satisfied with equality [5], which is given by

$$\frac{b_j^2}{\gamma_j^u}|\tilde{\mathbf{g}}_j^H\mathbf{g}_j|^2 = \sum_{j'\in\mathcal{K}_{\mathrm{u}}^{-j}} b_{j'}^2|\tilde{\mathbf{g}}_{j'}^H\mathbf{g}_j|^2 + I_{\mathrm{u},j}^{\mathrm{m}} + \sigma^2, \quad j\in\mathcal{K}_{\mathrm{u}}. \tag{3.18}$$

Therefore, $\{b_j^2\}$ can be obtained by solving the above K_{u} linear equations. Define a $K_{\mathrm{u}}\times K_{\mathrm{u}}$ matrix $\mathbf{V}$, where

$$[\mathbf{V}]_{jj'} = \begin{cases} \frac{1}{\gamma_j^{\mathrm{u}}}|\tilde{\mathbf{g}}_j^H\mathbf{g}_j|^2, & j'=j \\ -|\tilde{\mathbf{g}}_j^H\mathbf{g}_{j'}|^2, & j'\neq j. \end{cases} \tag{3.19}$$

Define $\tilde{\mathbf{b}} = [b_1^2, ..., b_{K_{\mathrm{u}}}^2]^T$ and $\boldsymbol{\zeta} = [I_{\mathrm{u},1}^{\mathrm{m}}+\sigma^2, ..., I_{\mathrm{u},K_{\mathrm{u}}}^{\mathrm{m}}+\sigma^2]^T$. Then, the solution to the

K_{u} linear equations in (3.18) can be compactly written as

$$\tilde{\mathbf{b}} = \mathbf{V}^{-1}\boldsymbol{\zeta}. \tag{3.20}$$

Finally, without loss of optimality, we obtain $b_j = \sqrt{[\tilde{\mathbf{b}}]_j}$, $j \in K_{\mathrm{u}}$. Thus, for $\mathcal{P}_{\mathrm{u}}$ being feasible, we obtain the weight vector solution $\mathbf{b}$ for $\mathcal{P}_{\mathrm{u}}$ in closed-form. Note that this is a key step that provides a substantial computational saving, especially for the scenarios where unicast users dominate the user population.

3.3.2 Obtaining Weight Vectors for Multicast Users

After obtaining $\mathbf{b}$, we now solve weight vectors $\{\mathbf{a}_i\}$ for the multicast users. With fixed $\mathbf{b}$, $\mathcal{P}'_{\mathrm{o}}$ is equivalent to the following optimization problem w.r.t $\{\mathbf{a}_i\}$:

$$\mathcal{P}_{\mathrm{m}} : \min_{\{\mathbf{a}_i\}} \sum_{i\in\mathcal{G}} \|\widetilde{\mathbf{H}}_i \mathbf{a}_i\|^2 \tag{3.21a}$$

$$\text{s.t.} \quad \frac{e_j^{\mathrm{u}}}{\sum_{i\in\mathcal{G}} |\mathbf{a}_i^H \widetilde{\mathbf{H}}_i^H \mathbf{g}_j|^2 + I_{\mathrm{u},j}^{\mathrm{u}} + \sigma^2} \geq \gamma_j^{\mathrm{u}}, \; j \in \mathcal{K}_{\mathrm{u}} \tag{3.21b}$$

$$\frac{|\mathbf{a}_i^H \widetilde{\mathbf{H}}_i^H \mathbf{h}_{ik}|^2}{\sum_{l\in\mathcal{G}^{-i}} |\mathbf{a}_l^H \widetilde{\mathbf{H}}_l^H \mathbf{h}_{ik}|^2 + I_{\mathrm{m},ik}^{\mathrm{u}} + \sigma^2} \geq \gamma_{ik}^{\mathrm{m}}, \; k \in \mathcal{K}_{\mathrm{m}}, i \in \mathcal{G} \tag{3.21c}$$

where $e_j^{\mathrm{u}} \triangleq b_j^2 |\tilde{\mathbf{g}}_j^H \mathbf{g}_j|^2$, and $I_{\mathrm{u},j}^{\mathrm{u}} \triangleq \sum_{j'\in\mathcal{K}_{\mathrm{u}}^{-j}} b_{j'}^2 |\tilde{\mathbf{g}}_{j'}^H \mathbf{g}_j|^2$ and $I_{\mathrm{m},ik}^{\mathrm{u}} \triangleq \sum_{j\in\mathcal{K}_{\mathrm{u}}} b_j^2 |\tilde{\mathbf{g}}_j^H \mathbf{h}_{ik}|^2$ represent the interference from other unicast users to the desired unicast user, and from all unicast users to the desired multicast user, respectively.

Note that $\mathcal{P}_{\mathrm{m}}$ is similar to the conventional multi-group multicast QoS problem (3.8). The difference is that $\mathcal{P}_{\mathrm{m}}$ has the additional SINR constraint in (3.21b) for the unicast users, which is convex quadratic w.r.t. $\{\mathbf{a}_i\}$. We adopt the similar SCA method in [27] to convexify $\mathcal{P}_{\mathrm{m}}$ and solve it iteratively. To improve the computational efficiency, we develop an ADMM-based fast algorithm to solve the subproblem in each

SCA iteration.

The SCA Method

SCA is an iterative algorithm that solves a non-convex problem using a sequence of convex approximation. For a non-convex problem with a convex objective function, SCA is proven to converge to a stationary point [88]. The SCA method for $\mathcal{P}_{\mathrm{m}}$ is summarized below.

Introduce a set of $G \times 1$ auxiliary vectors $\mathbf{v}_i$, $i \in \mathcal{G}$. For matrix $\mathbf{A} \succeq 0$, we have $(\mathbf{a}_i - \mathbf{v}_i)^H \mathbf{A}(\mathbf{a}_i - \mathbf{v}_i) \succeq 0$, or equivalently, $\mathbf{a}_i^H \mathbf{A} \mathbf{a}_i \geq 2\mathfrak{Re}\{\mathbf{a}_i^H \mathbf{A} \mathbf{v}_i\} - \mathbf{v}_i^H \mathbf{A} \mathbf{v}_i$, for any $\mathbf{v}_i$, $i \in \mathcal{G}$. We convexify the constraint in (3.21c) in $\mathcal{P}_{\mathrm{m}}$ by applying the above inequality to the numerator of the SINR expression in (3.21c) and arrive at the following problem:

$$
\begin{aligned}
&\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v}) : \min_{\{\mathbf{a}_i\}} \sum_{i \in \mathcal{G}} \|\widetilde{\mathbf{H}}_i \mathbf{a}_i\|^2 && (3.22\text{a})\\
&\text{s.t.} \sum_{i \in \mathcal{G}} |\mathbf{a}_i^H \mathbf{r}_{ij}|^2 \leq \frac{e_j^{\mathrm{u}}}{\gamma_j^{\mathrm{u}}} - I_{\mathrm{u},j}^{\mathrm{u}} - \sigma^2, \quad j \in \mathcal{K}_{\mathrm{u}} && (3.22\text{b})\\
&\quad \gamma_{ik}^{\mathrm{m}} \sum_{l \in \mathcal{G}_{-i}} |\mathbf{a}_l^H \mathbf{c}_{lik}|^2 - 2\mathfrak{Re}\{\mathbf{a}_i^H \mathbf{c}_{iik} \mathbf{c}_{iik}^H \mathbf{v}_i\} + |\mathbf{v}_i^H \mathbf{c}_{iik}|^2 \\
&\qquad\qquad + \gamma_{ik}^{\mathrm{m}} (I_{\mathrm{m},ik}^{\mathrm{u}} + \sigma^2) \leq 0, \quad k \in \mathcal{K}_{\mathrm{m}}, i \in \mathcal{G}, && (3.22\text{c})
\end{aligned}
$$

where $\mathbf{r}_{ij} \triangleq \widetilde{\mathbf{H}}_i^H \mathbf{g}_j$, $j \in \mathcal{K}_{\mathrm{u}}$, $i \in \mathcal{G}$, $\mathbf{c}_{lik} \triangleq \widetilde{\mathbf{H}}_l^H \mathbf{h}_{ik}$, $k \in \mathcal{K}_{\mathrm{m}}$, $i, l \in \mathcal{G}$, and $\mathbf{v} \triangleq [\mathbf{v}_1^H, \ldots, \mathbf{v}_G^H]^H$.

Since $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ is convex, it can be solved by the interior-point method [89] via a standard convex solver, such as CVX. Once the optimal solution $\mathbf{a}_i^\star(\mathbf{v})$ to $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ is obtained, we update each $\mathbf{v}_i$ by replacing it with $\mathbf{a}_i^\star(\mathbf{v})$, $i \in \mathcal{G}$. Then, $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ is iteratively solved until convergence. This iteratively procedure by SCA is proven to converge to a stationary point [88].

ADMM-Based Fast Algorithm for $\mathcal{P}_{\mathbf{mSCA}}(\mathbf{v})$

Although the interior-point method can be directly applied to solve $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$, it is a second-order algorithm, which has a high computational complexity especially when the problem size becomes large. As we expect the future systems to be in large-scale, we need to develop a more computationally inexpensive algorithm to find the solution. In this work, we propose an ADMM-based low complexity fast algorithm to solve $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$.

We introduce a set of auxiliary variables $d_{lik} \triangleq \mathbf{a}_l^H \mathbf{c}_{lik}$, for $l, i \in \mathcal{G}, k \in \mathcal{K}_{\mathrm{m}}$, and a set of $K_{\mathrm{m}} \times 1$ auxiliary vectors $\boldsymbol{\mu}_i \triangleq \mathbf{a}_i$, for $i \in \mathcal{G}$. Define $\mathbf{a} \triangleq [\mathbf{a}_1^H, ..., \mathbf{a}_G^H]^H$, $\boldsymbol{\mu} \triangleq [\boldsymbol{\mu}_1^H, ..., \boldsymbol{\mu}_G^H]^H$, and $\mathbf{d} = [\mathbf{d}_{11}^H, ..., \mathbf{d}_{GK_{\mathrm{m}}}^H]^H$ with $\mathbf{d}_{ik} = [d_{1ik}, ..., d_{Gik}]^T$. Then, $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ can be equivalently expressed as

$$
\begin{aligned}
\mathcal{P}_{\mathrm{mADMM}}: \ & \min_{\mathbf{a},\boldsymbol{\mu},\mathbf{d}} \sum_{i\in\mathcal{G}} \|\widetilde{\mathbf{H}}_i \mathbf{a}_i\|^2 && \text{(3.23a)} \\
\text{s.t.} \ \ & \boldsymbol{\mu}_i = \mathbf{a}_i \ , i \in \mathcal{G} && \text{(3.23b)} \\
& \sum_{i\in\mathcal{G}} |\boldsymbol{\mu}_i^H \mathbf{r}_{ij}|^2 \le \Big(\frac{e_j^{\mathrm{u}}}{\gamma_j^u} - I_{\mathrm{u},j}^{\mathrm{u}} - \sigma^2\Big), \ j \in \mathcal{K}_{\mathrm{u}} && \text{(3.23c)} \\
& d_{lik} = \mathbf{a}_l^H \mathbf{c}_{lik}, \quad l, i \in \mathcal{G}, \ k \in \mathcal{K}_{\mathrm{m}} && \text{(3.23d)} \\
& \gamma_{ik}^{\mathrm{m}} \sum_{l\in\mathcal{G}-i} |d_{lik}|^2 - 2\mathfrak{Re}\{d_{iik}\mathbf{c}_{iik}^H \mathbf{v}_i\} + |\mathbf{v}_i^H \mathbf{c}_{iik}|^2 \\
& \qquad\qquad + \gamma_{ik}^{\mathrm{m}} (I_{\mathrm{m},ik}^{\mathrm{u}} + \sigma^2) \le 0, \ k \in \mathcal{K}_{\mathrm{m}}, i \in \mathcal{G}. && \text{(3.23e)}
\end{aligned}
$$

Denote $\mathcal{D}$ and $\mathcal{F}$ as the feasible sets for the constraints in (3.23c) and (3.23e), respectively. Accordingly, define the indicator functions

$$
\mathbb{I}_{\mathcal{D}}(\boldsymbol{\mu}) = \begin{cases} 0 & \text{if } \ \boldsymbol{\mu} \in \mathcal{D} \\ \infty & \text{otherwise} \end{cases}, \quad \mathbb{I}_{\mathcal{F}}(\mathbf{d}) = \begin{cases} 0 & \text{if } \mathbf{d} \in \mathcal{F} \\ \infty & \text{otherwise} \end{cases}.
$$

Then, $\mathcal{P}_{\text{mADMM}}$ can be equivalently expressed as follows:

$$\mathcal{P}'_{\text{mADMM}} : \min_{\mathbf{a},\boldsymbol{\mu},\mathbf{d}} \sum_{i\in\mathcal{G}} \|\widetilde{\mathbf{H}}_i \mathbf{a}_i\|^2 + \mathbb{I}_{\mathcal{F}}(\mathbf{d}) + \mathbb{I}_{\mathcal{D}}(\boldsymbol{\mu})$$
$$\text{s.t. } (3.23\text{b}) \text{ and } (3.23\text{d}).$$

As we will see, the introduction of the sets of auxiliary variables and new equality constraints is useful in separating the original problem into subproblems with separate variables to solve iteratively. The augmented Lagrangian [74] of $\mathcal{P}'_{\text{mADMM}}$ is given by

$$\begin{aligned}\mathcal{L}_p(\mathbf{a},\boldsymbol{\mu},\mathbf{d},\mathbf{q},\mathbf{z}) = &\sum_{i\in\mathcal{G}} \|\widetilde{\mathbf{H}}_i \mathbf{a}_i\|^2 + \mathbb{I}_{\mathcal{F}}(\mathbf{d}) + \mathbb{I}_{\mathcal{D}}(\boldsymbol{\mu}) \\ &+ \frac{\rho}{2}\sum_{l\in\mathcal{G}}\sum_{i\in\mathcal{G}}\sum_{k\in\mathcal{K}_{\text{m}}} \left| d_{lik} - \mathbf{a}_l^H \mathbf{c}_{lik} + q_{lik} \right|^2 + \frac{\rho}{2}\sum_{i\in\mathcal{G}} \|\boldsymbol{\mu}_i - \mathbf{a}_i + \mathbf{z}_i\|^2,\end{aligned} \tag{3.24}$$

where $\rho > 0$ is the penalty parameter, and $\mathbf{q} \triangleq [\mathbf{q}_{11}^H, ..., \mathbf{q}_{GK_{\text{m}}}^H]^H$ with $\mathbf{q}_{ik} \triangleq [q_{1ik}, ..., q_{Gik}]^T$ and $\mathbf{z} \triangleq [\mathbf{z}_1^H, ..., \mathbf{z}_G^H]^H$ are the dual variables associated with the constraints in (3.23b) and (3.23d), respectively.

Note from (3.24) that the minimization of $\mathcal{L}_p(\mathbf{a},\boldsymbol{\mu},\mathbf{d},\mathbf{q},\mathbf{z})$ can be broken into two subproblems, one for $\{\mathbf{d},\boldsymbol{\mu}\}$ and the other for $\mathbf{a}$, to be solved alternatively. Thus, our proposed ADMM-based procedure has the following three updating steps at iteration $(n+1)$:

$$1)\ \{\mathbf{d}^{(n+1)}, \boldsymbol{\mu}^{(n+1)}\} = \arg\min_{\mathbf{d},\boldsymbol{\mu}} \mathcal{L}_p(\mathbf{a}^{(n)}, \boldsymbol{\mu}, \mathbf{d}, \mathbf{q}^{(n)}, \mathbf{z}^{(n)}) \tag{3.25}$$

$$2)\ \mathbf{a}^{(n+1)} = \arg\min_{\mathbf{a}} \mathcal{L}_p(\mathbf{a}, \boldsymbol{\mu}^{(n+1)}, \mathbf{d}^{(n+1)}, \mathbf{q}^{(n)}, \mathbf{z}^{(n)}) \tag{3.26}$$

$$\begin{aligned}3)\ &q_{lik}^{(n+1)} = q_{lik}^{(n)} + (d_{lik}^{(n+1)} - \mathbf{a}_l^{(n+1)H} \mathbf{c}_{lik}),\ l, i \in \mathcal{G}, k \in \mathcal{K}_{\text{m}} \\ &\mathbf{z}_i^{(n+1)} = \mathbf{z}_i^{(n)} + (\boldsymbol{\mu}_i^{(n+1)} - \mathbf{a}_i^{(n+1)}),\ i \in \mathcal{G}.\end{aligned} \tag{3.27}$$

The main advantage of our proposed ADMM-based algorithm is that we are able to obtain the closed-form solution for each variable update in (3.25)–(3.27), as we will show below.

Given $\mathbf{a}^{(n)}$, $\mathbf{q}^{(n)}$, and $\mathbf{z}^{(n)}$, we see that $\mathcal{L}_p(\mathbf{a}, \boldsymbol{\mu}, \mathbf{d}, \mathbf{q}, \mathbf{z})$ in (3.24) can be further grouped into two parts, one for $\mathbf{d}$ only and the other for $\boldsymbol{\mu}$ only. Thus, the optimization of $\mathbf{d}$ and $\boldsymbol{\mu}$ in (3.25) can be further split in two separate subproblems:

$$\mathbf{d}^{(n+1)} = \arg\min_{\mathbf{d}} \frac{\rho}{2} \sum_{l\in\mathcal{G}} \sum_{i\in\mathcal{G}} \sum_{k\in\mathcal{K}_{\mathrm{m}}} |d_{lik} - \mathbf{a}_l^{(n)H}\mathbf{c}_{lik} + q_{lik}^{(n)}|^2 + \mathbb{I}_{\mathcal{F}}(\mathbf{d}), \tag{3.28}$$

$$\boldsymbol{\mu}^{(n+1)} = \arg\min_{\boldsymbol{\mu}} \sum_{i\in\mathcal{G}} \|\boldsymbol{\mu}_i - \mathbf{a}_i^{(n)} + \mathbf{z}_i^{(n)}\|^2 + \mathbb{I}_{\mathcal{D}}(\boldsymbol{\mu}). \tag{3.29}$$

We now derive the closed-form solutions for the subproblems in (3.25) and (3.26).

i) **Updating $\mathbf{d}^{(n+1)}$**:

The minimization problem in (3.28) is equivalent to the following problem:

$$\begin{aligned} &\min_{\mathbf{d}} \sum_{l\in\mathcal{G}} \sum_{i\in\mathcal{G}} \sum_{k\in\mathcal{K}_{\mathrm{m}}} |d_{lik} - \mathbf{a}_l^{(n)H}\mathbf{c}_{lik} + q_{lik}^{(n)}|^2 \\ &\text{s.t} \quad (3.23\text{e}). \end{aligned} \tag{3.30}$$

The above problem can be decomposed to GK_{m} separate subproblems to solve, one for each $\mathbf{d}_{ik}$, $k \in \mathcal{K}_{\mathrm{m}}$, $i \in \mathcal{G}$, given by

$$\begin{aligned} &\min_{\mathbf{d}_{ik}} \sum_{l\in\mathcal{G}} \left|d_{lik} - \psi_{1,lik}^{(n)}\right|^2 \\ &\text{s.t} \; \psi_{2,ik+}\gamma_{ik}^{\mathrm{m}} \sum_{l\in\mathcal{G}^{-i}} |d_{lik}|^2 - 2\mathfrak{Re}\{d_{iik}\psi_{3,ik}\} \le 0, k \in \mathcal{K}_{\mathrm{m}} \end{aligned} \tag{3.31}$$

where $\psi_{1,lik}^{(n)} \triangleq \mathbf{a}_l^{(n)H}\mathbf{c}_{lik} - q_{lik,}^{(n)}$, $\psi_{2,ik} \triangleq \gamma_{ik}^{\mathrm{m}}(I_{\mathrm{m},ik}^{\mathrm{u}} + \sigma^2) + |\mathbf{v}_i^H\mathbf{c}_{iik}|^2$, and $\psi_{3,ik} \triangleq \mathbf{c}_{iik}^H\mathbf{v}_i$. Note that $\psi_{2,ik}$ is real, while both $\psi_{1,lik}$ and $\psi_{3,ik}$ are complex.

Note that problem (3.31) is a convex quadratically constrained quadratic program (QCQP) problem w.r.t. $\mathbf{d}_{ik}$, for which a closed-form solution may be derived using

Karush-Kuhn-Tucker (KKT) conditions [89]. In fact, a problem of a similar structure as in (3.31) has been considered in [23], for which the optimal solution has been derived in closed-form. We directly apply the result from [23, Appendix A] to our problem and present the solution below. Let $\alpha_{ik} \geq 0$ denote the Lagrange multiplier associated with the constraint in (3.31). The optimal solution $\mathbf{d}^{\star}_{ik}$ to (3.31) is given by [23]

$$d^{\star}_{lik} = \begin{cases} \psi^{(n)}_{1,iik} + \alpha^{\star}_{ik}\psi^{*}_{3,ik}, & \text{for } l = i \\ \dfrac{\psi^{(n)}_{1,lik}}{1 + \alpha^{\star}_{ik}\gamma^{\mathrm{m}}_{ik}} & \text{for } l \neq i \end{cases} \tag{3.32}$$

where the optimal $\alpha^{\star}_{ik} = 0$, if $\psi_{2,ik} + \gamma^{\mathrm{m}}_{ik}\sum_{l\neq i}|\psi^{(n)}_{1,lik}|^2 - 2\mathfrak{Re}\{\psi^{(n)}_{1,iik}\psi_{3,ik}\} \leq 0$; otherwise, α_{ik} is the unique real positive root of the following cubic equation:

$$\begin{aligned} &-2\gamma^{\mathrm{m}2}_{ik}|\psi_{3,ik}|^2\alpha^{\star\,3}_{ik} + (\gamma^{\mathrm{m}2}_{ik}\psi_{2,ik} - 2\gamma^{\mathrm{m}2}_{ik}\mathfrak{Re}\{\psi^{(n)}_{1,iik}\psi_{3,ik}\} - 4\gamma^{\mathrm{m}}_{ik}|\psi_{3,ik}|^2)\alpha^{\star\,2}_{ik} \\ &+\ 2(\gamma^{\mathrm{m}}_{ik}\psi_{2,ik} - 2\gamma^{\mathrm{m}}_{ik}\mathfrak{Re}\{\psi^{(n)}_{1,iik}\psi_{3,ik}\} - |\psi_{3,ik}|^2)\alpha^{\star}_{ik} \\ &+ \psi_{2,ik} - 2\mathfrak{Re}\{\psi^{(n)}_{1,iik}\psi_{3,ik}\} + \gamma^{\mathrm{m}}_{ik}\sum_{l\neq i}|\psi^{(n)}_{1,lik}|^2 = 0. \end{aligned} \tag{3.33}$$

The roots to the above equation are given by the cubic formula. Furthermore, the real positive root is guaranteed to be unique [23]. The update $\mathbf{d}^{(n+1)}$ is then obtained using the closed-form solution for $\mathbf{d}^{\star}_{ik}$ in (3.32), for each $k \in \mathcal{K}_{\mathrm{m}}, i \in \mathcal{G}$.

ii) **Updating $\boldsymbol{\mu}^{(n+1)}$**:

The update of $\boldsymbol{\mu}$ in (3.29) is equivalent to solving the following problem:

$$\begin{aligned} \mathcal{P}_{\boldsymbol{\mu}}:\ \min_{\boldsymbol{\mu}}\ & \sum_{i\in\mathcal{G}}\|\boldsymbol{\mu}_i - \tilde{\mathbf{a}}^{(n)}_i\|^2 \\ \text{s.t.}\ & \sum_{i\in\mathcal{G}}|\boldsymbol{\mu}^H_i\mathbf{r}_{ij}|^2 \leq\ \tilde{e}^{\mathrm{u}}_j,\ j\in\mathcal{K}_{\mathrm{u}} \end{aligned} \tag{3.34}$$

where $\tilde{\mathbf{a}}_i^{(n)} \triangleq \mathbf{a}_i^{(n)} - \mathbf{z}_i^{(n)}$, $i \in \mathcal{K}_\mathrm{m}$, and $\tilde{e}_j^\mathrm{u} \triangleq \frac{e_j^\mathrm{u}}{\gamma_j^\mathrm{u}} - I_{\mathrm{u},j}^\mathrm{u} - \sigma^2$, $j \in \mathcal{K}_\mathrm{u}$. Note that $\mathcal{P}_{\boldsymbol{\mu}}$ is a convex QCQP problem with K_u constraints. We discuss the solution in two cases:

Case 1: Let $\boldsymbol{\mu}_i = \tilde{\mathbf{a}}_i^{(n)}$, $i \in \mathcal{G}$. If the constraint in (3.34) is satisfied, then it is the optimal solution to $\mathcal{P}_{\boldsymbol{\mu}}$, and we have $\boldsymbol{\mu}_i^{(n+1)} = \tilde{\mathbf{a}}_i^{(n)}$, $i \in \mathcal{G}$. Otherwise, consider Case 2 below.

Case 2: In this case, since $\mathcal{P}_{\boldsymbol{\mu}}$ is a convex QCQP problem, it has zero-duality gap. Thus, we obtain the optimal solution in its Lagrange dual domain. The Lagrangian of $\mathcal{P}_{\boldsymbol{\mu}}$ is given by

$$\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu}) = \sum_{i \in \mathcal{G}} \|\boldsymbol{\mu}_i - \tilde{\mathbf{a}}_i^{(n)}\|^2 + \sum_{j \in \mathcal{K}_\mathrm{u}} \delta_j \Big(\sum_{i \in \mathcal{G}} |\boldsymbol{\mu}_i^H \mathbf{r}_{ij}|^2 - \tilde{e}_j^\mathrm{u}\Big)$$

where $\delta_j \geq 0$ is the Lagrange multiplier associated with the constraint in (3.34), for $j \in \mathcal{K}_\mathrm{u}$, and $\boldsymbol{\delta} \triangleq [\delta_1, ..., \delta_{K_\mathrm{u}}]^T$. Denote $\tilde{\mathbf{e}} \triangleq [\tilde{e}_1, ..., \tilde{e}_{K_\mathrm{u}}]^T$. Regrouping the terms in $\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu})$, we have

$$\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu}) = \sum_{i \in \mathcal{G}} \boldsymbol{\mu}_i^H \Big(\mathbf{I} + \sum_{j \in \mathcal{K}_\mathrm{u}} \delta_j \mathbf{r}_{ij} \mathbf{r}_{ij}^H\Big) \boldsymbol{\mu}_i - \boldsymbol{\delta}^T \tilde{\mathbf{e}} - 2 \sum_{i \in \mathcal{G}} \mathfrak{Re}\{\tilde{\mathbf{a}}_i^{(n)H} \boldsymbol{\mu}_i\} + \sum_{i \in \mathcal{G}} \|\tilde{\mathbf{a}}_i^{(n)}\|^2. \tag{3.35}$$

The Lagrange dual problem for $\mathcal{P}_{\boldsymbol{\mu}}$ is given by

$$\max_{\boldsymbol{\delta}} \Big[g(\boldsymbol{\delta}) \triangleq \min_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu}) \Big] \quad \text{s.t.} \quad \boldsymbol{\delta} \succeq 0. \tag{3.36}$$

For the inner minimization in (3.36), setting the gradient of $\mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu})$ w.r.t $\boldsymbol{\mu}_i^*$ to zero, we have

$$\nabla_{\boldsymbol{\mu}_i^*} \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu}) = \Big(\mathbf{I} + \sum_{j \in \mathcal{K}_\mathrm{u}} \delta_j \mathbf{r}_{ij} \mathbf{r}_{ij}^H\Big) \boldsymbol{\mu}_i - \tilde{\mathbf{a}}_i^{(n)} = 0, \tag{3.37}$$

which yields the optimal solution $\boldsymbol{\mu}_i^\mathrm{o}(\boldsymbol{\delta})$ as

$$\boldsymbol{\mu}_i^\mathrm{o}(\boldsymbol{\delta}) = \Big(\mathbf{I} + \sum_{j \in \mathcal{K}_\mathrm{u}} \delta_j \mathbf{r}_{ij} \mathbf{r}_{ij}^H\Big)^{-1} \tilde{\mathbf{a}}_i^{(n)}, \ i \in \mathcal{G}. \tag{3.38}$$

Substituting the above expression into (3.35), we have $g(\boldsymbol{\delta}) = \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu}^{\text{o}}(\boldsymbol{\delta}))$.

To solve the dual problem (3.36) for the optimal $\boldsymbol{\delta}^o$, we apply the projected subgradient method [90]. Define $(z)^+ \triangleq \max\{z, 0\}$ for $z \in \mathbb{R}$. The projected subgradient method at iteration t has the following update:

$$\delta_j^{t+1} = \left(\delta_j^t - \tau_t \phi_j^t\right)^+, \quad j \in \mathcal{K}_{\text{u}}, \tag{3.39}$$

where τ_t is the step size at iteration t, and ϕ_j^t is a subgradient of $-g(\boldsymbol{\delta})$ at $\boldsymbol{\delta}^t$, *i.e.*, $\phi_j^t \in \partial(-g(\boldsymbol{\delta}^t))$.[3] From (3.35), we consider the following subgradient

$$\phi_j^t = \frac{\partial - \mathcal{L}(\boldsymbol{\delta}, \boldsymbol{\mu}^t)}{\partial \delta_j}\bigg|_{\boldsymbol{\delta}=\boldsymbol{\delta}^t} = -\sum_{i \in \mathcal{G}} \left|\boldsymbol{\mu}_i^{tH} \mathbf{r}_{ij}\right|^2 + \tilde{e}_j^{\text{u}} \tag{3.40}$$

where $\boldsymbol{\mu}_i^t = \boldsymbol{\mu}^{\text{o}}(\boldsymbol{\delta}^t)$. Substituting (3.40) into (3.39) , we have the projected subgradient update for problem (3.36):

$$\delta_j^{t+1} = \left(\delta_j^t + \tau_t\left(\sum_{i \in \mathcal{G}} \left|\mathbf{r}_{ij}^H \boldsymbol{\mu}_i^{\text{o}}(\boldsymbol{\delta}^t)\right|^2 - \tilde{e}_j^{\text{u}}\right)\right)^+, \; j \in \mathcal{K}_{\text{u}}. \tag{3.41}$$

It has been shown in [90] that the projected subgradient method is guaranteed to converge to the optimal solution for non-summable diminishing step sizes $\{\tau_t\}$. For the constant step size, the method converges to the neighborhood of the optimal solution. Since $g(\boldsymbol{\delta})$ is differentiable, with the constant step size, the method converges to the optimal, provided $\{\phi_j^t\}$ is small enough [90].

Once we obtain the optimal $\boldsymbol{\delta}^o$ to problem (3.36), we have $\boldsymbol{\mu}_i^{\text{o}}(\boldsymbol{\delta}^o)$, $i \in \mathcal{G}$, as the optimal solution to $\mathcal{P}_{\boldsymbol{\mu}}$, and the update $\boldsymbol{\mu}_i^{(n+1)} = \boldsymbol{\mu}_i^{\text{o}}(\boldsymbol{\delta}^o)$, $i \in \mathcal{G}$.

Remark. We point out that as the proposed algorithm converges over iterations, between Cases 1 and 2, we expect that Case 1 becomes the most likely solution for

[3] The set of subgradients of $f(\mathbf{x})$ is denoted by $\partial f(\mathbf{x})$

$\mathcal{P}_{\boldsymbol{\mu}}$. To see this, we note that when the algorithm converges over iterations, we expect $\boldsymbol{\mu}_i^{(n)} \to \mathbf{a}_i^{(n)}$ as in (3.23b), and $\mathbf{z}_i^{(n)} \to \mathbf{0}$ as in (3.27). Thus, initially, Case 2 may be the solution for $\mathcal{P}_{\boldsymbol{\mu}}$. However, as n increases, $\boldsymbol{\mu}_i^{(n)}$ and $\tilde{\mathbf{a}}_i^{(n)}$ from previous iteration n become close. As a result, Case 1 becomes the typical solution. Thus, the required computational complexity becomes further reduced at the later iterations to be a direct closed-form solution for updating $\boldsymbol{\mu}^{(n+1)}$, avoiding the projected subgradient method.

iii) **Updating $\mathbf{a}^{(n+1)}$**:

From (3.24), the update of $\mathbf{a}$ in (3.26) can be obtained by solving the following problem:

$$\min_{\mathbf{a}} \sum_{i\in\mathcal{G}} \|\widetilde{\mathbf{H}}_i\mathbf{a}_i\|^2 + \frac{\rho}{2}\sum_{l\in\mathcal{G}}\sum_{i\in\mathcal{G}}\sum_{k\in\mathcal{K}_\mathrm{m}} \left|d_{lik}^{(n+1)} - \mathbf{a}_l^H\mathbf{c}_{lik} + q_{lik}^{(n)}\right|^2 + \frac{\rho}{2}\sum_{i\in\mathcal{G}} \|\boldsymbol{\mu}_i^{(n+1)} - \mathbf{a}_i + \mathbf{z}_i^{(n)}\|^2, \tag{3.42}$$

which by switching subscripts l and i in the second term, is equivalent to the following problem:

$$\min_{\mathbf{a}} \sum_{i\in\mathcal{G}} \|\widetilde{\mathbf{H}}_i\mathbf{a}_i\|^2 + \frac{\rho}{2}\sum_{i\in\mathcal{G}}\sum_{l\in\mathcal{G}}\sum_{k\in\mathcal{K}_\mathrm{m}} \left|d_{ilk}^{(n+1)} - \mathbf{a}_i^H\mathbf{c}_{ilk} + q_{ilk}^{(n)}\right|^2 + \frac{\rho}{2}\sum_{i\in\mathcal{G}} \|\boldsymbol{\mu}_i^{(n+1)} - \mathbf{a}_i + \mathbf{z}_i^{(n)}\|^2. \tag{3.43}$$

The above problem can be decomposed to G subproblems, one for each $\mathbf{a}_i, i \in \mathcal{G}$, given by

$$\min_{\mathbf{a}_i} \|\widetilde{\mathbf{H}}_i\mathbf{a}_i\|^2 + \frac{\rho}{2}\sum_{l\in\mathcal{G}}\sum_{k\in\mathcal{K}_\mathrm{m}} |d_{ilk}^{(n+1)} - \mathbf{a}_i^H\mathbf{c}_{ilk} + q_{ilk}^{(n)}|^2 + \frac{\rho}{2}\|\boldsymbol{\mu}_i^{(n+1)} - \mathbf{a}_i + \mathbf{z}_i^{(n)}\|^2. \tag{3.44}$$

The above problem is an unconstrained convex quadratic optimization problem for $\mathbf{a}_i$. By setting the derivative of the objective function to zero, we obtain the solution

Algorithm 1 ADMM-Based Algorithm for Solving $\mathcal{P}_{\text{mSCA}}(\mathbf{v})$

Initialization: Initialize $\mathbf{a}^{(0)} = \mathbf{v}$. Set $\mathbf{q}^{(0)} = \mathbf{0}$, and $\mathbf{z}^{(0)} = \mathbf{0}$. Set ρ. Set $n = 0$.
repeat

1. Update $\mathbf{d}^{(n+1)}$ by (3.32).
2. Update $\boldsymbol{\mu}_i^{(n+1)}$, $i \in \mathcal{G}$, by solving $\mathcal{P}_{\boldsymbol{\mu}}$using the solution in Case 1 or 2.
3. Update $\mathbf{a}_i^{(n+1)}$ by (3.45), for $i \in \mathcal{G}$.
4. Update $\mathbf{q}^{(n+1)}$ and $\mathbf{z}^{(n+1)}$ by (3.27).
5. Set $n \leftarrow n + 1$.

until convergence
Set $\mathbf{a}_i^\star(\mathbf{v}) = \mathbf{a}_i^{(n)}$, $i \in \mathcal{G}$.

Algorithm 2 The SCA for Solving $\mathcal{P}_{\text{m}}$

Initialization: Initialize $\mathbf{v}^{(0)}$. Set $t = 0$.
repeat

1. Obtain the optimal $\mathbf{a}_i^\star(\mathbf{v}^{(t)})$ to $\mathcal{P}_{\text{mSCA}}(\mathbf{v}^{(t)})$ via Algorithm 1.
2. Set $\mathbf{v}_i^{(t+1)} = \mathbf{a}_i^\star(\mathbf{v}^{(t)})$, $i \in \mathcal{G}$.
3. Set $t \leftarrow t + 1$.

until convergence

to (3.44) as follows:

$$\mathbf{a}_i^{(n+1)} = \frac{\rho}{2}\left(\widetilde{\mathbf{H}}_i\left(\mathbf{I} + \frac{\rho}{2}\sum_{l\in\mathcal{G}}\sum_{k\in\mathcal{K}_{\text{m}}}\mathbf{h}_{lk}\mathbf{h}_{lk}^H\right)\widetilde{\mathbf{H}}_i^H + \frac{\rho}{2}\mathbf{I}\right)^{-1} \cdot \left(\sum_{l\in\mathcal{G}}\sum_{k\in\mathcal{K}_{\text{m}}}\left(d_{ilk}^{(n+1)} + q_{ilk}^{(n)}\right)^*\widetilde{\mathbf{H}}_i^H\mathbf{h}_{lk} + \boldsymbol{\mu}_i^{(n+1)} + \mathbf{z}_i^{(n)}\right). \tag{3.45}$$

Our proposed ADMM-based algorithm for solving $\mathcal{P}_{\text{mSCA}}(\mathbf{v})$ per SCA iteration is summarized in Algorithm 1. For the sake of completeness, we summarize SCA for solving $\mathcal{P}_{\text{m}}$ in Algorithm 2 .

3.4 Summary of Algorithm

Our overall proposed AO-based algorithm to solve $\mathcal{P}_{\text{o}}$ is summarized in Algorithm 3. By adopting the optimal solution structures in (3.10) and (3.11), we convert the

Algorithm 3 The AO-Based Algorithm for Solving $\mathcal{P}_{\mathrm{o}}$

Initialization:
Obtain $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ in (3.13) by the fixed-point method in S1)-S4).
Convert $\mathcal{P}_{\mathrm{o}}$ to $\mathcal{P}'_{\mathrm{o}}$ using $\mathbf{R}(\boldsymbol{\lambda}, \boldsymbol{\nu})$ in (3.12).
//Solving $\mathcal{P}'_{\mathrm{o}}$
Initialize $\{\mathbf{a}_i^0\}$. Set $q = 0$.
repeat

1. Unicast: For fixed $\{\mathbf{a}_i^q\}$, update $\mathbf{b}^{q+1}$ using (3.20).
2. Multicast: For fixed $\mathbf{b}^{q+1}$, update $\{\mathbf{a}_i^{q+1}\}$ by solving $\mathcal{P}_{\mathrm{m}}$ via Algorithm 2.
3. Set $q \leftarrow q + 1$.

until convergence
Based on $\mathbf{b}^q$ and $\{\mathbf{a}_i^q\}$, compute $\{\mathbf{u}_j\}$ and $\{\mathbf{w}_i\}$ in (3.10) and (3.11), respectively.

original joint unicast and multicast beamforming problem $\mathcal{P}_{\mathrm{o}}$ to a joint weight optimization problem $\mathcal{P}'_{\mathrm{o}}$ w.r.t. the weight vectors $\mathbf{b}$ and $\{\mathbf{a}_i\}$. Using the AO approach, we separate $\mathcal{P}'_{\mathrm{o}}$ into two subproblems for unicast users and multicast users ($\mathcal{P}_{\mathrm{u}}$ and $\mathcal{P}_{\mathrm{m}}$), respectively and solve them alternatingly. A main benefit of having the two separate subproblems is that, for the unicast subproblem $\mathcal{P}_{\mathrm{u}}$, we obtained the closed-form solution for the weight vector $\mathbf{b}$ in (3.20). For the multicast subproblem $\mathcal{P}_{\mathrm{m}}$, we adopt SCA to solve $\mathcal{P}_{\mathrm{m}}$ iteratively for $\{\mathbf{a}_i\}$, where we propose an ADMM-based fast algorithm with the closed-form updates to solve convex approximation $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ for each SCA iteration (Algorithm 1).

As a result, our proposed algorithm is a three-layered algorithm. The outer layer is the AO approach alternating between the unicast and multicast subproblems. For the multicast subproblem, our algorithm contains another two-layered algorithm to solve $\mathcal{P}_{\mathrm{m}}$, where SCA is applied to solve $\mathcal{P}_{\mathrm{m}}$ and the ADMM-based algorithm is applied in each SCA iteration. The proposed algorithm is computational inexpensive due to the closed-form updates in each layer. In particular, since we obtain a simple closed-form solution for unicast users, the proposed algorithm is especially attractive for the systems with a large number of unicast users.

3.4.1 Convergence

Our proposed AO approach in Algorithm 3 is guaranteed to converge. To see this, at each AO iteration, we solve unicast subproblem $\mathcal{P}_{\mathrm{u}}$ optimally by the optimal solution in (3.20). For solving multicast subproblem $\mathcal{P}_{\mathrm{m}}$ using SCA, by the convergence guarantee of SCA, the objective value of $\mathcal{P}_{\mathrm{m}}$ is guaranteed to be always non-increasing from the previous iteration. Thus, the objective values in both unicast and multicast subproblem are ensured to be non-increasing at each AO iteration. As a result, since the objective function in $\mathcal{P}'_{\mathrm{o}}$ is bounded below, the proposed AO-based algorithm converges to a stationary-point of $\mathcal{P}_{\mathrm{o}}$.

3.4.2 Initialization

Algorithm 3 requires to a feasible initial point $\{\mathbf{a}_i^0\}$ for solving $\mathcal{P}'_{\mathrm{o}}$. As mentioned below $\mathcal{P}'_{\mathrm{o}}$, under the optimal beamforming structure, the converted weight optimization problem $\mathcal{P}'_{\mathrm{o}}$ has a relatively small problem size, depending only on the total number of users. Thus, we can use SDR with Gaussian randomization method to solve $\mathcal{P}'_{\mathrm{o}}$ to obtain an initial point, as suggested in [27].[4] For a problem of a small size, the computational cost of this method remains quite low. Furthermore, this method generally provides a good initial point, which leads to faster convergence of the entire algorithm.

In each AO iteration, to solve multicast problem $\mathcal{P}_{\mathrm{m}}$, SCA in Algorithm 2 also

[4] For this purpose, we treat $\mathcal{P}'_{\mathrm{o}}$ as a pure multicasting problem to apply SDR.

requires a feasible initial point $\mathbf{v}^{(0)}$ to $\mathcal{P}_{\mathrm{m}}$, for the sequential updates using $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$. We propose the following initialization method: In AO iteration $(q+1)$, use $\{\mathbf{a}_i^q\}$ from the previous AO iteration as the initial point $\{\mathbf{v}_i^{(0)}\}$ for Algorithm 2, *i.e.,* $\mathbf{v}_i^{(0)} = \mathbf{a}_i^q$, $i \in \mathcal{G}$. This initial point is guaranteed to be feasible for $\mathcal{P}_{\mathrm{m}}$ in this AO iteration. To see this, in AO iteration $(q+1)$, $\mathbf{b}^{q+1}$ is the optimal solution to $\mathcal{P}_{\mathrm{u}}$ for fixed $\{\mathbf{a}_i^q\}$. Therefore, $(\{\mathbf{a}_i^q\}, \mathbf{b}^{q+1})$ is feasible to $\mathcal{P}_{\mathrm{u}}$, satisfying (3.15b) and (3.15c). Since (3.21b) and (3.21c) are respectively the same constraints as (3.15b) and (3.15c), $(\{\mathbf{a}_i^q\}, \mathbf{b}^{q+1})$ is also feasible to $\mathcal{P}_{\mathrm{m}}$. Thus, setting $\mathbf{v}_i^{(0)} = \mathbf{a}_i^q$, $i \in \mathcal{G}$ ensures that the initial point $\mathbf{v}^{(0)}$ is feasible to $\mathcal{P}_{\mathrm{m}}$ with given $\mathbf{b}^{q+1}$.

3.4.3 Computational Complexity

We now analyze the computational complexity of the overall proposed algorithm. In Algorithm 3, the outer-layer AO approach has two main steps. For unicast subproblem, obtaining $\mathbf{b}^{q+1}$ in (3.20) requires a matrix inversion, which has the complexity of $O(K_{\mathrm{u}}^3)$. Note that, for the given user channels, this matrix inversion only needs to be performed once in Algorithm 3 to be used in all subsequent iterations. The rest computation in (3.20) only requires $O(K_{\mathrm{u}}^2)$ operations.

For solving multicast subproblem $\mathcal{P}_{\mathrm{m}}$ by SCA, at each SCA iteration, the ADMM-based method in Algorithm 1 has five closed-form updates. The update of $\mathbf{d}$ in (3.32) is a simple algebraic formula. Updating $\boldsymbol{\mu}$ requires solving $\mathcal{P}_{\boldsymbol{\mu}}$, for which the solution is given by either Case 1 or 2. As discussed in Remark 3.3.2, as the algorithm converges over iterations, the simple solution in Case 1 is the most likely case, which has no computation involved. At the initial iterations, the solution may be in Case 2, which requires iterative updates by (3.41) under the projected subgradient algorithm. The main computational complexity lies in the matrix inversion in (3.38) with the complexity of $O(K_{\mathrm{u}}^3)$. Finally, the update of $\mathbf{a}_i$ is given in (3.45). Although the expression requires a matrix inversion with the complexity of $O(K_{\mathrm{m}}^3)$, this matrix inversion is only a function of user channels and thus only needs to be computed once at the beginning of Algorithm 3. The remaining computation in (3.45) has the complexity of $O(K_{\mathrm{m}}^2)$.

From the above, excluding the matrix inversions in (3.20) and (3.45) performed at the beginning of Algorithm 3, the overall complexity in each AO iteration is $O(K_{\mathrm{u}}^3 + K_{\mathrm{m}}^2)$ for the beginning iterations and $O(K_{\mathrm{u}}^2 + K_{\mathrm{m}}^2)$ for the later iterations as the algorithm converges. We see that, for a large number of unicast users, our algorithm has a much lower computational complexity than directly solving the problem as a pure multi-group multicast problem as in (3.8).

Chapter 4

Simulation Results

In this chapter, we evaluate the performance of our proposed algorithm for joint downlink unicast and multi-group multicast transmission. In our simulation, unless otherwise specified, we set the default system setup as $G = 3$, $K_{\mathrm{m}} = 7$, $K_{\mathrm{u}} = 5$, and $\gamma_j^{\mathrm{u}} = \gamma_{ik}^{\mathrm{m}} = 10$ dB. The channel vectors of multicast and unicast users are generated i.i.d as $\mathbf{h}_{ik} \backsim \mathcal{CN}(\mathbf{0}, \beta_{ik}^{\mathrm{m}}\mathbf{I})$ and $\mathbf{g}_j \backsim \mathcal{CN}(\mathbf{0}, \beta_j^{\mathrm{u}}\mathbf{I})$, respectively, where β_{ik}^{m} and β_j^{u} represent the large-scale channel variation for multicast and unicast users, respectively, for $k \in \mathcal{K}_{\mathrm{m}}$, $i \in \mathcal{G}$, and $j \in \mathcal{K}_{\mathrm{u}}$. Based on the large-scale channel variation, we consider two types of channels in our experiments:

1. Pathloss channels: We set β_{ik}^{m} and β_j^{u} based on the pathloss model as $\beta_{ik}^{\mathrm{m}} = K_o d_{ik}^{\mathrm{m}-3}$, $\beta_j^{\mathrm{u}} = K_{\mathrm{o}} d_j^{\mathrm{u}-3}$, where d_{ik}^{m} (d_j^{u}) is the distance between the BS and multicast user k in group i (unicast user j), pathloss exponent is 3, and K_{o} is the pathloss constant. We set K_o such that the nominal average received signal-to-noise ratio (SNR) at the boundary of a cell of radius R is 0 dB, *i.e.,* $1 \cdot K_{\mathrm{o}} R^{-3}/\sigma^2 = 0$ dB.

2. Normalized channels: As a special case of pathloss channel, we consider that

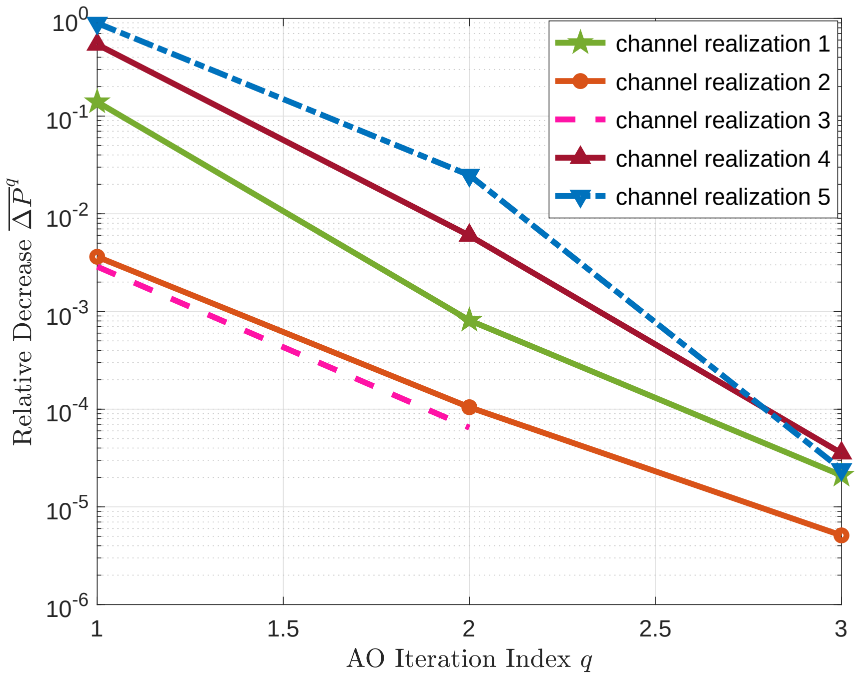

Figure 4.1: The convergence behavior of the outer-layer AO approach in Algorithm 3 $(N = 100, G = 3, K_{\mathrm{m}} = 7, K_{\mathrm{u}} = 5)$.

all unicast and multicast users have the same distance to the BS. Thus, we normalize the channel gain as $\beta^{\mathrm{m}}_{ik} = \beta^{u}_{j} = 1$ for all $k \in \mathcal{K}_{\mathrm{m}}$, $i \in \mathcal{G}$, and $j \in \mathcal{K}_{\mathrm{u}}$.

We average the performance results over 100 channel realizations per user (for pathloss channels, the performance result is averaged over 10 realizations of user locations). All the simulations are performed on MATLAB R2020a running on a Windows x64 machine with 2.90 GHz CPU and 16 GB RAM.

4.1 Convergence Behaviour

We first study the convergence behaviour of our proposed algorithm (UMBF-ADMM). Recall that the algorithm contains three layers of iterations: AO, SCA, and ADMM. The convergence behaviour of the outer-layer AO approach in Algorithm 3 is shown in

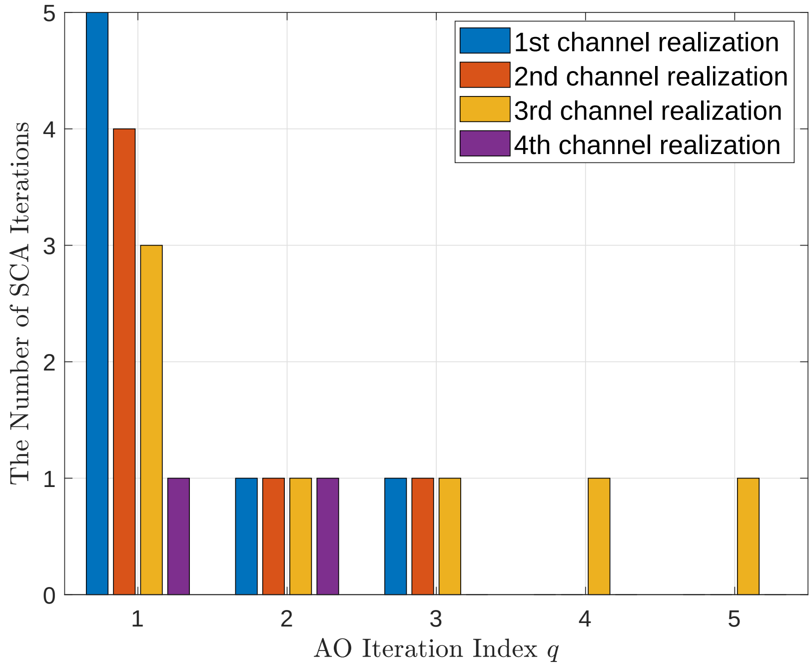

Figure 4.2: Convergence of the SCA layer for four random channel realizations ($N = 100, G = 3, K_{\mathrm{m}} = 7, K_{\mathrm{u}} = 5$)

Fig. 4.1, for five channel realizations. We define the relative decrease of the transmit power objective in $\mathcal{P}'_{\mathrm{o}}$ by

$$\overline{\Delta P}^q = |P_{\mathrm{tot}}^{q+1} - P_{\mathrm{tot}}^q|/P_{\mathrm{tot}}^q, \tag{4.1}$$

where P_{tot}^q denotes the total power obtained at iteration q. We see that the algorithm converges rapidly. Within four iterations, the relative decrease $\overline{\Delta P}^q$ drops below 10^{-4}. Fig. 4.2 shows the number of SCA iterations required to solve multicast subproblem $\mathcal{P}_{\mathrm{m}}$ in Algorithm 2 at different AO iterations, for four random channel realizations. We set the convergence threshold for the relative decrease of the objective function in $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ to be 10^{-4}. We observe that the number of SCA iterations needed for convergence decreases fast over AO iterations. Typically, for the first AO iteration, more SCA iterations are needed for convergence. After one AO iteration, mostly only

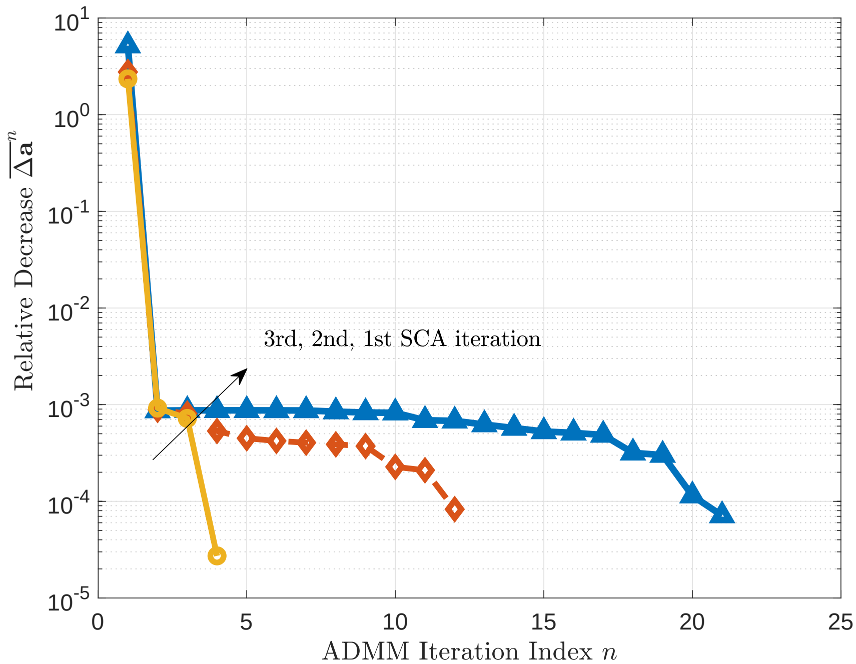

Figure 4.3: The convergence behavior of the ADMM algorithm (Algorithm 1) at different SCA iterations ($N = 100, G = 3, K_{\mathrm{m}} = 7, K_{\mathrm{u}} = 5$).

one iteration is enough to reach the convergence.

The convergence behaviour of ADMM in Algorithm 1 in the innermost layer of UMBF-ADMM is presented in Fig. 4.3. We define the relative error in the multicast weight vector update $\mathbf{a}^{(n)}$ as

$$\overline{\Delta \mathbf{a}}^{(n)} = \max_{1 \leq i \leq G} \{ \|\mathbf{a}_i^{(n+1)} - \mathbf{a}_i^{(n)}\| / \|\mathbf{a}_i^{(n)}\| \}. \tag{4.2}$$

Using this relative error as the convergence criterion, we set the convergence threshold to 10^{-4}. In Fig. 4.3, we plot the convergence curves associated with different SCA iterations (*e.g.*, $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v}^{(t)})$ for different iteration index t). We observe that the algorithm converges much faster over SCA iterations. The required iterations for the relative error of 10^{-4} changes from 20 iterations in the first SCA iteration to 4 iterations in the third SCA iteration. This trend is expected as SCA converges, the

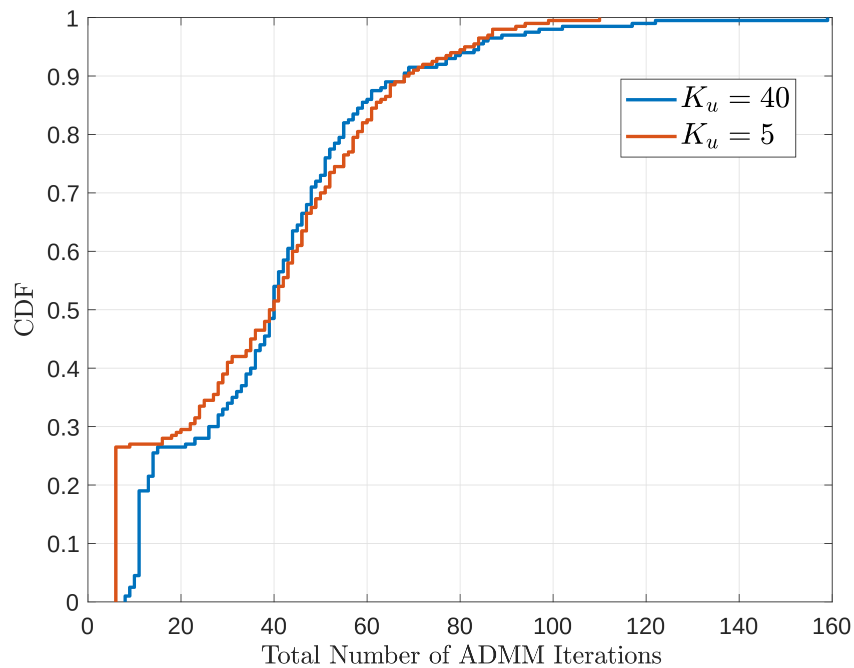

Figure 4.4: CDF of the total number of ADMM iterations ($N = 100, G = 3, K_{\mathrm{m}} = 7$)

initial point chosen in Algorithm 1 is closer to the optimal solution to the multicast subproblem.

The advantage of our proposed algorithm is that the convergence behaviour of our proposed algorithm is not affected by the number of unicast users, since the unicast problem is solved separately by a closed-form expression. To verify this, in Fig. 4.4, we plot the cumulative distribution function (CDF) of the total number of ADMM iterations of all AO iterations required for convergence, generated over 100 channel realizations. As expected, the total number of ADMM iterations remains roughly unchanged for $K_{\mathrm{u}} = 5$ and 40.

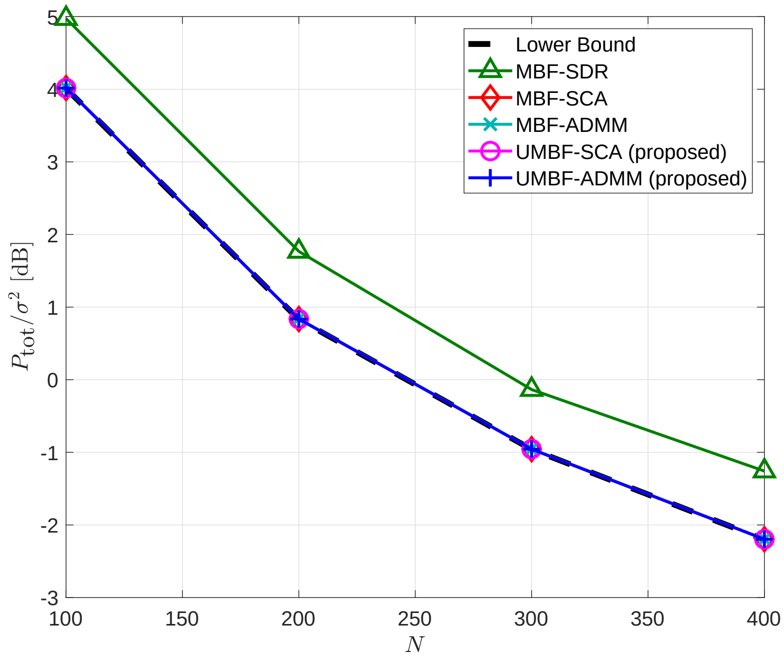

Figure 4.5: Normalized transmit power $P_{\mathrm{tot}}/\sigma^2$ vs. N for Normalized Channels ($G = 3, K_{\mathrm{u}} = 5, K_{\mathrm{m}} = 7$)

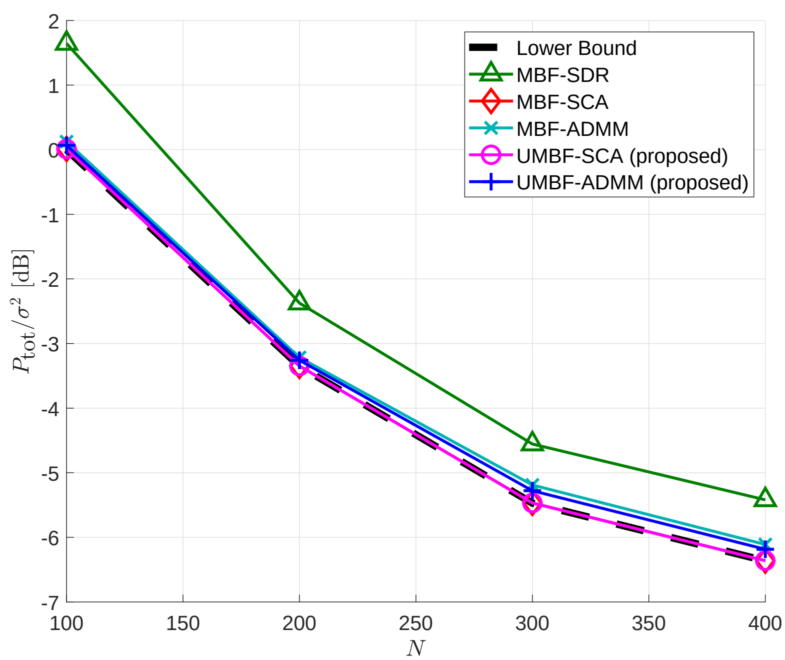

Figure 4.6: Normalized transmit power $P_{\mathrm{tot}}/\sigma^2$ vs. N for Pathloss Channels ($G = 3, K_{\mathrm{u}} = 5, K_{\mathrm{m}} = 9$)

Table 4.1: Average Computation Time over N (sec.) $(K_{\mathrm{u}} = 5, K_{\mathrm{m}} = 7, G = 3)$

N	100	200	300	400
UMBF-ADMM (proposed)	0.08	0.11	0.15	0.21
UMBF-SCA (proposed)	8.47	7.36	5.31	5.16
MBF-ADMM [80]	0.33	0.21	0.29	0.42
MBF-SCA [27]	15.79	13.59	11.31	10.33
MBF-SDR [27]	0.84	0.84	0.85	0.87

4.2 Performance Comparison

We now compare the performance of our proposed algorithm with other existing algorithms. We refer our proposed ADMM-base algorithm (Algorithms 1-3) for joint unicast and multicast beamforming as UMBF-ADMM. For comparison, for our proposed approach, instead of ADMM in Algorithm 1, we also consider directly solving $\mathcal{P}_{\mathrm{mSCA}}(\mathbf{v})$ in each SCA iteration using the classical interior-point method via standard convex solver CVX in MATLAB. We name this method as UMBF-SCA. Furthermore, we consider three alternative algorithms, which use the optimal beamforming structure but consider $\mathcal{P}'_{\mathrm{o}}$ as a pure multicasting problem to solve: 1) MBF-SDR [27]: Computing unicast and multicast weight vectors by SDR and Gaussian randomization. 2) MBF-SCA [27]: Computing unicast and multicast weight vectors by SCA, where the interior-point method is used in each SCA iteration. 3) MBF-ADMM [80]: The same as MBF-SCA, except that an ADMM-based method is used in each SCA iteration. Note that, different from our proposed algorithms, in all the above three methods, $\mathcal{P}'_{\mathrm{o}}$ is treated as a pure multicasting problem, and the unicast and multicast

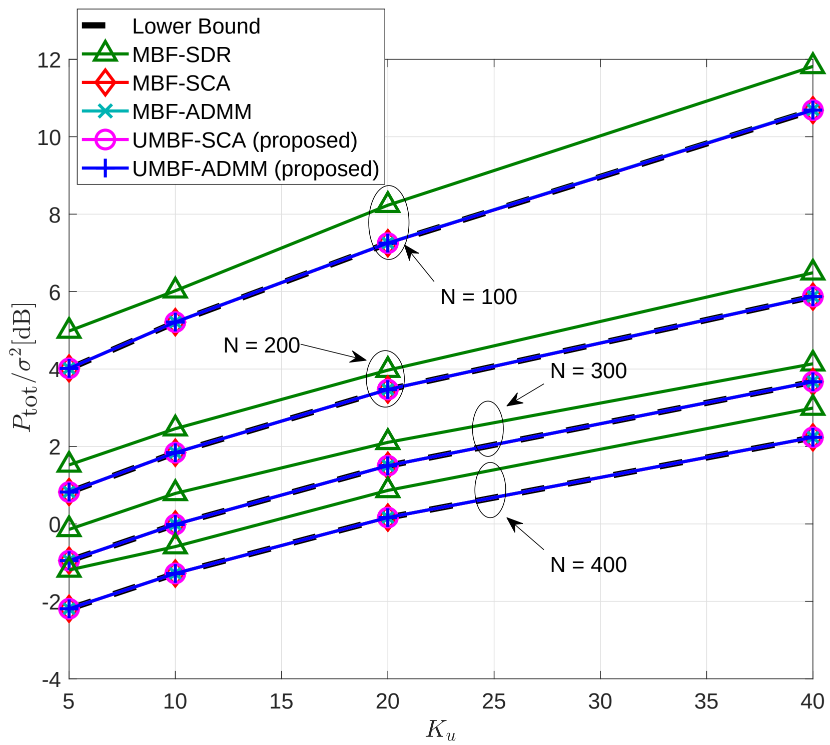

Figure 4.7: Normalized transmit power $P_{\mathrm{tot}}/\sigma^2$ vs. K_{u} for Normalized Channels $(G = 3, K_{\mathrm{m}} = 7)$

weight vectors are not separately computed. Finally, we consider a lower bound to $\mathcal{P}_{\mathrm{o}}$, which is obtained by solving $\mathcal{P}_{\mathrm{o}}$ directly via SDR. It is used as the performance benchmark for all performance plots.

In Fig. 4.5, we plot the total transmit power over noise, $P_{\mathrm{tot}}/\sigma^2$, vs. the number of antennas N obtained by different methods, where normalized channels are used. We observe that our proposed algorithms (UMBF-ADMM and UMBF-SCA) attain the lower bound for all values of N. The same is for MBF-SCA and MBF-ADMM. This shows that our proposed AO approach alternatingly solving unicast and multicast problems provides equal performance as the existing algorithms treating the problem as a pure multicasting problem. As expected, MBF-SDR has a constant gap of about 1 dB to the lower bound over N. This SDR-based method is an approximate method,

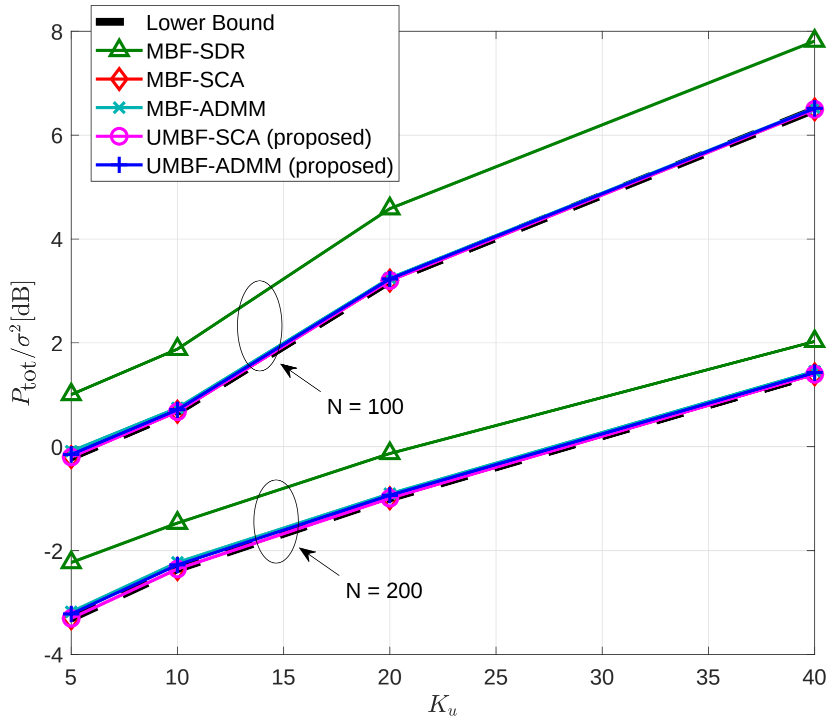

Figure 4.8: Normalized transmit power P_{tot}/σ^2 vs. K_{u} for Pathloss Channels ($G = 3, K_{\text{m}} = 9$)

whose performance deteriorates as the problem size, in terms of the number of constraints (*i.e.,* the total number of users), is relatively large. We also present P_{tot}/σ^2 vs. N in Fig. 4.6, where pathloss channels are considered. Similar performance as in Fig. 4.5 is observed, where the performance of UMBF-ADMM is near-optimal at all N values.

To see the computational advantage of our proposed method over other methods, we show the average computation time vs. N in Table 4.1. Since the optimal beamforming structure is employed in all methods, the computation time does not have noticeable increase over N for all methods.[1] Further more, our proposed method

[1]The computation time of the SCA-based methods decreases as N initially increases from 100 to 300. This is mainly due to fewer SCA iterations are needed for convergence as N changes from moderately large to large value, with a better initial point closer to the (local) optimum.

Table 4.2: Average Computation Time over K_{u} (sec.) $(K_{\mathrm{m}} = 7, G = 3, N = 100)$

K_{u}	5	10	20	40
UMBF-ADMM (proposed)	0.06	0.07	0.09	0.11
UMBF-SCA (proposed)	9.21	10.27	18.53	27.87
MBF-ADMM [80]	0.26	0.39	1.61	4.86
MBF-SCA [27]	16.44	24.22	71.28	185.08
MBF-SDR [27]	0.86	1.30	2.60	6.41

UMBF-ADMM is the fastest algorithm in obtaining the solution. It provides magnitudes of complexity reduction from that in MBF-SCA attributing to both ADMM and the AO approach. Note also that, as discussed in Section 3.4.2 on initialization, for a smaller problem size, SDR is much faster than SCA to compute a solution, but at the expense of a noticeably worse performance as in Fig. 4.5.

To further demonstrate the advantage of UMBF-ADMM in the scenario of having a large number of unicast users, we show $P_{\mathrm{tot}}/\sigma^2$ vs. the number of unicast users K_{u} in Fig. 4.7 for N ranges from 100 to 400. Normalized channels are considered. Again, UMBF-ADMM and UMBF-SCA show a near-optimal performance as compared to the lower bound for all values of K_{u} and N. Fig. 4.8 presents $P_{\mathrm{tot}}/\sigma^2$ vs. K_{u} for $N = 100$ and 200, when pathloss channels are used. Similar performance as in Fig. 4.7 is observed, where the performance of UMBF-ADMM is near-optimal at all K_u and N values.

With the favorable performance, Table 4.2 shows the computation time for the corresponding plots in Fig. 4.7. The computational advantage of UMBF-ADMM is clearly seen. As K_{u} increases, the computation time of UMBF-ADMM increases only slightly but roughly maintains the same order. In contrast, the computation time of other methods increases substantially with K_{u}. For relatively large K_{u}, UMBF-ADMM is significantly faster than other methods with several order of magnitude of complexity reduction. This shows the clear advantage of UMBF-ADMM as an attractive method for joint unicast and multicast beamforming in large-scale massive MIMO systems.

Chapter 5

Conclusions and Future Work

In this thesis, we proposed a low-complexity high-performing algorithm for the joint downlink unicast and multi-group multicast beamforming QoS problem, especially suitable for massive MIMO systems with a large number of unicast users present. We utilized the optimal multicast beamforming structure recently obtained to lower the problem dimension and proposed to separate the original problem into two subproblems for the respective unicast and multicast users to be solved alternatingly using the AO approach. By linking the unicast subproblem to the multiuser unicast power minimization, we obtained the solution to the unicast subproblem in closed-form. For the multicast subproblem, we apply SCA to solve it, where we proposed an ADMM-based fast algorithm to solve the convex approximation problem at each SCA iteration.

Our overall proposed algorithm is a multi-layered iterative algorithm using either closed-form updates or a fast ADMM-based algorithm to compute the updates. The algorithm has a low computational complexity, which grows very mildly with K_{u} and N, and thus, is especially suitable for massive MIMO systems with a large number of unicast users. Simulation results showed that our proposed algorithms achieve a near-optimal performance for a wide range of values of N and K_{u}, with the computation complexity that is several order of magnitude lower than existing algorithms.

Several extensions are possible on this work. To further analyze the performance of the proposed algorithm, we can consider the imperfect CSI. In reality, the BS can't obtain the perfect CSI and this would affect the performance of the beamforming algorithms. Especially, for massive MIMO systems, obtaining the accurate CSI is challenging in a fading environment. Therefore, we can extend our problem to develop a robust algorithm for joint unicast and multicast transmission, by taking the imperfect CSI into consideration. Moreover, we can extend our problem to the MMF problem, where beamforming vectors are designed to maximize the minimum SINR among all users, under a sum power constraint. We can solve the MMF problem by solving the QoS problem iteratively based on the inverse relation between MMF and QoS problems. Furthermore, this work can be extended to study the problem of the multi-cell joint unicast and multicast beamforming in cellular networks, where the inter-cell interference must be considered. In this work, we analyzed single-carrier scenario. However, in practical systems, multi-carrier transmission scheme has been widely used, and finding the efficient beamforming solution under this case is of practical importance. Thus, one possible extension can be obtaining the optimal solution for joint unicast and multicast beamforming problem in large-scale multi-carrier networks.

Bibliography

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[3] E. BjÃűrnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, 2016.

[4] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[5] E. BjÃűrnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Processing Mag.*, vol. 31, no. 4, pp. 142–148, 2014.

[6] N. D. Sidiropoulos, T. N. Davidson, and Zhi-Quan Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, 2006.

[7] 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.

[8] T. Chang, Z. Luo, and C. Chi, "Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3932–3943, 2008.

[9] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5132–5142, 2014.

[10] M. Jordan, X. Gong, and G. Ascheid, "Multicell multicast beamforming with delayed SNR feedback," in *Proc.IEEE Global Telecommn. Conf. (GLOBECOM)*, 2009, pp. 1–6.

[11] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, 2013.

[12] M. Dong and B. Liang, "Multicast relay beamforming through dual approach," in *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2013, pp. 492–495.

[13] N. Bornhorst, M. Pesavento, and A. B. Gershman, "Distributed beamforming for multi-group multicasting relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 221–232, 2012.

[14] M. R. A. Khandaker and Y. Rong, "Precoding design for MIMO relay multicasting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544–3555, 2013.

[15] Y. C. B. Silva and A. Klein, "Linear transmit beamforming techniques for the multigroup multicast scenario," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4353–4367, 2009.

[16] K. Cumanan, L. Musavian, S. Lambotharan, and A. B. Gershman, "SINR balancing technique for downlink beamforming in cognitive radio networks," *IEEE Signal Processing Lett.*, vol. 17, no. 2, pp. 133–136, 2010.

[17] A. H. Phan, H. D. Tuan, H. H. Kha, and D. T. Ngo, "Nonsmooth optimization for efficient beamforming in cognitive radio multicast transmission," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2941–2951, 2012.

[18] K. T. Phan, S. A. Vorobyov, N. D. Sidiropoulos, and C. Tellambura, "Spectrum sharing in wireless networks via qos-aware secondary multicast beamforming," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2323–2335, 2009.

[19] Y. Huang, Q. Li, W.-K. Ma, and S. Zhang, "Robust multicast beamforming for spectrum sharing-based cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 527–533, 2012.

[20] Q.-D. Vu, K.-G. Nguyen, and M. Juntti, “Weighted maxâĂŞmin fairness for c-ran multicasting under limited fronthaul constraints,” *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1534–1548, 2018.

[21] D. Christopoulos, S. Chatzinotas, and B. Ottersten, “Multicast multigroup beamforming for per-antenna power constrained large-scale arrays,” in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2015, pp. 271–275.

[22] M. Sadeghi, L. Sanguinetti, R. Couillet, and C. Yuen, “Reducing the computational complexity of multicasting in large-scale antenna systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2963–2975, 2017.

[23] E. Chen and M. Tao, “ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems,” *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2685–2698, 2017.

[24] B. Wu, K. C. Lin, K. Hsu, and H. Wei, “Hybridcast: Joint multicast-unicast design for multiuser MIMO networks,” in *Proc. IEEE Conf. on Computer Communications (INFOCOM)*, 2015, pp. 1724–1732.

[25] E. G. Larsson and H. V. Poor, “Joint beamforming and broadcasting in massive MIMO,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 3058–3070, 2016.

[26] M. Sadeghi, E. BjÃűrnson, E. G. Larsson, C. Yuen, and T. Marzetta, “Joint unicast and multi-group multicast transmission in massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6375–6388, 2018.

[27] M. Dong and Q. Wang, "Multi-group multicast beamforming: Optimal structure and efficient algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 3738–3753, 2020.

[28] M. Dong and Q. Wang, "Optimal multi-group multicast beamforming structure," in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2019, pp. 1–5.

[29] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Mag.*, vol. 27, no. 3, pp. 62–75, 2010.

[30] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over emerging 5G networks: Challenges and perspectives," *IEEE Network*, vol. 31, no. 2, pp. 80–89, 2017.

[31] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, 1998.

[32] C. Farsakh and J. Nossek, "Spatial covariance based downlink beamforming in an SDMA mobile radio system," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1497–1506, 1998.

[33] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," *Handbook of Antennas in Wireless Communications*, 01 2001.

[34] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Mag.*, vol. 27, no. 3, pp. 20–34, 2010.

[35] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 12–21, 2013.

[36] L. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Processing Lett.*, vol. 21, no. 1, pp. 114–117, 2014.

[37] G. Scutari, F. Facchinei, L. Lampariello, S. Sardellitti, and P. Song, "Parallel and distributed methods for constrained nonconvex optimization-part ii: Applications in communications and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1945–1960, 2017.

[38] G. Venkatraman, A. TÃűlli, M. Juntti, and L.-N. Tran, "Multigroup multicast beamformer design for MISO-OFDM with antenna selection," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5832–5847, 2017.

[39] N. Bornhorst and M. Pesavento, "An iterative convex approximation approach for transmit beamforming in multi-group multicasting," in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2011, pp. 426–430.

[40] Y. Gao and M. Schubert, "Group-oriented beamforming for multi-stream multicasting based on quality-of-service requirements," in *1st IEEE International*

Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005., 2005, pp. 193–196.

[41] J. Yu and M. Dong, “Low-complexity weighted MRT multicast beamforming in massive MIMO cellular networks,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2018, pp. 3849–3853.

[42] J. Yu and M. Dong, “Distributed low-complexity multi-cell coordinated multicast beamforming with large-scale antennas,” in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2018, pp. 1–5.

[43] C. Zhang, M. Dong, and B. Liang, “First-order fast algorithm for structurally optimal multi-group multicast beamforming in large-scale systems,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2021, pp. 4790–4794.

[44] ——, “Fast first-order algorithm for large-scale max-min fair multi-group multicast beamforming,” *arXiv preprint arXiv:2107.07540*, 2021.

[45] Y. C. B. Silva and A. Klein, “Adaptive beamforming and spatial multiplexing of unicast and multicast services,” in *IEEE Int. Symposium on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2006, pp. 1–5.

[46] L. Zhang, W. Li, Y. Wu, X. Wang, S. Park, H. M. Kim, J. Lee, P. Angueira, and J. Montalban, “Layered-division-multiplexing: Theory and practice,” *IEEE Trans. Broadcast*, vol. 62, no. 1, pp. 216–232, 2016.

[47] J. Zhao, D. GÃijndÃijz, O. Simeone, and D. GÃşmez-Barquero, "Non-orthogonal unicast and broadcast transmission via joint beamforming and LDM in cellular networks," *IEEE Trans. Broadcast*, vol. 66, no. 2, pp. 216–228, 2020.

[48] L. Fay, L. Michael, D. GÃşmez-Barquero, N. Ammar, and M. W. Caldwell, "An overview of the ATSC 3.0 physical layer specification," *IEEE Trans. Broadcast*, vol. 62, no. 1, pp. 159–171, 2016.

[49] D. GÃşmez-Barquero and O. Simeone, "LDM versus FDM/TDM for unequal error protection in terrestrial broadcasting systems: An information-theoretic view," *IEEE Trans. Broadcast*, vol. 61, no. 4, pp. 571–579, 2015.

[50] D. Kim, F. Khan, C. V. Rensburg, Z. Pi, and S. Yoon, "Superposition of broadcast and unicast in wireless cellular systems," *IEEE Commun. Mag.*, vol. 46, no. 7, pp. 110–117, 2008.

[51] E. Chen, M. Tao, and Y. Liu, "Joint base station clustering and beamforming for non-orthogonal multicast and unicast transmission with backhaul constraints," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6265–6279, 2018.

[52] L. You, J. Xiong, K. Li, W. Wang, and X. Gao, "Non-orthogonal unicast and multicast transmission for massive MIMO with statistical channel state information," *IEEE Access*, vol. 6, pp. 66 841–66 849, 2018.

[53] O. Tervo, L. Tran, S. Chatzinotas, M. Juntti, and B. Ottersten, "Energy-efficient joint unicast and multicast beamforming with multi-antenna user ter-

minals," in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2017, pp. 1–5.

[54] Y. Li, M. Xia, and Y. Wu, "Energy-efficient precoding for non-orthogonal multicast and unicast transmission via first-order algorithm," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4590–4604, 2019.

[55] Y. Liu, C. Lu, M. Tao, and J. Wu, "Joint multicast and unicast beamforming for the MISO downlink interference channel," in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2017, pp. 1–5.

[56] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: a promising phy-layer strategy for lte evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, 2016.

[57] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, 2019.

[58] H. Joudeh and B. Clerckx, "Rate-splitting for max-min fair multigroup multicast beamforming in overloaded systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7276–7289, 2017.

[59] O. Tervo, L. Trant, S. Chatzinotas, B. Ottersten, and M. Juntti, "Multigroup multicast beamforming and antenna selection with rate-splitting in multicell systems," in *Proc. IEEE Workshop on Signal Processing advances in Wireless Commun.(SPAWC)*, 2018, pp. 1–5.

[60] H. Chen, D. Mi, B. Clerckx, Z. Chu, J. Shi, and P. Xiao, "Joint power and subcarrier allocation optimization for multigroup multicast systems with rate splitting," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2306–2310, 2020.

[61] Y. Mao, B. Clerckx, and V. O. K. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *Int. Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–5.

[62] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[63] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[64] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, 2006, pp. 359–363.

[65] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, 2007.

[66] D. Gesbert, M. Kountouris, R. W. Heath, C.-b. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Mag.*, vol. 24, no. 5, pp. 36–46, 2007.

[67] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[68] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[69] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. ApelfrÃűjd, and T. Svensson, “The role of small cells, coordinated multipoint, and massive MIMO in 5G,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, 2014.

[70] R. Chataut and R. Akl, “Massive MIMO systems for 5G and beyond network-sâĂŤoverview, recent trends, challenges, and future research direction,” *Sensors*, vol. 20, no. 10, 2020.

[71] E. BjÃűrnson, E. G. Larsson, and M. Debbah, “Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?” *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, 2016.

[72] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[73] E. BjÃűrnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO has unlimited capacity,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, 2018.

[74] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers.* Now Publishers Inc, 2011.

[75] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect csi: An ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, 2012.

[76] K. Huang and N. D. Sidiropoulos, "Consensus-ADMM for general quadratically constrained quadratic programming," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5297–5310, 2016.

[77] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, 2015.

[78] N. Mohamadi, M. Dong, and S. ShahbazPanahi, "ADMM-based fast algorithm for robust multi-group multicast beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2021, pp. 4440–4444.

[79] ——, "Low-complexity ADMM-based algorithm for robust multi-group multicast beamforming in large-scale systems," *IEEE Trans. Signal Process.*, Aug. 2021, submitted.

[80] C. Zhang, "First-order fast algorithms for structurally optimal large-scale multi-group multicast beamforming," Master's thesis, University of Toronto, Aug. 2021.

[81] X. Zhang, S. Sun, F. Qi, R. Bo, R. Q. Hu, and Y. Qian, “Massive MIMO based hybrid unicast/multicast services for 5G,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[82] F. Tan, P. Wu, Y.-C. Wu, and M. Xia, “Energy-efficient non-orthogonal multicast and unicast transmission of cell-free massive MIMO systems with SWIPT,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 949–968, 2021.

[83] V. Saxena, G. Fodor, and E. Karipidis, “Mitigating pilot contamination by pilot reuse and power control schemes for massive MIMO systems,” in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015, pp. 1–6.

[84] Y. Li, Y.-H. Nam, B. L. Ng, and J. Zhang, “A non-asymptotic throughput for massive MIMO cellular uplink with pilot reuse,” in *Proc.IEEE Global Telecommn. Conf. (GLOBECOM)*, 2012, pp. 4500–4504.

[85] D. Verenzuela, E. BjÃűrnson, X. Wang, M. Arnold, and S. ten Brink, “Massive-MIMO iterative channel estimation and decoding (miced) in the uplink,” *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 854–870, 2020.

[86] E. Balevi, A. Doshi, and J. G. Andrews, “Massive MIMO channel estimation with an untrained deep neural network,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2079–2090, 2020.

[87] Z. Jiang, S. Chen, S. Zhou, and Z. Niu, “Joint user scheduling and beam selection optimization for beam-based massive MIMO downlinks,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2190–2204, 2018.

[88] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.

[89] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[90] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University, Autumn Quarter*, vol. 2004, pp. 2004–2005, 2003.