Robust Multi-group Multicast Beamforming Design and Antenna Selection for massive MIMO Systems

by

Niloofar Mohamadi

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

Doctor of Philosophy
in
Electrical and Computer Engineering

Faculty of Engineering and Applied Science
University of Ontario Institute of Technology (Ontario Tech University)
Oshawa, Ontario, Canada

April, 2023

THESIS EXAMINATION INFORMATION

Submitted by: **Niloofar Mohamadi**

Doctor of Philosophy in **Electrical and Computer Engineering**

Thesis Title: Robust Multi-Group Multicast Beamforming and Antenna Selection for Massive MIMO Systems

An oral defense of this thesis took place on March 2nd, 2023 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee:	Prof. Khalid Elgazzar
Research Supervisor:	Prof. Min Dong
Research Co-supervisor:	Prof. Shahram ShahbazPanahi
Examining Committee Member:	Prof. Ali Grami
Examining Committee Member:	Prof. Ying Wang
University Examiner:	Prof. Jing Ren
External Examiner:	Prof. Saeed Gazor, Queen's 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

In this dissertation, we use an Alternating Direction Method of Multipliers (ADMM) algorithm to design robust multi-group multicast beamforming scheme for massive multiple-input multiple-output (MIMO) systems for two scenarios; 1) all antennas are available at the base station (BS), 2) only subset of antennas are available due limitation on the number of RF chains (joint antenna selection and beamforming design). In both scenarios we assume only estimates of the channel covariance matrices are available at the base station with a bounded error. In the first scenario, we formulate the robust multicast beamforming optimization problem to minimize the transmit power while the minimum required quality-of-service (QoS) is met. We directly solve the formulated optimization problem. We develop a two-layered ADMM-based fast algorithm to directly tackle the non-convex problem, where we obtain closed-form or semi-closed-form solutions to each subproblem. Simulation results show that our proposed algorithm provides a favorable performance compared with existing alternative methods with considerable computational complexity reduction.

In the second scenario, we investigate the problem of joint antenna selection and robust multi-group multicast beamforming scheme for massive MIMO systems. We aim to obtain binary antenna selection and multicast beamforming vectors in order to minimize power consumption at the BS subject to per antenna transmit power limit, and the minimum worst case SINR requirements while limited number of antennas can be exploited for beamforming. To overcome the acquired mixed-Boolean problem, we replace the binary constraints with a continuous equivalent form. We use exact penalty function to transfer the obtained continuous problem into a more tractable problem. We develop a two phases solution, with phase one focusing on antenna selection, and second phase to obtain multicast beamforming vectors using the selected antennas in the first phase. In phase one, we propose two different approaches, SINR-based and SLR-based, to acquire antenna selection vector. We exploit a fast ADMM algorithm to directly tackle the problems in phase one and two without applying convex approximation technique that may result in performance degradation. Simulation results illustrate the advantage of our proposed algorithm

in terms of computational complexity over existing alternative methods while maintaining favorable performance.

Keywords: *Robust multicast beamforming; Antenna Selection; large-scale systems; alternating direction method of multipliers (ADMM); low complexity;*

AUTHORS DECLARATION

I, Niloofar Mohamadi, 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.

STATEMENT OF CONTRIBUTIONS

Part of this dissertation (described in Chapters 3) have been published as:

- N. Mohamadi, M. Dong and S. ShahbazPanahi, “Admm-Based Fast Algorithm for Robust Multi-Group Multicast Beamforming,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2021, pp. 4440-4444.

- N. Mohamadi, M. Dong and S. ShahbazPanahi, “Low-Complexity ADMM-Based Algorithm for Robust Multi-Group Multicast Beamforming in Large-Scale Systems,” *IEEE Trans. Signal Process*, vol. 70, pp. 2046-2061, 2022.

Part of this dissertation (described in Chapters 4) have been submitted as:

- N. Mohamadi , M. Dong and S. ShahbazPanahi. “Joint Antenna Selection and Robust Multi-group Multicast Beamforming in Massive MIMO Systems”, submitted to *IEEE Trans. Signal Process*, March, 2023.

I hereby certify that I am the sole author of this thesis. I have used standard referencing practices to acknowledge ideas, research techniques, or other materials that belong to others. Furthermore, I hereby certify that I am the sole source of the creative works and/or inventive knowledge described in this thesis.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep and sincere appreciation to my supervisors Prof. Min Dong and Prof. Shahram ShahbazPanahi for their unconditional support, patient and guidance. I cannot express how grateful I am for all the teachings, insightful feedbacks, and constructive criticism. This endeavor would not have been possible without their support and guidance.

Last but not least, my heartfelt gratitude to my husband Mahdi for his unconditional love , support, patient and unwavering belief in me at every stage of the research project, to my parents Nahid and Sediq for their unconditional love, giving me strength to chase my dreams, to my brothers, Moien , Matin and Saeed Khaki for their encouragement and love, and finally to all my friends for their support and kindness.

Contents

List of Acronyms

BS	Base Station
MIMO	Multi Input Multi Output
ADMM	Alternating Direction Method of Multipliers
QoS	Quality of Service
MMF	Max-Min Fair
QCQP	Quadratically Constrained Quadratic Programming
CSI	Channel State Information
SDR	Semi-Definite Relaxation
SCA	Successive Convex Approximation
SINR	Signal-to-Interference-plus-Noise Ratio
SLR	Signal to Leakage Ratio
TDD	Time-Division Duplex
BAB	Branch and Bound

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

1.1.1 5G Technology

The rapid growth of wireless connected devices and the emergence of new applications raise the challenge of bandwidth limit and higher data rate requirements. A solution to address the challenges of higher data rate demands is 5G technology. The Internet of Things (IoT) and mobile internet are the critical factors driving 5G technology. The 5G applications are categorized into three groups: Enhanced Mobile Broadband (eMBB), Massive Machine Type Communications (mMTC), and Ultra-Reliable and Low Latency Communications (URLLC) [1, 2]. In 5G-related applications, expected improvement areas include system throughput, energy efficiency, latency, and wireless network capacity [3]. To address the required improvement areas in 5G, many new technologies, such as heterogeneous cellular networks [4–6], cell-free architecture [7, 8], energy harvesting communications [9–12] are introduced. Two significant factors in selecting the candidate technologies to address the emerging challenges of the 5G applications are bandwidth (spectral efficiency) and energy efficiency. Energy efficiency is an important factor in wireless networks especially with increasing number of users in 5G networks [13].

Three of focus areas to address 5G technology requirements are 1) Multiple input multiple outputs (MIMO) for spectral efficiency improvement, 2) mm-wave signals usage to increase the bandwidth, 3) deployment of multi-layer and ultra-dense networks for geo-

graphical spectral usage [3]. This dissertation focus is efficient MIMO transmission design. Hence, we discuss these systems in more detail in the following sections.

1.1.2 Massive MIMO Systems in 5G

As noted above, wireless networks experience a drastic increase in data transmission with emerging 5G network applications, such as, machine-to-machine and video streaming applications [14, 15]. In the decades ahead, the traffic load on wireless networks have significant upward trend due to emergence of 5G applications. Hence, the high volume of data exchange imposes challenges in designing wireless network systems. These challenges include a rise in the required spectrum resources and the requirement for energy efficiency in transmitting signals [14]. As a result, the technologies chosen for massive MIMO systems in 5G must increase the transmission capacity without increasing the required bandwidth resources.

By exploiting antenna arrays with large number of antenna elements both at the transmitter and receiver, Massive MIMO systems are the significant potential technology to address the challenges in increasing number of applications in 5G systems [16, 17]. One of the benefits of using a massive number of antennas is near-deterministic wireless channels, i.e, the channels from BS to the users become near-orthogonal. Formation of the near-orthogonal channels helps to remove small-scale fading, intra-cell interference, and uncorrelated noise in the presence of many antennas [18]. Additionally, massive MIMO systems improve the data rate and energy efficiency compared with single antenna systems.

1.1.3 Multicasting

Wireless multicasting is the process of simultaneously sending common information to a group of users. Figure 1.1 illustrates the multicasting concept of a group of users receiving common data. Multicast techniques are necessary for large-scale group communication applications [19]. Multicasting improves communication efficiency by reducing the information processing load at the servers and network devices. Furthermore, multicasting optimizes performance by a reduction in redundant traffic. In multicasting, the BS exploits

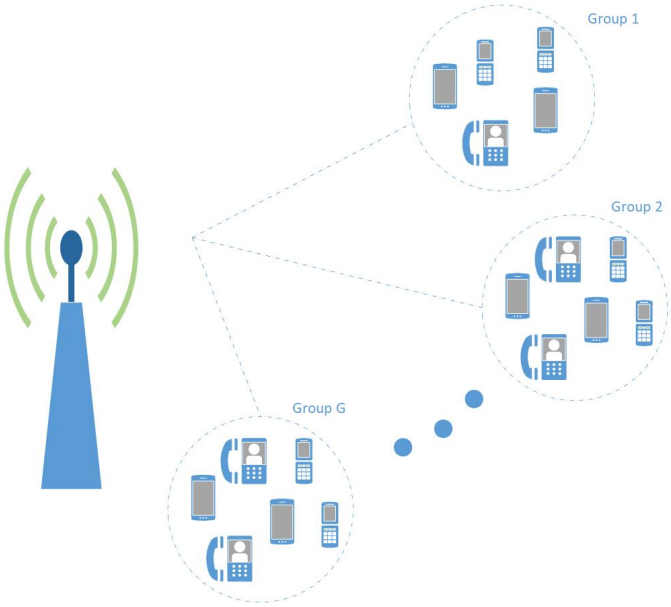

Figure 1.1: Multicasting technique

minimum network resources due to point-to-multipoint applications. Hence, wireless multicasting has become an essential technology for data delivery and content distribution in wireless services and applications in the future networks [20].

1.1.4 Beamforming

Beamforming is a technique that redirects the signal power to a specific direction rather than broadcasting the power to all directions. Figure 1.2 shows beamforming that directs power beams in a specific direction to cover the user with the intended message. The redirection of the signal power is based on the channel state information received at the BS [21]. Beamforming techniques improve spectral efficiency by increasing the signal-to-noise ratio. Furthermore, a combination of beamforming with techniques such as multiple-input multiple-output (MIMO) and intelligent antennas increase the capacity and cell range. Acquiring channel state information is challenging due to the noise of the environment. There are different techniques for estimating channel state information; however, in practice, the error associated with the estimation is unavoidable. Hence, to make the design robust to error, channel estimation error must be embedded in the design of the beamforming weights.

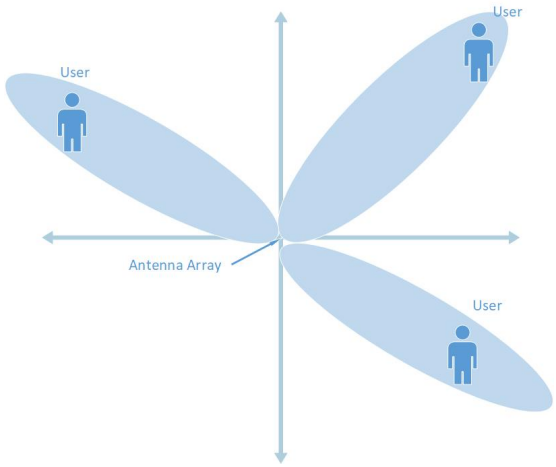

Figure 1.2: Beamforming

1.1.5 Multicast Beamforming

Physical-layer multicast beamforming is an efficient transmission technique to deliver common data to multiple users simultaneously [22–27]. It exploits the wireless broadcasting feature and directional signaling ability to boost achievable rate and improve the power and bandwidth efficiency [28].

Depending no the application, the design of multicast beamforming vectors at the BS, depending on the application, is commonly based on two approaches, first approach (known as the QoS problem), the objective is to minimize the transmit power subject to a minimum threshold required per user SINR [29–31]. It is common in the QoS problem to add per antenna power constraint to avoid excessive transmit power at each antenna [32]. The second approach (known as the max-min fair (MMF) problem) the goal is to maximize minimum SINR subject to transmit power limitation at the BS [25, 30, 33, 34]. The formulated problems for the two above approaches in the problem of beamforming design in multicast systems are non-convex and NP-hard. Consequently, they are challenging to solve. As previously stated in Section 1.1.2, with the advent of massive MIMO in 5G wireless communication technology, the size of problems to obtain the beamforming vector increases

considerably. This imposes the challenge of the development of multicast beamforming techniques that focus on practical and computationally efficient solutions suitable for such systems.

1.1.6 Robust Multicast Beamforming

One important factor in the design of multicast beamforming for both QoS and MMF problems is the availability of accurate channel state information (CSI) at the base station. However, in practice, the BS only has access to an estimation of the CSI in both frequency-division duplex (FDD) and time-division duplex (TDD) systems [35]. Furthermore, in massive MIMO systems with large number of antennas due to high interference, acquiring an accurate estimation of CSI is more challenging [36,37]. The common methods to estimate CSI include maximum-likelihood (ML) estimation [38], least-square (LS) estimation [39], and minimum-mean-square-error (MMSE) estimation [40]. In all the above methods, a known pilot signal is sent from a transmitter to receivers to estimate CSI. The pilot signal is affected by path loss, scattering, diffraction, fading, and shadowing, reflected in the quality of the channel, consequently, CSI. In the TDD system, both the transmitter and receiver use the same bandwidth, so downlink and uplink channels are reciprocal. However, due to different bandwidths in the downlink and uplink in FDD systems, acquiring the CSI is more challenging [41]. Thus, it is essential to incorporate the error in the CSI to have a robust design. The CSI estimation process involves high computational complexity, whereas, in massive MIMO systems with a large number of antennas, this complexity increases considerably [42, 43]. Furthermore, high variation in the instantaneous CSI causes short coherence time, leading to frequent channel acquisition [44]. The second-order statistics of the channel vectors, channel covariance matrix, evolves much slower than instantaneous CSI with higher coherence time. Hence, using a channel covariance matrix instead of instantaneous CSI remarkably reduces the computational complexity by saving the number of times required to obtain CSI [44].

In the design of multicast beamforming in the literature, it is common to consider two approaches for modeling error to guarantee a robust design. Most studies consider a

worst-case design assuming spherical bounded CSI error, where the worst-case SINR is considered for the QoS constraint for robust beamforming optimization [45]. The second approach is probabilistic design based on SINR outage [46].

1.1.7 Joint Antenna Selection and Robust Multicast Beamforming

The usage of all available antennas at the BS is restricted by the number of available RF chains devoted to each antenna. RF chain hardware has high cost and complexity as it includes a mixer, a power amplifier, and analog-to-digital (A/D) and digital-to-analog (D/A) converters. Hence, the number of available RF chains is less than the number of antennas as antenna production have low cost and complexity. Figure 1.3 illustrates such scenario with L available RF chains and M antennas at the base station serving multiple multicast groups.

Two techniques to exploit full capacity of available antennas and RF chains at the BS are soft and hard antenna selection approaches. In soft antenna selection technique (known as hybrid beamforming), a network of phase shifters connect antennas and RF chains. Soft selection approaches are categorized into two architectures, fully-connected and partly-connected [47]. In the fully-connected architecture each RF chain connects to all available antennas [47, 48]. The drawback of this structure is high power consumption as a large number of phase shifters are used. Nevertheless, due to the connection to all antennas, the fully-connected structure utilizes full array gain. In the partly-connected structure, each RF chains only connected to a subset of antennas, that leads to a lower circuit complexity and lower power consumption while have lower spectral efficiency compare with fully-connected structure [47]. In hard antenna selection techniques, the RF chains are connected to antennas through switches. In this technique a subset of antennas is switched to available RF chains. The hard antenna selection technique reduces the cost and complexity of the BS by exploiting switches.

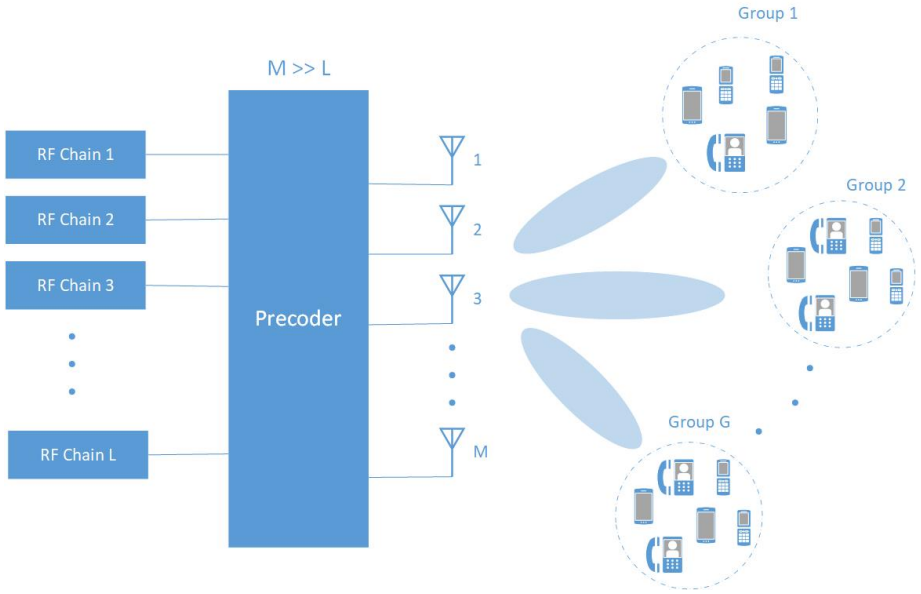

Figure 1.3: Antenna Selection scenario with L available RF chains at the base station, and M number of antennas, serving G multicast group.

1.2 Objectives and Contributions

The objectives and contributions of this dissertation are summarized below.

1.2.1 Objectives

The main objectives of this dissertation are summarized into two major completed works as follows:

- First, we provide a low-complexity robust multi-group multicast beamforming design suitable for massive MIMO systems. We assume that only the estimated channel covariance matrices are available at the BS and that the estimation error lies in a bounded hyper-sphere. Based on this, we consider the robust multicast quality of service (QoS) problem, aiming to minimize the transmit power subject to the worst-case (minimum) SINR guarantee.

- We then provide a low complexity solution for joint robust multi-group multicast beamforming design and antenna selection problem suitable for massive MIMO sys-

tems. We assume that only the estimated channel covariance matrices are available at the BS and the estimation error lies in a bounded hyper-sphere. Based on this assumption, we consider the robust multicast QoS problem, aiming to minimize the transmit power subject to the worst-case (minimum) SINR guaranteed and a limitation on the power radiated from each antennas.

1.2.2 Contributions

We divide the contribution of this dissertation into two parts for the two completed works on robust multi-group multicast beamforming design and joint robust multi-group multicast beamforming and antenna selection design. Our main contributions for the problem of robust multi-group multicast beamforming are summarized below:

- We develop our approach to directly solve the exact robust QoS problem. This is different from the existing common approach of using a lower bound on the worst-case SINR [49, 50]. Specifically, we reformulate the worst-case SINR constraint and obtain the exact worst-case expression for the bounded CSI error. Using this expression, we transform the original robust optimization into an equivalent non-convex optimization problem. This conversion is a crucial step for the subsequent development of our fast algorithm. Unlike the multicast QoS problem under perfect CSI, the converted optimization problem is not a QCQP problem, where the constraint function has a more complicated structure.

- We develop an ADMM-based fast algorithm to solve the converted non-convex problem directly with a convergence guarantee. Our proposed algorithm contains two layers of different ADMM procedures. We design the outer-layer ADMM to decompose the non-convex problem into two convex subproblems to be solved alternatingly. Efficiently solving these subproblems is still challenging. In particular, for one subproblem, we further develop an inner-layer consensus-ADMM-based algorithm to obtain the solution. A salient feature in our ADMM constructions at both layers is that by exploiting the structure of each subproblem and developing special

optimization techniques, we obtain either closed-form or semi-closed-form solutions to each subproblem. Therefore, all the ADMM updates can be performed using these derived solutions, significantly reducing the required computation and leading to a fast iterative algorithm. Based on the recently established result on the convergence of ADMM that covers a wide range of non-convex problems [51], we show that our proposed algorithm is guaranteed to converge to a stationary point of the original robust QoS problem.

- We conduct detailed simulation studies using two commonly considered models of the channel covariance matrix. Numerical studies show that our proposed ADMM-based algorithm provides fast convergence. It offers near-identical performance to that using the generic interior-point methods (IPM) offered by standard nonlinear solvers but with significantly lower complexity. Furthermore, our proposed algorithm provides a noticeable improvement over other existing robust design methods that rely on a lower bound on the worst-case SINR, with an improved likelihood of finding a feasible solution (*i.e.*, feasibility rate) and reduced transmit power consumption. Finally, as the number of antennas and users increases, our proposed algorithm maintains low computational complexity, with several orders of magnitude of time reduction as compared with other methods.

Our main contributions in the second work are summarized below:

- The problem of joint robust multi-group multicast beamforming design and antenna selection based on channel covariance matrix information has been studied for the first time based on best of authors knowledge. We use a Boolean selection vector for the antenna selection, followed by continuous relaxation to tackle mixed -Boolean problem. To avoid more relaxation leading to performance degradation we solve non-convex optimization problem directly using ADMM algorithm. This is different from existing studies using convex approximation to further simplify the optimization problem [52, 53]. Convex approximation causing either relaxation or tightening

of feasible set. In the former case, the solution of the optimization problem using relaxed feasible set can be infeasible to the original optimization problem, and the latter case causes extra power consumption in the BS due to limiting the feasible set.

- We use the result in [54] for exact worst case SINR to preserve extra power consumption due to conservative bound on the worst case SINR. Afterward, we develop a two phases algorithm to solve the problem of joint robust multicast beamforming and antenna selection. In the first phase due to continuous relaxation the obtained selection vector is not guaranteed to have binary entries. Therefore, the fucus of the first phase is to decide on the selected antennas by rounding the continuous selection vendor to binary values. In the initial phase to further reduce the computational complexity, we remove the associated term related to channel covariance matrix error, and consider a perfect case. This causes to improve the computational complexity considerably. The simulation results shows a near identical performance of the original problem with and without channel covariance matrix error consideration in the initial phase.

- As the scale of the optimization problem in first phase and second phase has direct relation with number of antennas and RF chains, respectively, the computational complexity of the first phase has the dominant portion of the computational complexity. Therefore, to decrease computational complexity in the first phase we propose to use SLR constraint with lower computational complexity compare with SINR. the simulation results shows near identical of the performance when SLR is used while dramatically lower the computational complexity.

- To tackle the non-convex equality resulting from RF chains, we use exact penalty method. The exact value of penalty constant is challenging to find, and depends on the problem structure. Exhaustive search for penalty term in this method imposes additional computational complexity. Our proposed ADMM-based algorithm eliminate the dependency of problem to penaltty constant, causes a considerable reduction in the computational complexity while keeping performance.

1.3 Outline of Dissertation

TThe rest of this thesis is organized as follows. In Chapter 2, we first review the literature study on multicast beamforming, the robust multicast beamforming cases. Moreover, we review the existing studies on joint antenna selection and multicast beamforming.

In Chapter 3 , we describe the problem of robust multicast beamforming. We describe our proposed approach to reformulate the robust multicast beamforming problem. We discuss our ADMM-based low-complexity algorithm to solve the reformulated non-convex optimization problem.

In Chapter 4 we present the problem of joint antenna selection and robust multicast beamforming. We propose our two-phase approach to solve the mixed-binary problem of joint antenna selection and robust multicast beamforming. We present the SINR-based and SLR-based approaches to solving phase one of the proposed two-phase approach to select the best antennas. We discuss our fast ADMM-based algorithm with low computational complexity to solve the acquired problem.

In Chapter 5, we discuss the conclusion of the two works, robust multicast beamforming and joint antenna selection and robust multicast beamforming, described in this thesis.

1.4 List of Publications

The research in this dissertation has resulted in the following list of publications [54, 55] .

1. N. Mohamadi, M. Dong and S. ShahbazPanahi, “Admm-Based Fast Algorithm for Robust Multi-Group Multicast Beamforming,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2021, pp. 4440-4444.

2. N. Mohamadi, M. Dong and S. ShahbazPanahi, “Low-Complexity ADMM-Based Algorithm for Robust Multi-Group Multicast Beamforming in Large-Scale Systems,” *IEEE Trans. Signal Process*, vol. 70, pp. 2046-2061, 2022.

3. N. Mohamadi , M. Dong and S. ShahbazPanahi. “Joint Antenna Selection and Robust Multi-group Multicast Beamforming in Massive MIMO Systems”, submitted to

IEEE Trans. Signal Process, March, 2023.

1.5 Notations

Hermitian, transpose, and conjugate are denoted by $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$, respectively. The Frobenius norm of a matrix is denoted by $\|\cdot\|_F$, and the Euclidean norm of a vector is denoted by $\|\cdot\|$. The Hadamard matrix product is denoted by $\odot$ .The inequality $\mathbf{A} \succcurlyeq 0$ means that matrix $\mathbf{A}$ is positive semi-definite, and tr$(\mathbf{A})$ denotes the trace of $\mathbf{A}$. The real part of a complex variable a is denoted by $\mathfrak{Re}\{a\}$. A complex Gaussian random vector $\mathbf{a}$ with zero mean and covariance matrix $\mathbf{C}$ is denoted by $\mathbf{a} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$.

Chapter 2

Literature Review

2.1 Massive MIMO

MIMO technology through improvement in the data throughput and range, has attracted a lot of attention. This improvement is without a considerable effect on bandwidth and transmits power. MIMO technology was first proposed in 1993 and 1994 [56]. The patent in [56] aimed to increase the throughput by co-locating several antennas at the transmitter. Next, in 1999, to illustrate the feasibility of MIMO in wireless networks, a practical laboratory experiment was done in [57].

To increase the gain of traditional MIMO, the concept of massive MIMO where several hundred antennas are exploited is introduced by Marzetta in [58]. The effect of the antenna array's shape, including spherical, cylindrical, and rectangular, on the signal propagation, studied in [59–61]. The achievable rate in uplink (UL) and downlink (DL) of non-cooperative multi-cellular time-division duplexing (TDD) systems is investigated in [62]. In [62], channel estimation, pilot contamination, path loss, and antenna correlation for each link are considered to study the effect of multiple antennas in MIMO systems. A research on The effects of multiple antennas on energy efficiency are conducted in [63]. In [63], the effect of narrowing the signal beam to a region where the user locates is studied. The challenges and overview of the massive MIMO systems in 5G are discussed extensively in [64].

2.2 Multicast Beamforming

Multicast beamforming design has been studied by many in the literature for a single-user group [22, 65], multiple groups [25, 27, 30, 66–69], multi-cell networks [29, 70–72], relay networks [73], and mixed unicast-multicast transmissions [26, 74] assuming perfect channel state information (CSI). Two design formulations are commonly considered: the quality-of-service (QoS) problem, *i.e.*, minimizing the transmit power with minimum QoS guaranteed, and the max-min fair (MMF) problem, *i.e.*, maximizing the minimum signal-to-interference-plus-noise ratio (SINR) subject to the transmit power constraint.

Since the family of multicast beamforming problems is NP-hard, numerical methods or signal processing techniques were developed to find good suboptimal solutions. Semi-definite relaxation (SDR) has been a common method for traditional multi-antenna systems [22, 29, 30, 75]. Lately, successive convex approximation (SCA) has become more popular for its advantage in both performance and computational efficiency for a larger problem size [76, 77]. However, due to their high computational costs, these methods still need to be more practical for large-scale massive MIMO systems.

For massive MIMO systems, low-complexity suboptimal schemes were considered [70, 71, 77], and an ADMM-based fast algorithm was proposed [69]. The optimal multi-group multicast beamforming structure has been obtained recently in [68], which shows an inherent low-dimensional structure. Based on this structure, fast first-order algorithms were developed for different multicast beamforming design problems, or scenarios in massive MIMO systems proposed [23, 26, 27, 74].

For most of the multicast beamforming problems the implementation is to rely on the assumption of perfect CSI. However, in practical scenarios, the base station has access to an estimation of the CSI due to imperfect channel estimation [78, 79]. In the presence of the CSI error, the required SINR target for the QoS problems can not be guaranteed.

2.3 Robust Multicast Beamforming

Downlink multicast beamforming design naturally requires that the CSI is supplied at the transmitter. However, acquiring perfect CSI is a difficult task in wireless systems. The CSI typically is collected through feedback channels from the receiver side using a training sequence or using the uplink phase directly where it is applicable in time division duplex (TDD) systems [80]. The errors associated with the estimation process are unavoidable in both cases in practice [78]. Furthermore, the limited capacity of the feedback channel causes the quantization effect that it requires to be embedded in the estimated CSI [79]. In the case of using the uplink values, the fast-changing wireless environment also causes outdated CSI estimation. The effects above and the importance of the CSI in the beamforming design make the robustness of the downlink precoding techniques an important research matter.

Two types of error models are typically considered for robust beamforming designs: one approach is considering a probabilistic error model. This method assures a chance of outage condition fulfillment. The techniques that deal with the outage-based approach are based on replacing the probabilistic outage constraints with a deterministic conservative convex approximation [81–83]. This strategy seeks to provide good approximation with high precision to satisfy the outage constraints. The second approach, the subject of this work, is the worst-case scenario, which tackles the imperfect CSI by considering the channel error located inside a bounded region [84]. This approach aims to design a precoder that satisfies the worst case of the QoS constraints under the prescribed channel error model. The worst-case scenario is mainly based on an approximation of the worst-case SINRs. The main challenge in the robust design of multicast beamforming is the conservative approximation on the worst-case that leads to unnecessary power consumption. Recently, small cell deployments with high density and massive MIMO [85] are the promising technologies for the future generations of wireless systems as data services providers for large devices.

Existing works on robust multicast beamforming are limited to traditional multi-antenna

systems. Most studies consider a worst-case design assuming spherical bounded CSI error, where the worst-case SINR is considered for the QoS constraint for robust beamforming optimization [86–88]. The probabilistic design based on SINR outage was considered in [88, 89]. A typical approach in these existing works is to apply S-lemma [90] to convert the robust QoS constraint into an equivalent set of finite linear matrix inequality constraints and then apply SDR to obtain a feasible solution. As SDR has high computational complexity, some studies proposed to use a lower bound on the worst-case SINR and solve the related problem by SCA to reduce the complexity [50, 91]. However, such lower bounds are rather conservative and lead to substantial performance degradation as the channel estimation error increases. Robust multicast beamforming design has also been considered for cognitive radio networks [87], and cloud radio access networks [92, 93], where SDR-based methods were proposed.

Since channel covariance matrices typically evolve much slower than instantaneous CSI, using channel covariance for downlink beamforming design has been a common approach in the literature [94, 95]. In TDD massive MIMO systems, acquiring channel covariance can reduce uplink pilot overhead substantially and has been considered in [96–98]. Robust multicast beamforming using channel covariance matrices has been considered for traditional multi-antenna systems in [99, 100], where SDR-based methods were proposed.

The existing techniques to tackle the multicast problem, including SDR, SCA, and convex-concave procedure, impose high computational convexity. These techniques for large-scale problems become impractical. Therefore, low complexity with high-performance algorithms is vital for the growing scale of these wireless systems.

The importance of the CSI error and the growth in the scale of the wireless systems intrigue us to investigate the robust problem of multicast beamforming in large-scale systems.

2.4 Joint Antenna Selection and Robust Multicast Beamforming

Antenna selection in MIMO systems initially studied in [101, 102]. The two techniques, soft and hard antenna selection to exploit full capacity of available antennas and RF chains at the BS proposed initially in [103, 104]. Antenna selection in MIMO systems initially studied in [101, 102]. In multicast scenario, one of the leading works for joint beamforming design and antenna selection proposed in [105] where MMF problem for single multicast group has been solved. One method to solve joint beamforming design and antenna selection problems is an exhaustive search over all possible subsets of antennas, followed by a beamforming design problem with the subset of selected antennas. This approach for large-scale problems in massive MIMO is impractical due to its high computational complexity. Branch and bound (BAB) is a novel search principal for optimal antenna subset selection [106, 107]. In BAB, a search tree is used where traversing down the tree at each level, an antenna is removed based on the objective function gradient. However, the BAB algorithm improves the performance by eliminating some subset of options from exhaustive searches, in massive MIMO it is not an effective method due to the high computational complexity at each level of the search tree. Branch and bound (BAB) as one of the method to solve the problem of joint antenna selection and multicast beamforming is a novel search principal for optimal antenna subset selection that studied in [106, 107]. BAB is not a practical solution for large-scale problem due to high complexity.

Two standard techniques with lower computational complexity to incorporate antenna selection constraints in problem formulation for hard antenna selection are; 1) Cardinality constraint, where zero norms of the beamforming vector need to be equal to the number of selected antennas, 2) Binary antenna selection vector where the one zero entries of this binary vector are equivalent with selection and not the selection of antenna, respectively. One common method to solve the joint multicast beamforming and antenna selection, including cardinality constraint, is to augment the cardinality constraints to objective function as a penalty term to increase sparsity [108]. After that, a convex approximation of zero norms is

considered to overcome the highly non-convex zero norm as a penalty term at the objective. One challenge in this approach is to yield an accurate penalty constant for the cardinality penalty term. The focus of this work is on hard antenna selection technique. Therefore, selecting the best subset of the antenna affects the wireless network's performance. In [108], a joint multicast beamforming design and antenna selection with cardinality constraint for a single multicast scenario group has been considered. The authors in [108] has used mixed $\ell_{1,2}$ norm as a convex approximation of the cardinality penalty. Furthermore, the acquired non-convex problem due to non-convex SINR has transferred into a convex subproblem using SCA technique, followed by a fast ADMM algorithm to solve problem at each iteration of the SCA algorithm. The problem of joint multicast beamforming and antenna selection for multi group scenario including cardinality constraint is considered in [109]. In [109], $\ell_{1,\infty}$ norm is proposed as a convex approximation of cardinality penalty. Furthermore, SDR has been exploited to solve obtained problem that imposes high computational complexity. On the other hand, the approach with introduction of a binary selection vector for antenna selection convert the joint multicast beamforming and antenna selection problem into a mixed-integer non-convex problem. One approach to tackle this mixed-integer problem is a tight continuous relaxation. In this approach due to relaxation of the binary constraint, the selection vector is not a binary vector. Hence, the problem of joint multicast beamforming and antenna selection using binary selection with continuous relaxation commonly studies in two phases. The first phase is to determine the selected antennas, and second phase for beamforming design. The problem of joint multicast beamforming and antenna selection for multi-cell multi-group scenario has studied in [52]. The authors in [52] consider the problem of maximizing energy efficiency subject to QoS and power constraints. Both the binary antenna selection and cardinality constraint has studied in [52] where SCA has used to solve these problems. In [53] a joint antenna selection and hybrid beamforming for multi-group multicast scenario has been considered. The users in [53] equipped with power splitting (PS) device to split energy harvester (EH) and the information decoder (ID) for simultaneous wireless information and power transfer. The authors in [53] consider

a QoS problem with adding binary antenna selection vector constraint added. To tackle mixed-integer problem in [53] a feasible point pursuit-successive convex approximation (FPP-SCA) algorithm is proposed.

In all of the above studies on joint multicast beamforming and antenna selection, the assumption is the BS has perfect knowledge of CSI. However, in reality, the BS has only access to an estimation of the CSI. The problem of joint robust multicast beamforming and antenna selection for single group is studied in [107]. A leaning based BAB algorithm has exploited to solve QoS problem. However, the BAB algorithm is impractical for massive MIMO due to large number of antennas. Channel covariance matrix due to its slow evolution is a better candidate for beamforming design compare to instantaneous CSI. To the best of authors knowledge, the problem of joint robust multicast beamforming and and antenna selection for multi-group scenario considering channel covariance matrix has not been studied.

2.5 Alternating Direction Method of Multipliers (ADMM)

Due to the increasing scale of problems in massive MIMO systems with many antennas, it is essential to develop low-complexity solutions. ADMM is a great candidate to solve large-scale convex and non-convex problems. The ADMM method was developed first in the 1970s [110]. ADMM is closely related to other algorithms, such as dual decomposition, the method of the multiplier, and others. ADMM approach exploits the decomposability of dual ascent and excellent convergence properties of the method of multipliers.

The ADMM method decomposition-coordination procedure is exploited to decompose the large-scale problem into small-scale subproblems to be solved. Due to the high theoretical complexity of proving the convergence of the ADMM algorithm for a wide range of optimization problems, a theoretical convergence is challenging to develop. However, in practice, ADMM performs well in terms of convergence for a wide variety of optimization problems [111,112]. The convergence results for the convex problem have been well established. For the class of convex optimization problems, there are many convergence results

in the literature [113–116]. A common type of ADMM is a two-block ADMM algorithm where the optimization variables divide into two blocks to be solved iteratively. However, multi-block of variables in the ADMM algorithm is studied in some literature [117, 118]. The multi-block ADMM method for convex problems is investigated in [119–121]. Although, the studies on the convergence of the ADMM for the non-convex are limited due to the complexity of the non-convex problems [122–124]. Recently, the convergence of the three-blocks ADMM for non-convex problems has been studied in [125]. In [125] the authors prove the convergence of three-block ADMM for the class of non-convex problem with linear equality constraints having Kurdyka-Lojasiewicz property. Note that most of the mentioned literature in the proof of the convergence for ADMM is subject to predefined conditions and specific problem structures.

Recently, ADMM-based methods have been a known low-complexity method to apply to multicast beamforming design problems in massive MIMO systems due to large-scale antennas. ADMM-based algorithms have been studied in the literature for multi-group multicast beamforming design in massive MIMO systems in [27, 32], and for joint unicast and multicast beamforming design in large-scale massive MIMO systems in [26]. All the above studies are based on perfect CSI at the base station. However, ADMM-based approaches have not been applied in the robust design of multicast beamforming, , where channel uncertainty needs to be taken into account. Furthermore, the ADMM-based approach in the above studies has only been applied to convex problems. In this work, we develop a robust multicast beamforming design for the massive MIMO systems, and develop techniques to apply ADMM directly to the non-convex optimization with convergence guarantee.

Chapter 3

Low-Complexity ADMM-Based Algorithm for Robust Multi-group Multicast Beamforming in Large-Scale Systems

3.1 Problem Formulation

3.1.1 System Model

We consider a downlink multi-group multicasting scenario where a BS, equipped with M antennas, transmits messages to G multicast groups. Let $\mathcal{G} \triangleq \{1, ..., G\}$ denote the index set of the multicast groups and $\mathcal{K}_i \triangleq \{1, \ldots, K_i\}$ the index set of the single-antenna users in group i, for $i \in \mathcal{G}$. Each group receives a common message that is independent of the messages sent to other groups. Users in different multicast groups are disjoint. Let $\mathbf{w}_i \in \mathbb{C}^{M\times 1}$ denote the multicast beamforming vector for group i. The signal received at user k in group i is given by

$$y_{ik} = \mathbf{w}_i^H \mathbf{h}_{ik} s_i + \sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H \mathbf{h}_{ik} s_l + n_{ik},\ k \in \mathcal{K}_i,\ i \in \mathcal{G}. \tag{3.1}$$

where $\mathbf{h}_{ik} \in \mathbb{C}^{M\times 1}$ is the channel vector from the BS to user k in group i, s_i is the message intended for group i, n_{ik} is the receiver additive white Gaussian noise at user k in group i with zero mean and variance σ^2, and $\mathcal{G}_{-i} \triangleq \mathcal{G}\backslash\{i\}$. The total transmit power at the BS is $\sum_{i=1}^{G} \|\mathbf{w}_i\|^2$.

For both TDD and frequency-division duplex (FDD) systems, the instantaneous downlink channel state information (CSI) may not be available at the BS. Acquiring CSI can be challenging in massive MIMO systems. One approach is to obtain the channel covariance matrix from training symbols over time [96–98]. In this work, we consider that the BS uses the channel covariance matrix information for the beamforming design. Let $\bar{\mathbf{R}}_{ik} \triangleq \mathbb{E}\{\mathbf{h}_{ik}\mathbf{h}_{ik}^H\}$ denote the *true* channel covariance matrix for user k in group i. The QoS problem for the multicast beamforming design is formulated as transmit power minimization subject to the minimum SINR requirements:

$$\begin{aligned} \mathcal{P}_{\text{perf}}: \min_{\{\mathbf{w}_i\}_{i=1}^G} \quad & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \\ \text{s.t.} \quad & \frac{\mathbf{w}_i^H \bar{\mathbf{R}}_{ik} \mathbf{w}_i}{\sum_{l \in \mathcal{G}_{-i}} \mathbf{w}_l^H \bar{\mathbf{R}}_{ik} \mathbf{w}_l + \sigma^2} \geq \gamma_{ik}, \ k \in \mathcal{K}_i, \ i \in \mathcal{G} \end{aligned} \tag{3.2}$$

where γ_{ik} is the minimum SINR target for user k in group i.

3.1.2 Robust Formulation Based on Channel Covariance

In practice, only the estimated channel covariance matrices are available at the BS. We consider a robust beamforming design that incorporates the uncertainty of channel covariance matrices to ensure a robust performance in the presence of the estimation error. Let $\mathbf{R}_{ik}$ denote the *estimated* channel covariance matrix for user k in group i. The estimation error in the channel covariance matrix is modeled as $\bar{\mathbf{R}}_{ik} = \mathbf{R}_{ik} + \mathbf{E}_{ik}$, where $\mathbf{E}_{ik}$ is the corresponding error matrix. We assume that $\mathbf{E}_{ik}$ is bounded within a hyper-spherical region as

$$\|\mathbf{E}_{ik}\|_F \leq \epsilon_{ik}, \quad k \in \mathcal{K}_i, i \in \mathcal{G} \tag{3.3}$$

where ϵ_{ik} is the error bound. This spherical error model is a common model considered in the robust designs [86–88, 94, 95].

Let $\mathcal{B}(\epsilon_{ik})$ denote the set of all error matrices satisfying (3.3):

$$\mathcal{B}(\epsilon_{ik}) \triangleq \{\mathbf{E}_{ik} : \ \|\mathbf{E}_{ik}\|_F \leq \epsilon_{ik}\}. \tag{3.4}$$

Based on $\mathcal{P}_{\text{perf}}$ and the channel estimation error in (3.4), we formulate the robust multicast beamforming optimization for the QoS problem as

$$\begin{aligned}\mathcal{P}_{\text{rob}}: \min_{\{\mathbf{w}_i\}_{i=1}^G} \quad & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \\ \text{s.t.} \quad & \frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \geq \gamma_{ik}, \forall\, \mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik}), \\ & k\in\mathcal{K}_i, i\in\mathcal{G}. \end{aligned} \tag{3.5}$$

Note that there is an implicit condition for $\mathbf{E}_{ik}$ in the constraint (3.3) that for given estimated channel covariance $\mathbf{R}_{ik}$, $\mathbf{E}_{ik}$ should be such that the true covariance matrix $\bar{\mathbf{R}}_{ik} \succcurlyeq 0$. If the error bound ϵ_{ik} is small enough (relative to $\|\bar{\mathbf{R}}_{ik}\|_F$), then this condition is satisfied for any point in $\mathcal{B}(\epsilon_{ik})$ in (3.4), and $\mathcal{P}_{\text{rob}}$ is a highly accurate approximation to the exact robust beamforming problem. On the other hand, if ϵ_{ik} is large, then the set of possible error matrices $\mathbf{E}_{ik}$'s (*i.e.*, those satisfy the condition $\bar{\mathbf{R}}_{ik} \succcurlyeq 0$), denoted by $\mathcal{E}_{ik}$, is only a subset of $\mathcal{B}(\epsilon_{ik})$. This means that the feasible solution set for $\mathcal{P}_{\text{rob}}$ satisfying the constraint in (3.5) may be smaller than that for the problem under the error set $\mathcal{E}_{ik}$. It follows that the minimum transmit power obtained from $\mathcal{P}_{\text{rob}}$ is an upper bound to that of the exact robust beamforming optimization problem for $\mathbf{E}_{ik} \in \mathcal{E}_{ik}$, and the solution to $\mathcal{P}_{\text{rob}}$ may be suboptimal.

Note that problem $\mathcal{P}_{\text{perf}}$ is known to be a non-convex QCQP problem and is NP-hard. As a result, the robust optimization problem $\mathcal{P}_{\text{rob}}$ is even more challenging to solve. Moreover, as the future systems are expected to be massive MIMO with $M \gg 1$, it is critical to provide computationally efficient solution for practical implementation, In what follows, we first reformulate $\mathcal{P}_{\text{rob}}$ and then propose a fast algorithm to find a stationary solution to it.

3.2 Robust Multicast Beamforming Reformulation

It is straightforward to see that the SINR constraint in (3.5) is equivalent to letting the worst-case SINR for $\mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik})$ satisfy the minimum SINR target. Thus, $\mathcal{P}_{\text{rob}}$ is equivalent to the following problem

$$\mathcal{P}_0 : \min_{\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2$$

$$\text{s.t.} \min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \geq \gamma_{ik}, k\in\mathcal{K}_i, i\in\mathcal{G}. \tag{3.6}$$

In the existing works on robust beamforming design, a common approach to handle the worst-case SINR constraints, such as those in (3.6), is to derive a *conservative* lower bound on the worst-case SINR and use this bound in the constraint instead [49, 50]. However, there are two main caveats for this approach: 1) The lower bound can be overly conservative, as it is loose when error bound ϵ_{ik} is moderately large, leading to significant performance degradation. 2) Using this conservative lower bound may render the problem infeasible to solve, even though the original problem is feasible and a solution exists. Below, we first present a conservative lower bound on the worst-case SINR for problem $\mathcal{P}_0$, and formulate the problem using this lower bound. Next, to overcome the issues in the conservative bounding method mentioned above, we reformulate $\mathcal{P}_0$ by obtaining the equivalent constraint for worst-case SINR in (3.6).

3.2.1 The Approach via Conservative SINR Lower Bound

A common approach to lower bound the worst-case SINR at the left hand side (LHS) of (3.6) is to use the following inequality:

$$\min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \geq \frac{\min\limits_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\max_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2}. \tag{3.7}$$

The minimization and the maximization problems with respect to (w.r.t.) $\mathbf{E}_{ik}$ at the respective numerator and denominator in (3.7) both have a linear objective function with a quadratic constraint, and their solutions are known in literature [95]. The corresponding optimal objective values for these two problems are given by

$$\min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i = \mathbf{w}_i^H(\mathbf{R}_{ik}-\epsilon_{ik}\mathbf{I})\mathbf{w}_i, \tag{3.8}$$

$$\max_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l = \mathbf{w}_l^H(\mathbf{R}_{ik}+\epsilon_{ik}\mathbf{I})\mathbf{w}_l \tag{3.9}$$

where the optimal solutions to (3.8) and (3.9) are $\mathbf{E}_{ik} = -\epsilon_{ik}\frac{\mathbf{w}_i\mathbf{w}_i^H}{\|\mathbf{w}_i\|^2}$ and $\mathbf{E}_{ik} = \epsilon_{ik}\frac{\mathbf{w}_i\mathbf{w}_i^H}{\|\mathbf{w}_i\|^2}$, respectively. Therefore, replacing the worst-case SINR at the LHS of (3.6) by the lower bound in (3.7), we have the following optimization problem

$$\begin{aligned}\mathcal{P}_{\text{consv}}: \ &\min_{\{\mathbf{w}_i\}_{i=1}^G} \sum_{i=1}^G \|\mathbf{w}_i\|^2 \\ \text{s.t. } &\frac{\mathbf{w}_i^H(\mathbf{R}_{ik}-\epsilon_{ik}\mathbf{I})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\epsilon_{ik}\mathbf{I})\mathbf{w}_l+\sigma^2} \geq \ \gamma_{ik}, \ k\in\mathcal{G}_i, i\in\mathcal{G}.\end{aligned}$$

Problem $\mathcal{P}_{\text{consv}}$ is still a non-convex QCQP problem. Note from (3.7) that a feasible solution to $\mathcal{P}_{\text{consv}}$ is also feasible to $\mathcal{P}_0$. Thus, the minimum objective value of $\mathcal{P}_{\text{consv}}$ provides an upper bound on the minimum objective value of $\mathcal{P}_0$. We will use $\mathcal{P}_{\text{consv}}$ as an alternative robust design method for performance comparison in the simulation studies. In Section 3.4, for comparison purpose, we briefly describe a commonly used approach to solve $\mathcal{P}_{\text{consv}}$.

3.2.2 Reformulation of Worst-Case SINR Constraint

As discussed earlier, the SINR lower bound in (3.7) can be loose, and the solution provided by $\mathcal{P}_{\text{consv}}$ can be highly suboptimal for $\mathcal{P}_0$. Instead of using the lower bound, in what follows, we directly examine the worst-case SINR constraint in (3.6). For $k\in\mathcal{K}_i$, $i\in\mathcal{G}$, constraint (3.6) is equivalent to

$$\frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \geq \gamma_{ik}, \quad \forall\ \mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik}),$$

which is further equivalent to

$$J(\mathbf{E}_{ik};\{\mathbf{w}_i\}) \geq \sigma^2\gamma_{ik}, \quad \forall\ \mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik}) \tag{3.10}$$

where

$$J(\mathbf{E}_{ik};\{\mathbf{w}_i\}) \triangleq \mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l.$$

The constraint in (3.10) can be further expressed as

$$\min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} J(\mathbf{E}_{ik};\{\mathbf{w}_i\}) \geq \sigma^2\gamma_{ik}. \tag{3.11}$$

Thus, we transform $\mathcal{P}_0$ into the following equivalent problem:

$$\begin{aligned}\mathcal{P}_0': \ &\min_{\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \\ &\text{s.t.} \ \min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} J(\mathbf{E}_{ik};\{\mathbf{w}_i\}) \geq \sigma^2\gamma_{ik}, \ k\in\mathcal{K}_i, i\in\mathcal{G}. \end{aligned} \tag{3.12}$$

Focusing on the LHS of (3.12), we can express it as the following minimization problem w.r.t. $\mathbf{E}_{ik}$:

$$\min_{\mathbf{E}_{ik}} J(\mathbf{E}_{ik};\{\mathbf{w}_i\}) \quad \text{s.t.} \ \|\mathbf{E}_{ik}\|_F^2 \leq \epsilon_{ik}^2. \tag{3.13}$$

Problem (3.13) is a convex optimization problem, which is similar to the problem in (3.8). Thus, similar to [95], we solve it via the KKT conditions [90] and obtain the optimal solution to problem (3.13) as (see Appendix A.1 for the derivation details)

$$\mathbf{E}_{ik}^{\star} = -\epsilon_{ik}\frac{\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H}{\left\|\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H\right\|_F} \tag{3.14}$$

where $\|\mathbf{E}_{ik}^{\star}\|_F = \epsilon_{ik}$. Define $\mathbf{W} \triangleq [\mathbf{w}_1, \ \ldots, \ \mathbf{w}_G] \in \mathbb{C}^{M\times G}$, and also define $\mathbf{D}_{ik}$ as a $G\times G$ diagonal matrix with the ith diagonal entry being 1 and the rest being $-\gamma_{ik}$. Then, we can express $\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H = \mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H \triangleq \mathbf{A}_{ik}(\mathbf{W})$. Following this, the minimum objective value of problem (3.13) under $\mathbf{E}_{ik}^{\star}$ (*i.e.,* the worst case error matrix) can now be obtained as

$$\begin{aligned}&J(\mathbf{E}_{ik}^{\star};\{\mathbf{w}_i\}) \\ &= \mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik}^{\star})\mathbf{w}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik}^{\star})\mathbf{w}_l \\ &= \mathbf{w}_i^H\mathbf{R}_{ik}\mathbf{w}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H\mathbf{R}_{ik}\mathbf{w}_l - (\mathbf{w}_i^H\frac{\epsilon_{ik}\mathbf{A}_{ik}(\mathbf{W})}{\|\mathbf{A}_{ik}(\mathbf{W})\|_F}\mathbf{w}_i \quad - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H\frac{\epsilon_{ik}\mathbf{A}_{ik}(\mathbf{W})}{\|\mathbf{A}_{ik}(\mathbf{W})\|_F}\mathbf{w}_l) \\ &= \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \text{tr}\Big(\frac{\epsilon_{ik}\mathbf{A}_{ik}(\mathbf{W})}{\|\mathbf{A}_{ik}(\mathbf{W})\|_F}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\Big) \\ &= \text{tr}(\mathbf{R}_{ik}\mathbf{A}_{ik}(\mathbf{W})) - \epsilon_{ik}\|\mathbf{A}_{ik}(\mathbf{W})\|_F. \end{aligned} \tag{3.15}$$

Substituting the expression in (3.15) into the LHS of (3.12), we transform $\mathcal{P}_0'$ into the following equivalent problem

$$\mathcal{P}_1: \min_{\mathbf{W}} \ \text{tr}(\mathbf{W}^H\mathbf{W})$$
$$\text{s.t. } \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{3.16}$$

Thus, the original robust multicast beamforming problem $\mathcal{P}_0$ is now equivalently transformed into $\mathcal{P}_1$, which we will focus on solving. Note that $\mathcal{P}_1$ is a non-convex optimization problem due to the non-convex constraint in (3.16). A popular approach to solve a wide range of non-convex problem is to apply the SCA method, which forms a convex approximation to the problem to be solved iteratively. However, due to the Frobenius norm in the constraint in (3.16), it is difficult to find a suitable and tight convex approximation to apply the SCA method to this problem.[1] In the next section, we design an ADMM-based fast algorithm to directly solve $\mathcal{P}_1$.

3.3 ADMM-Based Fast Algorithm

ADMM is a robust and fast numerical method for solving large-scale problems. While the convergence of ADMM for convex problems has been proven and well-understood (see [112, 126]), the convergence results for the non-convex problems are rather limited. A recent work in [51] has established the convergence of ADMM applicable to a wide range of non-convex problems. Utilizing this result, we develop a non-convex ADMM-based algorithm to solve $\mathcal{P}_1$ directly, for which its convergence to a stationary point is guaranteed.

Introducing auxiliary matrices $\mathbf{V}_{ik} \in \mathbb{C}^{M\times G}$, $k \in \mathcal{K}_i$, $i \in \mathcal{G}$, we transfer $\mathcal{P}_1$ to the following equivalent problem:

$$\mathcal{P}_2: \min_{\{\mathbf{V}_{ik}\},\mathbf{W}} \ \text{tr}\big(\mathbf{W}^H\mathbf{W}\big)$$
$$\text{s.t.} \quad \mathfrak{Re}\big\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}^H)\big\} - \epsilon_{ik}\|\mathbf{V}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G} \tag{3.17}$$
$$\mathbf{V}_{ik} = \mathbf{W}\mathbf{D}_{ik}, \quad k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{3.18}$$

[1] It is possible to find convex approximations for (3.16), but they usually are not tight, leading to inferior solutions and performance.

Let $\mathcal{C}$ denote the feasible set of $\big(\{\mathbf{V}_{ik}\}, \mathbf{W}\big)$ satisfying the constraint in (3.17). We define the indicator function for $\mathcal{C}$ as

$$\mathbb{I}_{\mathcal{C}}\big(\{\mathbf{V}_{ik}\}, \mathbf{W}\big) = \begin{cases} 0, & \text{if } (\mathbf{W}, \{\mathbf{V}_{ik}\}) \in \mathcal{C}, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.19}$$

Using (3.19), we transfer the constraint in (3.17) into the objective function in $\mathcal{P}_2$ and arrive at the following equivalent problem:

$$\begin{aligned} \mathcal{P}_2' : \min_{\mathbf{W}, \{\mathbf{V}_{ik}\}} \ & \mathrm{tr}\big(\mathbf{W}^H\mathbf{W}\big) + \mathbb{I}_{\mathcal{C}}\big(\{\mathbf{V}_{ik}\}, \mathbf{W}\big) \\ \text{s.t. } & \mathbf{V}_{ik} = \mathbf{W}\mathbf{D}_{ik}, \quad k \in \mathcal{K}_i, i \in \mathcal{G}. \end{aligned} \tag{3.20}$$

Since $\mathcal{P}_2'$ involves complex-valued optimization variables, its corresponding augmented Lagrangian is given by [127]

$$\begin{aligned} \mathcal{L}_\rho\big(\{\mathbf{V}_{ik}\}, \mathbf{W}, \{\mathbf{Z}_{ik}\}\big) = & \ \mathrm{tr}(\mathbf{W}^H\mathbf{W}) + \mathbb{I}_{\mathcal{C}}\big(\{\mathbf{V}_{ik}\}, \mathbf{W}\big) \\ & + \rho \sum_{i=1}^{G} \sum_{k=1}^{K_i} \|\mathbf{V}_{ik} - \mathbf{W}\mathbf{D}_{ik} + \mathbf{Z}_{ik}\|_F^2 \end{aligned} \tag{3.21}$$

where $\mathbf{Z}_{ik} \in \mathbb{C}^{M \times G}$ is the dual variable associated with the constraint in (3.20), and ρ is the penalty parameter.

Note that by introducing $\mathbf{V}_{ik}$ and its associated equality constraint in (3.18), we can now carry out the minimization of $\mathcal{L}_\rho(\{\mathbf{V}_{ik}\}, \mathbf{W}, \{\mathbf{Z}_{ik}\})$ w.r.t. $\{\mathbf{V}_{ik}\}$ and $\mathbf{W}$ separately. Our proposed ADMM-based algorithm for $\mathcal{P}_1$ is summarized in Algorithm 1. It consists of three updating steps. In Steps 1 and 2, $\mathcal{L}_\rho(\{\mathbf{V}_{ik}\}, \mathbf{W}, \{\mathbf{Z}_{ik}\})$ is minimized w.r.t. $\{\mathbf{V}_{ik}\}$ and $\mathbf{W}$, respectively, to obtain the updates. For each of these subproblems, we are able to derive a semi-closed-form solution. We present the details of the solution to each subproblem in Sections 3.3.1 and 3.3.2.

3.3.1 Updating $\{\mathbf{V}_{ik}\}$

From $\mathcal{L}_\rho\big(\{\mathbf{V}_{ik}\}, \mathbf{W}, \{\mathbf{Z}_{ik}\}\big)$ in (3.21), given $(\mathbf{W}^j, \{\mathbf{Z}_{ik}^j\})$ in iteration j, updating $\{\mathbf{V}_{ik}\}$ in (3.22) is equivalent to

$$\{\mathbf{V}_{ik}^{j+1}\} = \arg \min_{\{\mathbf{V}_{ik}\}} \mathbb{I}_{\mathcal{C}}(\mathbf{V}_{ik}, \mathbf{W}^j)$$

Algorithm 1 The ADMM-Based Algorithm for $\mathcal{P}_1$

Initialization: Set ρ; Set initial $\mathbf{W}^0$, $\mathbf{Z}_{ik}^0 = \mathbf{0}$; Set $j = 0$.

repeat

1) Update the auxiliary matrices $\{\mathbf{V}_{ik}^{j+1}\}$:

$$\{\mathbf{V}_{ik}^{j+1}\} = \arg\min_{\{\mathbf{V}_{ik}\}} \mathcal{L}_\rho\big(\{\mathbf{V}_{ik}\}, \mathbf{W}^j, \{\mathbf{Z}_{ik}^j\}\big). \tag{3.22}$$

2) Update beamforming matrix $\mathbf{W}^{j+1}$:

$$\mathbf{W}^{j+1} = \arg\min_{\mathbf{W}} \mathcal{L}_\rho\big(\{\mathbf{V}_{ik}^{j+1}\}, \mathbf{W}, \{\mathbf{Z}_{ik}^j\}\big). \tag{3.23}$$

3) Update dual variables $\{\mathbf{Z}_{ik}^{j+1}\}$:

$$\mathbf{Z}_{ik}^{j+1} = \mathbf{Z}_{ik}^j + \mathbf{V}_{ik}^{j+1} - \mathbf{W}^{j+1}\mathbf{D}_{ik},\ k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{3.24}$$

4) Set $j \leftarrow j + 1$.

until convergence

$$+ \rho \sum_{i=1}^{G} \sum_{k=1}^{K_i} \|\mathbf{V}_{ik} - \mathbf{W}^j\mathbf{D}_{ik} + \mathbf{Z}_{ik}^j\|^2, \tag{3.25}$$

which is further equivalent to the following problem:

$$\begin{aligned} \min_{\{\mathbf{V}_{ik}\}} \quad & \sum_{i=1}^{G} \sum_{k=1}^{K_i} \|\mathbf{V}_{ik} - \mathbf{W}^j\mathbf{D}_{ik} + \mathbf{Z}_{ik}^j\|_F^2 \\ \text{s.t.} \quad & \mathfrak{Re}\big\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}^H)\big\} - \epsilon_{ik}\|\mathbf{V}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. \end{aligned} \tag{3.26}$$

The above problem can be decomposed into $\sum_{i=1}^{G} K_i$ separate subproblems, one for each user $k \in \mathcal{K}_i$ in group $i \in \mathcal{G}$ as

$$\begin{aligned} \mathcal{P}_{\mathbf{V}}: \ \min_{\mathbf{V}_{ik}} \ & \|\mathbf{V}_{ik} - \mathbf{W}^j\mathbf{D}_{ik} + \mathbf{Z}_{ik}^j\|_F^2 \\ & \text{s.t. } \mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}^{jH})\} - \epsilon_{ik}\|\mathbf{V}_{ik}\mathbf{W}^{jH}\|_F \geq \sigma^2\gamma_{ik}. \end{aligned} \tag{3.27}$$

Note that $\|\cdot\|_F$ is a convex function and the expression in $\|\cdot\|_F$ is affine w.r.t. $\mathbf{V}_{ik}$. Thus, $\mathcal{P}_{\mathbf{V}}$ is a convex optimization problem and can be solved using IPM [90] implemented in standard convex solvers. However, the IPM is a second-order algorithm, whose computational complexity is high for large-scale problems. To reduce the computational complexity in solving $\mathcal{P}_{\mathbf{V}}$, we propose our method below to derive a semi-closed-form solution.

Semi-closed-form solution to $\mathcal{P}_{\mathbf{V}}$

We first vectorize each matrix term in $\mathcal{P}_{\mathbf{V}}$ and transform the problem into the following equivalent optimization problem in the vector form:

$$\begin{aligned}\mathcal{P}_{\mathbf{v}}: \min_{\mathbf{v}_{ik}} \ & (\mathbf{v}_{ik}-\mathbf{b}_{ik})^H(\mathbf{v}_{ik}-\mathbf{b}_{ik}) \\ \text{s.t.} \ & \|\mathbf{W}_{\mathrm{c},ik}^{j\frac{1}{2}}\mathbf{v}_{ik}\| \le \mathfrak{Re}\{\mathbf{r}_{ik}^H\mathbf{v}_{ik}\}-1\end{aligned}$$

where $\mathbf{v}_{ik} \triangleq \text{vec}(\mathbf{V}_{ik})$, $\mathbf{b}_{ik} \triangleq \text{vec}(\mathbf{W}^j\mathbf{D}_{ik}-\mathbf{Z}_{ik}^j)$, $\mathbf{r}_{ik} \triangleq \frac{1}{\sigma^2\gamma_{ik}}\text{vec}(\mathbf{R}_{ik}^H\mathbf{W}^j)$, and $\mathbf{W}_{\mathrm{c},ik}^j \triangleq \frac{\epsilon_{ik}^2}{(\sigma^2\gamma_{ik})^2}\mathbf{W}^{jT}\mathbf{W}^{j*}\otimes\mathbf{I}_M$. To derive the vectorized version of the second term at the LHS of the constraint in (3.27), we have used the following fact: $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T\otimes\mathbf{A})\text{vec}(\mathbf{B})$.

Since $\mathcal{P}_{\mathbf{v}}$ is for each $k \in \mathcal{K}_i$ and $i \in \mathcal{G}$, for notation simplicity, in the rest of this subsection, we remove subscript ik and superscript j for all variables in $\mathcal{P}_{\mathbf{v}}$ to derive the solution. The solution to $\mathcal{P}_{\mathbf{v}}$ may be one of the two cases discussed below.

Case 1): The constraint is inactive at the optimality of $\mathcal{P}_{\mathbf{v}}$: We have the optimal $\mathbf{v}^{\mathrm{o}} = \mathbf{b}$.

Case 2): The constraint is satisfied with equality at the optimality of $\mathcal{P}_{\mathbf{v}}$: In this case, if we impose $\mathfrak{Re}\{\mathbf{r}^H\mathbf{v}\}-1 \ge 0$, then solving $\mathcal{P}_{\mathbf{v}}$ is equivalent to solve the following problem:[2]

$$\begin{aligned}\mathcal{P}'_{\mathbf{v}}: \min_{\mathbf{v}} \ & (\mathbf{v}-\mathbf{b})^H(\mathbf{v}-\mathbf{b}) \\ \text{s.t.} \ & \mathbf{v}^H\mathbf{W}_{\mathrm{c}}\mathbf{v} = (\mathfrak{Re}\{\mathbf{r}^H\mathbf{v}\}-1)^2.\end{aligned}$$

The Lagrangian of $\mathcal{P}'_{\mathbf{v}}$ is given by

$$\mathcal{L}(\mathbf{v},\lambda) = (\mathbf{v}-\mathbf{b})^H(\mathbf{v}-\mathbf{b}) + \lambda(\mathbf{v}^H\mathbf{W}_{\mathrm{c}}\mathbf{v} - (\mathfrak{Re}\{\mathbf{r}^H\mathbf{v}\}-1)^2)$$

where λ is the Lagrange multiplier associated with the constraint in $\mathcal{P}'_{\mathbf{v}}$. Setting the gradient of $\mathcal{L}(\mathbf{v},\lambda)$ w.r.t. $\mathbf{v}^*$ to zero, we obtain

$$\nabla_{\mathbf{v}^*}\mathcal{L}(\mathbf{v},\lambda) = \mathbf{v}-\mathbf{b} + \lambda(\mathbf{W}_{\mathrm{c}}\mathbf{v} - (\mathfrak{Re}\{\mathbf{r}^H\mathbf{v}\}-1)\mathbf{r}) = 0, \tag{3.28}$$

and the optimal solution $\mathbf{v}^{\mathrm{o}}$ to $\mathcal{P}'_{\mathbf{v}}$ is obtained as

$$\mathbf{v}^{\mathrm{o}} = (\mathbf{I}+\lambda\mathbf{W}_{\mathrm{c}})^{-1}(\mathbf{b}+\lambda(\mathfrak{Re}\{z\}-1)\mathbf{r}) \tag{3.29}$$

[2]The constraint $\mathfrak{Re}\{\mathbf{r}^H\mathbf{v}\}-1 \ge 0$ is not explicitly included in $\mathcal{P}'_{\mathbf{v}}$. We will first solve $\mathcal{P}'_{\mathbf{v}}$ and then use this constraint to determine the solution.

where $z \triangleq \mathbf{r}^H \mathbf{v}^{\mathrm{o}}$. The optimal $\mathbf{v}^{\mathrm{o}}$ in (3.29) depends on $\mathfrak{Re}\{z\}$. To extract $\mathfrak{Re}\{z\}$ from (3.29), we multiply $\mathbf{r}^H$ and then take $\mathfrak{Re}\{\cdot\}$ on both sides of (3.29) to arrive at

$$\mathfrak{Re}\{z\} = \mathfrak{Re}\{\mathbf{r}^H(\mathbf{I} + \lambda \mathbf{W}_{\mathrm{c}})^{-1}(\mathbf{b} + \lambda(\mathfrak{Re}\{z\} - 1)\mathbf{r})\}. \tag{3.30}$$

Thus, we obtain $\mathfrak{Re}\{z\}$ as

$$\mathfrak{Re}\{z\} = \frac{\mathfrak{Re}\{\mathbf{r}^H \mathbf{C}(\lambda)\mathbf{b}\} - \lambda \mathbf{r}^H \mathbf{C}(\lambda)\mathbf{r}}{1 - \lambda \mathbf{r}^H \mathbf{C}(\lambda)\mathbf{r}} \tag{3.31}$$

where $\mathbf{C}(\lambda) \triangleq (\mathbf{I} + \lambda \mathbf{W}_{\mathrm{c}})^{-1}$.

To determine $\mathbf{v}^{\mathrm{o}}$ in (3.29), we now only need to find λ. Substituting the expression of $\mathbf{v}^{\mathrm{o}}$ in (3.29) into the equality constraint in $\mathcal{P}'_{\mathbf{v}}$, we have the following equation

$$\begin{aligned}\big(\mathfrak{Re}\{z\} - 1\big)^2 \big(1 - \lambda^2 \mathbf{r}^H \mathbf{P}(\lambda)\mathbf{r}\big) \\ - 2\lambda \big(\mathfrak{Re}\{z\} - 1\big)\mathfrak{Re}\{\mathbf{r}^H \mathbf{P}(\lambda)\mathbf{b}\}\big) - \mathbf{b}^H \mathbf{P}(\lambda)\mathbf{b} = 0\end{aligned} \tag{3.32}$$

where $\mathbf{P}(\lambda) \triangleq \mathbf{C}(\lambda)\mathbf{W}_{\mathrm{c}}\mathbf{C}(\lambda)$. Thus, we can obtain λ as the root of the equation in (3.32). Substituting $\mathfrak{Re}\{z\}$ in (3.31) into (3.32) and after some manipulation of the denominator terms, we can show that solving (3.32) is equivalent to solving the following equation (See Appendix A.2 for more details):

$$\begin{aligned}\big(\mathfrak{Re}\{\mathbf{r}^H \mathbf{C}(\lambda)\mathbf{b}\} - 1\big)^2 \big(1 - \lambda^2 \mathbf{r}^H \mathbf{P}(\lambda)\mathbf{r}\big) \\ - 2\lambda \big(\mathfrak{Re}\{\mathbf{r}^H \mathbf{C}(\lambda)\mathbf{b}\} - 1\big) \big(1 - \lambda \mathbf{r}^H \mathbf{C}(\lambda)\mathbf{r}\big) \mathfrak{Re}\{\mathbf{r}^H \mathbf{P}(\lambda)\mathbf{b}\} \\ - \big(1 - \lambda \mathbf{r}^H \mathbf{C}(\lambda)\mathbf{r}\big)^2 \mathbf{b}^H \mathbf{P}(\lambda)\mathbf{b} = 0.\end{aligned} \tag{3.33}$$

We now further simplify (3.33) by examining the structure of $\mathbf{C}(\lambda)$. Recall the expression of $\mathbf{W}_{\mathrm{c}} = \frac{\epsilon^2}{(\sigma^2\gamma)^2}\mathbf{W}^T\mathbf{W}^* \otimes \mathbf{I}_M$. We consider eigenvalue decomposition $\mathbf{W}^T\mathbf{W}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$, where $\mathbf{\Sigma} \triangleq \mathrm{diag}(\bar{\sigma}_1^2, \cdots, \bar{\sigma}_G^2)$ with $\bar{\sigma}_j^2$ being the jth eigenvalue of $\mathbf{W}^T\mathbf{W}^*$ and $\mathbf{U}$ is the unitary matrix of the corresponding eigenvectors. Then, we have $\mathbf{W}_{\mathrm{c}} = \frac{\epsilon^2}{(\sigma^2\gamma)^2}\mathbf{U}\mathbf{\Sigma}\mathbf{U}^H \otimes \mathbf{I}_M$. Using this structure of $\mathbf{W}_{\mathrm{c}}$, we simplify the matrix inversion in $\mathbf{C}(\lambda)$ through eigenvalue decomposition and apply it to $\mathbf{P}(\lambda)$. This leads to simplify the terms in (3.33) as

$$\mathfrak{Re}\{\mathbf{r}^H \mathbf{C}(\lambda)\mathbf{b}\} = \sum_{j=1}^{G} \frac{\mathfrak{Re}\{\bar{\mathbf{r}}_j^H \bar{\mathbf{b}}_j\}}{1 + \lambda\bar{\sigma}_j^2}, \ \mathbf{r}^H \mathbf{C}(\lambda)\mathbf{r} = \sum_{j=1}^{G} \frac{\|\bar{\mathbf{r}}_j\|^2}{1 + \lambda\bar{\sigma}_j^2}$$

$$\mathbf{r}^H\mathbf{P}(\lambda)\mathbf{r} = \sum_{j=1}^{G} \frac{\bar{\sigma}_j^2 \|\bar{\mathbf{r}}_j\|^2}{(1+\lambda\bar{\sigma}_j^2)^2},\ \mathbf{b}^H\mathbf{P}(\lambda)\mathbf{b} = \sum_{j=1}^{G} \frac{\bar{\sigma}_j^2 \|\bar{\mathbf{b}}_j\|^2}{(1+\lambda\bar{\sigma}_j^2)^2},$$
$$\mathfrak{Re}\{\mathbf{r}^H\mathbf{P}(\lambda)\mathbf{b}\} = \sum_{j=1}^{G} \frac{\bar{\sigma}_j^2 \mathfrak{Re}\{\bar{\mathbf{r}}_j^H\bar{\mathbf{b}}_j\}}{(1+\lambda\bar{\sigma}_j^2)^2}$$

where $\bar{\mathbf{b}} = [\bar{\mathbf{b}}_1^H, \ldots, \bar{\mathbf{b}}_G^H]^H \triangleq (\mathbf{U}\otimes\mathbf{I}_M)^H\mathbf{b}$ with $\bar{\mathbf{b}}_j \in \mathbb{C}^{M\times 1}$, $j = 1, \ldots, G$, and $\bar{\mathbf{r}} = [\bar{\mathbf{r}}_1^H, \ldots, \bar{\mathbf{r}}_G^H]^H \triangleq (\mathbf{U}\otimes\mathbf{I}_M)^H\mathbf{r}$ with $\bar{\mathbf{r}}_j \in \mathbb{C}^{M\times 1}$, $j = 1, \ldots, G$.[3] Therefore, (3.33) can be simplified as in (3.34).

$$\Big(\sum_{j=1}^{G} \frac{\mathfrak{Re}\{\bar{\mathbf{r}}_j^H\bar{\mathbf{b}}_j\}}{1+\lambda\bar{\sigma}_j^2} - 1\Big)^2 \Big(1 - \lambda^2 \sum_{j=1}^{G} \frac{\bar{\sigma}_j^2\|\bar{\mathbf{r}}_j\|^2}{(1+\lambda\bar{\sigma}_j^2)^2}\Big) - \Big(1 - \lambda\sum_{j=1}^{G} \frac{\|\bar{\mathbf{r}}_j\|^2}{1+\lambda\bar{\sigma}_j^2}\Big)^2 \sum_{j=1}^{G} \frac{\bar{\sigma}_j^2\|\bar{\mathbf{b}}_j\|^2}{(1+\lambda\bar{\sigma}_j^2)^2} -$$
$$2\lambda\Big(\sum_{j=1}^{G} \frac{\mathfrak{Re}\{\bar{\mathbf{r}}_j^H\bar{\mathbf{b}}_j\}}{1+\lambda\bar{\sigma}_j^2} - 1\Big) \cdot \Big(1 - \lambda\sum_{j=1}^{G} \frac{\|\bar{\mathbf{r}}_j\|^2}{1+\lambda\bar{\sigma}_j^2}\Big) \sum_{j=1}^{G} \frac{\bar{\sigma}_j^2 \mathfrak{Re}\{\bar{\mathbf{r}}_j^H\bar{\mathbf{b}}_j\}}{(1+\lambda\bar{\sigma}_j^2)^2} = 0. \tag{3.34}$$

The resulting function at the LHS of (3.34) is a smooth polynomial function of λ, whose roots can be easily found. Recall that we impose the constraint $\mathfrak{Re}\{z\} - 1 \geq 0$ to $\mathcal{P}'_{\mathbf{v}}$. Thus, we select the root of (3.34) that results in the minimum objective value of $\mathcal{P}'_{\mathbf{v}}$ while satisfying the condition $\mathfrak{Re}\{z\} - 1 \geq 0$ as λ.

To summarize, for each $k \in \mathcal{K}_i, i \in \mathcal{G}$, we obtain the optimal $\mathbf{V}_{ik}^{\mathrm{o}}$ to $\mathcal{P}_{\mathbf{V}}$ as follows: If $\mathbf{b}_{ik}$ in Case 1 is feasible, it is the optimal solution; otherwise, the semi-closed-form solution in (3.29) of Case 2 is the optimal solution. The update $\mathbf{V}_{ik}^{j+1}$ in (3.25) is given by this optimal solution.

3.3.2 Updating W

After we obtain $\{\mathbf{V}_{ik}^{j+1}\}$, the update of $\mathbf{W}$ in (3.23) is equivalent to solving the following problem:

$$\mathcal{P}_{\mathbf{W}}:\ \min_{\mathbf{W}} \mathrm{tr}\big(\mathbf{W}^H\mathbf{W}\big) + \sum_{i=1}^{G}\sum_{k=1}^{K_i} \rho\|\mathbf{V}_{ik}^{j+1} - \mathbf{W}\mathbf{D}_{ik} + \mathbf{Z}_{ik}^{j}\|_F^2$$
$$\text{s.t. } \mathfrak{Re}\big\{\mathrm{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}^{j+1}\mathbf{W}^H)\big\} - \epsilon_{ik}\|\mathbf{V}_{ik}^{j+1}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. \tag{3.35}$$

[3] Based on the definitions of $\mathbf{b}$, and $\mathbf{r}$, we have $(\mathbf{U}\otimes\mathbf{I}_M)^H\mathbf{b} = \mathrm{vec}((\mathbf{W}\mathbf{D} - \mathbf{Z})\mathbf{U}^*)$, and $(\mathbf{U}\otimes\mathbf{I}_M)^H\mathbf{r} = \frac{1}{\sigma^2\gamma}\mathrm{vec}(\mathbf{R}^H\mathbf{W}\mathbf{U}^*)$. These relations should be used for efficient computation.

Since $\|\cdot\|_F$ is convex, both objective and constraint functions are convex. Thus, $\mathcal{P}_{\mathbf{W}}$ is a convex optimization problem. However, the computational complexity of solving $\mathcal{P}_{\mathbf{W}}$ using standard convex solvers is very high due to the Frobenius norm in the constraint, and this would defeat the purpose of applying ADMM to the overall problem for fast computation. To overcome this issue, we propose using the consensus ADMM technique [112] to solve $\mathcal{P}_{\mathbf{W}}$ efficiently.

Introducing auxiliary matrices $\mathbf{W}_{ik} \in \mathbb{C}^{M\times G}$, $k \in \mathcal{K}, i \in \mathcal{G}$, we equivalently reformulate $\mathcal{P}_{\mathbf{W}}$ as

$$\begin{aligned}\mathcal{P}'_{\mathbf{W}}: \min_{\mathbf{W},\{\mathbf{W}_{ik}\}} \ & \mathbb{I}_{\mathcal{D}}\big(\{\mathbf{W}_{ik}\}\big) + \mathrm{tr}\big(\mathbf{W}^H\mathbf{W}\big) + \sum_{i=1}^{G}\sum_{k=1}^{K_i} \rho\|\mathbf{V}_{ik}^{j+1} - \mathbf{W}_{ik}\mathbf{D}_{ik} + \mathbf{Z}_{ik}^{j}\|_F^2 \\ \text{s.t.} \ & \mathbf{W}_{ik} = \mathbf{W}, \ k \in K_i, i \in \mathcal{G} \end{aligned} \tag{3.36}$$

where

$$\mathbb{I}_{\mathcal{D}}(\{\mathbf{W}_{ik}\}) = \begin{cases} 0 & \text{if } \mathbf{W}_{ik} \in \ \mathcal{D}, \forall k \in \mathcal{K}_i, i \in \mathcal{G} \\ \infty & \text{otherwise} \end{cases}$$

is the indicator function with $\mathcal{D}$ being the feasible set satisfying all constraints in (3.35). The augmented Lagrangian of $\mathcal{P}'_{\mathbf{W}}$ is given by

$$\begin{aligned}&\mathcal{L}_{\mu}\big(\{\mathbf{W}_{ik}\}, \mathbf{W}, \{\mathbf{Y}_{ik}\}\big) = \mathrm{tr}\big(\mathbf{W}^H\mathbf{W}\big) + \mathbb{I}_{\mathcal{D}}\big(\{\mathbf{W}_{ik}\}\big) \\ &+ \rho\sum_{i=1}^{G}\sum_{k=1}^{K_i}\|\mathbf{V}_{ik}^{j+1} - \mathbf{W}_{ik}\mathbf{D}_{ik} + \mathbf{Z}_{ik}^{j}\|_F^2 \\ &+ \mu\sum_{i=1}^{G}\sum_{k=1}^{K_i}\|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{Y}_{ik}\|_F^2 \end{aligned} \tag{3.37}$$

where $\mathbf{Y}_{ik} \in \mathbb{C}^{M\times G}$ is the dual variable associated with the constraint in (3.36), and μ is the penalty parameter. Using the consensus ADMM technique, we minimize $\mathcal{L}_{\mu}(\{\mathbf{W}_{ik}\}, \mathbf{W}, \{\mathbf{Y}_{ik}\})$ by separating it into two subproblems w.r.t. $\{\mathbf{W}_{ik}\}$ and $\mathbf{W}$, respectively, and solving them alternatingly. This consensus-ADMM-based method is summarized in Algorithm 2.[4] In the following subsections, we describe the first two updating steps in details.

[4]We use superscript (l) (*e.g.,* $\mathbf{W}^{(l)}$) as the iteration index in Algorithm 2 (inner-layer iteration) for $\mathcal{P}_{\mathbf{W}}$ to obtain the update $\mathbf{W}^{j+1}$, while superscript j (*e.g.,* $\mathbf{W}^{j}$) is used as the iteration index in Algorithm 1 (outer-layer iteration).

Algorithm 2 Consensus-ADMM-Based Algorithm for $\mathcal{P}_{\mathbf{W}}$

Initialization: Set μ; Set $\mathbf{W}^{(0)} = \mathbf{W}^j$, $\mathbf{Y}_{ik}^{(0)} = \mathbf{0}$; Set $l = 0$.

repeat

1. Update $\{\mathbf{W}_{ik}^{(l+1)}\}$

$$\{\mathbf{W}_{ik}^{(l+1)}\} = \arg\min_{\{\mathbf{W}_{ik}\}} \mathcal{L}_\mu\big(\{\mathbf{W}_{ik}\}, \mathbf{W}^{(l)}, \{\mathbf{Y}_{ik}^{(l)}\}\big). \tag{3.38}$$

2. Update $\mathbf{W}^{(l+1)}$

$$\mathbf{W}^{(l+1)} = \arg\min_{\mathbf{W}} \mathcal{L}_\mu\big(\{\mathbf{W}_{ik}^{(l+1)}\}, \mathbf{W}, \{\mathbf{Y}_{ik}^{(l)}\}\big). \tag{3.39}$$

3. Update dual variables $\{\mathbf{Y}_{ik}^{(l+1)}\}$

$$\mathbf{Y}_{ik}^{(l+1)} = \mathbf{Y}_{ik}^{(l)} + \mathbf{W}^{(l+1)} - \mathbf{W}_{ik}^{(l+1)}. \tag{3.40}$$

4. $l \leftarrow l + 1$.

until convergence.

Updating $\{\mathbf{W}_{ik}\}$

Similar to problem (3.26), given $\mathbf{W}^{(l)}$ and $\{\mathbf{Y}_{ik}^{(l)}\}$ at iteration l, the update of $\{\mathbf{W}_{ik}\}$ in (3.38) can be decomposed into $\sum_{i=1}^{G} K_i$ independent subproblems, one for each user $k \in \mathcal{K}_i$ in group $i \in \mathcal{G}$ as

$$\begin{aligned} &\min_{\mathbf{W}_{ik}} \ \rho\|\mathbf{W}_{ik}\mathbf{D}_{ik} - \overline{\mathbf{Z}}_{ik}^j\|_F^2 + \mu\|\mathbf{W}_{ik} - \overline{\mathbf{Y}}_{ik}^{(l)}\|_F^2 \\ &\text{s.t. } \mathfrak{Re}\{\text{tr}(\overline{\mathbf{R}}_{ik}^j \mathbf{W}_{ik}^H)\} - \epsilon_{ik}\|\mathbf{V}_{ik}^j \mathbf{W}_{ik}^H\|_F \geq \sigma^2 \gamma_{ik} \end{aligned} \tag{3.41}$$

where $\overline{\mathbf{Z}}_{ik}^j \triangleq \mathbf{V}_{ik}^{j+1} + \mathbf{Z}_{ik}^j$, $\overline{\mathbf{Y}}_{ik}^{(l)} \triangleq \mathbf{W}^{(l)} + \mathbf{Y}_{ik}^{(l)}$, and $\overline{\mathbf{R}}_{ik}^j \triangleq \mathbf{R}_{ik}\mathbf{V}_{ik}^j$. The update $\mathbf{W}_{ik}^{(l+1)}$ is the solution to problem (3.41).

Note that problem (3.41) is convex, and its structure is similar to $\mathcal{P}_{\mathbf{V}}$. Thus, we can apply a technique similar to the one used to solve $\mathcal{P}_{\mathbf{V}}$ and derive a semi-closed-form solution to problem (3.41). Specifically, we transform problem (3.41) into its vector form by replacing each matrix term with its respective vector representation. The equivalent problem after vectorization is given by

$$\begin{aligned} \mathcal{P}_{\mathbf{w}} : \min_{\mathbf{w}_{ik}} \ &\rho\|\overline{\mathbf{D}}_{ik}\mathbf{w}_{ik} - \bar{\mathbf{z}}_{ik}^j\|^2 + \mu\|\mathbf{w}_{ik} - \bar{\mathbf{y}}_{ik}^{(l)}\|^2 \\ \text{s.t. } &\mathfrak{Re}\{\bar{\mathbf{r}}_{ik}^{jH}\mathbf{w}_{ik}\} - \|\mathbf{V}_{\text{c},ik}^{j\frac{1}{2}}\mathbf{w}_{ik}\| \geq 1. \end{aligned} \tag{3.42}$$

where $\mathbf{w}_{ik} \triangleq \text{vec}(\mathbf{W}_{ik})$, $\bar{\mathbf{z}}_{ik}^{j} \triangleq \text{vec}(\overline{\mathbf{Z}}_{ik}^{j})$, $\bar{\mathbf{y}}_{ik}^{(l)} \triangleq \text{vec}(\overline{\mathbf{Y}}_{ik}^{(l)})$, $\overline{\mathbf{D}}_{ik} \triangleq \mathbf{D}_{ik} \otimes \mathbf{I}_M$, $\bar{\mathbf{r}}_{ik}^{j} \triangleq \frac{1}{\sigma^2\gamma_{ik}}\text{vec}(\overline{\mathbf{R}}_{ik}^{j})$, and $\mathbf{V}_{\text{c},ik}^{j} \triangleq \frac{\epsilon_{ik}^2}{(\sigma^2\gamma_{ik})^2}\mathbf{V}_{ik}^{jT}\mathbf{V}_{ik}^{j*} \otimes \mathbf{I}_M$. In the sequel, for notation simplicity in deriving the solution to $\mathcal{P}_{\mathbf{w}}$, we remove subscript ik and superscripts j and (l) for all variables in $\mathcal{P}_{\mathbf{w}}$. Using an approach similar to the one used to solve $\mathcal{P}_{\mathbf{v}}$ in Section 3.3.1, we derive the solution to $\mathcal{P}_{\mathbf{w}}$ in two cases, as described below.

Case 1): The constraint in (3.42) is inactive at the optimality: In this case, $\mathcal{P}_{\mathbf{w}}$ is equivalent to the unconstrained problem with only the objective function. Setting the gradient of the objective function in $\mathcal{P}_{\mathbf{w}}$ w.r.t. $\mathbf{w}^*$ to zero, we have

$$\rho\overline{\mathbf{D}}^H(\overline{\mathbf{D}}\mathbf{w} - \bar{\mathbf{z}}) + \mu(\mathbf{w} - \bar{\mathbf{y}}) = 0,$$

which yields

$$\mathbf{w}^{\text{o}} = (\rho\overline{\mathbf{D}}^H\overline{\mathbf{D}} + \mu\mathbf{I})^{-1}(\rho\overline{\mathbf{D}}^H\bar{\mathbf{z}} + \rho\bar{\mathbf{y}}). \tag{3.43}$$

Case 2): The constraint in (3.42) is satisfied with equality at the optimality, *i.e.*, $\mathfrak{Re}\{\bar{\mathbf{r}}^H\mathbf{w}\} - 1 = \|\mathbf{V}_{\text{c}}^{\frac{1}{2}}\mathbf{w}\|$: In this case, by imposing $\mathfrak{Re}\{\bar{\mathbf{r}}^H\mathbf{w}\} - 1 \geq 0$, we have the following equivalent problem to $\mathcal{P}_{\mathbf{w}}$

$$\begin{aligned}\mathcal{P}'_{\mathbf{w}} : \min_{\mathbf{w}}\ & \rho(\overline{\mathbf{D}}\mathbf{w} - \bar{\mathbf{z}})^H(\overline{\mathbf{D}}\mathbf{w} - \bar{\mathbf{z}}) + \mu(\mathbf{w} - \bar{\mathbf{y}})^H(\mathbf{w} - \bar{\mathbf{y}}) \\ \text{s.t. }& (\mathfrak{Re}\{\bar{\mathbf{r}}^H\mathbf{w}\} - 1)^2 - \mathbf{w}^H\mathbf{V}_{\text{c}}\mathbf{w} = 0.\end{aligned} \tag{3.44}$$

Note that $\mathcal{P}'_{\mathbf{w}}$ is convex. The Lagrangian of $\mathcal{P}'_{\mathbf{w}}$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{w}, \tilde{\lambda}) =& \rho(\overline{\mathbf{D}}\mathbf{w} - \bar{\mathbf{z}})^H(\overline{\mathbf{D}}\mathbf{w} - \bar{\mathbf{z}}) + \mu(\mathbf{w} - \bar{\mathbf{y}})^H(\mathbf{w} - \bar{\mathbf{y}}) \\ & + \tilde{\lambda}(\mathbf{w}^H\mathbf{V}_{\text{c}}\mathbf{w} - (\mathfrak{Re}\{\bar{\mathbf{r}}^H\mathbf{w}\} - 1)^2).\end{aligned} \tag{3.45}$$

Equating the gradient of $\mathcal{L}(\mathbf{w}, \tilde{\lambda})$ w.r.t. $\mathbf{w}^*$ to zero, we have

$$\begin{aligned}\nabla_{\mathbf{w}^*}\mathcal{L}(\mathbf{w}, \tilde{\lambda}) =& (\mu\mathbf{I} + \rho\overline{\mathbf{D}}^2)\mathbf{w} - (\rho\overline{\mathbf{D}}^H\bar{\mathbf{z}} + \mu\bar{\mathbf{y}}) \\ & + \tilde{\lambda}(\mathbf{V}_{\text{c}}\mathbf{w} - (\mathfrak{Re}\{\bar{\mathbf{r}}^H\mathbf{w}\} - 1)\bar{\mathbf{r}}) = 0.\end{aligned} \tag{3.46}$$

Define $\bar{\mathbf{I}} \triangleq \mu\mathbf{I} + \rho\overline{\mathbf{D}}^2$, and $\bar{\mathbf{b}} \triangleq \rho\overline{\mathbf{D}}^H\bar{\mathbf{z}} + \mu\bar{\mathbf{y}}$. Using $\bar{\mathbf{I}}$ and $\bar{\mathbf{b}}$, we see that (3.46) has a structure similar to (3.28). Following (3.29), we obtain the optimal $\mathbf{w}^{\text{o}}$ to $\mathcal{P}'_{\mathbf{w}}$ as

$$\mathbf{w}^{\text{o}} = (\bar{\mathbf{I}} + \tilde{\lambda}\mathbf{V}_{\text{c}})^{-1}(\bar{\mathbf{b}} + \tilde{\lambda}(\mathfrak{Re}\{\bar{z}\} - 1)\bar{\mathbf{r}}) \tag{3.47}$$

where $\bar{z} \triangleq \bar{\mathbf{r}}^H \mathbf{w}^{\mathrm{o}}$. Note that $\mathbf{w}^{\mathrm{o}}$ in (3.47) has exactly the same form as $\mathbf{v}^{\mathrm{o}}$ in (3.29). Using the same technique in Section 3.3.1, we can obtain the expression $\mathfrak{Re}\{\bar{z}\}$ as a function of $\tilde{\lambda}$, and numerically obtain $\tilde{\lambda}$ by finding the roots of a smooth polynomial function. To determine which root is to be selected for $\tilde{\lambda}$, we recall that the constraint $\mathfrak{Re}\{\bar{z}\} - 1 \geq 0$ is imposed to $\mathcal{P}'_{\mathbf{w}}$ in order to be equivalent to $\mathcal{P}_{\mathbf{w}}$. We select the root $\tilde{\lambda}$ that results in the minimum objective value of $\mathcal{P}'_{\mathbf{w}}$ while satisfying $\mathfrak{Re}\{\bar{z}\} - 1 \geq 0$. Since the derivations are similar to those in Section 3.3.1, we leave the details of obtaining $\mathfrak{Re}\{\bar{z}\}$ and $\tilde{\lambda}$ in Appendix A.3.

Finally, the optimal $\mathbf{w}^{\mathrm{o}}$ to $\mathcal{P}_{\mathbf{w}}$ is chosen between (3.43) or (3.47), whichever leads to the minimum objective value. After solving $\mathcal{P}_{\mathbf{w}}$, we obtain the update $\mathbf{W}_{ik}^{(l+1)}$ to problem (3.41) for each $k \in \mathcal{K}_i$ and $i \in \mathcal{G}$.

Updating W

Given $\{\mathbf{W}_{ik}^{(l+1)}\}$, the update of $\mathbf{W}$ in (3.39) is obtained by solving the following problem:

$$\min_{\mathbf{W}} \mathrm{tr}(\mathbf{W}^H \mathbf{W}) + \mu \sum_{i=1}^{G} \sum_{k=1}^{K_i} \|\mathbf{W} - \left(\mathbf{W}_{ik}^{(l+1)} - \mathbf{Y}_{ik}^{(l)}\right)\|_F^2, \tag{3.48}$$

which is an unconstrained quadratic program. Its closed-from solution can be easily obtained by setting gradient of the objective function to zero. As a result, the update $\mathbf{W}^{(l+1)}$ is given by

$$\mathbf{W}^{(l+1)} = \frac{\mu \sum_{i=1}^{G} \sum_{k=1}^{K_i} \left(\mathbf{W}_{ik}^{(l+1)} - \mathbf{Y}_{ik}^{(l)}\right)}{1 + \mu \sum_{i=1}^{G} K_i}. \tag{3.49}$$

3.3.3 Discussions

Summary of the algorithm

For the robust multicast beamforming problem $\mathcal{P}_1$, rather than resorting to the conventional convex approximation approach, which is difficult to apply, we explore the structure of this non-convex problem formulation and directly solve it by leveraging the ADMM technique. Our overall proposed ADMM-based fast algorithm to solve $\mathcal{P}_1$ contains two layers of ADMM procedures. The outer-layer ADMM given by Algorithm 1 solves the non-convex

problem $\mathcal{P}_1$ by decomposing the problem into three main subproblems (3.22)-(3.24), which are solved iteratively. The salient feature in our design is that for these subproblems formed by our ADMM construction in $\mathcal{P}_2$, we are able to develop special techniques to derive the semi-closed-form solutions to these subproblems. For the second subproblem, we develop the inner-layer consensus-ADMM-based algorithm (Algorithm 2) to solve it, where the semi-closed-form solutions are obtained for the updates in each ADMM iteration. These semi-closed-form solutions are essential in our ADMM-based algorithm for fast computing the solution to $\mathcal{P}_1$, especially for large-scale systems such as massive MIMO.

Remark 1. Although ADMM has been considered for multicast beamforming design for a single group [128] and multiple groups [69], there are a few key differences between our problem and those in [69, 128]. The problems considered in [69, 128] are under perfect CSI, which leads to a QCQP formulation. In [128], in the absence of interference, a consensus ADMM-based algorithm is developed for the QCQP problem directly, although its convergence cannot be guaranteed. In [69], the SCA method is used to convexify the QCQP problem, and then an ADMM algorithm is developed to solve the convex problem at each SCA iteration. As commented in [69], consensus ADMM may also be applied, but it results in a problem of larger size with higher computational complexity. In contrast, our problem $\mathcal{P}_1$ under robust formulation is not a QCQP problem. As discussed below $\mathcal{P}_1$, the constraint function is in a special non-convex form that existing ADMM algorithms are not applicable. Different from these existing algorithms, we directly design the ADMM procedure to solve $\mathcal{P}_1$, resulting in two layers of ADMM constructions. As explained above, the key novelty in our ADMM constructions (in $\mathcal{P}_2$ and $\mathcal{P}'_{\mathbf{W}}$) is that at each ADMM iteration, we can break the main problem into smaller subproblems, for which we obtain the solution in semi-closed form.

Convergence

For solving $\mathcal{P}_{\mathbf{W}}$ to update $\mathbf{W}$ in (3.23), since the problem is convex, the inner-layer consensus-ADMM in Algorithm 2 is guaranteed to converge to the optimal solution [112]. For the outer-layer ADMM, the objective function in $\mathcal{P}_1$ is Lipschitz differentiable w.r.t.

$\mathbf{W}$. The reformulated problem $\mathcal{P}_2'$ is a non-convex optimization problem with equality constraints. Based on [51], we have the following conclusion.

Proposition 1. The outer-layer ADMM in Algorithm 1 is guaranteed to converge to a stationary point of $\mathcal{P}_1$.

Proof. To prove this proposition, we mainly apply the convergence result of Theorem 1 in [51] to problem $\mathcal{P}_2'$. We need to verify assumptions A1-A5 in Theorem 1 of [51] hold for $\mathcal{P}_2'$. Consider the structure and notations of the optimization problem (7) in [51], where the objective function is $\phi(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = [\mathbf{x}_0^T, \ldots, \mathbf{x}_p^T]^T$ and $\mathbf{y}$ being the optimization variables, and the linear equality constraint is $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{b}$. We rewrite our problem $\mathcal{P}_2'$ in the format of problem (7) in [51]: the objective function of $\mathcal{P}_2'$ can be written as $\phi(\{\mathbf{V}_{ik}\}, \mathbf{W}) = f_0(\{\mathbf{V}_{ik}\}, \mathbf{W}) + h(\mathbf{W})$, where $h(\mathbf{W}) = \text{tr}(\mathbf{W}^H\mathbf{W})$ and $f_0(\{\mathbf{V}_{ik}\}, \mathbf{W}) = \mathbb{I}_{\mathcal{C}}(\{\mathbf{V}_{ik}\}, \mathbf{W})$ is an indicator function that is lower semi-continuous. Note that, in our problem, we treat $\{\mathbf{V}_{ik}\}$ as $\mathbf{x}_0$ and $\mathbf{W}$ as $\mathbf{y}$ (in our case, $\mathbf{x} = \mathbf{x}_0$). Also, the equality constraint in (3.20) of $\mathcal{P}_2'$ is $\mathbf{V}_{ik} - \mathbf{W}\mathbf{D}_{ik} = \mathbf{0}$, for all i, k.

Based on the above, we check assumptions A1-A5. A1 (coercivity) holds: as $\mathbf{W} \to \infty$, $h(\mathbf{W}) = \text{tr}\big(\mathbf{W}^H\mathbf{W}\big) \to \infty$, and thus, $\phi(\{\mathbf{V}_{ik}\}, \mathbf{W}) \to \infty$. A2 (feasibility) can be easily verified, since for the equality constraint (3.20), we have $\mathbf{A} = \mathbf{I}$ and $\mathbf{B} = -\mathbf{D}_{ik}$. A3 (Lipschitz sub-minimization paths): This assumption holds as both $-\mathbf{D}_{ik}$ and $\mathbf{I}$ are full column rank. A4 (objective-f regularity) holds, because we do not have the function g in our objective function ϕ, and $f_0 = \mathbb{I}_{\mathcal{C}}$ is lower semi-continuous. A5 (objective-h regularity): this can be easily verified as $h(\mathbf{W}) = \text{tr}(\mathbf{W}^H\mathbf{W})$ is Lipschitz differentiable.

We point out that the structure of our objective function ϕ is slightly different from that in the problem (7) of [51]. Specifically, our objective function is in the form of $\phi(\mathbf{x}, \mathbf{y}) = f_0(\mathbf{x}_0, \mathbf{y}) + h(\mathbf{y})$, where $f_0(\mathbf{x}_0, \mathbf{y}) = \mathbb{I}_{\mathcal{C}}(\mathbf{x}_0, \mathbf{y})$ is an indicator function of both $\mathbf{x}_0$ and $\mathbf{y}$, while in the problem (7) of [51], $f_0(\mathbf{x}_0)$ is a lower semi-continuous function of $\mathbf{x}_0$ only. Nonetheless, Theorem 1 of [51] still holds under this modification. That is, we can verify that all the proofs, including the lemmas for Theorem 1, still hold. We explain the reason briefly: this is because f_0 in our problem is a specific indicator function; at each ADMM

iteration, the updates for $\mathbf{x}_0$ and $\mathbf{y}$ are the solutions of the minimization of the augmented Lagrangian function w.r.t. $\mathbf{x}_0$ and $\mathbf{y}$, respectively. As a result of the minimization, these updates for $\mathbf{x}_0$ and $\mathbf{y}$ should satisfy $(\mathbf{x}_0, \mathbf{y}) \in \mathcal{C}$ and $f_0(\mathbf{x}_0, \mathbf{y}) = 0$. Using this fact, we verify that all the steps in the proof of Theorem 1 still hold. Thus, as stated in Theorem 1 of [51], the ADMM algorithm converges to a stationary point. □

For ADMM convergence, two types of convergence criteria are considered in the literature [112]: 1) Residual convergence: this refers to the equality constraint of the auxiliary matrices introduced in (3.18). That is, $\|\mathbf{V}_{ik}^{j+1} - \mathbf{W}^{j+1}\mathbf{D}_{ik}\|_F \to 0$, $\forall k \in \mathcal{K}_i, i \in \mathcal{G}$, as $j \to \infty$; 2) Objective convergence: this refers to the convergence of the objective function over iterations. Note that the residual and objective quantities converge simultaneously. Therefore, we can use either of the two quantities as the convergence criterion.

Initialization

One advantage of ADMM for our problem is that it does not require a feasible initial point for convergence [51, 112]. Thus, our proposed ADMM-based algorithm can use a random initial point with robust convergence, although a good initial point facilitates faster convergence. This is in contrast to existing popular convex approximation approaches, such as SCA, where a feasible initial point is required, which imposes a challenge for the algorithm design and implementation.

3.3.4 Computational Complexity

As discussed earlier, our proposed algorithm includes two layers of iterations. The computations in each iteration are based on the (semi-)closed-form expressions. Algorithm 1 for the outer layer iteratively updates $\{\mathbf{V}_{ik}\}$, $\mathbf{W}$, $\{\mathbf{Z}_{ik}\}$, where $\mathbf{W}$ is computed by Algorithm 2. At each iteration, $\mathbf{V}_{ik}$ is updated using (3.29) (through the vectorized version $\mathbf{v}_{ik}$). This requires the eigenvalue decomposition of $\mathbf{W}_{\mathrm{c}}$, which is needed for the matrix inversion and for simplifying (3.33) into (3.34) to compute λ. As discussed below (3.33), $\mathbf{W}_{\mathrm{c}}$ is a block diagonal matrix. Thus, the eigenvalue decomposition is only needed for $\mathbf{W}^T\mathbf{W}^*$ (a $G \times G$

matrix), which has the complexity of $O(G^3)$. Once we have the eigenvalue decomposition of $\mathbf{W}_{\mathrm{c}}$, it can be directly used to compute the matrix inversion in (3.29). The complexity for the remaining matrix-vector multiplications is $O(MG^2)$ for each $\mathbf{V}_{ik}$. Note that the eigenvalue decomposition only needs to be computed once for all $\mathbf{V}_{ik}$'s. Thus, the overall complexity to compute all $\mathbf{V}_{ik}$'s is $O(G^3 + (\sum_{i=1}^{G} K_i)MG^2)$. From (3.24), updating $\{\mathbf{Z}_{ik}\}$ requires $O((\sum_{i=1}^{G} K_i)MG)$. Thus, the computational complexity for updating $\{\mathbf{V}_{ik}\}$ and $\{\mathbf{Z}_{ik}\}$ at each iteration is $O(G^3 + (\sum_{i=1}^{G} K_i)MG^2)$.

Algorithm 2 for the inner layer iteratively updates $\mathbf{W}, \{\mathbf{W}_{ik}\}, \{\mathbf{Y}_{ik}\}$. Among these quantities, updating $\mathbf{W}_{ik}$ in (3.47) contains the most computational intensive operations. Note that $\mathbf{W}_{ik}$ in (3.47) has a structure similar to $\mathbf{V}_{ik}$ (3.29). Following the similar analysis above, the computation involves eigenvalue decomposition and matrix-vector operations in (3.47) and has the complexity of $O(G^3 + MG^2)$ for each $\mathbf{W}_{ik}$. Different from that in the outer layer, eigenvalue decomposition is required for each $\mathbf{W}_{ik}$, as $\mathbf{V}_{\mathrm{c},ik}^{j}$ is different for each $\mathbf{w}_{ik}$ in $\mathcal{P}_{\mathbf{w}}$ (or equivalently, $\mathbf{V}_{ik}^{j}$ is different for each $\mathbf{W}_{ik}$ in (3.41)). Thus, the total computational complexity for updating all $\mathbf{W}_{ik}$'s is $O((\sum_{i=1}^{G} K_i)(G^3 + MG^2))$. The complexity for updating $\mathbf{W}$ in (3.49) and $\{\mathbf{Y}_{ik}\}$ in (3.40) is $O((\sum_{i=1}^{G} K_i)MG)$. Thus, the computational complexity per iteration in Algorithm 2 is $O((\sum_{i=1}^{G} K_i)(G^3 + MG^2))$.

From the above analysis for Algorithms 1 and 2, the overall computational complexity of our proposed algorithm in each outer layer iteration is $O((\sum_{i=1}^{G} K_i)(G^3 + MG^2))$.

3.4 Lower Bound and Alternative Approach

For comparison purpose, in this section, we consider a lower bound for $\mathcal{P}_1$ and briefly describe another possible approach for the robust multicast beamforming problem $\mathcal{P}_{\mathrm{rob}}$. These methods will be used for performance comparison with our proposed algorithm in Section 3.5 for the simulation study.

3.4.1 Performance Benchmark for $\mathcal{P}_1$

As the performance benchmark, we consider a lower bound on the objective function of $\mathcal{P}_1$ by solving the relaxed version of $\mathcal{P}_1$. Let $\mathbf{X}_i \triangleq \mathbf{w}_i\mathbf{w}_i^H$, $i \in \mathcal{G}$. By dropping the rank-1 constraint on $\mathbf{X}_i$, we relax $\mathcal{P}_1$ into the following problem

$$\begin{aligned}\mathcal{P}_1^{\text{lb}}: \min_{\{\mathbf{X}_i\}} \ & \sum_{i=1}^{G} \text{tr}(\mathbf{X}_i) \\ \text{s.t. } & \text{tr}(\mathbf{R}_{ik}\mathbf{X}_i) - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}} \text{tr}(\mathbf{R}_{ik}\mathbf{X}_l) - \epsilon_{ik}\|\mathbf{X}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}} \mathbf{X}_l\|_F \geq \sigma^2\gamma_{ik},\ k \in \mathcal{K}_i, i \in \mathcal{G} \\ & \mathbf{X}_i \succcurlyeq 0,\ i \in \mathcal{G}. \end{aligned} \tag{3.50}$$

Note that $\mathcal{P}_1^{\text{lb}}$ is convex w.r.t. $\{\mathbf{X}_i\}$.[5] Thus, it can be solved using the standard convex solvers. The optimal objective value of $\mathcal{P}_1^{\text{lb}}$ provides a lower bound to $\mathcal{P}_1$.

3.4.2 The SCA Approach to Solve $\mathcal{P}_{\text{consv}}$

Recall the alternative optimization problem $\mathcal{P}_{\text{consv}}$ discussed in Section 3.2.1, which is formulated using the conservative SINR lower bound. It can be solved via the SCA approach. The SCA approach is a popular method to solve a non-convex problem via successive convex approximation of the problem. We briefly describe the SCA approach below.

Consider auxiliary variables $\mathbf{u}_i$, $i \in \mathcal{G}$. Using the fact that $(\mathbf{u}_i - \mathbf{w}_i)^H\mathbf{A}(\mathbf{u}_i - \mathbf{w}_i) \geq 0$ holds for any $\mathbf{A} \succcurlyeq 0$, we have $\mathbf{w}_i^H\mathbf{A}\mathbf{w}_i \geq 2\Re\{\mathbf{u}_i^H\mathbf{A}\mathbf{w}_i\} - \mathbf{u}_i^H\mathbf{A}\mathbf{u}_i$. Applying this inequality to the numerator of the constraint in $\mathcal{P}_{\text{consv}}$, we convexify the constraint and have the following problem for given $\{\mathbf{u}_i\}$:

$$\begin{aligned}\min_{\{\mathbf{w}_i\}} \ & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 \\ \text{s.t. } & \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}} \mathbf{w}_l^H(\mathbf{R}_{ik} + \epsilon_{ik}\mathbf{I})\mathbf{w}_l + \epsilon_{ik}\mathbf{w}_i^H\mathbf{w}_i + \sigma^2\gamma_{ik} \\ & -2\mathfrak{Re}\{\mathbf{u}_i^H\mathbf{R}_{ik}\mathbf{w}_i\} + \mathbf{u}_i^H\mathbf{R}_{ik}\mathbf{u}_i \leq 0, k \in \mathcal{K}_i, i \in \mathcal{G}. \end{aligned} \tag{3.51}$$

[5] Note that $\mathcal{P}_1^{\text{lb}}$ is not a semi-definite programming (SDP) problem due to the third term in the constraint. Thus, this relaxation approach is different from the typical semi-definite relaxation commonly considered.

Note that any set of $\{\mathbf{w}_i\}$ that satisfies the constraints in (3.51) also satisfies the constraints in $\mathcal{P}_{\text{consv}}$ and thus are feasible to $\mathcal{P}_{\text{consv}}$. Since problem (3.51) is convex, the solution $\{\mathbf{w}_i\}$ can be easily obtained using the standard convex solvers. Then, letting $\mathbf{u}_i = \mathbf{w}_i$, $i \in \mathcal{G}$, we iteratively solve problem (3.51) until convergence.

3.5 Simulation Results

We consider the downlink multi-group multicast beamforming problem with a symmetric setup, where $K_i = K, \forall\, i \in \mathcal{G}$, and the SINR target per user $\gamma_{ik} = \gamma, \forall\, k \in \mathcal{K}_i, i \in \mathcal{G}$. We consider the default setup as $G = 3$ groups, $K = 4$ users per group, $M = 100$ antennas, and $\gamma = 5$ dB, unless otherwise stated. Noise variance is set to $\sigma^2 = 1$. We consider the error bound $\epsilon_{ik} = \eta\|\bar{\mathbf{R}}_{ik}\|_F$ in (3.3), where η is the normalized error bound that we use to study the robust beamforming performance. We consider two models for the channel covariance matrix that are commonly considered in the literature:

Model 1

Consider the channel vector $\mathbf{h}_{ik}$ between BS and each user $k \in \mathcal{K}_i, i \in \mathcal{G}$. Let $\hat{\mathbf{h}}_{ik}$ be the estimated channel vector, where $\hat{\mathbf{h}}_{ik} = \mathbf{h}_{ik} + \mathbf{e}_{ik}$, with $\mathbf{e}_{ik}$ being the error vector. We model the channel covariance matrix as $\bar{\mathbf{R}}_{ik} = \mathbf{h}_{ik}\mathbf{h}_{ik}^H$. The estimated covariance matrix is given as $\mathbf{R}_{ik} = \hat{\mathbf{h}}_{ik}\hat{\mathbf{h}}_{ik}^H$. Then, the error matrix is obtained as $\mathbf{E}_{ik} = \bar{\mathbf{R}}_{ik} - \mathbf{R}_{ik}$. We generate independent channel vectors as $\mathbf{h}_{ik} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Consider the bounded estimation error model as in (3.3). To model the error more realistically, we first generate the independent errors as $\mathbf{e}_{ik} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$, where σ_e^2 is the error variance. Then, we set $\epsilon_{ik} = \eta\|\bar{\mathbf{R}}_{ik}\|_F$, where η is a cut-off bound, which is set such that $\|\mathbf{E}_{ik}\|_F \leqslant \eta\|\bar{\mathbf{R}}_{ik}\|_F$ holds for 90% of realizations of $\mathbf{e}_{ik}$. Those realizations with $\mathbf{E}_{ik}$ satisfying (3.3) are considered in the simulation.

Model 2

We consider a narrow-band mmWave channel model. The number of multi-path clusters between the transmitter and each receiver is assumed $N_{\text{c}} = 2$, and the number of rays

per cluster is assumed $N_{\rm r} = 3$. Between the transmitter and user k in group i, the steering vector of ray p in cluster c is given by $\mathbf{a}(\theta_{ik}^{cp}) \triangleq \frac{1}{\sqrt{M}}[1, e^{j2\pi \frac{\Delta}{\lambda_{\rm c}} \cos(\theta_{ik}^{cp})}, \ldots, e^{j2\pi \frac{\Delta}{\lambda_{\rm c}}(M-1)\cos(\theta_{ik}^{cp})}]$, where θ_{ik}^{cp} is the angle of departure (AoD) of this ray, Δ is the antennas separation, and $\lambda_{\rm c}$ is the carrier wavelength. The rays in each cluster are generated within an angular spread around the cluster center. For user k in group i, the AoD of ray p in cluster c is modeled as $\theta_{ik}^{cp} = \bar{\theta}_{ik}^{c} + \Delta\theta_{ik}^{cp}$, where $\bar{\theta}_{ik}^{c} \in [0^o, 360^o)$ is the AoD of the cluster center randomly generated, and $\Delta\theta_{ik}^{cp} \sim$ Uniform$[-2^o, 2^o]$ is the angle of ray p from the cluster center $\bar{\theta}_{ik}^{c}$ generated independently from other rays. The channel covariance matrix for user k in group i is given by $\bar{\mathbf{R}}_{ik} = \sum_{c=1}^{N_{\rm c}} \sum_{p=1}^{N_{\rm r}} \xi_{ik}^{cp} \mathbf{a}(\theta_{ik}^{cp}) \mathbf{a}^H(\theta_{ik}^{cp})$, where $\xi_{ik}^{cp} \in [0, 1]$ is the relative weight of ray p in cluster c for the user, with $\sum_{c=1}^{N_{\rm c}} \sum_{p=1}^{N_{\rm r}} \xi_{ik}^{cp} = 1$. For the estimated channel covariance matrix $\mathbf{R}_{ik}$, we consider AoD estimation. The AoD estimate of ray p in cluster c is $\hat{\theta}_{ik}^{cp} = \bar{\theta}_{ik}^{c} + \Delta\theta_{ik}^{cp} + \Delta\phi^{cp}$, where the estimation error $\Delta\phi_{ik}^{cp} \sim$ Uniform$[-\frac{\phi_{\max}}{2}, \frac{\phi_{\max}}{2}]$, with $\phi_{\max}$ being maximum error spread. Thus, $\mathbf{R}_{ik}$ is given by $\mathbf{R}_{ik} = \sum_{c=1}^{N_{\rm c}} \sum_{p=1}^{N_{\rm r}} \xi_{ik}^{cp} \mathbf{a}(\hat{\theta}_{ik}^{cp}) \mathbf{a}^H(\hat{\theta}_{ik}^{cp})$. Similar to Model 1, the error bound for the error matrix $\mathbf{E}_{ik}$ is set as $\epsilon_{ik} = \eta \|\bar{\mathbf{R}}_{ik}\|_F$. Specifically, for a given value of $\phi_{\max}$, η is set such that $\|\mathbf{E}_{ik}\|_F \leqslant \eta \|\bar{\mathbf{R}}_{ik}\|_F$ holds for in 90% of the error realizations. Those realizations with $\mathbf{E}_{ik}$ satisfying (4.4) are considered in the simulation.

3.5.1 Algorithm Convergence Behavior

We first study the convergence behaviour of the proposed algorithm. Algorithm 1 contains the outer-layer ADMM-based iterations to solve $\mathcal{P}_1$, and Algorithm 2 contains inner-layer consensus-ADMM-based iterations to solve $\mathcal{P}_{\mathbf{W}}$. As discussed in Section 3.3.3, we consider both objective convergence and residual convergence. Let P^j denote the value of the power objective function in $\mathcal{P}_1$ in iteration j. We define the relative objective difference in iteration j as $\Delta P^j \triangleq \frac{|P^j - P^{j-1}|}{P^{j-1}}$. The relative residual for the equality constraint in (3.18) is defined as $\Delta V^j \triangleq \max_{i,k} \left\{ \frac{\|\mathbf{V}_{ik}^j - \mathbf{D}_{ik}\mathbf{W}^j\|_F}{\|\mathbf{V}_{ik}^j\|_F} \right\}$. Model 1 for $\mathbf{R}_{ik}$'s is used to show the convergence performance. A random initial point $\mathbf{W}^{(0)}$ for Algorithm 1 is used. We set $\rho = \mu = \frac{7}{\sqrt{M}}$, which are found to provide fast convergence under various values of M, K, and G in our experimental study. Figs. 3.1 and 3.2 show the trajectory of ΔP^j and ΔV^j vs.

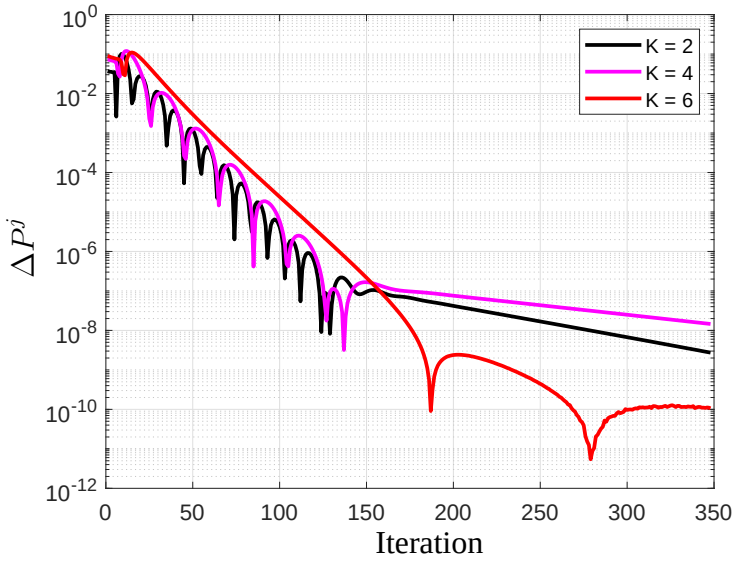

Figure 3.1: The convergence behavior of relative objective difference ΔP^j of $\mathcal{P}_1$ by Algorithm 1 (Model 1).

iteration j, respectively, for $K = 2, 4, 6$. We see that both the values of ΔP^j and ΔV^j drop below 10^{-3} after only 50 iterations, demonstrating fast convergence. When using Algorithm 2 to solve $\mathcal{P}_{\mathbf{W}}$ in the inner layer, we define the relative difference of the residual for the equality constraint in (3.36) in iteration l as $\Delta W^{(l)} \triangleq \max_{i,k} \left\{ \frac{\|\mathbf{W}_{ik}^{(l)} - \mathbf{W}^{(l)}\|_F}{\|\mathbf{W}_{ik}^{(l)}\|_F} \right\}$. Fig. 3.3 shows the trajectory of $\Delta W^{(l)}$ vs. iteration l at the outer-layer iteration $j = 3, 10, 20, 50$. We see that within 10 to 25 iterations, the value of $\Delta W^{(l)}$ drops below 10^{-3}, showing fast convergence. The inner-layer convergence rate becomes faster as the outer-layer iteration j increases. This is because the initial value $\mathbf{W}^{(0)}$ in Algorithm 2 is set as the output $\mathbf{W}^j$ from Algorithm 2 from the previous outer-layer iteration j. Thus, the initial value becomes closer to the stationary point of the problem, as Algorithm 1 converges.

3.5.2 Effect of Normalized Error Bound η on Performance

We now study the effect of channel covariance estimation error on the performance. For comparison, besides the solution to $\mathcal{P}_1$ obtained by our proposed algorithm, we consider the solution to $\mathcal{P}_{\text{consv}}$ found by the SCA approach in Section 3.4.2. Note that common to all QoS problems, $\mathcal{P}_0$ may not always be feasible. It depends on the SINR target considered and the

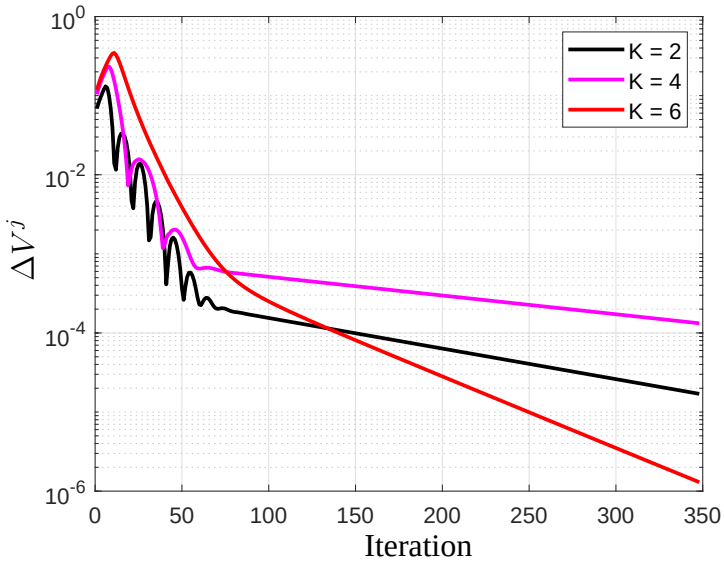

Figure 3.2: The convergence behavior of relative residual ΔV^j by Algorithm 1 (Model 1).

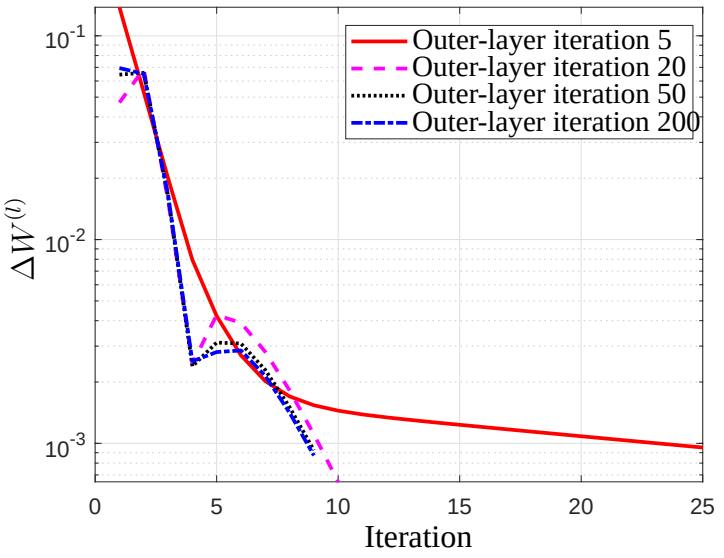

Figure 3.3: The convergence behavior of relative residual $\Delta W^{(l)}$ by Algorithm 2 to solve $\mathcal{P}_{\mathbf{W}}$ (Model 1).

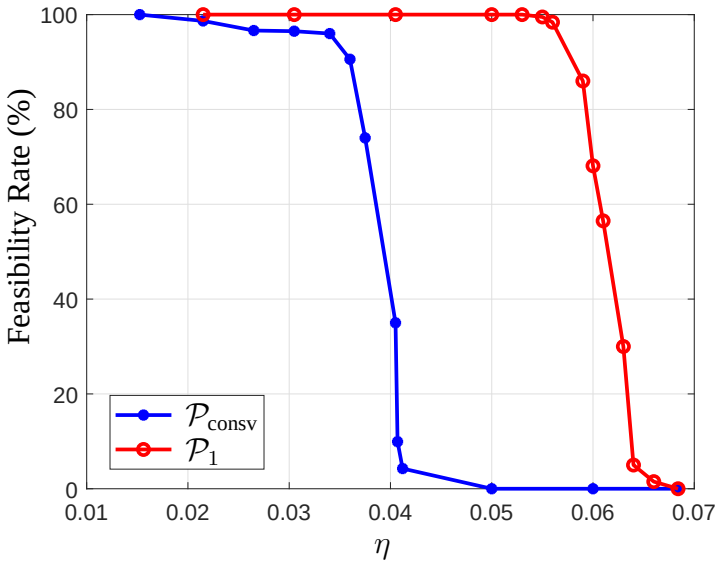

Figure 3.4: The feasibility rate vs. normalized error bound η (Model 1; $G = 3$, $K = 4$, $M = 100$).

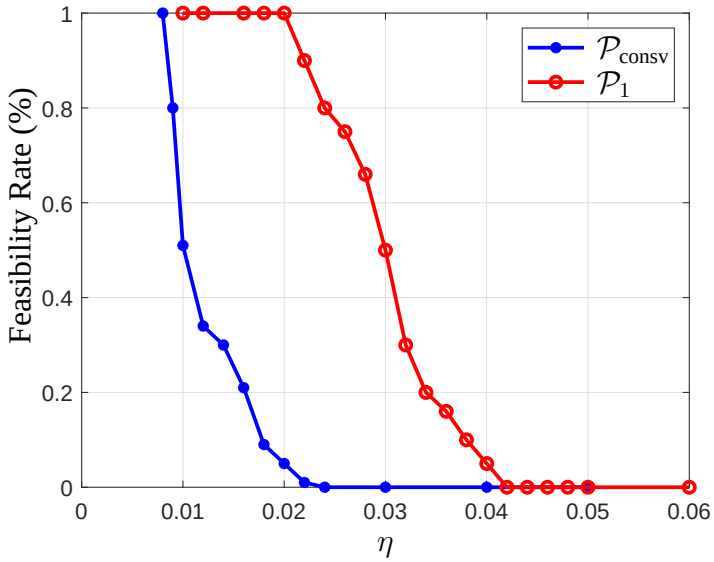

Figure 3.5: The feasibility rate vs. normalized error bound η (Model 2; $G = 3$, $K = 4$, $M = 100$).

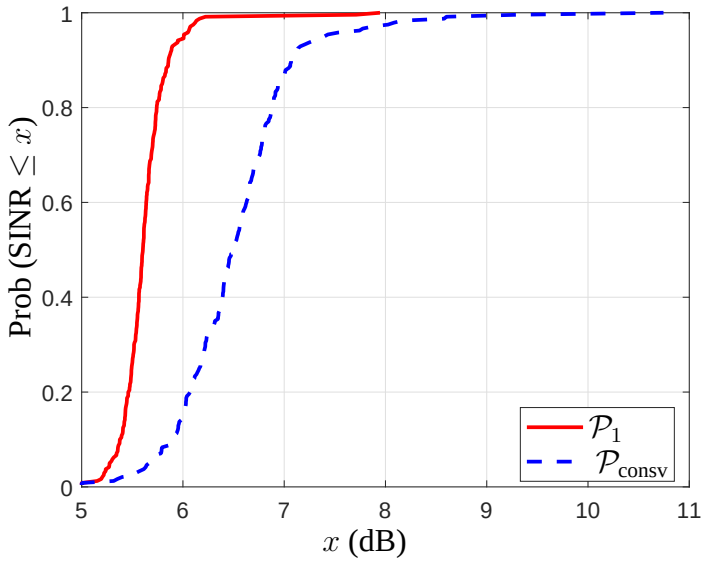

Figure 3.6: The CDF of the actual SINR at users under the robust design solution (Model 1; $\eta = 0.04$).

level of the estimation error. Thus, an important performance measure for robust design is to evaluate the percentage of channel realizations that the considered algorithm can find a feasible solution, which we refer to as the feasibility rate. We first evaluate the feasibility rate for a given error bound η. Fig. 3.4 shows the feasibility rate vs. the normalized error bound η for Model 1 over 200 channel realizations. Common in robust design, we see that the feasibility rate is highly dependent on the error η of the channel covariance estimation; it drops to 0 when η exceeds a certain value. We observe the improved feasibility offered by our proposed algorithm over the approach via solving $\mathcal{P}_{\text{consv}}$. This is because $\mathcal{P}_{\text{consv}}$ uses a conservative SINR lower bound to replace the original SINR constraint, while our solution uses the original SINR constraint. A similar comparison of the feasibility rate is observed for Model 2, as shown in Fig. 3.5.

With $\{\mathbf{w}_i\}$ obtained by our robust design (Algorithms 1 and 2) for $\mathcal{P}_1$ and by SCA for $\mathcal{P}_{\text{consv}}$, we compare the actual SINR received by all users. Fig. 3.6 plots the CDF of the actual SINR at the LHS of (3.2) of all users under Model 1 for $\eta = 0.04$. It is evident that since $\mathcal{P}_{\text{consv}}$ uses a conservative SINR lower bound, the actual SINRs among users exceed the SINR target by a larger margin as compared to our proposed algorithm via the exact

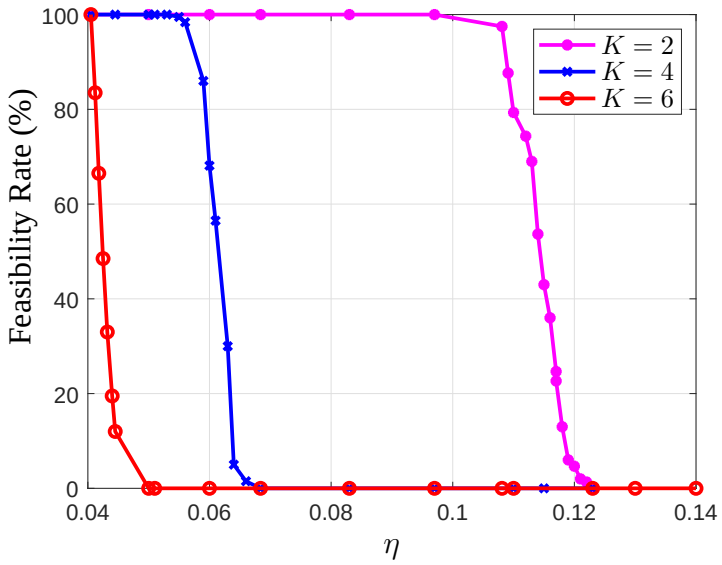

Figure 3.7: Feasibility rate vs. η

worst-case SINR constraint. A similar performance is observed for Model 2 and thus is not shown.

Next, in Fig. 3.7, we show the performance for different number of users per group $K = 2, 4, 6$ for Model 1. Fig. 3.7 shows the feasibility rate vs. η over 200 channel realizations. We see that for the same SINR target γ, the feasibility curve shifts to the left as K increases, indicating that the tolerable margin of channel uncertainty reduces in the robust design. Note that this trend is consistent with the behavior of multicast beamforming under perfect CSI, where the minimum SINR within a group reduces as K increases. Fig. 3.8 shows the corresponding transmit power objective value in $\mathcal{P}_1$ achieved by our algorithm normalized against noise P/σ^2 vs. η. As η increases, the transmitted power increases significantly. When the feasibility rate drops to 0, we can equivalently consider the transmit power being $P = +\infty$. The same performance trend as in Figs. 3.7 and 3.8 is observed for Model 2 as well.

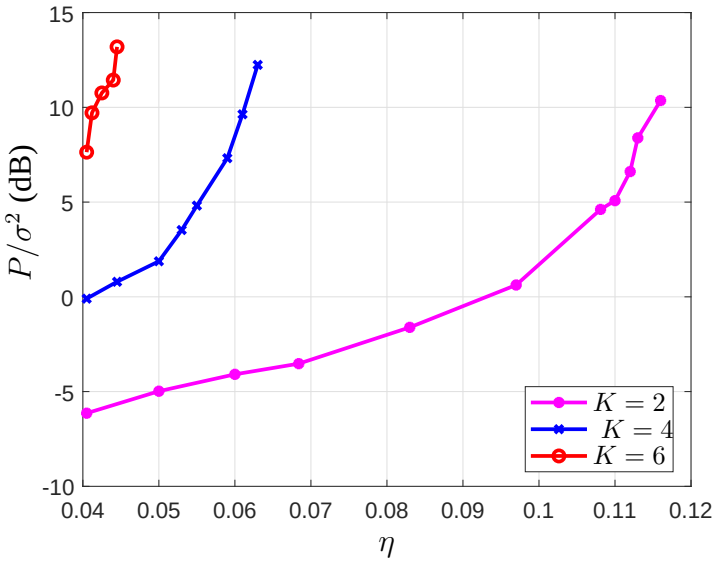

Figure 3.8: Normalized transmit power vs. η

3.5.3 Performance Comparison

We now evaluate the performance of our proposed algorithm (Algorithms 1 and 2) for $\mathcal{P}_1$ for different system parameters. For comparison, we also consider the following approaches: 1) $\mathcal{P}_1$ via IPM: solving $\mathcal{P}_1$ using IPM via the generic nonlinear program solver `fmincon` in MATLAB.[6]; 2) Lower Bound ($\mathcal{P}_1^{\text{lb}}$): the lower bound on the objective function of $\mathcal{P}_1$, which is obtained by solving $\mathcal{P}_1^{\text{lb}}$ in Section 3.4.1; 3) $\mathcal{P}_{\text{consv}}$ via SCA: solving $\mathcal{P}_{\text{consv}}$ by SCA; 4) Perfect Case ($\mathcal{P}_{\text{perf}}$): perfect channel covariance matrix, which is equivalent to solve $\mathcal{P}_{\text{perf}}$.

Performance vs. number of antennas M

Figs. 3.9 and 3.10 show the average normalized transmit power P/σ^2 vs. M for Model 1 ($\eta = 0.04$) and Model 2 ($\eta = 0.02$), respectively. The performance of our ADMM-based algorithm is near identical to that of the IPM solution. Also, our algorithm substantially outperforms the solution to $\mathcal{P}_{\text{consv}}$, providing significant power saving under both channel

[6]Note that there is no other existing algorithm to solve the robust design problem $\mathcal{P}_{\text{rob}}$ (equivalent to $\mathcal{P}_1$) directly. Thus, we resort to the generic IPM to solve $\mathcal{P}_1$ for comparison.

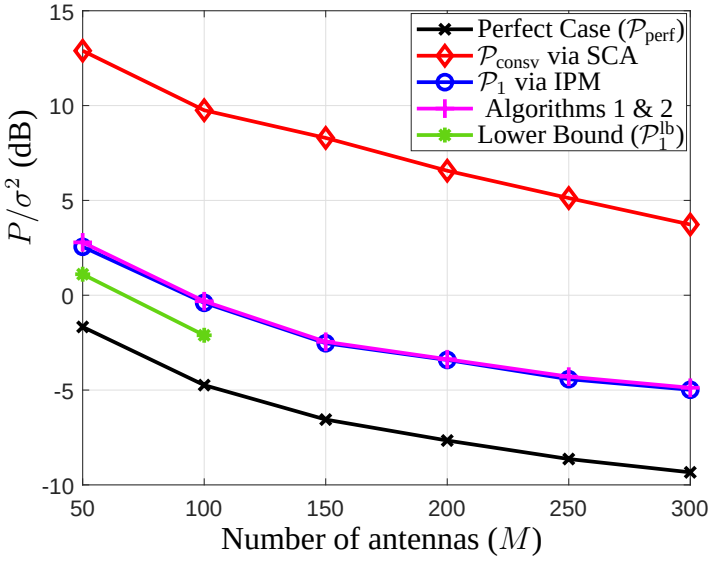

Figure 3.9: Normalized transmit power vs. M, Model 1, $\eta = 0.04$.

models. It is observed that larger values of η result in bigger performance gap between our solution to $\mathcal{P}_1$ and the solution to $\mathcal{P}_{\text{consv}}$. Note that the lower bound is only provided for $M \leq 100$. This is because solving $\mathcal{P}_1^{\text{lb}}$ using standard convex solvers incurs very high computational complexity for a large problem size, and thus becomes impractical for $M \geq 100$.

The computational advantage of our proposed algorithm is shown in Table 3.1, where we provide the average computation time for the plots in Fig. 3.9. Our proposed algorithm is fast in computing the solution, and the increment of computation time over M is mild. This is in particular attractive for massive MIMO systems. In contrast, both using IPM for $\mathcal{P}_1$ and using SCA for $\mathcal{P}_{\text{consv}}$ have high computational complexity, which increases with M significantly and thus are impractical for moderately large value of M.

Performance vs. number of users K

Figs. 3.11 and 3.12 show the normalized transmitted power vs. K for Model 1 and Model 2, respectively. Again, we see that our proposed ADMM-based algorithm achieves the same performance as IPM for $\mathcal{P}_1$. For K from 2 to 10, our algorithm always obtains a solution,

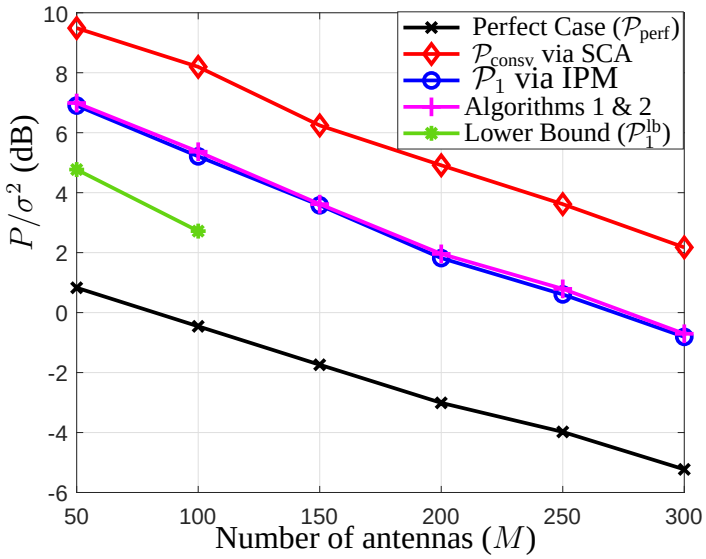

Figure 3.10: Normalized transmit power vs. M, Model 2, $\eta = 0.02$.

Table 3.1: Average Computation Time Over M (sec.)

M	50	100	150	200	250	300
Algorithms 1 & 2	0.23	0.48	0.55	0.72	6.44	32.45
IPM ($\mathcal{P}_1$)	69.85	413.3	1617	2788	9954	17490
$\mathcal{P}_{\text{consv}}$ via SCA	195.51	970.7	3017	7643	2.3e4	8.7e4

while $\mathcal{P}_{\text{consv}}$ via SCA increasingly fails to find a feasible solution, and thus, the transmit power increases sharply and goes to infinity for $K > 8$ (*i.e.*, no feasible solution can be found). In contrast, the transmit power required by our solution is significantly lower. Note that as K becomes larger, the lower bound becomes more loose, as the solution to $\mathcal{P}_1^{\text{lb}}$ is less likely to be rank-1. Thus, a larger gap is observed between our algorithm and the lower bound.

Table 3.2 provides the corresponding computation time of the plots in Fig. 3.11. Again, our proposed ADMM-based algorithm has a substantially lower computational complexity as compared with the other two methods. The computation time of our algorithm only increases with K slightly. This indicates that our proposed algorithm is suitable for massive MIMO systems.

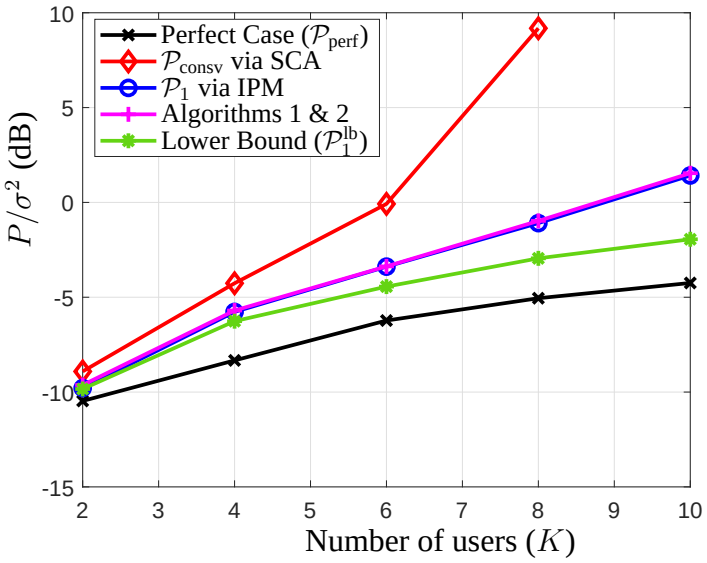

Figure 3.11: Normalized transmit power vs. K, Model 1, $\eta = 0.04$.

Table 3.2: Average Computation Time Over K (sec.)

K	2	4	6	8	10
Algorithms 1 & 2	0.29	0.48	1.60	2.16	11.49
$\mathcal{P}_1$ via IPM	90.9	413.2	561	1076	3023
$\mathcal{P}_{\text{consv}}$ via SCA	267	1032	4061	7043	N/A

Performance vs. number of groups G

Figs. 3.13 and 3.14 show the average normalized transmit power vs. G for Model 1 and Model 2, respectively. Similar relative performance as in Figs. 3.11 and 3.12 can be observed. Our proposed ADMM-based algorithm yields a solution with a much lower transmit power than that of $\mathcal{P}_{\text{consv}}$ via SCA. The latter fails to find a feasible solution for $G > 4$. The computation time of different algorithms that generate the plots in Fig. 3.13 is shown in Table 3.3. We see that the computation time of our proposed algorithm only increases very mildly over G and is significantly lower than that of other algorithms in comparison.

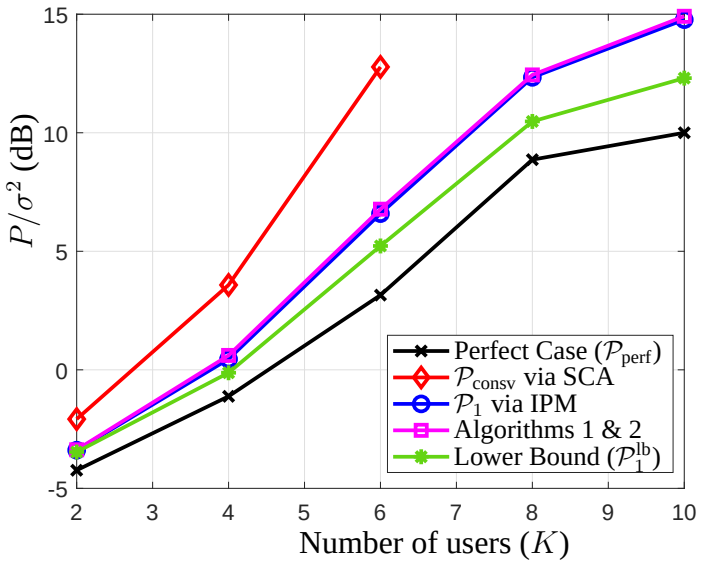

Figure 3.12: Normalized transmit power vs. K, Model 2, $\eta = 0.02$.

Table 3.3: Average Computation Time Over G (sec.)

G	2	3	4	5
Algorithms 1 & 2	0.17	0.48	0.57	0.92
$\mathcal{P}_1$ via IPM	134.0	413.2	1189	2884
$\mathcal{P}_{\text{consv}}$ via SCA	627	1032	3.8e3	N/A

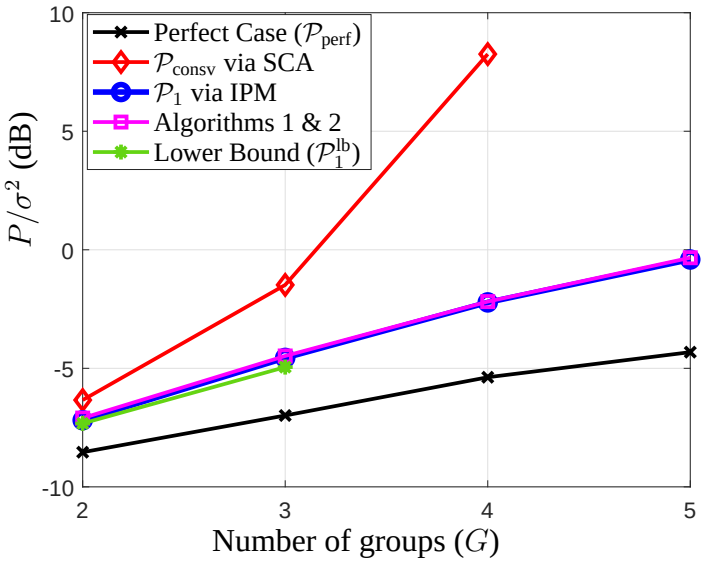

Figure 3.13: Normalized transmit power vs. G, Model 1, $\eta = 0.04$.

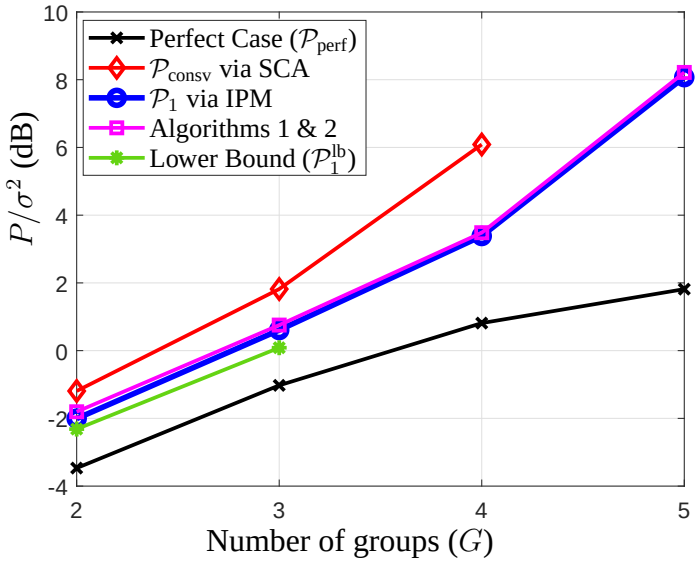

Figure 3.14: Normalized transmit power vs. G, Model 2, $\eta = 0.02$.

Chapter 4

Joint Antenna Selection and Robust Multi-group Multicast Beamforming in Massive MIMO Systems

4.1 System Model and Problem Formulation

4.1.1 System Model

We consider a downlink multi-group multicasting scenario in a massive MIMO system, where a multi-antenna BS serves G multicast groups. Users in each group receive a common message, which is independent of messages sent to other groups. Let $\mathcal{G} \triangleq \{1, ..., G\}$ denote the index set of the multicast groups and $\mathcal{K}_i \triangleq \{1, \ldots, K_i\}$ the index set of the single-antenna users in group i, for $i \in \mathcal{G}$. Each user is associated with only one multicast group.

The BS is equipped with M antennas ($M \gg 1$) and L transmit RF chains, where $L \leq M$. With a limited number of RF chains, the BS selects L antennas, one for each RF chain, for downlink transmission. Let $\mathbf{h}_{ik} \in \mathbb{C}^{M\times 1}$ denote the channel vector from the BS to user k in group i. Let $\mathbf{w}_i = [w_{i1}, \cdots, w_{iM}]^T \in \mathbb{C}^{M\times 1}$ denote the multicast beamforming vector for group i. Let $\mathbf{a} \triangleq [a_1, \ldots, a_M]^T$ denote the antenna selection vector, where $a_m \in \{0, 1\}$ is the selection indicator for antenna $m \in \mathcal{M} \triangleq \{1, \ldots, M\}$, with 1 being the antenna is selected and 0 otherwise. Given antenna selection $\mathbf{a}$, the signal received at user

k in group i is given by

$$y_{ik} = \mathbf{w}_i^H \mathbf{A}\mathbf{h}_{ik} s_i + \sum_{l \in \mathcal{G}_{-i}} \mathbf{w}_l^H \mathbf{A}\mathbf{h}_{ik} s_l + n_{ik},\ k \in \mathcal{K}_i,\ i \in \mathcal{G}.$$

where $\mathbf{A} = \text{diag}(\mathbf{a})$, s_i is the symbol intended for group i, n_{ik} is the receiver additive white Gaussian noise at user k in group i with zero mean and variance σ^2, and $\mathcal{G}_{-i} \triangleq \mathcal{G} \backslash \{i\}$.

In a massive MIMO system, acquiring the instantaneous downlink channel state information (CSI) can be challenging at the BS. Since the channel covariance matrix evolves much slower than instantaneous CSI, one approach in the literature is to estimate the channel covariance matrix from training symbols over time [96–98]. Following this, in this work, we consider that the BS performs antenna selection and beamforming design based on the channel covariance matrix information. Let $\bar{\mathbf{R}}_{ik} \triangleq \mathbb{E}\{\mathbf{h}_{ik}\mathbf{h}_{ik}^H\}$ denote the *true* full channel covariance matrix between the BS and user k in group i. The BS computes SINR of user k in group i based on channel covariance matrix. With antenna selection $\mathbf{a}$, it is given as follows:

$$\text{SINR}_{ik} = \frac{\mathbf{w}_i^H \mathbf{A}\bar{\mathbf{R}}_{ik}\mathbf{A}\mathbf{w}_i}{\sum_{l \in \mathcal{G}_{-i}} \mathbf{w}_l^H \mathbf{A}\bar{\mathbf{R}}_{ik}\mathbf{A}\mathbf{w}_l + \sigma^2}, k \in \mathcal{K}_i,\ i \in \mathcal{G}. \tag{4.1}$$

With antenna selection, the transmit power constraint can be imposed by incorporating antenna selection $\mathbf{a}$. Specifically, we assume that the maximum transmit power at each selected antenna is P_{max} and 0 at unselected antennas. Thus, we have the per antenna power constraint as follows:

$$\sum_{i=1}^{G} |w_{im}|^2 \leq a_m P_{\text{max}}, \quad m \in \mathcal{M}. \tag{4.2}$$

From (4.2), it immediately follows that for $a_m = 0$, $w_{im} = 0, \forall i \in \mathcal{G}$. Under constraint (4.2), we can remove selection matrix $\mathbf{A}$ from the SINR expression in (4.1) without affecting the expression.

Our goal is to jointly determine antenna selection and multicast beamforming vectors for multi-group multicasting to minimize the transmit power, subject to per antenna

transmit power limit and the minimum SINR requirements. This problem is formulated as follows:

$$\mathcal{P}_{\text{perf}}: \min_{\mathbf{a},\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2$$

$$\text{s.t.} \quad \frac{\mathbf{w}_i^H \bar{\mathbf{R}}_{ik} \mathbf{w}_i}{\sum_{l \in \mathcal{G}_{-i}} \mathbf{w}_l^H \bar{\mathbf{R}}_{ik} \mathbf{w}_l + \sigma^2} \geq \gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \tag{4.3a}$$

$$\sum_{i=1}^{G} |w_{im}|^2 \leq a_m P_{\max}, \ m \in \mathcal{M}, \tag{4.3b}$$

$$\sum_{m=1}^{M} a_m = L, \tag{4.3c}$$

$$a_m \in \{0, 1\}, \ m \in \mathcal{M} \tag{4.3d}$$

where γ_{ik} is the minimum SINR target for user k in group i.

4.1.2 Robust Formulation Based on Channel Covariance

In practical systems, the BS only has the estimated channel covariance matrices for the transmission design. Thus, we consider a robust design in antenna selection and multicast beamforming, which accounts the uncertainty of channel covariance matrices in the presence of the estimation error to ensure a robust performance. Let $\mathbf{R}_{ik}$ denote the $M \times M$ *estimated* channel covariance matrix for user k in group i. We model the estimation error in the channel covariance matrix as $\bar{\mathbf{R}}_{ik} = \mathbf{R}_{ik} + \mathbf{E}_{ik}$, where $\mathbf{E}_{ik}$ is the corresponding error matrix. We follow a spherical error model for channel uncertainty commonly considered in the robust designs [86–88, 94, 95]. In particular, we assume $\mathbf{E}_{ik}$ is bounded within a hyper-spherical region as follows:

$$\|\mathbf{E}_{ik}\|_F \leq \epsilon_{ik}, \quad k \in \mathcal{K}_i, i \in \mathcal{G} \tag{4.4}$$

where ϵ_{ik} is the error bound. Let $\mathcal{B}(\epsilon_{ik})$ denote the set of all error matrices satisfying (4.4), given by

$$\mathcal{B}(\epsilon_{ik}) \triangleq \{\mathbf{E}_{ik}: \ \|\mathbf{E}_{ik}\|_F \leq \epsilon_{ik}\}. \tag{4.5}$$

Following the above error model, instead of $\mathcal{P}_{\text{perf}}$, we formulate the joint optimization of antennas selection and robust multi-group multicast beamforming problem where the SINR constraint in (4.3a) is replaced by the worst-case SINR for any $\mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik})$, given as follows:[1]

$$\mathcal{P}_{\text{rob}}: \min_{\mathbf{a},\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2$$

$$\text{s.t.} \quad \min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \geq \gamma_{ik}, k\in\mathcal{K}_i, i\in\mathcal{G}, \tag{4.6}$$

$$\sum_{i=1}^{G}|w_{im}|^2 \leq a_m P_{\text{max}},\ m\in\mathcal{M}, \tag{4.7}$$

$$\sum_{m=1}^{M} a_m = L, \tag{4.8}$$

$$a_m \in \{0,1\},\ m\in\mathcal{M} \tag{4.9}$$

Note that both $\mathcal{P}_{\text{perf}}$ and $\mathcal{P}_{\text{rob}}$ are mixed-integer programming problems due to antenna selection. In $\mathcal{P}_{\text{perf}}$, for given $\mathbf{a}$, the problem becomes a non-convex QCQP problem, which is NP-hard. Compared with $\mathcal{P}_{\text{perf}}$, the robust formation in SINR constraint (4.6) makes the problem even more challenging to solve than $\mathcal{P}_{\text{perf}}$ under the perfect knowledge of channel covariance matrices. Furthermore, the problem is in large-scale for massive MIMO systems with $M \gg 1$. Therefore, it is critical to design an effective and scalable solution with low computational complexity. With this goal, in the following section, we reformulate problem $\mathcal{P}_{\text{rob}}$, and drive a low complexity solution.

[1]We note that for the true channel covariance matrix $\bar{\mathbf{R}}_{ik} \succcurlyeq 0$, there is an implicit condition on $\mathbf{E}_{ik}$ for a given estimated channel covariance $\mathbf{R}_{ik}$, such that $\mathbf{R}_{ik}+\mathbf{E}_{ik} \succcurlyeq 0$. When the error bound ϵ_{ik} is sufficiently small, this condition is satisfied for any $\mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik})$, and $\mathcal{P}_{\text{rob}}$ accurately reflects the exact robust beamforming problem. When ϵ_{ik} is large, the set of possible $\mathbf{E}_{ik}$'s is a subset of $\mathcal{B}(\epsilon_{ik})$. In this case, $\mathcal{P}_{\text{rob}}$ provides an upper bound of the exact robust beamforming problem.

4.2 Joint Antenna Selection and Robust Beamforming

4.2.1 Problem Reformulation

In problem $\mathcal{P}_{\text{rob}}$, we first reformulate the worst-case SINR constraint (4.6) for a given antenna selection vector $\mathbf{a}$ by following the approach proposed in section 3.2.1 for a pure robust multicast beamforming problem. In particular, constraint (4.6) for $k \in \mathcal{K}_i$, $i \in \mathcal{G}$ can be equivalently written as

$$\frac{\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i}{\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l+\sigma^2} \geq \gamma_{ik},\ \forall \mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik}),$$

which is equivalent to

$$\mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l \geq \sigma^2\gamma_{ik}, \forall \mathbf{E}_{ik} \in \mathcal{B}(\epsilon_{ik}). \tag{4.10}$$

The above constraint can be further expressed as the following equivalent constraint:

$$\min_{\mathbf{E}_{ik}\in\mathcal{B}(\epsilon_{ik})} \mathbf{w}_i^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_i - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l^H(\mathbf{R}_{ik}+\mathbf{E}_{ik})\mathbf{w}_l \geq \sigma^2\gamma_{ik}. \tag{4.11}$$

The LHS of constraint (4.11) is a convex optimization problem over $\mathbf{E}_{ik}$, which yields a closed-form solution given by $\mathbf{E}_{ik}^\star = -\epsilon_{ik}\dfrac{\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H}{\left\|\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H\right\|_F}$ as in section 3.2.2. Define $\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_G] \in \mathbb{C}^{M\times G}$. Define $\mathbf{D}_{ik} \in \mathbb{C}^{G\times G}$ as a diagonal matrix with the ith diagonal entry being 1 and the rest being $-\gamma_{ik}$. Then, we have $\mathbf{w}_i\mathbf{w}_i^H - \gamma_{ik}\sum_{l\in\mathcal{G}_{-i}}\mathbf{w}_l\mathbf{w}_l^H = \mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H$. Substituting the expression of $\mathbf{E}_{ik}^\star$ into the LHS of (4.11), we then have the following equivalent constraint to constraint (4.6) for $k \in \mathcal{K}_i, i \in \mathcal{G}$:

$$\text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}. \tag{4.12}$$

For the per antenna power constraint in (4.7), let $\mathbf{e}_m$ denote an $M \times 1$ selection vector with the mth entry being 1 and the rest 0's. Then, $\mathbf{e}_m^T\mathbf{W}$ selects the mth row of $\mathbf{W}$, and we have $\sum_{i=1}^G |w_{im}|^2 = \|\mathbf{e}_m^T\mathbf{W}\|^2$. Based on the above, we transform $\mathcal{P}_{\text{rob}}$ into the following equivalent problem:

$$\mathcal{P}_1 : \min_{\mathbf{a},\mathbf{W}}\ \text{tr}(\mathbf{W}\mathbf{W}^H)$$

$$\text{s.t.} \quad \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \tag{4.13a}$$

$$\|\mathbf{e}_m^T\mathbf{W}\|^2 \leq a_m P_{\text{max}}, \ m \in \mathcal{M}, \tag{4.13b}$$

$$\sum_{m=1}^{M} a_m = L. \tag{4.13c}$$

$$a_m \in \{0,1\}, \ m \in \mathcal{M}. \tag{4.13d}$$

Compared with $\mathcal{P}_{\text{rob}}$, $\mathcal{P}_1$ has a more tractable form where the constraint in (4.13a) is now an explicit function of $\mathbf{W}$. In the rest of the paper, we will focus on solving $\mathcal{P}_1$.

4.2.2 Proposed Approach for Antenna Selection and Robust Multicast Beamforming

Even though constraint (4.13a) is in a more tractable form, $\mathcal{P}_1$ is still a challenging mixed-integer programming problem in large-scale. To address this, we propose our low-complexity approach for antenna selection and robust multicast beamforming. To further simply $\mathcal{P}_1$, we consider replacing the binary variables $\{a_m\}$ by continuous variables in $[0,1]$. In particular, it is shown in [53, Lemma 1] that the following three constraints together are equivalent to the binary constraints (4.13c) and (4.13d) for $a_{m,t}$'s:

$$\sum_{m=1}^{M} a_m^2 = L, \ \sum_{m=1}^{M} a_m = L, \ \ 0 \leq a_m \leq 1, m \in \mathcal{M}. \tag{4.14a}$$

Based on the above, we can equivalently replace constraints (4.13c) and (4.13d) by the set of three constraints in (4.14a) and convert the mixed-integer programming problem into the following equivalent non-convex optimization problem with continuous variables:

$$\mathcal{P}_2: \ \min_{\mathbf{a},\mathbf{W}} \text{tr}(\mathbf{W}\mathbf{W}^H)$$

$$\text{s.t.} \quad \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \tag{4.14b}$$

$$\|\mathbf{e}_m^T\mathbf{W}\|^2 \leq a_m P_{\text{max}}, \ m \in \mathcal{M}, \tag{4.14c}$$

$$\sum_{m=1}^{M} a_m = L, \tag{4.14d}$$

$$\sum_{m=1}^{M} a_m^2 = L, \tag{4.14e}$$

$$0 \leq a_m \leq 1,\ m \in \mathcal{M}. \tag{4.14f}$$

Note that even though $\mathcal{P}_2$ are now with all continuous variables, it is still a non-convex and NP-hard problem. To further make the problem more tractable, we relax $\mathcal{P}_2$ by transferring the equality constraint (4.14e) to the objective function as a penalty term with penalty parameter $\zeta > 0$, as shown below:

$$\begin{aligned}
\mathcal{P}_3 : \min_{\mathbf{a},\mathbf{W}}\ & \text{tr}(\mathbf{W}\mathbf{W}^H) + \zeta|\mathbf{a}^T\mathbf{a} - L| \\
\text{s.t.}\ & \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) - \epsilon_{ik}\|\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, && \text{(4.15a)} \\
& \|\mathbf{e}_m^T\mathbf{W}\|^2 \leq a_m P_{\max},\ m \in \mathcal{M}, && \text{(4.15b)} \\
& \mathbf{1}^T\mathbf{a} = L, && \text{(4.15c)} \\
& \mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1} && \text{(4.15d)}
\end{aligned}$$

where the constraints on $\mathbf{a}$ are now shown in the vector forms.

Note that $\mathcal{P}_3$ is relaxed as compare with $\mathcal{P}_2$, where the selection vector $\mathbf{a}$ is no longer guaranteed to be binary. Considering this relation, we propose a two-phase approach:[2]

- **Phase 1 – Antenna selection**: We first determine antenna selection by solving $\mathcal{P}_3$ for $(\mathbf{a}, \mathbf{W})$ jointly. Note that since $\mathcal{P}_3$ is a relaxed problem, the solution $\mathbf{a}$ may not be binary, and as a result, the beamforming solution $\mathbf{W}$ may not be feasible to original problem $\mathcal{P}_2$. Our purpose in this phase is to determine the binary antenna selection vector $\mathbf{a}$ efficiently. Towards this, we propose two approaches:
  - We first propose an SINR-based approach and develop a fast algorithm to obtain $\mathbf{a}$.
  - To further reduce the computational complexity, we then propose an SLR-based approach with a fast algorithm to determine $\mathbf{a}$.

[2]Note that one may consider using the alternating optimization approach to solve $\mathbf{W}$ and $\mathbf{a}$ iteratively. However, $\mathcal{P}_3$ is non-convex w.r.t. $\mathbf{W}$, and the alternating procedure dose not guarantee to converge. Also, we need to make a choice to map solution $\mathbf{a}$ to binary variables of 0 and 1, without violating any constraints, which is challenging to do.

- **Phase 2 – Robust multicast beamforming solution**: After obtaining $\mathbf{a}$ in Phase 1, we solve the robust multicast beamforming problem in $\mathcal{P}_1$ for $\mathbf{W}$ with the selected antennas.

4.3 Antenna Selection: SINR-Based Approach

In Phase 1, we determine antenna selection $\mathbf{a}$ based on $\mathcal{P}_3$. Since the problem is large-scale with $M \gg 1$, we need to develop an algorithm to compute $\mathbf{a}$ efficiently. The main difficulty in solving $\mathcal{P}_3$ is in the non-convex constraint (4.15a), which is the worst-case SINR in the robust formulation. In particular, the Frobenius norm $\|\cdot\|_F$ w.r.t. $\mathbf{W}$ reflects the impact of maximum covariance matrix error ϵ_{ik} from (4.4) on SINR. This term complicates the algorithm design to solve $\mathcal{P}_3$ efficiently.

Since our main goal in this phase is to determine antenna selection $\mathbf{a}$, while computing $\mathbf{W}$ in this phase is to facilitate the determination of $\mathbf{a}$, we propose to simplify $\mathcal{P}_3$ for a more tractable problem. Specifically, we ignore the second term in (4.15a), which accounts for robustness consideration, and only use first term. In other words, we ignore the estimation error and express SINR under the estimated channel covariance matrices $\{\mathbf{R}_{ik}\}$ to find antenna selection vector $\mathbf{a}$. This leads to the following modified problem from $\mathcal{P}_3$:

$$
\begin{aligned}
\mathcal{P}_3^{\text{SINR}}: \ & \min_{\mathbf{a},\mathbf{W}} \ \text{tr}(\mathbf{W}\mathbf{W}^H) + \zeta|\mathbf{a}^T\mathbf{a} - L| \\
\text{s.t.} \ & \ \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \\
& \ \|\mathbf{e}_m^T\mathbf{W}\|^2 \leq a_m P_{\text{max}}, \ m \in \mathcal{M}, \\
& \ \mathbf{1}^T\mathbf{a} = L, \\
& \ \mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}
\end{aligned}
$$

Although $\mathcal{P}_3^{\text{SINR}}$ is still non-convex, it is more amenable to solve. ADMM is a robust and fast numerical method to solve large-scale problems. We design an ADMM-based fast algorithm to directly solve $\mathcal{P}_3^{\text{SINR}}$. The convergence of ADMM for a wide range of non-convex problems is recently established in [51]. Following this, we develop an ADMM-based algorithm to solve $\mathcal{P}_3^{\text{SINR}}$ with low computational complexity.

The ADMM-Based Algorithm for $\mathcal{P}_3^{SINR}$: To construct ADMM blocks, we introduce auxiliary variable $\mathbf{Y} \in \mathbb{C}^{M\times G}$ and transfer $\mathcal{P}_3^{\text{SINR}}$ to the following equivalent problem:

$$\begin{aligned}
\mathcal{P}_{\text{ADMM}}^{\text{SINR}}: \min_{\mathbf{a},\mathbf{W},\mathbf{Y}}\ & \text{tr}(\mathbf{W}\mathbf{W}^H) + \zeta|\mathbf{a}^T\mathbf{a} - L| \\
\text{s.t. }\ & \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, && \text{(4.16a)}\\
& \|\mathbf{e}_m^T\mathbf{Y}\|^2 \leq a_m P_{\max}, m \in \mathcal{M}, && \text{(4.16b)}\\
& \mathbf{1}^T\mathbf{a} = L, && \text{(4.16c)}\\
& \mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}, && \text{(4.16d)}\\
& \mathbf{Y} = \mathbf{W}. && \text{(4.16e)}
\end{aligned}$$

We denote $\mathcal{C}$ the feasible set of $(\mathbf{a},\mathbf{W},\mathbf{Y})$ that satisfies constraints (4.16a)–(4.16d) in $\mathcal{P}_{\text{ADMM}}$ and define the indicator function for $\mathcal{C}$ as

$$\mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y}) = \begin{cases} 0, & \text{if } (\mathbf{a},\mathbf{W},\mathbf{Y}) \in \mathcal{C} \\ \infty, & \text{otherwise.} \end{cases} \tag{4.17}$$

Then, by transferring the constraints (4.16a)–(4.16d) into the objective function in $\mathcal{P}_{\text{ADMM}}$, we arrive at the following equivalent problem

$$\begin{aligned}
\min_{\mathbf{a},\mathbf{W},\mathbf{Y}}\ & \text{tr}(\mathbf{W}\mathbf{W}^H) + \zeta|\mathbf{a}^T\mathbf{a} - L| + \mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y}) && \text{(4.18)}\\
\text{s.t. }\ & \mathbf{Y} = \mathbf{W}.
\end{aligned}$$

The augmented Lagrangian for problem (4.18) is given by

$$\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z}) = \text{tr}(\mathbf{W}\mathbf{W}^H) + \zeta|\mathbf{a}^T\mathbf{a} - L| + \mathbb{I}_{\mathcal{C}}(\mathbf{a},\mathbf{W},\mathbf{Y}) + \rho\|\mathbf{W} - \mathbf{Y} + \mathbf{Z}\|_F^2. \tag{4.19}$$

where $\mathbf{Z} \in \mathbb{C}^{M\times G}$ is the dual variable associated with constraint (4.16e), and ρ is penalty parameter.

With the auxiliary variable $\mathbf{Y}$ and constraint (4.16e), we now can minimize $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})$ w.r.t. $\mathbf{a}$, $\mathbf{W}$, and $\mathbf{Y}$ separately in an iterative fashion. Finally, we round the elements in $\mathbf{a}$ to the nearest 0 or 1 for antenna selection. Our proposed ADMM-based algorithm is summarized in Algorithm 3. The algorithm consists of two main updating blocks. In the first two blocks, we minimize $\mathcal{L}_\rho(\mathbf{a},\mathbf{W},\mathbf{Y},\mathbf{Z})$ w.r.t. $(\mathbf{a},\mathbf{W})$ and $\mathbf{Y}$, respectively, to obtain

Algorithm 3 The ADMM-Based Algorithm for $\mathcal{P}_3^{\text{SINR}}$

Initialization: Set ρ; Set initial $\mathbf{W}^0$, $\mathbf{a}^0$, $\mathbf{Z}^0 = \mathbf{0}$; Set $l = 0$.

repeat

1) Update beamforming matrix $\mathbf{W}^{l+1}$, $\mathbf{a}^{l+1}$:

$$(\mathbf{W}^{l+1}, \mathbf{a}^{l+1}) = \arg\min_{\mathbf{a},\mathbf{W}} \mathcal{L}_\rho\big(\mathbf{a}, \mathbf{W}, \mathbf{Y}^l, \mathbf{Z}^l\big). \tag{4.20}$$

2) Update the auxiliary matrices $\mathbf{Y}^{l+1}$:

$$\mathbf{Y}^{l+1} = \arg\min_{\mathbf{Y}} \mathcal{L}_\rho\big(\mathbf{a}^{l+1}, \mathbf{W}^{l+1}, \mathbf{Y}, \mathbf{Z}^l\big). \tag{4.21}$$

3) Update dual variables $\mathbf{Z}^{l+1}$:

$$\mathbf{Z}^{l+1} = \mathbf{Z}^l + \mathbf{Y}^{l+1} - \mathbf{W}^{l+1}. \tag{4.22}$$

4) Set $l \leftarrow l + 1$.

until convergence

Obtain $\mathbf{a}$: $a_m = \text{round}(a_m^l)$, $m \in \mathcal{M}$.

the updates. We will show that the solutions to these two optimization subproblems can be computed efficiently using the semi-closed-form expressions.

For the first block in Algorithm 3, given $(\mathbf{Y}^l, \mathbf{Z}^l)$ in iteration l, the joint optimization subproblem (4.20) for $(\mathbf{a}, \mathbf{W})$ can be further decoupled into two subproblems for $\mathbf{W}$ and $\mathbf{a}$ separately. This can be seen from $\mathcal{L}_\rho(\mathbf{a}, \mathbf{W}, \mathbf{Y}, \mathbf{Z})$ in (4.19), where the terms contain $\mathbf{W}$ or $\mathbf{a}$ separately. Thus, in the following subsections, we will first describe the solution for updating $\mathbf{W}$ in Section 4.3.1, and then we derive a closed-form solution for updating $\mathbf{a}$ in Section 4.3.2. The solution to subproblem (4.21) for updating $\mathbf{Y}$ is described in Section 4.3.3.

4.3.1 Update W

Given $(\mathbf{Y}^l, \mathbf{Z}^l)$ in iteration l, in the subproblem (4.20), we first minimize $\mathcal{L}_\rho(\mathbf{a}, \mathbf{W}, \mathbf{Y}, \mathbf{Z})$ w.r.t. $\mathbf{W}$. After removing the terms in (4.19) that are irrelevant to $\mathbf{W}$, the problem is equivalent to

$$\min_{\mathbf{W}} \text{tr}(\mathbf{W}\mathbf{W}^H) + \rho\|\mathbf{W} - \mathbf{Y}^l + \mathbf{Z}^l\|_F^2 \tag{4.23}$$

$$\text{s.t. } \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) \geq \sigma^2\gamma_{ik},\ k \in \mathcal{K}_i, i \in \mathcal{G}.$$

Subproblem (4.23) is a non-convex NP-hard problem.[3] To obtain a good solution that is also computationally efficient, we propose an inner-layer consensus-ADMM-based algorithm for subproblem (4.23). The approach for ADMM construction is similar to in Algorithm 3 for $\mathcal{P}^{\text{SINR}}_{\text{ADMM}}$. First, after expanding the second term of the objective function in (4.23) and removing the constant terms, problem (4.23) is equivalent to the following problem

$$\begin{aligned}\mathcal{P}_{\mathbf{W}}: \min_{\mathbf{W}}\ & (\rho+1)\text{tr}(\mathbf{W}\mathbf{W}^H) + 2\rho\mathfrak{Re}\{\text{tr}(\mathbf{B}^l\mathbf{W}^H)\} \\ \text{s.t. }\ & \text{tr}(\mathbf{R}_{ik}\mathbf{W}\mathbf{D}_{ik}\mathbf{W}^H) \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}.\end{aligned}$$

where $\mathbf{B}^l \triangleq -\mathbf{Y}^l + \mathbf{Z}^l$.

Introducing auxiliary variables $\mathbf{V}_{ik} \in \mathbb{C}^{M\times G}$ and $\mathbf{W}_{ik} \in \mathbb{C}^{M\times G}$, for $k \in \mathcal{K}_i$, $i \in \mathcal{G}$, we have the following equivalent problem to $\mathcal{P}_{\mathbf{W}}$:

$$\begin{aligned}\mathcal{P}^{\text{ADMM}}_{\mathbf{W}}: \min_{\mathbf{W},\{\mathbf{V}_{ik},\mathbf{W}_{ik}\}}\ & (\rho+1)\text{tr}(\mathbf{W}\mathbf{W}^H) + 2\rho\mathfrak{Re}\{\text{tr}(\mathbf{B}^l\mathbf{W}^H)\} \\ \text{s.t. }\ & \mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H)\} \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. && \text{(4.24a)} \\ & \mathbf{W}_{ik} = \mathbf{W}, k \in \mathcal{K}_i, i \in \mathcal{G}. && \text{(4.24b)} \\ & \mathbf{V}_{ik} = \mathbf{W}\mathbf{D}_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}. && \text{(4.24c)}\end{aligned}$$

Denote the feasibility set of $\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}$ satisfying constraints in (4.24a) as $\mathcal{F}$, and define the indicator function $\mathbb{I}_{\mathcal{F}}(\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}) \triangleq \{0 : \text{if } \{\mathbf{V}_{ik}, \mathbf{W}_{ik}\} \in \mathcal{F}; \infty : \text{ otherwise}\}$. We transfer $\mathcal{P}^{\text{ADMM}}_{\mathbf{W}}$ to the following equivalent problem with only linear equality constraints:

$$\begin{aligned}\min_{\mathbf{W},\{\mathbf{V}_{ik},\mathbf{W}_{ik}\}}\ & (\rho+1)\text{tr}(\mathbf{W}\mathbf{W}^H) + 2\rho\mathfrak{Re}\{\text{tr}(\mathbf{B}^l\mathbf{W}^H)\} + \mathbb{I}_{\mathcal{F}}\big(\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}\big) && \text{(4.25)} \\ \text{s.t. }\ & \mathbf{W}_{ik} = \mathbf{W}, k \in \mathcal{K}_i, i \in \mathcal{G}. \\ & \mathbf{V}_{ik} = \mathbf{W}\mathbf{D}_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}.\end{aligned}$$

The augmented Lagrangian to problem (4.25) is given by

$$\tilde{\mathcal{L}}_{\tilde{\rho}}\big(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\}\big) = (\rho+1)\text{tr}(\mathbf{W}\mathbf{W}^H) + 2\rho\mathfrak{Re}\{\text{tr}(\mathbf{B}^l\mathbf{W}^H)\}$$

[3]The constraint in (4.23) is similar to those in the multicast beamforming QoS problems.

Algorithm 4 Consensus-ADMM-based Algorithm for $\mathcal{P}_{\mathbf{W}}$

Initialization: Set $\tilde{\rho}$; Set initial $\mathbf{W}^0$, $\mathbf{a}^0$, $\mathbf{S}_{ik}^0 = \mathbf{0}$, $\mathbf{T}_{ik}^0 = \mathbf{0}$, $\forall k, i$; Set $j = 0$.

repeat

1) Update the auxiliary matrices $\{\mathbf{W}_{ik}^{j+1}\}$:

$$\{\mathbf{W}_{ik}\}^{j+1} = \arg\min_{\{\mathbf{W}_{ik}\}} \tilde{\mathcal{L}}_{\rho}\big(\mathbf{W}^j, \{\mathbf{V}_{ik}^j, \mathbf{W}_{ik}, \mathbf{S}_{ik}^j, \mathbf{T}_{ik}^j\}\big). \tag{4.27}$$

2) Update the auxiliary matrices $\{\mathbf{V}_{ik}^{l+1}\}$:

$$\{\mathbf{V}_{ik}\}^{j+1} = \arg\min_{\{\mathbf{V}_{ik}\}} \tilde{\mathcal{L}}_{\tilde{\rho}}\big(\mathbf{W}^j, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}^{j+1}, \mathbf{S}_{ik}^j, \mathbf{T}_{ik}^j\}\big). \tag{4.28}$$

3) Update beamforming matrix $\mathbf{W}^{j+1}$:

$$\mathbf{W}^{j+1} = \arg\min_{\mathbf{W}} \tilde{\mathcal{L}}_{\tilde{\rho}}\big(\mathbf{W}, \{\mathbf{V}_{ik}^{j+1}, \mathbf{W}_{ik}^j, \mathbf{S}_{ik}^j, \mathbf{T}_{ik}^j\}\big). \tag{4.29}$$

4) Update dual variables $\mathbf{S}_{ik}^{j+1}$, $\mathbf{T}_{ik}^{j+1}$:

$$\mathbf{S}_{ik}^{j+1} = \mathbf{S}_{ik}^j + \mathbf{W}^{j+1} - \mathbf{W}_{ik}^{j+1}, \tag{4.30}$$

$$\mathbf{T}_{ik}^{j+1} = \mathbf{T}_{ik}^j + \mathbf{W}^{j+1}\mathbf{D}_{ik} - \mathbf{V}_{ik}^{j+1}. \tag{4.31}$$

5) Set $j \leftarrow j + 1$.

until convergence

$$+ \mathbb{I}_{\mathcal{F}}\big(\{\mathbf{V}_{ik}, \mathbf{W}_{ik}\}\big) + \tilde{\rho} \sum_{i=1}^{G} \sum_{k=1}^{K} \|\mathbf{W}\mathbf{D}_{ik} - \mathbf{V}_{ik} + \mathbf{T}_{ik}\|_F^2 +$$

$$\tilde{\rho} \sum_{i=1}^{G} \sum_{k=1}^{K} \|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{S}_{ik}\|_F^2. \tag{4.26}$$

where $\{\mathbf{S}_{ik}\}$ and $\{\mathbf{T}_{ik}\}$ are the dual variables associated with the constraints in (4.24b) and (4.24c), respectively, and $\bar{\rho}$ is the penalty parameter.

Based on the above, we construct three ADMM blocks to update these variables iteratively by minimizing $\tilde{\mathcal{L}}_{\tilde{\rho}}\big(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\}\big)$ w.r.t. $\{\mathbf{W}_{ik}\}$, $\{\mathbf{V}_{ik}\}$, and $\mathbf{W}$, respectively, to solve $\mathcal{P}_{\mathbf{W}}^{\text{ADMM}}$. This consensus-ADMM-based algorithm is summarized in Algorithm 4. The three updating blocks involve solving optimization problems (4.27), (4.28), and (4.29), respectively. Below, we develop a closed-form solution for each of these three subproblems. For notation simplicity, we remove the superscript for iteration index, with the understanding that the solutions are used for updates in iteration j.

Update $\mathbf{W}_{ik}$

We first fix $(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\})$ and minimize $\tilde{\mathcal{L}}_{\tilde{\rho}}\big(\mathbf{W}, \{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\}$ w.r.t. $\{\mathbf{W}_{ik}\}$ in (4.27). From (4.26), we see that the problem can be decomposed into $\sum_{i=1}^{G} K_i$ separate subproblems, one for each k and i:

$$\begin{aligned}\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}} : \min_{\mathbf{W}_{ik}} \ & \|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{S}_{ik}\|_F^2. \\ \text{s.t. } & \mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H)\} \geq \sigma^2\gamma_{ik}.\end{aligned}$$

Since $\mathcal{P}_{\mathbf{W}_{ik}}$ is a convex problem with one linear constraint, we can obtain its solution in closed-form. The optimal $\mathbf{W}_{ik}^\star$ is one of the following two solutions: i) $\mathbf{W}_{ik}^\star = \mathbf{S}_{ik} + \mathbf{W}$: it is the optimal solution to $\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}}$ if it satisfies the constraint. Otherwise: ii) The constraint is active. We solve $\mathcal{P}_{\mathbf{W}_{ik}}$ via the KKT conditions. The Lagrangian for $\mathcal{P}_{\mathbf{W}_{ik}}$ is given by

$$\mathcal{L}(\mathbf{W}_{ik}, \lambda_{ik}) = \|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{S}_{ik}\|_F^2 + \lambda_{ik}(-\mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H)\} + \sigma^2\gamma_{ik}).$$

where λ_{ik} is the Lagrangian multiplier associated with the constraint. Set the gradient w.r.t. $\mathbf{W}_{ik}^H$ to zero, *i.e.*, $\nabla_{\mathbf{W}_{ik}^*}\mathcal{L}(\mathbf{W}_{ik}, \lambda_{ik}) = 0$, we obtain the optimal $\mathbf{W}_{ik}^\star$ as

$$\mathbf{W}_{ik}^\star = \mathbf{W} + \mathbf{S}_{ik} + \frac{1}{2}\lambda_{ik}\mathbf{R}_{ik}\mathbf{V}_{ik}. \tag{4.32}$$

Substitute the expression of $\mathbf{W}_{ik}^\star$ in (4.32) into the constraint with equality: $\mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H)\} - \sigma^2\gamma_{ik} = 0$, we obtain the optimal $\lambda_{ik}^\star$:

$$\lambda_{ik}^\star = \frac{2\big(-\mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}(\mathbf{W} + \mathbf{S}_{ik})^H)\} + \sigma^2\gamma_{ik}\big)}{\|\mathbf{R}_{ik}\mathbf{V}_{ik}\|_F^2}. \tag{4.33}$$

Note that $\lambda_{ik}^\star \geq 0$ in (4.33); otherwise, the solution in i) holds.

Update $\mathbf{V}_{ik}$

To update $\mathbf{V}_{ik}$ for fixed $(\mathbf{W}, \{\mathbf{S}_{ik}, \mathbf{T}_{ik}\})$, the optimization problem (4.28) can again be decomposed into $\sum_{i=1}^{G} K_i$ separate subproblems, one for each k and i:

$$\begin{aligned}\mathcal{P}^{\text{in}}_{\mathbf{V}_{ik}} : \min_{\mathbf{V}_{ik}} \ & \|\mathbf{W}\mathbf{D}_{ik} - \mathbf{V}_{ik} + \mathbf{T}_{ik}\|_F^2. \\ \text{s.t. } & \mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}\mathbf{V}_{ik}\mathbf{W}_{ik}^H)\} \geq \sigma^2\gamma_{ik}.\end{aligned}$$

The form of $\mathcal{P}^{\text{in}}_{\mathbf{V}_{ik}}$ is the same as $\mathcal{P}^{\text{in}}_{\mathbf{W}_{ik}}$. Thus, we directly have the solution as follows:

i) $\mathbf{V}^{\star}_{ik} = \mathbf{W}\mathbf{D}_{ik} + \mathbf{T}_{ik}$: it is the optimal solution to $\mathcal{P}^{\text{in}}_{\mathbf{V}_{ik}}$ if it satisfies the constraint; otherwise:

ii) The constraint is active, and we have

$$\mathbf{V}^{\star}_{ik} = \mathbf{W}\mathbf{D}_{ik} + \mathbf{T}_{ik} + \frac{1}{2}\tilde{\lambda}^{\star}_{ik}\mathbf{R}^{H}_{ik}\mathbf{W}_{ik} \tag{4.34}$$

where $\tilde{\lambda}^{\star}_{ik} = \dfrac{2(-\mathfrak{Re}\{\text{tr}(\mathbf{R}_{ik}(\mathbf{W}\mathbf{D}_{ik} + \mathbf{T}_{ik})\mathbf{W}^{H}_{ik})\} + \sigma^2\gamma_{ik})}{\|\mathbf{W}^{H}_{ik}\mathbf{R}_{ik}\|^2_F}$.

Update W

Problem (4.29) to solve $\mathbf{W}$ for fixed $\{\mathbf{V}_{ik}, \mathbf{W}_{ik}, \mathbf{S}_{ik}, \mathbf{T}_{ik}\}$ is an unconstrained quadratic optimization problem, given by

$$\min_{\mathbf{W}} (\rho+1)\text{tr}(\mathbf{W}\mathbf{W}^H) + 2\rho\mathfrak{Re}\{\text{tr}(\mathbf{B}^l\mathbf{W}^H)\} + \tilde{\rho}\sum_{i=1}^{G}\sum_{k=1}^{K}\|\mathbf{W}\mathbf{D}_{ik} - \mathbf{V}_{ik} + \mathbf{T}_{ik}\|^2_F + \tilde{\rho}\sum_{i=1}^{G}\sum_{k=1}^{K}\|\mathbf{W} - \mathbf{W}_{ik} + \mathbf{S}_{ik}\|^2_F.$$

Taking the derivative of the objective function and setting it to zero, we obtain

$$\mathbf{W}^{\star} = (\bar{\rho}\mathbf{C} - \rho\mathbf{B}^l)\Big(\tilde{\rho}\sum_{i=1}^{G}\sum_{k=1}^{K}\mathbf{D}^2_{ik} + \big(\tilde{\rho}\sum_{i=1}^{G}K_i + \rho + 1\big)\mathbf{I}\Big)^{-1}. \tag{4.35}$$

where $\mathbf{C} \triangleq \sum_{i=1}^{G}\sum_{k=1}^{K}(\mathbf{W}_{ik} - \mathbf{S}_{ik} + (\mathbf{V}_{ik} - \mathbf{T}_{ik})\mathbf{D}_{ik})$.

4.3.2 Update a

For subproblem (4.20), we now minimize $\mathcal{L}_{\rho}(\mathbf{a}, \mathbf{W}, \mathbf{Y}^l, \mathbf{Z}^l)$ w.r.t. $\mathbf{a}$. After removing the terms in (4.19) that are not related to $\mathbf{a}$, we have the following equivalent problem:

$$\min_{\mathbf{a}} \zeta|\mathbf{a}^T\mathbf{a} - L| \tag{4.36a}$$

$$\text{s.t. } \|\mathbf{e}^T_m\mathbf{Y}\|^2 \le a_m P_{\max},\ m \in \mathcal{M}, \tag{4.36b}$$

$$\mathbf{1}^T\mathbf{a} = L, \tag{4.36c}$$

$$\mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}. \tag{4.36d}$$

We further reformulate the above problem. First, note that by (4.36c) and (4.36d), we have $\mathbf{a}^T\mathbf{a} \leq \mathbf{1}^T\mathbf{a} = L$. Thus, minimizing the objective function in (4.36a) is equivalent to maximizing $\mathbf{a}^T\mathbf{a}$. Next, we can combine constraints (4.36b) and (4.36d) to determine the lower bound on a_m. The resulting equivalent problem to problem (4.36) is given by

$$\begin{aligned} \mathcal{P}_{\mathbf{a}} : \max_{\mathbf{a}} \ & \mathbf{a}^T\mathbf{a} \\ \text{s.t. } & \mathbf{p} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}, \\ & \mathbf{1}^T\mathbf{a} = L. \end{aligned}$$

where $\mathbf{p} \triangleq [p_1, \ldots, p_M]^T$ with $p_m \triangleq \frac{\|\mathbf{e}_m^T\mathbf{Y}\|^2}{P_{\max}}$. Note that $p_m \leq 1$ for $\mathcal{P}_{\mathbf{a}}$ being feasible.

Since $\mathcal{P}_{\mathbf{a}}$ contains only linear constraints, it meets the linearity constraint qualification (LCQ), one of the regularity conditions under which the KKT conditions is necessary for optimality [129]. Thus, we use the KKT conditions to derive the solution to $\mathcal{P}_{\mathbf{a}}$. In particular, we show there is a unique solution satisfying the KKT conditions, and thus it is the optimal solution to $\mathcal{P}_{\mathbf{a}}$. The Lagrangian for $\mathcal{P}_{\mathbf{a}}$ is

$$\mathcal{L}(\mathbf{a}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \mu) = -\mathbf{a}^T\mathbf{a} + \boldsymbol{\lambda}_1^T(\mathbf{p} - \mathbf{a}) + \boldsymbol{\lambda}_2^T(\mathbf{a} - \mathbf{1}) + \mu(\mathbf{1}^T\mathbf{a} - L) \tag{4.37}$$

where $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$, and μ are the Lagrangian multipliers associated with the constraints in $\mathcal{P}_{\mathbf{a}}$, respectively. The KKT conditions are listed below:

$$\begin{aligned} &\text{C1} : \nabla_{\mathbf{a}}\mathcal{L}(\mathbf{a}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \mu) = \mathbf{0} \Rightarrow -2\mathbf{a} - \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 + \mu\mathbf{1} = \mathbf{0}. \\ &\text{C2} : \boldsymbol{\lambda}_1 \odot (\mathbf{p} - \mathbf{a}) = \mathbf{0}, \\ &\text{C3} : \boldsymbol{\lambda}_2 \odot (\mathbf{a} - \mathbf{1}) = \mathbf{0}, \\ &\text{C4} : \mathbf{p} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}, \\ &\text{C5} : \mathbf{1}^T\mathbf{a} - L = 0. \end{aligned}$$

From C2–C4, we have the following three situations for a_m, $m \in \mathcal{M}$:

S1) If $\lambda_{2m} \neq 0$, then $a_m = 1$ and $\lambda_{1m} = 0$.

S2) If $p_m < a_m < 1$, then $\lambda_{1m} = \lambda_{2m} = 0$; and by C1, $a_m = \mu/2 \triangleq \bar{a}$, where $\bar{a} \in (p_m, 1)$.

S3) If $\lambda_{1m} \neq 0$, then $a_m = p_m$, and $\lambda_{2m} = 0$.

From the above three possible values for a_m, we divide $\{a_m\}$ into three sets. Define the following index sets: $\mathcal{S}_1 \triangleq \{m : \ a_m = 1, m \in \mathcal{M}\}$, $\mathcal{S}_2 = \{m : \ a_m = \bar{a}, m \in \mathcal{M}\}$, $\mathcal{S}_3 = \{m : \ a_m = p_m, m \in \mathcal{M}\}$. Also, denote $N_1 \triangleq |\mathcal{S}_1|$ and $N_2 \triangleq |\mathcal{S}_2|$. We have $|\mathcal{S}_3| = M - N_1 - N_2$. With these three sets, C5 can be rewritten as

$$N_1 + N_2\bar{a} + \sum_{m \in \mathcal{S}_3} p_m = L. \tag{4.38}$$

Similarly, we also rewrite the objective function in $\mathcal{P}_{\mathbf{a}}$ as

$$\mathbf{a}^T\mathbf{a} = N_1 + N_2\bar{a}^2 + \sum_{m \in \mathcal{S}_3} p_m^2. \tag{4.39}$$

Thus, to solve $\mathcal{P}_{\mathbf{a}}$, we need to determine N_1 and N_2, *i.e.*, the sizes of these sets, and the value $\bar{a}$ for $\mathcal{S}_2$.

Note that in general, if $\sum_{m \in \mathcal{S}_3} p_m \notin \mathbb{Z}^+$, then for (4.38) holds, we must have $N_2 \geq 1$ ($\mathcal{S}_2 \neq \emptyset$). Thus, we have the following two possible cases for $\mathcal{S}_2$: 1) $N_2 \geq 1$; 2) $N_2 = 0$. In what follows, we first focus on $N_2 \geq 1$, which is the most likely case.

Case $N_2 \geq 1$

First, assume that $\mathcal{S}_3$ is given. Let $z \triangleq N_1 + N_2$, where $z > N_1$ for $N_2 \geq 1$. Also, since $\mathcal{S}_3$ is given, z is fixed. From (4.38), we have

$$N_1 + N_2\bar{a} = L - \sum_{m \in \mathcal{S}_3} p_m \triangleq L' \tag{4.40}$$

where L' is fixed for given $\mathcal{S}_3$. Then, we have $\bar{a} = \frac{L' - N_1}{z - N_1}$. From (4.39), let $y \triangleq N_1 + N_2\bar{a}^2$. Replacing $\bar{a}$ with the expression above, we can write y as a function of N_1 as

$$y(N_1) = N_1 + \frac{(L' - N_1)^2}{z - N_1}. \tag{4.41}$$

Following this, the objective function is given by

$$\mathbf{a}^T\mathbf{a} = y(N_1) + \sum_{m \in \mathcal{S}_3} p_m^2 \tag{4.42}$$

where the second term at the right hand side (RHS) is fixed for given $\mathcal{S}_3$. Thus, $\mathcal{P}_{\mathbf{a}}$ is equivalent to

$$\max_{N_1:N_1<z} y(N_1), \tag{4.43}$$

for which we only need to check the difference in $y(N_1)$ w.r.t. N_1. Consider $N_1+1<z$. From (4.41), we have

$$y(N_1+1)-y(N_1)=\frac{(z-L')^2}{(z-N_1)(z-N_1-1)}>0. \tag{4.44}$$

Thus, for any given $z>0$, $y(N_1)$ is an increasing function of N_1, and therefore for (4.43), N_1 should be maximized. This leads to $N_1^\star=z-1$, and $N_2^\star=1$, for given $z>0$. Thus, we obtain $N_2^\star$ for $\mathcal{S}_2$, *i.e.*, only one $a_m=\bar{a}$ with inactive inequality constraints.

Next, we select $\mathcal{S}_3$ and determine $N_1^\star$. Note that for $N_2^\star=1$, the objective function in (4.39) becomes

$$\mathbf{a}^T\mathbf{a}=N_1+\bar{a}^2+\sum_{m\in\mathcal{S}_3}p_m^2\triangleq g(N_1,\mathbf{p}) \tag{4.45}$$

where $|\mathcal{S}_3|=M-N_1-1$. Now we examine $g(N_1,\mathbf{p})$ w.r.t. N_1. By (4.40), we have

$$N_1=L-\big(\bar{a}+\sum_{m\in\mathcal{S}_3}p_m\big). \tag{4.46}$$

Substituting the above expression into $g(N_1,\mathbf{p})$, we have

$$g(N_1,\mathbf{p})=L-\bar{a}+\bar{a}^2-\sum_{m\in\mathcal{S}_3}p_m+\sum_{m\in\mathcal{S}_3}p_m^2.$$

Note that increasing N_1 to N_1+1 means removing an element from $\mathcal{S}_3$ and reducing $|\mathcal{S}_3|$ by one. Let m' be the entry that is removed from $\mathcal{S}_3$. Then, we have

$$g(N_1+1,\mathbf{p})-g(N_1,\mathbf{p})=p_{m'}-p_{m'}^2\geq 0, \tag{4.47}$$

where the last inequality is due to $0\leq p_m\leq 1$, $\forall m$. Thus, with $\bar{a}$ given, $g(N_1,\mathbf{p})$ is an increasing function of N_1, and maximizing N_1 leads to the maximum objective value $\mathbf{a}^T\mathbf{a}$ in (4.45).

Now consider N_1 in (4.46) obtained from the equality constraint C5. For any given size $|\mathcal{S}_3|$, N_1 in (4.46) is maximized by choosing elements in $\mathcal{S}_3$ whose p_m's have the lowest values. Hence, for any N_1, $\mathcal{S}_3$ can be determined as follows. Sort $\{p_m\}$ in the ascending order: $p_{m_1} \leq p_{m_2} \leq \cdots \leq p_{m_M}$. Then, $\mathcal{S}_3 = \{m_1, m_2, \ldots, m_{M-N_1-1}\}$ and

$$N_1 = L - \left(\bar{a} + \sum_{j=1}^{M-N_1-1} p_{m_j}\right). \tag{4.48}$$

To determine $N_1^\star$, we recall that $\bar{a} < 1$. From (4.48), we have the following bounds: $L - 1 < N_1^\star + \sum_{j=1}^{M-N_1^\star-1} p_{m_j} < L$. This leads to a simple search for the optimal $N_1^\star$: Starting from $N_1 = L$, we reduce N_1 until the above inequalities are met.

Once $N_1^\star$ is obtained, the optimal $\bar{a}^\star$ is given by

$$\bar{a}^\star = L - N_1^\star - \sum_{j=1}^{M-N_1^\star-1} p_{m_j}. \tag{4.49}$$

Case $N_2 = 0$

In the above, for general $\mathbf{p}$, we have assumed $N_2 \geq 1$ in (4.40) and obtain $N_2^\star = 1$. In the less likely case, it is possible that $N_2 = 0$ for some $\mathcal{S}_3$, where $\sum_{m\in\mathcal{S}_3} p_m \in \mathbb{Z}^+$. Denote such special set as $\mathcal{S}_3^{\text{z}}$. Then, N_1 in (4.46) is reduced to $N_1 = L - \sum_{m\in\mathcal{S}_3^{\text{z}}} p_m$. Comparing it with $N_1^\star$ in (4.48), since $0 < \bar{a} < 1$, it is not difficult to see that unless $\mathcal{S}_3^{\text{z}}$ happens to contain the indices of the lowest values in $\{p_m\}$, $N_1^\star$ in (4.48) under $N_2^\star = 1$ still has the maximum value and is optimal for $\mathcal{P}_{\mathbf{a}}$. If $\mathcal{S}_3^{\text{z}}$ happens to contain the indices of the lowest values in $\{p_m\}$, *i.e.*, $\mathcal{S}_3^{\text{z}} = \{m_1, \ldots, m_{M-N_1}\}$, then

$$N_1 = L - \sum_{j=1}^{M-N_1} p_{m_j}. \tag{4.50}$$

Compared the above expression with that in (4.48), we see that in fact we can treat the solution above as the special case of (4.48) for $N_2 \geq 1$, where substituting N_1 in (4.50) into (4.49), we have $\bar{a}^\star = p_{m_{M-N_1^\star}}$. In other words, in this special case, $\mathcal{S}_2$ and $\mathcal{S}_3$ are merged into $\mathcal{S}_3$.

Algorithm 5 The optimal solution $\mathbf{a}^\star$ for $\mathcal{P}_{\mathbf{a}}$.

1) Sort $\mathbf{p}$ in ascending order: $p_{m_1} \leq p_{m_2} \leq \cdots \leq p_{m_M}$.
2) Initialize $N_1 = L - 1$.
while $N_1 + \sum_{j=1}^{M-N_1-1} p_{m_j} \geq L$ **do**
$N_1 \leftarrow N_1 - 1$.
end while
3) Set $N_1^\star = N_1$.
4) Determine the optimal $\mathbf{a}^\star$:
a) $a^\star_{m_j} = p_{m_j}$, $j = 1, \ldots, M - N_1^\star - 1$.
b) $a^\star_{m_{M-N_1}} = \bar{a}^\star$ as in (4.49).
c) $a^\star_{m_j} = 1$, $j = M - N_1^\star + 1, \ldots, M$.

Based on both cases discussed above, we summarize the optimal solution $\mathbf{a}^\star$ to $\mathcal{P}_{\mathbf{a}}$ in Algorithm 5. In this algorithm, only sorting of elements in $\mathbf{p}$ is required; otherwise, $\mathbf{a}^\star$ is provided in closed-form.

Remark 2. In the reformulated problem $\mathcal{P}_3^{\text{SINR}}$ (and $\mathcal{P}_3$), the objective function contains an additional penalty term $\zeta|\mathbf{a}^T\mathbf{a} - L|$, where ζ is the penalty parameter. We point out that as a feature of our proposed ADMM structure in Algorithm 3, no tuning of ζ is needed. Specifically, as discussed earlier, the joint optimization in (4.20) can be separated for $\mathbf{W}$ in problem (4.23) and $\mathbf{a}$ in problem (4.36). Problem (4.36) is equivalent to $\mathcal{P}_{\mathbf{a}}$, where ζ does not affect the optimal solution $\mathbf{a}$ and can be ignored.

4.3.3 Update $\mathbf{Y}$

Subproblem (4.21) for updating $\mathbf{Y}$ can be decomposed into M sperate subproblems, one for each $\mathbf{y}_m \triangleq \mathbf{e}_m^T\mathbf{Y}$, $m \in \mathcal{M}$, given by

$$\begin{aligned} \mathcal{P}^{\text{in}}_{\mathbf{y}_m} : \min_{\mathbf{y}_m} \ & \rho\|\mathbf{e}_m^T\mathbf{W} - \mathbf{y}_m + \mathbf{e}_m^T\mathbf{Z}\|_F^2 \\ \text{s.t.} \ & \|\mathbf{y}_m\|^2 \leq a_m P_{\max}. \end{aligned} \tag{4.51}$$

Since $\mathcal{P}^{\text{in}}_{\mathbf{y}_m}$ is a convex QCQP problem with a single constraint, it yields a closed-form solution with two possible forms:
i) $\mathbf{y}_m^\star = \mathbf{e}_m^T(\mathbf{W}+\mathbf{Z})$: if it satisfies constraint (4.51); otherwise ii) $\mathbf{y}_m^\star = \sqrt{a_m P_{\max}}\dfrac{\mathbf{e}_m^T(\mathbf{W}+\mathbf{Z})}{\|\mathbf{e}_m^T(\mathbf{W}+\mathbf{Z})\|}$.

4.3.4 Discussions

Summary of Algorithm

For the mixed-integer joint robust multicast beamforming and antenna selection in problem $\mathcal{P}_1$, we exploit an equivalent set of constraints to transfer the problem into problem $\mathcal{P}_2$ with all continuous variables. To transfer problem $\mathcal{P}_2$ into a more tractable problem, we use the exact penalty term to relax the problem by transferring the quadratic equality constraint in problem $\mathcal{P}_2$ to the objective function as in problem $\mathcal{P}_3$. Problem $\mathcal{P}_3$ due to relaxation does not guarantee a binary variable. Hence, we develop a two-phase approach, the first phase focus on antenna selection, followed by the second phase to design the beamforming matrix with the selected antennas. In Phase 1 rather than resorting to the conventional convex approximation approach, which is difficult to apply, we explore the structure of this non-convex problem formulation and directly solve it by leveraging the ADMM technique. Our overall proposed ADMM-based fast algorithm to solve $\mathcal{P}_{\text{ADMM}}^{\text{SINR}}$ contains two layers of ADMM procedures. The outer-layer ADMM given by Algorithm 3 solves the non-convex problem $\mathcal{P}_3^{\text{SINR}}$ by decomposing the problem into three main subproblems (4.20)- (4.22), which are solved iteratively. The salient feature in our design is that for these subproblems formed by our ADMM construction in $\mathcal{P}_3^{\text{SINR}}$, we can develop special techniques to derive the semi-closed-form solutions to these subproblems. For the second subproblem, we develop the inner-layer consensus-ADMM-based three-block algorithm (Algorithm 4) to solve it, where the closed-form solutions are obtained for the updates in each ADMM iteration. These closed-form solutions are essential in our ADMM-based algorithm for fast computing the solution to $\mathcal{P}_3^{\text{SINR}}$, especially for large-scale systems such as massive MIMO. One challenge in the exact penalty approach used to transfer problem $\mathcal{P}_2$ to $\mathcal{P}_3$ is to obtain the optimum value for the penalty parameter ζ. One common method in the literature to find the optimum value of ζ is an exhaustive search over a range of ζ values. An exhaustive search for large-scale problems adds considerable computational complexity due to solving the problem frequently to find the optimum value for ζ. One other significant benefit of our proposed ADMM solution in Algorithm 3 is to remove the dependency of the solution in

problem $\mathcal{P}_3^{\text{SINR}}$ to the value of ζ. In problem 4.3.2, ζ is a positive coefficient of the objective function. Consequently, ζ does not affect the optimum solution of the optimization problem 4.3.2. Hence, problem $\mathcal{P}_3^{\text{SINR}}$ does not need to be solved iteratively for the optimum value of ζ. This leads to a considerable improvement in the computational complexity of problem $\mathcal{P}_3^{\text{SINR}}$.

4.4 Antenna Selection: SLR-Based Approach

4.4.1 The SLR-Based Problem Formulation

Our previous approach for antenna selection in $\mathcal{P}_3^{\text{SINR}}$ is based on SINR under the estimated channel covariance matrix. To further reduce the computational complexity involved in antenna selection in Phase 1, in this section, we propose a different approach for the antenna selection based on signal-to-leakage-ratio (SLR). SLR is a heuristic performance metric often used for downlink multiuser (unicast) beamforming design in the literature [130]. It is shown to lead to a simpler problem and yet provides an effective solution with good performance, since the metric heuristically resembles the optimal beamforming structure.

In our problem, the SLR for user k in group i is defined as the ratio of the received signal power at the user to the interference that the beamformer of group i causes to the other groups:

$$\text{SLR}_{ik} = \frac{\mathbf{w}_i^H \mathbf{R}_{ik} \mathbf{w}_i}{\sum_{l \in \mathcal{G}_{-i}} \sum_{j=1}^{K_l} \mathbf{w}_i^H \mathbf{R}_{lj} \mathbf{w}_i + \sigma^2}. \tag{4.52}$$

Then, instead of $\mathcal{P}_3^{\text{SINR}}$, we formulate the antenna selection problem for Phase 1 by setting the minimum SLR requirements, given by

$$\begin{aligned}
\mathcal{P}_3^{\text{SLR}} : \min_{\mathbf{W},\mathbf{a}} \; & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 + \zeta |\mathbf{a}^H \mathbf{a} - L| \\
\text{s.t.} \; & \frac{\mathbf{w}_i^H \mathbf{R}_{ik} \mathbf{w}_i}{\sum_{l \in \mathcal{G}_{-i}} \sum_{j=1}^{K} \mathbf{w}_i^H \mathbf{R}_{lj} \mathbf{w}_i + \sigma^2} \geq \hat{\gamma}_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}, \\
& \|\mathbf{e}_m^T \mathbf{W}\|^2 \leq a_m P_{\max}, \; m \in \mathcal{M}, \\
& \mathbf{1}^T \mathbf{a} = L,
\end{aligned}$$

$$\mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}$$

where $\hat{\gamma}_{ik}$ is the minimum SLR target.

Remark 3. Note that due to the difference of SINR and SLR metrics, the value of the SLR target $\hat{\gamma}_{ik}$ may be different from the SINR target γ_{ik}, although we do not expect the difference to be significant. Since Phase 1 focuses on the antenna selection, the problem is not sensitive to the value of $\hat{\gamma}_{ik}$. Hence, we chose $\hat{\gamma}_{ik} = \gamma_{ik}$. In our simulation study, we will compare two approaches for antenna selection the Phase 1 and show that they achieve near identical performance.

4.4.2 The ADMM-Based Fast Algorithm

We again develop an ADMM-based fast algorithm to solve $\mathcal{P}_3^{\text{SLR}}$ with low computational complexity. The construction procedure is similar to $\mathcal{P}_{\text{ADMM}}^{\text{SINR}}$ for $\mathcal{P}_3^{\text{SINR}}$ in Section 4.3. Thus, we will only describe the procedure briefly. Using auxiliary variables $\overline{\mathbf{Y}} \in \mathbb{C}^{M\times G}$, we have the following equivalent problem to $\mathcal{P}_3^{\text{SLR}}$:

$$\begin{aligned}
\mathcal{P}_{\text{ADMM}}^{\text{SLR}}: \min_{\mathbf{W},\mathbf{a}} \ & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L| \\
\text{s.t.} \ & \mathbf{w}_i^H(\mathbf{R}_{ik} - \gamma_{ik}\overline{\mathbf{R}}_{-i})\mathbf{w}_i \geq \sigma_{ik}^2\hat{\gamma}_{ik}, \ k \in \mathcal{K}_i, i \in \mathcal{G}, && \text{(4.53a)} \\
& \|\mathbf{e}_m^T\overline{\mathbf{Y}}\|^2 \leq a_m P_{\max}, \ j \in \mathcal{M}, && \text{(4.53b)} \\
& \mathbf{1}^T\mathbf{a} = L, && \text{(4.53c)} \\
& \mathbf{0} \preccurlyeq \mathbf{a} \preccurlyeq \mathbf{1}, && \text{(4.53d)} \\
& \overline{\mathbf{Y}} = \mathbf{W} && \text{(4.53e)}
\end{aligned}$$

where $\overline{\mathbf{R}}_{-i} \triangleq \sum_{l\in\mathcal{G}_{-i}}\sum_{j=1}^{K_l}\mathbf{R}_{lj}$. Denote the feasibility set of $(\mathbf{W}, \mathbf{a}, \overline{\mathbf{Y}})$ satisfying constraints in (4.53a)–(4.53d) as $\mathcal{H}$ and define the indicator function $\mathbb{I}_{\mathcal{H}}(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}})$ similarly as in (4.17). We transform $\mathcal{P}_{\text{ADMM}}^{\text{SLR}}$ to the following equivalent problem:

$$\begin{aligned}
\min_{\mathbf{W},\mathbf{a},\overline{\mathbf{Y}}} \ & \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L| + \mathbb{I}_{\mathcal{H}}\big(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}\big) \\
\text{s.t.} \ & \overline{\mathbf{Y}} = \mathbf{W}
\end{aligned}$$

for which the augmented Lagrangian is given by

$$\bar{\mathcal{L}}_{\bar{\rho}}\big(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}\big) = \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 + \zeta|\mathbf{a}^H\mathbf{a} - L| + \mathbb{I}_{\mathcal{H}}\big(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}\big) + \bar{\rho}\|\mathbf{W} - \overline{\mathbf{Y}} + \overline{\mathbf{Z}}\|_F^2$$

where $\overline{\mathbf{Z}} \in \mathbb{C}^{M\times G}$ is the dual variable associated with constraint (4.53e) and $\bar{\rho}$ is the penalty parameter.

Similar to Algorithm 3, we construct two ADMM blocks to minimize $\bar{\mathcal{L}}_{\bar{\rho}}\big(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}\big)$ alternatingly w.r.t. $\mathbf{a}$, $\mathbf{W}$, and $\overline{\mathbf{Y}}$. Once the convergence is reached, we round the elements in $\mathbf{a}$ to the nearest 0 or 1 for antenna selection. The algorithm is summarized in Algorithm 6. Again, in the first block of Algorithm 6, the joint optimization of $(\mathbf{a}, \mathbf{W})$ in subproblem (4.54) can be decoupled into separate subproblems for $\mathbf{W}$ and $\mathbf{a}$, respectively. In particular, for the first two ADMM blocks:

- Update $\mathbf{a}$: In subproblem (4.54), the minimization of $\bar{\mathcal{L}}_{\bar{\rho}}(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}^l, \overline{\mathbf{Z}}^l)$ w.r.t. $\mathbf{a}$ is the same as in problem (4.36) in Section 4.3. Thus, Algorithm 5 provides the optimal solution $\mathbf{a}^\star$ for updating $\mathbf{a}$.

- Update $\overline{\mathbf{Y}}$: Subproblem (4.55) is the same as subproblem (4.21). Thus, the solution in Section 4.3.3 follows.

Thus, only the minimization of $\bar{\mathcal{L}}_{\bar{\rho}}(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}^l, \overline{\mathbf{Z}}^l)$ w.r.t. $\mathbf{W}$ in subproblem (4.54) in the first ADMM block is different from that for $\mathcal{P}^{\text{SINR}}_{\text{ADMM}}$ due to the SLR constraints in (4.53a). Below, we show $\mathbf{W}$ update can be computed efficiently using the semi-closed-form expressions. In particular, we will show that the SLR metric allows us to further decompose the optimization problem into smaller subproblems, which yields (semi)-closed-form updates with lower computational complexity.

Update $\mathbf{W}$

Recall $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_G]$. Similarly, we denote $\overline{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_G]$ and $\overline{\mathbf{Z}} = [\bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_G]$. Then, optimizing $\mathbf{W}$ in subproblem (4.54) can be written as follows:

$$\min_{\{\mathbf{w}_i\}} \sum_{i=1}^{G} \|\mathbf{w}_i\|^2 + \bar{\rho}\sum_{i=1}^{G} \|\mathbf{w}_i - \bar{\mathbf{y}}_i + \bar{\mathbf{z}}_i\|^2 \tag{4.57}$$

Algorithm 6 The ADMM-Based Algorithm for $\mathcal{P}_3^{\text{SLR}}$

Initialization: Set $\bar{\rho}$; Set initial $\mathbf{W}^0$, $\mathbf{a}^0$, $\bar{\mathbf{Z}}^0 = \mathbf{0}$; Set $l = 0$.
repeat
1) Update beamforming matrix $\mathbf{W}^{l+1}$, $\mathbf{a}^{l+1}$:

$$(\mathbf{W}^{l+1}, \mathbf{a}^{l+1}) = \arg\min_{\mathbf{a},\mathbf{W}} \bar{\mathcal{L}}_{\bar{\rho}}\big(\mathbf{a}, \mathbf{W}, \overline{\mathbf{Y}}^l, \overline{\mathbf{Z}}^l\big). \tag{4.54}$$

2) Update the auxiliary matrices $\mathbf{Y}^{l+1}$:

$$\overline{\mathbf{Y}}^{l+1} = \arg\min_{\overline{\mathbf{Y}}} \bar{\mathcal{L}}_{\bar{\rho}}\big(\mathbf{a}^{l+1}, \mathbf{W}^{l+1}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}^l\big). \tag{4.55}$$

3) Update dual variables $\mathbf{Z}^{l+1}$:

$$\overline{\mathbf{Z}}^{l+1} = \overline{\mathbf{Z}}^l + \overline{\mathbf{Y}}^{l+1} - \mathbf{W}^{l+1}. \tag{4.56}$$

4) Set $l \leftarrow l + 1$.
until convergence
Obtain $\mathbf{a}$: $a_m = \text{round}(a_m^l)$, $m \in \mathcal{M}$.

$$\text{s.t. } \mathbf{w}_i^H(\mathbf{R}_{ik} - \hat{\gamma}_{ik}\overline{\mathbf{R}}_{-i})\mathbf{w}_i \geq \sigma^2\hat{\gamma}_{ik},\ k \in \mathcal{K}_i, i \in \mathcal{G}.$$

The above problem can be decomposed into G sperate subproblems, one for each $\mathbf{w}_i$, given by

$$\begin{aligned}\mathcal{P}_{\mathbf{w}_i} : \min_{\mathbf{w}_i}\ & \|\mathbf{w}_i\|^2 + \bar{\rho}\|\mathbf{w}_i - \bar{\mathbf{y}}_i + \bar{\mathbf{z}}_i\|^2 \\ \text{s.t. }\ & \mathbf{w}_i^H(\mathbf{R}_{ik} - \hat{\gamma}_{ik}\overline{\mathbf{R}}_{-i})\mathbf{w}_i \geq \sigma^2\hat{\gamma}_{ik},\ k \in \mathcal{K}_i.\end{aligned}$$

Remark 4. Compared with $\mathcal{P}_{\mathbf{W}}$ in Algorithm 3 for solving $\mathcal{P}_3^{\text{SINR}}$, the SLR constraints in $\mathcal{P}_3^{\text{SLR}}$ allow us to decompose problem (4.57) into G subproblems to solve separately, which leads to simpler subproblems with closed-form solutions.

Problem $\mathcal{P}_{\mathbf{w}_i}$ is a non-convex QCQP problem. We again develop an inner consensus-ADMM-based fast algorithm to solve it. For notation simplicity, in the discussion below, we remove subscript i from all variables in $\mathcal{P}_{\mathbf{w}_i}$, with the understanding that they are all for group i. Using auxiliary variables $\hat{\mathbf{w}}_k \in \mathbb{C}^{M\times 1}$, $k \in \mathcal{K}$, we have the equivalent problem to $\mathcal{P}_{\mathbf{w}_i}$ as shown below. :

$$\min_{\mathbf{w},\{\hat{\mathbf{w}}_k\}} \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2$$

Algorithm 7 Consensus-ADMM-Based Algorithm for $\mathcal{P}_{\mathbf{w}_i}$

Initialization: Set $\hat{\rho}$; Set initial $\mathbf{w}^0$, $\{\tilde{\mathbf{z}}_k\}^0 = \mathbf{0}$; Set $q = 0$.

repeat

1) Update the auxiliary matrices $\{\hat{\mathbf{w}}_k^{q+1}\}$:

$$\{\hat{\mathbf{w}}_k^{q+1}\} = \arg\min_{\{\hat{\mathbf{w}}_k\}} \hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}^q, \{\hat{\mathbf{w}}_k, \hat{\mathbf{z}}_k^q\}\big). \tag{4.59}$$

2) Update beamforming matrix $\mathbf{w}^{q+1}$:

$$\mathbf{w}^{q+1} = \arg\min_{\mathbf{w}} \hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k^{q+1}, \hat{\mathbf{z}}_k^q\}\big). \tag{4.60}$$

3) Update dual variables $\{\hat{\mathbf{z}}_k^{q+1}\}$:

$$\hat{\mathbf{z}}_k^{q+1} = \hat{\mathbf{z}}_k^q + \mathbf{w}^{q+1} - \mathbf{w}_k^{q+1}. \tag{4.61}$$

4) Set $q \leftarrow q + 1$.

until convergence

$$\begin{aligned} \text{s.t. } & \mathbf{w}^H \overline{\mathbf{R}} \mathbf{w} - \frac{1}{\hat{\gamma}_k} \hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k + \sigma^2 \le 0,\ k \in \mathcal{K}, \\ & \hat{\mathbf{w}}_k = \mathbf{w}, k \in \mathcal{K}. \end{aligned} \tag{4.58}$$

Define $\mathcal{V}$ as the feasible set satisfying the inequality constraints in the above problem and define the indicator function $\mathbb{I}_{\mathcal{V}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) \triangleq \{0 : \text{if } (\mathbf{w}, \{\hat{\mathbf{w}}_k\}) \in \mathcal{V}; \infty : \text{otherwise}\}$. Then, the above problem is equivalent to

$$\begin{aligned} \min_{\mathbf{w}, \{\bar{\mathbf{w}}_k\}} \ & \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \mathbb{I}_{\mathcal{V}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) \\ \text{s.t. } & \hat{\mathbf{w}}_k = \mathbf{w}, k \in \mathcal{K}. \end{aligned}$$

for which the augmented Lagrangian is given by

$$\hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k, \hat{\mathbf{z}}_k\}\big) = \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \mathbb{I}_{\mathcal{V}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k\}\big) + \hat{\rho} \sum_{k=1}^{K} \|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2$$

where $\hat{\mathbf{z}}_k$ is the dual variable associated with constraint (4.58) and $\hat{\rho}$ is the penalty parameter. Again, we construct ADMM blocks to minimize $\hat{\mathcal{L}}_{\hat{\rho}}\big(\mathbf{w}, \{\hat{\mathbf{w}}_k, \hat{\mathbf{z}}_k\}\big)$ alternatingly. We show below that each optimization problem yields a (semi)-closed-form solution. The consensus-ADMM-based algorithm is illustrated in Algorithm 7.

i) Update $\{\hat{\mathbf{w}}_k\}$: The subproblem (4.59) can be decomposed into K $(= |\mathcal{K}|)$ separate subproblems, each given by

$$\begin{aligned} \min_{\hat{\mathbf{w}}_k} \ & \|\mathbf{w} + \hat{\mathbf{z}}_k - \hat{\mathbf{w}}_k\|^2 \\ \text{s.t. } & \hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k \geq \gamma_k(\mathbf{w}^H \overline{\mathbf{R}} \mathbf{w} + \sigma^2) \end{aligned} \tag{4.62}$$

Problem (4.62) is a QCQP problem with a single concave constraint function. It satisfies the linear independence constraint qualification (LICQ) under which the KKT conditions is necessary for optimality [129]. Thus, we use the KKT conditions to derive the solution and obtain a unique semi-closed-form solution, which is optimal to problem (4.62). There are two possible solutions:

i) $\hat{\mathbf{w}}_k^\circ = \mathbf{w} + \hat{\mathbf{z}}_k$: This is the solution if it satisfies the constraint in (4.62). Otherwise,

ii) The constraint in (4.62) holds with equality. Let $b_k \triangleq \hat{\gamma}_k(\mathbf{w}^H \bar{\mathbf{R}} \mathbf{w} + \sigma^2)$. The Lagrangian of problem (4.62) is given by

$$\mathcal{L}(\hat{\mathbf{w}}_k, \hat{\lambda}_k) = \|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 + \hat{\lambda}_k(-\hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k + b_k)$$

where $\hat{\lambda}_k$ is Lagrange multiplier associated with the constraint in (4.62). Set the gradient of $\mathcal{L}(\hat{\mathbf{w}}_k, \hat{\lambda}_k)$ to 0, we have

$$\nabla_{\hat{\mathbf{w}}_k^*} \mathcal{L}(\hat{\mathbf{w}}_k, \hat{\lambda}_k) = -\mathbf{w} + \hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k - \hat{\lambda}_k \mathbf{R}_k \hat{\mathbf{w}}_k = 0, \tag{4.63}$$

which leads to

$$\hat{\mathbf{w}}_k = (\hat{\lambda}_k \mathbf{R}_k + \mathbf{I})^{-1}(\mathbf{w} + \hat{\mathbf{z}}_k). \tag{4.64}$$

Since $-\hat{\mathbf{w}}_k^H \mathbf{R}_k \hat{\mathbf{w}}_k + b_k = 0$, we obtain $\hat{\lambda}_k$ as the solution of the following equality:

$$(\mathbf{w} + \hat{\mathbf{z}}_k)^H (\hat{\lambda}_k \mathbf{R}_k + \mathbf{I})^{-1} \mathbf{R}_k (\hat{\lambda}_k \mathbf{R}_k + \mathbf{I})^{-1} (\mathbf{w} + \hat{\mathbf{z}}_k) = b_k. \tag{4.65}$$

Consider eigenvalue decomposition $\mathbf{R}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{U}_k^H$, where $\mathbf{\Sigma}_k \triangleq \text{diag}(\tilde{\sigma}_{1k}^2, \ldots, \tilde{\sigma}_{Mk}^2)$. Then, (4.65) is equivalent to

$$\sum_{m=1}^{M} \frac{\tilde{\sigma}_{mk}^2 |\beta_{mk}|^2}{(1 + \hat{\lambda}_k \tilde{\sigma}_{mk}^2)^2} = b_k \tag{4.66}$$

where $\boldsymbol{\beta}_k \triangleq [\beta_{1k}, \ldots, \beta_{Mk}]^T = \mathbf{U}_k^H(\mathbf{w} + \hat{\mathbf{z}}_k)$. Since $\hat{\lambda}_k \geq 0$, the LHS of (4.66) is a decreasing function of $\hat{\lambda}_k \geq 0$. Thus, the solution $\hat{\lambda}_k$ can be found via bi-section search. Since the solution is unique, it is the optimal solution to problem (4.62).

ii) Update $\mathbf{w}$: The subproblem (4.60) for $\mathbf{w}$ is given by

$$\begin{aligned} \min_{\mathbf{w}} \ & \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \hat{\rho}\sum_{k=1}^{K}\|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 \\ \text{s.t.} \ & \mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} \leq \min_{k\in\mathcal{K}} \frac{1}{\hat{\gamma}_k}\hat{\mathbf{w}}_k^H\mathbf{R}_k\hat{\mathbf{w}}_k - \sigma^2. \end{aligned} \tag{4.67}$$

It is a convex QCQP problem with a single constraint, for which we obtain the closed-form solution. There are two possible solutions:

i) $\mathbf{w}^\circ = \frac{1}{1+\rho+K\bar{\rho}}(\bar{\rho}\sum_{k=1}^{K}(\hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k) + \rho(\bar{\mathbf{y}} - \bar{\mathbf{z}}_k))$: This is the solution if it satisfies the constraint in (4.67). Otherwise, ii) The constraint in (4.67) holds with equality. Define $d \triangleq \min_k \frac{1}{\hat{\gamma}_k}\hat{\mathbf{w}}_k^H\mathbf{R}_k\hat{\mathbf{w}}_k - \sigma^2$. The Lagrangian for problem (4.67) is

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \bar{\lambda}) =& \|\mathbf{w}\|^2 + \bar{\rho}\|\mathbf{w} - \bar{\mathbf{y}} + \bar{\mathbf{z}}\|^2 + \hat{\rho}\sum_{k=1}^{K}\|\mathbf{w} - \hat{\mathbf{w}}_k + \hat{\mathbf{z}}_k\|^2 \\ & + \bar{\lambda}(\mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} - d). \end{aligned}$$

where $\bar{\lambda}$ is the Lagrange multiplier for the constraint in (4.67). Set the gradient of $\mathcal{L}(\mathbf{w}, \bar{\lambda})$ to 0, we have

$$\nabla_{\hat{\mathbf{w}}^*}\mathcal{L}(\mathbf{w}, \bar{\lambda}) = \big(\bar{\lambda}\overline{\mathbf{R}} + (1 + \bar{\rho} + K\hat{\rho})\mathbf{I}\big)\mathbf{w} - \mathbf{c} = 0$$

where $\mathbf{c} \triangleq \hat{\rho}\sum_{k=1}^{K}(\hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k) + \bar{\rho}(\bar{\mathbf{y}} - \bar{\mathbf{z}})$. This leads to

$$\mathbf{w} = (\bar{\lambda}\overline{\mathbf{R}} + (1 + \bar{\rho} + K\hat{\rho})\mathbf{I})^{-1}\mathbf{c}. \tag{4.68}$$

Since $\mathbf{w}^H\overline{\mathbf{R}}\mathbf{w} - d = 0$, we obtain $\bar{\lambda}$ as the solution of the following inequality.

$$\mathbf{c}^H(\bar{\lambda}\overline{\mathbf{R}} + (1 + \bar{\rho} + K\hat{\rho})\mathbf{I})^{-1}\overline{\mathbf{R}}(\bar{\lambda}\overline{\mathbf{R}} + (1 + \bar{\rho} + K\hat{\rho})\mathbf{I})^{-1}\mathbf{c} = b. \tag{4.69}$$

Again, applying eigenvalue decomposition on $\overline{\mathbf{R}} = \overline{\mathbf{U}}\,\overline{\mathbf{\Sigma}}\,\overline{\mathbf{U}}^H$ with $\overline{\mathbf{\Sigma}} \triangleq \text{diag}(\bar{\sigma}_1^2, \ldots, \bar{\sigma}_M^2)$, we can rewrite (4.69) as the following equation:

$$\sum_{m=1}^{M} \frac{\bar{\sigma}_m^2|\bar{c}_m|^2}{(1 + \rho + K\bar{\rho} + \bar{\lambda}\bar{\sigma}_m^2)^2} = b. \tag{4.70}$$

where $\bar{\mathbf{c}} \triangleq [c_1, \ldots, c_M]^T = \bar{\mathbf{U}}^H \mathbf{c}$. The solution $\bar{\lambda} \geq 0$ to equation (4.70) can be found via bi-section search.

4.4.3 Computational Complexity Comparison

The SINR-based approach via $\mathcal{P}_3^{\text{SINR}}$

The computational complexity of the proposed ADMM-based algorithm in Algorithm 3 mainly occurs at updating $\mathbf{W}$ in $\mathcal{P}_{\mathbf{W}}$ in Section 4.3.1, which is computed by the consensus-ADMM- based algorithm in Algorithm 4. The updates of $\mathbf{a}$ and $\mathbf{Y}$ both yield closed-form solutions as shown in Algorithm 5 and in Section 4.3.3, respectively, where only simple matrix-vector operations are involved. Since Algorithm 5 involves sorting with average complexity $\mathcal{O}(M \log(M))$, the overall complexity for $\mathbf{a}$ and $\mathbf{Y}$ updates is $\mathcal{O}(M \log(M) + MG)$. Algorithm 4 contains three ADMM blocks in each iteration for optimizing $\{\mathbf{W}_{ik}\}$, $\{\mathbf{V}_{ik}\}$, and $\mathbf{W}$, whose closed-form solutions are given in (4.32), (4.34), and (4.35), respectively. Thus, the overall computational complexity in each iteration is $\mathcal{O}(M^2 G \sum_{i=1}^{G} K_i)$.[4]

The SLR-based approach via $\mathcal{P}_3^{\text{SLR}}$

The computational complexity of the ADMM-based algorithm in Algorithm 6 mainly occurs at updating $\mathbf{W}$ in Section 4.4.2, which is the consensus-ADMM- based algorithm in Algorithm 7. As discussed in Section 4.4.2, the updates of of $\mathbf{a}$ and $\overline{\mathbf{Y}}$ are the same as the SINR-based approach in Algorithm 3. Thus, the only difference of the two approaches is in the computational complexity incurred for updating $\mathbf{W}$. In particular, Algorithm 7 for updating $\mathbf{w}_i$ contains three ADMM blocks in each iteration for optimizing $\{\hat{\mathbf{w}}_{ik}\}$ and $\{\mathbf{w}_i\}$, whose semi-closed-form solutions are given in (4.64) and (4.67), respectively. Note that the matrix inversions in (4.64) and (4.67) only need to be performed once at the beginning of Algorithm 6. The overall computational complexity in each iteration is $\mathcal{O}(M^2 \sum_{i=1}^{G} K_i)$. Compare this with that for the SINR-based approach above, we see that the SLR-based approach has lower computational complexity in each iteration, leading to faster computation

[4]Note that $\mathbf{D}_{ik}$ is a diagonal matrix. We use this structure in calculating the required matrix multiplication operation in (4.32), (4.34), and (4.35). Also, the diagonal matrix inversion in (4.35) only needs to be computed once at the beginning of Algorithm 3.

of a solution. Furthermore, in Section 3.5, we will show in our simulation study that the SLR-based approach also results in faster convergence than the SINR-based approach.

4.5 Robust Multicast Beamforming

Given the antenna selection vector $\mathbf{a}$ obtained in Phase 1, in Phase 2, we find the robust beamforming solution for the selected antennas. Denote the channel covariance matrix of the selected antennas by $\mathbf{R}_{ik}^{\text{s}} \in \mathbb{C}^{L\times L}$; it is obtained by selecting the columns and rows from $\mathbf{R}_{ik}$, for which $a_m = 1$. Similarly, denote $\mathbf{W}^{\text{s}} \in \mathbb{C}^{L\times G}$ as the beamforming matrix of selected antennas, which corresponds to the L selected rows of $\mathbf{W}$, for which $a_m = 1$. Then, removing the constraints (4.13b)–(4.13d) on $\mathbf{a}$ from $\mathcal{P}_1$, we have the following equivalent robust multicast beamforming optimization problem:

$$\begin{aligned}\mathcal{P}_4 : &\min_{\mathbf{W}^{\text{s}}} \ \text{tr}(\mathbf{W}^{\text{s}}\mathbf{W}^{\text{s}H}) \\ &\text{s.t.} \ \ \text{tr}(\mathbf{R}_{ik}^{\text{s}}\mathbf{W}^{\text{s}}\mathbf{D}_{ik}\mathbf{W}^{\text{s}H}) - \epsilon_{ik}\|\mathbf{W}^{\text{s}}\mathbf{D}_{ik}\mathbf{W}^{\text{s}H}\|_F \geq \sigma^2\gamma_{ik}, k \in \mathcal{K}_i, i \in \mathcal{G}.\end{aligned}$$

The above problem is the same as the regular robust multicast beamforming optimization problem considered in section 3.2.2, for which an ADMM-based fast algorithm has been developed to solve $\mathcal{P}_4$. Thus, we can directly apply Algorithms 1 and 2 in Chapter **??** to solve $\mathcal{P}_4$.

4.6 Convergence Discussion of Proposed Algorithms

Problems $\mathcal{P}_3^{\text{SINR}}$ and $\mathcal{P}_3^{\text{SLR}}$ in Phase 1 and $\mathcal{P}_4$ in Phase 2 are all non-convex problems. The convergence results of ADMM for the non-convex problems are rather limited in the literature. Typically numerical results are used to study the convergence and performance of ADMM for a specific optimization problem. Nonetheless, in a recent work [51], the convergence of ADMM to a wide range of non-convex problems is shown. Based on this result, we have proven in 3.3.3 that the convergence of the proposed ADMM-based algorithm for robust multicast beamforming is guaranteed. That is, the algorithm to solve $\mathcal{P}_4$ is guaranteed to converge. Furthermore, we point out that for the subproblems solved by

Algorithms 3, 4, and 6, their structure (both the objective and constraint functions) are similar to optimization problem in (3.16) of section 3.2.2 , and the construction of two ADMM blocks is used in each algorithm. Thus, the convergence of Algorithms 3, 4, and 6 is also guaranteed by applying the similar convergence proof. Algorithm 7 uses three ADMM blocks for the non-convex subproblem, where the convergence guarantee is more challenging to show theoretically. Instead, we resort to simulation study in Section 3.5 to show the convergence of this algorithm.

4.7 Simulation Results

We consider a downlink multicast beamforming scenario with G groups and $K_i = K$, $\forall\, i \in \mathcal{G}$, and $\gamma_{ik} = \gamma$, $\forall\, k \in \mathcal{K}_i, i \in \mathcal{G}$. Unless otherwise stated, the default setup is $G = 3$ groups, $K = 4$ users per group, $M = 100$ antennas, $L = 10$ RF chains, and $\gamma = 3$ dB. We set noise variance to $\sigma^2 = 1$.

For the error bound in (4.4), we express it as $\epsilon_{ik} = \eta\|\overline{\mathbf{R}}_{ik}\|_F$, where η is the normalized error bound. We use η to study the robust beamforming performance. For modeling the channel covariance matrix, consider the channel vector $\mathbf{h}_{ik}$ between user $k \in \mathcal{K}_i, i \in \mathcal{G}$ and BS, and the estimated channel $\hat{\mathbf{h}}_{ik} = \mathbf{h}_{ik} + \mathbf{e}_{ik}$, where $\mathbf{e}_{ik}$ is estimation error. We model the channel covariance matrix and estimated channel covariance matrix as $\bar{\mathbf{R}}_{ik} = \mathbf{h}_{ik}\mathbf{h}_{ik}^H$ and $\mathbf{R}_{ik} = \hat{\mathbf{h}}_{ik}\hat{\mathbf{h}}_{ik}^H$, respectively. The error matrix is obtained as $\mathbf{E}_{ik} = \overline{\mathbf{R}}_{ik} - \mathbf{R}_{ik}$. We generate each channel vector independently as $\mathbf{h}_{ik} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and the error vectors independently as $\mathbf{e}_{ik} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(\mathbf{0}, \sigma_e^2\mathbf{I})$, where σ_e^2 is the corresponding variance. To model the bounded error in (4.4) more realistically, we obtain $\mathbf{E}_{ik}$ as the above. We then set η as a cut-off bound such that $\|\mathbf{E}_{ik}\|_F \leq \eta\|\overline{\mathbf{R}}_{ik}\|_F = \epsilon_{ik}$ holds for 90% of realizations of $\mathbf{e}_{ik}$. Finally, those realizations with $\mathbf{E}_{ik}$ satisfying (4.4) are considered in the simulation. Unless otherwise stated, we set the default value to $\eta = 0.01$.

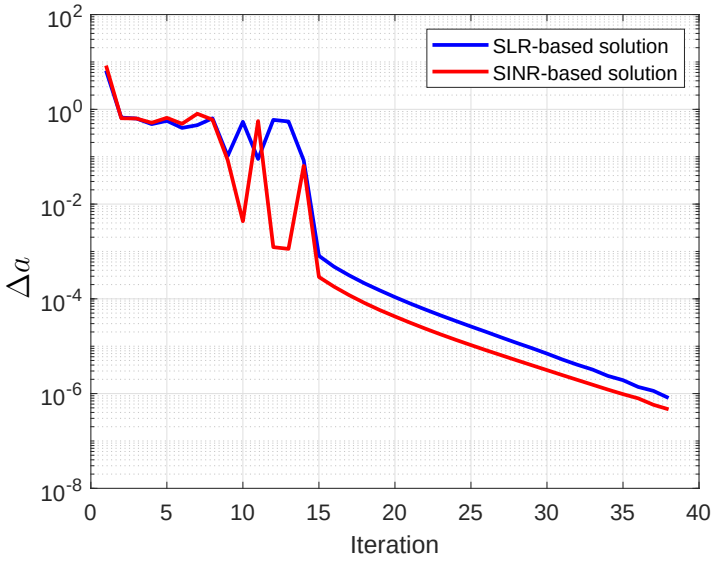

Figure 4.1: The convergence behavior of relative residual Δa by Algorithms 3 and 6.

4.7.1 Convergence Behaviour of Algorithms for Phase 1

For the proposed SINR-based and SLR-based approaches for antenna selection in Phase 1, we study the convergence behaviour of the ADMM-based algorithm developed for each approach.

For the SINR-based approach, Algorithm 3 provides the outer-layer ADMM procedure for solving $\mathcal{P}_3^{\text{SINR}}$, where the update for $\mathbf{W}$ in (4.20) is obtained by the inner-layer ADMM algorithm in Algorithm 4. Similarly, for the SLR-based approach, Algorithm 6 provides the outer-layer ADMM procedure for solving $\mathcal{P}_3^{\text{SLR}}$, where the update for $\mathbf{W}$ in (4.54) is obtained by the inner-layer ADMM algorithm in Algorithm 7. We first consider the outer-layer ADMM-based algorithm in each approach. We consider both residual convergence and objective convergence. Define the relative difference for $\mathbf{a}$ between iterations $l-1$ and l as $\Delta a \triangleq \frac{\|\mathbf{a}^l-\mathbf{a}^{l-1}\|}{\|\mathbf{a}^l\|}$, and the relative residual for the equality constraint in (4.16e) as $\Delta W \triangleq \frac{\|\mathbf{W}^l-\mathbf{Y}^{l-1}\|_F}{\|\mathbf{W}^l\|_F}$ (for the equality constraint in (4.53e) as $\Delta W \triangleq \frac{\|\mathbf{W}^l-\overline{\mathbf{Y}}^{l-1}\|_F}{\|\mathbf{W}^l\|_F}$). Let $P^l = \text{tr}(\mathbf{W}^l\mathbf{W}^{lH})$ denote the value of the power objective function in $\mathcal{P}_1$ in iteration l. We define the relative objective convergence as $\Delta P = \frac{|P^l-P^{l-1}|}{P^l}$. Figs. 4.1 and 4.2 show the trajectories of Δa and ΔW over iterations for $\mathbf{a}$ and $\mathbf{W}$, respectively, by both SINR-based

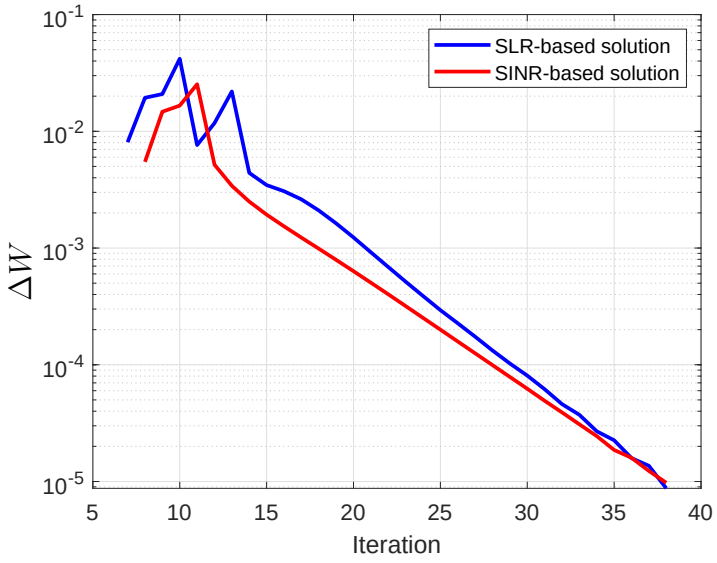

Figure 4.2: The convergence behavior of ΔW by Algorithms 3 and 6.

and SLR-based approaches. We see that the convergence behavior for both approaches are similar: the values of Δa and ΔW drop below 10^{-3} after 20 iterations, indicating fast convergence. Fig. 4.3 shows the trajectory of ΔP over iterations. Again, we see that the value of ΔP drops below 10^{-3} after 15 iterations.

Next, we study the convergence of the inner-layer ADMM-based algorithm to update $\mathbf{W}$ in each approach: Algorithm 4 for $\mathcal{P}_3^{\text{SINR}}$ and Algorithm 7 for $\mathcal{P}_3^{\text{SLR}}$. For the SINR-based approach, Algorithm 4 iterates among three update blocks $\mathbf{W}$, $\{\mathbf{W}_{ik}\}$, and $\{\mathbf{V}_{ik}\}$. Define the relative residuals in iteration j for the equality constraints in (4.24b) and (4.24c) as $\Delta\overline{W} \triangleq \max_{i,k}\{\frac{\|\mathbf{W}_{ik}^j-\mathbf{W}^j\|_F}{\|\mathbf{W}_{ik}^j\|_F}\}$ and $\Delta V = \max_{i,k}\{\frac{\|\mathbf{V}_{ik}^j-\mathbf{W}^j\mathbf{D}_{ik}\|_F}{\|\mathbf{V}_{ik}^j\|_F}\}$, respectively. Figs. 4.4 and 4.5 plot the trajectories of $\Delta\overline{W}$ and ΔV over iterations, respectively, for $M = 50, 100, 200$. We see that their values drop below 10^{-3} within about 200 iterations for all value of M. For the SLR-based approach, Algorithm 7 for $\mathcal{P}_3^{\text{SLR}}$ has two update blocks $\mathbf{w}_i$ and $\{\mathbf{w}_{ik}\}$ (with subscript i reinstated). Define the relative residual for the equality constraint in (4.58) as $\Delta w_i = \max_k\{\frac{\|\mathbf{w}_{ik}^j-\mathbf{w}_i^j\|}{\|\mathbf{w}_{ik}^j\|}\}$. Fig. 4.6 show the trajectory of Δw_i for $M = 50, 100, 200$. We see that the value of Δw_i drops below 10^{-3} after 100 iterations. Compared with Fig. 4.4, we see that the convergence speed of Algorithm 7 is

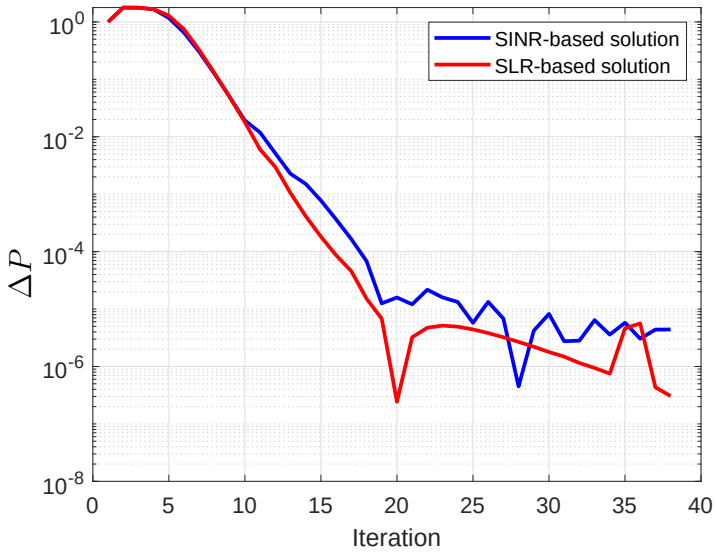

Figure 4.3: The convergence behavior of the relative objective difference ΔP of $\mathcal{P}_1$ by Algorithm 4.

approximately two times faster than that of Algorithm 4 for the SINR-based approach. As analyzed in Section 4.4.3, the computational complexity of Algorithm 7 is lower than that of Algorithm 4 per ADMM iteration. Thus, overall the SLR-based approach is a faster algorithm with lower complexity than the SINR-based approach to compute the solution.

4.7.2 Impact of Approximated Approach in Phase 1

In our proposed two-phase approach, to determine antenna selection in Phase 1, we simplify the problem by proposing $\mathcal{P}_3^{\text{SINR}}$, where we ignore the second term of the constraint (4.15a) in $\mathcal{P}_3$. Now, we study the effect of such approximation as compared to the original problem $\mathcal{P}_3$. Specifically, in Phase 1, we solve $\mathcal{P}_3$ directly via the interior point method (IPM) to obtain the antenna selection vector $\mathbf{a}$ and then round its elements to the nearest 0 or 1 for binary selection. Phase 2 is carried the same way as before. IPM is implemented via the genetic nonlinear program solver `fmincon` in MATLAB.[5] We tune the value of ζ from 0.005 to 2 and select the one resulting in the lowest transmit power. Fig. 4.7-top shows the transmit power normalized by noise P/σ^2 vs. the normalizes error bound η for

[5]We resort to the generic IPM as there is no other existing algorithm to solve the problem.

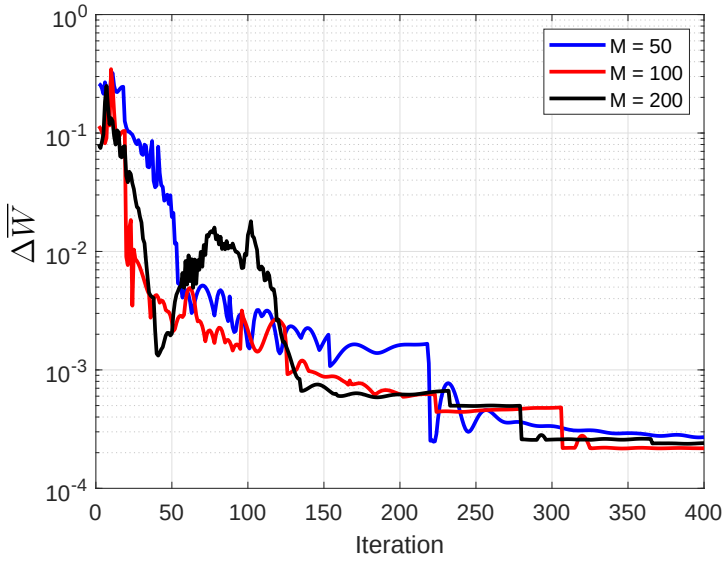

Figure 4.4: The inner-layer ADMM of SINR-based approach: the convergence behavior of relative residual $\Delta\overline{W}$ in Algorithm 4.

channel covariance. We see that using the approximate version $\mathcal{P}_3^{\text{SINR}}$ for antenna selection results in less than 0.3 dB performance gap to that of directly solving $\mathcal{P}_3$. Fig. 4.7-middle and Fig. 4.7-bottom show P/σ^2 vs. L and K, respectively. The performance of our SINR-based approach using $\mathcal{P}_3^{\text{SINR}}$ is near identical to that using $\mathcal{P}_3$ directly in Phase 1. These demonstrate that the effectiveness of our proposed approach in Phase 1 for antenna selection, which has small to negligible impact on the overall performance, but with significantly less computational complexity as we will see in the next subsection.

4.7.3 Performance Comparison

We now evaluate the performance of our proposed algorithms for antenna selection and robust multicast beamforming. For comparison, we consider the following methods to solve $\mathcal{P}_3^{\text{SINR}}$ and $\mathcal{P}_3^{\text{SLR}}$ for antenna selection in Phase 1: i) SINR-ADMM and SLR-ADMM: our proposed SINR-based approach with Algorithms 4 and SLR-based approach with Algorithm 7; ii) SINR-IPM and SLR-IPM: solving $\mathcal{P}_3^{\text{SINR}}$ and $\mathcal{P}_3^{\text{SLR}}$ using IPM via the genetic nonlinear program solver `fmincon` in MATLAB; iii) Heuristic selection: antenna selection based on the average channel strength, where we compute the average channel power

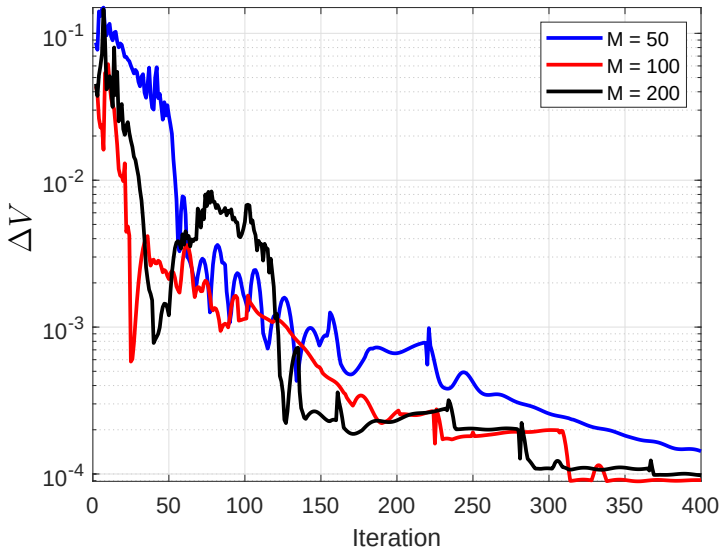

Figure 4.5: The inner-layer ADMM of SINR-based approach: the convergence behavior of relative residual ΔV in Algorithm 4.

gain from each antenna m to all users: $\bar{g}_m = \frac{1}{\sum_{i=1}^{G} K_i} \sum_{i=1}^{G} \sum_{k=1}^{K_i} [\mathbf{R}_{ik}]_{mm}$, and select the antennas with the L largest gains; iv) Random search-100: instead of exhaustive search of L antennas, we consider 100 trials of random selection, and choose the best selection among these trials that gives the lowest transmit power; v) All antennas: all M antennas are used for multicast beamforming and Phase 1 is not used. We use this benchmark as a lower bound for all antenna selection methods.

Performance vs. number of selected antennas L

Fig. 4.8 shows the normalized transmit power $\frac{P}{\sigma^2}$ vs. the number of selected antennas L for the default setup. As expected, the transmit power decreases as L increases for all the antenna selection methods. Our proposed two-phased solutions (SINR-ADMM and SLR-ADMM) outperforms the heuristic selection and random search-100. The performance gap becomes significant when L becomes smaller, as the good antenna selection becomes critical when only a few antennas are selected. Furthermore, comparing between the SINR-based and SLR-based approaches in Phase 1, from the zoom-in plot, we see the SLR-ADMM shows a negligible loss as compared to SINR-ADMM; the two approaches result in

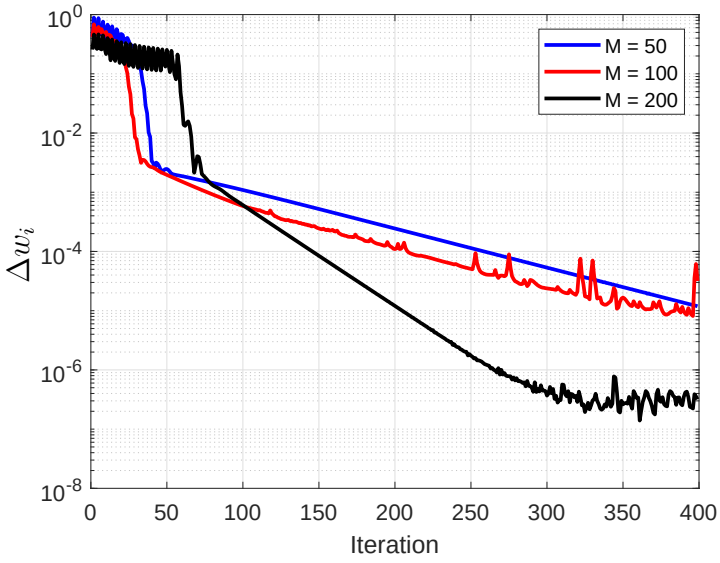

Figure 4.6: The inner-layer ADMM of SLR-based approach: the convergence behavior of relative residual Δw_i in Algorithm 7.

Table 4.1: Average Computation Time Over L (sec.)

L	5	10	15	20
SINR-ADMM	98	100	102	104
SLR-ADMM	22.3	25	27.8	30.6
SINR-IPM	10178	12119	12396	13016
SLR-IPM	6069.5	6983.1	7754.7	7906.7

nearly identical performance. Comparing the two approaches with their IPM counterpart, wee see that the performance of our ADMM-based algorithms is nearly identical to that produced by IPM. This demonstrates the effectiveness of our proposed fast algorithms.

The computational advantage of our proposed ADMM-based algorithms is shown in Table 4.1, where we provide the average computation time for the plots of SINR-based and SLR-based approaches in Fig. 4.8. Our proposed ADMM-based algorithms are significantly faster in computing a solution than IPM for both of the SINR-based and SLR-based approaches. Furthermore, SLR-ADMM has lower computational complexity than SINR-ADMM for all L values, with nearly identical performance. This demonstrates the effectiveness of our proposed SLR-based approach.

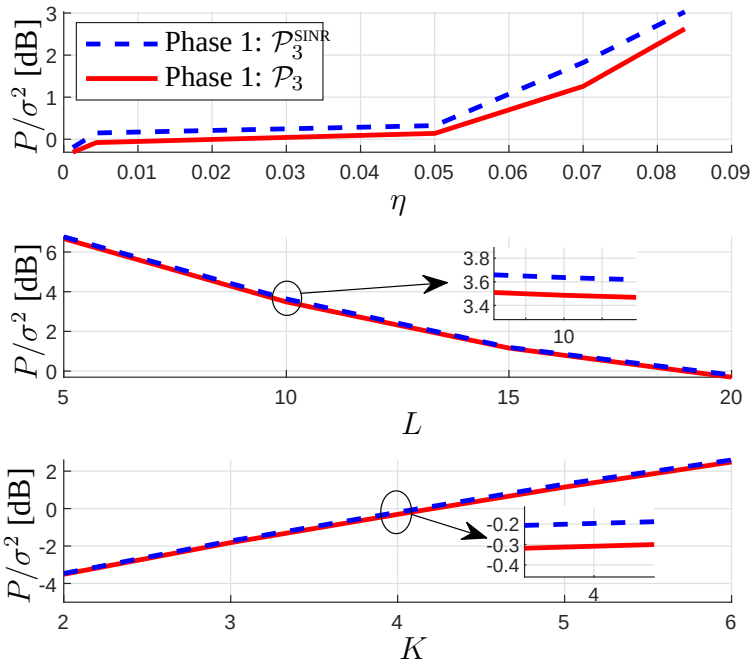

Figure 4.7: Top: normalized transmit power $\frac{P}{\sigma^2}$ vs. η; middle: normalized transmit power $\frac{P}{\sigma^2}$ vs. L; bottom: normalized transmit power $\frac{P}{\sigma^2}$ vs. K.

Performance vs. number of users K

Fig. 4.9 shows the normalized transmit power $\frac{P}{\sigma^2}$ vs. the number of users per group K for the default setup. Our proposed two-phased solutions outperforms the heuristic selection and random search-100 for all values of K considered. Also, the performance of SLR-based and SINR-based approaches is nearly identical. The performance of our proposed SINR-ADMM and SLR-ADMM is close but slightly better than SINR-IPM and SLR-IPM.

Table 4.2 shows the computational advantage of our proposed ADMM-based fast algorithms for difficult values of K. Similar to Table 4.1, our ADMM-based algorithms are substantially faster than IPM, and the SLR-based approach provides lower complexity than the SINR-based approach, with a similar performance.

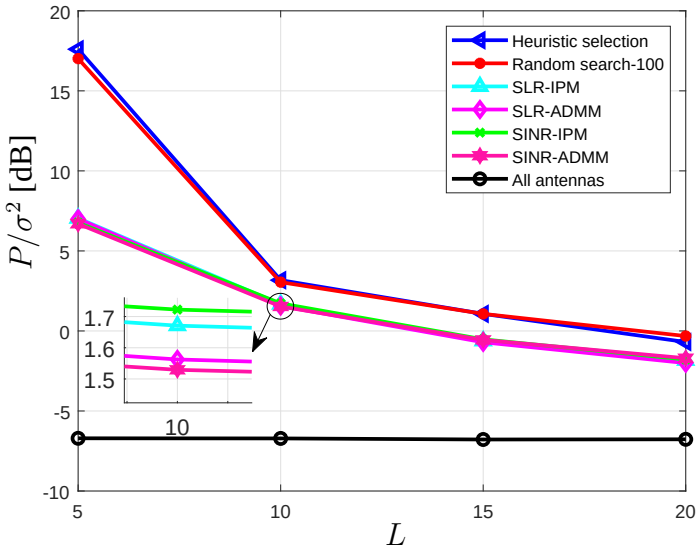

Figure 4.8: Normalized transmit power $\frac{P}{\sigma^2}$ vs. L ($M = 100$, $G = 3$, $K = 4$, $\gamma = 3$ dB, $\eta = 0.01$).

Table 4.2: Average Computation Time Over K (sec.)

K	2	3	4	5	6
SINR-ADMM	46	70	100	110	124
SLR-ADMM	11.5	17.5	25	27.5	31
SINR-IPM	6314.9	9566.8	12119	14769	16869
SLR-IPM	4031	5588.3	6983.1	9372.7	11177

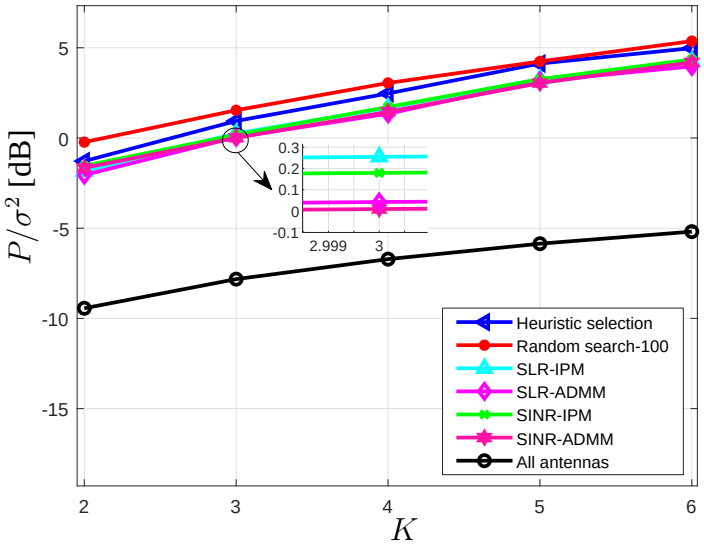

Figure 4.9: Normalized transmit power $\frac{P}{\sigma^2}$ vs. K ($M = 100$, $G = 3$ $L = 10$, $\gamma = 3$ dB, $\eta = 0.01$).

Chapter 5

Conclusion and Future Works

In this chapter, summarize the main results in this dissertation and outline some possible future works.

5.1 Robust multicast Beamforming Design for Massive MIMO

In this work, we have considered the robust multi-group multicast beamforming design under channel uncertainty and proposed an ADMM-based fast algorithm to find a solution. Using channel covariance matrix for beamforming design, we have considered the QoS problem using the worst-case SINR constraints, under the assumption that the estimation error is within a spherical region. We have directly solved this formulated non-convex problem by developing an ADMM-based fast algorithm with convergence guarantee. The algorithm contains two layers of ADMM procedures. The outer-layer ADMM breaks the problem into two subproblems, which are solved iteratively. We have developed special techniques to derive closed-form or semi-closed-form solutions to these subproblems. For one of the subproblems, we have developed an inner-layer consensus-ADMM-based algorithm, which leads to a semi-closed-form solution in each iterative update. The semi-closed-form solution significantly reduces computational complexity and allows us to compute the solution fast for large-scale systems with for a large number of antennas and users. Our simulation results have shown significant power saving offered by our solution compared with existing conservative methods for robust design, while showing substantial

computational saving over the existing methods.

5.2 Joint Antenna Selection and Robust Multicast Beamforming Design for Massive MIMO

In this work, we have considered joint antenna selection and robust multi-group multicast beamforming design in massive MIMO systems with a limited number of RF chains. Our design is based on the estimated channel covariance matrices available at the BS with a given error bound. Our design is to ensure the worst-case SINR meets the target at each user. We reformulate this NP-hard joint robust optimization problem, into a more tractable form, for which we proposed a two-phased approach with antenna selection phase, followed by multicast beamforming optimization. For the antenna selection phase, we have proposed the SINR-based approach and develop a two-layered ADMM-based fast algorithm to solve the joint optimization problem. In particular, our algorithm contains (semi-)closed-form solutions in each iteration update. To further lower the complexity, we have proposed the SLR-based approach for antenna selection and similarly developed a two-layered ADMM-based fast algorithm with closed-form updates in each iteration. With the selected antennas, we can apply the fast algorithm in Chapter **??** for robust multicast beamforming design. The simulation results show that our proposed approaches outperforms other antenna selection methods, and our ADMM-based algorithms provide fast computation of both antenna selection and beamforming solution.

5.3 Future work: Robust Multi-group Multicast Beamforming in Cloud Radio Access Networks(C-RAN)

Due to the nature of our work in the previous studies on large-scale systems, one possible extension of our current work is to extend our algorithms to C-RAN. One of the concerns in wireless communication is the raising demand for the mobile data due to evolution of the wireless communication and the progress in the devices capability of media streaming [131]. On the other hand, increasing the number of users requesting the same content

leads to overwhelming interference in the network [132]. One solution to these problems is C-RAN, in which high-speed backhaul links connect the BSs to cooperate through a baseband unit (BBU) to manage the interference and improve the spectrum efficiency [132]. However, requesting massive CSI for interference management imposes challenges in the C-RAN architecture [132]. Furthermore, the limited capacity of the backhaul links in large-scale systems results in joint multicast beamforming design and backhaul links selection [133].

One possible extension of our work is to tackle the joint robust quality of service problem and backhaul link selection. We can exploit our result on the exact worst-case SINR in the C-RAN to formulate the robust QoS problem. Due to the limited capacity of the backhaul links, a backhaul capacity constraint is added to the problem. This constraint guarantees the maximum data rate from the BBU to remote radio heads (RRHs) is less than the available capacity of the links. A transmit power constraint on each RRH, also, is imposed due to the high interference in the C-RAN. Compared with our previous work, the power consumption in the C-RAN is affected by both the BS and the backhaul links. The challenge here is solving the obtained NP-hard, non-convex, mixed-integer variables problem with low computational complexity while maintaining high performance.

Appendix A

Chapter 3 Appendices

A.1 Derivation of (3.14)

The Lagrangian for problem (3.13) is given by $\mathcal{L}(\mathbf{E}_{ik}, \nu_{ik}) = J(\mathbf{E}_{ik}; \{\mathbf{w}_i\}) + \nu_{ik}(\|\mathbf{E}_{ik}\|_F^2 - \epsilon_{ik}^2)$, where ν_{ik} is the Lagrange multiplier associated with the constraint on $\mathbf{E}_{ik}$. By the KKT conditions, we set the derivative of $\mathcal{L}(\mathbf{E}_{ik}, \nu_{ik})$ w.r.t. $\mathbf{E}_{ik}$ to zero and obtain $\mathbf{E}_{ik}$ as

$$\mathbf{E}_{ik} = -\frac{1}{2\nu_{ik}} \frac{\mathbf{w}_i \mathbf{w}_i^H - \gamma_{ik} \sum_{l \in \mathcal{G}_{-i}} \mathbf{w}_l \mathbf{w}_l^H}{\| \mathbf{w}_i \mathbf{w}_i^H - \gamma_{ik} \sum_{l \in \mathcal{G}_{-i}} \mathbf{w}_l \mathbf{w}_l^H \|_F}. \tag{A.1}$$

From (A.1), we have $\nu_{ik} > 0$; otherwise, the error bound $\|\mathbf{E}_{ik}\|_F^2 \leq \epsilon_{ik}^2$ will not hold. By the complimentary slackness condition $\nu_{ik}(\|\mathbf{E}_{ik}\|_F^2 - \epsilon_{ik}^2) = 0$, It follows that $\|\mathbf{E}_{ik}\|_F^2 = \epsilon_{ik}^2$. Bring this to (A.1), we have (3.14).

A.2 Derivation of (3.33)

Substituting $\mathfrak{Re}\{z\}$ in (3.31) into (3.32), we have

$$\left(\frac{\mathfrak{Re}\{\mathbf{r}\mathbf{C}(\lambda)\mathbf{b}\}-1}{1-\lambda\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{r}}\right)^2(1-\lambda^2\mathbf{r}^H\mathbf{P}(\lambda)\mathbf{r})$$
$$-2\lambda\left(\frac{\mathfrak{Re}\{\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{b}\}-1}{1-\lambda\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{r}}\right)\mathfrak{Re}\{\mathbf{r}^H\mathbf{P}(\lambda)\mathbf{b}\}-\mathbf{b}^H\mathbf{P}(\lambda)\mathbf{b}=0.$$

If $1-\lambda\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{r}\neq 0$, multiplying both sides of the above equation by $(1-\lambda\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{r})^2$, we arrive at the equation in (3.33). Note that as $\lambda\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{r}\rightarrow 1$, the LHS of the above equation goes to either $-\infty$ or $+\infty$. Thus, λ satisfying $\lambda\mathbf{r}^H\mathbf{C}(\lambda)\mathbf{r}=1$ is not the root of (3.32). Therefore, solving (3.32) is equivalent to solving (3.33).

A.3 Derivation of $\mathfrak{Re}\{\bar{z}\}$ and $\tilde{\lambda}$ for $\mathbf{w}^{\mathrm{o}}$ in (3.47)

Since $\mathbf{w}^{\mathrm{o}}$ in (3.47) has exactly the same form as $\mathbf{v}^{\mathrm{o}}$ in (3.29), the expression of $\mathfrak{Re}(\bar{z})$ has the same structure as in (3.31), given by

$$\mathfrak{Re}\{\bar{z}\} = \frac{\mathfrak{Re}\{\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{b}}\} - \tilde{\lambda}\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{r}}}{1 - \tilde{\lambda}\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{r}}} \tag{A.2}$$

where $\bar{\mathbf{C}}(\tilde{\lambda}) \triangleq (\bar{\mathbf{I}} + \tilde{\lambda}\mathbf{V}_{\mathrm{c}})^{-1}$. To determine $\mathbf{w}^{\mathrm{o}}$, we only need to find $\tilde{\lambda}$. Similar to (3.33), by substituting the expressions of $\mathbf{w}^{\mathrm{o}}$ in (3.47) and $\mathfrak{Re}(\bar{z})$ in (A.2) into the equality constraint in (3.44) of $\mathcal{P}'_{\mathbf{w}}$, we obtain $\tilde{\lambda}$ as a root of the following equation:

$$\begin{aligned}(\mathfrak{Re}\{\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\mathbf{b}\} - 1)^2(1 - \tilde{\lambda}^2 \bar{\mathbf{r}}^H \bar{\mathbf{P}}(\tilde{\lambda})\bar{\mathbf{r}}) \\ - 2\tilde{\lambda}(\mathfrak{Re}\{\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{b}}\} - 1)(1 - \tilde{\lambda}\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{r}})\mathfrak{Re}\{\bar{\mathbf{r}}^H \bar{\mathbf{P}}(\tilde{\lambda})\bar{\mathbf{b}}\}) \\ - \bar{\mathbf{b}}^H \bar{\mathbf{P}}(\tilde{\lambda})\bar{\mathbf{b}}(1 - \tilde{\lambda}\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{r}})^2 = 0\end{aligned} \tag{A.3}$$

where $\bar{\mathbf{P}}(\tilde{\lambda}) \triangleq \bar{\mathbf{C}}(\tilde{\lambda})\mathbf{V}_{\mathrm{c}}\bar{\mathbf{C}}(\tilde{\lambda})$.

We further simplify (A.3) by examining the structure of $\bar{\mathbf{C}}(\tilde{\lambda})$. Define

$$\bar{\mathbf{V}}_{\mathrm{c}} \triangleq \bar{\mathbf{I}}^{-\frac{1}{2}} \mathbf{V}_{\mathrm{c}} \bar{\mathbf{I}}^{-\frac{1}{2}}. \tag{A.4}$$

Then, we have $\bar{\mathbf{C}}(\tilde{\lambda}) = \bar{\mathbf{I}}^{-\frac{1}{2}}(\mathbf{I} + \tilde{\lambda}\bar{\mathbf{V}}_{\mathrm{c}})^{-1}\bar{\mathbf{I}}^{-\frac{1}{2}}$. Applying this into the definition of $\mathbf{P}(\tilde{\lambda})$, we have $\bar{\mathbf{P}}(\tilde{\lambda}) = \bar{\mathbf{I}}^{-\frac{1}{2}}(\mathbf{I} + \tilde{\lambda}\bar{\mathbf{V}}_{\mathrm{c}})^{-1}\bar{\mathbf{V}}_{\mathrm{c}}(\mathbf{I} + \tilde{\lambda}\bar{\mathbf{V}}_{\mathrm{c}})^{-1}\bar{\mathbf{I}}^{-\frac{1}{2}}$. Consider the eigenvalue decomposition of $\bar{\mathbf{V}}_{\mathrm{c}}$ in (A.4). Recall that $\bar{\mathbf{I}} = \mu\mathbf{I} + \rho\overline{\mathbf{D}}^2$, where $\overline{\mathbf{D}} = \mathbf{D} \otimes \mathbf{I}_M$. We can rewrite $\bar{\mathbf{I}} = \mu\mathbf{I}_G \otimes \mathbf{I}_M + \rho\mathbf{D}^2 \otimes \mathbf{I}_M = \mathbf{D}_{\mathrm{c}} \otimes \mathbf{I}_M$, where $\mathbf{D}_{\mathrm{c}} \triangleq \mu\mathbf{I}_G + \rho\mathbf{D}^2$. Recall that $\mathbf{V}_{\mathrm{c}} = \frac{\epsilon^2}{(\sigma^2\gamma)^2}\mathbf{V}^T\mathbf{V}^{j*} \otimes \mathbf{I}_M$. Substituting the expressions of $\mathbf{V}_{\mathrm{c}}$ and $\bar{\mathbf{I}}$ in (A.4), we have

$$\bar{\mathbf{V}}_{\mathrm{c}} = \frac{\epsilon^2}{(\sigma^2\gamma)^2}\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{V}^*\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}} \otimes \mathbf{I}_M. \tag{A.5}$$

Consider the following eigenvalue decomposition $\frac{\epsilon^2}{(\sigma^2\gamma)^2}\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{V}^*\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}} = \mathbf{U}_{\mathrm{v}}\mathbf{\Sigma}_{\mathrm{v}}\mathbf{U}_{\mathrm{v}}^H$, where $\mathbf{\Sigma}_{\mathrm{v}} = \mathrm{diag}(\tilde{\sigma}_1^2, \cdots, \tilde{\sigma}_G^2)$, with $\tilde{\sigma}_j^2$ being the jth eigenvalue of $\mathbf{V}_{\mathrm{v}}$ and $\mathbf{U}_{\mathrm{v}}$ being a unitary matrix. Then, we have $\bar{\mathbf{V}}_{\mathrm{c}} = \mathbf{U}_{\mathrm{v}}\mathbf{\Sigma}_{\mathrm{v}}\mathbf{U}_{\mathrm{v}}^H \otimes \mathbf{I}_M$. Using this structure of $\bar{\mathbf{V}}_{\mathrm{c}}$, we have $\bar{\mathbf{C}}(\tilde{\lambda}) = \mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{U}_{\mathrm{v}}(\mathbf{I} + \tilde{\lambda}\mathbf{\Sigma}_{\mathrm{v}})^{-1}\mathbf{U}_{\mathrm{v}}^H\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}} \otimes \mathbf{I}_M$ and $\bar{\mathbf{P}}(\tilde{\lambda}) = \mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{U}_{\mathrm{v}}(\mathbf{I} + \tilde{\lambda}\mathbf{\Sigma}_{\mathrm{v}})^{-1}\tilde{\lambda}\mathbf{\Sigma}_{\mathrm{v}}(\mathbf{I} + \tilde{\lambda}\mathbf{\Sigma}_{\mathrm{v}})^{-1}\mathbf{U}_{\mathrm{v}}^H\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}} \otimes \mathbf{I}_M$. This leads to the simplification of the terms in (A.3) as follows:

$$\mathfrak{Re}\{\bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{b}}\} = \sum_{j=1}^{G} \frac{\mathfrak{Re}\{\tilde{\mathbf{r}}_j^H \tilde{\mathbf{b}}_j\}}{1 + \tilde{\lambda}\tilde{\sigma}_j^2},\ \bar{\mathbf{r}}^H \bar{\mathbf{C}}(\tilde{\lambda})\bar{\mathbf{r}} = \sum_{j=1}^{G} \frac{\|\tilde{\mathbf{r}}_j\|^2}{1 + \tilde{\lambda}\tilde{\sigma}_j^2}$$

$$\bar{\mathbf{r}}^H\bar{\mathbf{P}}(\tilde{\lambda})\bar{\mathbf{r}} = \sum_{j=1}^{G}\frac{\tilde{\sigma}_j^2\|\tilde{\mathbf{r}}_j\|^2}{(1+\tilde{\lambda}\tilde{\sigma}_j^2)^2},\ \bar{\mathbf{b}}^H\bar{\mathbf{P}}(\tilde{\lambda})\bar{\mathbf{b}} = \sum_{j=1}^{G}\frac{\tilde{\sigma}_j^2\|\tilde{\mathbf{b}}_j\|^2}{(1+\tilde{\lambda}\tilde{\sigma}_j^2)^2},$$
$$\mathfrak{Re}\{\bar{\mathbf{r}}^H\bar{\mathbf{P}}(\tilde{\lambda})\bar{\mathbf{b}}\} = \sum_{j=1}^{G}\frac{\tilde{\sigma}_j^2\mathfrak{Re}\{\tilde{\mathbf{r}}_j^H\tilde{\mathbf{b}}_j\}}{(1+\tilde{\lambda}\tilde{\sigma}_j^2)^2}$$

where $\tilde{\mathbf{b}} \triangleq (\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{U}_{\mathrm{v}}\otimes\mathbf{I}_M)^H\bar{\mathbf{b}} = [\tilde{\mathbf{b}}_1^H,\ldots,\tilde{\mathbf{b}}_G^H]^H$, with $\tilde{\mathbf{b}}_j\in\mathbb{C}^{M\times 1}$, $j=1,\ldots,M$, and $\tilde{\mathbf{r}} \triangleq (\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{U}_{\mathrm{v}}\otimes\mathbf{I}_M)^H\bar{\mathbf{r}} = [\tilde{\mathbf{r}}_1^H,\ldots,\tilde{\mathbf{r}}_G^H]^H$, with $\tilde{\mathbf{r}}_j\in\mathbb{C}^{M\times 1}$, $j=1,\ldots,M$. Note that from the definitions of $\bar{\mathbf{b}}$ and $\bar{\mathbf{r}}$, we can efficiently compute $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{r}}$ as $\tilde{\mathbf{b}} = \mathrm{vec}((\rho\overline{\mathbf{Z}}\mathbf{D}_{\mathrm{c}} + \mu\overline{\mathbf{Y}})\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{U}_{\mathrm{v}}^*)$ and $\tilde{\mathbf{r}} = \frac{1}{\sigma^2\gamma}\mathrm{vec}(\overline{\mathbf{R}}\mathbf{D}_{\mathrm{c}}^{-\frac{1}{2}}\mathbf{U}_{\mathrm{v}}^*)$. Based on the above, (A.3) can be simplified as

$$\begin{aligned}
&\Big(\sum_{j=1}^{G}\frac{\mathfrak{Re}\{\tilde{\mathbf{r}}_j^H\tilde{\mathbf{b}}_j\}}{1+\tilde{\lambda}\tilde{\sigma}_j^2}-1\Big)^2\Big(1-\tilde{\lambda}^2\sum_{j=1}^{G}\frac{\tilde{\sigma}_j^2\|\tilde{\mathbf{r}}_j\|^2}{(1+\tilde{\lambda}\tilde{\sigma}_j^2)^2}\Big)\\
&-2\tilde{\lambda}\Big(\sum_{j=1}^{G}\frac{\mathfrak{Re}\{\tilde{\mathbf{r}}_j^H\tilde{\mathbf{b}}_j\}}{1+\tilde{\lambda}\tilde{\sigma}_j^2}-1\Big)\Big(1-\tilde{\lambda}\sum_{j=1}^{G}\frac{\|\tilde{\mathbf{r}}_j\|^2}{1+\tilde{\lambda}\tilde{\sigma}_j^2}\Big)\sum_{j=1}^{G}\frac{\tilde{\sigma}_j^2\|\tilde{\mathbf{r}}_j\|^2}{(1+\tilde{\lambda}\tilde{\sigma}_j^2)^2}\\
&-\sum_{j=1}^{G}\frac{\tilde{\sigma}_j^2\|\tilde{\mathbf{b}}_j\|^2}{(1+\tilde{\lambda}\tilde{\sigma}_j^2)^2}\Big(1-\tilde{\lambda}\sum_{j=1}^{G}\frac{\|\tilde{\mathbf{r}}_j\|^2}{1+\tilde{\lambda}\tilde{\sigma}_j^2}\Big)^2=0.
\end{aligned}\tag{A.6}$$

As the LHS of equation (A.6) is a smooth polynomial function of $\tilde{\lambda}$, its roots can be easily found. With the constraint $\mathfrak{Re}\{\bar{z}\}-1\geq 0$ imposed on $\mathcal{P}'_{\mathbf{w}}$, we select the root of (A.6) that results in the minimum objective value of $\mathcal{P}'_{\mathbf{w}}$ while satisfying $\mathfrak{Re}\{\bar{z}\}-1\geq 0$.

Bibliography

[1] W. Xiang, K. Zheng, and X. S. Shen, *5G mobile communications*. Springer, 2016.

[2] K. Zheng, L. Zhang, W. Xiang, and W. Wang, "Architecture of heterogeneous vehicular networks," in *Heterogeneous Vehicular Networks*. Springer, 2016, pp. 9–24.

[3] L. Zhao, H. Zhao, K. Zheng, and W. Xiang, *Massive MIMO in 5G networks: selected applications*. Springer, 2018.

[4] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo *et al.*, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, 2012.

[5] Y. S. Soh, T. Q. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 5, pp. 840–850, 2013.

[6] N. Deng, W. Zhou, and M. Haenggi, "Heterogeneous cellular network models with dependence," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2167–2181, 2015.

[7] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[8] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, 2020.

[9] F. Amirnavaei and M. Dong, "Online power control optimization for wireless transmission with energy harvesting and storage," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4888–4901, 2016.

[10] M. Dong, W. Li, and F. Amirnavaei, "Online joint power control for two-hop wireless relay networks with energy harvesting," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 463–478, 2017.

[11] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, 2012.

[12] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, 2011.

[13] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 167–178, 2012.

[14] R. C. de Lamare, "Massive MIMO systems: Signal processing challenges and future trends," *URSI Radio Science Bulletin*, vol. 2013, no. 347, pp. 8–20, 2013.

[15] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[16] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[17] 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 Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[18] 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.

[19] C. de Morais Cordeiro, H. Gossain, and D. P. Agrawal, "Multicast over wireless mobile ad hoc networks: present and future directions," *IEEE network*, vol. 17, no. 1, pp. 52–59, 2003.

[20] J.-M. Vella and S. Zammit, "A survey of multicasting over wireless access networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 718–753, 2012.

[21] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[22] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, pp. 2239–2251, Jun. 2006.

[23] 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 ICASSP*, 2021, pp. 4790–4794.

[24] J. Fink, R. L. G. Cavalcante, and S. Staczak, "Multi-group multicast beamforming by superiorized projections onto convex sets," *IEEE Trans. Signal Process.*, vol. 69, pp. 5708–5722, 2021.

[25] C. Zhang, M. Dong, and B. Liang, "Fast first-order algorithm for large-scale max-min fair multi-group multicast beamforming," *IEEE Wireless Commun. Lett.*, 2022.

[26] S. Mohammadi, M. Dong, and S. ShahbazPanahi, "Fast algorithm for joint unicast and multicast beamforming for large-scale massive MIMO," *IEEE Trans. Signal Process.*, vol. 70, pp. 5413–5428, 2022.

[27] C. Zhang, M. Dong, and B. Liang, "Ultra-low-complexity algorithms with structurally optimal multi-group multicast beamforming in large-scale systems," *arXiv preprint arXiv:2206.01846*, 2022.

[28] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, pp. 186–195, Feb. 2014.

[29] Z. Xiang, M. Tao, and X. Wang, "Coordinated multicast beamforming in multicell networks," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 12–21, Jan. 2013.

[30] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, pp. 1268–1279, Mar. 2008.

[31] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4916–4927, 2007.

[32] E. Chen and M. Tao, "ADMM-based fast algorithm for multi-group multicast beamforming in large-scale wireless systems," *IEEE Trans. Commun.*, vol. 65, pp. 2685–2698, Jun. 2017.

[33] 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.

[34] A. Z. Yalccın and Y. Yapıcı, "Max-min fair beamforming for cooperative multigroup multicasting with rate-splitting," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 254–268, 2020.

[35] T. Weber, A. Sklavos, and M. Meurer, "Imperfect channel-state information in MIMO transmission," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 543–552, 2006.

[36] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5g wireless communications: A deep learning approach," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 227–236, 2020.

[37] J. P. González-Coma, P. Suárez-Casal, P. M. Castro, and L. Castedo, "FDD channel estimation via covariance estimation in wideband massive MIMO systems," *Sensors*, vol. 20, no. 3, p. 930, 2020.

[38] Z. Du, X. Song, J. Cheng, and N. C. Beaulieu, "Maximum likelihood based channel estimation for macrocellular OFDM uplinks in dispersive time-varying channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 176–187, 2010.

[39] E. Karami, "Tracking performance of least squares MIMO channel estimation algorithm," *IEEE Trans. Commun.*, vol. 55, no. 11, pp. 2201–2209, 2007.

[40] J. Ma, S. Zhang, H. Li, N. Zhao, and A. Nallanathan, "Iterative LMMSE individual channel estimation over relay networks with multiple antennas," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 423–435, 2017.

[41] W. Utschick, V. Rizzello, M. Joham, Z. Ma, and L. Piazzi, "Learning the CSI recovery in FDD systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6495–6507, 2022.

[42] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, 2019.

[43] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Massive MIMO with spatially correlated rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, 2019.

[44] K. Upadhya and S. A. Vorobyov, "Covariance matrix estimation for massive mimo," *IEEE Signal Processing Lett.*, vol. 25, no. 4, pp. 546–550, 2018.

[45] Q. Gong, Z. Chen, and G. Wei, "Downlink multicasting beamforming with imperfect CSI on both transceiver sides," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, 2012, pp. 1472–1477.

[46] M.-C. Yue, S. X. Wu, and A. M.-C. So, "A robust design for MISO physical-layer multicasting over line-of-sight channels," *IEEE Signal Processing Lett.*, vol. 23, no. 7, pp. 939–943, 2016.

[47] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, 2016.

[48] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, 2016.

[49] I. Wajid, M. Pesavento, Y. C. Eldar, and D. Ciochina, "Robust downlink beamforming with partial channel state information for conventional and cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3656–3670, 2013.

[50] G. Zang, H. V. Cheng, Y. Cui, W. Liu, F. Yang, L. Ding, and H. Liu, "Low complexity algorithms for robust multigroup multicast beamforming," *IEEE Commun. Lett.*, vol. 23, no. 8, pp. 1409–1412, 2019.

[51] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. of Sci. Comput.*, vol. 78, no. 1, pp. 29–63, 2019.

[52] O. Tervo, L.-N. Tran, H. Pennanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multicell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, 2018.

[53] O. T. Demir and T. E. Tuncer, "Antenna selection and hybrid beamforming for simultaneous wireless information and power transfer in multi-group multicasting systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6948–6962, 2016.

[54] N. Mohamadi, M. Dong, and S. ShahbazPanahi, "Low-complexity ADMM-based algorithm for robust multi-group multicast beamforming in large-scale systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2046–2061, 2022.

[55] ——, "ADMM-based fast algorithm for robust multi-group multicast beamforming," in *Proc. IEEE ICASSP*, 2021, pp. 4440–4444.

[56] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," Sep. 1994, uS Patent 5,345,599.

[57] G. D. Golden, C. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics letters*, vol. 35, no. 1, pp. 14–16, 1999.

[58] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE transactions on wireless communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[59] D. NTT, "Requirements, candidate solutions & technology roadmap for lte rel-12 onward," *3GPP RWS-120010*, 2012.

[60] Samsung, "Technologies for Rel-12 and onward," Nov. 2013.

[61] HUAWEI and HiSilicon, "Views on Rel-12 and onwards for LTE and UMTS," 2013.

[62] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the ul/dl of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[63] 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.

[64] A. Shaikh and M. J. Kaur, "Comprehensive survey of massive MIMO for 5g communications," in *2019 Advances in Science and Engineering Technology International Conferences (ASET)*, 2019, pp. 1–5.

[65] S. X. Wu, W. Ma, and A. M. So, "Physical-layer multicasting by stochastic transmit beamforming and Alamouti space-time codings," *IEEE Trans. Signal Process.*, pp. 4230–4245, Sep. 2013.

[66] M. Sadeghi, E. Bjrnson, E. G. Larsson, C. Yuen, and T. L. Marzetta, "Maxmin fair transmit precoding for multi-group multicasting in massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1358–1373, 2018.

[67] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Weighted fair multicast multigroup beamforming under per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 62, pp. 5132–5142, Oct. 2014.

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

[69] 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, Jun. 2017.

[70] J. Yu and M. Dong, "Low-complexity weighted MRT multicast beamforming in massive MIMO cellular networks," in *Proc. IEEE ICASSP*, Apr. 2018, pp. 3849–3853.

[71] ——, "Distributed low-complexity multi-cell coordinated multicast beamforming with large-scale antennas," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun.*, Jun. 2018, pp. 1–5.

[72] O. Tervo, L. Tran, H. Pennanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multicell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, 2018.

[73] M. Dong and B. Liang, "Multicast relay beamforming through dual approach," in *Proc. IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, Dec. 2013, pp. 492–495.

[74] S. Mohammadi, M. Dong, and S. ShahbazPanahi, "Fast algorithm for joint unicast and multicast beamforming in large-scale systems," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2021, pp. 91–95.

[75] 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.

[76] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup beamforming for per-antenna power constrained large-scale arrays," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun.*, Jun. 2015, pp. 271–275.

[77] 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, pp. 2963–2975, May 2017.

[78] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[79] F. A. P. De Figueiredo, F. A. C. M. Cardoso, I. Moerman, and G. Fraidenraich, "Channel estimation for massive MIMO TDD systems assuming pilot contamination and frequency selective fading," *IEEE Access*, vol. 5, pp. 17 733–17 741, 2017.

[80] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 714–725, Oct. 2009.

[81] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, pp. 5690–5705, 2014.

[82] F. Sohrabi and T. N. Davidson, "Coordinate update algorithms for robust power loading for the MU-MISO downlink with outage constraints," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2761–2773, 2016.

[83] X. He and Y.-C. Wu, "Probabilistic QoS constrained robust downlink multiuser MIMO transceiver design with arbitrarily distributed channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6292–6302, 2013.

[84] A. Pascual-Iserte, D. Palomar, A. Perez-Neira, and M. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, 2006.

[85] 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, pp. 40–60, Jan. 2013.

[86] Q. Gong, Z. Chen, and G. Wei, "Downlink multicasting beamforming with imperfect CSI on both transceiver sides," in *Proc. IEEE PIMRC*, 2012, pp. 1472–1477.

[87] 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.

[88] M.-C. Yue, S. X. Wu, and A. M.-C. So, "A robust design for MISO physical-layer multicasting over line-of-sight channels," *IEEE Signal Processing Lett.*, vol. 23, no. 7, pp. 939–943, 2016.

[89] L. You, A. Liu, W. Wang, and X. Gao, "Outage constrained robust multigroup multicast beamforming for multi-beam satellite communication systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 352–355, 2019.

[90] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[91] Q.-D. Vu, K.-G. Nguyen, and M. Juntti, "Weighted maxmin fairness for C-RAN multicasting under limited fronthaul constraints," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1534–1548, 2018.

[92] Y. Shi, J. Zhang, and K. B. Letaief, "Robust group sparse beamforming for multicast green cloud-RAN with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4647–4659, 2015.

[93] Y. Chen, S. He, Y. Huang, J. Ren, and L. Yang, "Robust multigroup multicast beamforming design for backhaul-limited cloud radio access network," *IEEE Signal Processing Lett.*, vol. 26, no. 1, pp. 189–193, 2019.

[94] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," *Handbook of Ant. in Wireless Communi.,[ed] Lal C. Godara, CRC Press*, pp. 18–1–18–33, 2001.

[95] S. Shahbazpanahi, A. B. Gershman, Zhi-Quan Luo, and Kon Max Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, 2003.

[96] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[97] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexingthe large-scale array regime," *IEEE Trans. Inform. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[98] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Processing Lett.*, vol. 25, no. 6, pp. 863–867, 2018.

[99] K. L. Law, I. Wajid, and M. Pesavento, "Robust downlink beamforming in multigroup multicasting using trace bounds on the covariance mismatches," in *Proc. IEEE ICASSP*, 2012, pp. 3229–3232.

[100] W. Zhang, Z. Chen, Q. Gong, and G. Wei, "Robust multicast beamforming using covariance channel state information," in *IEEE Int. Conf. on Commun. in China (ICCC)*, 2012, pp. 636–641.

[101] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, 2004.

[102] A. Dua, K. Medepalli, and A. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2353–2357, 2006.

[103] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, 2005.

[104] S. Payami, M. Ghoraishi, M. Dianati, and M. Sellathurai, "Hybrid beamforming with a reduced number of phase shifters for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4843–4851, 2018.

[105] S. Y. Park and D. J. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2524–2534, 2008.

[106] P. Somol, P. Pudil, and J. Kittler, "Fast branch and bound algorithms for optimal feature selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 900–912, 2004.

[107] Z. Zhang and M. Tao, "A learning based branch-and-bound algorithm for single-group multicast beamforming," in *Proc. IEEE Global Telecommn. Conf. (GLOBECOM)*, 2021, pp. 1–6.

[108] M. S. Ibrahim, A. Konar, and N. D. Sidiropoulos, "Fast algorithms for joint multicast beamforming and antenna selection in massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 1897–1909, 2020.

[109] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, 2013.

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

[111] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 83–88, 2012.

[112] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[113] L. Chen, D. Sun, and K.-C. Toh, "A note on the convergence of ADMM for linearly constrained convex optimization problems," *Computational Optimization and Applications*, vol. 66, no. 2, pp. 327–343, 2017.

[114] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, "A general analysis of the convergence of ADMM," in *International Conference on Machine Learning*. PMLR, 2015, pp. 343–352.

[115] B. He, S. Xu, and X. Yuan, "Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints," *arXiv preprint arXiv:2107.01897*, 2021.

[116] Y. Xu, H. Deng, and W. Zhu, "Synchronous distributed ADMM for consensus convex optimization problems with self-loops," *Information Sciences*, vol. 614, pp. 185–205, 2022.

[117] W. Deng, M.-J. Lai, Z. Peng, and W. Yin, "Parallel multi-block ADMM with o (1/k) convergence," *Journal of Scientific Computing*, vol. 71, no. 2, pp. 712–736, 2017.

[118] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of admm for multi-block convex minimization problems is not necessarily convergent," *Mathematical Programming*, vol. 155, no. 1, pp. 57–79, 2016.

[119] Y. Shen, Q. Gao, and X. Yin, "A multi-parameter parallel ADMM for multi-block linearly constrained separable convex optimization," *Applied Numerical Mathematics*, vol. 171, pp. 369–388, 2022.

[120] Y. Shen, X. Zhang, and X. Zhang, "A partial ppa block-wise ADMM for multi-block linearly constrained separable convex optimization," *Optimization*, vol. 70, no. 3, pp. 631–657, 2021.

[121] X. Chang, S. Liu, P. Zhao, and X. Li, "Convergent prediction–correction-based ADMM for multi-block separable convex programming," *Journal of Computational and Applied Mathematics*, vol. 335, pp. 270–288, 2018.

[122] X. Wang, J. Yan, B. Jin, and W. Li, "Distributed and parallel ADMM for structured nonconvex optimization problem," *IEEE transactions on cybernetics*, vol. 51, no. 9, pp. 4540–4552, 2019.

[123] M. Hong, "A distributed, asynchronous, and incremental algorithm for nonconvex optimization: An admm approach," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 935–945, 2017.

[124] J. Guo, G. Hug, and O. Tonguz, "Asynchronous ADMM for distributed non-convex optimization in power systems," *arXiv preprint arXiv:1710.08938*, 2017.

[125] C. Zhang, Y. Song, X. Cai, and D. Han, "An extended proximal ADMM algorithm for three-block nonconvex optimization problems," *Journal of Computational and Applied Mathematics*, vol. 398, p. 113681, 2021.

[126] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Math. Program.*, vol. 162, no. 1-2, pp. 165–199, 2017.

[127] L. Li, X. Wang, and G. Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables," *Math. Prob. in Eng.*, 2015.

[128] 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.

[129] A. F. Izmailov and M. V. Solodov, "Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of newton-type methods," *Mathematical Programming*, vol. 95, no. 3, pp. 631–650, 2003.

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

[131] Y. Li, M. Xia, and Y.-C. Wu, "First-order algorithm for content-centric sparse multicast beamforming in large-scale c-ran," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 5959–5974, 2018.

[132] Y. Chen, S. He, Y. Huang, J. Ren, and L. Yang, "Robust multigroup multicast beamforming design for backhaul-limited cloud radio access network," *IEEE Signal Processing Lett.*, vol. 26, no. 1, pp. 189–193, 2019.

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