OULUN YLIOPISTO
UNIVERSITY of OULU
DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

# MASTER'S THESIS

## PRECODER DESIGN FOR MULTI-ANTENNA TRANSMISSION IN MU-MIMO WITH QoS REQUIREMENTS

| | |
|---|---|
| Author | Ayswarya Padmanabhan |
| Supervisor | Prof. Markku Juntti |
| Second Supervisor | Dr. Antti Tölli |
| Technical Advisor | Dr. Le-Nam Tran |

**February 2016**

**Ayswarya P. (2016) Precoder Design for Multi-Antenna Transmission in MU-MIMO with QoS Requirements.** University of Oulu, Center for Wireless Communications - Radio Technologies, Master's Degree Program in Wireless Communications Engineering. Master's thesis, 64 p.

# ABSTRACT

**A multiple-input multiple-output (MIMO) interference broadcast channel (IBC) channel is considered. There are several base stations (BSs) transmitting useful information to their own users and unwanted interference to its neighboring BS users. Our main interest is to maximize the system throughput by designing transmit precoders with weighted sum rate maximization (WSRM) objective for a multi-user (MU)-MIMO transmission. In addition, we include the quality of service (QoS) requirement in terms of guaranteed minimum rate for the users in the system. Unfortunately, the problem considered is nonconvex and known to be non-deterministic polynomial (NP) hard. Therefore, to determine the transmit precoders, we first propose a centralized precoder design by considering two closely related approaches, namely, direct signal-to-interference-plus-noise-ratio (SINR) relaxation via sequential parametric convex approximation (SPCA), and mean squared error (MSE) reformulation. In both approaches, we adopt successive convex approximation (SCA) technique to solve the nonconvex optimization problem by solving a sequence of convex subproblems. Due to the huge signaling requirements in the centralized design, we propose two different distributed precoder designs, wherein each BS determines only the relevant set of transmit precoders by exchanging minimal information among the coordinating BSs. Initially, we consider designing precoders in a decentralized manner by using alternating directions method of multipliers (ADMM), wherein each BS relaxes inter-cell interference as an optimization variable by including it in the objective. Then, we also propose a distributed precoder design by solving the Karush-Kuhn-Tucker (KKT) expressions corresponding to the centralized problems. Numerical simulations are provided to compare different system configurations with QoS constraints for both centralized and distributed algorithms.**

**Keywords: ADMM, AO, IBC, KKT, MSE reformulation, MU-MIMO, nonconvex optimization, precoder design, QoS, SCA, WSRM.**

# TABLE OF CONTENTS

# FOREWORD

This master's thesis is focused on distributed precoder design for MU-MIMO with QoS constraints. I would like to thank Centre for wireless Communication (CWC) and University of Oulu for providing me the chance to do the thesis work.

I would like to express my sincere gratitude to Prof. Markku Juntti for providing me an opportunity to work under his guidance and support. I am very grateful to Dr. Le-Nam Tran for his meticulous support and encouragement during my thesis work. The knowledge that he has provided me during my work is invaluable and incommensurable. Being a WCE student, I cannot forget to mention Dr. Antti Tölli, who has taught wireless communications II course, which has laid a strong foundation on both linear algebra and optimization in addition to the wireless concepts. I also take this opportunity to thank all the professors at CWC and University of Oulu for their lectures and teachings. Finally, I would like to thank Prof. Kari Kärkkäinen for providing me an opportunity to join WCE program at University of Oulu. It would be impossible for me to study in this prestigious university without his constant support and advices.

It is said that one should never thank your husband Ganesh Venkatraman but I have no better word than thanking him who has been my equestrian in helping me chase and follow my dream. He stood by me as a guide a mentor and also a teacher. Every time I broke down he encouraged me with ideas and corrected both my work and my understanding.

Last but not the least, I dedicate this thesis to my son Adhithya Sriram, and also extend my sincere appreciation to my father Padmanabhan, mother Santhi and brother Ram Prasad for always being a supportive pillar. Finally, I also thank my husbands family who trusted me and gave me constant encouragement to do the studies and for their love and support.

# LIST OF ABBREVIATIONS AND SYMBOLS

## Abbreviations

| | |
|---|---|
| ADMM | Alternating directions method of multipliers |
| AO | Alternating optimization |
| AWGN | Additive white Gaussian noise |
| BER | Bit error rate |
| BS | Base station |
| CDMA | Code division multiple access |
| CSI | Channel State Indormation |
| DL | Downlink |
| DPC | Dirty paper coding |
| DoF | Degree of freedom |
| FDM | Frequency division multiplexing |
| FDMA | Frequency division multiple access |
| GBS | Guaranteed bit rate |
| IBC | Interference broadcast channel |
| IC | Interference channel |
| IA | Interference Alignment |
| IID | Independent and identically distributed |
| ISI | Inter-symbol interference |
| KKT | Karush-Kuhn-Tucker |
| LTE | Long term evolution |
| MIMO | Multiple-input multiple-output |
| MU | Multi-user |
| MISO | Multiple-input single-output |
| MSE | Mean squared error |
| MMSE | Minimum mean squared error |
| MRT | Maximum ratio transmission |
| ML | Maximum likelihood |
| NP | Non-deterministic polynomial |
| OTA | Over-the-air |
| OFDM | Orthogonal frequency division multiplexing |
| PL | Pathloss |
| QoS | Quality of Service |
| SDMA | Space-division multiple access |
| SVD | Singular value decomposition |

| | |
|---|---|
| SNR | Signal-to-noise-ratio |
| SIR | Signal-to-interference-ratio |
| SINR | Signal-to-interference-plus-noise-ratio |
| STBC | Space-time block codes |
| STTC | Space-time trellis codes |
| SOC | Second order cone |
| SOCP | Second order cone programming |
| SCA | Successive convex approximation |
| SPCA | Sequential parametric convex approximation |
| TDM | Time-division multiplexing |
| TDMA | Time-division multiplex access |
| UL | Uplink |
| VoIP | Voice over IP |
| WSRM | Weighted sum rate maximization |
| ZF | Zero-forcing |

## Set Representations

| | |
|---|---|
| $\mathbb{R}$ | Real Number |
| $\mathbb{S}$ | A subset of users |
| $\mathbb{C}$ | Complex number |

## Scalars Vectors and Matrices

| | |
|---|---|
| $K$ | Total number of users in the system |
| $a_k, e_k, c_k, r_k$ | Coupling variables |
| $d_k$ | Data symbol corresponding to user $k$ |
| $d'_k$ | Estimated data symbol of user $k$ |
| $\mathbf{x}_k$ | Transmitted signal vector corresponding to user $k$ |
| $\mathbf{y}_k$ | Received signal at user $k$ |
| $b_i$ | BS that serves user $i$ |
| $\mathbf{h}_{b_i,k}$ | Channel (row) vector from BS $b_i$ to user $k$ |
| $\mathbf{n}_k$ | Noise vector at the receiver |
| $\mathbf{u}_k$ | Receive beamformer corresponding to user k |
| $\mathbf{w}_k$ | Transmit beamformer for user $k$ |
| $\mathbf{w}_i$ | Transmit beamformer of interfering user $i$ |
| $\mathcal{B}$ | Coordinated BS |
| $\mathcal{U}_b$ | Set of all coordinating BS indices |

| | |
|---|---|
| $R_k$ | Minimum mean squared error receiver |
| $P_b$ | Power at BS $b$ |
| $C$ | Capacity of a MIMO channel |
| $\mathcal{CN}$ | Circularly symmetric complex Gaussian distribution |
| $\mathbf{I}$ | Identity matrix |
| $N_{\text{B}}$ | Number of BS |
| $N_{\text{T}}$ | Number of transmit antennas |
| $N_{\text{R}}$ | Number of recieve antennas |
| $N_{\text{min}}$ | $\min(N_{\text{T}}, N_{\text{R}})$ |
| $\mathbf{H}_{b,k}$ | MIMO Channel between user $k$ and BS $b$ |
| $\alpha_k$ | A positive weighting factor |
| $\beta_k$ | Sum of total Interference and Noise |
| $\eta_k$ | MSE for a data symbol $d_k$ |
| $\bar{\eta}_k$ | Fixed SCA for the MSE |
| $\phi_k$ | Parametric constant |
| $\gamma_k$ | SINR experienced by user $k$ |
| $\delta^{b}_{b_k,k}$ | Represents the actual interference caused by BS $b$ to user $k$, which is served by BS $b_k$ |
| $\delta^{G}_{b_k,k}$ | Represents the global consensus of interference value at user $k$ |
| $\delta^{b_k}_{b_i,k}$ | Used to represent the interference caused by BS $b_i$ to user $k$, which is maintained in BS $b_k$ |
| $\lambda^{b_k}_{b_i,k}$ | Dual variable for the interference caused by BS $b_i$ to user $k$, which is maintained in BS $b_k$ |
| $\rho$ | Step size used in subgradient update |
| $\sigma$ | Standard deviation of Gaussian noise |

## Mathematical Operators & Symbols

| | |
|---|---|
| $\Re(.)$ | Real part of a function |
| $\Im(.)$ | Imaginary part of a function |
| $\min(x, y)$ | Minimum between x and y |
| $\max(x, y)$ | Maximum between x and y |
| $\|.\|$ | Absolute value of a complex number |
| $\|\|.\|\|_2$ | $l_2$ norm |
| $(\cdot)^{-1}$ | Inverse of a Matrix |
| $(\cdot)^{T}$ | Transpose of a Matrix |
| $(\cdot)^{H}$ | Hermitian of a Matrix |
| $\mathbf{E}_x\{.\}$ | Expectation of variable over $x$ |

# 1. INTRODUCTION

Traditionally, wireline communications are used to provide secured connectivity between any two interconnected terminals. Due to the lack of interference in the wireline transmissions, the signal-to-noise-ratio (SNR) is used as a performance measure while accessing the channel for transmissions. In order to provide multiple users to access wireline systems, either time-division multiplex access (TDMA) or frequency-division multiple access (FDMA) is employed among the contending users [1]. However, in spite of all the above mentioned merits, the connectivity is often limited by the interconnectedness of the networking entities. Moreover, the cost of laying the cables to facilitate wireline communications require huge capital investment, which predominantly limit the underlying benefits of wireline systems.

On the contrary, optical fibers or high capacity cables are used to interconnect only the base stations (BSs) in a wireless communication systems, thereby avoiding huge capital investment to create a ecosystem of devices and BSs. Currently, wireless communication is gaining more importance due to its seamless and ubiquitous connectivity. In addition, it also provides various other advantages such as improved data rate though spatial multiplexing and range extensions by utilizing channel diversity through multi-antenna transmission [2]. In our day-to-day life, the dependency on wireless services has increased with the advent of smart phones due to on-demand availability and easy accessibility of the desired contents.

However, due to the broadcast nature of wireless transmissions, inter-cell interference cannot be ignored while designing the transmission. Even though it can be minimized by the use of frequency reuse factors, it often reduces the achievable throughput due to the limited utilization of the available spectrum [3]. With the advent of multiple antenna transmission, *i.e.*, by using multiple-input multiple-output (MIMO) technique, both spatial multiplexing and range extension through diversity combining can be performed to improve the overall throughput of the network [4]. Multiplexing multiple user data streams over spatial dimension improves the achievable rate tremendously without increasing the available spectrum or power, which is termed as multi-user (MU)-MIMO transmission.

In order to facilitate MU-MIMO transmission by multiplexing different user data streams, precoders are to be designed efficiently to minimize the inter-user interference in addition to the inter-cell interference that exists in the wireless transmissions. Therefore, with the advent of every capacity improving schemes for a MIMO systems, the overall system complexity and the overhead involved in obtaining the relevant information, *i.e.*, the channel state information (CSI) knowledge, increases significantly [4]. In addition, the availability of wireless spectrum is limited and different for each country, therefore, mobile devices are obligated to support multiple frequency bands in order to facilitate the roaming of user, which is one of the core feature of wireless service.

In wireless model, multiple user data streams are multiplexed over both time and frequency by using both time-division multiplexing (TDM) and orthogonal frequency division multiplexing (OFDM) (superior over frequency-division multiplexing (FDM) technique due to zero guard band transmission). Moreover, in MU-MIMO technique, the available users are also multiplexed across the spatial dimension by using transmit precoders. By properly designing the transmit precoders at the BS, receiver complexity can be greatly reduced. However, to design transmit precoders, the knowledge of CSI corresponding to each user is required at the BS [4]. It is often obtained by transmitting orthogonal pilots in both uplink and downlink to measure CSI with respect to each user. Upon obtaining the CSI between each user and the respective serving BS, transmit precoders are designed with the objective of maximizing or minimizing certain utility function. Since the design problem involves additional system limitations as constraints, it is often formulated as an optimization problem, which is solved by using existing solvers or can be solve iteratively by solving a subproblem of original problem in each iteration. In order to design precoders for MU-MIMO scenario, the interference model can be considered as either interference broadcast channel (IBC) or interference channel (IC). Even though both IBC and IC are used to model wireless systems [5], however, in practice, IBC is often used to model cellular transmission scenarios.

In the MIMO IBC with multiple BSs, each transmitting data streams to users in the respective cell by interfering the transmissions on neighboring cells. Some practical examples that can be modeled as MIMO IBC are Cognitive Radio systems, ad-hoc wireless networks, wireless cellular communication, and etc. For a MIMO IBC scenario, dirty paper coding (DPC) is known to be the capacity-achieving scheme [6, 7], however, it requires the knowledge of interference, and the receiver complexity is significantly higher due to the requirement of interference cancellation based detectors. Therefore, to reduce the design complexity, we rely on linear precoding techniques only at both transmitters and receivers.

In this thesis, we consider weighted sum rate maximization (WSRM) as the objective while designing the transmit precoders for the spatially multiplexed users in MU-MIMO technique. The precoder design with the WSRM objective is studied in the literature by considering various extensions. Since we know that the WSRM problem is nonconvex and NP-hard even for single-antenna receivers [8], there exists a class of beamformer designs which are based on achieving the necessary optimal conditions of the WSRM problem, as can be seen in, [9, 10, 11, 12]. In [13], transmit precoders are designed by using branch and bound technique to solve WSRM problem via feasibility subproblems for a given signal-to-interference-plus-noise-ratio (SINR). It is also numerically shown that the suboptimal designs that achieve the necessary optimal conditions of the WSRM problem perform very close to the optimal design.

In this thesis, we analyze the WSRM objective for a MU-MIMO transmission. The beamformers (or precoders) are designed to maximize the sum rate of all users with minimal or no interference among the multiplexed users. Note that the existing WSRM formulation presented in [14] utilizes the inequality between arithmetic mean and geometric mean to reduce the nonconvex constraint into a series of convex constraints. However, as the SINR of a user approaches zero, the problem becomes unstable and the algorithm may not converge at all. In order to address this design issue, we propose alternative formulations to overcome the same.

Moreover, the main contribution of this thesis is on the design of transmit precoders with the WSRM objective in addition to the user specific quality of service (QoS) requirements in terms of guaranteed minimum rate [15, 16, 17]. We first propose the centralized precoder design for the aforementioned problem, which is followed by the distributed approaches for a practical implementation. In both design problems, we consider two different approaches to solve the precoder design problem, namely, by directly relaxing the SINR constraint, and by utilizing mean squared error (MSE) equivalence with the SINR expression upon using minimum mean squared error (MMSE) receivers. The effectiveness of the proposed algorithm is evaluated in the numerical experiments and is discussed in the simulation results section.

In order to improve the convergence speed of distributed precoder design, we use the corresponding Karush-Kuhn-Tucker (KKT) expression of the centralized problem by associating the coupling variables across the respective BSs. The KKT method of the distributed precoder design is discussed for both direct SINR relaxation and also for the MSE reformulation approach. For comparison purposes, we also provide alternating directions method of multipliers (ADMM) based distributed precoder design. We compare the performance of various algorithms by using numerical simulations.

Even though there exists several methods to obtain optimal beamformers [13, 18, 19], they may not be practically feasible, since the complexity of finding optimal designs grows exponentially with the problem size. Hence, the need of computationally conducive suboptimal solutions to the WSRM problem still remains, as discussed in [14]. In [9], the iterative coordinated beamforming algorithm was proposed by manipulating the KKT equations. However, this method is not provably convergent. On the contrary, [11, 12] solved the WSRM problem by utilizing the relation between SINR and MSE expression upon using MMSE receivers. The resulting problem of joint transceiver design is then solved using alternating optimization (AO) approach between transmit and receive beamforming. Similarly, in [20], the WSRM problem is solved by employing successive convex approximation (SCA) method for the MSE reformulated problem. It has been shown that the algorithm proposed in [20] has a better initial convergence than the methods proposed in [11, 12] in spite of reaching the same objective upon convergence. As we show by numerical results, these methods have a slower convergence rate compared to our proposed design, since the MSE based methods involves receiver updates even for single receive antenna.

The thesis is outlined as follows. The central object of interest is the transmit precoder design. Chapter 2 reviews background and literature. It also covers the introductory details on the MIMO model and its capacity. Starting from MIMO basics we discuss various models like the MIMO-IC and MIMO-IBC, in brief we also discuss the MIMO channel capacity. In addition, Chapter 2 also introduces the mathematical preliminaries of an optimization problem that will be used extensively in the remaining chapters. This section provides insight about convex functions and sets, by explaining a convex optimization problem. The goal is to have a small set of example model in our report to evaluate the performance in our further chapters.

Chapter 3 introduces the proposed two centralized MU-MIMO precoder design formulations with additional user specific quality of service (QoS) requirements. As, a baseline we look into the design of transmit precoders with WSRM objective in a MU-MIMO scenario, for which the system model and problem formulation is discussed. We propose two different centralized formulations (i) Direct SINR Relaxation for AP-GP Approach and (ii) Reformulation via MSE. Based on these methods we propose a distributed solution in Chapter 4. We also propose the ADMM based decentralization scheme for the comparison purposes. In Chapter 5, we compare the performance of various proposed algorithms using numerical simulations. Finally, conclusions are drawn and summary is given in Chapter 6.

# 2. BACKGROUND

## 2.1. MIMO Communications

The MIMO technology is a breakthrough in wireless communication that utilizes the available spatial dimension provided by $N_{\mathrm{T}}$ transmit and $N_{\mathrm{R}}$ receive antennas. However, due to the contention for accessing the wireless resources by a multitude of devices, efficient utilization and scheduling of the available resources to the contending devices or users has become a challenging task. Therefore, to utilize the available wireless resources such as spatial, temporal and frequency dimensions efficiently, knowledge of each user CSI is required to be known at the transmitter prior to any transmission or resource allocation. Moreover, due to the increase in number of devices accessing the available resources, complexity of both transmitter and receiver side algorithms has become significantly complex such that the processing power requirement can be compared to that of a personal computer.

A typical MIMO system is presented in Figure 1, which consists of various system blocks that performs a specific task in the order as shown [2]. The underlying assumption on the input bits is that the source coding is already performed on the incoming bits to remove any redundancy to make the input source distribution Gaussian. Then, the incoming data of each user is modified by channel coding to increase the redundancy and then interleaved to avoid the ill-effects of burst errors in the channel. Note that the channel coding performs the opposite of source coding technique, *i.e.*, by increasing the redundancy. It is carried out to ensure that the transmitted symbols can be decoded correctly at the receiver, thereby reducing the bit error probability.

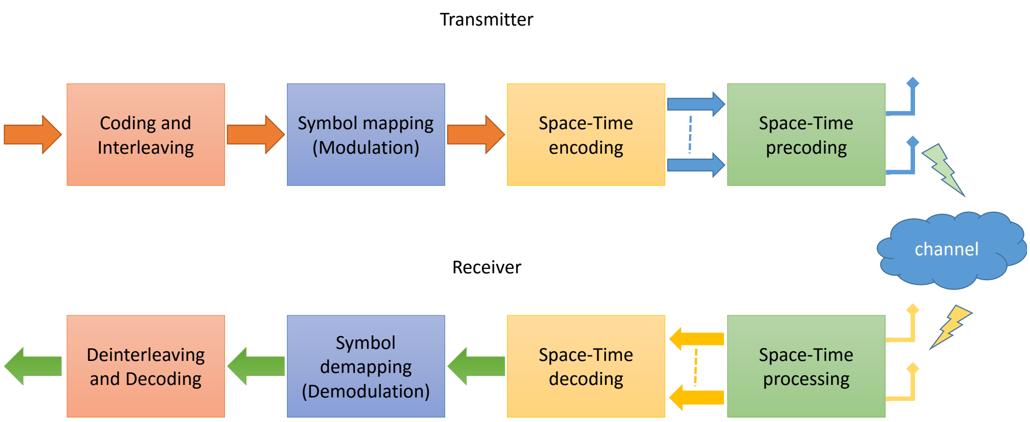


Figure 1: MIMO system model.

The channel entries corresponding to each independent transmit-receive pair is denoted by a complex entry $h_{ij} \in \mathbb{C}$, where $i$ corresponds to the receive element and $j$ denotes the transmit element. Using this notation, MIMO channel matrix $\mathbf{H} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$ is given as

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1M} \\ h_{21} & h_{22} & \dots & h_{2M} \\ \dots & \dots & \dots & \dots \\ h_{N1} & h_{N2} & \dots & h_{NM} \end{pmatrix}. \tag{1}$$

Let $\mathbf{x} \in \mathbb{C}^{N_T \times 1}$ be the transmitted symbol that can possibly includes the data corresponding to

- single stream for providing the diversity and beamforming benefits (transmit diversity mode), or
- multiple data streams corresponding to single user for increasing the total throughput, *i.e.*, single-user MIMO, or
- multiple data streams, where each stream is intended for different user MU-MIMO transmission mode.

In all the above mode of transmissions, the knowledge of CSI at the transmitter is assumed. Even if the CSI is not available at the transmitter, we can still achieve the above mentioned benefits but with noticeably worse performance than the one with CSI at the transmitter. Note that $x_i$, $\forall i \in \{1, 2, \dots, N_T\}$ corresponds to the transmitted signal from each antenna element $i$. Using the above notations, the received signal $\mathbf{y} \in \mathbb{C}^{N_R \times 1}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{n} \sim \mathcal{CN}(0, N_0)$ denotes the complex Gaussian noise with zero-mean and variance $N_0$.

In order to characterize the benefit of using multiple antenna elements, let us now consider the capacity improvements provided by the MIMO system under the assumption that CSI is known at the transmitter. Note that the capacity $C$ is defined as the maximum data rate at which the reliable communication is possible. Therefore, for an additive white Gaussian noise (AWGN) channel, the capacity is given by

$$C_{\text{AWGN}} = \log(1 + \gamma) \text{ bps/Hz}, \tag{3}$$

where $\gamma$ is the link SNR, and (3) measures the maximum achievable spectral efficiency through the AWGN channel as a function of the SNR.

## 2.2. MIMO Configuration

The main issue in the study of MIMO transmission schemes is how to mitigate multi-user interference. We know that the interference is a major set back and a limiting factor in the wireless communication networks. In practice, there are several commonly used methods for dealing with interference arising due to inter and intra cell transmissions. The problem of interference is in general dealt with planning of radio resource management (RRM). Initially, we can treat the interference as a noise and just focus on extracting the desired signals as discussed in [21, 22] or we can design the transmit covariance matrix and receive equalizers to consider the interference present in the network.

The BS consisting of multiple transmit antennas can serves more users simultaneously by utilizing the spatial degrees of freedom, which is the underlying concept of space-division multiple access (SDMA) or MU-MIMO transmissions. However, MU-MIMO systems impose precise requirements for CSI at the transmitter, which is often difficult to acquire in practice than the knowledge of CSI at receivers. Therefore, we consider only the downlink (DL) systems due to the challenges involved in the design of broadcast precoders that can multiplex different users data streams spatially. In cellular systems, one can distinguish between the in-cell users, where the SINR is mainly limited by the intra-cell transmissions, and the cell-edge users, where in addition inter-cell interference should also be considered while performing resource allocation to maximize the network throughput or to provide fairness among the users.

Depending on the type of scenario, we can characterize the MIMO system models as MIMO IBC or MIMO interference channel (IC) model. The MIMO IBC consists of both in-cell and cell-edge users in the consideration. The spatial streams provided by the MIMO channel can be used for either single-user MIMO or MU-MIMO transmission. On the contrary, the MIMO IC scenario considers only the cell-edge users, thereby allocates the spatial streams by considering both intra- and inter-cell interference from the neighboring BSs.

### *2.2.1. MIMO-IC*

A systematic study of the performance of cellular communication systems where each cell communicates multiple streams to its users while causing interference from and to the neighboring cells due to transmission over a common shared resource known as, MIMO-IC. A $K$-user MIMO-IC model consists of a network of $K$ transmit-receive pairs where each transmitter communicates multiple data streams to its respective re-

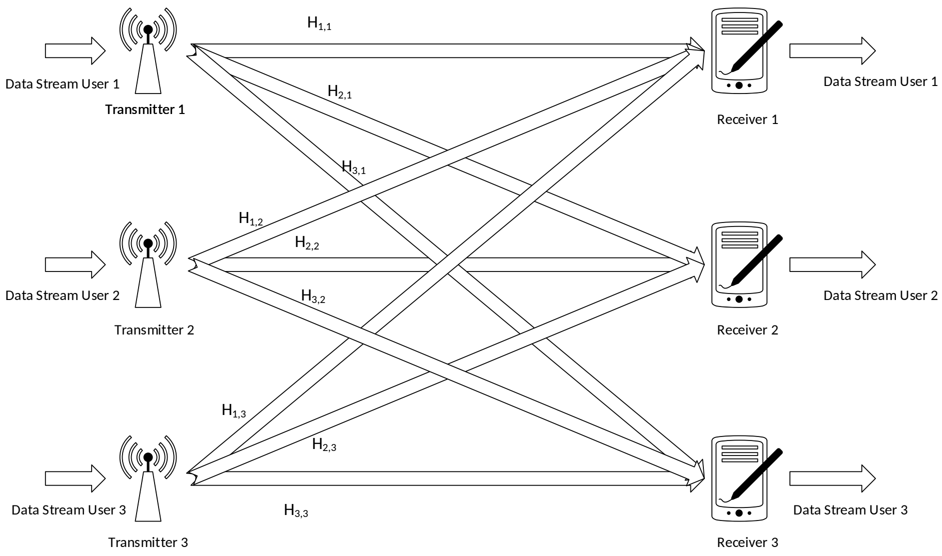


Figure 2: MIMO-IC model.

ceiver. In doing so, it generates interference at all other receivers present in the system. This MIMO model is mentioned in [23] and [24]. In [5], precoder design for such a system has discussed based on interference alignment (IA) concept.

#### *2.2.2. MIMO-IBC*

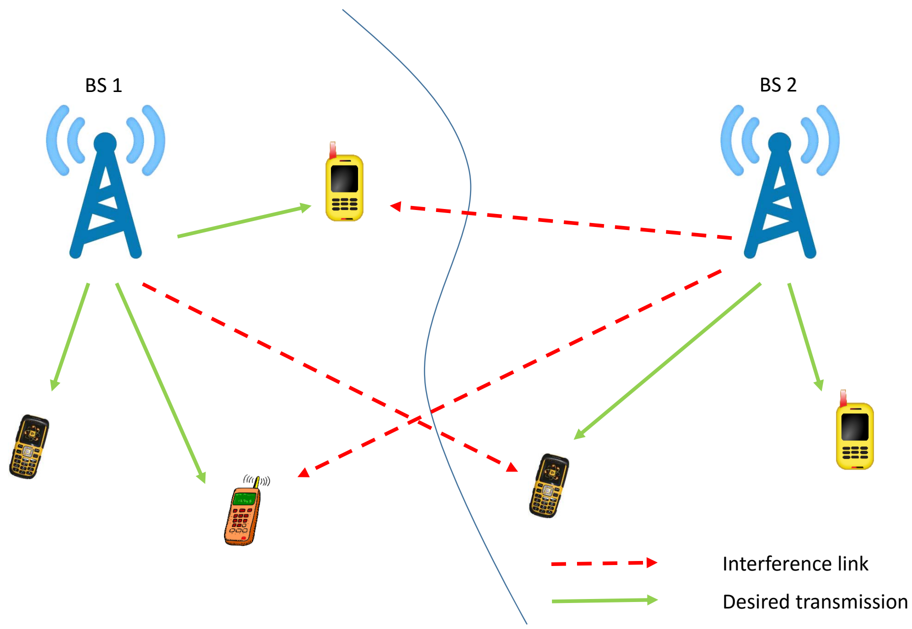


Figure 3: MIMO-IBC model.

A linear transceiver design problem is considered in MIMO-IBC whereby multiple BSs in a cellular network simultaneously transmit signals to a group of users in their own cells while causing interference to the users in other cells. Both the BSs and the users are equipped with multiple antennas, and they share the same time/frequency resource for transmission. This model is used in [12] to design transmit and receive precoders by using the equivalence between MSE and SINR expression while using MMSE receivers.

## 2.3. MIMO Precoding Design with CSI at Transmitter

In this thesis report optimal precoder design for WSRM in MIMO interference networks is studied. For this well known non-convex optimization problem, convex approximations based on interference alignment are developed, for multi-beam cases. Considering that each user treats interference from other users as noise. It is well known that, due to interference coupling, the problem is a non-convex optimization and is hard to solve. In the high SNR regime, there has been recent progress on maximizing the sum degrees of freedom, exploiting the idea of interference alignment. It has been shown that maximizing the sum degrees of freedom is still an NP hard problem [25].

## 2.4. Mathematical Preliminaries - Convex Optimization

### *2.4.1. Convexity*

Convex analysis is the study of mathematics dealing about convex sets and functions [26]. Convex analysis is considered to be the core for optimization. This plays a major role in study of statistics, mathematical economics, and also has several applications in the field of wireless communication such as MIMO precoder designs, user scheduling algorithms, wireless resource allocation problems, energy efficiency designs, and sparse solutions etc.

We note that set $\mathcal{K} \subset \mathbb{R}^n$ is said to be convex, if any line segment through the points $x, y$ belongs to $\mathcal{K}$ [26, 27]. If a set is defined by the intersection of several convex sets, then the resulting set is convex, whereas the union of two or more convex sets is not necessarily convex. Furthermore, when a set is not convex then it is called as nonconvex set, *i.e.*, every points in a line segment joining $x, y$ need not be in set $\mathcal{K}$. A set is said to be affine, iff any two points in $\mathcal{K}$ lies in $\mathcal{K}$. Every affine set is also convex, since it contains the entire line between the two distinct points in it [26]. Few examples of convex sets are triangle, rectangle, polyhedron and quadratic functions such as $f(x) = ax^2 + bx + c$ is convex if and only if $a \geq 0$.

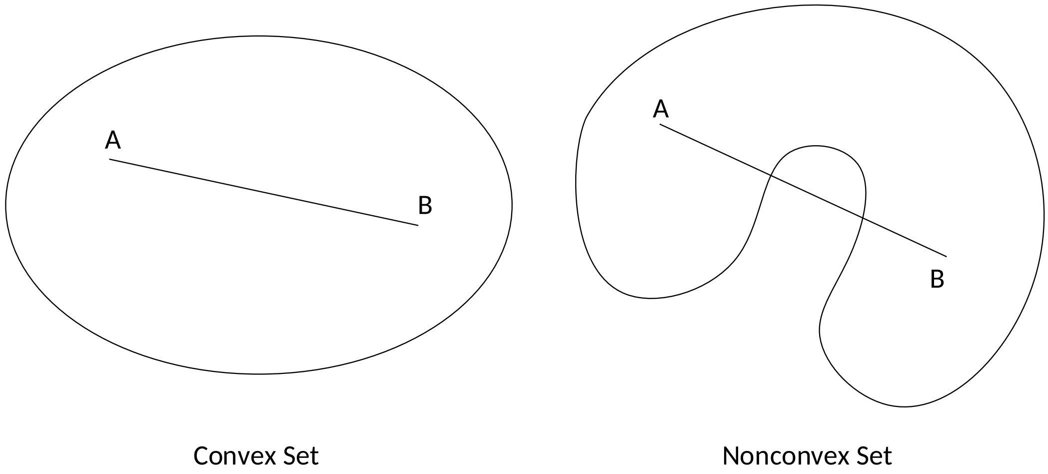


Figure 4: Representation of Convex and Nonconvex Sets.

A function is convex if and only if the region above the graph as shown in figure is convex set. As mentioned in [26], a function $f$ is convex if $\forall x, y \in \mathcal{K}$, $\forall \theta \in [0, 1]$:

$$f(\theta x + (1 - \theta)y) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}). \quad (4)$$

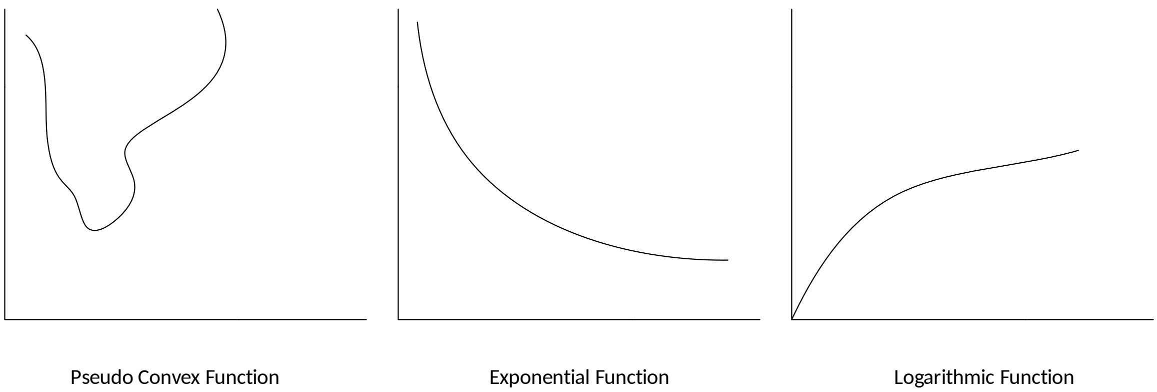


Figure 5: Representation of Convex Functions.

Depending on the type of convexity as discussed in [26, 27, 28], we can classify convex functions further as follows.

- Function $f$ is said to be strictly convex if strict inequality holds in (4) when ever $\forall x \neq y \in \mathcal{K}$, $\forall \theta \in (0, 1)$. A function $f$ is said to be concave if function

$-f$ is convex and strictly concave when function $-f$ is strictly convex. Strict convexity means that the graph of $f$ lies below the segment $\mathcal{S}$. Certain examples of strict convex functions are exponential and quadratic function.

- For affine functions, there is equality in (4), so an affine function is said to be both convex and concave.

- A function is said to be strongly convex, whenever $\forall x, y \in \mathcal{K}$ and $\theta \in (0,1)$ there exists a constant $c > 0$, so that,

$$f(\theta x + (1-\theta)y) \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}) - \frac{c}{2}\theta\,(1-\theta)\,\|x-y\|^2. \quad (5)$$

The relationship between strict strong and convex function as in [27] can be outlined as: a strongly convex function is strictly convex which is convex, but the reverse is not possible. For instance, a linear function is a convex function that is not strictly convex, an exponential function is strictly convex but not strongly convex, and a quadratic function is an example of strong convex function.

### *2.4.2. Optimization Problem Formulation*

A generic optimization problem is similar to linear programming problem, that can be solved quickly depending on the variables and the constraints. A standard optimization problem can be written as

$$\underset{x}{\text{minimize}} \quad f(\mathbf{x}) \quad (6a)$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{K}, \quad (6b)$$

where $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is the objective function that has to be minimized with respect to the constraint $\mathbf{x}$ and $\mathcal{X} \subset \mathbb{R}^n$ is the feasible set. A feasible point $\mathbf{x}^* \in \mathcal{K}$ is said to be optimal if $f(\mathbf{x}^*) \leq f(\mathbf{x}) \, \forall \, \mathbf{x} \in \mathcal{K}$ *i.e.*, $\mathbf{x}^*$ has the least value of $f$ amongst all vectors that satisfies the constraint. It is always assumed that $\mathcal{K}$ is closed and convex and the function $f$ is differentiable on $\mathcal{K}$ [27]. Similarly, a maximization problem can be written by negating the objective function.

Generic optimization problems are used in the data fitting problem, device sizing in electronic circuits, and portfolio optimization etc. However, it is considered that general optimization problems are time consuming, complex in finding solutions and at times it is seen to not provide the actual solutions. Nevertheless, there are certain class of problems such as least square, linear programming and general convex optimization problems, which are discussed in [26], can be solved efficiently in polynomial time.

A convex optimization problem can be defined so that all of its constraints are convex functions, and the objective is a convex function as well. The problem can be minimizing a convex function, or maximizing a concave function. In general, linear functions are convex so the linear programming problem is a convex problem. A general convex optimization problem can be written as

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(\mathbf{x}) && \text{(7a)} \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, ...m && \text{(7b)} \\ & h_j(\mathbf{x}) = 0, j = 1, ...p, && \text{(7c)} \end{aligned}$$

where, $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function or the cost function and $\mathcal{X} \subset \mathbb{R}^n$ is the feasible set and is called convex when $\mathcal{X}$ is closed convex set and $f(\mathbf{x})$ is convex on $\mathbb{R}^n$. (7b) is an inequality constraint and the corresponding function $g_i(\mathbf{x})$ is the inequality constraint function, and (7c) is the equality constraint. Function $h_j(\mathbf{x})$ is the equality constraint function in the optimization problem.

To find a solution for an unconstrained objective, we differentiate the objective function with respect to the optimization variable x and equate it to zero as $\nabla f(\mathbf{x}) = 0$. However, for an constrained problem as in (7), we solve the Lagrangian of the problem (7) as

$$\underset{\boldsymbol{\lambda},\boldsymbol{\mu}}{\text{maximize}} \ \underset{\mathbf{x}}{\text{minimize}} \quad L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}), \qquad (8)$$

where $\lambda_i \geq 0$ and $\mu_j$ are Lagrange multipliers. The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ denotes the stacked entries of dual variables $\lambda_i$ and $\mu_j$, respectively. Now, the Lagrangian in (8) for (7) is solved by using KKT conditions as

- $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}^T$,
- $\nabla_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}^T$,
- $\nabla_{\boldsymbol{\mu}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}^T$,
- $\boldsymbol{\lambda} \geq 0$,
- and complementary slackness conditions $\lambda_1 g_1 = 0, \lambda_2 g_2 = 0, \ldots, \lambda_m g_m = 0$ and $\mu_1 h_1 = 0, \mu_2 h_2 = 0, \ldots, \mu_p h_p = 0$.

Using the above system of the KKT expressions, (7) can be solve for an optimal solution if the problem is convex. However, if the considered formulation is not convex, then we can only solve for a stationary point. More details on solving nonconvex problems will be discussed in the forthcoming sections.

# 3. CENTRALIZED FRAMEWORK FOR PRECODER DESIGNS

## 3.1. Introduction to Precoder design

Precoding can be explained as a transmitter side processing that can provide multi-stream transmission in a MIMO communication model. Multiple data streams are sent from the transmit antenna elements with appropriate weights to maximize certain network utility by improving the receiver side SINR. The precoder design can aim at maximizing certain utility, for example, the sum rate of a multi-user system by performing spatial multiplexing or by extending the coverage and the received signal quality. Closed loop precoding techniques are used for both point-to-point single-user and also for multi-user MIMO scenarios in the current wireless standards.

In a point-to-point MIMO system, the transmitter is equipped with multiple antenna which transmit spatially multiplexed data to a receiver equipped with multiple antennas. Since both single-user and multi-user MIMO technique require huge processing complexity, performing multi-user detection over an inter-symbol interference (ISI) channel demands exponential complexity. However, upon using OFDM based transmission, spatially multiplexed MIMO techniques are possible in real-time by the virtue of narrow-band channels provided by the OFDM transmission [29], which increases the symbol duration to obtain narrow-band sub-carriers.

If the transmitter knows only the statistical information about CSI and the receivers are aware of the respective channel matrices, then the optimal precoding technique is to broadcast data symbols by considering the covariance of channel matrices. However, by knowing the complete CSI at the transmitter, eigen-beamforming based on singular value decomposition (SVD) achieves the full MIMO capacity for a given configuration. Note that as mentioned in [4] the transmitter emits multiple streams over all Eigen directions corresponding to the channel matrix and the power allocation across each stream is based on water-filling solution.

In MU-MIMO system, a multi-antenna transmitter communicates simultaneously with multiple receivers with one or more antennas known as SDMA. Precoding algorithms for the SDMA systems can be sub-divided into linear and nonlinear precoding types. The capacity achieving algorithms are nonlinear but linear precoding approaches usually achieves reasonable performance with much lower complexity. Linear precoding strategies include maximum ratio transmission (MRT), zero-forcing (ZF) precoding, and transmit Wiener precoding [4]. In addition, there are also precoding strategies for low-rate feedback of channel state information, which is usually the channel covariance feedback [30, 31].

Nonlinear precoding is designed based on the concept of DPC [6], which shows that any known interference at the transmitter can be subtracted without wasting any of the available transmit power. The transmitter only needs to know the interference to cancel it from the users in the network. However, to perform DPC, the transmitter requires the CSI knowledge of all served users. Due to nonlinear processing, it is often difficult to implement in practice. Thus, linear precoding techniques are often considered.

## 3.2. System Model and Problem Formulation

### *3.2.1. System Model*

Let us consider a downlink MIMO IBC system consisting of $N_{\mathrm{B}}$ coordinating BSs with $N_{\mathrm{T}}$ transmit antennas each and $K$ single antenna receivers. By coordination, we mean that all BSs design the transmit precoders to minimize the inter-cell interference without sharing the data symbols among them. The set of all $K$ user indices is denoted by $\mathcal{U} = \{1, 2, \ldots, K\}$. We assume that data for the $k^{th}$ user is transmitted from one BS, which is denoted by $b_k \in \mathcal{B}$, where $\mathcal{B} \triangleq \{1, 2, \ldots, N_B\}$ is the set of all coordinating BS indices. The set of all users served by BS $b$ is denoted by $\mathcal{U}_b$. Assuming flat fading channel conditions, the input-output relation for the $k^{th}$ user channel is given as

$$y_k = \mathbf{h}_{b_k,k}\mathbf{x}_k + \sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{h}_{b_i,k}\mathbf{x}_i + n_k \tag{9}$$

where $\mathbf{h}_{b_i,k} \in \mathbb{C}^{1 \times N_T}$ is the channel coefficient between BS $b_i$ and user $k$. Note that $n \sim \mathcal{CN}(0, \sigma^2)$ is zero-mean circularly symmetric complex Gaussian noise with variance $\sigma^2$, and $\mathbf{x}_k \in \mathbb{C}^{N_T \times 1}$ is the transmit symbol corresponding to user $k$. Without loss of generality, we assume that all receivers know the corresponding CSI between serving BS, *i.e.*, $\mathbf{h}_{b_k,k}$, to decode the transmitted symbols associated with each user $k$.

Under the assumption that linear precoding is used for spatial multiplexing, transmitted symbol from BS $b$ is given by

$$\sum_{k \in \mathcal{U}_b} \mathbf{x}_k = \sum_{k \in \mathcal{U}_b} \mathbf{w}_k d_k \tag{10}$$

where $d_k$ is the normalized data symbol, and $\mathbf{w}_k \in \mathbb{C}^{N_T \times 1}$ is the linear precoding vector. Now, by using (10) in the expression (9), the received symbol is written as

$$y_k = \mathbf{h}_{b_k,k}\mathbf{w}_k d_k + \sum_{i=1, i\neq k}^{K} \mathbf{h}_{b_i,k}\mathbf{w}_i d_i + n_k. \tag{11}$$

The term $\sum_{i=1,i\neq k}^{K} \mathbf{h}_{b_i,k}\mathbf{w}_i d_i$ in (11) includes both intra-cell and inter-cell interference components. The total transmit power of BS $b$ is given by the constraint $\sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|^2 \leq P_b$ with $P_b$ as the maximum available transmit power budget, and the SINR $\gamma_k$ corresponding to user $k$ is given as

$$\gamma_k = \frac{|\mathbf{h}_{b_k,k}\mathbf{w}_k|^2}{\sigma^2 + \sum_{i=1,i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2}. \tag{12}$$

### *3.2.2. Problem Formulation*

In order to formulate the problem of designing linear transmit precoders with WSRM objective, we consider including the constraint on total transmit power. By doing so, the WSRM problem can be formulated as

$$\underset{\gamma_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{13a}$$

$$\text{subject to} \quad \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|^2 \leq P_b, \forall\, b \in \mathcal{B} \tag{13b}$$

$$\frac{|\mathbf{h}_{b_k,k}\mathbf{w}_k|^2}{\sigma^2 + \sum_{i=1,i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2} \geq \gamma_k,\ \forall k \tag{13c}$$

where $\alpha_k$ is a positive weighting factor for user $k$ which are typically introduced to maintain a certain degree of fairness among the users. Then, note that the constraint (13c) is an over-estimator for the SINR term $\gamma_k$, since the expression in (12) cannot be used directly in the optimization framework. However, note that at optimal solution, the relaxed SINR expression in (13c) will be tight.

The precoder design for the MIMO-IBC scenario is difficult due to the non convex nature of the problem formulation [8]. In general, the rate maximizing beamformer designs has an inherent complexity due to existence of optimization variables, *i.e.*, transmit precoders, in both the numerator and in the denominator of the SINR expression. In addition, the beamformer design can be classified into centralized and distributed approaches depending on the type of processing, *i.e.*, whether the design is performed by a centralized entity or by each BS independently through some coupling information exchange. In the centralized approach, the common controller is assumed to have the complete CSI of all BS-user links in order to design precoders for all BSs. In the distributed approach, practical difficulties of distributing CSI over the backhaul network and high complexity of joint precoding design motivates the analysis. The beamforming and power allocation strategies can be computed locally using only the local CSI in a distributed design. In particular, for a single receive antenna scenario, the goal of transmit precoding is to maximize the received signal power at the intended terminal while minimizing the interference caused to the others.

The core prior work on the centralized design can be found in [9, 10, 11, 12] and references therein solve the problem of precoder design by the centralized approaches. Moreover, [11, 12, 20] addressed the WSRM problem as a MSE minimization problem by using the relation between MSE and SINR while using MMSE receivers at the user terminals. At first, we propose a precoder design based on [32], which utilizes the relation between arithmetic and geometric mean. This method has been utilized in [14] to design transmit precoders in an iterative manner. In the following discussion, we present two different approaches of designing transmit precoders based on the above approximations.

### 3.3. Direct SINR Relaxation via SPCA

At first, we discuss centralized transmit precoder design based on sequential parametric convex approximation (SPCA) algorithm proposed in [32] and further extended to wireless systems in [14]. The centralized coordinated DL transmission requires CSI to be fed back from the users to their respective serving BS, and aggregated at the central coordination node to form the channel matrix for precoding, so that interference can be mitigated. Before discussing the solutions, let us look at the existing WSRM algorithm for centralized precoder design with constraints required to formulate it as an optimization problem.

Let us consider the problem in (13), where we relaxed the SINR expression in (12) by introducing inequality constraints as

$$\underset{w_k,\gamma_k,\beta_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{14a}$$

$$\text{subject to} \quad \frac{|\mathbf{h}_{b_k,k}\mathbf{w}_k|^2}{\beta_k} \geq \gamma_k, \forall k \in \mathcal{U} \tag{14b}$$

$$\beta_k \geq \sigma^2 + \sum_{i=1, i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2, \forall k \in \mathcal{U}, \tag{14c}$$

$$\sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} \tag{14d}$$

where the SINR expression in (12) is relaxed by using the inequalities in (14b) and (14c). We can see that (14b) is an under-estimator of SINR and (14c) provides an upper-bound on the total interference seen by all the users $k \in \mathcal{U}_b$, denoted as $\beta_k$. Thus, we can replace problem (13) by an equivalent and tractable formulation in (14) to solve the WSRM problem.

In order to find an optimal solution for problem (14), we can observe that (14d) and (14c) are convex constraints with the involved variables. Moreover, we note that (14b) is the only nonconvex constraint in (14). In order to solve the nonconvex problem (14), we find a convex subset for the nonconvex constraint (14b). To do so, we consider the following equivalent representation for constraint (14b) as

$$\Re(\mathbf{h}_{b_k,k}\mathbf{w}_k) \;\geq\; \sqrt{\gamma_k\beta_k} \tag{15a}$$

$$\Im(\mathbf{h}_{b_k,k}\mathbf{w}_k) \;=\; 0, \forall k \in \mathcal{U} \tag{15b}$$

where (15b) is used to restrict the transmit phase of $\mathbf{w}_k$ without affecting the objective. Moreover, making the imaginary part to zero does not affect the optimality of (14), since phase rotation on $\mathbf{w}_k$ will result in the same objective while satisfying all constraints. Secondly, we can also show that all the constraints in (14) hold with equality at optimum. It follows from the fact that to maximize sum rate, $\gamma_k$ has to be maximized, *i.e.*, the interference limit term $\beta_k$ has to decrease. In order to reduce $\beta_k$, (14c) must be tight, thereby making the above relaxation to hold with equality at optimum. Using the above equivalent representations, we can reformulate (14) to find the transmit beamformers $\mathbf{w}_k$ as

$$\underset{w_k,\gamma_k,\beta_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{16a}$$

$$\text{subject to} \quad \Re(\mathbf{h}_{b_k,k}\mathbf{w}_k) \geq \sqrt{\gamma_k\beta_k}, \forall k \in \mathcal{U} \tag{16b}$$

$$\Im(\mathbf{h}_{b_k,k}\mathbf{w}_k) = 0, \forall k \in \mathcal{U}, \tag{16c}$$

$$\beta_k \geq \sigma^2 + \sum_{i=1, i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_k|^2, \forall k \in \mathcal{U}, \tag{16d}$$

$$\sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B}. \tag{16e}$$

Since the r.h.s of (16b) is a geometric mean, we can bound it by a suitable convex upper approximation, *i.e.*, the arithmetic mean, as

$$\sqrt{\gamma_k \beta_k} \;\leq\; \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}} \;\triangleq\; f(\gamma_k, \beta_k, \phi_k^{(i)}) \tag{17}$$

where $\phi_k^{(i)}$ is a parametric constant, which is given as

$$\phi_k^{(i)} = \sqrt{\frac{\beta_k^{(i-1)}}{\gamma_k^{(i-1)}}}. \tag{18}$$

Note that $\beta_k^{(i)}$ and $\gamma_k^{(i)}$ are the solution obtained by solving (16) with the approximation (17) in $i$th iteration for $\beta_k$ and $\gamma_k$, respectively. Using (17) and (18), we can easily show that

$$\lim_{i\to\infty} f(\gamma_k, \beta_k, \phi_k^{(i)}) \to \sqrt{\gamma_k^* \beta_k^*} \tag{19}$$

where $\sqrt{\gamma_k^* \beta_k^*}$ is the optimal value upon convergence of the SPCA procedure [32]. Finally, the approximate SPCA based iterative precoder design problem with the objective of WSRM is given as

$$\underset{w_k,\gamma_k,\beta_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{20a}$$

$$\text{subject to} \quad \Re(\mathbf{h}_{b_k,k}\mathbf{w}_k) \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \forall k \in \mathcal{U} \tag{20b}$$

$$\Im(\mathbf{h}_{b_k,k}\mathbf{w}_k) = 0, \forall k \in \mathcal{U}, \tag{20c}$$

$$\beta_k \geq \sigma^2 + \sum_{i=1, i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2, \forall k \in \mathcal{U}, \tag{20d}$$

$$\sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B}. \tag{20e}$$

The above iterative problem is solved until convergence, *i.e.*, $i \to \infty$. Upon the convergence of above algorithm, the approximation in (17) will be tight, and the KKT condition of (20) is equivalent to (16) as shown in Appendix 9.1.

### 3.4. Reformulation via MSE

As an alternative method of solving the WSRM problem subject to convex transmit power constraint, we exploit the relationship between the MSE and the achievable

SINR while using the MMSE receivers at the user terminals [11, 12]. Let us denote MSE as $\epsilon_k$, which is defined for data symbol $d_k$ as

$$\epsilon_k = \mathbb{E}\left[(d'_k - d_k)^2\right] = |1 - u_k^* \mathbf{h}_{b_k,k}\mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_k^* \mathbf{h}_{b_k,i}\mathbf{w}_i|^2 + \bar{N}_0 \tag{21}$$

where $d'_k$ is the estimated data symbol of the corresponding transmit symbol $d_k$, $\bar{N}_0 = |u_k|^2\sigma^2$, and $u_k \in \mathbb{C}$ is the receive beamformer of user $k$. For a fixed receivers, (21) is a convex function in terms of transmit beamformers $\mathbf{w}_k \forall k$. Upon solving for transmit precoders, the receive beamformers $u_k \forall k$ can be solved directly by using the MMSE receiver, which is defined as

$$R_k = \sum_{i=1}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2 + \sigma^2, \tag{22}$$

$$u_k = R_k^{-1}\mathbf{h}_{b_k,k}\mathbf{w}_k, \tag{23}$$

where $R_k \in \mathbb{C}$ corresponds to both inter-cell and intra-cell interference term, since the users are equipped with single receive antenna.

Note that the optimal receive beamformers turn out to be the MMSE receivers, since the relation between the MSE and the received SINR is due to the assumption that the receivers are based on the MMSE criterion. The MMSE receiver in (23) can also be used without compromising the performance.

Now, by using (23) in the MSE expression (21), we obtain the following relation with the corresponding SINR as

$$\epsilon_k = (1 + \gamma_k)^{-1}. \tag{24}$$

Therefore, we utilize the above relation in (16) to reformulate the WSRM problem as

$$\underset{\gamma_k}{\text{maximize}} \sum_{i=1}^{K} \alpha_k \log(1 + \gamma_k) \;\Leftrightarrow\; \underset{\epsilon_k}{\text{minimize}} \sum_{i=1}^{K} \alpha_k \log(\epsilon_k). \tag{25}$$

Using the relation (25) in (13), we obtain the following reformulated problem

$$\underset{\epsilon_k}{\text{minimize}} \quad \sum_{k=1}^{K} \alpha_k \log(\epsilon_k) \tag{26a}$$

$$\text{subject to} \quad \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|^2 \le P_b, \forall\, b \in \mathcal{B} \tag{26b}$$

$$|1 - u_k^* \mathbf{h}_{b_k,k}\mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_k^* \mathbf{h}_{b_k,i}\mathbf{w}_i|^2 + \bar{N}_0 \le \epsilon_k,\ \forall k \tag{26c}$$

where (26c) is a relaxed over-estimator for the MSE expression in (21).

In spite of using the MSE formulation, the problem is still nonconvex. Therefore, (26) cannot be solved directly. Thus, we resort to SCA approach by relaxing the non-convex constraint by a sequence of convex approximations [33]. In order to find a suitable convex upper bound for the logarithmic objective function, we linearize the objective as mentioned in [20] with proper variable change and perform SCA for the difference of convex constraint as in (26c) around some fixed MSE point, say, $\bar{\epsilon}_k$, as

$$\log(\epsilon_k) \leq \left\{ \log \bar{\epsilon}_k + \frac{\epsilon - \bar{\epsilon}_k}{\bar{\epsilon}_k} \right\}. \tag{27}$$

where $\log(\bar{\epsilon}_k)$ is a constant and $(\bar{\epsilon}_k)^{-1}\epsilon_k$ is the first-order linear approximation $\log(\epsilon_k)$. The above inequality follows from the fact that concave functions are upper bounded by the first order Taylor approximation. Now, by using the above approximation, the iterative MSE minimization problem for the $i$th SCA step is given as

$$\underset{w_k, \epsilon_k, t_k}{\text{minimize}} \quad \sum_{i=1}^{K} \alpha_k \left( \bar{\epsilon}_k^{(i)} \right)^{-1} \epsilon_k \tag{28a}$$

$$\text{subject to} \quad \epsilon_k \geq |1 - u_k^* \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_i^* \mathbf{h}_{b_k,i} \mathbf{w}_i|^2 + \bar{N}_0 \tag{28b}$$

$$\sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B}. \tag{28c}$$

The above problem is solved iteratively by first fixing the receive beamformers $\mathbf{u}_k$ and optimized for transmit precoders $\mathbf{w}_k$. After each SCA iteration, the receiver beamformers are updated by using the expression in (23) for the fixed transmit precoders obtained at that step. The above procedure of alternating the optimization variables in each step is called as AO. Upon convergence of the above procedure, the optimal solution satisfies the KKT expression of the original nonconvex problem (13) as shown in Appendix 9.1. Alternatively, SCA steps can be iterated until convergence for fixed receiver update as well. However, by doing so, the total number of iterations required for the overall convergence is significantly large.

# 4. DISTRIBUTED PRECODER DESIGN VIA ADMM

So far, we have discussed the centralized precoder design solutions. In this chapter, we discuss the distributed precoder design techniques wherein the precoders are designed at each BS with the local CSI knowledge by exchanging some coupling variables among the coordinating BSs. In addition, we also consider certain guaranteed QoS requirement in the form of minimum rate as an additional constraint to each user in the system.

Note that to distribute the precoder design procedure, we can use either backhaul to exchange the coupling interference variables or by using over-the-air (OTA) technique to update the respective precoders at each BS. We will explore both possibilities by analyzing different techniques and the procedure to update the involved variables. Even though we claim that the distributed schemes require significantly less overhead as compared to the centralized one, it is not always true. For example, if we consider a semi-static fading scenario, where the channel remains relatively constant over multiple transmission slots, centralized scheme would be more beneficial as there is only one time overhead involved in updating CSI knowledge at the centralized controller.

However, in practice, we can always limit the number of iterations in the distributed scenario to have a compromise between the achievable rate to the involved overhead. Moreover, it is often enough to update only once per frame if the time correlated fading is slow enough, since the operating point can be initialized by the precoders obtained from previous transmission. Therefore, it would be beneficial to consider both procedures and to understand the update procedure in order to minimize the involved overhead. We consider the ADMM based distributed design due to its fast convergence properties [34]. Then, we study the distributed design of SPCA and MSE based centralized approach by using KKT expressions. In all cases, we consider both with and without guaranteed rate requirement constraint.

## 4.1. SPCA Formulation without QoS Requirements

In this section, we consider the problem of distributed precoder design using the ADMM with WSRM objective. In this decomposition scheme, the precoders are designed at each BS by exchanging the coupling interference information across backhaul that interconnects the coordinating BSs. In this procedure, users are not involved in the precoder design unlike the OTA based approach discussed in the following chapter. Let us consider the convex subproblem (20) for the $i$th iteration, rewritten as

$$\underset{w_k,\gamma_k,\beta_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{29a}$$

$$\text{subject to} \quad \Re\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \forall k \in \mathcal{U} \tag{29b}$$

$$\beta_k \geq \sigma^2 + \sum_{i=1, i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2, \forall k \in \mathcal{U}, \tag{29c}$$

$$\sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} \tag{29d}$$

where $\Im\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} = 0$ is implicitly assumed. In order to perform a distributed design of the above convex subproblem in (29), we adopt ADMM technique by introducing additional optimization variables [35] as

$$\underset{w_k,\gamma_k,\beta_k,\delta_{b,j}}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{30a}$$

$$\text{subject to} \quad \Re\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \ \forall k \in \mathcal{U} \tag{30b}$$

$$\sigma^2 + \sum_{\substack{i\in\mathcal{U}_{b_k} \\ i\neq k}} |\mathbf{h}_{b_k,k}\mathbf{w}_i|^2 + \sum_{b\in\mathcal{B}_{b_k}} \delta_{b,k} \leq \beta_k, \forall k \in \mathcal{U}_{b_k}, \tag{30c}$$

$$\delta_{b,k} \geq \sum_{i\in\mathcal{U}_b} |\mathbf{h}_{b,k}\mathbf{w}_i|^2, \ \forall k \in \mathcal{U}_{b_k}, \forall b \in \mathcal{B}_{b_k} \tag{30d}$$

$$\sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B}. \tag{30e}$$

where $\mathcal{B}_b = \{1, 2, \ldots, b-1, b, \ldots, N_B\}$, and $\delta_{b,i}$ is the interference caused from BS $b$ to user $i$. Equation (30d) is a relaxed interference constraint used to favor the distributed implementation.

Even after relaxing the interference terms from the respective neighboring BSs for each of the user $k$, (30) is still not in the form to be distributed across the coordinating BSs. Therefore, we now introduce additional BS specific variables that hold the local copy of the interference caused by the neighboring BS transmissions as

$$\underset{w_k,\gamma_k,\beta_k,\delta_{b,j}}{\text{maximize}} \quad \sum_{b\in\mathcal{B}} \sum_{k\in\mathcal{U}_b} \alpha_k \log(1+\gamma_k) \tag{31a}$$

$$\text{subject to} \quad \Re\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \ \forall k \in \mathcal{U} \tag{31b}$$

$$\sigma^2 + \sum_{\substack{i \in \mathcal{U}_{b_k} \\ i \neq k}} |\mathbf{h}_{b_k,k}\mathbf{w}_i|^2 + \sum_{b \in \mathcal{B}_{b_k}} \delta_{b,k}^{b_k} \leq \beta_k, \forall k \in \mathcal{U}_{b_k}, \tag{31c}$$

$$\sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} \tag{31d}$$

$$\delta_{b,k}^{b_k} \geq \sum_{i \in \mathcal{U}_b} |\mathbf{h}_{b,k}\mathbf{w}_i|^2, \; \forall k \in \mathcal{U}_{b_k}, \forall b \in \mathcal{B}_{b_k} \tag{31e}$$

$$\delta_{b_k,i}^{b_k} \geq \sum_{j \in \mathcal{U}_{b_k}} |\mathbf{h}_{b_k,i}\mathbf{w}_j|^2, \; \forall i \in \bar{\mathcal{U}}_{b_k} \tag{31f}$$

$$\delta_{b,k}^{b_k} = \delta_{b,k}, \; \forall k \in \mathcal{U}_{b_k}, \forall b \in \mathcal{B}_{b_k} \tag{31g}$$

$$\delta_{b_k,i}^{b_k} = \delta_{b_k,i}, \; \forall i \in \mathcal{U}_{b_k} \tag{31h}$$

where $\bar{\mathcal{U}}_b = \mathcal{U} \backslash \mathcal{U}_b$, $\delta_{b,k}^{b_k}$ denotes the local copy of the total interference caused by BS $b$ to user $k$, which is served by BS $b_k$. Similarly, $\delta_{b_k,i}^{b_k}$ represents the local copy of the actual interference caused by BS $b_k$ to user $i$, which is served by some other BS, say, $b$. The constraints in (31g) and (31h) are used to ensure that the local copies of the interference terms maintained at each BS are equal, *i.e.*, it ensures

$$\delta_{b,k}^{b} = \delta_{b,k}^{b_k} \tag{32}$$

which relates the actual interference $\delta_{b,k}^{b}$ caused by BS $b$ to user $k \in \mathcal{U}_{b_k}$ to the one assumed by BS $b_k$ for user $k$ as $\delta_{b,k}^{b_k}$.

Note that in order for the distributed implementation to be identical with the centralized design, (32) must be satisfied at the optimum. Therefore, to decentralize the precoder design, we consider using a partial Lagrangian for the equality constraint (32) and by collecting the variables that are relevant to BS $b_k$ as

$$\begin{aligned}
\underset{w_k,\gamma_k,\beta_k,\delta_{b,k}^{b_k},\delta_{b_k,i}^{b_k}}{\text{maximize}} \quad & \sum_{k \in \mathcal{U}_{b_k}} \alpha_k \log(1+\gamma_k) + \sum_{k \in \mathcal{U}_{b_k}} \left(\delta_{b,k}^{b_k} - \delta_{b,k}\right) \nu_{b,k} \\
& + \sum_{i \in \bar{\mathcal{U}}_{b_k}} \left(\delta_{b_k,i}^{b_k} - \delta_{b_k,i}\right) \nu_{b_k,i} && (33a) \\
\text{subject to} \quad & \Re\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \; \forall k \in \mathcal{U}_{b_k} && (33b) \\
& \sigma^2 + \sum_{\substack{i \in \mathcal{U}_{b_k} \\ i \neq k}} |\mathbf{h}_{b_k,k}\mathbf{w}_i|^2 + \sum_{b \in \mathcal{B}_{b_k}} \delta_{b,k}^{b_k} \leq \beta_k, \forall k \in \mathcal{U}_{b_k} && (33c) \\
& \delta_{b_k,i}^{b_k} \geq \sum_{i \in \bar{\mathcal{U}}_b} |\mathbf{h}_{b,k}\mathbf{w}_i|^2, \; \forall i \in \bar{\mathcal{U}}_{b_k} && (33d) \\
& \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall k \in \mathcal{U}_{b_k}. && (33e)
\end{aligned}$$

where $\bar{\mathcal{U}}_b = \mathcal{U} \backslash \mathcal{U}_b$, the dual variable corresponding to the constraint (32) is denoted by $\nu^i_{b,k}$, and (33d) denotes the total interference caused by BS $b_k$ transmission to the neighboring user $i \in \bar{\mathcal{U}}_{b_k}$. Additionally, note that $\delta_{b,k}$ is the global consensus interference corresponding to user $k$ from BS $b$. The above formulation is called as dual decomposition, and due to the instability involved in updating the consensus variable, we rely on the robust ADMM counterpart [36].

The ADMM based distributed precoder design is obtained by augmenting a strongly convex proximal term in the objective of each BS $b_k$ as

$$\begin{aligned}\underset{w_k,\gamma_k,\beta_k,\delta^{b_k}_{b,k},\delta^{b_k}_{b_k,i}}{\text{maximize}} \quad & \sum_{k\in\mathcal{U}_{b_k}} \alpha_k \log(1+\gamma_k) + \sum_{k\in\mathcal{U}_{b_k}} \left(\delta^{b_k}_{b,k} - \delta_{b,k}\right)\nu_{b,k} + \sum_{i\in\bar{\mathcal{U}}_{b_k}} \left(\delta^{b_k}_{b_k,i} - \delta_{b_k,i}\right)\nu_{b_k,i} \\ & + \frac{\rho}{2}\sum_{k\in\mathcal{U}_{b_k}} \|\delta^{b_k}_{b,k} - \delta_{b,k}\|^2 + \frac{\rho}{2}\sum_{i\in\bar{\mathcal{U}}_{b_k}} \|\delta^{b_k}_{b_k,i} - \delta_{b_k,i}\|^2. \end{aligned} \tag{34}$$

The proximal term $\|\delta^{b_k}_{b,k} - \delta_{b,k}\|^2$ ensures the uniqueness of the final solution and also stabilizes the update expression. Now, by using (34) in (33), we obtain the ADMM based distributed precoder design for each BS $b_k$ as

$$\begin{aligned}\underset{w_k,\gamma_k,\beta_k,\delta^{b_k}_{b,k},\delta^{b_k}_{b_k,i}}{\text{maximize}} \quad & \sum_{k\in\mathcal{U}_{b_k}} \alpha_k \log(1+\gamma_k) + \sum_{k\in\mathcal{U}_{b_k}} \left(\delta^{b_k}_{b,k} - \delta_{b,k}\right)\nu^i_{b,k} \\ & + \frac{\rho}{2}\sum_{k\in\mathcal{U}_{b_k}} \|\delta^{b_k}_{b,k} - \delta_{b,k}\|^2 + \sum_{i\in\bar{\mathcal{U}}_{b_k}} \left(\delta^{b_k}_{b_k,i} - \delta_{b_k,i}\right)\nu^i_{b_k,i} \\ & + \frac{\rho}{2}\sum_{i\in\bar{\mathcal{U}}_{b_k}} \|\delta^{b_k}_{b_k,i} - \delta_{b_k,i}\|^2 \qquad (35a) \\ \text{subject to} \quad & \Re\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} \geq \gamma_k \frac{\phi^{(i)}_k}{2} + \beta_k \frac{1}{2\phi^{(i)}_k}, \forall k \in \mathcal{U}_{b_k} \qquad (35b) \\ & \sigma^2 + \sum_{\substack{i\in\mathcal{U}_{b_k} \\ i\neq k}} |\mathbf{h}_{b_k,k}\mathbf{w}_i|^2 + \sum_{b\in\mathcal{B}_{b_k}} \delta^{b_k}_{b,k} \leq \beta_k, \forall k \in \mathcal{U}_{b_k} \qquad (35c) \\ & \delta^{b_k}_{b_k,i} \geq \sum_{i\in\bar{\mathcal{U}}_b} |\mathbf{h}_{b,k}\mathbf{w}_i|^2, \ \forall i \in \bar{\mathcal{U}}_{b_k} \qquad (35d) \\ & \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall k \in \mathcal{U}_{b_k}. \qquad (35e) \end{aligned}$$

The problem (35) is solved for fixed dual variable $\nu_{b,k}$ as $\nu^i_{b,k}$ and $\delta_{b,k}$. Upon solving (35) independently across each BS, the coupling variables such as $\delta_{b,k}$, $\nu^i_{b,k}$ need to be updated for obtaining the centralized solution. In order to do so, we need to exchange the interference variables $\delta^{b_k}_{b,k}$ and $\delta^{b}_{b,k}$ across the BSs $b$ and $b_k$. Upon obtaining the coupling interference variables, the global consensus variable $\delta_{b,k}$ is updated at the corresponding BSs $b$ and $b_k$ as

$$\delta_{b,k} = \frac{\delta^{b}_{b,k} + \delta^{b,k}_{b,k}}{2}. \tag{36}$$

Once the consensus terms are updated, the corresponding dual variable $\nu_{b,k}$ is modified by the subgradient update at BS $b_k$ as

$$\nu^{i+1}_{b,k} = \nu^{i}_{b,k} - \rho\left(\delta^{b_k}_{b,k} - \delta_{b,k}\right). \tag{37}$$

The algorithmic representation of the distributed ADMM based precoder design is outlined in Algorithm 1. The total number of variables, *i.e.*, the consensus interference $\delta_{b,k}$, that is exchanged across the coordinating BSs is given by $(N_B - 1) \times K$, since each user will see interference from $N_B - 1$ BSs excluding the serving BS.

**Algorithm 1** ADMM Method

**Input:** $\alpha_k, \mathbf{h}_{b_k,k}, \forall b \in \mathcal{B}, \forall k \in \mathcal{U}_b$.
**Output:** $\mathbf{w}_k, \forall k \in \{1, 2, \ldots, K\}$
**Initialization:** $i = 0$ and $\mathbf{w}_k$ by satisfying total transmit power constraint
initialize global interference vector $\delta^{0}_{b,k} = 0^T$
initialize the dual variables $\nu \forall b \in \mathcal{B}, \forall k$
**for each** BS $b \in \mathcal{B}$ perform the following procedure
**repeat**
  begin with $j = 0$
  **repeat**
    solve the precoders $\mathbf{w}_k$ and local interference $\delta^{b}_{b,k}$ using (35)
    exchange $\delta^{b}_{b,k}$ and $\delta^{b_k}_{b,k}$ among the coordinating BSs $b$ and $b_k$ via backhaul
    update the global consensus interference term as in (36)
    update the dual variables $\nu$ using (37)
  **until** do until convergence
  update the operating point $\phi^{(i)}_k$ with (18) by using the solution obtained from ADMM design
**until** perform until SPCA problem convergence

## 4.2. SPCA Formulation with QoS Requirements

In this section, we consider the above problem of maximizing the sum rate of all users with an additional QoS constraint in the form of minimum guaranteed rate requirement [15, 16, 17] The QoS requirements are usually guided by the service type associated with each transmission. For example, in order to provide an appreciable call quality in voice over IP (VoIP) service, a BS should ensure certain minimum guaranteed rate requirement for VoIP users. Furthermore, in long term evolution (LTE), guaranteed bit rate service (GBS) is one of the service qualifiers for data transmission.

In order to formulate the WSRM problem with a guaranteed rate requirement for each user, we include an additional constraint that ensures it as

$$\log(1+\gamma_k) \geq R_k,\ \forall k \in \mathcal{U} \tag{38}$$

where $R_k$ is the user specific minimum rate requirement. Since our objective is to distribute the precoder design with the minimum rate requirement, we reuse the final distributed precoder design formulation in (35) to provide the guaranteed minimum rate requirement. Because (38) is convex and it includes the optimization variable $\gamma_k$ that is associated to the respective user $k$ only, therefore, we can write the precoder design problem associated with BS $b_k$ as

$$\begin{aligned}
\underset{w_k,\gamma_k,\beta_k,\delta^{b_k}_{b,k},\delta^{b_k}_{b_k,i}}{\text{maximize}} \quad & \sum_{k\in\mathcal{U}_{b_k}} \alpha_k \log(1+\gamma_k) + \sum_{k\in\mathcal{U}_{b_k}} \left(\delta^{b_k}_{b,k} - \delta_{b,k}\right)\nu^i_{b,k} \\
& + \frac{\rho}{2}\sum_{k\in\mathcal{U}_{b_k}} \|\delta^{b_k}_{b,k} - \delta_{b,k}\|^2 + \sum_{i\in\bar{\mathcal{U}}_{b_k}} \left(\delta^{b_k}_{b_k,i} - \delta_{b_k,i}\right)\nu^i_{b_k,i} \\
& + \frac{\rho}{2}\sum_{i\in\bar{\mathcal{U}}_{b_k}} \|\delta^{b_k}_{b_k,i} - \delta_{b_k,i}\|^2 && \text{(39a)} \\
\text{subject to} \quad & \Re\{\mathbf{h}_{b_k,k}\mathbf{w}_k\} \geq \gamma_k \frac{\phi^{(i)}_k}{2} + \beta_k \frac{1}{2\phi^{(i)}_k},\ \forall k \in \mathcal{U}_{b_k} && \text{(39b)} \\
& \sigma^2 + \sum_{\substack{i\in\mathcal{U}_{b_k}\\ i\neq k}} |\mathbf{h}_{b_k,k}\mathbf{w}_i|^2 + \sum_{b\in\mathcal{B}_{b_k}} \delta^{b_k}_{b,k} \leq \beta_k, \forall k \in \mathcal{U}_{b_k} && \text{(39c)} \\
& \delta^{b_k}_{b_k,i} \geq \sum_{i\in\bar{\mathcal{U}}_b} |\mathbf{h}_{b,k}\mathbf{w}_i|^2,\ \forall i \in \bar{\mathcal{U}}_{b_k} && \text{(39d)} \\
& \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall k \in \mathcal{U}_{b_k} && \text{(39e)} \\
& \log(1+\gamma_k) \geq R_k,\ \forall k \in \mathcal{U}_b. && \text{(39f)}
\end{aligned}$$

The problem in (39) is convex in each SCA step $i$. Moreover, it involves only the variables that are associated to BS $b_k$. Therefore, (39) can be performed in parallel among each BS in $\mathcal{B}$ until convergence. Upon the convergence of the ADMM iterations, the SPCA update for variable $\phi^{(i)}_k$ is performed to proceed with the next iteration. The above procedure is performed until convergence of the objective sequence. The iterative procedure is similar to that of Algorithm 1.

Similarly, the centralized precoder design problem via MSE reformulation in (28) can also be performed in a decentralized manner by following the same steps as mentioned in Section 4.1. Since the ADMM procedure is straightforward, we refer the interested readers to [15]. The total number of variables, *i.e.*, the consensus interference $\delta_{b,k}$, that are exchanged across the coordinating BSs is given by $(N_B - 1) \times K$, since each user will see interference from $N_B - 1$ BSs excluding the serving BS.

# 5. DISTRIBUTED PRECODER DESIGN VIA KKT EXPRESSIONS

Even though the ADMM method of distributed precoder design has better convergence behavior compared to the other schemes like the primal and dual decomposition, the number of iterations and the overhead involved in the signaling limit its practical usage. In this chapter, we discuss an alternative precoder design by solving the KKT expressions in each SCA step across the coordinating BSs. Unlike the ADMM technique, where the precoders are designed at the BSs in a coordinated manner, the KKT approach includes users as well in the precoder design procedure via OTA, thereby reducing the utilization of the backhaul.

We formulate distributed precoder designs for the centralized problems in (20) and (28) by solving the respective KKT expressions. In addition, we also discuss the distributed precoder design to provide guaranteed minimum rate to all users in the system [15, 16, 17]. The main objective of the distributed design is to obtain a set of transmit precoders to all users in the system by reducing the amount of signaling overhead. It is due to the fact that the ADMM requires significant number of iterations before convergence, the overhead involved is dependent on the size of system, *i.e.*, the number users and BSs in the network. Even though distributed designs via the KKT expressions depends on the system size, the overhead involved and the number of iterations required to converge are significantly smaller compared to ADMM scheme. Therefore, the proposed methods are more suitable for the practical implementation.

## 5.1. SINR Relaxation via SPCA without QoS Requirements

Before proceeding with the distributed precoder design, we note that the OTA approach is not required to perform the SPCA based design while considering single-antenna receiver at user terminals. The precoders for the SPCA method can be designed explicitly by exchanging the coupling interference variables among the coordinating BSs via the backhaul. Even though it is similar to that of ADMM based distributed design, the number of iterations required to obtain an efficient set of precoders is significantly less when compared to the former approach. It follows from the fact that ADMM requires multiple iterations in each SCA step whereas the KKT based solution updates only at each SCA iteration. However, when the number of receive antenna is greater than one, then OTA based training can be considered as a viable practical implementation.

Let us proceed with the distributed design by writing the SPCA based centralized problem (20) along with the respective dual variables for each of the constraint as

$$\underset{w_k,\gamma_k,\beta_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) \tag{40a}$$

Subject to

$$a_k: \quad |\mathbf{h}_{b_k,k}\mathbf{w}_k| \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \forall k \in \mathcal{U} \tag{40b}$$

$$e_k: \quad \beta_k \geq \sigma^2 + \sum_{i=1, i\neq k}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2, \forall k \in \mathcal{U}, \tag{40c}$$

$$c_b: \quad \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} \tag{40d}$$

where $a_k$, $e_k$, and $c_b$ are dual variables corresponding to constraints (40b), (40c), and (40b), respectively. Note that we drop the $\Re\{.\}$ operator from constraint (40b), since we solve (40) by using the KKT expressions. Moreover, we drop the weights from objective function for clarity.

Because the problem defined by (40) is convex, it can be solved by using the KKT expressions. Let us write the Lagrangian of (40) with the corresponding dual variables as

$$L(\gamma_k, \beta_k, \mathbf{w}_k, a_k, e_k, c_b) = -\sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) + \sum_{b\in\mathcal{B}} c_b \left( \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|^2 - \beta_k \right)$$
$$+ \sum_{k=1}^{K} e_k \left( \sigma^2 + \sum_{\substack{i=1,\\ i\neq k}}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2 - \beta_k \right) + \sum_{k=1}^{K} a_k \left( \frac{1}{2\phi_k^{(i)}}\gamma_k + \frac{\phi_k^{(i)}}{2}\beta_k - |\mathbf{h}_{b_k,k}\mathbf{w}_k| \right). \tag{41}$$

Note that dual variable $c_b$ is associated with each BS whereas the dual variables $a_k$ and $e_k$ are related to each user. Now, the optimization problem is given by

$$\underset{a_k,e_k,c_b}{\text{maximize}} \quad \underset{w_k,\gamma_k,\beta_k}{\text{minimize}} \quad L(\gamma_k, \beta_k, \mathbf{w}_k, a_k, e_k, c_b) \tag{42}$$

where the solution is obtained by differentiating (41) with respect to each of the associated optimization and dual variables as presented in Appendix 9.3.1. Note that the objective is reversed in the Lagrangian expression due to the negative operator before the actual sum rate objective in (41).

Upon solving the KKT expressions in Appendix 9.3.1, we obtain the following system of update equations to design transmit precoders with fixed operating point

$$\phi_k^{(i)} = \sqrt{\frac{\beta_k^{(i-1)}}{\gamma_k^{(i-1)}}}. \tag{43}$$

Now, by using fixed $\phi_k^{(i)}$, other optimization variables are updated as

$$a_k^{(i)} = \frac{\alpha_k \, \phi_k^{(i)}}{1 + \gamma_k^{(i-1)}} \tag{44a}$$

$$e_k^{(i)} = \frac{a_k^{(i)} \phi_k^{(i)}}{2} \tag{44b}$$

$$\mathbf{w}_k^{(i)} = \frac{a_k^{(i)}}{2} \left( \sum_{i \neq K} e_i^{(i)} \mathbf{h}_{b_k,i}^H \mathbf{h}_{b_k,i} + c_b \mathbf{I}_{N_T} \right)^{-1} \mathbf{h}_{b_k,k}^H \tag{44c}$$

$$\beta_k^{(i)} = \sigma^2 + \sum_{j=1, j \neq k}^{K} |\mathbf{h}_{b_j,k} \mathbf{w}_j^{(i)}|^2 \tag{44d}$$

$$\gamma_k^{(i)} = 2\phi_k^{(i)} \left( |\mathbf{h}_{b_k,k} \mathbf{w}_k^{(i)}| - \frac{\phi_k^{(i)} \beta_k^{(i)}}{2} \right). \tag{44e}$$

Since the dual variable $a_k^{(i)}$ depends on $\phi_k^{(i)}$, the initial operating point $\phi_k^{(1)}$ is fixed by using some feasible transmit precoders $\mathbf{w}_k^{(0)}$. It follows from the fact that $\phi_k^{(1)}$ is given by (43) in which $\gamma_k^{(0)}$ and $\beta_k^{(0)}$ can be obtained for a fixed transmit precoder $\mathbf{w}_k^{(0)}$. Once $\phi_k^{(1)}$ is fixed, rest of the variables are updated in the order as outlined in (44). The dual variable $c_b$ is obtained at each BS such that the total power budget $P_b$ is satisfied by the transmit precoders $\mathbf{w}_k$. It is usually found by using the bisection search.

To obtain a practical distributed precoder design, we note that (44c) is the only constraint that involves the neighboring BSs dual variables $e_k^{(i)}$. Since the coupling dual variable is a scalar, it can shared among the respective BSs in $\mathcal{B}$ via backhaul to evaluate (44c). Upon obtaining the coupling dual variables $e_k^{(i)}$, the respective transmit precoders can be designed locally by updating the rest of the equations in (44). Once the transmit precoders $\mathbf{w}_k^{(i)}$ are evaluated, the interference variable $\beta_k^{(i)}$ needs to be identified at each BS to update $\gamma_k^{(i)}$. Since $\beta_k^{(i)}$ in (44d) has transmit precoders of the neighboring BSs, it cannot be obtained directly by using the available local information. Therefore, we require an additional backhaul transmission from all the neighboring BSs to notify the total interference caused by the current transmit precoders. Now, (44d) can be equivalently written as

$$\beta_k^{(i)} = \sigma^2 + \sum_{j \in \mathcal{U}_{b_k} \setminus k} |\mathbf{h}_{b_k,k} \mathbf{w}_j|^2 + \sum_{b \in \mathcal{B}_{b_k}} \sum_{j \in \mathcal{U}_b} |\mathbf{h}_{b,k} \mathbf{w}_j|^2 \tag{45a}$$

$$\delta_{b,k}^{(i)} = \sum_{j \in \mathcal{U}_b} |\mathbf{h}_{b,k}\mathbf{w}_j|^2, \ \forall b \in \mathcal{B}_{b_k} \quad (45b)$$

$$\beta_k^{(i)} = \sigma^2 + \sum_{j \in \mathcal{U}_{b_k} \setminus k} |\mathbf{h}_{b_k,k}\mathbf{w}_j|^2 + \sum_{b \in \mathcal{B}_{b_k}} \delta_{b,k}^{(i)} \quad (45c)$$

where $\delta_{b,k}^{(i)}$ is evaluated at each BS $b$ and notified to the corresponding serving BS. Upon receiving $\delta_{b,k}^{(i)}$ with the updated precoders, each BS then evaluates (44d) via (45) and proceed with the $\gamma_k^{(i)}$ calculation. The above procedure is performed until convergence in the same sequence as outlined in (44). The practical implementation of the above distributed scheme is outlined in Algorithm 2.

**Algorithm 2** SINR Relaxation via SPCA without QoS Requirements

**Input:** $\alpha_k$, $\mathbf{h}_{b_k,k}, \forall b \in \mathcal{B}, \forall k \in \mathcal{U}_b$.
**Output:** $\mathbf{w}_k, \forall k \in \mathcal{U}$
**Initialization:** $i = 1$, $\phi_k^{(1)}$ using some feasible transmit precoders $\mathbf{w}_k^{(0)}$
**repeat**
$\forall$ BSs $b \in \mathcal{B}$
solve (44a) and (44b) with $\phi_k^{(i)}$ using (43), $\forall k \in \mathcal{U}_b$
perform backhaul exchange among coordinating BSs to notify dual variables $e_k^{(i)}$
with the updated dual variables $e_k^{(i)}$ from neighboring BSs, solve (44c)
upon finding $\mathbf{w}_k^{(i)}$, all BSs evaluate $\delta_{b,k}, \forall k \in \bar{\mathcal{U}}_b$ and notify the same to respective serving BS
after obtaining $\delta_{b,k}$ from each $b \in \mathcal{B}_{b_k}$ and for all $k \in \mathcal{U}_{b_k}$, BSs then update $\beta_k^{(i)}$ and $\gamma_k^{(i)}$ by using (45) and (44e) locally
**until** convergence

## 5.2. SINR Relaxation via SPCA with QoS Requirements

Now, we extend the above method of precoder design with SPCA method to include an additional QoS constraint. By including a minimum guaranteed rate constraint as $\log(1 + \gamma_k) \geq R_k$, the problem in (40) can be rewritten by representing the dual variables as

$$\underset{w_k, \gamma_k, \beta_k}{\text{maximize}} \quad \sum_{k=1}^{K} \alpha_k \log(1 + \gamma_k) \quad (46a)$$

Subject to

$$a_k : \quad |\mathbf{h}_{b_k,k}\mathbf{w}_k| \geq \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}}, \forall k \in \mathcal{U} \quad (46b)$$

$$e_k: \quad \beta_k \geq \sigma^2 + \sum_{j=1, j\neq k}^{K} |\mathbf{h}_{b_j,k}\mathbf{w}_j|^2, \forall k \in \mathcal{U}, \tag{46c}$$

$$c_b: \quad \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} \tag{46d}$$

$$d_k: \quad \log(1+\gamma_k) \geq R_k \tag{46e}$$

where the additional dual variable $d_k$ corresponds to the guaranteed minimum rate requirement associated with user $k$.

Following the similar procedure as in Section 5.1, we formulate the Lagrangian of (46) as

$$\begin{aligned} L(\gamma_k, \beta_k, \mathbf{w}_k, a_k, e_k, c_b, d_k) = &-\sum_{k=1}^{K} \alpha_k \log(1+\gamma_k) + \sum_{b\in\mathcal{B}} c_b \left( \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right) \\ &+ \sum_{k=1}^{K} a_k \left( \frac{1}{2\phi_k^{(i)}}\gamma_k + \frac{\phi_k^{(i)}}{2}\beta_k - |\mathbf{h}_{b_k,k}\mathbf{w}_k| \right) + \sum_{k=1}^{K} d_k \Big( R_k - \log(1+\gamma_k) \Big) \\ &+ \sum_{k=1}^{K} e_k \left( \sigma^2 + \sum_{\substack{i=1,\\ i\neq k}}^{K} |\mathbf{h}_{b_i,k}\mathbf{w}_i|^2 - \beta_k \right) \end{aligned} \tag{47}$$

and the corresponding optimization problem is given by

$$\underset{a_k,b_k,c_b,d_k}{\text{maximize}} \quad \underset{w_k,\gamma_k,\beta_k}{\text{minimize}} \quad L(\gamma_k, \beta_k, \mathbf{w}_k, a_k, e_k, c_b, d_k) \tag{48}$$

To solve the above optimization problem, we equate the derivative of (47) with respect to each of the optimization and dual variables to zero as shown in Appendix 9.3.1.

Upon solving the KKT expression in (67) and (68b), we obtain the following update expressions with $\phi_k^{(i)} = \sqrt{\frac{\beta_k^{(i-1)}}{\gamma_k^{(i-1)}}}$ as

$$a_k^{(i)} = \frac{\alpha_k \, \phi_k^{(i)} \left(1 + d_k^{(i-1)}\right)}{1+\gamma_k^{(i-1)}} \tag{49a}$$

$$e_k^{(i)} = \frac{a_k^{(i)} \phi_k^{(i)}}{2} \tag{49b}$$

$$\mathbf{w}_k^{(i)} = \frac{a_k^{(i)}}{2} \left( \sum_{i\neq K} e_i^{(i)} \mathbf{h}_{b_k,i}^H \mathbf{h}_{b_k,i} + c_b \mathbf{I}_{N_T} \right)^{-1} \mathbf{h}_{b_k,k}^H \tag{49c}$$

$$\beta_k^{(i)} = \sigma^2 + \sum_{j=1, j\neq k}^{K} |\mathbf{h}_{b_j,k}\mathbf{w}_j^{(i)}|^2 \tag{49d}$$

$$\gamma_k^{(i)} = 2\phi_k^{(i)}\left(|\mathbf{h}_{b_k,k}\mathbf{w}_k^{(i)}| - \frac{\phi_k^{(i)}\beta_k^{(i)}}{2}\right) \quad (49e)$$

$$d_k^{(i)} = d_k^{(i-1)} + \rho\left[R_k - \log(1+\gamma_k^{(i)})\right]^+ \quad (49f)$$

where (49f) denotes the subgradient update for dual variable $d_k$ and $\rho$ is some step size. The operator $[x]^+$ in (49f) is defined as $[x]^+ = \max(x, 0)$, which ensures $d_k^{(i)} \geq 0$. The step size parameter $\rho$ can either be a constant or diminishing one by following the discussions in [34]. The dual variable $c_b$ is obtained such that the total power budget $P_b$ is satisfied by the transmit precoders $\mathbf{w}_k$. It is usually found by using the bisection search. To conclude this section, a practical way of implementing the above set of equations in (49), we can follow Algorithm 2.

Even though the distributed methods proposed to solve (44) and (49) requires only the backhaul exchanges to update the coupling dual variables $e_k^{(i)}$, it is not the case when the receivers are equipped with multiple antennas. The receiver side beamformers are also required to design an efficient set of transmit precoders, since (44c) and (49c) will include receive beamformers of all the interfering users. Therefore, for a multi-antenna receive scenario, the overhead involved in feeding back all the interference channels seen by users to the respective BSs is significantly larger, thereby requiring high capacity backhaul links between BSs.

Due to the limited capacity of the existing backhaul fibers, it is often not possible to carry out the iterative distributed precoder designs via backhaul alone. As an alternative, we can consider including users in the precoder design via OTA signaling as discussed in [37]. In such a case, users will perform all the necessary updates based on the downlink precoded pilot transmissions from all BSs to design receive beamformer. It is then notified to all BSs through uplink sounding pilots so as to update the respective transmit precoders at the BSs for next iteration. This procedure is discussed below on distributed designs for MSE reformulated problem, since MMSE equalizer is required even for single antenna receivers to utilize the relation between MSE and SINR. The algorithmic representation is presented in Algorithm 3.

The total number of variables exchanged via backhaul can be calculated as follows. At first, we consider dual variable $e_k$ that corresponds to each user in the system, therefore, it contributes $K$ scalar variables. Secondly, we have $\delta_{b,k}$ that corresponds to each interference link, contributing $(N_B - 1) \times K$. Therefore, the overall number of scalar entries that are exchanged via backhaul is $(N_B - 1) \times K + K$, which is larger than the ADMM overhead by $K$. However, the ADMM involves multiple iterations for each SCA update point unlike the KKT method, which operates at SCA step. Unless the number of ADMM iterations is one, the KKT based schemes will always have less signaling for the given throughput improvement.

**Algorithm 3** SINR relaxation via SPCA with QoS Requirements

**Input:** $\alpha_k$, $\mathbf{h}_{b_k,k}, \forall b \in \mathcal{B}, \forall k \in \mathcal{U}_b$.
**Output:** $\mathbf{w}_k, \forall k \in \mathcal{U}$
**Initialization:** $i = 1$, $\phi_k^{(1)}$ using some feasible transmit precoders $\mathbf{w}_k^{(0)}$
**repeat**
$\forall$ BSs $b \in \mathcal{B}$
solve (49a) and (49b) with $\phi_k^{(i)}$ using (43), $\forall k \in \mathcal{U}_b$
perform backhaul exchange among coordinating BSs to notify dual variables $e_k^{(i)}$
with the updated dual variables $e_k^{(i)}$ from neighboring BSs, solve (49c)
upon finding $\mathbf{w}_k^{(i)}$, all BSs evaluate $\delta_{b,k}, \forall k \in \bar{\mathcal{U}}_b$ and notify the same to respective serving BS
after obtaining $\delta_{b,k}$ from each $b \in \mathcal{B}_{b_k}$ and for all $k \in \mathcal{U}_{b_k}$, BSs then update $\beta_k^{(i)}$, $\gamma_k^{(i)}$, and $d_k^{(i)}$ by using (45), (49e), and (49f) locally
**until** convergence

### 5.3. MSE Reformulation without QoS Requirements

In this section, we discuss an alternative approach of designing the transmit precoders by considering the MSE based problem in (28). Let us proceed further by rewriting the MSE reformulated convex subproblem for the $i$th SCA iteration with dual variables as

$$\underset{w_k,\epsilon_k,t_k}{\text{minimize}} \quad \sum_{i=1}^{K} \alpha_k \left(\bar{\epsilon}_k^{(i)}\right)^{-1} \epsilon_k \tag{50a}$$

subject to

$$a_k : \quad \epsilon_k \geq |1 - u_k^* \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_i^* \mathbf{h}_{b_k,i} \mathbf{w}_i|^2 + \bar{N}_0 \tag{50b}$$

$$c_b : \quad \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} \tag{50c}$$

where $a_k$ and $c_b$ are dual variables corresponding to the constraints (50b) and (50c), respectively. The SCA operating point is given by fixed MSE point obtained from previous SCA iteration as $\bar{\epsilon}_k^{(i)}$.

In order to solve the convex subproblem (50), we write the corresponding Lagrangian with the dual variables as

$$\begin{aligned} L(\epsilon_k, \mathbf{w}_k, a_k, c_b) = \sum_{k=1}^{K} \alpha_k \left(\bar{\epsilon}_k^{(i)}\right)^{-1} + \sum_{b \in \mathcal{B}} c_b \left( \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right) \\ \sum_{k=1}^{K} a_k \left( |1 - u_k^* \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_i^* \mathbf{h}_{b_k,i} \mathbf{w}_i|^2 + \bar{N}_0 - \epsilon_k \right) \end{aligned} \tag{51}$$

The dual variable $c_b$ is associated with each BS, and $a_k$ is the dual variable associated with the MSE constraint of each user $k$. Using this, the optimization problem is given as

$$\underset{a_k, c_b}{\text{maximize}} \quad \underset{w_k, \epsilon_k}{\text{minimize}} \quad L(\epsilon_k, \mathbf{w}_k, a_k, c_b) \tag{52}$$

and the convex subproblem is solved by equating the derivative of the Lagrangian with respect to each of the optimization and dual variables to zero. For further details on the KKT conditions, readers are referred to Appendix 9.4.1.

By using the gradient conditions (69), primal constraints in (50), and the slackness criterion in (70), we obtain the following iterative solution as

$$a_k^{(i)} = \frac{\alpha_k}{\bar{\epsilon_k}^{(i-1)}} \tag{53a}$$

$$\mathbf{w}_k^{(i)} = a_k^{(i)} \left( \sum_{j=1}^{K} a_j^{(i)} \mathbf{h}_{b_j,k}^H u_j^{(i-1)} u_j^{*(i-1)} \mathbf{h}_{b_k,k} + c_b \mathbf{I} \right)^{-1} u_k^{*(i-1)} \mathbf{h}_{b_k,k} \tag{53b}$$

$$u_k^{(i)} = \left( \sum_{j=1}^{K} \mathbf{h}_{b_j,k} \mathbf{w}_j^{(i)} \mathbf{w}_j^{H(i)} \mathbf{h}_{b_j,k}^H + \sigma^2 \right)^{-1} \mathbf{h}_{b_k,k} \mathbf{w}_k^{(i)} \tag{53c}$$

$$\epsilon_k^{(i)} = |1 - u_k^{*(i)} \mathbf{h}_{b_k,k} \mathbf{w}_k^{(i)}|^2 + \sum_{j \in \bar{\mathcal{U}}_b} |u_j^{*(i)} \mathbf{h}_{b_k,j} \mathbf{w}_j^{(i)}|^2 + |u_k^{(i)}|^2 N_0. \tag{53d}$$

The dual variable $c_b$ is obtained such that the total power budget $P_b$ is satisfied by the transmit precoder $\mathbf{w}_k$. It is usually found by using the bisection search. The above KKT expressions are solved iteratively until convergence to obtain an efficient set of transmit and receive beamformers.

In order to the evaluate the precoder $\mathbf{w}_k \forall k$, the MMSE receivers corresponding to all interfering users are required at each BS $b \in \mathcal{B}$. Moreover, to evaluate the MMSE receivers associated with each user in the system, complete interference channel information should be available at the BSs. Therefore, implementing the MSE based precoder design via backhaul exchange is not viable as it requires complexity similar to that of the centralized schemes.

However, by following various approaches presented in [37], the OTA based precoder training procedure is a viable option for practical implementation. It is achieved by the following procedure. Each BS evaluate the transmit precoders $\mathbf{w}_k$ of the associated users from (53b) with arbitrary dual variables $a_k$. Upon finding the transmit precoders, all BSs will then transmit precoded pilots in an orthogonal manner with $\mathbf{w}_k$ as precoders in the downlink training phase. All users in the system receive orthogonal pilot transmissions from each BS and then evaluate the effective channels, *i.e.*, user $k$ will estimate $\mathbf{h}_{b_j,k} \mathbf{w}_j^{(i)}$ from each BS $b_i \in \mathcal{B}$.

Now, each user updates the respective MMSE receivers by using (53c) with the desired $\mathbf{h}_{b_k,k}\mathbf{w}_k^{(i)}$ and interfering $\mathbf{h}_{b_j,k}\mathbf{w}_j^{(i)}$ effective channels. Once the receivers are updated as $u_k^{(i)}$, each user then proceeds to update the MSE operating point $\epsilon_k^{(i)}$ with (53d). Note that users can also consider using an alternate MSE expression for (53d), which is given as

$$\epsilon_k^{(i)} = 1 - u_k^{(i)}\mathbf{h}_{b_k,k}\mathbf{w}_k^{(i)}. \tag{54}$$

Since all the necessary variables are already present at the user terminal. Upon evaluating $\epsilon_k^i$, each user then update $a_k^{(i+1)}$ by using (53a). Now, users will send an uplink precoded pilot to inform both receive beamformer $u_k^{(i)}$ and dual variable $a_k^{(i+1)}$. It is achieved by using $\sqrt{a_k^{(i+1)}}u_k^{*(i)}$ and $a_k^{(i+1)}u_k^{*(i)}$ as precoders for pilots that are transmitted orthogonally in the uplink direction, where $u_k^*$ denotes the complex conjugate of $u_k$. Each BS then receives an equivalent channel as

$$\mathbf{h}_{b,k}^T\sqrt{a_k^{(i+1)}}u_k^{*(i)} \tag{55}$$

where $\mathbf{h}_{b,k}^T$ is the reciprocal uplink channel with dimension $\mathbb{C}^{N_T\times 1}$ since $\mathbf{h}_{b,k} \in \mathbb{C}^{1\times N_T}$.

By using the effective uplink channel, each BS then updates the corresponding transmit precoders $\mathbf{w}_k^{(i+1)}$ by using (53b). The above discussed procedure is carried out until convergence or for limited number of iterations depending on the signaling overhead involved. This is also briefly outlined in Algorithm 4.

**Algorithm 4** MSE Reformulation without QoS Requirements

**Input:** $\alpha_k, \mathbf{h}_{b,k}, \forall b \in \mathcal{B}, \forall k \in \mathcal{U}$.
**Output:** $\mathbf{w}_k, u_k, \forall k \in \mathcal{U}$
**Initialization:** $i = 1$ , dual variables ${a_k}^{(0)} = 1$, transmit precoders ${\mathbf{w_k}}^{(0)}$, MMSE receivers $u_k^{(0)}$ using (53c)
**repeat**
for all BS $b \in \mathcal{B}$, update $\mathbf{w}_k^{(i)}$ with (53b) and use it to perform downlink precoded pilot transmission
for all user $k \in \mathcal{U}$, execute the following steps
evaluate MMSE receiver $u_k^{(i)}$ by using (53c) with effective downlink channel
update MSE operating point $\epsilon_k^{(i)}$ with (54)
evaluate dual variable $a_k^{(i+1)}$ by using (53a)
using $\sqrt{a_k^{(i+1)}}u_k^{*(i)}$ and $a_k^{(i+1)}u_k^{*(i)}$ as precoders, uplink precoded pilots are sent from each user orthogonally to all BS in $\mathcal{B}$
**until** convergence

### 5.4. MSE Reformulation with QoS Requirements

In this section, we discuss the MSE reformulated distributed precoder design presented in section 5.3 with the additional guaranteed minimum rate requirement for all users [15, 16]. Since the formulation is very much similar to (50), we rewrite with the corresponding dual variables as

$$\begin{aligned}
&\underset{w_k,\epsilon_k,t_k}{\text{minimize}} && \sum_{i=1}^{K} \alpha_k \left(\bar{\epsilon}_k^{(i)}\right)^{-1} \epsilon_k && \text{(56a)}\\
&\text{subject to} && && \\
&a_k: && \epsilon_k \geq |1 - u_k^* \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_i^* \mathbf{h}_{b_k,i} \mathbf{w}_i|^2 + \bar{N}_0 && \text{(56b)}\\
&c_b: && \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} && \text{(56c)}\\
&d_k: && -\log \epsilon_k \geq R_k && \text{(56d)}
\end{aligned}$$

where $a_k, c_b$ and $d_k$ are dual variables corresponding to the constraints (56b), (56c), and (56d), respectively. The SCA operating point is given by fixed MSE point obtained from previous SCA iteration as $\bar{\epsilon}_k^{(i)}$. However, note that (56d) is nonconvex due to the involved variables, therefore, (56) is a nonconvex problem.

In order to solve (56), we use first order Taylor approximation for the convex function $-\log(\epsilon_k)$ as

$$-\log(\bar{\epsilon}_k) - \frac{\epsilon_k - \bar{\epsilon}_k}{\bar{\epsilon}_k} \geq R_k \tag{57}$$

which is a convex constraint. Let us now replace (56d) with (57) to obtain a convex problem as

$$\begin{aligned}
&\underset{w_k,\epsilon_k,t_k}{\text{minimize}} && \sum_{i=1}^{K} \alpha_k \left(\bar{\epsilon}_k^{(i)}\right)^{-1} \epsilon_k && \text{(58a)}\\
&\text{subject to} && && \\
&a_k: && \epsilon_k \geq |1 - u_k^* \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_i^* \mathbf{h}_{b_k,i} \mathbf{w}_i|^2 + \bar{N}_0 && \text{(58b)}\\
&c_b: && \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 \leq P_b, \forall b \in \mathcal{B} && \text{(58c)}\\
&d_k: && \log(\bar{\epsilon}_k) + \frac{\epsilon_k - \bar{\epsilon}_k}{\bar{\epsilon}_k} + R_k \leq 0 && \text{(58d)}
\end{aligned}$$

and the associated Lagrangian with the corresponding dual variables is given by

$$L(\epsilon_k, \mathbf{w}_k, a_k, c_b, d_k) = \sum_{k=1}^{K} \alpha_k \left(\bar{\epsilon}_k^{(i)}\right)^{-1} + \sum_{b \in \mathcal{B}} c_b \left( \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right)$$
$$\sum_{k=1}^{K} a_k \left( |1 - u_k^* \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} |u_i^* \mathbf{h}_{b_k,i} \mathbf{w}_i|^2 + \bar{N}_0 - \epsilon_k \right)$$
$$+ \sum_{k=1}^{K} d_k \left( \log(\bar{\epsilon}_k) + \frac{\epsilon_k - \bar{\epsilon}_k}{\bar{\epsilon}_k} + R_k \right) \quad (59)$$

The dual variable $c_b$ is associated with each BS, and $a_k, d_k$ are dual variables corresponding to MSE and QoS constraint of each user $k$. Using this, the optimization problem is given by

$$\underset{a_k, c_b, d_k}{\text{maximize}} \quad \underset{w_k, \epsilon_k}{\text{minimize}} \quad L(\epsilon_k, \mathbf{w}_k, a_k, c_b, d_k) \quad (60)$$

and the convex subproblem is solved by equating the derivative of the Lagrangian with respect to each of the optimization and dual variables to zero. The further details of the KKT conditions are included in Appendix 9.4.2.

By using the gradient conditions (71), the primal constraints in (58), and the slackness criterion in (72), we obtain the following iterative solution as

$$a_k^{(i)} = -\frac{d_k^{(i-1)}}{\log\left(\epsilon_k{}^{(i-1)}\right)} + \frac{\alpha_k}{\epsilon_k{}^{(i-1)}} \quad (61a)$$
$$\mathbf{w}_k^{(i)} = a_k^{(i)} \left( \sum_{j=1}^{K} a_j^{(i)} \mathbf{h}_{b_j,k}^H u_j^{(i-1)} u_j^{H(i-1)} \mathbf{h}_{b_k,k} + c_b \mathbf{I} \right)^{-1} \mathbf{u}_k^{H(i-1)} \mathbf{h}_{b_k,k} \quad (61b)$$
$$u_k^{(i)} = \left( \sum_{j=1}^{K} \mathbf{h}_{b_j,k} \mathbf{w}_j^{(i)} \mathbf{w}_j^{H(i)} \mathbf{h}_{b_j,k}^H + \sigma^2 \mathbf{I}_{N_R} \right)^{-1} \mathbf{h}_{b_k,k} \mathbf{w}_k^{(i)} \quad (61c)$$
$$\epsilon_k^{(i)} = |1 - u_k^{*(i)} \mathbf{h}_{b_k,k} \mathbf{w}_k^{(i)}|^2 + \sum_{j \in \bar{\mathcal{U}}_b} |u_j^{*(i)} \mathbf{h}_{b_k,j} \mathbf{w}_j^{(i)}|^2 + |u_k^{(i)}|^2 N_0 \quad (61d)$$
$$d_k^{(i)} = d_k^{(i-1)} + \rho_k^{(i)} \left[ \log(\epsilon_k^{(i-1)}) + \frac{\epsilon_k^{(i)} - \bar{\epsilon}_k^{(i-1)}}{\epsilon_k^{(i-1)}} + R_k \right]^+ . \quad (61e)$$

where $\rho_k^{(i)}$ is the subgradient step size, which can either be constant or diminishing in each update step. The dual variable $c_b$ is obtained such that the total power budget $P_b$ is satisfied by the transmit precoders $\mathbf{w}_k$. It is usually found by using the bisection search. The distributed implementation follows the same procedure as that of the one presented in Section 5.3 without QoS constraint. For completeness, the algorithmic representation is presented in Algorithm 5 to illustrate the distributed precoder design to provide guaranteed minimum rate.

**Algorithm 5** MSE Reformulation with QoS Requirements

**Input:** $\alpha_k, \mathbf{h}_{b,k}, \forall b \in \mathcal{B}, \forall k \in \mathcal{U}$.
**Output:** $\mathbf{w}_k, u_k, \forall k \in \mathcal{U}$
**Initialization:** $i = 1$ , dual variables $a_k{}^{(0)} = 1$, transmit precoders $\mathbf{w_k}^{(0)}$, MMSE receivers $u_k^{(0)}$ using (53c), and dual variables $d_k^{(0)} = 1$.
**repeat**
for all BS $b \in \mathcal{B}$, update $\mathbf{w}_k^{(i)}$ with (61b) and use it to perform downlink precoded pilot transmission
for all user $k \in \mathcal{U}$, execute the following steps
evaluate MMSE receiver $u_k^{(i)}$ by using (61c) with effective downlink channel
update MSE operating point $\epsilon_k^{(i)}$ with (54)
dual variable $d_k^{(i)}$ is updated by using subgradient method in (61e)
evaluate dual variable $a_k^{(i+1)}$ by using (61a)
using $\sqrt{a_k^{(i+1)}}u_k^{*(i)}$ and $a_k^{(i+1)}u_k^{*(i)}$ as precoders, uplink precoded pilots are sent from each user orthogonally to all BS in $\mathcal{B}$
**until** convergence

Since the subgradient method is used to update the dual variable $d_k$ corresponding to the guaranteed minimum rate constraint, the convergence is typically slower as compared to Algorithm 4. However, performing distributed updates until convergence is not efficient. It follows from the fact that performance improvement is noticeably large in the first few iterations than in the final stages. Therefore, it is worthwhile to perform distributed approach for few iterations and use it as a starting point for the forthcoming transmission slots by considering time-correlated fading nature [35]. Note that there is no need for any backhaul transmission in MSE based design unlike SPCA method.

# 6. NUMERICAL RESULTS

In this section, we analyze the performance of MIMO IBC precoder design for various scenarios. At first, we demonstrate the sum rate behavior of various precoder designs presented in Sections 4.1, 5.1, and 5.3, then, we present the same with minimum QoS requirement.

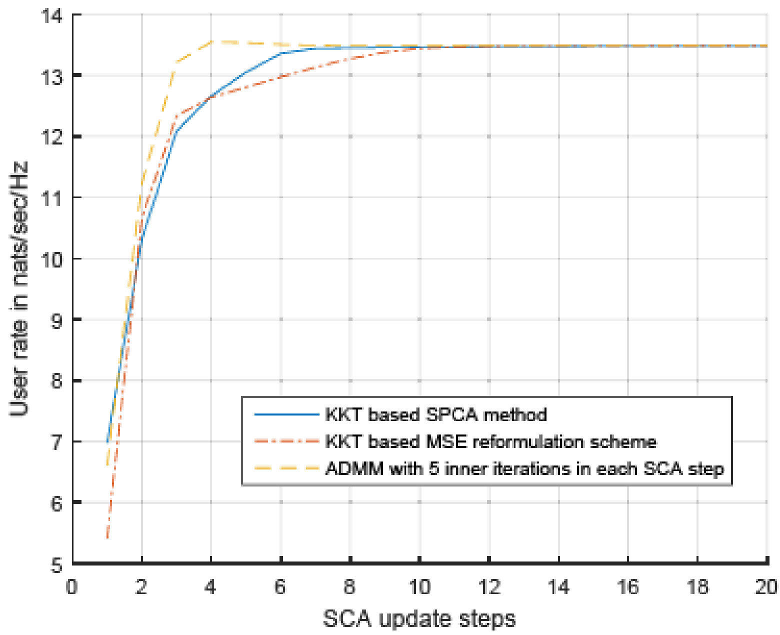


Figure 6: Sum rate performance for $N_{\mathrm{T}} = 8, N_{\mathrm{B}} = 2, K = 4$ model at 10 dB SNR with users at cell-edge.

Figure 6 illustrates the convergence of the proposed algorithms by distributing the users around the cell-edge. The scenario considered involves $N_{\mathrm{B}} = 2$ BSs, having $N_{\mathrm{T}} = 8$ transmit antennas, operating at 10dB SNR and serving $K = 2$ single receive antenna users in a coordinated manner. Figure 6 demonstrates the sum rate performance of various algorithms. The sum rate plot is shown only at the SCA update points.

Figure 7 demonstrates a scenario with $N_{\mathrm{B}} = 2$ BSs, each equipped with $N_{\mathrm{T}} = 8$ transmit elements operating at 10 dB SNR, serving $K = 6$ single receive antenna users in a coordinated manner. The users are assumed to be distributed with signal-to-interference-ratio (SIR) in $[0, 12]$ dB. The performances of the ADMM, SPCA, and MSE based distributed precoder designs are shown in Figure 7. The initial values of the precoders are generated based on single user transmit beamformer. The KKT based SPCA and MSE reformulation methods converge monotonically, whereas the convergence of ADMM need not be monotonic in each the ADMM update.

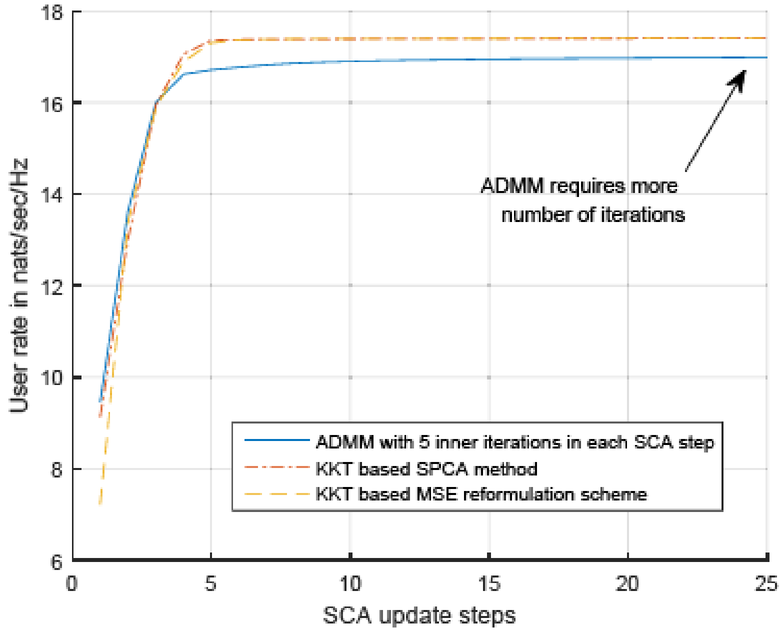


Figure 7: Sum rate performance for $N_{\rm T} = 8, N_{\rm B} = 2, K = 6$ model with pathloss in $[0, -6]$ dB at $10$ dB SNR.

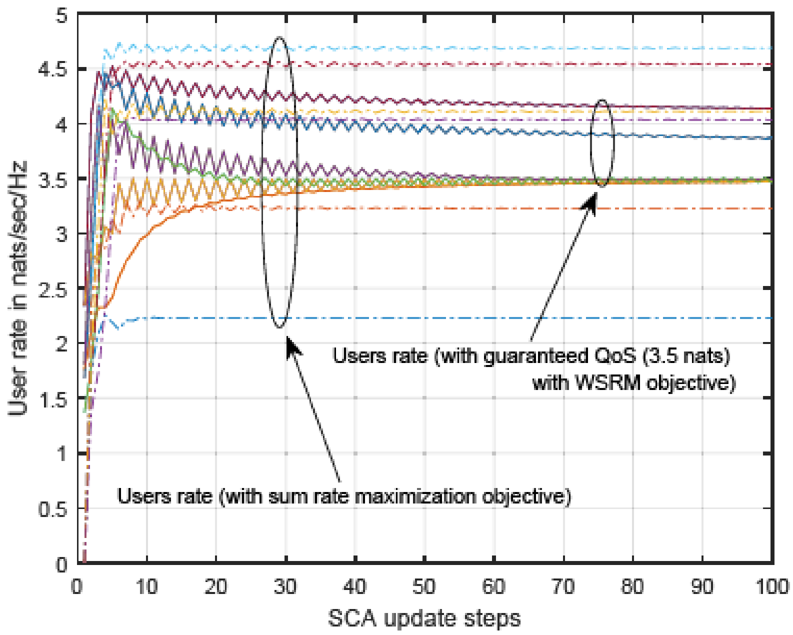


Figure 8: Behavior of users at $10$ dB for SPCA Approach with and without QoS constraints $N_{\rm T} = 8, N_{\rm B} = 2, K = 6$ model.

Figures 8-11 demonstrates the behavior of the proposed algorithms with and without QoS constraints in the formulation. A cell edge scenario is illustrated. Each user with a QoS constraint is subject to a minimum guaranteed rate in nats. The convergences for the proposed algorithms are studied with a constant or diminishing step size and the user rates are updated after every iteration.

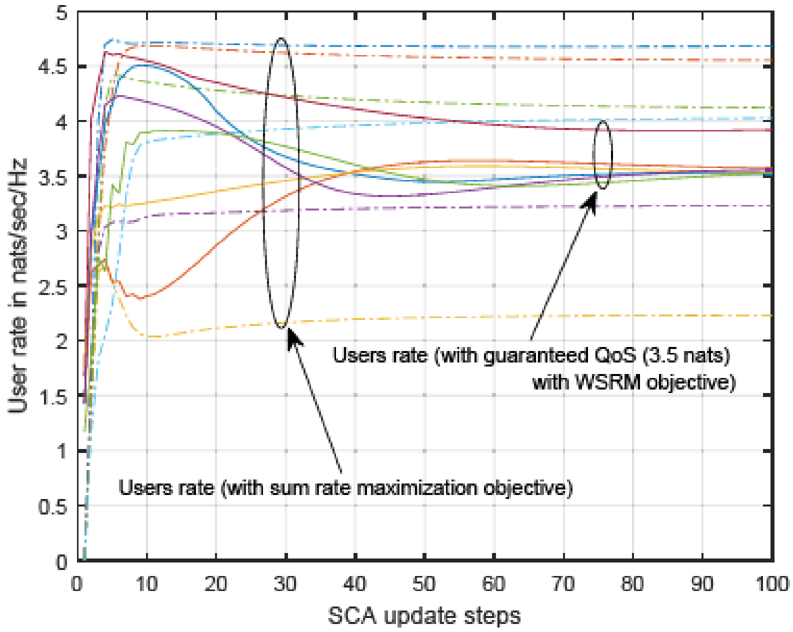


Figure 9: Behavior of users at 10 dB for MSE Approach with and without QoS constraints $N_{\mathrm{T}} = 8, N_{\mathrm{B}} = 2, K = 6$ model.

In Figure 8, the behavior for SINR relaxation via SPCA with QoS requirement, considering $N_{\mathrm{B}} = 2$ BSs, each having $N_{\mathrm{T}} = 8$ transmit antennas, and serving $K = 6$ users in the system. For this setup, we observe that all the $6$ users are provided a minimum QoS rate of $3.5$ nats, where most of the users attain the QoS requirement. In addition, it can be observed that the convergence has several glitches and is not monotonic due to the sub gradient update used for the QoS constraint.

Similarly, Figure. 9 exemplifies the behavior of the MSE reformulation with QoS requirement, where the system consists of $N_{\mathrm{B}} = 2$ BSs, each having $N_{\mathrm{T}} = 8$ transmit antennas, operating at 10 dB SNR and serving $K = 6$ users. A minimum rate of $3.5$ nats is provided for each user in the system. We observe that all the $6$ users in the system obtain the minimum QoS rate. Similar to the above discussion, the convergence may not be monotonic due to the sub gradient update.

Figure 10, demonstrates the infeasible case of QoS requirements, *i.e.*, by fixing minimum guaranteed rate requirement of $3.5$ nats for each user. The users are distributed in such a way that the SIR seen by any user lies in $[0, 12]$ dB. Since the algorithm cannot guarantee the QoS requirement for the given transmit power, it can be seen from Figure 10 that one of the user rate is in fact decreasing as highlighted in the figure.

Figure 11 illustrates the performance of the ADMM scheme with and without QoS constraints. The model involves $N_{\mathrm{B}} = 2$ BSs, equipped with $N_{\mathrm{T}} = 8$ transmit ele-

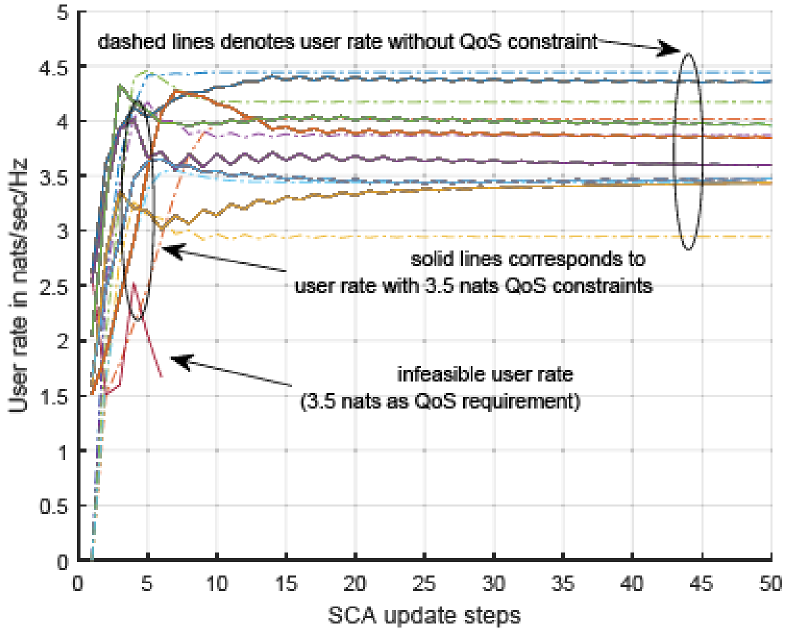


Figure 10: Behavior of users at 10 dB for MSE Approach with and without QoS constraints $N_{\mathrm{T}} = 8, N_{\mathrm{B}} = 3, K = 6$ model.

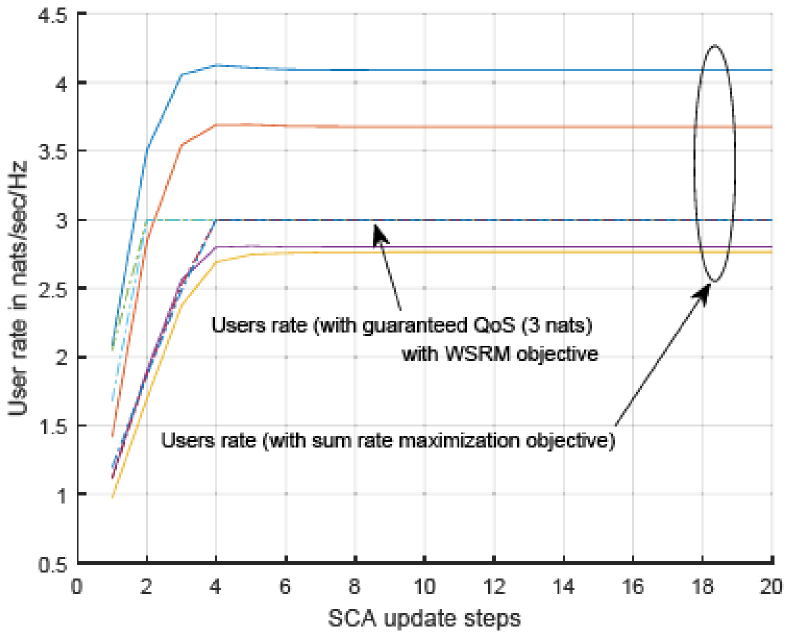


Figure 11: Behavior of users at 10 dB for ADMM method with and without QoS constraints $N_{\mathrm{T}} = 8, N_{\mathrm{B}} = 2, K = 4$ model.

ments serving $K = 4$ single antenna users. As seen from Figure 11 that the minimum rate provided for a user is $\approx 2.75$ nats with the WSRM objective. However, when we included an additional guaranteed rate constraint of 3 nats to each user, the proposed

ADMM based distributed algorithm provided the required QoS of 3 nats to all users as highlighted in Figure 11.

# 7. SUMMARY AND CONCLUSIONS

In this work, we designed transmit precoders for a multi-cell multi-user multiple-input multiple-output system with user specific minimum rate as quality of service requirements. In general, the goal of precoding is to maximize the signal power at the intended terminal while minimizing the interference caused at other terminals. In order to design transmit precoders, we considered three distributed precoder designs and analyzed the practical feasibility of the proposed schemes.

In this thesis, we first discussed centralized precoder design by using two approaches, namely, direct signal-to-interference-plus-noise-ratio relaxation via sequential parametric convex approximation and mean squared error reformulation. In the centralized approach, the common controller is assumed to have the complete channel state information of all base station-user links in order to design precoders for all base station. We adopted successive convex approximation technique to handle the nonconvex nature of the original problem by solving a sequence of convex subproblems.

We proposed two related distributed precoder design algorithms, wherein the precoders are designed at each base station with the local channel state information knowledge by exchanging coupling variables among the coordinating base stations. In addition, we also considered the problem of imposing certain minimum quality of service requirements in the form of guaranteed rate to the users in the system. For the proposed algorithms, the interference exchange to update the precoders at each base station is carried out via either backhaul or over-the-air.

In distributed designs, we initially designed transmit precoders at each base station in a decentralized manner by employing alternating directions method of multipliers (ADMM) technique. It is carried out by relaxing the inter-cell interference as an optimization variable by including it in each base station objective. The precoders are designed in each base station (BS) by exchanging the interference information via backhaul which interconnects two base stations. The direct relaxation of signal-to-interference-plus-noise-ratio via sequential parametric convex approximation formulation was formulated for both with and without guaranteed quality of service rate constraints.

We further investigated an alternative precoder design by solving the Karush-Kuhn-Tucker expressions in each successive convex approximation step across the coordinating base station. Using the centralized approaches we formulated a distributed precoder design by solving the respective Karush-Kuhn-Tucker expressions. In addition, we also discussed the distributed precoder design to provide guaranteed minimum rate to all users in the system. The reason for considering the Karush-Kuhn-Tucker based distributed precoder design is that the alternating directions method of multipliers requires several steps of iterations in each successive convex approximation step. Moreover, the complexity involved in backhaul exchange increases with the system size. Due to the above said reasons, the Karush-Kuhn-Tucker based solutions are more preferable for practical implementation owing to the closed form updates in each successive convex approximation iteration.

In the Karush-Kuhn-Tucker approach, we begin with the sequential parametric convex approximation based centralized design was extended to the Karush-Kuhn-Tucker based distributed for both with and without quality of service rate constraint. Since we considered a single antenna receive, the interference coupling variable is exchanged via backhaul. Similarly the mean squared error design was also extended to the Karush-Kuhn-Tucker based approach for both with and without quality of service user rate constraint, over-the-air based precoder training procedure is considered as an viable option for practical implementation.

Numerical analysis for the alternating directions method of multipliers and the Karush-Kuhn-Tucker based algorithms without quality of service constraints suggested that the convergence is monotonic. However, algorithm involving the Karush-Kuhn-Tucker expressions is shown to converge quickly when compared to alternating directions method of multipliers approach. Similarly, for minimum quality of service rate methods, we observed that all the users attained certain rate above the guaranteed minimum rate due to the available transmit power budget.

As future work, we consider extending the precoder design for the following scenarios. At first, we extend the precoder design with the guaranteed rate requirement for primary users in a cognitive radio framework. In this case, the secondary users may take any rate without affecting the primary users quality of service requirements. Secondly, we can consider extending the precoder design over a time correlated fading scenario, wherein we perform the partial over-the-air and backhaul based exchanges to design precoders. As a final extension, we can consider selecting a subset of user for precoder design to reduce the number of iterations required for the precoder convergence.

# 8. REFERENCES


[1] Goldsmith A. (2005) Wireless communications. Cambridge university press.

[2] Biglieri E., Calderbank R., Constantinides A., Goldsmith A., Paulraj A. & Poor H.V. (2007) MIMO wireless communications. Cambridge university press.

[3] Rappaport T.S. et al. (1996) Wireless communications: principles and practice, vol. 2. prentice hall PTR New Jersey.

[4] Tse D. & Viswanath P. (2005) Fundamentals of Wireless Communication. Wiley Series in Telecommunications, Cambridge University Press.

[5] Cadambe V.R. & Jafar S.A. (2008) Interference Alignment and Degrees of Freedom of the K-User Interference Channel. IEEE Transactions on Information Theory 54, pp. 3425–3441.

[6] Costa M.H. (1983) Writing on dirty paper (corresp.). IEEE Transactions on Information Theory 29, pp. 439–441.

[7] Weingarten H., Steinberg Y. & Shamai S. (2004) The capacity region of the gaussian mimo broadcast channel. In: IEEE International Symposium on Information Theory, pp. 174–174.

[8] Luo Z.Q. & Zhang S. (2008) Dynamic spectrum management: Complexity and duality. IEEE Journal of Selected Topics in Signal Processing 2, pp. 57–73.

[9] Venturino L., Prasad N. & Wang X. (2010) Coordinated linear beamforming in downlink multi-cell wireless networks. IEEE Transactions on Wireless Communications 9, pp. 1451–1461.

[10] Ng C.T. & Huang H. (2010) Linear precoding in cooperative mimo cellular networks with limited coordination clusters. IEEE Journal on Selected Areas in Communications 28, pp. 1446–1454.

[11] Christensen S.S., Agarwal R., Carvalho E. & Cioffi J. (2008) Weighted Sum-Rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design. IEEE Transactions on Wireless Communications 7, pp. 4792–4799.

[12] Shi Q., Razaviyayn M., Luo Z.Q. & He C. (2011) An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel. IEEE Transactions on Signal Processing 59, pp. 4331–4340.

[13] Joshi S.K., Weeraddana P.C., Codreanu M. & Latva-Aho M. (2012) Weighted sum-rate maximization for miso downlink cellular networks via branch and bound. IEEE Transactions on Signal Processing 60, pp. 2090–2095.

[14] Tran L.N., Hanif M., Tölli A. & Juntti M. (2012) Fast Converging Algorithm for Weighted Sum Rate Maximization in Multicell MISO Downlink. IEEE Signal Processing Letters 19, pp. 872–875.

[15] Kaleva J., Tölli A. & Juntti M. (2013) Decentralized Beamforming for Weighted Sum Rate Maximization with Rate Constraints. In: 24th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops), IEEE, pp. 220–224.

[16] Kaleva J., Tölli A. & Juntti M. (2013) Primal Decomposition based Decentralized Weighted Sum Rate Maximization with QoS Constraints for Interfering Broadcast Channel. In: IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), IEEE, pp. 16–20.

[17] Kaleva J., Tolli A. & Juntti M. (2015) Rate constrained decentralized beamforming for mimo interfering broadcast channel. In: Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on, IEEE, pp. 376–380.

[18] Bjornson E., Zheng G., Bengtsson M. & Ottersten B. (2012) Robust monotonic optimization framework for multicell miso systems. IEEE Transactions on Signal Processing 60, pp. 2508–2523.

[19] Liu L., Zhang R. & Chua K.C. (2012) Achieving global optimality for weighted sum-rate maximization in the k-user gaussian interference channel with multiple antennas. IEEE Transactions on Wireless Communications 11, pp. 1933–1945.

[20] Kaleva J., Tölli A. & Juntti M. (2012) Weighted Sum Rate Maximization for Interfering Broadcast Channel via Successive Convex Approximation. In: Global Communications Conference, IEEE, pp. 3838–3843.

[21] Etkin R.H., Tse D.N. & Wang H. (2008) Gaussian interference channel capacity to within one bit. IEEE Transactions on Information Theory 54, pp. 5534–5562.

[22] Annapureddy V.S. & Veeravalli V.V. (2009) Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region. IEEE Transactions on Information Theory 55, pp. 3032–3050.

[23] Shannon C.E. et al. (1961) Two-way communication channels. In: 4th Berkeley Symposium on Math, Statistics, and Probability, vol. 1, Citeseer, vol. 1, pp. 611–644.

[24] Shannon C.E. (2001) A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5, pp. 3–55.

[25] Luo Z.Q. & Zhang S. (2008) Dynamic Spectrum Management: Complexity and Duality. IEEE Journal of Selected Topics in Signal Processing 2, pp. 57–73.

[26] Boyd S. & Vandenberghe L. (2004) Convex optimization. Cambridge university press.

[27] Scutari G., Palomar D.P., Facchinei F. & Pang J.S. (2010) Convex optimization, game theory, and variational inequality theory. IEEE Signal Processing Magazine 27, pp. 35–49.

[28] Bertsekas D.P., Nedi A., Ozdaglar A.E. et al. (2003) Convex analysis and optimization .

[29] Sesia S., Toufik I. & Baker M. (2009) LTE: the UMTS long term evolution. Wiley Online Library.

[30] Srinivasa S., Jafar S. et al. (2007) The optimality of transmit beamforming: A unified view. IEEE Transactions on Information Theory 53, pp. 1558–1564.

[31] Jorswieck E. & Boche H. (2004) Channel capacity and capacity-range of beamforming in mimo wireless systems under correlated fading with covariance feedback. IEEE Transactions on Wireless Communications 3, pp. 1543–1553.

[32] Beck A., Ben-Tal A. & Tetruashvili L. (2010) A sequential parametric convex approximation method with applications to nonconvex truss topology design problems. Journal of Global Optimization 47, pp. 29–51.

[33] Marks B.R. & Wright G.P. (1978) A General Inner Approximation Algorithm for Nonconvex Mathematical Programs. Operations Research 26, pp. 681–683.

[34] Boyd S., Parikh N., Chu E., Peleato B. & Eckstein J. (2011) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends® in Machine Learning 3, pp. 1–122.

[35] Tölli A., Pennanen H. & Komulainen P. (2011) Decentralized Minimum Power Multi-Cell Beamforming with Limited Backhaul Signaling. IEEE Transactions on Wireless Communications 10, pp. 570–580.

[36] Shen C., Chang T.H., Wang K.Y., Qiu Z. & Chi C.Y. (2012) Distributed Robust Multicell Coordinated Beamforming With Imperfect CSI: An ADMM Approach. IEEE Transactions on Signal Processing 60, pp. 2988–3003.

[37] Komulainen P., Tölli A. & Juntti M. (2013) Effective CSI Signaling and Decentralized Beam Coordination in TDD Multi-Cell MIMO Systems. IEEE Transactions on Signal Processing 61, pp. 2204–2218.

[38] Bertsekas D.P. (1999) Nonlinear Programming. Athena Scientific, 2nd ed.

[39] Zangwill W. (1969) Nonlinear Programming: A Unified Approach. Prentice-Hall International Series in Management, Prentice-Hall.

[40] Meyer R. (1976) Sufficient Conditions for the Convergence of Monotonic Mathematical Programming Algorithms. Journal of Computer and System Sciences 12, pp. 108–121.

[41] Bertsekas D.P. & Tsitsiklis J.N. (1989) Parallel and Distributed Computation: Numerical Methods. Prentice Hall Englewood Cliffs, NJ.

# 9. APPENDICES

## 9.1. Convergence Proof for Centralized Algorithms

The centralized problem formulations in (13) and (26) are nonconvex, therefore, we adopt successive convex approximation (SCA) technique to solve the problem in an iterative manner by solving a sequence of convex subproblems. The centralized algorithms outlined in Algorithm 1 generates a sequence of objective values and a corresponding sequence of beamformer iterates. The convergence of the iterate sequence follows Theorem 1.

**Theorem 1.** *Every limit point of the sequence generated by above algorithms is a stationary point.*

*Proof.* In order to prove the above statement about the sequence of iterates generated by the centralized algorithm, let us consider a following generalized problem structure as

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) && \text{(62a)} \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m && \text{(62b)} \\ & h_j(\mathbf{x}) = 0, j = 1, \ldots, p. && \text{(62c)} \end{aligned}$$

where $\mathbf{x}$ is a vector formed by stacking all the optimization variables of nonconvex problems (13) and (26), respectively. Without loss of generality, let us proceed further with the following assumptions. Let the inequality constraints $g_i(\mathbf{x})$, $\forall i \in \{1, \ldots, n\}$ are all differentiable and convex functions, $g_i(\mathbf{x})$, $\forall i \in \{n+1, \ldots, m\}$ are all differentiable and possibly nonconvex, and the linearity constraints are all affine.

In order to solve the above nonconvex problem (62), we refer to SCA method discussed in [33, 32]. The problem (62) is solved by approximating the nonconvex set by a convex subset and solved iteratively by updating the convex subset in each iteration. As shown in [33], the inner SCA algorithm for the minimization problem can be done in the following steps,

- Set a starting point for the variable and constraint $\mathbf{x}^0 \in F$ and set $h^0 = g_0(\mathbf{x}^0)$. Let $A^0 = \{\mathbf{x} | h^0 = g_0(\mathbf{x}) \text{ and } \mathbf{x} \in F\}$, where $F$ can be defined as the feasible region.

- In the $k^{th}$ iteration replace the constraint $g_i(\mathbf{x}) \leq 0,\ i = (n+1), \ldots, m$, by $\bar{g}_i(\mathbf{x}, \mathbf{x}^k) \leq 0$, where $\bar{g}_i(\mathbf{x}, \mathbf{x}^k)$ is a differentiable convex function and $\mathbf{x}^k \in \mathcal{A}^{k-1}$. Each function $\bar{g}_i(\mathbf{x}, \mathbf{x}^k)$ must have the following properties

$$g_i(\mathbf{x}) \leq \bar{g}_i(\mathbf{x}, \mathbf{x}^k) \quad \forall \mathbf{x} \in F^k \tag{63a}$$
$$g_i(\mathbf{x}) = \bar{g}_i(\mathbf{x}^k, \mathbf{x}^k) \tag{63b}$$
$$\nabla g_i(\mathbf{x}^k) = \nabla \bar{g}_i(\mathbf{x}^k, \mathbf{x}^k),\ \forall j = 1, \ldots, n \tag{63c}$$

The feasible region $F^k = \{\mathbf{x} \,|\, g_i(\mathbf{x}) \leq 0,\ \forall i = 1, \ldots, n,$ and $\bar{g}_i(x, x^k) \leq 0,\ \forall i = n+1, \ldots, m\}$ should satisfy slaters constraint qualification condition for convex programs.

- Solve the approximation convex program

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & g_0(\mathbf{x}) & \text{(64a)} \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0, i = 1, \ldots, n & \text{(64b)} \\ & \bar{g}_i(\mathbf{x}, \mathbf{x}^k) \leq 0, i = n+1, \ldots, m. & \text{(64c)} \end{aligned}$$

Let $h^k = \min\{g_0(\mathbf{x}) | \mathbf{x} \in F^k\}$.

- If $h^k = h^{(k-1)}$, then $\mathbf{x}^k$ is a Karush-Kuhn-Tucker (KKT) solution for the minimization problem. Otherwise, let $a^k = \{\mathbf{x} \,|\, h^k = g_0(\mathbf{x})$ and $\mathbf{x} \in F^k\}$ and return to step 1.

Note that the monotonic decrease of sequence $\{f(\mathbf{x}^k)\}$ is guaranteed by using the following argument. Since each subproblem (64) includes the solution from previous iteration, *i.e.*, $\mathbf{x}^{k-1} \in \mathcal{F}^k$ (see (63)), $f(\mathbf{x}^k) \leq f(\mathbf{x}^{k-1})$. Therefore, monotonic decrease of the objective sequence is guaranteed. Now, by using [38 Prop. A.3], we can show that $\{f(\mathbf{x}^k)\}$ is bounded and monotonically decreasing, therefore, it converges as $k \to \infty$.

Now by following [32, 39, 40], we can show that the sequence of iterates converges to a set of limit points, since in each SCA step, the problem (64) is convex, and therefore can have multiple minimizers. Due to this, we can have oscillatory behavior in the sequence of iterates, which may lead to lack of convergence. However, we note that as $\lim_{k\to\infty} \mathbf{x}^k \to \mathcal{F}^*$, where $\mathcal{F}^*$ is the set of all limit points. Therefore, by using [40 Theorem 3.1], we can show that $\{\mathbf{x}^k\}$ converges to a continuum of limit points, or every limit point is a stationary point. The stationarity of limit points can be easily established by considering that $\mathbf{x}^k$ is a solution of (64), therefore, as $k \to \infty$, by using [32], we can show that every point in $\mathcal{F}^*$ is a stationary point of (62).

$\square$

## 9.2. Convergence Proof for Distributed Algorithms

The convergence of ADMM based distributed precoder design can be guaranteed by the following.

- In each SCA step, the subproblem considered for ADMM is convex
- The ADMM updates are performed until convergence.

Then, the convergence of ADMM follows the discussions in [34].

Similarly, for the KKT based distributed precoder designs, if the number of iterations is significantly large or iterated as $i \to \infty$, convergence is guaranteed by following Theorem 1. It is due to the fact that KKT based design solves the convex subproblem in each SCA iteration using system of equations, which is similar to the respective centralized algorithms.

On the contrary, in case of KKT based precoder designs with quality of service (QoS) constraints, convergence cannot be ensured directly. However, if the step size parameter $\rho$ used in the subgradient update of the dual variable $d_k$ is diminishing in each step, *i.e.*, $\sum_{k=0}^{\infty} \rho^{(i)} = \infty$, and $\lim_{i\to\infty} \rho^{(i)} \to 0$, then the convergence can be ensured by following [41 Prop. 8.2.6].

## 9.3. KKT Conditions for SPCA method

In order to obtain an iterative precoder design algorithm, the KKT expressions of problem (40) and (49) are required, which is found by differentiating the associated Lagrangian function with respect to each of the optimization and dual variables.

### *9.3.1. SPCA without QoS Requirements*

Upon differentiating and grouping the associated variables of (40), we obtain the following relations as

$$\nabla_{\gamma_k} \quad : \quad -\frac{1}{1+\gamma_k} + \frac{a_k}{2\phi_k^{(i)}} = 0 \tag{65a}$$

$$\nabla_{\beta_k} \quad : \quad -\frac{a_k \phi_k^{(i)}}{2} - e_k = 0 \tag{65b}$$

$$\nabla_{\mathbf{w}_k} \quad : \quad 2\,\mathbf{w}_k \left( \sum_{i \neq k} e_i \mathbf{h}_{b_k,i}^H \mathbf{h}_{b_k,i} + c_b \mathbf{I}_{N_T} \right) = a_k \mathbf{h}_{b_k,k}^H. \tag{65c}$$

In addition to (65), it also includes the complementary slackness conditions as

$$a_k \left( \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}} - |\mathbf{h}_{b_k,k}\mathbf{w}_k| \right) = 0 \tag{66a}$$

$$e_k \left( \sigma^2 + \sum_{i=1, i\neq k}^{K} \|\mathbf{h}_{b_i,k}\mathbf{w}_k\|^2 - \beta_k \right) = 0 \tag{66b}$$

$$c_b \left( \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right) = 0. \tag{66c}$$

Note that to obtain a tractable solution, we consider the following assumptions on dual variables as $a_k \neq 0$ and $e_k \neq 0$, thereby making the respective constraints to be tight or satisfies with equality.

#### *9.3.2. SPCA with QoS Requirements*

Upon differentiating and grouping the associated variables of (40), we obtain the following relations as

$$\nabla_{\gamma_k} \quad : \quad -\frac{1}{1+\gamma_k} - d_k + \frac{a_k}{2\phi_k^{(i)}} = 0 \tag{67a}$$

$$\nabla_{\beta_k} \quad : \quad -\frac{a_k\phi_k^{(i)}}{2} - e_k = 0 \tag{67b}$$

$$\nabla_{\mathbf{w}_k} \quad : \quad 2\,\mathbf{w}_k \left( \sum_{i\neq k} e_i \mathbf{h}_{b_k,i}^H \mathbf{h}_{b_k,i} + c_b \mathbf{I}_{N_T} \right) = a_k \mathbf{h}_{b_k,k}^H. \tag{67c}$$

In addition to (67), it also includes the complementary slackness conditions as

$$a_k \left( \gamma_k \frac{\phi_k^{(i)}}{2} + \beta_k \frac{1}{2\phi_k^{(i)}} - |\mathbf{h}_{b_k,k}\mathbf{w}_k| \right) = 0 \tag{68a}$$

$$e_k \left( \sigma^2 + \sum_{i=1, i\neq k}^{K} \|\mathbf{h}_{b_i,k}\mathbf{w}_k\|^2 - \beta_k \right) = 0 \tag{68b}$$

$$c_b \left( \sum_{k\in\mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right) = 0 \tag{68c}$$

$$d_k \left( R_k - \log(1+\gamma_k) \right) = 0. \tag{68d}$$

Note that to obtain a tractable solution, we consider the following assumptions on dual variables as $a_k \neq 0$ and $e_k \neq 0$, thereby making the respective constraints to be tight or satisfies with equality. However, (68d) cannot be assumed with equality, therefore, we need to adopt subgradient approach to find the optimal dual variable $d_k$.

## 9.4. KKT Conditions for mean squared error (MSE) Reformulation

### *9.4.1. MSE without QoS Requirements*

By considering the Lagrangian function in (51), we obtain the following expression by taking the gradient of (51) with respect to each of the associated optimization variables as

$$\nabla_{\epsilon_k} \quad : \quad \frac{1}{\bar{\epsilon}_k} - a_k = 0 \tag{69a}$$

$$\nabla_{\mathbf{w}_k} \quad : \quad \mathbf{w}_k \left( a_k \sum_{i \neq k} \mathbf{h}_{b_k,i}^H \mathbf{u}_i^H \mathbf{u}_i \mathbf{h}_{b_k,i} + c_b \mathbf{I}_{N_T} \right) = a_k \mathbf{u}_k^H \mathbf{h}_{b_k,k}. \tag{69b}$$

In addition to (69), it also includes the complementary slackness conditions as

$$a_k \left( |1 - \mathbf{u}_k^H \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}}_b} \|\mathbf{u}_i^H \mathbf{h}_{b_k,i} \mathbf{w}_i\|^2 + \bar{N}_0 - \epsilon_k \right) = 0 \tag{70a}$$

$$c_b \left( \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right) = 0. \tag{70b}$$

In addition to (70), the primal constraints given in (50) are also considered while designing an iterative approach to solve the involved variables.

### *9.4.2. MSE with QoS Requirements*

By considering the Lagrangian function in (59), we obtain the following expression by taking the gradient of (59) with respect to each of the associated optimization variables as

$$\nabla_{\epsilon_k} \quad : \quad \frac{1}{\bar{\epsilon}_k} - \frac{d_k}{\log \bar{\epsilon}_k} - a_k = 0 \tag{71a}$$

$$\nabla_{\mathbf{w}_k} \quad : \quad \mathbf{w}_k \left( a_k \sum_{i \neq k} \mathbf{h}_{b_k,i}^H \mathbf{u}_i^H \mathbf{u}_i \mathbf{h}_{b_k,i} + c_b \mathbf{I}_{N_T} \right) = a_k \mathbf{u}_k^H \mathbf{h}_{b_k,k} \tag{71b}$$

(71c)

In addition to (71), it also includes the complementary slackness conditions as

$$a_k \left( |1 - \mathbf{u}_k^H \mathbf{h}_{b_k,k} \mathbf{w}_k|^2 + \sum_{i \in \bar{\mathcal{U}_b}} \|\mathbf{u}_i^H \mathbf{h}_{b_k,i} \mathbf{w}_i\|^2 + \bar{N}_0 - \epsilon_k \right) = 0 \tag{72a}$$

$$c_b \left( \sum_{k \in \mathcal{U}_b} \|\mathbf{w}_k\|_2^2 - P_b \right) = 0 \tag{72b}$$

$$d_k \left( \log(\bar{\epsilon}_k) + \frac{\epsilon_k - \bar{\epsilon}_k}{\bar{\epsilon}_k} + R_k \right) \tag{72c}$$

In addition to (72), the primal constraints given in (58) are also considered while designing an iterative approach to solve the involved variables.