Spring 4-1-2015

Optimization and Performance of MIMO B-MAC Interference Networks

Xing Li
University of Colorado Boulder, lewiexing@gmail.com

Follow this and additional works at: https://scholar.colorado.edu/ecen_gradetds

Part of the Systems and Communications Commons, and the Theory and Algorithms Commons

Recommended Citation
https://scholar.colorado.edu/ecen_gradetds/118

This Dissertation is brought to you for free and open access by Electrical, Computer & Energy Engineering at CU Scholar. It has been accepted for inclusion in Electrical, Computer & Energy Engineering Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Optimization and Performance of MIMO B-MAC Interference Networks

by

Xing Li

B.S., Peking University, 2005
M.S., Peking University, 2008

A thesis submitted to
the Faculty of the Graduate School of the
University of Colorado at Boulder
in fulfillment of the requirements for the degree of
Doctor of Philosophy
Department of Electrical, Computer and Energy Engineering

2015
This thesis entitled:
Optimization and Performance
of MIMO B-MAC Interference Networks
written by Xing Li
has been approved for the Department of Electrical, Computer and Energy Engineering

Prof. Youjian Liu

Prof. Lijun Chen

Prof. Peter Mathys

Prof. Behrouz Touri

Prof. Mahesh K. Varanasi

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
Li, Xing (Ph.D., Electrical, Computer and Energy Engineering)

Optimization and Performance of MIMO B-MAC Interference Networks

Thesis directed by Prof. Youjian Liu

Abstract

This thesis studies optimization and performance of the MIMO B-MAC interference networks which includes broadcast channel (BC), multiaccess channel (MAC), interference channels, X networks, and many practical wireless networks as special cases. A 3D channel model for distributed MIMO system is set up, based on which the antenna correlations can be characterized in analytic form. We propose a new algorithm, named Dual Link algorithm, for the classic problem of weighted sum-rate maximization for MIMO multiaccess channels (MAC), broadcast channels (BC), and general MIMO interference channels with Gaussian input and a total power constraint. For MIMO MAC/BC, the algorithm finds optimal signals to achieve the capacity region boundary. For interference channels with Gaussian input assumption, two of the previous state-of-the-art algorithms are the WMMSE algorithm and the polite water-filling (PWF) algorithm. The WMMSE algorithm is provably convergent, while the PWF algorithm takes the advantage of the optimal transmit signal structure and converges the fastest in most situations but is not guaranteed to converge in all situations. It is highly desirable to design an algorithm that has the advantages of both algorithms. The proposed dual link algorithm is such an algorithm. Its fast and guaranteed convergence is important to distributed implementation and time varying channels. In addition, the technique and a scaling invariance property used in the convergence proof may find applications in other non-convex problems in communication networks. The dual link algorithm is also further modified to fit practical applications. Since the centralized algorithm is not scalable as network size increases, the optimization algorithm needs to be working in a mainly distributed fashion to avoid having huge signaling overheads. We’ve proposed the distributed dual link algorithm for time division duplex (TDD) interference networks. It replaces direct and cross channel information feedbacks with iterations of forward and reverse link pilots.
training whose complexity grows linearly with the number of users in the network. By totally avoiding channel state knowledge feedback, the distributed dual link algorithm has significant lower signaling overhead compared to the traditional methods, especially in networks with large number of interfering users. However, the real TDD channels are not reciprocal because the transmit and receive RF chains are different in a transceiver. To solve this issue, we proposed a simple method of channel calibration to restore TDD channel reciprocity for MIMO interference networks that is essential to the distributed implementation of the Dual Link algorithm and other algorithms that require reciprocity. On the other hand, the channel knowledge is generally imperfect in a realistic scenario. To study its impact, we introduce a simple channel uncertainty model that characterizes different levels of channel uncertainty. Based on this model, the ergodic weighted sum-rate maximization problem is studied. The ergodic dual link algorithm is proposed to analytically solve the optimization problem. We also propose the robust dual link algorithm which is sub-optimal but has good performance under all channel uncertainty levels and is suitable for online distributed implementation. Finally, we study the physical layer transmission and reception schemes in a cellular system where the basestations are equipped with infinitely many antennas. It is shown that the uplink Signal to Interference Ratio (SIR) and the downlink SIR corresponding to a given basestation (BS) - mobile station (MS) pair are identical.
To my wife and my parents.
Acknowledgment

Special thanks to Prof. Youjian Liu for the supervision of my thesis.

I greatly appreciate suggestions and insights from Prof. Lijun Chen and Prof. Behrouz Touri.

I dearly thank Prof. Peter Mathys and Prof. Mahesh K. Varanasi to serve on my dissertation committee.

I also thank the following individuals, Seungil You, Jun Xiao, Prasanna Madhusudhanan, Yao Wang, Zhe Feng, Reza Aliehmmati, Farhad Pouramali-Anaraki, Hanchao Qi for discussions and support.
Contents

Abstract iii
Acknowledgment vi

List of Figures ix

Chapter 1. Introduction 1

Chapter 2. 3D Channel Model for Distributed MIMO Satellite Systems 5
  2.1. Introduction 5
  2.2. Three-Dimensional Channel Model 7
  2.3. Rate and Correlation 13
  2.4. Performance Evaluation 15
  2.5. Conclusion 16

Chapter 3. The Dual Link Algorithm 18
  3.1. Introduction 18
  3.2. Preliminaries 20
  3.3. The Dual Link Algorithm 26
  3.4. Extension for Environment with Colored Noise 36
  3.5. Simulation Results 38
  3.6. Complexity Analysis 42
  3.7. Conclusion 44

Chapter 4. Distributed Implementation of the Dual Link Algorithm 46
  4.1. Introduction 46
  4.2. Distributed Dual Link Algorithm and Distributed Channel Estimation 46
4.3. Conclusion 49

Chapter 5. Antenna Calibration for TDD Channel Reciprocity 51
  5.1. Introduction 51
  5.2. Internal Calibration 51
  5.3. Internal Channel Calibration for Distributed Dual Link Algorithm 52
  5.4. Conclusion 54

Chapter 6. Robust Algorithm Design for Channel Uncertainty 55
  6.1. Introduction 55
  6.2. Problem Formulation 56
  6.3. Ergodic WMMSE Algorithm 57
  6.4. Ergodic Dual Link Algorithm 59
  6.5. Robust Dual Link Algorithm 67
  6.6. Simulation Results 68
  6.7. Conclusion 71

Chapter 7. Uplink and Downlink Analysis of Massive MIMO Cellular Systems 74
  7.1. Introduction 74
  7.2. System Model 74
  7.3. Uplink and Downlink Analysis 76
  7.4. Conclusion 79

Chapter 8. Conclusion 80

Bibliography 82
List of Figures

2.1 The 3-D channel model for the channel between satellite antennas and ground station antennas with the projected distance $A_{ij}$ 8

2.1 The information rate of SIMO(1×3), MIMO(2×3), MIMO(3×3) channels versus the projected distance $A_{12}$ between the 1st and 2nd satellites 16

3.1 An example of B-MAC network. The solid lines represent data links and the dash lines represent interference. 21

3.1 The monotonic convergence of the forward and reverse link weighted sum-rates of the Dual Link algorithm with $P_T = 100$ and $g_{l,k} = 0$dB, $\forall l, k$. 39

3.2 PWF algorithm vs. WMMSE algorithm vs. dual link algorithm under weak interference with $P_T = 100$, $g_{l,l} = 0$dB and $g_{l,k} = -10$dB for $l \neq k$. 40

3.3 PWF algorithm vs. WMMSE algorithm vs. dual link algorithm under moderate interference with $P_T = 100$ and $g_{l,k} = 0$dB, $\forall l, k$. 41

3.4 PWF algorithm vs. WMMSE algorithm vs. dual link algorithm under strong interference with $P_T = 100$, $g_{l,l} = 0$dB and $g_{l,k} = 10$dB for $l \neq k$. 42

3.5 An example that PWF, WMMSE, and dual link algorithms converge to different stationary points. $P_T = 100$, $g_{l,l} = 0$dB and $g_{l,k} = 10$dB for $l \neq k$. 43

3.6 Dual Link algorithm vs. Iterative water-filling on MAC and BC channels 44

6.1 Ergodic Dual Link Algorithm vs. Ergodic WMMSE Algorithm 69

6.2 Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 0$ 70

6.3 Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 0.5$ 71

6.4 Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 0.9$ 72
6.5 Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 1$  73

7.1 User distribution and pilots assignments  75
CHAPTER 1

Introduction

As the demand of cell phone data service keeps increasing, one of the most effective solutions is deploying more base stations or access points in the high-user-density area. However, the path loss versus distance curve is flatter at shorter distance than that of longer distance. As a result, the inter-cell interference becomes significant as the cell size/coverage area shrinks. Therefore, joint transmit signal design that manages interference is a key technology for the next generation wireless communication systems. Traditionally, interference between different users is handled by enforcing certain orthogonality between them, which can be frequency, time, space or even code. In the sense of the optimization problem that allocates all the available resources to these users in order to maximize the overall performance (e.g. the weighted sum rate), introducing orthogonality means adding more constraints to the original problem and may end up with a sub-optimal solution. Instead of degrading the problem for simplicity, this thesis is aimed to find a reliable way to achieve the optimal solution of the general weighted sum rate maximization problem with only power constraint.

A general interference network named MIMO B-MAC network is considered in our work. In such a network, a transmitter may send independent data to different receivers, like in BC, and a receiver may receive independent data from different transmitters, like in MAC. To have a better understanding of the MIMO channels, we first set up a 3D channel model for distributed MIMO satellite system to study the correlation characteristics between antenna elements. The numerous geometric parameters of the locations of transmit antennas, receive antennas, and scatterers are abstracted into a single key parameter, projected distance, which determines the spatial correlation. Information rate results are provided to demonstrate that the distributed MIMO system can have significant rate gain over SIMO system for most situations.
We design a new algorithm that jointly optimizes the covariance matrices of transmit signals of multiple transmitters in order to maximize the weighted sum-rate of the data links for the MIMO B-MAC networks, assuming Gaussian transmit signal and the availability of all channel state information. The problem is non-convex, and various algorithms have been proposed for various cases, e.g., [5–7, 9, 14, 23, 24, 32–36]. Among the state-of-the-art algorithms, we have proposed the polite water-filling (PWF) algorithms [14]. Because it takes advantage of the optimal transmit signal structure for an achievable rate region, the polite water-filling structure, the PWF algorithm has the lowest complexity and the fastest convergence when it converges. However, in some strong interference cases, it has small oscillation. Another state-of-the-art algorithm is the WMMSE algorithm in [24]. It was proposed for beamforming matrix design for the MIMO interfering broadcast channels but could be readily applied to the more general B-MAC networks and input covariance matrix design. It transforms the weighted sum-rate maximization into an equivalent weighted sum-MSE cost minimization problem, which has three sets of variables and is convex when any two variable sets are fixed. With the block coordinate descent technique, the WMMSE algorithm is guaranteed to converge to a stationary point, though the convergence is observed in simulations to be relatively slower than the PWF algorithm.

It is thus highly desirable to have an algorithm with the advantages of both PWF and WMMSE algorithms, i.e., fast convergence by taking advantage of the optimal transmit signal structure and provable convergence for the general interference network. The main contribution of our work is to present such an algorithm, named dual link algorithm. It exploits the forward-reverse link rate duality in a new way. We also show that the dual link algorithm can be easily modified to deal with the weighted sum-rate maximization in an environment with colored noise. Numerical experiments demonstrate that the dual link algorithm is almost as fast as the PWF algorithm and can be a few iterations or more than ten iterations faster than the WMMSE algorithm, depending on the desired accuracy with respect to the local optimum. Note that being faster even by a couple iterations will be critical in distributed implementation in time division duplex (TDD) networks with time
varying channels, where the overhead of each iteration costs significant signaling resources between the transmitters and the receivers. The faster the convergence is, the faster channel variations can be accommodated by the algorithm.

Indeed, the dual link algorithm is highly scalable and suitable for distributed implementation because, for each data link, only its own channel state and the aggregated interference plus noise covariance needs to be estimated no matter how many interferers are there. By taking advantage of the reciprocity of TDD channels, we design the distributed dual link algorithm and a unique distributed channel estimation method that can avoid the massive signaling overheads caused by feeding back channel state information (CSI) in a MIMO interference network.

For real application of the distributed dual link algorithm, some key issues are also considered, including the differences in the RF-chains and the channel uncertainty. We propose a simple and effective antenna calibration method for TDD MIMO interference networks to maintain the reciprocity. And we design robust versions of the distributed algorithm to optimize the average performance.

We also study the physical layer transmission and reception schemes in a cellular system where the BSs are equipped with infinitely many antennas, the pilot sequences can be non-orthogonal, and number of users in the cell can be arbitrary. It is shown that the uplink SIR and the downlink SIR corresponding to a given BS-MS pair are identical.

The rest of the thesis is organized as follows. Chapter 2 presents the 3D channel model for distributed MIMO satellite systems and close form solution of antenna correlations. Chapter 3 proposes the dual link algorithm, establishes its monotonic convergence, and provides numerical examples as well as complexity analysis. Chapter 4 proposes a distributed version of the dual link algorithm and a corresponding distributed channel estimation method. In Chapter 5, an antenna calibration method is presented to maintain the TDD channel reciprocity property required by distributed implementations. Chapter 6 introduces a simple channel uncertainty model and proposes the ergodic dual link algorithm and robust dual link algorithm to improve the average performance under channel uncertainty. Chapter 7 studies
the physical layer transmission and reception schemes in a massive MIMO cellular system and presents the uplink SIR and the downlink SIR duality. Chapter 8 concludes.
CHAPTER 2

3D Channel Model for Distributed MIMO Satellite Systems

2.1. Introduction

Multi-antenna at both transmitter and receiver in a multiple-input multiple-output (MIMO) system can significantly increase spectral efficiency [26]. Given the success of MIMO techniques in land mobile systems, it is expected that space/air communication systems will take advantage of it in the future. Experimental results show that significant improvement in rate can be obtained in 2x2 MIMO system consists of two emulated co-cluster low elevation satellites and a vehicle with two polarized antennas compared with single satellite single-input single-output (SISO) system [10]. In [18], compact MIMO antenna array configurations in conjunction with satellite diversity techniques is employed. The simulation results indicates that although the rate is degraded by correlation and mutual coupling, significant rate gain can still be achieved. In [21] an approach is presented for the construction of a rate-optimized MIMO line of sight (LOS) satellite channel between two geostationary satellites and multiple ground station antennas.

In this chapter, we consider distributed MIMO satellite systems, where through high speed (optical) links between satellites, antennas on different satellites cooperatively form a transmit/receive antenna array. Note that when the transmit antennas are at different directions, MIMO systems have rate gain even without rich scattering. This is different from the traditional systems where the transmit antennas are at the same direction and rich scattering is needed for MIMO rate gain. To evaluate whether the performance gain over the traditional SISO or SIMO (single-input multiple-output) systems justifies the increased complexity, channel model has to be built to study the correlation of the channel matrix in order to calculate the rate. The channel model is desired to be, without losing the essence of reality, simple enough so that the antenna correlation functions can be found in analytic
form in order to gain insight into how they depends on the geometric parameters. The results in this paper can be applied to both satellites and air based communication systems such as High Altitude Platforms (HAPs).

Satellite channels and 2-D/3-D (two/three dimensional) fading channel models have been extensively studied. The SISO satellite channel model and measurement has been studied in [15] on both the line of sight (LOS) component and the multipath component. Most of the research on satellite channels has focused on LOS component [21, 29], which normally has far higher power. However, when the satellite is at low orbit or the ground station is in urban area, it is possible that multipath signal component is more important. Thus, it is necessary to build a channel model under which both LOS and multipath components can be studied. To simplify the scattering environment and examine the spatial fading correlation for land mobile systems, 2-D channel models with a 2-D scatterer ring have been proposed in [3, 25]. For satellite systems, 3-D channel models is needed. In existing 3-D MIMO channel models [1, 19, 20], it is often assumed that the distance between the transmitter and receiver is far larger than the antenna separations at both ends and thus, the transmit antennas are at the same direction. Also, existing 3-D MIMO channel models often have a large amount of geometric parameters for the locations of the transmit and receive antennas, which makes it difficult and complicated to evaluate the overall performance of the system, so the reduction of the model parameters becomes important. Another new characteristics of the distributed MIMO satellite systems is that Doppler shifts at multiple antennas cannot be simultaneously canceled. Since channel model is the focus of this paper, for the rate calculation, the bandwidth of the signal is assumed to be much larger than the Doppler shift. The rate and detection problems, where the bandwidth and Doppler shift are comparable, will be derived in future works.

The main contributions of the work in this chapter are as follows.

- A simple 3-D channel model is proposed to study the performance of distributed MIMO satellite system, in which both LOS and multipath signal components are considered. Both 2-D scatterer ring and 3-D scatterer ring are considered.
• The numerous geometric parameters of the locations of transmit antennas, receive
antennas, and scatterers are abstracted into a single key parameter, projected distance,
which determines the spatial correlation.

• Information rate results are provided to demonstrate that the distributed MIMO satel-
lite system can have significant rate gain over SIMO system for most situations.

The chapter is organized as follows. After presenting the assumptions used in this chap-
ter, we propose the 3-D channel model and provide the spatial and temporal correlation
function results in Section 2.2. In Section 2.3, the rate is analyzed and the condition for rate
degradation due to correlation is discussed. In Section 2.4, we show some typical numeric
results to demonstrate the rate gain of the distributed MIMO system. Section 2.5 concludes.

2.2. Three-Dimensional Channel Model

Due to channel reciprocity, we only describe the downlink case. All the satellite antennas
are referred to as transmit antennas, while the ground station antennas are referred to as
receive antennas. Both LOS and multipath components are taken into account. We consider
\( n_T \) transmit antennas, which may belong to different satellites, and \( n_R \) receive antennas,
which belong to one ground station in the center of a scatterer ring. The relative locations
of satellites, the ground station, and the scatterer ring are shown in Figure 2.1. Following
notations are used throughout the paper.

\( \beta_k \) : elevation angle of \( k \)th transmit antenna;
\( h_k \) : height of \( k \)th transmit antenna;
\( \varphi_k \) : direction angle of \( k \)th transmit antenna;
\( f_{D_k} \) : Doppler spread frequency caused by the movement of \( k \)th transmit antenna;
\( \rho_p \) : distance from the \( p \)th receive antenna to the center of the scatterer ring;
\( \theta_p \) : angle of the \( p \)th receive antenna;
\( \xi \) : angle of the ground station moving speed;
\( f_D \) : Doppler spread frequency caused by the movement of the ground station;
\( \lambda \) : carrier wavelength;
$N$: total number of the scatterers on the ring;
$a$: radius of the scatterer ring;
$\alpha_n$: angle of the $n$th scatterer $sc_n$ on the ring ($\alpha_n = \frac{2\pi n}{N}$);
$\phi_n$: initial phase shift of the $n$th scatterer.

### 2.2.1. Line of sight (LOS) component.

The LOS components of the channel coefficient between the $k$th transmit antenna and the $p$th receive antenna $c_{pk}^{(LOS)}(t)$ is given below. The distance between the antennas is

$$d_{pk} = \sqrt{h_k^2 + \left( \frac{h_k}{\tan \beta_k} \right)^2 + \rho_p^2 - 2 \frac{h_k}{\tan \beta_k} \rho_p \cos(\varphi_k - \theta_p)}$$

(2.2.1)$$
\approx \frac{h_k}{\sin \beta_k} - \rho_p \cos \beta_k \cos(\varphi_k - \theta_p),$$
where the approximation is based on the fact that the distance from the transmit antenna to the center of the scatterer ring is far greater than the distance from the receive antennas to the center of the scatterer ring \((\frac{h_k}{\sin \beta_k} \gg \rho_p)\). The corresponding channel matrix entry is

\[
C_{pk}^{(\text{LOS})}(t) = e^{j2\pi f_D t \cos(\xi - \phi_k) \cos \beta_k + \frac{d_{pk}}{\lambda}}.
\]

Here we assume that the received signal at each ground station antenna has the same unitary power strength and only differs in phase caused by different length of propagation path, because the receive antennas are relatively close to each other, and the attenuation of the signal from different satellites can be normalized into their transmit powers.

### 2.2.2. Multipath Signal Component.

In this subsection, we consider the multipath component and the corresponding channel matrix entries. Similar to the 2-D multipath fading channel model proposed in [3], a scatterer ring is placed around the ground station to model the multipath environment. The scatterers are assumed to be uniformly distributed on the ring, and each scatterer has an independent, uniformly distributed initial phase \(\phi_n\) over \([-\pi, \pi]\). Summing signals bounced from the scatters, the multipath component of the channel coefficient between the \(k\)th transmit antenna and the \(p\)th receive antenna is

\[
C_{pk}(t) = \frac{1}{\sqrt{N}} e^{j2\pi f_D t} \sum_{n=1}^{N} e^{j[2\pi f_D t \cos(\xi - \alpha_n) + \phi_k^{(\text{Sa})} + \phi_k^{(\text{GS})} + \phi_{pn}^{(\text{BS})}]}
\]

where \(N\) is the number of scatterers; \(\phi_k^{(\text{Sa})} = -2\pi s_{kn}/\lambda\); \(s_{kn}\) is the distance between \(k\)th transmit antenna and the \(n\)th scatterer; \(\phi_{pn}^{(\text{GS})} = -2\pi L_{pn}/\lambda\); \(L_{pn}\) is the distance between \(p\)th receive antenna and \(n\)th scatterer. When \(N\) approaches infinity, the distribution of \(C_{pk}(t)\) converges to Gaussian because of the uniformly distributed initial phase.

#### 2.2.2.1. Transmit Antenna Correlation (2-D Scatterer Ring Case).

In the following, we derive the correlation between two channel coefficients of two arbitrary satellite antennas and one ground station antenna. Without loss of generality, the two satellite antennas are indexed by subscript 1 and 2 and the ground station antenna index is omitted. By Eq. (2.2.3), the channel coefficients are
\[ (2.2.4) \quad c_i(t) = \frac{1}{\sqrt{N}} e^{j2\pi f_{1n}t} \sum_{n=1}^{N} e^{j[2\pi f_{D} t \cos(\xi - \alpha_n) + \phi_n + \phi_n^{(Sa)}]} . \]

Let \( \Delta \phi_n^{(Sa)} \) be the phase difference caused by the propagation path length difference from \( n \)th scatterer to two transmit antennas, then \( \Delta \phi_n^{(Sa)} = \phi_1^{(Sa)} - \phi_2^{(Sa)} = -2\pi(s_{1n} - s_{2n})/\lambda \). The distance between the \( i \)th transmit antenna and the \( n \)th scatterer \( s_{in} \) is

\[ (2.2.5) \quad s_{in} = \sqrt{(\frac{h_i}{\sin \beta_i})^2 + a^2 - 2a \frac{h_i}{\tan \beta_i} \cos(\varphi_i - \alpha_n)} . \]

For typical cases of wireless communication between satellite and ground station, \( h_i \gg a \) always holds. Thus Eq. (2.2.5) can be simplified to

\[ (2.2.6) \quad s_{in} \approx \frac{h_i}{\sin \beta_i} - a \cos \beta_i \cos(\varphi_i - \alpha_n) . \]

Define \( \Delta h_{12} \triangleq \frac{h_2}{\sin \beta_2} - \frac{h_1}{\sin \beta_1} \) and \( \Delta \varphi_{12} \triangleq \varphi_2 - \varphi_1 \). Then the path length difference from \( n \)th scatterer to the two transmit antennas is

\[ \Delta s_n \triangleq \frac{h_2}{\sin \beta_2} - \frac{h_1}{\sin \beta_1} - s_{1n} - s_{2n} \]

\[ \approx \Delta h_{12} + a[(\cos \beta_1 - \cos \beta_2 \cos \Delta \varphi_{12}) \cos(\varphi_1 - \alpha_n)] \]

\[ + (\cos \beta_2 \sin \Delta \varphi_{12}) \sin(\varphi_1 - \alpha_n)] . \]

Define \textit{projected distance}

\[ A_{12} \triangleq \sqrt{\cos^2 \beta_2 + \cos^2 \beta_1 - 2 \cos \beta_2 \cos \beta_1 \cos \Delta \varphi_{12}} , \]

and \( \gamma_{12} \triangleq \arctan \left( \frac{-\cos \beta_2 \sin \Delta \varphi_{12}}{\cos \beta_2 \cos \varphi_{12} - \cos \beta_1} \right) \).

Eq. (2.2.7) becomes

\[ (2.2.8) \quad \Delta s_n \approx \Delta h_{12} + a A_{12} \cos(\gamma_{12} + \varphi_1 - \alpha_n) , \]

which can be used to calculate \( \Delta \phi_n^{(Sa)} \) and eventually the correlation function.
PROPOSITION 1. Assuming $N \to \infty$, the cross-correlation between two transmit antennas for the 2-D scatterer ring case can be shown to be

$$R_{c_1c_2} = e^{2\pi j [f_{d_1} - f_{d_2}]t + f_{d_2} \tau + \frac{\Delta \lambda_{12}}{\lambda}}$$

$$\cdot J_0 \left\{ 2\pi \sqrt{\left[ f_{d_2} \cos \xi + \frac{a}{\lambda} A_{12} \cos (\gamma_{12} + \varphi_1) \right]^2} + \left[ f_{d_2} \sin \xi + \frac{a}{\lambda} A_{12} \sin (\gamma_{12} + \varphi_1) \right]^2 \right\},$$

(2.2.9)

where $J_0(\cdot)$ is the Bessel function of the first kind of order zero.

Under following conditions, the correlation of multipath components between two transmit antennas can be parameterized by $A_{12}$, instead of all the geometric parameters of the satellites, such as $\beta_i, h_i$ and $\varphi_i$. If the ground station is not moving, then by Proposition 1, the correlation is

$$R_{c_1c_2} = e^{2\pi j [f_{d_1} - f_{d_2}]t + f_{d_2} \tau + \frac{\Delta \lambda_{12}}{\lambda}} \cdot J_0 \left( \frac{2\pi}{\lambda} a A_{12} \right).$$

(2.2.10)

The norm of the correlation is $|R_{c_1c_2}| = |J_0 \left( \frac{2\pi}{\lambda} a A_{12} \right)|$, which is solely determined by $A_{12}$. When $A_{12}$ is small, the correlation between the two transmit antennas' multipath components is high. As shown in Figure 2.1, the physical meaning of $A_{12}$ is the distance between $\vec{v}_1$ and $\vec{v}_2$, the projections of two transmit antennas' unit length direction vectors $\vec{d}_1$ and $\vec{d}_2$. Thus, we call it “projected distance”, and it indicates how close the directions of these two transmit antennas are. Since high correlation leads to decreasing of the rate, this parameter is useful for Distributed MIMO system rate evaluation.

2.2.2.2. Transmit Antenna Correlation (3-D Scatterer Ring Case). Here we assume that the scatterers are uniformly distributed on the 3-D scatterer ring in the shape of a cylinder. This is useful for an urban area. For a scatterer $sc_{nm}$, its direction angle is $\alpha_n = 2\pi n/N$, $n = 1, 2, \ldots, N$, and its height is $h_{Rm} = h_L + (h_H - h_L) \frac{m-1}{M-1}$, $m = 1, 2, \ldots, M$, where $h_L$ and $h_H$ are corresponding to bottom and top of the scatterer ring. Thus the total number of scatterers is $NM$. Considering of the phase difference brought by the height of the scatterers, Proposition 1 can be extended to the 3-D version.
Proposition 2. Assuming $N \to \infty$ and $M \to \infty$, the cross-correlation between two transmit antennas for the 3-D scatterer ring case is

$$R_{c_1c_2} = e^{j2\pi[(f_{D_1}-f_{D_2})t+D_2\tau+\Delta h_{12}]}$$

$$\cdot e^{-j\pi \frac{1}{\sin \beta_1} - \frac{1}{\sin \beta_2}}(h_L+h_H)$$

$$\cdot \text{sinc}(\lambda \left(\frac{1}{\sin \beta_1} - \frac{1}{\sin \beta_2}\right)(h_H-h_L))$$

$$\cdot J_0\left[2\pi \sqrt{|f_D\tau \cos \xi + \frac{a}{\lambda}A_{12} \cos(\gamma_{12} + \phi_1)|^2}$$

$$+ |f_D\tau \sin \xi + \frac{a}{\lambda}A_{12} \sin(\gamma_{12} + \phi_1)|^2\right].$$

(2.2.11)

The cross-correlation function for the 3-D case only differ in a phase shift and in the sinc function part compared to 2-D case. The projected distance again plays an essential role in determining the value of the cross-correlation function. But it should be noticed that the sinc function part can reduce the correlation even if the projected distance is small.

2.2.2.3. Transmit and Receive Antenna Correlation. Finally we consider the correlation between any two entries of the channel matrix. For the 2-D scatterer ring case, since we assume all the receive antennas are near the center of the scatterer ring and the distances between them are far smaller than the radius ($a \gg d_{pq}^{(GS)}$), it’s reasonable to assume that the two arbitrary receive antennas are symmetric with respect to the center of the scatterer ring. Let the line connecting the two receive antennas be the Y-axis. For arbitrary receive antennas $p$ and $q$ and transmit antennas $k$ and $l$, assume coordinates of the first receive antenna in the XYZ-axis is $(0, \frac{d_{pq}^{(GS)}}{2}, 0)$ and coordinates of the second one is $(0, -\frac{d_{pq}^{(GS)}}{2}, 0)$. Let $\Delta \phi_n^{(GS)} = \phi_q^{(GS)} - \phi_p^{(GS)} = -2\pi(L_{qn} - L_{pn})/\lambda$. Since $a \gg d_{pq}^{(GS)}$, we have $L_{pn} \approx a - \frac{d_{pq}^{(GS)}}{2} \cos \alpha_n$ and $L_{qn} \approx a + \frac{d_{pq}^{(GS)}}{2} \cos \alpha_n$, thus $\Delta \phi_n^{(GS)} \approx -\frac{2\pi}{\lambda}d_{pq}^{(GS)} \cos \alpha_n$. By further considering $\Delta \phi_n^{(GS)}$ in Eq. (2.2.3), we generalize Proposition 1 and 2 to the correlation between two arbitrary antenna pairs.
Proposition 3. Letting $N \to \infty$, the cross-correlation between two channel coefficients in the 2-D scatterer ring case is

$$
R_{c_{pk}c_{ql}} = e^{j2\pi[(f_{D_k} - f_{D_l})t + f_{D_l}\tau + \Delta h_{kl}]}
\cdot J_0\left\{2\pi \sqrt{f_{D_l}\tau \cos \xi - \frac{d_{pq}^{(GS)}}{\lambda} + \frac{a}{\lambda} A_{kl} \cos(\gamma_{kl} + \varphi_k)}\right\}^2
+ [f_{D_l}\tau \sin \xi + \frac{a}{\lambda} A_{kl} \sin(\gamma_{kl} + \varphi_k)]^2\}.
$$

(2.2.12)

Note that $R_{c_{pk}c_{pd}}$ or $R_{c_{pk}c_{qk}}$ can be obtained by letting $d_{pq}^{(GS)} = 0$ or $\Delta h_{kl} = 0$, $A_{kl} = 0$, $f_{D_k} = f_{D_l}$.

Proposition 4. Letting $N \to \infty$ and $M \to \infty$, the cross-correlation between two arbitrary antenna pairs for the 3-D scatterer ring case is

$$
R_{c_{pk}c_{ql}} = \frac{1}{h_H - h_L} e^{j2\pi[(f_{D_k} - f_{D_l})t + f_{D_l}\tau + \Delta h_{kl}]}
\cdot \int_{h_L}^{h_H} e^{-j\frac{2\pi}{\lambda} \left(\frac{1}{\sin \gamma_k} - \frac{1}{\sin \gamma_l}\right)} \cdot J_0\left\{2\pi \sqrt{f_{D_l}\tau \cos \xi - \frac{d_{pq}^{(GS)}}{\lambda} + \frac{a}{\lambda} A_{kl} \cos(\gamma_{kl} + \varphi_k)}\right\}^2
+ [f_{D_l}\tau \sin \xi + \frac{a}{\lambda} A_{kl} \sin(\gamma_{kl} + \varphi_k)]^2\} dx.
$$

(2.2.13)

Although it doesn’t have a easy close form expression as the 2-D case, numerical method can be used to calculate its value.

2.3. Rate and Correlation

Assuming high speed optical link among the satellites, they can form a distributed MIMO transmitter. The received signal at the ground station is $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$, where $\mathbf{x} \in \mathbb{C}^{n_t \times 1}$ is the transmitted signal; $\mathbf{y} \in \mathbb{C}^{n_r \times 1}$ is the received signal; and $\mathbf{w} \in \mathbb{C}^{n_r \times 1}$ is circularly symmetric complex Gaussian noise; $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ is the channel matrix. The channel matrix
is a linear combination of the LOS component and multipath component as

\[
H = \sqrt{\frac{1}{K + 1}} H_{MP} + \sqrt{\frac{K}{K + 1}} H_{LOS},
\]

(2.3.1)

where \(H_{MP} = \begin{bmatrix} c \end{bmatrix}^{(LOS)} \) is the multipath component; \(H_{LOS} = [c_{pk}]\) is the LOS component; and \(K\) is the power ratio between them.

We evaluate the information rate assuming no channel state information (CSI) at the transmitter and perfect CSI at the receiver because of large feedback delay from the receiver to the transmitter. Since the transmitters do not know the correlation of the channel matrix, a strategy that will maximize the minimum average mutual information is suggested in [27] to transmit an isotropic signal with covariance matrix \(E[xx^\dagger] = \frac{P}{n_T} I_{n_T}\). For a specific real scenario, the powers of entries of \(x\) should be scaled to reflect the path losses. But for the purpose of illustration in this paper, we assume all transmit antennas have the same distance from the receiver. Under the power constraint \(E[\|x\|^2] \leq P\), the information rate is [26]

\[
C(H) = E[H][\log \det(I_{n_T} + (P/n_T)HH^\dagger)],
\]

(2.3.2)

where we have assumed wideband signals so that the phase changes caused by the Doppler shifts are not significant within a symbol period. Then the rate does not depend the Doppler frequency. For narrow band signal, the rate will be studied in future works.

The projected distance \(A_{ij}\) will determine the performance of the system, since both the cross-correlation function of the multipath component and the phase of the LOS component depend on \(A_{ij}\). For example, if one \(A_{ij} \to 0\), the rate will be reduced to the case of \(n_T - 1\) satellite antennas.

The projected distance can be used to help the system design:

- For one satellite, one antenna is enough if the ground station antennas are in the same scatterer ring and close to each other. On the other hand, if the ground station antennas have very large separation, like in the case of high speed linked cooperative ground stations at different locations, multiple antennas on one satellite is useful. This fits the design rule in [21] that small-sized antenna arrays can be realized at
one link end only at the cost of larger element displacements at the opposite link end.

- The projected distances of the scatterer ring of a ground station can be used to determine when the satellites are worth cooperating. Note that the projected distances are not the same for different scatterer rings at different locations. If cooperative multiple ground stations at different locations are used, the system will have better performance if the projected distances are large for at least some ground stations.

### 2.4. Performance Evaluation

A scenario involving three one antenna satellites and a ground station with three antennas is considered to show how the projected distances $A_{ij}$ affect the rate of the distributed MIMO satellite system. In the urban area, the multipath signal energy is no longer negligible due to the rich scattering environment, $K$ can vary from 3.9−11.9dB in the city [16]. Thus, the simulation parameters are set as $K = 7$dB, $SNR = 10$dB, carrier frequency $f_c = 1.2$GHz, symbol period $T_s = 50$ns, scatterer ring radius $a = 20$m, number of scatterers $N = 64$, ground station moving speed and direction $v_{GS} = 80$km/hr, $\xi = \frac{3\pi}{5}$, three ground station antennas’ location $\rho_1 = \rho_2 = \rho_3 = 0.3$m, $\theta_1 = 0$, $\theta_2 = \frac{2\pi}{3}$, $\theta_3 = \frac{4\pi}{3}$, three LEO satellites’ location with respect to the scatterer ring $h_1 = 900$km, $\beta_1 = \frac{\pi}{7}$, $\varphi_1 = (0, \pi]$, $h_2 = 900$km, $\beta_2 = \frac{\pi}{4}$, $\varphi_2 = 0$, $h_3 = 950$km, $\beta_3 = \frac{\pi}{3}$, $\varphi_3 = \frac{\pi}{4}$. Three different cases are compared: SIMO(1 × 3) system using satellite 1, MIMO(2 × 3) system using satellite 1 and 2, MIMO(3 × 3) system using all three satellites. The locations of the 2nd and 3rd satellites are fixed, while the location of the 1st satellite changes to examine the influence of satellite positions on the rate of different systems.

As shown in Fig. 2.1, significant rate gain can be obtained by employing the distributed MIMO transmission. Since the SIMO(1 × 3) system only involves one satellite, its rate won’t change. The MIMO systems have significant rate gain depending on the channel correlation, which in turn depends on the projected distance. From the simulation results, we can see that when $A_{12} \rightarrow 0$, the performance of $(n_T \times 3)$ system will drop to $((n_T - 1) \times 3)$ system
level, just as stated in Sec. 2.3. The rate gain only have significant degradation when the projected distance is very close to zero.

Therefore, the average rate gain can be indicated by the probability that the projected distance is smaller than a threshold, \( P(\{A_{ij} < A_0\}) \). From Figure 2.1, recall that \( A_{ij} \) is determined by length of the projections of two satellite antennas’ unitary direction vectors and the angle between them. Let \( A_0 = 0.1 \) be the threshold, above which Fig. 2.1 shows significant rate gain. We have \( P(\{A_{ij} < 0.1\}) \approx 1.43\% \) for uniformly distributed satellites, which suggests that the distributed MIMO satellite system can have significant rate gain in most situations.

2.5. Conclusion

In this chapter, a 3-D channel model for both LOS and multipath components of distributed MIMO communication systems is proposed. The analytic form of the cross-correlation functions of multipath components are derived. The numerous geometric parameters of the system are captured by a single key parameter, the projected distance, which appears in
the correlation functions which affects the rate. Numeric results show that the distributed MIMO satellite system has significant rate gain in most situations compared to single satellite systems.
CHAPTER 3

The Dual Link Algorithm

3.1. Introduction

One of the main approaches to accommodating the explosive growth in mobile data is
to reduce the cell size and increase the base station/access point density. However, the path
loss versus distance curve is flatter at shorter distance than that of longer distance. As a
result, the inter-cell interference becomes significant as the cell size/coverage area shrinks.
Therefore, joint transmit signal design that manages interference is a key technology for
the next generation wireless communication systems. Traditionally, interference between
different users is handled by enforcing certain orthogonality between them, which can be
frequency, time, space or even code. In the sense of the optimization problem that allocates
all the available resources to these users in order to maximize the overall performance (e.g.
the weighted sum rate), introducing orthogonality means adding more constraints to the
original problem and may end up with a sub-optimal solution. Instead of degrading the
problem for simplicity, our goal here is to find a reliable way to achieve the optimal solution
of the general weighted sum rate maximization problem with only power constraint.

In this chapter, we design a new algorithm that jointly optimizes the covariance matrices
of transmit signals of multiple transmitters in order to maximize the weighted sum-rate of
the data links for the MIMO B-MAC networks, assuming Gaussian transmit signal and the
availability of all channel state information. The MIMO B-MAC network model [14] includes
broadcast channel (BC), multiaccess channel (MAC), interference channels, X networks, and
many practical wireless networks as special cases. The problem is non-convex, and various
algorithms have been proposed for various cases, e.g., [5–7, 9, 14, 23, 24, 32–36]. Among
the state-of-the-art algorithms, we have proposed the polite water-filling (PWF) algorithms [14].
Because it takes advantage of the optimal transmit signal structure for an achievable rate
region, the polite water-filling structure, the PWF algorithm has the lowest complexity and the fastest convergence when it converges. However, in some strong interference cases, it has small oscillation. Another state-of-the-art algorithm is the WMMSE algorithm in [24]. It was proposed for beamforming matrix design for the MIMO interfering broadcast channels but could be readily applied to the more general B-MAC networks and input covariance matrix design. It transforms the weighted sum-rate maximization into an equivalent weighted sum-MSE cost minimization problem, which has three sets of variables and is convex when any two variable sets are fixed. With the block coordinate descent technique, the WMMSE algorithm is guaranteed to converge to a stationary point, though the convergence is observed in simulations to be relatively slower than the PWF algorithm.

It is thus highly desirable to have an algorithm with the advantages of both PWF and WMMSE algorithms, i.e., fast convergence by taking advantage of the optimal transmit signal structure and provable convergence for the general interference network. The main contribution of this paper is to present such an algorithm, named dual link algorithm. It exploits the forward-reverse link rate duality in a new way. Numerical experiments demonstrate that the dual link algorithm is almost as fast as the PWF algorithm and can be a few iterations or more than ten iterations faster than the WMMSE algorithm, depending on the desired accuracy with respect to the local optimum. Note that being faster even by a couple iterations will be critical in distributed implementation in time division duplex (TDD) networks with time varying channels, where the overhead of each iteration costs significant signaling resources between the transmitters and the receivers. The faster the convergence is, the faster channel variations can be accommodated by the algorithm. Indeed, the dual link algorithm is highly scalable and suitable for distributed implementation because, for each data link, only its own channel state and the aggregated interference plus noise covariance needs to be estimated no matter how many interferers are there. We have also shown that the dual link algorithm can be easily modified to deal with the weighted sum-rate maximization in an environment with colored noise.
Another contribution of this paper is the proof of the monotonic convergence of the algorithm. It uses only very general convex analysis, as well as a particular scaling invariance property that we identify for the weighted sum-rate maximization problem. We expect that the scaling invariance holds for and our proof technique applies to many non-convex problems in communication networks.

The centralized version of dual link algorithm for total power constraint has been generalized to multiple linear constraints using a minimax approach [2], and has stimulated the design of another monotonic convergent algorithm based on convex-concave procedure [30] which has slower convergence but can handle nonlinear convex constraints. Nevertheless, the dual link algorithm uses a different derivation approach, which is based on the optimal transmit signal structure, and easily leads a low complexity distributed algorithm. Thus, the special case of total power constraint provides a different view and insight than the general multiple linear constraint case.

The rest of this chapter is organized as follows. Section 3.2 presents the system model, formulates the problem, and briefly reviews the related results on the rate duality and polite water-filling structure. Section 3.3 proposes the new algorithm and establishes its monotonic convergence. Section 3.4 shows how to modify the dual link algorithm for the environment with colored noise. Numerical examples are presented in Section 3.5. Complexity analysis is provided in Section 3.6. Section 3.7 concludes.

3.2. Preliminaries

In this section, we describe the system model and formulate the optimization problem, then briefly review some related results on the polite water-filling, which leads to the design of the dual link algorithm that has fast, monotonic convergence.

3.2.1. B-MAC Interference Networks. We consider a general interference network named MIMO B-MAC network with multiple transmitters and receivers [12, 14]. A transmitter in the MIMO B-MAC network may send independent data to different receivers, like in BC, and a receiver may receive independent data from different transmitters, like in MAC.
Assume there are totally $L$ mutually interfering data links in a B-MAC network. Link $l$’s physical transmitter is $T_l$, which has $L_{T_l}$ many antennas. Its physical receiver is $R_l$, which has $L_{R_l}$ many antennas. Figure 3.1 shows an example of B-MAC networks with five data links. Link 2 and 3 have the same receiver. Link 3 and 4 have the same transmitter. When multiple data links have the same receiver or the same transmitter, interference cancellation techniques such as successive decoding and cancellation or dirty paper coding can be applied [14]. The received signal at $R_l$ is

$$y_l = \sum_{k=1}^{L} H_{l,k} x_k + w_l,$$

where $x_k \in \mathbb{C}^{L_{T_k} \times 1}$ is the transmit signal of link $k$ and is modeled as a circularly symmetric complex Gaussian vector; $H_{l,k} \in \mathbb{C}^{L_{R_l} \times L_{T_k}}$ is the channel state information (CSI) matrix between $T_k$ and $R_l$; and $w_l \in \mathbb{C}^{L_{R_l} \times 1}$ is a circularly symmetric complex Gaussian noise vector with identity covariance matrix. The circularly symmetric assumption of the transmit signal can be dropped easily by applying the proposed algorithm to real Gaussian signals with twice the dimension.

Figure 3.1. An example of B-MAC network. The solid lines represent data links and the dash lines represent interference.
3.2.2. Problem Formulation. Assuming the channels are known at both the transmitters and receivers (CSITR), an achievable rate of link \( l \) is

\[
I_l (\Sigma_{1:L}) = \log \left| I + H_{l,l} \Sigma_l H_l^H \Omega_l^{-1} \right|
\]

where \( \Sigma_l \) is the covariance matrix of \( x_l \); and \( \Omega_l \) is the interference-plus-noise covariance matrix of the \( l \)th link,

\[
\Omega_l = I + \sum_{k=1}^{L} H_{l,k} \Sigma_k H_{l,k}^H.
\]

If the interference from link \( k \) to link \( l \) is completely canceled using successive decoding and cancellation or dirty paper coding, we can simply set \( H_{l,k} = 0 \) in (3.2.3). Otherwise, the interference is treated as noise. This allows this paper to cover a wide range of communication techniques.

The optimization problem that we want to solve is the weighted sum-rate maximization under a total power constraint:

\[
\text{WSRM}_\text{TP}: \max_{\Sigma_{1:L}} \sum_{l=1}^{L} w_l I_l (\Sigma_{1:L})
\]

\[
\text{s.t.} \quad \Sigma_l \succeq 0, \forall l,
\]

\[
\sum_{l=1}^{L} \text{Tr}(\Sigma_l) \leq P_T,
\]

where \( w_l > 0 \) is the weight for link \( l \). The generalization to multiple linear constraints as in [13] is given in [2], which covers the individual power constraints as a special case.

3.2.3. Rate Duality and Polite Water-filling. We review the relevant results on the non-convex optimization (3.2.4) given in [14]. Dual network, reverse links, and rate duality was introduced. The optimal structure of the transmit signal covariance matrices is polite water-filling structure, whose definition involves the reverse link interference plus noise covariance matrices. It suggests an iterative polite water-filling algorithm, which is
compared with the new algorithm in this paper. The polite water-filling structure was used to derive a dual transformation, based on which the new algorithm in this paper has been designed.

A Dual Network and the Reverse Links. A virtual dual network can be created from the original B-MAC network by reversing the roles of all transmitters and receivers and replacing the channel matrices with their conjugate transpose. The data links in the original networks are denoted as forward links while those in the dual network are denoted as reverse links. We use \(^{\cap}\) to denote the corresponding terms in the reverse links. The interference-plus-noise covariance matrix of reverse link \(l\) is

\[
\hat{\Omega}_l = I + \sum_{k=1}^{L} H^\dagger_{k,l} \hat{\Sigma}_k H_{k,l},
\]

where \(\hat{\Sigma}_k\) is the transmit signal covariance matrix of reverse link \(k\). The achievable rate of reverse link \(l\) is

\[
\hat{I}_l \left( \hat{\Sigma}_{1:L} \right) = \log \left| I + H^\dagger_{l,l} \hat{\Sigma}_l H_{l,l} \hat{\Omega}_l^{-1} \right|.
\]

A dual optimization problem corresponding to 3.2.4 can be formulated as

\[
\text{WSRM \_ TP \_ D: } \max_{\Sigma_{1:L}} \sum_{l=1}^{L} w_l \hat{I}_l \left( \hat{\Sigma}_{1:L} \right) \quad \text{s.t. } \hat{\Sigma}_l \succeq 0, \forall l, \sum_{l=1}^{L} \text{Tr} \left( \hat{\Sigma}_l \right) \leq P_T.
\]

Rate Duality. The rate duality states that the achievable rate regions of the forward link channels \(\left[ H_{l,k} \right], \sum_{l=1}^{L} \text{Tr} \left( \Sigma_l \right) \leq P_T \) and reverse link channels \(\left[ H^\dagger_{k,l} \right], \sum_{l=1}^{L} \text{Tr} \left( \hat{\Sigma}_l \right) \leq P_T \) are the same [14]. The achievable rate regions are defined using rates in (3.2.2) and (3.2.6). A covariance transformation in [14] calculates the reverse link input covariance matrices \(\hat{\Sigma}_l\)'s from the forward ones \(\Sigma_l\)'s. The rate duality is proved by showing that these calculated
\( \hat{\Sigma}_t \)'s achieves equal or higher rates than the forward link rates employing \( \Sigma_t \)'s under the same value of power constraint \( P_T [14] \).

**Polite Water-filling Structure.** We review the polite water-filling results from [14]. The Lagrange function of problem (3.2.4) is

\[
L (\mu, \Theta_{1:L}, \Sigma_{1:L}) = \sum_{l=1}^L w_l \log \left| I + H_{l,l} \Sigma_l H_{l,l}^\dagger \Omega_l^{-1} \right| + \sum_{l=1}^L \text{Tr} (\Sigma_l \Theta_l) + \mu \left( P_T - \sum_{l=1}^L \text{Tr} (\Sigma_l) \right),
\]

where \( \Theta_{1:L} \) and \( \mu \) are Lagrange multipliers. The KKT conditions are

\[
\nabla_{\Sigma_l} L = \begin{cases} w_l H_{l,l}^\dagger \left( \Omega_l + H_{l,l} \Sigma_l H_{l,l}^\dagger \right)^{-1} H_{l,l} + \Theta_l - \mu I \\ - \sum_{k \neq l} w_k H_{k,l}^\dagger \left( \Omega_k^{-1} - \left( \Omega_k + H_{k,k} \Sigma_k H_{k,k}^\dagger \right)^{-1} \right) H_{k,l} \end{cases} = 0,
\]

(3.2.8)

\[
\mu \left( P_T - \sum_{l=1}^L \text{Tr} (\Sigma_l) \right) = 0,
\]

\[ \text{tr} (\Sigma_l \Theta_l) = 0, \]

\[ \Sigma_l, \Theta_l \succ 0, \mu \geq 0. \]

At a stationary point of problem (3.2.4), the transmit signal covariance matrices \( \Sigma_{1:L} \) have the polite water-filling structure [14]. Recall that in a single user MIMO channel, the optimal \( \Sigma \) is a water-filling over channel \( \mathbf{H} \), i.e., the eigenvectors of \( \Sigma \) are the right singular vectors of \( \mathbf{H} \) and the eigenvalues are calculated using water-filling of parallel channels with singular values of \( \mathbf{H} \) as channel gains. The polite water-filling structure is that the equivalent transmit covariance matrix \( \Omega_l^\dagger \Sigma_l \Omega_l^\dagger \) is a water-filling over the equivalent post- and pre-whitened channel \( \hat{\mathbf{H}}_l = \Omega_l^{-\frac{1}{2}} \mathbf{H}_l \Omega_l^{-\frac{1}{2}} \), where the reverse link interference plus noise covariance \( \hat{\Omega}_l \) is
calculated from $\hat{\Sigma}_{1:L}$ and $\hat{\Sigma}_{1:L}$ are calculated from $\Sigma_{1:L}$ using the above mentioned covariance transformation. Simultaneously, these $\hat{\Sigma}_{1:L}$ also have the polite water-filling structure and are the stationary point of the reverse link optimization problem (3.2.7). In the case of parallel channels, the polite water-filling will reduce to the traditional water-filling.

**Polite Water-filling Algorithm.** The polite water-filling structure naturally suggests the following iterative polite water-filling algorithm, Algorithm PP, in [14]. It works as follows. After initializing the reverse link interference plus noise covariance matrices $\hat{\Omega}_{1:L}$, we perform a forward link polite water-filling to obtain $\Sigma_{1:L}$. The reverse link polite water-filling is performed to obtain $\hat{\Sigma}_{1:L}$. This finishes one iteration. The iterations stops when the change of the objective function is less than a threshold or when a predetermined number of iterations is reached. Because the algorithm enforces the optimal signal structure at each iteration, it converges very fast if it converges. In particular, for parallel channels, it gives the optimal solution in half an iteration with initial values $\hat{\Omega}_l = I$, $\forall l$. Unfortunately, this algorithm is not guaranteed to converge, especially in very strong interference cases.

**Dual Transformation.** The following relation between $\Sigma_{1:L}$ and $\hat{\Sigma}_{1:L}$ at stationary points are proved using the polite water-filling structure in [14]. We name them dual transformation in this paper:

\begin{equation}
\hat{\Sigma}_l = \frac{w_l}{\hat{\mu}} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,l}^\dagger \Sigma_l H_{l,l} \right)^{-1} \right), \quad l = 1, \ldots, L;
\end{equation}

\begin{equation}
\Sigma_l = \frac{w_l}{\mu} \left( \Omega_l^{-1} - \left( \hat{\Omega}_l + H_{l,l}^\dagger \hat{\Sigma}_l H_{l,l} \right)^{-1} \right), \quad l = 1, \ldots, L,
\end{equation}

where the Lagrange multipliers $\mu$ and $\hat{\mu}$ are the Lagrange multipliers of the forward and reverse links for the power constraints. Equation (3.2.9) can be substituted it into the KKT condition (3.2.8) to recover the polite water-filling solution to the KKT conditions. In past works, the term $\frac{w_l}{\mu} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,l}^\dagger \Sigma_l H_{l,l} \right)^{-1} \right)$ in the KKT condition has always been the obstacle to an elegant solution. The dual transformation is used in the next section to design a new convergent algorithm.
3.3. The Dual Link Algorithm

3.3.1. The Algorithm. We propose a new algorithm, named Dual Link Algorithm, for the weighted sum-rate maximization problem (3.2.4). It has fast and monotonic convergence. The main idea is that, since we already know the optimal input covariance matrices $\Sigma_{1:L}$ and $\hat{\Sigma}_{1:L}$ must satisfy the dual transformation (3.2.9) and (3.2.10), we can directly use these the dual transformation to update $\hat{\Sigma}_{1:L}$ and $\Sigma_{1:L}$, instead of solving the KKT conditions and enforce the polite water-filling structure of $\hat{\Sigma}_{1:L}$ and $\Sigma_{1:L}$ as in the polite water-filling algorithms [14].

It is well known that equality $\sum_{l=1}^{L} \text{Tr} (\Sigma_l) = P_T$ holds when $\Sigma_{1:L}$ is a stationary point of problem (3.2.4), e.g., [12, Theorem 8 (item 3)]. This is because of the nonzero noise variance. It indicates that the full power should always be used. Since the covariance transformation [14, Lemma 8] preserves total power, we also have $\sum_{l=1}^{L} \text{Tr} \left( \hat{\Sigma}_l \right) = P_T$. The Lagrange multipliers $\mu$ and $\hat{\mu}$ should be chosen to satisfy the power constraint $\sum_{l=1}^{L} \text{Tr} (\Sigma_l) = P_T$ as

$$\mu = \frac{1}{P_T} \sum_{l=1}^{L} w_l \text{tr} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,j} \hat{\Sigma}_l H_{l,j}^\dagger \right)^{-1} \right).$$

(3.3.1) $$\hat{\mu} = \frac{1}{P_T} \sum_{l=1}^{L} w_l \text{tr} \left( \hat{\Omega}_l^{-1} - \left( \hat{\Omega}_l + H_{l,j}^\dagger \hat{\Sigma}_l H_{l,j} \right)^{-1} \right).$$

(3.3.2)

The above suggests the Dual Link Algorithm in Table Algorithm 1, that takes advantage of the structure of the weighted sum-rate maximization problem. A node who knows global channel state information runs the algorithm. The algorithm starts by initializing $\Sigma_l$’s as random matrices or scaled identity matrices, which can be used to calculate forward link interference plus noise covariance $\Omega_l$’s. Then, $\Sigma_l$’s of the virtual reverse links can be calculated by the dual transformation (3.2.9) with $\mu$ given in (3.3.1). These $\hat{\Sigma}_l$’s are used to calculate virtual reverse link interference plus noise covariance matrices $\hat{\Omega}_l$’s. Then, $\Sigma_l$’s of the forward links can be calculated by the dual transformation (3.2.10) with $\hat{\mu}$ given in (3.3.2). The above is repeated until the weighted sum rate converges or a fixed number of iterations are reached.
Algorithm 1 Dual Link Algorithm

1. Initialize $\Sigma_l$’s, s.t. $\sum_{l=1}^{L} \text{Tr}(\Sigma_l) = P_T$
2. $R \leftarrow \sum_{l=1}^{L} w_l L_l(\Sigma_{1:L})$
3. Repeat
4. $R' \leftarrow R$
5. $\Omega_l \leftarrow I + \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^H$
6. $\Sigma_l \leftarrow \frac{P_{l} w_l (\Omega_l^{-1} - (\Omega_l + H_{l,l} \Sigma_l H_{l,l}^H)^{-1})}{\sum_{l=1}^{L} w_l \text{tr}(\Omega_l^{-1} - (\Omega_l + H_{l,l} \Sigma_l H_{l,l}^H)^{-1})}$
7. $\hat{\Omega}_l \leftarrow I + \sum_{k \neq l} H_{l,k}^H \Sigma_k H_{k,l}$
8. $\hat{\Sigma}_l = \frac{P_{l} w_l (\Omega_l^{-1} - (\hat{\Omega}_l + H_{l,l} \Sigma_l H_{l,l}^H)^{-1})}{\sum_{l=1}^{L} w_l \text{tr}(\Omega_l^{-1} - (\hat{\Omega}_l + H_{l,l} \Sigma_l H_{l,l}^H)^{-1})}$
9. $R \leftarrow \sum_{l=1}^{L} w_l L_l(\Sigma_{1:L})$
10. until $|R - R'| \leq \epsilon$ or a fixed number of iterations are reached.

The most important properties of the dual link algorithm is that, unlike other algorithms for this problem, it is extremely well suited for distributed implementation and is scalable to network size. In TDD systems, the virtual reverse links can be replaced by physical reverse links and the distributed algorithm are almost identical to the centralized version. The covariance matrices $\Omega_l$ and $H_{l,l} \Sigma_l H_{l,l}^H$ can be estimated at the forward link receiver of link $l$. Note that $\Omega_l$ is calculated by the channel for us for free, no matter how many interfering links are there. It is similar for the reverse links. All the calculations only need local covariance information. No matter how large the total network size is, as long as the local interfering links of significance at a node is upper bounded. The details of the distributed algorithm is given in an upcoming paper.

As confirmed by the proof and numerical experiments, Dual Link Algorithm has monotonic convergence and is almost as fast as the polite water-filling (PWF) algorithm. It converges to a stationary point of both problem (3.2.4) and its dual (3.2.7) simultaneously, and both (3.2.9) and (3.2.10) achieve the same sum-rate at the stationary point. We will analyze the convergence properties of Algorithm 1 in the next section.

3.3.2. Preliminaries of the Convergence Proof. In the following sections, we will prove the monotonic convergence of Algorithm 1. As will be seen later, the proof uses only
very general convex analysis, as well as a particular scaling invariance property that we identify for the weighted sum-rate maximization problem. We expect that the scaling invariance holds for and our proof technique applies to many non-convex problems in communication networks that involve the rate or throughput maximization.

3.3.2.1. Equivalent Problem and the Lagrange Function. The weighted sum-rate maximization problem (3.2.4) is equivalent to the following problem by considering the interference plus noise covariance matrices as additional variables with additional equality constraints:

\[
\begin{align*}
\max_{\Sigma_{1:L}, \Omega_{1:L}} & \quad \sum_{l=1}^{L} w_l \left( \log |\Omega_l + H_{l,l} \Sigma_l H_{l,l}^\dagger| - \log |\Omega_l| \right) \\
\text{s.t.} & \quad \Sigma_l \succeq 0, \ \forall l, \\
& \quad \sum_{l=1}^{L} \text{Tr}(\Sigma_l) \leq P_T, \\
& \quad \Omega_l = I + \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^\dagger, \ \forall l,
\end{align*}
\]

which is still non-convex. Consider the Lagrangian of the above problem

\[
F(\Sigma, \Omega, \Lambda, \mu) = \sum_{l=1}^{L} w_l \left( \log |\Omega_l + H_{l,l} \Sigma_l H_{l,l}^\dagger| - \log |\Omega_l| \right) \\
+ \mu \left( P_T - \sum_{l=1}^{L} \text{Tr}(\Sigma_l) \right) \\
+ \sum_{l=1}^{L} \text{Tr} \left( \Lambda_l \left( \Omega_l - I - \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^\dagger \right) \right),
\]

where $\Sigma$ represents $\Sigma_{1:L}$; $\Omega$ represents $\Omega_{1:L}$; $\Lambda$ represents $\Lambda_{1:L}$; the domain of $F$ is $\{ \Sigma, \Omega, \Lambda, \mu | \Sigma_l \in \mathbb{H}_{+}^{L_T \times L_T}, \Omega_l \in \mathbb{H}_{+}^{L_R \times L_R}, \Lambda_l \in \mathbb{H}^{L_R \times L_R}, \mu \in \mathbb{R}^+, \forall l \}$. Here $\mathbb{H}^{n \times n}$, $\mathbb{H}_{+}^{n \times n}$, and $\mathbb{H}_{++}^{n \times n}$ are the sets of $n \times n$ Hermitian matrices, positive semidefinite matrices, and positive definite matrices respectively.
One can easily verify that the function $F$ is concave in $\Sigma$ and convex in $\Omega$. Furthermore, the gradients are given by

$$
\nabla_{\Sigma_i} F = w_i H_{i,l}^\dagger \left( \Omega_i + H_{i,l} \Sigma_i H_{i,l}^\dagger \right)^{-1} H_{i,l} 
- \mu I - \sum_{k \neq l} H_{k,l}^\dagger \Lambda_k H_{k,l},
$$

$$
\nabla_{\Omega_i} F = w_l \left( \left( \Omega_i + H_{i,l} \Sigma_i H_{i,l}^\dagger \right)^{-1} - \Omega_i^{-1} \right) + \Lambda_l.
$$

Now suppose that we have the pair $(\Sigma, \Omega)$ such that

$$
\sum_{l=1}^L \text{Tr}(\Sigma_l) = P_T, \\
\Omega_l = I + \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^\dagger,
$$

then,

$$
F(\Sigma_{1:L}, \Omega_{1:L}, \Lambda_{1:L}, \mu) = \sum_{l=1}^L w_l \left( \log |\Omega_l + H_{l,l} \Sigma_l H_{l,l}^\dagger| - \log |\Omega_l| \right),
$$

which is the original weighted sum-rate function. For notational simplicity, denote the weighted sum-rate function by $V(\Sigma)$, i.e.,

$$
V(\Sigma) = \sum_{l=1}^L w_l \left( \log \left| I + \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^\dagger + H_{l,l} \Sigma_l H_{l,l}^\dagger \right| 
- \log \left| I + \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^\dagger \right| \right).
$$

3.3.2.2. Solution of the first-order condition. Suppose that we want to solve the following system of equations in terms of $(\Sigma, \Omega)$ for given $(\Lambda, \mu)$:

$$
\nabla_{\Sigma_i} F = 0,
\nabla_{\Omega_i} F = 0.
$$
\[ \nabla_{\Omega_t} F = 0. \]

Define
\[ \hat{\Sigma}_t = \frac{1}{\mu} \Lambda_t, \]
\[ \hat{\Omega}_t = I + \sum_{k \neq l} H_{k,l}^\dagger \hat{\Sigma}_t H_{k,l}, \]
the above system of equations becomes
\[ \hat{\Sigma}_t = \frac{w_t}{\mu} \left( \Omega_t^{-1} - \left( \Omega_t + H_{t,l} \hat{\Sigma}_t H_{t,l}^\dagger \right)^{-1} \right), \]
\[ \hat{\Omega}_t = \frac{w_t}{\mu} H_{t,l}^\dagger \left( \Omega_t + H_{t,l} \hat{\Sigma}_t H_{t,l}^\dagger \right)^{-1} H_{t,l}. \]

An explicit solution to this system of equations is given by
\[ \Sigma_t = \frac{w_t}{\mu} \left( \Omega_t^{-1} - \left( \Omega_t + H_{t,l} \hat{\Sigma}_t H_{t,l}^\dagger \right)^{-1} \right) \]
\[ \Omega_t = \frac{w_t}{\mu} H_{t,l}^\dagger \left( H_{t,l}^\dagger \hat{\Sigma}_t H_{t,l} + \Omega_t \right)^{-1} H_{t,l}. \]

The detailed proof of this solution can be found in [2,31].

**Remark 5.** (3.3.6) and (3.3.7) are actually the first-order optimality conditions of (3.3.3)’s dual problem which is equivalent to (3.2.7). Algorithm 1 uses (3.3.4) and (3.3.6) to update \( \hat{\Sigma}_{1:L} \) and \( \Sigma_{1:L} \). When it converges, equations (3.3.4)-(3.3.7) will all hold, and the KKT conditions of problem (3.3.3) and its dual will all be satisfied.

**3.3.3. Convergence Results.** We are ready to present the following two main convergence results regarding Algorithm 1. Denote by \( \Sigma^{(n)} \) the \( \Sigma \) value at the \( n \)-th iteration of Algorithm 1.

**Theorem 6.** The objective value, i.e., the weighted sum-rate, is monotonically increasing in Algorithm 1, i.e.,
\[ V(\Sigma^{(n)}) \leq V(\Sigma^{(n+1)}). \]
From the above theorem, the following conclusion is immediate.

**Corollary 7.** The sequence \( V_n = V(\Sigma^{(n)}) \) converges to some limit point \( V_{\infty} \).

**Proof.** Since \( V(\Sigma) \) is a continuous function and its domain \( \{\Sigma | \Sigma_l \succeq 0, \text{Tr(} \Sigma \text{)} \leq P_T, \forall l\} \) is a compact set, \( V_n \) is bounded above. From Theorem 6, \( \{V_n\} \) is a monotone increasing sequence, therefore there exists a limit point \( V_{\infty} \) such that \( \lim_{n \to \infty} V_n = V_{\infty} \). \( \square \)

If we define a stationary point \( (\Sigma^*_l) \) of Algorithm 1, \( \Sigma^{(n)} = \Sigma^* \) implies \( \Sigma^{(n+k)} = \Sigma^* \) for all \( k = 0, 1, \cdots \), then we have the following result.

**Theorem 8.** Algorithm 1 converges to a stationary point \( \Sigma^*_1 : L \).

The above implies that both the weighted sum rate and the transmit signal covariance matrices converge. The proof of Theorems 6 and 8 will be presented later in this section. Before that, we first establish a few inequalities and identify a particular scaling property of the Lagrangian \( F \).

**3.3.3.1. The first inequality.** Suppose that we have a feasible point \( \Sigma^{(n)} \succeq 0 \), and

\[
(3.3.8) \quad \sum_{l=1}^{L} \text{Tr} \left( \Sigma^{(n)}_l \right) = P_T.
\]

In Algorithm 1, we generate \( \Omega^{(n)}_l \) such that

\[
(3.3.9) \quad \Omega^{(n)}_l = I + \sum_{k \neq l} H_{l,k} \Sigma^{(n)}_k H_{l,k}^\dagger.
\]

Now we have a pair \( (\Sigma^{(n)}, \Omega^{(n)}) \). Using this pair, we can compute \( (\Lambda^{(n)}_1, \mu^{(n)}) \) as

\[
\Lambda^{(n)}_l = w_l \left( \Omega^{(n)}_l \right)^{-1} - \left( \Omega^{(n)}_l + H_{l,l} \Sigma^{(n)}_l H_{l,l}^\dagger \right)^{-1},
\]

\[
\mu^{(n)} = \frac{1}{P_T} \sum_{l=1}^{L} \text{Tr} \left( \Lambda^{(n)}_l \right).
\]

Note that \( \Sigma^{(n)}_l \) in Algorithm 1 is equal to
\[
\hat{\Sigma}_l^{(n)} = \frac{\Lambda_l^{(n)}}{\mu^{(n)}},
\]

From this and (3.3.4), the gradient of \( F \) with respect to \( \Omega \) at the point \( (\Sigma^{(n)}, \Omega^{(n)}) \) vanishes, i.e.,

\[
\nabla_{\Omega} F(\Sigma^{(n)}, \Omega, \Lambda^{(n)}, \mu^{(n)})|_{\Omega^{(n)}} = 0.
\]

Since \( F \) is convex in \( \Omega \), if we fix \( \Sigma = \Sigma^{(n)} \), then \( \Omega^{(n)} \) is a global minimizer of \( F \). In other words,

\[(3.3.10) \quad F(\Sigma^{(n)}, \Omega^{(n)}, \Lambda^{(n)}, \mu^{(n)}) \leq F(\Sigma^{(n)}, \Omega, \Lambda^{(n)}, \mu^{(n)})\]

for all \( \Omega > 0 \).

3.3.3.2. Scaling invariance of \( F \). We will identify a remarkable scaling invariance property of \( F \), which plays a key role in the convergence proof of Algorithm 1. For given \( (\Sigma^{(n)}, \Omega^{(n)}, \Lambda^{(n)}, \mu^{(n)}) \), we have

\[(3.3.11) \quad F\left(\frac{1}{\alpha} \Sigma^{(n)}, \frac{1}{\alpha} \Omega^{(n)}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}\right) = F(\Sigma^{(n)}, \Omega^{(n)}, \Lambda^{(n)}, \mu^{(n)})\]

for all \( \alpha > 0 \). To show this scaling invariance property, note that

\[
\Omega_l^{(n)} = \sum_{k \neq l} H_{l,k} \Sigma_k^{(n)} H_{l,k}^\dagger = \mathbf{I},
\]

\[
\sum_{l=1}^{L} \text{Tr}(\Sigma_l^{(n)}) = P_T,
\]

\[
P_T \mu^{(n)} = \sum_{l=1}^{L} \text{Tr}(\Lambda_l^{(n)}).
\]
Applying the above equalities and some mathematical manipulations, we have

\[
F(\frac{1}{\alpha} \Sigma^{(n)}, \frac{1}{\alpha} \Omega^{(n)}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)})
\]

\[
= \sum_{l=1}^{L} w_l \left( \log \left| \Omega_{l}^{(n)} + H_{l,l} \Sigma_{l}^{(n)} H_{l,l}^\dagger \right| - \log \left| \Omega_{l}^{(n)} \right| \right)
\]

\[
+ \alpha \mu^{(n)} \{ P_T - \frac{1}{\alpha} P_T \} + \sum_{l=1}^{L} \text{Tr} \left( \alpha \Lambda_{l}^{(n)} \left( \frac{1}{\alpha} I - I \right) \right)
\]

\[
= \sum_{l=1}^{L} w_l \left( \log \left| \Omega_{l}^{(n)} + H_{l,l} \Sigma_{l}^{(n)} H_{l,l}^\dagger \right| - \log \left| \Omega_{l}^{(n)} \right| \right)
\]

\[
+ (\alpha - 1) \mu^{(n)} P_T + (1 - \alpha) \sum_{l=1}^{L} \text{Tr}(\Lambda_{l}^{(n)})
\]

\[
= \sum_{l=1}^{L} w_l \left( \log \left| \Omega_{l}^{(n)} + H_{l,l} \Sigma_{l}^{(n)} H_{l,l}^\dagger \right| - \log \left| \Omega_{l}^{(n)} \right| \right)
\]

\[
= F(\Sigma^{(n)}, \Omega^{(n)}, \Lambda^{(n)}, \mu^{(n)}),
\]

where the first equality uses the fact that

\[
\log \left| \frac{1}{\alpha} \left( \Omega_{l}^{(n)} + H_{l,l} \Sigma_{l}^{(n)} H_{l,l}^\dagger \right) \right| - \log \left| \frac{1}{\alpha} \Omega_{l}^{(n)} \right|
\]

\[
= \log \left| \Omega_{l}^{(n)} + H_{l,l} \Sigma_{l}^{(n)} H_{l,l}^\dagger \right| - \log \left| \Omega_{l}^{(n)} \right|.
\]

Furthermore,

\[
\nabla_{\Omega_l} F(\frac{1}{\alpha} \Sigma^{(n)}, \Omega, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}) \big|_{\Omega^{(n)}}
\]

\[
= w_l \left( \left( \frac{1}{\alpha} \Omega_{l}^{(n)} + H_{l,l} \frac{1}{\alpha} \Sigma_{l}^{(n)} H_{l,l}^\dagger \right)^{-1} - \left( \frac{1}{\alpha} \Omega_{l}^{(n)} \right)^{-1} \right)
\]

\[
+ \alpha \nabla_{\Omega_l} F(\Sigma^{(n)}, \Omega, \Lambda^{(n)}, \mu^{(n)}) \big|_{\Omega^{(n)}}
\]

\[
(3.3.12)
\]

\[
= 0, \quad \forall l.
\]

Therefore, \( \frac{1}{\alpha} \Omega^{(n)} \) is a global minimizer of \( F(\frac{1}{\alpha} \Sigma^{(n)}, \Omega, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}) \), as \( F \) is convex in \( \Omega \).
3.3.3.3. The second and third inequalities. Given \((\alpha \Lambda^{(n)}, \alpha \mu^{(n)})\), we generate \(\hat{\Sigma}, \hat{\Omega}\) using equation (3.3.6) and (3.3.7). If we choose \(\alpha\) so that

\[
\sum_{l=1}^{L} \text{Tr}(\hat{\Sigma}_{l}) = P_T,
\]

then \(\hat{\Sigma} = \Sigma^{(n+1)}\) in Algorithm 1. Since \((\Sigma^{(n+1)}, \hat{\Omega})\) is chosen to make the gradients zero:

\[
\nabla_{\Sigma} F(\Sigma, \hat{\Omega}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)})|_{\Sigma^{(n+1)}} = 0,
\]

\[
\nabla_{\Omega} F(\Sigma^{(n+1)}, \Omega, \alpha \Lambda^{(n)}, \alpha \mu^{(n)})|_{\hat{\Omega}} = 0,
\]

we conclude that \(\Sigma^{(n+1)}\) is a global maximizer, i.e.,

\[
F(\Sigma, \hat{\Omega}, \alpha \Lambda^{(n+1)}, \alpha \mu^{(n+1)}) \leq F(\Sigma^{(n+1)}, \hat{\Omega}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)})
\]

for all \(\Sigma \succeq 0\); and \(\hat{\Omega}\) is a global minimizer, i.e.,

\[
F(\Sigma^{(n+1)}, \hat{\Omega}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}) \leq F(\Sigma^{(n+1)}, \Omega, \alpha \Lambda^{(n)}, \alpha \mu^{(n)})
\]

for all \(\Omega > 0\).

3.3.3.4. Proof of Theorem 6. With the three inequalities (3.3.10, 3.3.14, 3.3.15) obtained above, we are ready to prove Theorem 6. As in Algorithm 1

\[
\Omega_l^{(n+1)} = I + \sum_{k \neq l} H_{l,k} \Sigma_k^{(n+1)} H_{l,k}^*,
\]

we have
\[ V(\Sigma^{(n)}) \]

(3.3.17) \[ = F(\Sigma^{(n)}, \Omega^{(n)}, \Lambda^{(n)}, \mu^{(n)}) \]

(3.3.18) \[ = F\left(\frac{1}{\alpha} \Sigma^{(n)}, \frac{1}{\alpha} \Omega^{(n)}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}\right) \]

(3.3.19) \[ \leq F\left(\frac{1}{\alpha} \Sigma^{(n)}, \tilde{\Omega}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}\right) \]

(3.3.20) \[ \leq F(\Sigma^{(n+1)}, \tilde{\Omega}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}) \]

(3.3.21) \[ \leq F(\Sigma^{(n+1)}, \Omega^{(n+1)}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}) \]

(3.3.22) \[ = V(\Sigma^{(n+1)}), \]

where (3.3.17) follows from the satisfied constraints (3.3.8, 3.3.9); (3.3.18) follows from the scaling invariance (3.3.11); (3.3.19) follows from convexity and scaling invariance (3.3.10, 3.3.12); (3.3.20) follows from the second inequality (3.3.14); (3.3.21) follows from the third inequality (3.3.15); (3.3.22) follows from the satisfied constraints (3.3.13, 3.3.16).

3.3.3.5. Proof of Theorem 8. We have shown in Corollary 7 that \( V_n \) converges to a limit point under Algorithm 1. To show the convergence of the algorithm, it is enough to show that if \( V(\Sigma^{(n)}) = V(\Sigma^{(n+1)}) \), then \( \Sigma^{(n+1)} = \Sigma^{(n+k)} \) for all \( k = 1, 2, \cdots \). Suppose \( V(\Sigma^{(n)}) = V(\Sigma^{(n+1)}) \), then from the proof in the above, we have

\[ F(\Sigma^{(n+1)}, \Omega^{(n+1)}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}) \]

\[ = F(\Sigma^{(n+1)}, \tilde{\Omega}, \alpha \Lambda^{(n)}, \alpha \mu^{(n)}). \]
Since $\hat{\Omega}$ is a global minimizer, the above equality implies $\Omega^{(n+1)}$ is a global minimizer too. From the first order condition for optimality, we have

$$
\nabla_{\Omega} F(\Sigma^{(n+1)}, \Omega, \alpha \Lambda^{(n)}, \alpha \mu^{(n+1)})|_{\Omega^{(n+1)}}
= w_l \left( \left( \Omega^{(n+1)}_l + H_{l,l} \Sigma^{(n+1)}_l H_{l,l}^l \right)^{-1} - \left( \Omega^{(n+1)}_l \right)^{-1} \right)
+ \alpha \Lambda^{(n)}_l
= 0.
$$

On the other hand, we generate $\Lambda^{(n+1)}$ such that

$$
\Lambda^{(n+1)}
= w_l \left( \Omega^{(n+1)}_l^{-1} - \left( \Omega^{(n+1)}_l + H_{l,l} \Sigma^{(n+1)}_l H_{l,l}^l \right)^{-1} \right)
= \alpha \Lambda^{(n)}_l.
$$

This shows $\hat{\Sigma}^{(n+1)} \propto \hat{\Sigma}^{(n)}$. However, since the trace of each matrix is same, we conclude that

$$
\hat{\Sigma}^{(n+1)} = \hat{\Sigma}^{(n)}.
$$

From this it is obvious that $\hat{\Sigma}^{(n)} = \hat{\Sigma}^{(n+1)} = \ldots$ and $\Sigma^{(n+1)} = \Sigma^{(n+2)} = \ldots$.

Remark 9. When the algorithm converges, the pair $(\Sigma, \Omega)$ satisfies the first order optimality condition for $F$. Moreover, since $\sum_{l=1}^L \text{Tr} (\Sigma_l) = P_T$, and $\Omega_l = I + \sum_{k \neq l} H_{l,k} \Sigma_k H_{l,k}^l$, \forall $l$, $(\alpha \Lambda, \alpha \mu)$ also satisfies the first order optimality condition for $F$. This implies that the pair $(\Sigma, \Omega, \alpha \Lambda, \alpha \mu)$ is a saddle point of $F$, which means that they are indeed a primal-dual point that solves the KKT system of the weighted sum-rate maximization.

3.4. Extension for Environment with Colored Noise

In reality, due to other non-cooperating networks, the noise at the receivers may no longer be white as stated in (3.2.3). However, the dual link algorithm can be easily adjusted to
solve this weighted sum-rate maximization problem with colored noise. The key idea is to use proper noise whitening method to convert the problem back to the white noise form while still keeping the reciprocity property of the forward and reverse link channels so that the original dual link algorithm can be readily applied.

Assuming the noise covariance matrix of forward/reverse link \( l \) is \( \mathbf{N}_l \) instead of \( \mathbf{I} \). Let \( \mathbf{T}_l = \mathbf{N}_l^{\frac{1}{2}} \), the original weighted sum-rate maximization problem with colored noise is

\[
(3.4.1) \quad \max_{\Sigma_1, \ldots, \Sigma_L} \sum_{l=1}^{L} w_l \log \left| \mathbf{I} + \mathbf{H}_{l,l} \Sigma_l \mathbf{H}_{l,l}^\dagger \Omega_l^{-1} \right|
\]

s.t. \( \Sigma_l \succeq 0, \forall l, \sum_{l=1}^{L} \text{Tr} (\Sigma_l) \leq P_T, \)

\[\Omega_l = \mathbf{N}_l + \sum_{k=1, k \neq l}^{L} \mathbf{H}_{l,k} \Sigma_k \mathbf{H}_{l,k}^\dagger,\]

which is equivalent to

\[
(3.4.2) \quad \max_{\Sigma_1, \ldots, \Sigma_L} \sum_{l=1}^{L} w_l \log \left| \mathbf{I} + \mathbf{T}_l^{-1} \mathbf{H}_{l,l} \Sigma_l \mathbf{H}_{l,l}^\dagger \mathbf{T}_l^{-1} \Omega_l^{-1} \right|
\]

s.t. \( \Sigma_l \succeq 0, \forall l, \sum_{l=1}^{L} \text{Tr} (\Sigma_l) \leq P_T, \)

\[\Omega_l = \mathbf{I} + \sum_{k=1, k \neq l}^{L} \mathbf{T}_l^{-1} \mathbf{H}_{l,k} \Sigma_k \mathbf{H}_{l,k}^\dagger \mathbf{T}_l^{-1}.\]

Therefore the forward link receiver \( R_l \) should multiply the whitening matrix \( \mathbf{T}_l^{-1} \) to the received signal.

The dual problem of (3.4.2) is

\[
(3.4.3) \quad \max_{\hat{\Sigma}_1, \ldots, \hat{\Sigma}_L} \sum_{l=1}^{L} w_l \log \left| \mathbf{I} + \mathbf{H}_{l,l}^\dagger \mathbf{T}_l^{-1} \hat{\Sigma}_l \mathbf{T}_l^{-1} \mathbf{H}_{l,l} \hat{\Omega}_l^{-1} \right|
\]

s.t. \( \hat{\Sigma}_l \succeq 0, \forall l, \sum_{l=1}^{L} \text{Tr} (\hat{\Sigma}_l) \leq P_T, \)

\[\hat{\Omega}_l = \mathbf{I} + \sum_{k=1, k \neq l}^{L} \mathbf{H}_{k,l}^\dagger \mathbf{T}_k^{-1} \hat{\Sigma}_k \mathbf{T}_k^{-1} \mathbf{H}_{k,l}.\]
This means the reverse link transmitter $R_l$ should multiply the whitening matrix $\mathbf{T}_l^{-1}$ to the signal before transmitting. Note that the noise covariance matrix in dual problem (3.4.3) is still $\mathbf{I}$.

Therefore, problem (3.4.1) with colored noise can be transformed into the equivalent problem (3.4.2) which can be solved using Algorithm 1. It should also be pointed out that $\mathbf{N}_l$ is only needed in $R_l$ where it can be acquired locally, which is important for distributed implementation.

3.5. Simulation Results

In this section, we provide numerical examples to compare the proposed dual link algorithm with the PWF algorithm [14] and the WMMSE algorithm [24]. Consider a B-MAC network with 10 data links among 10 transmitter-receiver pairs that fully interfere with each other. Each link has 3 transmit antennas and 4 receive antennas. For each simulation, the channel matrices are independently generated and fixed by one realization of $\mathbf{H}_{l,k} = \sqrt{g_{l,k}} \mathbf{H}_{l,k}^{(W)}$, $\forall k, l$, where $\mathbf{H}_{l,k}^{(W)}$ has zero-mean i.i.d. complex Gaussian entries with unit variance and $g_{l,k}$ is the average channel gain. The weights $w_i$’s are uniformly chosen from 0.5 to 1. The total transmit power $P_T = 100$.

Fig. 3.1 shows the convergence of the Dual Link algorithm for a network with $g_{l,k} = 0$ dB, $\forall l, k$. From the proof of Theorem 6, the weighted sum-rate of the forward links and that of the reverse links not only increase monotonically over iterations, but also increase over each other over half iterations. In Algorithm 1, the reverse link transmit signal covariance matrices are updated in the first half of each iteration (line 6), and the forward link transmit signal covariance matrices are updated in the second half (line 8). From Fig. 3.1, we clearly see that the weighted sum-rates of the forward links and reverse links increase in turns until they converge to the same value, which also confirms that problem (3.2.4) and its dual problem (3.2.7) reach their stationary points at the same time.

Fig. 3.2 shows a typical case of rate versus number of iterations under the weak interference setting ($g_{l,l} = 0$ dB and $g_{l,k} = -10$ dB for $l \neq k$). Under a channel realization and
Figure 3.1. The monotonic convergence of the forward and reverse link weighted sum-rates of the Dual Link algorithm with $P_T = 100$ and $g_{l,k} = 0$dB, $\forall l,k$.

An initial point, PWF algorithm converges slightly faster than the dual link algorithm. In contrast, the WMMSE algorithm's convergence is slower, e.g., eight iterations to achieve what dual link algorithm achieves in four iterations and more than ten iterations slower to reach some higher value of the weighted sum-rate.

When the gain of the interfering channels are comparable to that of the desired channel, as shown in Fig. 3.3, the difference in the convergence speed between the PWF/Dual Link algorithm and the WMMSE algorithm is less than that of the weak interference case. But around five iteration difference for some high value of the weighted sum-rate is still significant.

Under strong interference setting, the PWF algorithm may oscillate and no longer converge as shown in Fig. 3.4. The dual link algorithm still converges slightly faster than the WMMSE algorithm.
Figure 3.2. PWF algorithm vs. WMMSE algorithm vs. dual link algorithm under weak interference with $P_T = 100$, $g_{l,l} = 0$dB and $g_{l,k} = -10$dB for $l \neq k$.

<table>
<thead>
<tr>
<th>$g_{l,k}$ (dB)</th>
<th>PWF (90%)</th>
<th>DL (90%)</th>
<th>WMMSE (90%)</th>
<th>PWF (95%)</th>
<th>DL (95%)</th>
<th>WMMSE (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>1.199</td>
<td>1.653</td>
<td>2.521</td>
<td>2.140</td>
<td>2.781</td>
<td>5.067</td>
</tr>
<tr>
<td>0</td>
<td>6.075</td>
<td>7.211</td>
<td>9.537</td>
<td>11.465</td>
<td>12.408</td>
<td>15.864</td>
</tr>
</tbody>
</table>

Table 3.1. Average number of iterations needed to reach 90% and 95% of the local optimum for PWF, WMMSE, and dual link algorithms. $P_T = 100$, $g_{l,l} = 0$dB and $g_{l,k} = -10/0/10$dB for $l \neq k$.

Given the same initial point, these three algorithms may converge to different stationary points, as shown in Fig. 3.5. Since the original weighted sum-rate maximization problem is non-convex, a stationary point is not necessarily a global maximum. In practical applications, one may run such algorithms multiple times starting from different initial points and choose the best.

Table 3.1 shows the average performance of these three algorithms under different interference levels. The results are obtained by averaging over 1000 independent channel
realizations under each interference setting. (It should be mentioned that, under the strong interference setting \( g_{l,k} = 10 \)dB), only the converged cases (837 out of 1000) are considered for the PWF algorithm.) We can see that the dual link algorithm outperforms WMMSE algorithm in all the settings. While the PWF algorithm is slightly ahead under weak and moderate interference settings, it is slower than the dual link algorithm under the strong interference setting.

Figure 3.6 compares the dual link algorithm with the sum-rate iterative water-filling algorithm proposed in [7]. The iterative water-filling algorithm is known to converge to the sum-rate boundary point of capacity region for both MAC and BC channels [7, 32]. Here we consider a BC channel and its dual MAC channel with the sum power constraint. The BC channel contains 1 transmitter and 10 receivers, each equipped with 5 antennas. Entries of the channel matrices are generated from i.i.d. zero mean Gaussian distribution with unit variance. The background noise covariance is identity matrix. We apply the iterative water-filling algorithm on the dual MAC channel and apply the dual link algorithm on both the dual MAC channel and the BC channel. As shown in Figure 3.6, they all converge to the
Figure 3.4. PWF algorithm vs. WMMSE algorithm vs. dual link algorithm under strong interference with $P_T = 100$, $g_{l,l} = 0$dB and $g_{l,k} = 10$dB for $l \neq k$.

same capacity region boundary point that maximizes the sum-rate, which confirms that the dual link algorithm achieves the capacity region boundary for MAC and BC channels. The dual link algorithm converges faster than the iterative water-filling algorithm.

3.6. Complexity Analysis

We have evaluated in the above convergence properties of the proposed new algorithm, the PWF algorithm and the WMMSE algorithm in terms of the number of iterations. We now analyze the complexity of each iteration for these algorithms.

Recall that $L$ is the number of users or links, and for simplicity, assume that each user has $N$ transmit (and receive) antennas, so the resulting $\Sigma_l$ (and $\hat{\Sigma}_l$) is a $N \times N$ matrix. Suppose that we use the straightforward matrix multiplication and inversion. Then the complexity of these operations are $O(N^3)$. For the new algorithm, at each iteration, $\Omega_l$ incurs a complexity
Figure 3.5. An example that PWF, WMMSE, and dual link algorithms converge to different stationary points. $P_T = 100$, $g_{t,l} = 0 \text{dB}$ and $g_{t,k} = 10 \text{dB}$ for $l \neq k$.

of $O(LN^3)$ and $\Omega_l + \mathbf{H}_{l,l} \hat{\Sigma}^{(n+1)} \mathbf{H}_{l,l}^\dagger$ incurs a complexity of $O(LN^3)$. To obtain $\hat{\Sigma}_l$, we have to invert a $N \times N$ matrix, which incurs a complexity of $O(N^3)$. Therefore, the total complexity for calculating a $\hat{\Sigma}_l$ is given by $O(LN^3)$, and the complexity of generating $\hat{\Sigma}$ is given by $O(L^2N^3)$. As calculating $\Sigma$ incurs the same complexity as calculating $\hat{\Sigma}$, the complexity of the new algorithm is $O(L^2N^3)$ for each iteration.

The PWF algorithm uses the same calculation to generate $\Omega_l$ and incurs a complexity of $O(LN^3)$ for each $\Omega_l$. Then, it uses the singular value decomposition of $\Omega_l^{-\frac{1}{2}} \mathbf{H}_{l,l} \hat{\Omega}_l^{-\frac{1}{2}}$, which incurs a complexity of $O(N^3)$. Since we need $L$ of these operations, the total complexity of the PWF algorithm is $O(L^2N^3)$. For the WMMSE algorithm, it is shown in [24] that its complexity is $O(L^2N^3)$.

So, all three algorithms have the same computational complexity per iteration if we use $O(N^3)$ matrix multiplication. Recently, Williams [28] presents an $O(N^{2.3727})$ matrix
Figure 3.6. Dual Link algorithm vs. Iterative water-filling on MAC and BC channels multiplication and inversion method. If we use this algorithm, then the new algorithm and the WMMSE algorithm have $O(L^2 N^{2.3727})$ complexity since the $N^3$ factor comes from the matrix multiplication and inversion. However, in addition to $L^2$ number of matrix multiplications and inversions, the PWF algorithm has $L$ number of $N$ by $N$ matrix singular value decompositions. Therefore the complexity of the PWF algorithm is $O(L^2 N^{2.3727} + LN^3)$.

3.7. Conclusion

We have proposed a new algorithm, the dual link algorithm, to solve the weighted sum-rate maximization problem in general interference channels. Based on the polite water-filling results and the rate duality [14], the proposed dual link algorithm updates the transmit signal covariance matrices in the forward and reverse links in a symmetric manner and has fast and guaranteed convergence. We have given a proof for its monotonic convergence, and the proof
technique may be generalized for other problems. Numerical examples have demonstrated that the new algorithm has a convergence speed close to the fastest PWF algorithm which is however not guaranteed to converge in all situations. The dual link algorithm is scalable and well suited for distributed implementation. It can also be easily modified to accommodate colored noise environments.
CHAPTER 4

Distributed Implementation of the Dual Link Algorithm

4.1. Introduction

In practical applications, distributed algorithm is much more important because centralized algorithm is not scalable as network size increases. Firstly, the CSI (channel state information) of all direct and cross links needed at central processor will result in huge signaling overhead; Secondly, all the optimization is performed at central processor, which means huge computation burden. If too much time is spent on feedbacks and computation, there won’t be enough time left for actual data transmission before the channels change. Therefore, for real applications in a MIMO interference network that contains many transmitter/receiver pairs, the optimization algorithm needs to be working in a mainly distributed fashion.

4.2. Distributed Dual Link Algorithm and Distributed Channel Estimation

The centralized dual link algorithm is ideal for distributed implementation, especially in TDD networks. In TDD networks, assuming reciprocity of channels, the physical reverse links are the same as the virtual reverse links. In each iteration of the dual link algorithm, the forward link input covariance matrix $\Sigma_l$ is updated as follows:

$$
\Sigma_l \leftarrow \frac{P_T w_l \left( \hat{\Omega}_l^{-1} - \left( \hat{\Omega}_l + H_{l,l}^\dagger \hat{\Sigma}_l H_{l,l} \right)^{-1} \right)}{\sum_{i=1}^L w_i \text{tr} \left( \hat{\Omega}_i^{-1} - \left( \hat{\Omega}_l + H_{l,l}^\dagger \hat{\Sigma}_l H_{l,l} \right)^{-1} \right)}.
$$

(4.2.1)

Notice that the denominator is just a scalar used for power adjustment and is same for any $l$. Thus, the reverse link interference-plus-noise covariance matrix $\hat{\Omega}_l$ and signal covariance matrix $H_{l,l}^\dagger \hat{\Sigma}_l H_{l,l}$ are the only two matrices required to calculate $\Sigma_l$. Moreover, both $\hat{\Omega}_l$ and $H_{l,l}^\dagger \hat{\Sigma}_l H_{l,l}$ can be estimated locally at $T_l$, because the channels have done the summation
of the signals for free regardless of the number of users. This means that we can exactly implement the dual link algorithm in distributed fashion with only scalar message exchange to satisfy the total power constraint. After every $T_l$ has estimated $\hat{\Omega}_l$ and $H_{l,t}^\dagger \Sigma_l H_{l,t}$, it can broadcast $w_t \text{tr} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,t}^\dagger \Sigma_l H_{l,t} \right)^{-1} \right)$ to each other and then update its $\Sigma_l$ using only local channel covariance $\hat{\Omega}_l$ and $H_{l,t}^\dagger \Sigma_l H_{l,t}$.

Similarly, in the forward link, $R_l$ can also estimate local channel covariance $\Omega_l$ and $H_{l,t}^\dagger \Sigma_l H_{l,t}^\dagger$, broadcast $w_t \text{tr} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,t}^\dagger \Sigma_l H_{l,t}^\dagger \right)^{-1} \right)$ to each other, and update its own $\hat{\Sigma}_l$:

$$\hat{\Sigma}_l \leftarrow \frac{P_t w_t \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,t}^\dagger \Sigma_l H_{l,t} \right)^{-1} \right)}{\sum_{i=1}^L w_t \text{tr} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,t}^\dagger \Sigma_l H_{l,t} \right)^{-1} \right)}.$$  

(4.2.2)

Now the problem left is how to estimate local channel covariance efficiently at each forward/reverse link receivers. In order to estimate each link’s own signal covariance and interference-plus-noise covariance separately, we can use orthogonal pilots for different links.

Without loss of generality, we can consider the forward links. Assuming $\tilde{x}_{l,i}$ is the pilot signal vector transmitted by $T_l$ at time $i$, if the total length is $N$, then the whole pilot signal from $T_l$ can be written as

$$X_l = [\tilde{x}_{l,1}, \tilde{x}_{l,2}, \ldots, \tilde{x}_{l,N}].$$

$X_l$’s need to satisfy:

$$\mathbb{E} [\tilde{x}_{k,n} \cdot \tilde{x}_{l,n}^\dagger] = \frac{1}{N} X_k X_l^\dagger = \begin{cases} \Sigma_l, & k = l \\ 0, & k \neq l \end{cases}$$

Thus, we can choose

$$X_l = U_l D_l^{\frac{1}{2}} Z_l$$
where $U_l$ and $D_l$ are calculated from SVD of $\Sigma_l$ ($\Sigma_l = U_l D_l U_l^\dagger$, $U_l$ is unitary, $D_l$ is diagonal with non-negative entries), and rows of all $Z_l$’s are orthogonal to each other ($\frac{1}{N}Z_k Z_l^\dagger = \begin{cases} I, & k = l \\ 0, & k \neq l \end{cases}$).

Then the received signal at $R_k$ is

$$Y_k = \sum_{l=1}^L H_{k,l} X_l + W_k$$

$$= \sum_{l=1}^L H_{k,l} U_l D_l^\frac{1}{2} Z_l + W_k$$

where $W_k$ is the background Gaussian noise at $R_k$ with zero mean and covariance $I$.

Since pilots $Z_k$ are known at both link $k$’s transmitter and receiver, we have:

$$\frac{1}{N} Y_k \cdot Y_k^\dagger = \frac{1}{N} \left( \sum_{l=1}^L H_{k,l} U_l D_l^\frac{1}{2} Z_l + W_k \right) \cdot \left( \sum_{l=1}^L H_{k,l} U_l D_l^\frac{1}{2} Z_l + W_k \right)^\dagger$$

$$\rightarrow \sum_{l=1}^L H_{k,l} U_l D_l^\frac{1}{2} U_l^\dagger H_{k,l}^\dagger + I$$

$$= H_{k,k} \Sigma_k H_{k,k}^\dagger + \Omega_k,$$

(4.2.3)

$$\frac{1}{N} Y_k \cdot Z_k = \frac{1}{N} \left( \sum_{l=1}^L H_{k,l} U_l D_l^\frac{1}{2} Z_l + W_k \right) \cdot Z_k$$

(4.2.4)

$$\rightarrow H_{k,k} U_k D_k^\frac{1}{2}.$$

By combining (4.2.3) and (4.2.4), we can get estimations of $\Omega_k$ and $H_{k,k} \Sigma_k H_{k,k}^\dagger$ at $R_k$. Similarly, in the reverse link, we’ll be able to estimate $\hat{\Omega}_l$ and $H_{l,l}^\dagger \hat{\Sigma}_l H_{l,l}$ at $T_k$.

Using the above techniques, the distributed dual link algorithm for TDD networks is shown as Algorithm 2.
Algorithm 2 Distributed Dual Link Algorithm for TDD network

01. $T_i$ initialize $\Sigma_i$, s.t. $\sum_{l=1}^{L} \text{Tr}(\Sigma_l) = P_T$
02. Repeat
   (Forward Link)
   03. All $T_i$ transmits their pilots $X_i = [\tilde{x}_{i,1}, \tilde{x}_{i,2}, \ldots \tilde{x}_{i,N}]$ at the same time
   04. Each $R_k$ estimates $A_{k,k} = H_{k,k} \Sigma_k H_{k,k}^H$ and $\Omega_k = I + \sum_{l \neq k} H_{k,l} \Sigma_l H_{k,l}^H$
   05. Each $R_k$ calculates $\tilde{\Sigma}_k = w_k \left( \Omega_k^{-1} - (A_{k,k} + \Omega_k)^{-1} \right)$ and $\hat{p}_k = \text{tr} \left( \tilde{\Sigma}_k \right)$
   06. For $k = 1, 2, \ldots, L$
   07. $R_k$ broadcasts $\hat{p}_k$ to each other
08. Next $k$
09. Each $R_k$ calculates $\bar{\Sigma}_k = \frac{p_k \hat{p}_k}{\sum_{i=1}^{L} p_i} \tilde{\Sigma}_k$
   (Reverse Link)
10. All $R_k$ transmits their pilots $\bar{X}_k = [\tilde{x}_{k,1}, \tilde{x}_{k,2}, \ldots \tilde{x}_{k,N}]$ at the same time
11. Each $T_i$ estimates $\hat{A}_{i,l} = H_{i,l}^H \tilde{\Sigma}_l H_{i,l}$ and $\hat{\Omega}_l = I + \sum_{k \neq l} H_{k,l}^H \tilde{\Sigma}_k H_{k,l}$
12. Each $T_i$ calculates $\hat{\Sigma}_l = w_l \left( \hat{\Omega}_l^{-1} - (\hat{A}_{i,l} + \hat{\Omega}_l)^{-1} \right)$ and $p_l = \text{tr} \left( \hat{\Sigma}_l \right)$
13. For $l = 1, 2, \ldots, L$
14. $T_i$ broadcasts $p_l$ to each other
15. Next $l$
16. Each $T_i$ calculates $\Sigma_l = \frac{p_l p_i}{\sum_{i=1}^{L} p_i} \hat{\Sigma}_l$
17. Until convergence or max iteration number

4.3. Conclusion

The distributed dual link algorithm of a TDD network performs the iterations of equation (4.2.1, 4.2.2) on physical forward and reverse link channels instead of virtual reverse link channels, by sending pilot signals on both links. It replaces direct and cross channel information feedbacks with iterations of forward and reverse link pilots training whose complexity grows linearly with the number of users in the network. By totally avoiding channel state knowledge feedback, the distributed dual link algorithm has significant lower signaling overhead compared to the traditional methods, especially in networks with large number of interfering users.

It needs to be mentioned that in reality, although we can properly choose the pilots to ensure orthogonality among different antennas of the same transmitter/receiver, it’s difficult
to maintain pilot orthogonality between different users due to symbol timing mismatch especially in uplink cases. So a practical way is using long scramble code to keep approximate pilot orthogonality between different transmitters/receivers.
CHAPTER 5

Antenna Calibration for TDD Channel Reciprocity

5.1. Introduction

Network densification has been used to accommodate the explosive growth of wireless data traffic. However, interference becomes severe due to less scatters in shrinking cells. We have proposed a joint beamforming algorithm, the Dual Link algorithm [11], for optimal interference management in multiple-input-multiple-output (MIMO) interference networks. The algorithm has very fast and monotonic convergence. Taking advantage of the forward and reverse channel reciprocity of time division duplex (TDD) channels, the Dual Link algorithm is very suitable for distributed implementation. It requires little coordination among the nodes, while the overhead of coordination has hindered other joint beamforming algorithms.

However, the real TDD channels are not reciprocal because the transmit and receive RF chains are different in a transceiver. Various channel calibration methods have been studied to estimate and compensate the differences [4, 8, 22]. A novel internal calibration method was proposed in [22] that works for multiple-input-single-output (MISO) systems and does not require feedback between transmitters and receivers.

In this chapter, inspired by the internal calibration method, we present a simple method of channel calibration for MIMO interference networks that is essential to the distributed implementation of the Dual Link algorithm [11] and other algorithms that require reciprocity.

5.2. Internal Calibration

The internal calibration method in [22] works as follows. Let $h_{i,T_i \rightarrow j,R_k} = h_{j,R_k \rightarrow i,T_i} \in \mathbb{C}$ be the reciprocal wireless channel from $i$-th antenna of $T_i$ to $j$-th antenna of $R_k$, $t_{i,T_i} \in \mathbb{C}$ be the channel of the transmit RF chain of the $i$-th antenna of $T_i$, and $r_{j,R_k} \in \mathbb{C}$ be the channel...
of the receive RF chain of $j$-th antenna of $R_k$. The overall forward link channel is

$$\tilde{h}_{i,T_i \rightarrow j,R_k} = t_{i,T_i} \cdot h_{i,T_i \rightarrow j,R_k} \cdot r_{j,R_k}.$$  

Similarly, the overall reverse link channel from $j$-th antenna of $R_k$ to $i$-th antenna of $T_i$ is

$$\tilde{h}_{j,R_k \rightarrow i,T_i} = t_{j,R_k} \cdot h_{j,R_k \rightarrow i,T_i} \cdot r_{i,T_i}.$$  

By letting the antennas of a transceiver take turns to transmit pilots while other antennas of the same transceiver listen, the calibration coefficients

$$(5.2.1) \quad b_{1,T_i \rightarrow j,T_j} = \frac{\tilde{h}_{1,T_i \rightarrow j,T_j}}{\tilde{h}_{j,T_j \rightarrow 1,T_i}} = \frac{t_{1,T_i} \cdot r_{j,T_j}}{r_{1,T_i} \cdot t_{j,T_j}}, \quad \forall j,$$

can be measured internally. For a MISO system, $\tilde{h}_{1,R \rightarrow i,T_i}$ can be measured from reverse link pilot training. Then, the forward link $l$’s channel can be learned by

$$\tilde{h}_{i,T_i \rightarrow 1,R_l} = \frac{\tilde{h}_{1,R_l \rightarrow i,T_i} t_{1,R_k}}{b_{1,T_i \rightarrow i,T_i} t_{1,R_k}}$$

up to a factor $\frac{r_{1,R_k}}{r_{i,R_k}}$, which does not depend on $i$ and thus does not affect the forward link beamforming.

**5.3. Internal Channel Calibration for Distributed Dual Link Algorithm**

As we’ve shown in the previous chapter, the distributed dual link algorithm relies heavily on the TDD channel reciprocity for the reverse link pilot training. The distributed Dual Link algorithm of a TDD network performs the iterations of following equations:

$$(5.3.1) \quad \hat{\Sigma}_l \leftarrow \frac{P_{T_i} \text{tr} \left( \Omega_l^{-1} - \left( \Omega_l + H_{l,l} \Sigma_l H_{l,l}^\dagger \right)^{-1} \right) }{\sum_{l=1}^{L} w_l \text{tr} \left( \hat{\Omega}_l^{-1} - \left( \hat{\Omega}_l + \hat{H}_{l,l} \hat{\Sigma}_l \hat{H}_{l,l}^\dagger \right)^{-1} \right)}$$

$$(5.3.2) \quad \Sigma_l \leftarrow \frac{P_{T_i} \text{tr} \left( \hat{\Omega}_l^{-1} - \left( \hat{\Omega}_l + \hat{H}_{l,l}^\dagger \hat{\Sigma}_l \hat{H}_{l,l} \right)^{-1} \right) }{\sum_{l=1}^{L} w_l \text{tr} \left( \hat{\Omega}_l^{-1} - \left( \hat{\Omega}_l + \hat{H}_{l,l}^\dagger \hat{\Sigma}_l \hat{H}_{l,l} \right)^{-1} \right)}$$
on physical forward and reverse link channels, not on virtual reverse link channels, by sending pilot signals on both links. The terms $\Omega_l, \Omega_i + \mathbf{H}_{l,i} \Sigma_i \mathbf{H}_{l,i}^\dagger$, and their reverse link counterparts can be estimated locally, because the channels have done the summation of the signals for free regardless of the number of users. But it does require the reciprocity of not only the desired MIMO channels $\mathbf{H}_{l,i}$ but also the interfering channels $\mathbf{H}_{l,k}$. Simply applying the method of [22] by transmitting pilots on each antenna does not work here.

The simple and key idea of compensating the RF chain differences for MIMO interference channel is that the compensation needs to be done before the pilot and information transmission because the signals from different links will be mixed together at a receiver with noise. The new method is to multiply the calibration coefficients before sending pilots or information signals on both forward and reverse links. The resulting overall forward channel from $i$-th antenna of $T_l$ to $j$-th antenna of $R_k$ becomes

$$
\tilde{h}_{i,T_l \rightarrow j,R_k} = b_{i,T_l \rightarrow j,T_l} \cdot t_{i,T_l} \cdot h_{i,T_l \rightarrow j,R_k} \cdot r_{j,R_k}
$$

(5.3.3)

Similarly, the corresponding overall reverse link channel from $j$-th antenna of $R_k$ to $i$-th antenna of $T_l$ becomes

$$
\tilde{h}_{j,R_k \rightarrow i,T_l} = b_{j,R_k \rightarrow j,R_k} \cdot t_{j,R_k} \cdot h_{j,R_k \rightarrow i,T_l} \cdot r_{i,T_l}
$$

(5.3.4)

$$
= \frac{t_{1,R_k}}{r_{1,R_k}} \left( r_{j,R_k} \cdot h_{j,R_k \rightarrow i,T_l} \cdot t_{i,T_l} \right)
$$

$$
= b_{j,R_k \rightarrow j,T_l} \tilde{h}_{i,T_l \rightarrow j,R_k}.
$$

Note that the terms in the brackets are equal since $h_{i,T_l \rightarrow j,R_k} = h_{j,R_k \rightarrow i,T_l}$. Then the effective forward link channel from $T_l$ to $R_k$ is $\mathbf{H}_{k,l} = \left[ \tilde{h}_{i,T_l \rightarrow j,R_k} \right]_{j,i}$ and the effective reverse link channel from $R_k$ to $T_l$ is $\mathbf{H}_{l,k} = b_{1,R_k \rightarrow 1,T_l} \mathbf{H}_{k,l}^T$.

If all $T_l$’s are the same kind of hardware and all $R_k$’s are the same kind of hardware, then $|b_{1,R_k \rightarrow 1,T_l}|$ is approximately constant for different $k$ and $l$. It only affects the power constraint by a constant factor. Let $b_{1,R_k \rightarrow 1,T_l} = |b_{1,R_k \rightarrow 1,T_l}| e^{j\phi_{kl}}$. In the dual link algorithm,
replacing $\mathbf{H}_{k,l}$ by $\hat{\mathbf{H}}_{k,l}$ and replacing $\mathbf{H}_{k,l}^\dagger$ by $\hat{\mathbf{H}}_{l,k}^*$, the extra phase part $e^{j\phi_{k,l}}$ will not affect the distributed Dual Link algorithm or other algorithms that require duality, because
\[
\left(e^{-j\phi_{k,l}}\hat{\mathbf{H}}_{k,l}^\dagger\right)\hat{\Sigma}_k\left(e^{j\phi_{k,l}}\hat{\mathbf{H}}_{k,l}\right) = \hat{\mathbf{H}}_{k,l}^\dagger\hat{\Sigma}_k\hat{\mathbf{H}}_{k,l}.
\]

5.4. Conclusion

In this chapter, a channel calibration method for MIMO interference channel is provided to achieve TDD channel reciprocity if all $T_l$’s are the same kind of hardware and all $R_k$’s are the same kind of hardware. If $T_l$’s are different kind of hardware or $R_k$’s are different kind of hardware, $\frac{t_{l,T_l}}{r_{l,T_l}}$ of (5.3.3) and $\frac{t_{l,R_k}}{r_{l,R_k}}$ of (5.3.4) needs to be estimated and canceled.
CHAPTER 6

Robust Algorithm Design for Channel Uncertainty

6.1. Introduction

In a realistic scenario, the channel knowledge is generally imperfect. On one hand, the time-vary nature of the wireless channel will always introduce channel uncertainty over time. Thus, it requires us to estimate the channels, feedback necessary information, calculate and distribute optimized signal structures in a timely manner so that there will be enough time left for the actual data transmission in the limited channel coherent time window. On the other hand, estimation errors can also be viewed as part of the channel uncertainty since it also brings difference between the real channel and the estimated channel. The pilots used for channel estimation can be contaminated with background noise, other users’ pilots or even interference from other networks. Although increasing the pilot length can improve the accuracy, it will also proportionally increase the training time, which results in high signaling overheads and limits the data transmission.

The dual link algorithm has fast convergence speed and low complexity, which saves computation time. The distributed implementation and the corresponding covariance matrices estimation method can significantly lower the signaling overhead caused by channel state knowledge feedback in networks with large number of interfering users. Both our previous results help to secure enough data transmission time in the channel coherent time window while updating the optimal signal structure to improve the data rate. But in order to maintain reliable data transmission, robust algorithm design should be further applied to optimize the average performance under channel uncertainty.
6.2. Problem Formulation

Assume $\hat{H}_{l,k}$ is the known channel matrix between transmitter $T_k$ and transmitter $R_l$, it can be an out-dated channel from the past time slots or an estimated channel with i.i.d. zero mean Gaussian entries which combines all the effects of time-varying and estimation errors. Thus, the real accurate channel is defined as

\begin{equation}
H_{l,k} = \sqrt{\rho} \hat{H}_{l,k} + \sqrt{1 - \rho} \tilde{H}_{l,k}
\end{equation}

where $0 \leq \rho \leq 1$ indicates the level of channel uncertainty.

The achievable rate of forward link $l$ under a realization $H$ is

\begin{equation}
\mathcal{I}_{l,H} (\Sigma_{1:L}, H) = \log \left| I + H_{l,l} \Sigma_l H_{l,l}^\dagger \Omega_{l,H}^{-1} \right|
\end{equation}

where $\Sigma_l$ is the input covariance matrix; and $\Omega_{l,H}$ is the interference-plus-noise covariance matrix of the $l^{th}$ forward link under the channel realization $H$:

\begin{equation}
\Omega_{l,H} = I + \sum_{k=1, k \neq l}^{L} H_{l,k} \Sigma_k H_{l,k}^\dagger.
\end{equation}

Accordingly, the achievable rate of reverse link $l$ under the realization $H$ is

\begin{equation}
\hat{\mathcal{I}}_{l,H} (\hat{\Sigma}_{1:L}, H) = \log \left| I + \hat{H}_{l,l} \hat{\Sigma}_l \hat{H}_{l,l}^\dagger \hat{\Omega}_{l,H}^{-1} \right|
\end{equation}

where $\hat{\Sigma}_l$ is the input covariance matrix; and $\hat{\Omega}_{l,H}$ is the interference-plus-noise covariance matrix of the $l^{th}$ reverse link under the channel realization $H$:

\begin{equation}
\hat{\Omega}_{l,H} = I + \sum_{k=1, k \neq l}^{L} \hat{H}_{k,l} \hat{\Sigma}_k \hat{H}_{k,l}.
\end{equation}

Our goal is to optimize the ergodic weighted sum-rate over the same $\Sigma_l$ for all channel realization:
(6.2.6) \textbf{EWSRM\_TP:} \quad \max_{\Sigma_{l,l}} \quad \mathbb{E}_{H} \left[ \sum_{l=1}^{L} w_l \mathcal{I}_l (\Sigma_{1:L}, H) \right] \\
\text{s.t.} \quad \Sigma_l \geq 0, \forall l, \\
\sum_{l=1}^{L} \text{Tr} (\Sigma_l) \leq P_T

where \( w_l > 0 \) is the weight for link \( l \).

6.3. Ergodic WMMSE Algorithm

The WMMSE algorithm proposed in [24] can be modified to solve problem (6.2.6).

The MSE matrix of link \( l \) under the channel realization \( H \) is given by

\[
\mathbf{E}_{l,H} = \left( \mathbf{I} - \mathbf{U}_{l,H}^\dagger \mathbf{H}_{l,l} \mathbf{V}_l \right) \left( \mathbf{I} - \mathbf{U}_{l,H}^\dagger \mathbf{H}_{l,l} \mathbf{V}_l \right)^\dagger + \mathbf{U}_{l,H}^\dagger \mathbf{\Omega}_{l,H} \mathbf{U}_{l,H}
\]

where \( \mathbf{U}_{l,H} \) is the receive beamformer of link \( l \) under the channel realization \( H \), \( \mathbf{V}_l \) is the transmit beamformer of link \( l \), and \( \Sigma_l = \mathbf{V}_l \mathbf{V}_l^\dagger \).

Fixing all the transmit beamformers \( \mathbf{V}_l \)'s and minimizing weighted sum-MSE \( \sum_{l=1}^{L} \text{tr} (\mathbf{W}_l \mathbf{E}_l) \) lead to the MMSE receiver:

(6.3.1) \quad \mathbf{U}_{l,H}^{\text{mmse}} = \left( \mathbf{H}_{l,l} \mathbf{V}_l \mathbf{V}_l^\dagger \mathbf{H}_{l,l}^\dagger + \mathbf{\Omega}_{l,H} \right)^{-1} \mathbf{H}_{l,l} \mathbf{V}_l.

Using the MMSE receiver, the corresponding MES matrix is

\[
\mathbf{E}_{l,H}^{\text{mmse}} = \mathbf{I} - \mathbf{V}_l^\dagger \mathbf{H}_{l,l}^\dagger \left( \mathbf{H}_{l,l} \mathbf{V}_l \mathbf{V}_l^\dagger \mathbf{H}_{l,l}^\dagger + \mathbf{\Omega}_{l,H} \right)^{-1} \mathbf{H}_{l,l} \mathbf{V}_l
\]

\[
= \left( \mathbf{I} + \mathbf{V}_l^\dagger \mathbf{H}_{l,l}^\dagger \mathbf{\Omega}_{l,H}^{-1} \mathbf{H}_{l,l} \mathbf{V}_l \right)^{-1}
\]
THEOREM 10. Let $W_{l,H} \succeq 0$ be a weight matrix for receiver $l$ under channel realization $H$. The problem

$$
(6.3.2) \quad \min_{U_{l,H}, V_l} \quad \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \left( \text{tr} (W_{l,H} E_{l,H}) - \log \det (W_{l,H}) \right) \right]
$$

$W_{l,H} \succeq 0$

$$
\sum_l \text{tr} (V_l V_l^\dagger) \leq P_T
$$

is equivalent to problem (6.2.6), in the sense that the optimal solution $V_l$ for the two problems are the same ($\Sigma_l = V_l V_l^\dagger$).

PROOF. The optimal $U_{l,H}$ for minimizing (6.3.2) is given by $U_{l,H}^{\text{mmse}}$ in (6.3.1). Fixing $U_{l,H}$ and $V_l$, the objective function in (6.3.2) is convex with respect to $W_{l,H}$. By checking the first order optimality condition for $W_{l,H}$, we get

$$
W_{l,H}^{\text{opt}} = E_{l,H}^{-1}
$$

Put in optimal $U_{l,H}$ and $W_{l,H}$, we have an equivalent optimization problem:

$$
\max \quad \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \log \det \left( (E_{l,H}^{\text{mmse}})^{-1} \right) \right]
$$

$\sum_l \text{tr} (V_l V_l^\dagger) \leq P_T$

$$
= \max \quad \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \log \det \left( I + V_l^\dagger H_{ll}^{-1} \Omega_{l,l}^{-1} H_{ll} V_l \right) \right]
$$

$\sum_l \text{tr} (V_l V_l^\dagger) \leq P_T$

$$
= \max \quad \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \log \det \left( I + H_{ll} V_l V_l^\dagger H_{ll}^{-1} \Omega_{l,l}^{-1} \right) \right]
$$

$\sum_l \text{tr} (V_l V_l^\dagger) \leq P_T$

$$
= R_0
$$

which is the same as problem (6.2.6). \hfill \blacksquare

Fixing $W_{l,H}$ and $U_{l,H}$, problem (6.3.2) becomes a convex quadratic optimization problem:

$$
\min_{\sum_l \text{tr} (V_l V_l^\dagger) \leq P_T} \quad \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \left( \text{tr} \left( W_{l,H} \left( I - U_{l,H}^\dagger H_{ll} V_l \right) \left( I - U_{l,H}^\dagger H_{ll} V_l \right)^\dagger \right.ight.ight.

+ U_{l,H}^\dagger \Omega_{l,H} U_{l,H} \left) \right) - \log \det (W_{l,H}) \right]
$$
which can be solved by Lagrange multipliers method. The optimal $V_l$ is given by

$$
V_l^{\text{opt}} = \left( \mu I + \mathbb{E}_H \left[ \sum_{k=1}^{L} w_k \left( H_{k,l} U_k H W_k U_{k,H} H_{k,l} \right) \right] \right)^{-1} \cdot \mathbb{E}_H \left[ w_l H_{l,l}^\dagger U_{l,H} W_{l,H}^\dagger \right]
$$

where the Lagrange multiplier $\mu$ is chosen s.t. $\sum_l \text{tr} \left( V_l V_l^\dagger \right) \leq P_T$ via bisection method.

The corresponding ergodic WMMSE algorithm is shown in Algorithm (3). Although it’s a centralized algorithm and is impractical in a large scale MIMO interference network, it can serve as a benchmark to evaluate our distributed ergodic dual link algorithm.

**Algorithm 3** Ergodic WMMSE algorithm for MIMO B-MAC interference network

1. Initialize $V_l$'s s.t. $\text{tr} \left( V_l V_l^\dagger \right) = \frac{P_T}{L}$
2. Repeat
3. $V_l' \leftarrow V_l$, $\forall l$
4. $U_{l,H} \leftarrow \left( H_{l,l} V_l V_l^\dagger H_{l,l}^\dagger + \Omega_{l,H} \right)^{-1} H_{l,l} V_l$, $\forall l, H$
5. $W_{l,H} \leftarrow \left( I - U_{l,H}^\dagger H_{l,l} V_l \right)^{-1}$, $\forall l, H$
6. $V_l \leftarrow \left( \mu I + \mathbb{E}_H \left[ \sum_{k=1}^{L} w_k \left( H_{k,l} U_k H W_k U_{k,H} H_{k,l} \right) \right] \right)^{-1} \cdot \mathbb{E}_H \left[ w_l H_{l,l}^\dagger U_{l,H} W_{l,H}^\dagger \right]$, $\forall l, \mu$ is chosen s.t. $\sum_l \text{tr} \left( V_l V_l^\dagger \right) \leq P_T$
7. Until $\sum_{i,j} \left| V_i' - V_i \right|_{i,j}^2 \leq \epsilon$

## 6.4. Ergodic Dual Link Algorithm

In order to apply the dual link algorithm to the ergodic problem (6.2.6), we need to take advantage of rate duality between the physical forward and reverse links. But in problem (6.2.6), $\Sigma_t$ are fixed for all the channel realizations, and such rate duality no longer exists. Thus, we transform problem (6.2.6) into the following equivalent optimization problem by using separate $\Sigma_{t,H}$ for different channel realization $H$ and adding equality constraints to make them equal:
\[(6.4.1) \text{DEWSRM TP: } \max_{\Sigma_{1:L,H}} \mathbb{E}_{H}\left[ \sum_{l=1}^{L} w_l I_l (\Sigma_{1:L,H}, H) \right]
\]
\[
\text{s.t. } \Sigma_{l,H} \succeq 0, \forall l, \\
\sum_{l=1}^{L} \text{Tr}(\Sigma_{l,H}) \leq P_T \\
\Sigma_{l,H} = \Sigma_{l,H_0}, \forall H \neq H_0
\]

where $H_0$ is one of the channel realizations served as the reference. In this way, we break the problem (6.4.1) into many small weighted sum-rate maximization problems under each channel realization $H$ and can optimize $\Sigma_{l,H}$ in the same distributed fashion as the distributed dual link algorithm we proposed in the previous chapter. The task left is finding the method to enforcing the constraint $\Sigma_{l,H} = \Sigma_{l,H_0}, \forall H \neq H_0$ while keeping the convergence property.

We’ll apply the similar techniques used to prove the convergence of the dual link algorithm to solve problem.

By relaxing $\Omega_{l,H}$, the problem (6.4.1) can be further rewritten into an equivalent problem:

\[(6.4.2) \max_{\Sigma_{1:L,H}, \Omega_{1:L,H}} \mathbb{E}_{H}\left[ \sum_{l=1}^{L} w_l \left( \log \left| \Omega_{l,H} + H_{l,l} \Sigma_{l,H} H_{l,l}^H \right| - \log |\Omega_{l,H}| \right) \right]
\]
\[
\text{s.t. } \Sigma_{l,H} \succeq 0, \forall l, \sum_{l=1}^{L} \text{Tr}(\Sigma_{l,H}) \leq P_T, \\
\Omega_{l,H} = I + \sum_{k \neq l} H_{l,k} \Sigma_{k,H} H_{l,k}^H, \forall l, \\
\Sigma_{l,H} = \Sigma_{l,H_0}, \forall H \neq H_0
\]
Consider the Lagrangian of the above problem

\[
F(\Sigma, \Omega, \Lambda, \mu, \Delta) = \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \left( \log \left| \Omega_{l,H} + H_{l,l} \Sigma_{l,H} \Sigma_{l,H}^\dagger \right| - \log \left| \Omega_{l,H} \right| \right) \right] \\
+ \sum_{H} \mu_H \left( P_T - \sum_{l=1}^{L} \text{Tr}(\Sigma_{l,H}) \right) \\
+ \sum_{H} \sum_{l=1}^{L} \text{Tr} \left( \Lambda_{l,H} \left( \Omega_{l,H} - I - \sum_{k \neq l} H_{l,k} \Sigma_{k,H} \Sigma_{k,H}^\dagger \right) \right) \\
+ \sum_{H \neq H_0} \sum_{l=1}^{L} \text{Tr} \left( \Delta_{l,H} (\Sigma_{l,H} - \Sigma_{l,H_0}) \right)
\]

We’ll combine the convex-concave approach used in the convergence proof of the dual link algorithm with gradient descent method to find the stationary point. In the outer loop, we update \( \Delta_{l,H} \) using gradient descent method:

\[
\Delta_{l,H}^{(n+1)} = \Delta_{l,H}^{(n)} - \sigma \nabla_{\Delta_{l,H}} F
\]

where \( \sigma \) is a positive step size. In the inner loop, we treat \( \Delta_{l,H} \) as constant and use dual link algorithm to maximize \( F \).

**Remark 11.** Function \( F \) is concave in \( \Sigma \) and convex in \( \Omega \).

**Remark 12.** Denote the ergodic weighted sum-rate function by \( V(\Sigma) \), i.e.,

\[
V(\Sigma) = \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \left( \log \left| \Omega_{l,H} + H_{l,l} \Sigma_{l,H} \Sigma_{l,H}^\dagger \right| - \log \left| \Omega_{l,H} \right| \right) \right].
\]

We have

\[
F(\Sigma, \Omega, \Lambda, \mu, \Delta) = V(\Sigma)
\]

for arbitrary \((\Lambda, \mu, \Delta)\), given any pair \((\Sigma, \Omega)\) satisfies

\[
\sum_{l=1}^{L} \text{Tr}(\Sigma_{l,H}) = P_T,
\]
\[ \Omega_{l,H} = I + \sum_{k \neq l} H_{l,k} \Sigma_{k,l} H_{l,k}^\dagger, \]

and

\[ \Sigma_{l,H} = \Sigma_{l,H_0}. \]

**Remark 13.** \( F(\Sigma, \Omega, \Lambda, \mu, \Delta) \) has a scaling invariance property, i.e.,

\[ F(\Sigma, \Omega, \Lambda, \mu, \Delta) = F\left( \frac{1}{\alpha} \Sigma, \frac{1}{\alpha} \Omega, \alpha \Lambda, \alpha \mu, \tilde{\Delta} \right) \]

for any \( \alpha \), if

\[ \sum_{l=1}^{L} \text{Tr}\left( \frac{1}{\mu_H} \Lambda_{l,H} \right) = P_T \]

and \( \tilde{\Delta} \) is chosen s.t.

\[ \sum_{H \neq H_0} \frac{1}{\alpha_H} \sum_{l=1}^{L} \text{Tr}\left( \Delta_{l,H} (\Sigma_{l,H} - \Sigma_{l,H_0}) \right) = \sum_{H \neq H_0} \sum_{l=1}^{L} \text{Tr}\left( \Delta_{l,H} (\Sigma_{l,H} - \Sigma_{l,H_0}) \right). \]

From first order optimality conditions, \( \frac{\partial F}{\partial \Omega_{l,H}} = 0 \) and \( \frac{\partial F}{\partial \Sigma_{l,H}} = 0 \), we have,

(6.4.3) \[ \Lambda_{l,H} = w_l \left( \Omega_{l,H}^{-1} - \left( \Omega_{l,H} + H_{l,l} \Sigma_{l,H} H_{l,l}^\dagger \right)^{-1} \right), \]

(6.4.4) \[ \mu_H I + \sum_{k \neq l} H_{k,l}^\dagger \Lambda_{k,H} H_{k,l} + \Delta_{l,H} = w_l H_{l,l}^\dagger \left( \Omega_{l,H} + H_{l,l} \Sigma_{l,H} H_{l,l}^\dagger \right)^{-1} H_{l,l}. \]

where \( \Delta_{l,H_0} \triangleq - \sum_{H \neq H_0} \Delta_{l,H}. \)

Define \( \Phi_{l,H} \triangleq \mu_H I + \sum_{k \neq l} H_{k,l}^\dagger \Lambda_{k,H} H_{k,l} + \Delta_{l,H}, \) \( \hat{\Sigma}_{l,H} \triangleq \frac{1}{\mu_H} \Lambda_{l,H}, \) \( \hat{\Omega}_{l,H} \triangleq I + \sum_{k \neq n} H_{k,l}^\dagger \hat{\Sigma}_{k,H} H_{k,l}. \)

For feasible \( (\mu_H, \Phi_{l,H}) \), an explicit solution to this system of equations is given by [2]

(6.4.5) \[ \Sigma_{l,H} = w_l \left( \Phi_{l,H}^{-1} - \left( \Phi_{l,H} + H_{l,l}^\dagger \Lambda_{l,H} H_{l,l} \right)^{-1} \right), \]

(6.4.6) \[ \Omega_{l,H} = w_l H_{l,l} \left( H_{l,l}^\dagger \Lambda_{l,H} H_{l,l} + \Phi_{l,H} \right)^{-1} H_{l,l}^\dagger. \]

A feasible \( \mu_H \geq 0 \) is chosen s.t. the total power constraint is satisfied. But a feasible \( \Phi_{l,H} \) requires \( \Phi_{l,H} \geq 0 \), which may not be true given certain choice of \( \Delta_{l,H}. \) Therefore, when
we update $\Delta_{t, H}$ in the outer loop:

$$
\Delta_{t, H}^{(n+1)} = \Delta_{t, H}^{(n)} - \sigma \nabla \Delta_{t, H} F,
$$

$$
= \Delta_{t, H}^{(n)} - \sigma \left( \Sigma_{t, H}^{(n)} - \Sigma_{t, H_0}^{(n)} \right)^*.
$$

$\Delta_{t, H}^{(n+1)}$ needs to be projected to a feasible subspace so that $\Phi_{t, H} \succeq 0$ will be satisfied:

$$(6.4.7) \quad \tilde{\Delta}_{t, H}^{(n+1)} = \operatorname{argmin} \sum_{H} \sum_{l} \left\| \Delta_{t, H}^{(n+1)} - \Delta_{t, H}^{(n+1)} \right\|^2_2$$

s.t. $\mu_H I + \sum_{k \neq l} H_{k,l}^H \Lambda_{k,H} H_{k,l} + \tilde{\Delta}_{t, H}^{(n+1)} \succeq 0$

which can be numerically solved by semi-positive definite programming.

In the inner loop, given the projected $\tilde{\Delta}_{t, H}^{(n+1)}$, Lagrangian function $F(\Sigma, \Omega, \Lambda, \mu, \Delta)$ will monotonically increase.

We start with $\Sigma_{t, H}^{(n)}$ from last iteration, which satisfy

$$(6.4.8) \quad \sum_{l=1}^L \operatorname{Tr}(\Sigma_{t, H}^{(n)}) = P_T, \forall H,$$

$\Omega_{t, H}^{(n)}$ is calculated as defined:

$$(6.4.9) \quad \Omega_{t, H} = I + \sum_{k \neq l} H_{k,l} \Sigma_{k,H} H_{k,l}^H.$$

Since constraints (6.4.8) and (6.4.9) are both satisfied, we can choose arbitrary $\Lambda$ and $\mu$ without changing the value of $F(\Sigma_{t, H}^{(n)}, \Omega_{t, H}^{(n)}, \Lambda, \mu, \Delta_{t, H}^{(n)}).

Next, $\Lambda_{t, H}^{(n)}$ is calculated using (6.4.3). Since (6.4.3) implies $\frac{\partial F}{\partial \Omega_{t, H}} = 0$ and $F$ is convex in $\Omega$, we have

$$F(\Sigma_{t, H}^{(n)}, \Omega_{t, H}^{(n)}, \Lambda_{t, H}^{(n)}, \mu, \Delta_{t, H}^{(n)}) \leq F(\Sigma_{t, H}^{(n)}, \Omega_{t, H}^{(n)}, \Lambda_{t, H}^{(n)}, \mu, \Delta_{t, H}^{(n)}), \forall \Omega.$$

Then, $\mu_{t, H}^{(n)}$ is chosen s.t.

$$(6.4.10) \quad \frac{1}{\mu_{t, H}^{(n)}} \sum_{l=1}^L \operatorname{Tr}(\Lambda_{t, H}^{(n)}) = P_T$$
Let
\[
\Phi^{(n)}_{l,H} = \mu^{(n)}_H \mathbf{I} + \sum_{k \neq l} H^\dagger_{k,l} \Lambda^{(n)}_{H} H_{k,l} + \Delta^{(n)}_{l,H}.
\]

Since in the previous outer loop, \( \Delta^{(n)}_{l,H} \) is projected to ensure the feasibility, we have \( \Phi^{(n)}_{l,H} \succeq 0 \).

Then \( \tilde{\Sigma}^{(n+1)}_{l,H} \) and \( \tilde{\Omega}^{(n+1)}_{l,H} \) are calculated as explicit solution to (6.4.3) and (6.4.4):
\[
(6.4.11) \quad \tilde{\Sigma}^{(n+1)}_{l,H} = w_l \left( \left( \Phi^{(n)}_{l,H} \right)^{-1} - \left( \Phi^{(n)}_{l,H} + H^\dagger_{l,t} \Lambda^{(n)}_{H} H_{l,t} \right)^{-1} \right)
\]
\[
(6.4.12) \quad \tilde{\Omega}^{(n+1)}_{l,H} = w_l H_{l,t} \left( H^\dagger_{l,t} \Lambda^{(n)}_{H} H_{l,t} + \Phi^{(n)}_{l,H} \right)^{-1} H^\dagger_{l,t}
\]

Notice that (6.4.11) and (6.4.12) together imply
\[
(6.4.13) \quad \frac{\partial F(\tilde{\Sigma}^{(n+1)}_{l,H}, \tilde{\Omega}^{(n+1)}_{l,H}, \Lambda^{(n)}_{l,H}, \mu^{(n)}_H, \Delta^{(n)}_{l,H})}{\partial \Omega^{(n+1)}_{l,H}} \bigg|_{\Omega^{(n+1)}_{l,H} = \tilde{\Omega}^{(n+1)}_{l,H}} = 0
\]
and
\[
(6.4.14) \quad \frac{\partial F(\Sigma^{(n)}_{l,H}, \Omega^{(n+1)}_{l,H}, \Lambda^{(n)}_{l,H}, \mu^{(n)}_H, \Delta^{(n)}_{l,H})}{\partial \Sigma^{(n+1)}_{l,H}} \bigg|_{\Sigma^{(n+1)}_{l,H} = \tilde{\Sigma}^{(n+1)}_{l,H}} = 0.
\]

Because of (6.4.14) and that fact that \( \Sigma \) is concave in \( F \), we have
\[
F(\Sigma^{(n)}_{l,H}, \Omega^{(n+1)}_{l,H}, \Lambda^{(n)}_{l,H}, \mu^{(n)}_H, \Delta^{(n)}_{l,H}) \leq F(\tilde{\Sigma}^{(n+1)}_{l,H}, \tilde{\Omega}^{(n+1)}_{l,H}, \Lambda^{(n)}_{l,H}, \mu^{(n)}_H, \Delta^{(n)}_{l,H}).
\]

Then since (6.4.13) holds, and \( \Omega \) is convex in \( F \), we have
\[
F(\Sigma^{(n+1)}_{l,H}, \Omega^{(n+1)}_{l,H}, \Lambda^{(n)}_{l,H}, \mu^{(n)}_H, \Delta^{(n)}_{l,H}) \leq F(\tilde{\Sigma}^{(n+1)}_{l,H}, \tilde{\Omega}^{(n+1)}_{l,H}, \Lambda^{(n)}_{l,H}, \mu^{(n)}_H, \Delta^{(n)}_{l,H}), \forall \tilde{\Omega}^{(n+1)}_{l,H}.
\]

Let
\[
\alpha_H = \frac{P_T}{\sum_{l=1}^L \text{Tr}(\bar{\Sigma}^{(n+1)}_{l,H})}
\]
and

\[
\begin{align*}
\Sigma_{l,H}^{(n+1)} &= \alpha_H \Sigma_{l,H}^{(n+1)} \\
\Omega_{l,H}^{(n+1)} &= \mathbf{I} + \sum_{k \neq l} H_{l,k} \Sigma_{k,H}^{(n)} H_{l,k}^T \\
\tilde{\Omega}_{l,H}^{(n+1)} &= \frac{1}{\alpha_H} \Omega_{l,H}^{(n+1)}
\end{align*}
\]

also adjust \( \Delta_{l,H}^{(n)} \) to \( \tilde{\Delta}_{l,H}^{(n)} = \beta_{l,H} \Delta_{l,H}^{(n)} \) s.t.

\[
\sum_{H \neq H_0} \sum_{l=1}^{L} \text{Tr} \left( \tilde{\Delta}_{l,H}^{(n)} \left( \Sigma_{l,H}^{(n+1)} - \Sigma_{l,H_0}^{(n+1)} \right) \right) = \sum_{H \neq H_0} \sum_{l=1}^{L} \text{Tr} \left( \Delta_{l,H}^{(n)} \left( \Sigma_{l,H}^{(n+1)} - \Sigma_{l,H_0}^{(n+1)} \right) \right).
\]

Then by remark 13, we have

\[
F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \Lambda_{l,H}^{(n)}, \mu_{l,H}^{(n)}, \Delta_{l,H}^{(n)}) = F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \alpha_H \Lambda_{l,H}^{(n)}, \alpha_H \mu_{l,H}^{(n)}, \tilde{\Delta}_{l,H}^{(n)})
\]

\[
= F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \Lambda, \mu, \tilde{\Delta}_{l,H}^{(n)}), \forall \Lambda, \mu.
\]

To summarize,

\[
F(\Sigma_{l,H}^{(n)}, \Omega_{l,H}^{(n)}, \Lambda, \mu, \Delta_{l,H}^{(n)}), \forall \Lambda, \mu
\]

\[
= F(\Sigma_{l,H}^{(n)}, \Omega_{l,H}^{(n)}, \Lambda_{l,H}^{(n)}, \mu_{l,H}^{(n)}, \Delta_{l,H}^{(n)})
\]

\[
\leq F(\Sigma_{l,H}^{(n)}, \Omega_{l,H}^{(n+1)}, \Lambda_{l,H}^{(n)}, \mu_{l,H}^{(n)}, \Delta_{l,H}^{(n)})
\]

\[
\leq F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \Lambda_{l,H}^{(n)}, \mu_{l,H}^{(n)}, \Delta_{l,H}^{(n)})
\]

\[
\leq F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \Lambda_{l,H}^{(n)}, \mu_{l,H}^{(n)}, \tilde{\Delta}_{l,H}^{(n)})
\]

\[
= F(\frac{1}{\alpha_H} \Sigma_{l,H}^{(n+1)}, \frac{1}{\alpha_H} \Omega_{l,H}^{(n+1)}, \Lambda_{l,H}^{(n)}, \mu_{l,H}^{(n)}, \tilde{\Delta}_{l,H}^{(n)})
\]

\[
= F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \alpha_H \Lambda_{l,H}^{(n)}, \alpha_H \mu_{l,H}^{(n)}, \tilde{\Delta}_{l,H}^{(n)})
\]

\[
= F(\Sigma_{l,H}^{(n+1)}, \Omega_{l,H}^{(n+1)}, \Lambda, \mu, \tilde{\Delta}_{l,H}^{(n)}), \forall \Lambda, \mu.
\]

After the inner loop, \( \tilde{\Delta}_{l,H}^{(n)} \) is updated and projected to \( \Delta_{l,H}^{(n+1)} \) in the outer loop which makes sure that \( \Phi_{l,H}^{(n+1)} \) is feasible for the next inner loop.
The ergodic dual link algorithm is shown in Algorithm 4.

**Algorithm 4 Ergodic Dual Link Algorithm**

1. Initialize $\Sigma_{l,H}$’s, s.t. $\sum_{l=1}^{L} \text{Tr} (\Sigma_{l,H}) = P_T$, $\Delta_{l,H}$’s, s.t. $\Phi_{l,H} \succ 0$.
2. $R \leftarrow \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \mathcal{I}_l (\Sigma_{l:L,H_0}, H) \right]$
3. Repeat
4. $R' \leftarrow R$
5. $\Omega_{l,H} \leftarrow I + \sum_{k \neq l} H_{l,k} \Sigma_{k,H} H_{l,k}^\dagger$
6. $\Lambda_{l,H} \leftarrow w_l \left( \Omega_{l,H}^{-1} - \left(\Omega_{l,H} + H_{l,l} \Sigma_{l,H} H_{l,l}^\dagger\right)^{-1} \right)$
7. $\mu_{H} \leftarrow \frac{1}{P_T} \sum_{l=1}^{L} \text{Tr}(\Lambda_{l,H})$
8. $\Phi_{l,H} \leftarrow \mu_{H} I + \sum_{k \neq l} H_{k,l} \Sigma_{k,H} H_{k,l}^\dagger + \Delta_{l,H}$
9. $\Sigma_{l,H} \leftarrow w_l \left( \Phi_{l,H}^{-1} - \left(\Phi_{l,H} + H_{l,l} \Sigma_{l,H} H_{l,l}^\dagger\right)^{-1} \right)$
10. $\alpha_{H} \leftarrow \frac{1}{P_T} \sum_{l=1}^{L} \text{Tr}(\Sigma_{l,H})$
11. $\Sigma_{l,H} \leftarrow \frac{\alpha_{H}}{\beta_{l,H}} \Sigma_{l,H}$
12. chose $\beta_{l,H}$ s.t.
\[
\sum_{H \neq H_0} \sum_{l=1}^{L} \beta_{l,H} \text{Tr} (\Delta_{l,H} (\Sigma_{l,H} - \Sigma_{l,H_0}))
= \sum_{H \neq H_0} \sum_{l=1}^{L} \text{Tr} (\Delta_{l,H} (\Sigma_{l,H} - \Sigma_{l,H_0}))
\]
13. $\Delta_{l,H} = \arg\min_{H \neq H_0} \sum_{l} \sum_{H} \| \Delta_{l,H} - \beta_{l,H} \Delta_{l,H} + \sigma (\Sigma_{l,H} - \Sigma_{l,H_0})^* \|_2^2$
\[\text{s.t.} \quad \mu_{H} I + \sum_{k \neq l} H_{k,l} \Lambda_{k,H} H_{k,l}^\dagger + \Delta_{l,H} \succ 0, \forall H\]
14. $R \leftarrow \mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \mathcal{I}_l (\Sigma_{l:L,H_0}, H) \right]$
15. until $|R - R'| \leq \epsilon$.

However, Algorithm 4 isn’t exactly suitable for distributed implementation. The main reason is that the Algorithm 4 needs to access the variables related to all the channel realization $H$ in each its iteration. While centralized algorithm can store all the past channel realization examples in the memory, distributed dual link algorithm relies on the physical
channels to updates the corresponding variable, but it only have access to one physical channel realization at a time. Although it can still make improvement based on the current channel realization, the effects on the ergodic rate may be very limited.

6.5. Robust Dual Link Algorithm

The original distributed dual link algorithm is motivated by the fact that the centralized algorithm uses signal covariance $\mathbf{H}_{l,l} \Sigma_l \mathbf{H}_{l,l}^\dagger$ / $\mathbf{H}_{l,l}^\dagger \Sigma_l \mathbf{H}_{l,l}$ and interference-plus-noise covariance $\Omega_l$ / $\tilde{\Omega}_l$ to update $\Sigma$ / $\hat{\Sigma}$. After introducing channel uncertainty as $\mathbf{H}_{l,k} = \sqrt{\rho_\mathcal{H}_{l,k}} + \sqrt{1-\rho_\mathcal{H}_{l,k}}$, could we use the expected covariance $\mathbb{E}_H \left[ \mathbf{H}_{l,k} \Sigma_k \mathbf{H}_{l,k}^\dagger \right]$ in the centralized dual link algorithm to obtain $\Sigma_l$'s that gives us a good ergodic weighted sum-rate $\mathbb{E}_H \left[ \sum_{l=1}^{L} w_l \mathcal{I}_l \left( \Sigma_{1:L}, \mathbf{H} \right) \right]$? This motivates the robust dual link algorithm shown as Algorithm 5.

Algorithm 5 Robust Dual Link Algorithm
1. Initialize $\Sigma_l$'s, s.t. $\sum_{l=1}^{L} \text{Tr} \left( \Sigma_l \right) = P_T$
2. $R \leftarrow \sum_{l=1}^{L} w_l \mathcal{I}_l \left( \Sigma_{1:L} \right)$
3. Repeat
4. $R' \leftarrow R$
5. $\Omega_l \leftarrow \mathbb{E}_H \left[ I + \sum_{k \neq l} \mathbf{H}_{l,k} \Sigma_k \mathbf{H}_{l,k}^\dagger \right]$
6. $\hat{\Sigma}_l \leftarrow \frac{P_T w_l \left( \Omega_l^{-1} - (\Omega_l + \mathbb{E}_H [H_{l,l} \Sigma_l H_{l,l}^\dagger])^{-1} \right)}{\sum_{l'=1}^{L} w_{l'} \left( \Omega_{l'}^{-1} - (\Omega_{l'} + \mathbb{E}_H [H_{l',l'} \Sigma_l H_{l',l'}^\dagger])^{-1} \right)}$
7. $\hat{\hat{\Sigma}}_l \leftarrow \mathbb{E}_H \left[ I + \sum_{k \neq l} \mathbf{H}_{k,l}^\dagger \hat{\Sigma}_k \mathbf{H}_{k,l} \right]$
8. $\hat{\hat{\Sigma}}_l \leftarrow \frac{P_T w_l \left( \hat{\hat{\Sigma}}_l^{-1} - (\hat{\hat{\Sigma}}_l + \mathbb{E}_H [H_{l,l} \hat{\Sigma}_l H_{l,l}^\dagger])^{-1} \right)}{\sum_{l'=1}^{L} w_{l'} \left( \hat{\hat{\Sigma}}_{l'}^{-1} - (\hat{\hat{\Sigma}}_{l'} + \mathbb{E}_H [H_{l',l'} \hat{\Sigma}_l H_{l',l'}^\dagger])^{-1} \right)}$
9. $R \leftarrow \sum_{l=1}^{L} w_l \mathcal{I}_l \left( \Sigma_{1:L} \right)$
10. until $\left| R - R' \right| \leq \epsilon$.

The robust dual link algorithm can be justified as follows. Since $\hat{\mathbf{H}}_{l,k}$ has i.i.d. zero mean Gaussian entries, the expected signal or interference covariance can be calculated as:
\[
\mathbb{E}_H \left[ H_{l,k} \Sigma_k H_{l,k}^\dagger \right] \\
= \mathbb{E}_H \left[ (\sqrt{\rho} H_{l,k} + \sqrt{1 - \rho} \bar{H}_{l,k}) \Sigma_k (\sqrt{\rho} \bar{H}_{l,k} + \sqrt{1 - \rho} \bar{H}_{l,k})^\dagger \right] \\
= \rho \bar{H}_{l,k} \Sigma_k \bar{H}_{l,k}^\dagger + (1 - \rho) \text{Tr}(\Sigma_k) I
\]

When \( \rho = 1 \) (no channel uncertainty), \( H_{l,k} \) becomes fixed matrices \( H_{l,k} \), \( \mathbb{E}_H \left[ H_{l,k} \Sigma_k H_{l,k}^\dagger \right] = \bar{H}_{l,k} \Sigma_k \bar{H}_{l,k}^\dagger \), and robust dual link algorithm (Algorithm 5) becomes the centralized dual link algorithm which is optimal given fixed channel matrices \( \bar{H}_{l,k} \).

Although the robust dual link algorithm is an ad-hoc algorithm only guaranteed to converge to a stationary point when \( \rho = 1 \), it has low complexity and acceptable performance under any channel uncertainty levels as we’ll show later in the next section. More importantly, the robust dual link algorithm can be easily made into an distributed online algorithm. We can estimate \( \mathbb{E}_H \left[ H_{l,k} \Sigma_k H_{l,k}^\dagger \right] \) by sending multiple pilots or taking advantage of the previous data to calculate a moving average. The similar training and signaling method can be used to distributedly implement the robust dual link algorithm as we did for the centralized dual link algorithm.

### 6.6. Simulation Results

In this section, numerical examples are provided to evaluate the performance of the proposed ergodic dual link algorithm (Algorithm 4) and robust dual link algorithm (Algorithm 5). We use the ergodic WMMSE algorithm as the benchmark since it’s guaranteed to converge to a local optimum.

Consider a B-MAC network with 10 data links among 10 transmitter-receiver pairs that fully interfere with each other. Each link has 4 transmit antennas and 4 receive antennas. For simplicity, we first chose random \( \bar{H}_{l,k} \)’s and fixed them, then independently generated 20 samples of each \( \bar{H}_{l,k} \). We calculate \( H_{l,k} \) from equation (6.2.1) and use them as realizations of equal probability to form the distribution. \( \bar{H}_{l,k} \) has zero-mean i.i.d. complex Gaussian
entries with unit variance, $\rho = 0.8$, and $P_T = 100$. The best curve out of 10 random initial points is chosen for each algorithm. As shown in Figure 6.1, the ergodic dual link algorithm monotonically converge to a stationary point as designed, but the ergodic WMMSE algorithm converges to a better rate. That’s mainly because in each iteration we have to project the updated $\Delta_{l,H}$ to a feasible subspace to make sure that the system of equations (6.4.3) and (6.4.4) has non-trivial solutions. But if projection (6.4.7) ends at the boundaries of the feasible subspace, $\Phi_{l,H}$ will be singular, and we have to chose a very small positive number as approximation of zero to numerically calculate $\Sigma_{l,H}$ from equation (6.4.5). This approximation causes the small performance gap between the ergodic dual link algorithm and the ergodic WMMSE algorithm.

For the robust dual link algorithm, we use the same settings except different $\rho$ to evaluate its performance under different channel uncertainty levels. As shown in Figure 6.2 and
6.3. the robust dual link algorithm converges faster and to the same local optimum as the ergodic WMMSE algorithm when channel uncertainty is large. It’s not surprising because when \( \rho = 0 \), \( \mathbb{E}_H \left[ H_{l,k} \Sigma_k H_{l,k}^\dagger \right] = \text{Tr} (\Sigma_k) I \), and the iterations of Algorithm 5 will generate diagonal \( \Sigma_i \) and \( \Sigma_i^\dagger \). And we can apply the same technique used in [26] to show that using diagonal matrices is sufficient to obtain the optimal ergodic weighted sum-rate when the channel matrices all have i.i.d. complex Gaussian entries. When the channel uncertainty become smaller as shown in Figure 6.4, the robust dual link algorithm doesn’t perform as well as the ergodic WMMSE algorithm but is still in the acceptable range. The performance drop can be a reasonable trade-off for an online distributed algorithm. Finally, when there is zero uncertainty, the robust dual link algorithm becomes the original dual link algorithm and thus can always converge to the stationary point as shown in Figure 6.5.
Figure 6.3. Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 0.5$

6.7. Conclusion

In this chapter, a simple channel uncertainty model is introduced to study the influence of the imperfect channel knowledge to the dual link algorithm. Ergodic dual link algorithm is proposed to solve the ergodic weighted sum-rate maximization problem to improve the average performance under channel uncertainty. And we also proposed the robust dual link algorithm, which is suitable for online distributed implementation and has good performance under all channel uncertainty level.
Figure 6.4. Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 0.9$
Figure 6.5. Robust Dual Link Algorithm vs. Ergodic WMMSE Algorithm, $\rho = 1$
CHAPTER 7

Uplink and Downlink Analysis of Massive MIMO Cellular Systems

7.1. Introduction

In this Chapter, we begin with a brief discussion about the physical layer transmission and reception schemes in a cellular system where the BSs are equipped with infinitely many antennas. It is shown that the effect of background noise is completely eliminated and the cellular system is essentially interference-limited in both the uplink and the downlink. The main result is that the SIR in the uplink and the downlink are identical. This section is an extension of the results in [17] to the case when the pilot sequences are non-orthogonal and for an arbitrary number of users in the cell.

7.2. System Model

7.2.1. Transmission-Reception Schemes. We consider a large cellular area that is divided into non-overlapping cells of radius $R$. The MSs in each cell are served by the BS that is at the center of the cell. Each MS has a single antenna and each BS is equipped with $M$ antennas ($M \to \infty$). The BSs employ $\Delta$-frequency reuse ($\Delta = 1, 3, \text{or} 7$) in order to further reduce the inter-cell interference. Fig. 7.1 illustrates the scenario.

All the BSs in the system share the same set of $P$ pilot sequences to serve up to $P$ MSs simultaneously. Each pilot sequence is a $K$ length unit-norm vector denoted as $a_i \in \mathbb{C}^{K \times 1}$, $i = 1, 2, \cdots, P$, and the correlation between two pilot sequences $a_i$, $a_j$ is $a_i^\dagger a_j = \alpha_{ij}$, such that $0 \leq |\alpha_{ij}| < 1$ and $\alpha_{ii} = 1$, $\forall i = 1, \cdots, P$. The pilots are said to be orthogonal if the correlation is zero. The BSs and the MSs communicate with each other via time division duplexing (TDD).

7.2.2. Pilot Signaling and Channel Estimation. The communication begins with the training phase when all the MSs in the cell transmit their respective pilot sequences to the
serving BS. For the sake of simplicity, it is assumed that all the MSs transmit synchronously. The BSs utilize the reverse-link pilot transmissions to estimate the reverse-link channel to each of its MSs. Consequently, the forward and reverse-link channels are identical due to the channel reciprocity induced by the TDD operation of the BSs and the MSs. The received signal at the $k^{th}$ BS corresponding to one pilot sequence (consisting of $K$ symbols) transmission period may be represented by $\mathbf{Y}_k \in \mathbb{C}^{M \times K}$:

\[
\mathbf{Y}_k = \sqrt{\rho_p} \sum_{l=1}^{\infty} \sum_{n=1}^{P} b_{kl} \mathbf{h}_{kn} a_n^\dagger \mathcal{I}(n, l) + \mathbf{W}_k
\]

where $M$ is the number of antennas at BS, $K$ is the length of a pilot sequence, $b_{kl} = 1$ if the $l^{th}$ BS operates in the same frequency band as the $k^{th}$ BS and $b_{kl} = 0$ otherwise. $\mathbf{h}_{kn} \in \mathbb{C}^{M \times 1}$ is the channel corresponding to the wireless link from $n^{th}$ MS of the $l^{th}$ BS to the $k^{th}$ BS, $\rho_p$ is the pilot signal-to-noise ratio (SNR), $\mathbf{W}_k$ is the i.i.d. zero-mean noise at $k$th BS, and $\mathcal{I}(n, l)$ is the indicator function

\[
\mathcal{I}(n, l) = \begin{cases} 
1, & n^{th} \text{ pilot is used in } l^{th} \text{ cell} \\
0, & \text{otherwise}
\end{cases}
\]
Further, $h_{kn} = \beta_{kn} R_{kn}^{-\epsilon/2} g_{kn}$, $g_{kn} \in \mathbb{C}^{M \times 1}$ represents the fast fading coefficients between the MS and each antenna of the $k^{th}$ BS with i.i.d. zero mean and unit variance entries, $\beta_{kn}$ is the shadow fading coefficient, generally modeled as a log-normal random variable with 0 mean and variance $\sigma^2$ dB, $R_{kn}$ is the distance between the MS and $k$th BS and $\epsilon$ is the path-loss exponent of the power-law path-loss model.

From the received signal in (7.2.1), the $k^{th}$ BS estimates the channel to the MS transmitting the $m^{th}$ pilot sequence as

$$
\hat{h}_{km} = \frac{Y_k a_m}{\sqrt{\rho P}} = h_{km} + \sum_{n=1, n \neq m}^{P} \alpha_{nm} h_{kn} I(n, k) + \sum_{l=1}^{\infty} \sum_{n=1}^{P} \alpha_{nm} b_{kl} h_{kn} I(n, l) + \frac{W_k a_m}{\sqrt{\rho P}},
$$

(7.2.2)

where $m = 1, 2, \cdots, P$, the first term is the desired channel coefficients, the second term is the contamination due to non-orthogonal pilots used by other MSs served by the $k^{th}$ BS, third term is the contamination due to the pilot transmissions of the MSs belonging to the other cells and the last term corresponds to the background noise. Next, we focus on the uplink data transmission and decoding scheme used by each BS to recover the data transmitted by each of its MSs.

### 7.3. Uplink and Downlink Analysis

#### 7.3.1. Uplink Data Transmission and Maximum-Ratio Combining

Following the pilot signaling stage is the reverse link data transmission stage, when all the MSs transmit data symbols to their corresponding BS, and the composite signal as received by the $k^{th}$ BS is given by

$$
Y_k = \sqrt{\rho_U} \sum_{l=1}^{\infty} \sum_{n=1}^{P} b_{kl} h_{kn} d_{ln} I(n, l) + w_k,
$$

(7.3.1)

where $\rho_U$ is the uplink SNR, $d_{ln}$ is data symbol transmitted by the $n^{th}$ MS of the $l^{th}$ BS’s signal, and $w_k$ is the i.i.d. zero mean and unit variance background noise at the $k^{th}$ BS
antennas. From the above received signal, the $k^{th}$ BS decodes the symbols corresponding to each of its MSs using maximum-ratio combining, by left-multiplying the received signal by the conjugate-transpose of the channel estimate of the corresponding MS. Further, in the limits $M \to \infty$, the decoded symbol $\hat{d}_{km}$ corresponding to the $m^{th}$ MS of the $k^{th}$ BS take a relatively simple form as shown below.

$$\hat{d}_{km} = \lim_{M \to \infty} \frac{\hat{h}_{kkm}^\dagger \mathbf{y}_k}{M \sqrt{p_U}}$$

$$= (a) \quad \lim_{M \to \infty} \frac{1}{M} \left( \sum_{l=1}^{\infty} \sum_{n=1}^{P} \alpha_{n}^{*} b_{kl} \mathbf{h}_{kl}^\dagger \mathbf{I} (n, l) + \frac{\mathbf{a}_m^\dagger \mathbf{W}_k^\dagger}{\sqrt{p_U}} \right)$$

$$\cdot \left( \sum_{s=1}^{\infty} \sum_{t=1}^{P} \mathbf{h}_{skt} b_{ks} d_{st} \mathbf{I} (s, t) + \frac{w_k}{\sqrt{p_U}} \right)$$

$$= (b) \quad \frac{\beta_{kkm} d_{km}}{R_{kkm}} + \sum_{n=1, \ n \neq m}^{P} \frac{\alpha_{nm}^{*} \beta_{knm} d_{kn}}{R_{knm}} \mathbf{I} (n, k)$$

$$+ \sum_{l=1, \ l \neq k}^{P} \sum_{n=1}^{P} \frac{\alpha_{nm}^{*} b_{kl} \beta_{knm} d_{ln}}{R_{knm}} \mathbf{I} (n, l),$$

(7.3.2)

where $m = 1, 2, \ldots, P$, $(a)$ is obtained by substituting for $\hat{h}_{kkm}$ and $\mathbf{y}_k$ from (7.2.2) and (7.3.1), respectively, and $(b)$ is obtained by noting that $\lim_{M \to \infty} \frac{\mathbf{h}_{kl}^\dagger \mathbf{h}_{kl}}{M} = \sqrt{\beta_{kl} R_{kl}} \lim_{M \to \infty} \frac{\mathbf{g}_{kl}^\dagger \mathbf{g}_{kl}}{M}$,

$$= b_{kl} \beta_{kl} R_{kl}^{-\frac{1}{2}} \delta (l = s, n = t), \lim_{M \to \infty} \frac{\mathbf{h}_{kl}^\dagger \mathbf{w}_k}{M} = 0, \lim_{M \to \infty} \frac{\mathbf{w}_k^\dagger \mathbf{h}_{kl}}{M} = 0, \text{ and } \lim_{M \to \infty} \frac{\mathbf{w}_k^\dagger \mathbf{w}_k}{M} = 0$$

by applying the law of large numbers, since $\mathbf{g}_{kl}$, $\mathbf{w}_k$, $\mathbf{W}_k$ all contain i.i.d. zero mean unit variance entries. Further, the first term in $(b)$ is the desired data symbol, the second is intra-cell interference term and the last is inter-cell interference term. Next, we study the downlink transmission scheme in detail.

### 7.3.2. Precoding and Downlink Data Transmission

In the downlink, the BS precodes the data symbol intended for each MS with the corresponding wireless links channel estimate, and transmits the sum of the precoded signals of all its MSs through the $M$ antennas. The received signal at the $m^{th}$ MS of the $k^{th}$ BS is

$$y_{km} = \sqrt{p_U} \sum_{l=1}^{\infty} b_{kl} \mathbf{h}_{kl}^\dagger x_l + w_{km},$$
where $\rho_D$ is the downlink SNR, $x_l = \sum_{n=1}^{P} \hat{h}_{ln} d_{ln} \mathcal{I} (n, l)$ is the signal transmitted by the $l^{th}$ BS, $d_{ln}$ is the data symbol intended for its $n^{th}$ MS which is precoded by the corresponding channel estimate $\hat{h}_{ln}$. Due to channel reciprocity induced by the TDD operation, the channel between the $l^{th}$ BS and the $m^{th}$ MS of the $k^{th}$ BS is $h_{km}^{\dag}$, and lastly, $w_{km}$ is a zero mean, unit variance random variable representing the background noise. Each MS performs a relatively simple processing to recover the data symbol transmitted by the serving BS. The downlink data symbol is decoded as

$$
\hat{d}_{km} = \lim_{M \to \infty} \frac{y_{km}}{M \sqrt{\rho_D}} \\
= \lim_{M \to \infty} \sum_{l=1}^{\infty} \sum_{n=1}^{P} \frac{h_{km}^{\dag} \hat{h}_{ln}}{M} d_{ln} \mathcal{I} (n, l) + \frac{w_{km}}{M \sqrt{\rho_D}} \\
= \sum_{l=1}^{\infty} \sum_{n=1}^{P} \alpha_{nm}^{\ast} \beta_{km} R_{km}^{-\varepsilon} b_{kl} d_{ln} \mathcal{I} (n, l),
$$

(7.3.3)

where $m = 1, 2, \cdots, M$, $\lim_{M \to \infty} \frac{h_{km}^{\dag} \hat{h}_{ln}}{M} = \alpha_{nm} \beta_{km} R_{km}^{-\varepsilon}$, $\alpha_{mn} = \alpha_{nm}^{\ast}$, and $\lim_{M \to \infty} \frac{w_{km}}{M \sqrt{\rho_D}} = 0$.

### 7.3.3. Uplink-Downlink SIR Duality

Notice that the resultant system is again interference-limited and moreover, by comparing in (7.3.2) and (7.3.3), we find that the expressions for the decoded symbol in the uplink and downlink are identical. Hence, the uplink and downlink SIRs are also identical, and the following lemma provides the corresponding expression.

**Lemma 14. Uplink-Downlink SIR Duality:** In a cellular system consisting of infinitely many BSs, each equipped with $M$ antennas ($M \to \infty$) and serving up to $P$ MSs simultaneously using $P$ pilot sequences that is common across all the BSs, with the transmission and reception schemes as studied above, the uplink SIR and the downlink SIR corresponding to a given BS-MS pair are identical and is given by

$$
SIR_{km} = \sum_{l=1}^{\infty} \sum_{n=1}^{P} \frac{\beta_{km}^{2} R_{km}^{-2\varepsilon}}{\sum_{(l,n) \neq (k,m)} \beta_{kl}^{2} \alpha_{nm}^{2} R_{kl}^{-2\varepsilon} \mathcal{I} (n, l)},
$$

(7.3.4)
and the per-user achievable rate is \( R_{km} = \frac{B}{\Delta} \log_2 (1 + \text{SIR}_{km}) \) bits/sec/user, where \( B \) is the entire allocated bandwidth and \( \Delta \) is the frequency reuse factor.

7.4. Conclusion

We study the physical layer transmission and reception schemes in a cellular system where the BSs are equipped with infinitely many antennas, the pilot sequences can be non-orthogonal, and number of users in the cell can be arbitrary. It is shown that the uplink SIR and the downlink SIR corresponding to a given BS-MS pair are identical.
CHAPTER 8

Conclusion

This thesis studies optimization and performance of the MIMO B-MAC interference networks. A 3D channel model for distributed MIMO system is set up, based on which the antenna correlations can be characterized in analytic form. We propose a new algorithm, named Dual Link algorithm, for the classic problem of weighted sum-rate maximization for MIMO multiaccess channels (MAC), broadcast channels (BC), and general MIMO interference channels with Gaussian input and a total power constraint. For MIMO MAC/BC, the algorithm finds optimal signals to achieve the capacity region boundary. For interference channels with Gaussian input assumption, the dual link algorithm is guaranteed to converge to the boundary of the achievable rate region. Since it takes advantage of the optimal transmit signal structure known as the polite water-filling structure, the convergence speed is very fast. The dual link algorithm is also further modified to fit practical applications. We’ve proposed the distributed dual link algorithm for time division duplex (TDD) interference networks. It replaces direct and cross channel information feedbacks with iterations of forward and reverse link pilots training whose complexity grows linearly with the number of users in the network. By totally avoiding channel state knowledge feedback, the distributed dual link algorithm has significant lower signaling overhead compared to the traditional methods, especially in networks with large number of interfering users. The real TDD channels are not reciprocal because the transmit and receive RF chains are different in a transceiver. So we propose a simple and effective antenna calibration method for TDD MIMO interference networks to maintain the reciprocity. A simple channel uncertainty model is introduced to characterize different levels of channel uncertainty. The ergodic dual link algorithm is proposed to analytically solve the ergodic weighted sum-rate maximization problem. We also propose the robust dual link algorithm which is sub-optimal but has good performance under
all channel uncertainty levels and is suitable for online distributed implementation. Finally, we study the physical layer transmission and reception schemes in a massive MIMO cellular system, and the uplink SIR and the downlink SIR duality is identified.
Bibliography


