OPTIMIZATION IN MULTI-RELAY WIRELESS NETWORKS

A Thesis Submitted
to the College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Electrical and Computer Engineering
University of Saskatchewan

by

Duy H. N. Nguyen
B. Eng. (Hons I)

Saskatoon, Saskatchewan, Canada

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, it is agreed that the Libraries of this University may make it freely available for inspection. Permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised this thesis work or, in their absence, by the Head of the Department of Electrical and Computer Engineering or the Dean of the College of Graduate Studies and Research at the University of Saskatchewan. Any copying, publication, or use of this thesis, or parts thereof, for financial gain without the written permission of the author is strictly prohibited. Proper recognition shall be given to the author and to the University of Saskatchewan in any scholarly use which may be made of any material in this thesis.

Request for permission to copy or to make any other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Electrical and Computer Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
S7N 5A9

Acknowledgments

Over the years, I have been fortunate to have the experience working with and getting support from many great persons. A few lines here are certainly not enough to express my sincere appreciation to all of them.

First of all, I would like to gratefully acknowledge and express a sincere thank you to my supervisor, Professor Ha Hoang Nguyen for his invaluable support and guidance during my studies at the University of Saskatchewan. I am truly privileged to have learned from his remarkable technical knowledge and research enthusiasm. He has not only introduced and taught me many aspects of the interesting field of wireless communications, but also laid down the path for me to keep exploring new knowledge in the field. This thesis would not be finished without his support and encouragement.

I would like to thank Professor Hoang Duong Tuan of University of New South Wales, NSW, Australia for his collaboration on various works in optimization theory. His deep and thorough knowledge in optimization theory has helped bring to light to collaborated works with him.

I am also grateful to Dr. Jochen Trumpf of Research School of Information and System Engineering, Australian National University (ANU), ACT, Australia, for his encouragement to my pursuit of higher education in North America. His guidance during the exciting summer of 2004-2005 at ANU has definitely boosted my confidence in pursuing research in the electrical engineering field.

I would like to say thank you to my two longtime friends back to the days in Australia: Khoa (Kevin) T. Phan of California Institute of Technology, and Quoc V. Le of Stanford University. The endless discussions on various topics in communication and networking theories with Khoa have inspired me to lots of open problems in the fields. With Quoc, I have had a great pleasure to talk with him about optimization theory, which has helped me to get more in-depth understanding to the theory.

I would like to thank all my labmates: Tung, Nam, Ha, Zohreh, Simin, and Quang for the wonderful time at the Communication Theory Research Group (CTRG).

My deepest love and gratitude is devoted to all of my family members: Mom and Dad, brothers Khanh and Ngoc, sister-in-law Huong, and beloved niece Giang, who always support me in each and every endeavor in my life. I truly owe all of my successes to them. A special thank you must go to my girlfriend, Ngan, for her love, support, and tenderness. I enjoy every single moment talking with her, planning our future, and I look forward to having her on the long journey ahead.

Finally, I would not forget to gratefully acknowledge the NSERC Discovery Grant and the Department of Electrical and Computer Engineering, University of Saskatchewan for the financial support of my studies.

Abstract

The concept of cooperation in communications has drawn a lot of research attention in recent years due to its potential to improve the efficiency of wireless networks. This new form of communications allows some users to act as relays and assist the transmission of other users' information signals. The aim of this thesis is to apply optimization techniques in the design of multi-relay wireless networks employing cooperative communications. In general, the thesis is organized into two parts: "Distributed space-time coding" (DSTC) and "Distributed beamforming", which cover two main approaches in cooperative communications over multi-relay networks.

In Part I of the thesis, various aspects of distributed implementation of space-time coding in a wireless relay network are treated. First, the thesis proposes a new fully-diverse distributed code which allows noncoherent reception at the destination. Second, the problem of coordinating the power allocation (PA) between source and relays to achieve the optimal performance of DSTC is studied and a novel PA scheme is developed. It is shown that the proposed PA scheme can obtain the maximum diversity order of DSTC and significantly outperform other suboptimal PA schemes. Third, the thesis presents the optimal PA scheme to minimize the mean-square error (MSE) in channel estimation during training phase of DSTC. The effect of imperfect channel estimation to the performance of DSTC is also thoroughly studied.

In Part II of the thesis, optimal distributed beamforming designs are developed for a wireless multiuser multi-relay network. Two design criteria for the optimal distributed beamforming at the relays are considered: (i) minimizing the total relay power subject to a guaranteed Quality of Service (QoS) measured in terms of signal-to-noise-ratio (SNR) at the destinations, and (ii) jointly maximizing the SNR margin at the destinations subject to power constraints at the relays. Based on convex optimization techniques, it is shown that these problems can be formulated and solved via second-order conic programming (SOCP). In addition, this part also proposes simple and fast iterative algorithms to directly solve these optimization problems.

Table of Contents

List of Tables

List of Figures

Abbreviations

4G	Fourth Generation
AF	Amplify-and-Forward
AWGN	Additive White Gaussian Noise
CSI	Channel State Information
CSNR	Channel Signal-to-Noise-Ratio
dB	Decibel
DF	Decode-and-Forward
DFT	Discrete Fourier Transform
DSTC	Distributed Space-Time Coding
DUSTM	Distributed Space-Time Modulation
GLRT	Generalized Likelihood Ratio Test
FDMA	Frequency Division Multiple Access
KKT	Karush-Kuhn-Tucker
LP	Linear Programming
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
MSE	Mean-Square Error
OFDM	Orthogonal Frequency Division Multiplexing
OSTBC	Orthogonal Space-Time Block Code
PA	Power Allocation

pdf	Probability Density Function
PEP	Pairwise Error Probability
QCQP	Quadratically Constrained Quadratic Programming
QoS	Quality of Service
QOSTBC	Quasi-Orthogonal Space-Time Block Code
QP	Quadratic Programming
SDP	Semidefinite Programming
SER	Symbol-Error-Rate
SIMO	Single-Input Multiple-Output
SINR	Signal-to-Interference-plus-Noise-Ratio
SNR	Signal-to-Noise-Ratio
SOC	Second-Order Cone
SOCP	Second-Order Conic Programming
SR	Selection Relaying
ST	Space-Time
STC	Space-Time Coding
TDMA	Time Division Multiple Access
USTM	Unitary Space-Time Modulation

1. Introduction

In recent years, the rapid expansion of wireless communications has put a significant pressure to the current wireless network infrastructure to cope with demands for higher throughput, higher robustness and better coverage. It is expected that such demands would even be stronger in the future fourth-generation (4G) wireless networks. No longer limited to a medium for only voice transmission, wireless communications have assumed an important role in transmitting data at higher rates and streaming multimedia services, such as video, at higher Quality of Service (QoS) requirements as well. These demands have posed tough technical challenges for the current and future wireless networks.

A fundamental aspect of wireless communication that makes the design of robust wireless networks challenging is the phenomenon of fading. Basically, fading in a wireless channel refers to the time and frequency variations of the channel quality. The fading effect of a wireless channel can be categorized into two types: large-scale fading, and small-scale fading [1]. The large-scale fading is due to the signal attenuation as a function of distance and shadowing effects caused by large objects such as buildings, hills, obstacles, etc. The small-scale fading is due to the constructive and destructive interferences of the multiple paths between the transmitter and receiver. How to deal with fading, especially the small-scale fading, is critical to the design of any robust wireless communication system.

As illustrated in Figure 1.1, the received signal in a wireless channel is a composite of the transmitted signal over several different paths due to reflection, diffraction, and scattering from buildings, moving objects like cars, trees, etc. When there is a

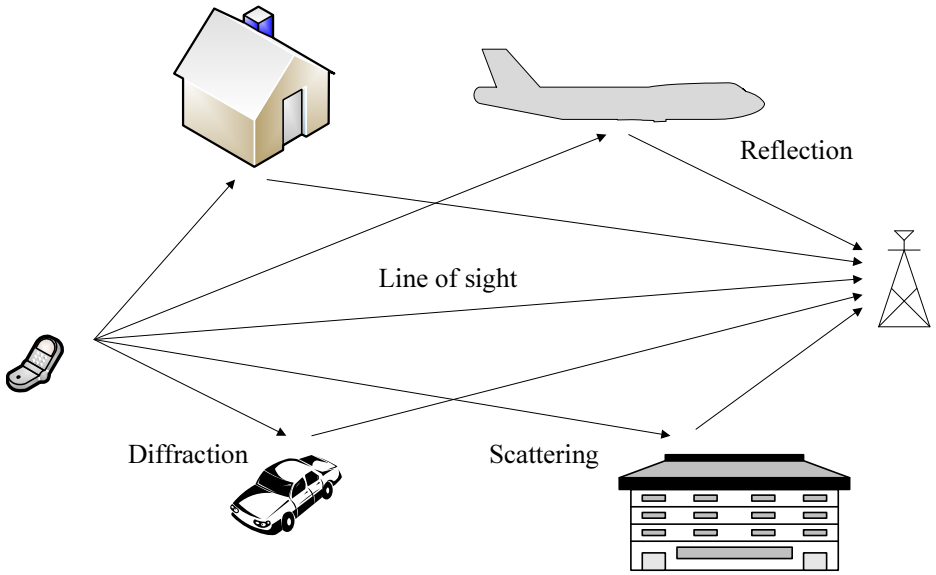

Figure 1.1 A typical transmission wireless environment.

large number of scatterers that contribute to the signal at the receiver, the central limit theorem leads to a Gaussian process model for the channel coefficient. If the process is zero-mean, the absolute value of the channel response at any time instant has a Rayleigh probability distribution. Mathematically, given $x[m]$ as the transmitted signal, the received signal $y[m]$ over a fading channel (in discrete base-band representation) at time m is given by

$$y[m] = h[m]x[m] + z[m], \tag{1.1}$$

where $z[m]$ is the additive white Gaussian noise (AWGN), and $h[m]$ is the channel coefficient, which is commonly referred to as the "channel state information" (CSI). In a typical Rayleigh fading model of the wireless channel, $h[m]$ is a zero-mean circularly symmetric complex Gaussian random variable.

It is noted that in an AWGN channel, where $h[m] = 1$, errors in detection occur due to the additive noise component $z[m]$. Uncoded signaling schemes, such as binary phase-shift keying (BPSK), can perform very well in an AWGN channel, where the detection error decays exponentially with the signal-to-noise ratio (SNR) [1]. On the other hand, in a fading channel, the BPSK signaling scheme fails completely even in

the absence of noise if the receiver has no knowledge of channel coefficient $h[m]$. This is because there is nothing in the received signal $y[m]$ which can be used to distinguish the transmitted signal $x[m]$. As the phase of the channel coefficient $h[m]$ is uniformly distributed between 0 and 2π, the phase of $y[m]$ is also uniformly distributed between 0 and 2π. Furthermore, the amplitude of $y[m]$ is independent of the transmitted BPSK symbol $x[m]$.

Suppose now that the channel coefficient is tracked such that it is known at the receiver. This can be practically done by sending a known sequence (called training), and performing "channel estimation" at the receiver using the training sequence. As a result, *coherent* detection can be performed as in the case with an AWGN channel. Even so, the communication scheme over the fading channel still suffers from a much poorer performance than that over the AWGN channel [2]. This is due to the fact that the channel gain is random and there is a certain probability that the channel is "in a deep fade" [1]. More specifically, when the channel attenuation is large (the amplitude of the channel coefficient is small), the low instantaneous received SNR leads to a high probability in detection error.

However, if the receiver is provided with several replicas of the same information signal transmitted over independent fading channels, the probability that all the received signal components are in deep fades simultaneously is much smaller. This approach to combat fading, called "diversity technique", has been commonly applied in wireless communications. In order to quantify the effectiveness of a diversity technique, the relationship between the average SNR and the average error probability P_e is determined. A common measure is the "diversity order", defined as follows:

$$G_d = -\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e}{\log \mathrm{SNR}}. \tag{1.2}$$

Obviously, the higher the diversity order is, the more reliable the wireless communication system is. There are several approaches to provide diversity techniques in a wireless system as outlined next. Regardless of the approach used, it is important that the implemented technique is capable of obtaining the system's "maximum diversity order". More precisely, the diversity technique should take the full advantage

of the multiple independent received copies of the same transmitted signal.

One diversity technique is to transmit the same information signal over multiple frequency bands, where the separation between successive bands equals or exceeds the coherence bandwidth of the channel. Here, the coherence bandwidth measures the frequency range over which the channel responses are correlated. This technique is called *frequency diversity*. A second technique to obtain independently faded versions of the same information signal is to transmit the signal in different time slots, where the separation between successive time slots equals or exceeds the coherence time of the channel. Here, the coherent time refers to the time duration over which the channel responses are correlated. This technique is known as *time diversity*. It is noted that both frequency and time diversity techniques are inefficient since frequency diversity requires bandwidth expansion, whereas time diversity needs extra time slots for transmission. Another popular approach to obtain diversity is to deploy multiple antennas in reception and/or transmission. This technique is called *spatial diversity*. Spatial diversity, albeit requiring extra costs for multiple antenna deployment, is much more spectral efficient and can overcome the drawbacks of both frequency diversity and time diversity.

Figure 1.2 demonstrates examples of providing spatial diversity in an uplink channel from a mobile unit to a base station. Spatial diversity can be obtained by deploying multiple antennas at the receiving end, referred to as *receive diversity*, or at the transmitting end, referred to as *transmit diversity*, or a combination of both. While it is straightforward to realize receive diversity by simple combining techniques at the receiver, such as the maximal ratio combining (MRC), achieving transmit diversity requires a more complex technique at the transmitter [1]. The recent invention of *space-time coding* [3, 4] (in the 1990s) allows a simple, yet elegant method to obtain transmit diversity with a very high bandwidth efficiency. In addition, space-time coding is capable of achieving the maximum diversity order of the multiple-antenna system. These recent developments in multi-input multi-output (MIMO) systems have been a significant step forward with a lot of potentials in meeting the technical

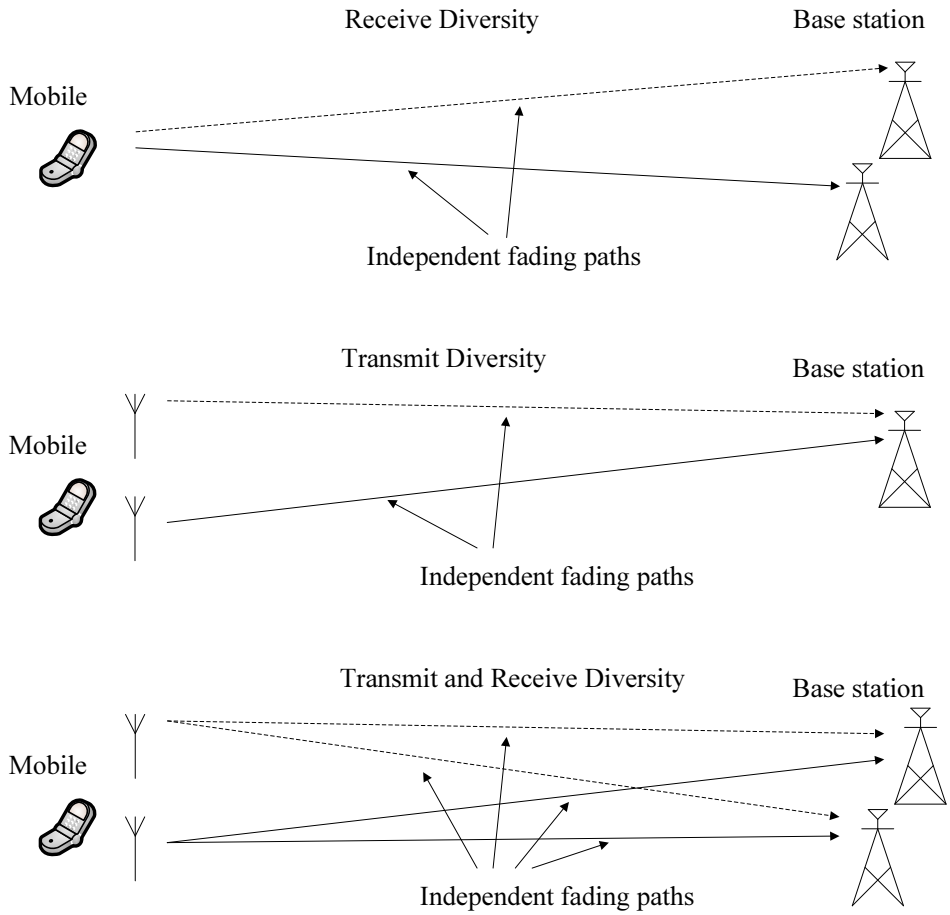

Figure 1.2 Illustration of spatial diversity in an uplink channel.

challenges of the current and next generation of wireless communications. However, the implementation of multiple antennas at mobile units faces a serious technical challenge due to the size and power limitation of the mobile units. This is because the multiple antennas have to be placed several wavelengths apart so that the channel between each transmit and receive antenna experiences independent fading.

More recently, cooperative communication has been proposed to provide a different implementation of multiple antennas which can allow future communication systems to overcome the aforementioned drawback [5]. In this new form of communications, the single-antenna users (or nodes) cooperate to relay each other's information signals, create a virtual array of transmit antennas, and, thus achieve spatial diversity. Such cooperation can significantly improve the reliability of signal transmission from each

user [6]. User cooperation also enables the system to enhance its capacity and extend its coverage [7]. Due to the tremendous potential of cooperative communication, there has been a lot of research efforts in the last few years to study both the theoretical performance and practical implementation of this new communication scheme.

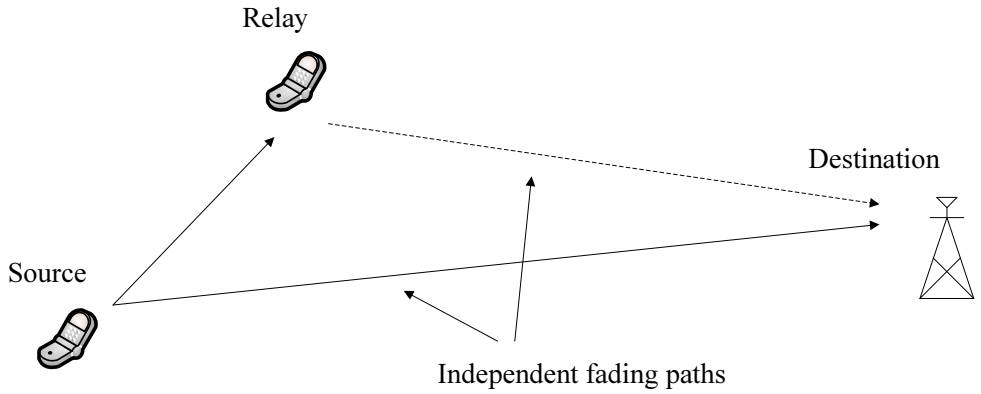

Figure 1.3 Example of a relay channel.

Figure 1.3 illustrates an example of cooperative communication in its simplest form. In particular, one user, called the source user, wants to send its information signal to another user, called the destination user. Another user, which acts as the relay, also receives the transmitted signal. The relay *processes* the received signal in some way, and then retransmits to the destination. As a result, the original signal experiences independent transmissions from the source to destination. At the relay, several relaying protocols, e.g., amplify-and-forward (AF), decode-and-forward (DF) [5], could be applied to *process* the received signal. As its name suggests, in AF protocol, the relay simply amplifies the received signal and then forwards to the destination. On the other hand, in DF protocol, the relay first decodes the received signal, re-encodes it, and then forwards to the destination. Figure 1.4 visually describes the AF and DF protocols at the relay.

In general, multiple relays could simultaneously assist the transmission from a source to a destination. Each relay could employ a dedicated channel to retransmit its received signal to the destination. However, this method is spectrally inefficient since it requires the number of dedicated channels to be at least equal to the number

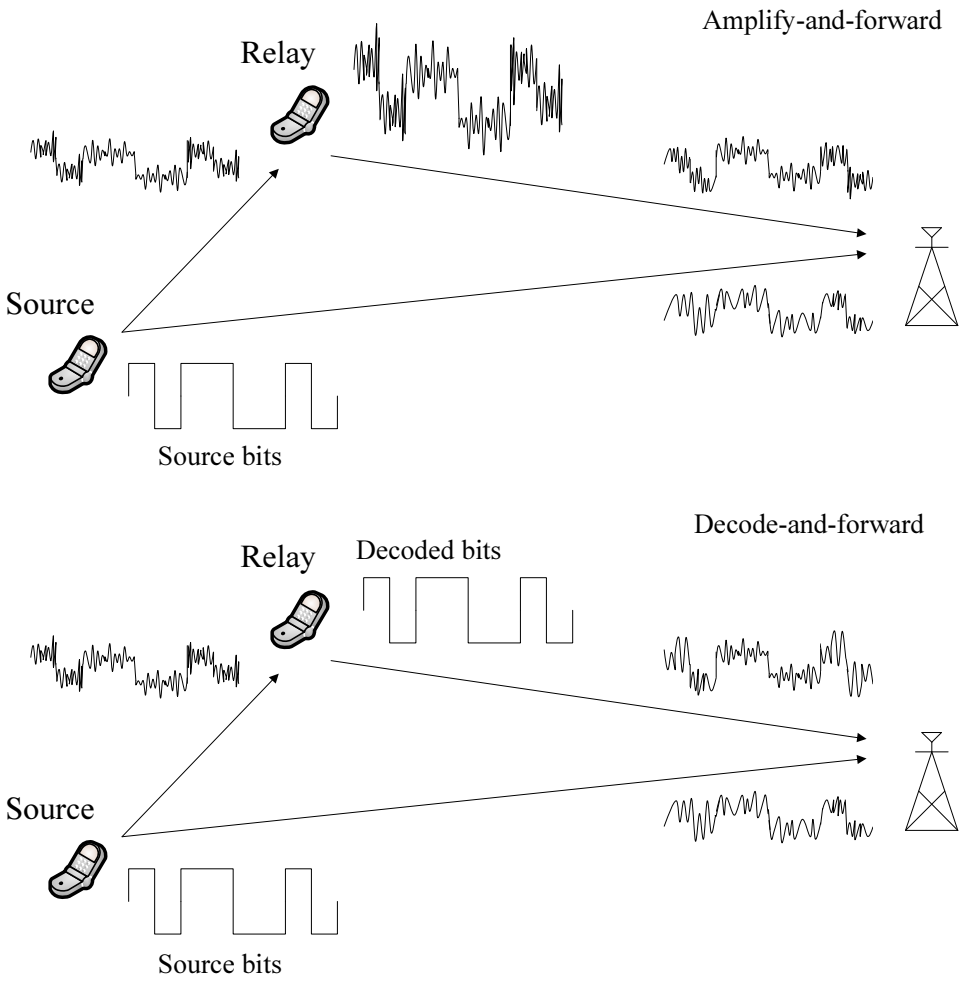

Figure 1.4 Amplify-and-forward and decode-and-forward protocols.

of relays. In [8], a new cooperative strategy using AF protocol, referred to as "distributed space-time coding" (DSTC), was proposed, where the conventional space-time coding designed for co-located antennas was implemented between the relays in a distributed manner. The new strategy allows a space-time transmission of source signals to the destination on the same channel and at the same time, and hence is more spectrally efficient. In addition, the new scheme is capable of obtaining the maximum diversity order promised by cooperative communications. Various aspects of the new cooperative strategy have led to many interesting research problems, such as code designs and performance analysis [8–10]. The main focuses of Part I of this thesis are to study fully-diverse code designs for DSTC with different types of CSI assumptions of the relay networks, optimal power allocation (PA) schemes between

the users, channel estimation, and performance analysis of mismatched decoding in DSTC.

It should be noted that DSTC can be implemented without the need of full channel state information (CSI) at the relays. However, should the CSI be available at the relays, i.e., each relay knows the CSI of the channel links connected to it, the relays can and should compensate for the phase changes introduced by the channels. As a result, the received signal at the destination can be coherently constructed. Such relaying strategy, known as "distributed beamforming", was first proposed in [11–14]. The works in [11–14] study the optimal distributed beamforming strategies to either maximize the signal-to-noise (SNR) at the destination subject to power constraints as the relays or minimize the total relay power subject to the QoS requirement at the destination. Motivated by the early works on distributed beamforming in a one-source one-destination network, Part II of this thesis generalizes the distributed beamforming designs to a multi-source multi-destination network. By applying convex optimization techniques, this second part of the thesis proposes resource allocation schemes at the relays in order to optimally assist multiple source-destination pairs.

1.1 Thesis Contribution and Outline

The thesis is divided into two main parts, titled *Distributed space-time coding* and *Distributed beamforming*, and is organized as follows.

Part I of the thesis, comprises Chapters 2, 3, and 4, is concerned with several important aspects of DSTC, including code designs, performance analysis, power allocation, channel estimation, and mismatched decoding. An important characteristic of DSTC is that it allows the destination to exploit cooperative diversity without the availability of CSI at the relays. Chapter 2 considers the design of DSTC in wireless multi-relay networks with different CSI assumptions at the destination. First, the chapter reviews the DSTC design criteria in coherent relay networks. It then presents the design and the performance analysis of DSTC when full CSI at the destination is unavailable. Proposed in the chapter is a new fully-diverse distributed

code for the partially coherent and the noncoherent relay networks. The proposed distributed unitary space-time modulation (DUSTM), which relies on Fourier-based USTM, allows noncoherent detection at the destination. Developed are the maximum likelihood (ML) receiver for DUSTM over partially coherent relay networks and the generalized likelihood-ratio test (GLRT) receiver for the noncoherent relay networks. Performance analysis of the DUSTM over the two types of relay networks reveals a surprising result that the knowledge of the relay-to-destination channel information has a very little impact on the code performance.

Chapter 3 considers the optimal PA problem for DTSC based on the second-order statistics of each source-to-relay ($S \to R$) and relay-to-destination ($R \to D$) channels, subject to a total power budget at the source and relays. The chapter first examines the optimal PA scheme to maximize the effective average SNR in an arbitrary relay network. Interestingly, it shows that maximizing the average SNR is not sufficient to optimize the performance of the DSTC. More specifically, at the optimal solution for such a PA scheme, some of the relays might not be active, and thus compromise the distributed code's full diversity order. It then introduces the concept of amount of fading to the relay networks and establishes the condition of the transmitted power at each relay such that the fading statistics of each $S \to R \to D$ link is balanced. A novel and simple PA scheme is proposed at the chapter's end and proved to be capable of obtaining the maximum diversity order in coherent, partially coherent, and noncoherent DSTC systems at high SNR. Finally, simulation results are given to confirm the analysis and show a significant performance improvement by the proposed PA scheme.

Chapter 4 is concerned with the optimal design in the training phase and the impact of channel estimation and mismatched decoding on DSTC. The chapter first makes use of the results in [15] that orthogonal training is optimal to minimize the total mean-square error (MSE) for both the maximum likelihood (ML) and minimum mean-square error (MMSE) estimations. It then studies an optimal PA scheme to further minimize the total MSE for both the estimation schemes. The result shows

that the optimal PA scheme is the same as that obtained in Chapter 3 under the minimum amount of fading constraint. The impact of imperfect channel estimation on the error performance of DSTC is also analyzed and it is proved that the mismatched decoding of DSTC is able to achieve the same diversity order as the coherent decoding of DSTC.

Part II of the thesis, consists of Chapters 5 and 6, is devoted to multi-source multi-destination networks employing distributed beamforming. An important remark about distributed beamforming is that it requires full CSI knowledge at the relays to perform beamforming. Two main problems in optimal distributed beamforming designs addressed in this part are (i) minimize the total relay power with guaranteed QoS in terms of SNR at the destinations, and (ii) jointly maximize the SNR margin at the destinations subject to power constraints at the relays. The problems are sequentially investigated and shown to be closely related to each other.

Chapter 5 studies optimal distributed beamforming designs to minimize the total relay power with guaranteed QoS. The chapter exploits convex optimization techniques to find the optimal beamformers in a relay network with or without per-relay power constraints. First, the chapter shows that these problems can be formulated and solved via second-order conic programming (SOCP). Although the optimal solutions to the problems can be obtained by any conic solution package, the contribution of this chapter is a proposal of simple and fast iterative algorithms to efficiently solve them. The feasibility conditions of the two optimization problems are also studied in the chapter. With different assumptions on orthogonality in $S \rightarrow R$ and $R \rightarrow D$ transmission phases, several potential future works on distributed beamforming designs are recommended at the chapter's end.

Chapter 6 studies the distributed beamforming problems that are inverse to the ones in Chapter 5. With the constraints on either the sum relay power or the per-relay power, optimal distributed beamforming designs are studied to jointly maximize the SNR margin at the destinations. Although the two optimization problems can be solved effectively by the bisection methods via SOCP feasibility problems, the chapter

proposes two simple and fast iterative algorithms to directly solve the two problems without the need of a standard conic solution package. Future works on sum-rate maximization of the relay networks are also suggested.

Chapter 7 draws the conclusion and gives suggestions for further studies.

1.2 Notations

The notations in this thesis are quite standard and explained as follows:

- $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively.
- A column vector is formatted in lower-case and bold, e.g., $\boldsymbol{x}$; whereas a matrix is in upper-case and bold, e.g., $\boldsymbol{A}$.
- $\mathbf{S}^n$ denotes the set of symmetric $n \times n$ matrices, whereas $\mathbf{S}_+^n$ denotes the set of symmetric positive semidefinite matrix. $\boldsymbol{A} \in \mathbf{S}_+^n$ is denoted as $\boldsymbol{A} \succeq \mathbf{0}$.
- $\boldsymbol{I}_M$ stands for the $M \times M$ identity matrix.
- $\mathrm{diag}(d_1, d_2, \ldots, d_M)$ denotes an $M \times M$ diagonal matrix with diagonal elements $d_1, d_2, \ldots, d_M$.
- $\det(\cdot)$ and $\mathrm{tr}(\cdot)$ denote the determinant and trace of a square matrix, respectively.
- $[\boldsymbol{x}]_i$ denotes the ith element of the column vector $\boldsymbol{x}$, whereas $[\boldsymbol{A}]_{ij}$ denotes the element at row i, column j of matrix $\boldsymbol{A}$.
- Superscripts $(\cdot)^{\mathcal{T}}$, $(\cdot)^*$, $(\cdot)^{\mathcal{H}}$, and $(\cdot)^\dagger$ stand for transpose, complex conjugate, complex conjugate transpose, and matrix pseudo-inverse operations, respectively.
- $\mathbb{E}_x[\cdot]$ and $\mathrm{var}_x[\cdot]$ indicate the expectation and variance of random variable x, respectively; whereas $x^\star$ denotes the optimal value of variable x.

- $\mathcal{CN}(0, \sigma^2)$ denotes a circularly symmetric complex Gaussian random variable with variance σ^2.
- $|\cdot|$ and $\|\cdot\|$ denote the absolute value and standard Euclidean norm, respectively.

Part I

Distributed Space-Time Coding

2. Distributed Space-Time Coding: Design Criteria and Performance Analysis

2.1 Introduction

Distributed space-time coding (DSTC) [6, 8, 9, 18] refers a cooperative strategy, where the conventional space-time coding for co-located antennas is implemented between the relays in a distributed manner. More specifically, when the source signal is received at the relays, it is linearly processed and then retransmitted to the destination in the form of a space-time codeword. Like the conventional space-time coding in MIMO systems, where transmit diversity is exploited without the need of CSI at the transmitter (see Appendix A), DSTC in a wireless relay network allows the relays to exploit cooperative diversity without the availability of CSI at the relays. As a result, it is well known that the transmission reliability of the source signal over the relay network can be significantly improved.

In the last few years, there has been a lot of research studies in DSTC, for both AF and DF protocols. With the DF protocol, DSTC was first studied in [6,10,19]. In [19], the authors present a new type of distributed space-time block codes (DSTBC) for wireless networks with a large number of users, where each user is assigned a unique signature vector. With the AF protocol, DSTC has been investigated in [8–10,20–22]. Performance analysis and design criteria of DSTC are thoroughly studied in [8, 10].

It is noted that most of the works on cooperative communications in the literature assume the availability of perfect channel state information (CSI) of all the channels

[0]The contributions in this chapter are also presented in [16, 17].

at the relays and/or destination, and only a few of them have considered the scenarios where only imperfect channel estimation or no CSI is available. Reference [23] investigated the noncoherent and mismatched coherent detectors for distributed STBC with one relay, where it is shown that the system can achieve a diversity order of 2. However, the derivations (see (19) and (20) in [23]) for the suboptimal receiver in the noncoherent detection turn the problem into the partially coherent one. The partially coherent relay network was investigated in detail in [22], where a differential coding scheme was proposed to take advantage of the cooperative diversity. For noncoherent relay networks, a fully-diverse distributed coding scheme based on division algebra was proposed in [20, 21]. Similar to the work in [8], references [20–22] consider AF protocol with linear processing at the relays.

Noncoherent reception for DSTC was also proposed in [24], where the decode-and-forward (DF), selection relaying (SR), incremental DF, and incremental SR protocols were employed, and USTM was implemented in a distributed fashion among the relays. A similar approach using DF protocol was also reported in [19, 25]. However, the drawback with these approaches is that the relay only forwards if it decodes the source signal correctly in the SR protocol, or it always forwards in the DF protocol. Given the random nature of the channels in wireless relay networks, for instance, when the channels between the source and some of the relays are bad, it is highly probable that the relays would decode incorrectly and thus not forward in the SR protocol, or forward the incorrect version of the source signal in the DF protocol. This will compromise the diversity advantage of DSTC. Moreover, the incremental DF and incremental SR require feedback from the destination to all the relays. The AF protocol is generally preferred to other protocols since it is always able to achieve the maximum diversity order and feedback from the destination is not required. Moreover, the AF protocol requires much less delay tolerance and infers no security problem as in the DF procotol.

This chapter focuses on the DTSC design in wireless relay networks with the AF protocol. The first part of the chapter reviews the DSTC with full CSI at the desti-

nation, i.e., coherent relay networks. From the performance analysis of the DSTC in coherent networks, the design criteria of a good DSTC scheme are then presented. It will be shown that the space-time code that achieves the full diversity and maximum coding gain in traditional MIMO systems also achieves the diversity order in DSTC implementation. This chapter then concentrates on the design and performance analysis of DSTC over the partially coherent networks (only relay-to-destination CSI is available) and the noncoherent networks (no CSI is available). It shows how to incorporate the Fourier-based USTM into wireless relay networks in a distributed manner. Developed are the maximum likelihood (ML) receiver for the distributed USTM (DUSTM) over partially coherent relay networks and the generalized likelihood-ratio test (GLRT) receiver for the noncoherent relay networks. Performance comparison of the DUSTM over the two types of relay networks reveals that, although the knowledge of relay-to-destination channels improves the symbol error rate (SER) compared to the case of fully noncoherent networks, this advantage diminishes as the total system power becomes large enough. In fact, it is shown that their performances are asymptotically the same when all the relays are active. The full diversity order, equal to the number of relays, is achievable in both networks if the coherence time is larger than twice the number of relays.

2.2 System Model

Consider a wireless relay network with $R+2$ nodes, as illustrated in Figure 2.1. The system has one source node, one destination node, and R relay nodes. Each node is equipped with only one antenna, which can be used for both reception and transmission in the half-duplex mode. There is no direct link from the source to the destination, and in order to facilitate communications between the source and the destination, the source signals are assisted by all the relays. Let $\tilde{f}_i \sim \mathcal{CN}(0, \tilde{\sigma}_{F_i}^2)$, and $\tilde{g}_i \sim \mathcal{CN}(0, \tilde{\sigma}_{G_i}^2)$ be the channel coefficients from the source to the ith relay, and from the ith relay to the destination, for $i = 1, \ldots, R$. These channel coefficients are assumed to be independent of each other, and constant over the coherence time T_C.

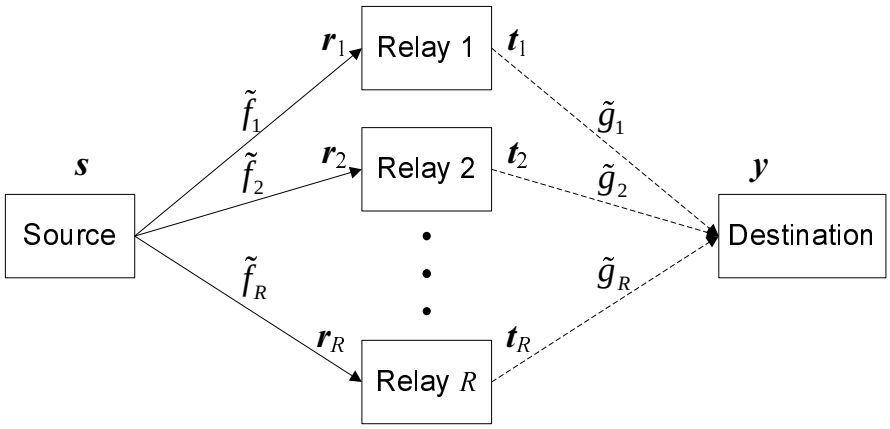

Figure 2.1 Block diagram of a distributed space-time coding system with $R + 2$ nodes.

Let $\mathcal{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L\}$ be the codebook consisting of L distinguished codewords of length $T < T_C$ employed by the source, where $\mathbb{E}[\boldsymbol{s}_k^{\mathcal{H}} \boldsymbol{s}_k] = 1$, for $k = 1, \ldots, L$. In the first stage, the source transmits vector $\sqrt{P_0 T}\boldsymbol{s}$ over T symbol intervals, such that P_0 is the average power per transmission. The received signal at relay-i can be written as

$$\boldsymbol{r}_i = \sqrt{P_0 T}\tilde{f}_i \boldsymbol{s} + \boldsymbol{z}_{R_i}, \tag{2.1}$$

where the noise vector $\boldsymbol{z}_{R_i}$ contains identical and independently distributed (i.i.d.) $\mathcal{CN}(0, \sigma_R^2)$ random variables. The AF protocol [5] with linear signal processing is applied at each relay. In particular, similar to [8], a *unitary relay matrix* $\boldsymbol{A}_i$ of size $T \times T$ is used to linearly process the received signal at the ith relay and form the retransmitted signal as

$$\boldsymbol{t}_i = \sqrt{\frac{P_i}{P_0 \tilde{\sigma}_{F_i}^2 + \sigma_R^2}} \boldsymbol{A}_i \boldsymbol{r}_i^{(*)} = \sqrt{\frac{\varepsilon_i}{\sigma_R^2}} \boldsymbol{A}_i \boldsymbol{r}_i^{(*)}, \quad i = 1, \ldots, R, \tag{2.2}$$

where the normalization factor $\varepsilon_i = P_i/(P_0 \sigma_{F_i}^2 + 1)$ with $\sigma_{F_i}^2 = \tilde{\sigma}_{F_i}^2/\sigma_R^2$ is to maintain the average transmitted power of P_i at the ith relay. Herein, $(\cdot)^{(*)}$ denotes the entity itself if the relay operates on $(\cdot)$, whereas it denotes the conjugate of the entity if the relay operates on $(\cdot)^*$. In the second stage, all the relays simultaneously retransmit to the destination. Let $\boldsymbol{z}_D$, whose elements are i.i.d. $\mathcal{CN}(0, \sigma_D^2)$, represent the AWGN vector at the destination. With perfectly synchronized transmissions from the relays,

the received signal at the destination can be formed as

$$\begin{aligned}\tilde{\boldsymbol{y}} &= \sum_{i=1}^{R} \tilde{g}_i \boldsymbol{t}_i + \boldsymbol{z}_D \\ &= \frac{\sqrt{P_0 T}}{\sigma_R} \sum_{i=1}^{R} \sqrt{\varepsilon_i} \tilde{f}_i^{(*)} \tilde{g}_i \boldsymbol{A}_i \boldsymbol{s}^{(*)} + \frac{1}{\sigma_R} \sum_{i=1}^{R} \sqrt{\varepsilon_i} \tilde{g}_i \boldsymbol{A}_i \boldsymbol{z}_{R_i}^{(*)} + \boldsymbol{z}_D. \end{aligned} \tag{2.3}$$

Normalize both sides by $\sqrt{P_0 T \sigma_D^2}$, and denote $\sigma_{G_i}^2 = \tilde{\sigma}_{G_i}^2 / \sigma_D^2$, one has

$$\boldsymbol{y} = \frac{\tilde{\boldsymbol{y}}}{\sqrt{P_0 T \sigma_D^2}} = \boldsymbol{X} \boldsymbol{\Lambda} \boldsymbol{h} + \boldsymbol{z}, \tag{2.4}$$

where

$$\begin{aligned} \boldsymbol{X} &= \left[\boldsymbol{A}_1 \boldsymbol{s}^{(*)}, \ldots, \boldsymbol{A}_R \boldsymbol{s}^{(*)} \right] \\ \boldsymbol{\Lambda} &= \operatorname{diag}\left(\sqrt{\varepsilon_1 \sigma_{F_1}^2 \sigma_{G_1}^2}, \ldots, \sqrt{\varepsilon_R \sigma_{F_R}^2 \sigma_{G_R}^2} \right) \\ \boldsymbol{h} &= \left[f_1^{(*)} g_1, \ldots, f_R^{(*)} g_R \right]^{\mathcal{T}} \\ \boldsymbol{z} &= \frac{1}{\sqrt{P_0 T \sigma_R^2}} \sum_{i=1}^{R} \sqrt{\varepsilon_i \sigma_{G_i}^2} g_i \boldsymbol{A}_i \boldsymbol{z}_{R_i}^{(*)} + \frac{1}{\sqrt{P_0 T \sigma_D^2}} \boldsymbol{z}_D. \end{aligned} \tag{2.5}$$

From the system model in (2.4), it can be interpreted that the $T \times R$ matrix $\boldsymbol{X}$ works like a space-time codeword in a multiple-antenna system. In this thesis, $\boldsymbol{X}$ is called as a distributed space-time codeword. The diagonal matrix $\boldsymbol{\Lambda}$ contains power allocation factors, which can be treated separately from the effective channel vector $\boldsymbol{h}$. It is noted that the ith element of $\boldsymbol{h}$ is a multiplication of the normalized channel factor f_i and g_i, where $f_i = \tilde{f}_i / \sigma_{F_i}$ and $g_i = \tilde{g}_i / \sigma_{G_i}$ are now i.i.d. $\mathcal{CN}(0,1)$. Finally, the noise vector $\boldsymbol{z}$, conditioned on $\{g_i\}$, contains i.i.d. Gaussian variables with zero-mean and variance:

$$\gamma = \frac{1}{P_0 T} \left(1 + \sum_{i=1}^{R} \varepsilon_i \sigma_{G_i}^2 |g_i|^2 \right). \tag{2.6}$$

The aim of Part I of this thesis is to design the distributed space-time codewords depending on the availability of CSI at the destination and the optimal power allocation scheme to maximize the performance of the relay networks. It is clear that the design of $\boldsymbol{X}$ can be divided into designing the structure of the source signal $\boldsymbol{s}$ and the relaying matrices $\boldsymbol{A}_i$, whereas the power allocation solely depends on $\boldsymbol{\Lambda}$.

2.3 Distributed Space-Time Coding in Coherent Relay Networks

In a coherent relay network, the CSI is fully known at the destination. Thus, coherent detection is possible. Let $\boldsymbol{s}_k$ be the source signal (sequence) and $\boldsymbol{X}_k$ be the distributed ST codeword associated with $\boldsymbol{s}_k$. As both $\{f_i\}$ and $\{g_i\}$ are known at the destination, the received vector $\boldsymbol{y}$ at the destination is Gaussian distributed with the mean $\boldsymbol{X}_k\boldsymbol{\Lambda h}$, and variance $\gamma \boldsymbol{I}_T$. Thus, one has the following conditional probability density function (pdf) for $\boldsymbol{y}$:

$$p\left(\boldsymbol{y}|\boldsymbol{s}_k\right) = \frac{\exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{X}_k\boldsymbol{\Lambda h}\|^2}{\gamma}\right)}{\pi^T\gamma^T} \tag{2.7}$$

In order to decode the original source signal, the coherent maximum likelihood (ML) decoder performs

$$\hat{\boldsymbol{s}} = \arg\max_{\boldsymbol{s}_k\in\mathcal{S}} p(\boldsymbol{y}|\boldsymbol{s}_k) = \arg\min_{\boldsymbol{s}_k\in\mathcal{S}} \|\boldsymbol{y}-\boldsymbol{X}_k\boldsymbol{\Lambda h}\|^2 . \tag{2.8}$$

From the ML decoding rule of coherent DSTC in (2.8), suppose that $\boldsymbol{X}_k$ is the transmitted codeword, the probability of decoding the received signal as the codeword $\boldsymbol{X}_l$ conditioned on the effective channel vector $\boldsymbol{h}$ can be calculated and bounded as:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l|\boldsymbol{h}) &= \mathbb{P}\left(\|\boldsymbol{y}-\boldsymbol{X}_k\boldsymbol{\Lambda h}\|^2 \geq \|\boldsymbol{y}-\boldsymbol{X}_l\boldsymbol{\Lambda h}\|^2\right) \\
&= \mathbb{P}\left(\|\boldsymbol{\Delta\Lambda h}\|^2 + 2\mathrm{Re}\{\boldsymbol{z}^{\mathcal{H}}\boldsymbol{\Delta\Lambda h}\} \leq 0\right) \\
&= Q\left(\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{\sqrt{2\gamma\|\boldsymbol{\Delta\Lambda h}\|^2}}\right) \\
&= Q\left(\sqrt{\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{2\gamma}}\right) \\
&< \frac{1}{2}\exp\left(-\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{4\gamma}\right),
\end{aligned} \tag{2.9}$$

where $\boldsymbol{\Delta} = \boldsymbol{X}_k - \boldsymbol{X}_l$, and the last inequality comes from $Q(x) < \frac{1}{2}\exp\left(-\frac{x^2}{2}\right)$.[1]

[1]The Q-function is defined as $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-\frac{u^2}{2}\right)\mathrm{d}u$.

In order to find the average PEP $\mathbb{P}(\boldsymbol{X}_k \rightarrow \boldsymbol{X}_l)$, $\mathbb{P}(\boldsymbol{X}_k \rightarrow \boldsymbol{X}_l|\boldsymbol{h})$ needs to be averaged over the distribution of $\boldsymbol{h}$, that is

$$\mathbb{P}(\boldsymbol{X}_k \rightarrow \boldsymbol{X}_l) = \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}} \left[Q\left(\sqrt{\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{2\gamma}} \right) \right] < \frac{1}{2} \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}} \left[\exp\left(-\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{4\gamma} \right) \right]. \quad (2.10)$$

It is clear from the above evaluation that in order to minimize the upper-bound of the PEP, $\boldsymbol{\Delta}$ has to be full rank and the determinant of $\boldsymbol{\Delta}^{\mathcal{H}}\boldsymbol{\Delta}$ has to be maximized. This is equivalent to maximizing the diversity and coding gains of the DTSC.[2] Interestingly, the space-time code designed to achieve the full diversity and optimum coding gains in conventional MIMO systems is also able to achieve the optimum performance in DSTC implementation. Thus, the design criteria applied to space-time coding in the MIMO conventional systems [4] are the same for designing DSTC over a relay network [8]. Moreover, in space-time coding, it is required that the transmission time to be at least equal to the number of transmit antennas to achieve the maximum diversity order [4]. Thus, in DSTC, the transmission T needs to be at least equal to the number of relays R in order to achieve the maximum diversity order of R.

Due to the distributed implementation of DSTC in relay networks, it is possible that some of the relay nodes are not available or in *failure* to assist the source-destination communication. Reference [8] argues that a good DSTC scheme should be *scale-free*. More specifically, the DSTC scheme is required to perform well even if some of the relays are not working. In addition, the diversity order of the distributed code should equal the number of the remaining working nodes.

Since good space-time coding schemes in traditional MIMO systems also work well with DSTC implementation, many known good ST designs have been applied to the relay networks. For instance, distributed linear dispersion (LD) code was considered in [8], whereas distributed space-time block code (STBC) was considered in [9]. However, given that STBC in MIMO systems possesses many advantages such as simple decoding (linearly for orthogonal codes or in a pairwise manner for quasi-

[2] A more detailed discussion on diversity and coding gains of space-time coding is given in Appendix A.

orthogonal codes) and *scale-free* property (removing some columns of an orthogonal code does not affect the orthogonality of the remaining columns), it is an obvious choice for DSTC. In the following, two examples of applying STBC to relay networks in a distributed fashion are given.

For a two-relay network, the distributed Alamouti code can be implemented as follows:

- The source signal is formed as $\boldsymbol{s} = [s_1,\ s_2]^{\mathcal{T}}$.
- The relaying matrices are:

$$\boldsymbol{A}_1 = \boldsymbol{I}_2,\ \boldsymbol{A}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

If the first relay operates on $\boldsymbol{s}$ and the second relay operates on $\boldsymbol{s}^*$, the codeword at the destination is

$$\boldsymbol{X} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix},$$

which is a typical Alamouti codeword for 2 transmit antennas (see Appendix A).

For a four-relay network, distributed quasi-orthogonal STBC (QOSTBC) can be implemented as follows:

- The source signal is formed as $\boldsymbol{s} = [s_1, s_2, s_3, s_4]^{\mathcal{T}}$.
- The relay matrices are:

$$\boldsymbol{A}_1 = \boldsymbol{I}_4, \qquad \boldsymbol{A}_2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\boldsymbol{A}_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},\ \boldsymbol{A}_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

If the first and the fourth relays operate on $\boldsymbol{s}$, and the other two operate on $\boldsymbol{s}^*$, then the effective codeword at the destination is

$$\boldsymbol{X} = \begin{bmatrix} s_1 & -s_2^* & -s_3^* & s_4 \\ s_2 & s_1^* & -s_4^* & -s_3 \\ s_3 & -s_4^* & s_1^* & -s_2 \\ s_4 & s_3^* & s_2^* & s_1 \end{bmatrix},$$

which is a typical QOSTBC for 4 transmit antennas [26].

2.4 Distributed Unitary Space-Time Modulation in Partially Coherent and Noncoherent Relay Networks

This section considers the design of DSTC when full CSI is not available at the destination and noncoherent detection is required. First, a brief review of Fourier-based unitary space-time constellation designs for noncoherent communication with multiple co-located transmit antennas, originally proposed in [27], is given. In this system model, the transmitter is equipped with M antennas, and the channel remains constant over T symbol times. The USTM design constructs a constellation of L unitary matrices, $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_L$ (each of size $T \times M$), such that $\boldsymbol{\Phi}_k^{\mathcal{H}} \boldsymbol{\Phi}_k = \boldsymbol{I}_M$, for $k = 1, \ldots, L$. From this pool of unitary matrices, the transmitted signal matrix is formed as $\sqrt{T}\boldsymbol{\Phi}_k$. Let $\boldsymbol{y}$ be the received signal vector. The ML receiver for USTM with *noncoherent reception* was shown in to be (see Appendix A):

$$\hat{\boldsymbol{\Phi}} = \arg \max_{\boldsymbol{\Phi}_k = \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_L} \operatorname{tr} \left\{ \boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^{\mathcal{H}} \boldsymbol{y} \right\}. \tag{2.11}$$

In [27], the authors proposed a Fourier-based approach in designing the unitary constellations, which uses ideas from signal processing theory. In this design, the kth constellation point, $\boldsymbol{\Phi}_k$, can be obtained from the first constellation point $\boldsymbol{\Phi}_1$ as

$$\boldsymbol{\Phi}_k = \boldsymbol{\Theta}^{k-1} \boldsymbol{\Phi}_1 \tag{2.12}$$

where

$$\boldsymbol{\Theta} = \operatorname{diag}[\mathrm{e}^{j\frac{2\pi}{L}u_1}, \ldots, \mathrm{e}^{j\frac{2\pi}{L}u_T}], \quad 0 \leq u_1, \ldots, u_T \leq L-1 \tag{2.13}$$

and $\mathbf{\Phi}_1$ is constructed by selecting M columns of a $T \times T$ DFT matrix. The optimal values of the so-called *frequencies* $u_1, \ldots, u_T$ are also given in [27] for the cases of 1, 2 and 3 transmit antennas.

Next, consider the application of USTM for distributed space-time coding in wireless relay networks, where full CSI is not available neither at the source, the relays, nor at the destination, and hence noncoherent detection is required. First, the kth codeword vector $\boldsymbol{s}_k$ from the source is formed by taking the diagonal elements of matrix $\mathbf{\Theta}^{k-1}$ and scaling by $1/\sqrt{T}$ to meet the power constraint. That is $\boldsymbol{s}_k = 1/\sqrt{T}[\mathrm{e}^{j\frac{2\pi}{L}u_1(k-1)}, \ldots, \mathrm{e}^{j\frac{2\pi}{L}u_T(k-1)}]^{\mathcal{T}}$. Second, the unitary matrix at each relay is constructed by diagonalizing one of the columns of $\mathbf{\Phi}_1$. As an example, consider the following design of $\mathbf{\Phi}_1$ for $M = 3$ transmit antennas, and $T = 8$, given in [27]:

$$\mathbf{\Phi}_1 = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}5} & \mathrm{e}^{j\frac{2\pi}{8}6} \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}2} & \mathrm{e}^{j\frac{2\pi}{8}4} \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}7} & \mathrm{e}^{j\frac{2\pi}{8}2} \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}4} & 1 \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}1} & \mathrm{e}^{j\frac{2\pi}{8}6} \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}6} & \mathrm{e}^{j\frac{2\pi}{8}4} \\ 1 & \mathrm{e}^{j\frac{2\pi}{8}3} & \mathrm{e}^{j\frac{2\pi}{8}2} \end{bmatrix}.$$

Then $\boldsymbol{A}_i$'s for a three-relay network are formed as

$$\begin{aligned} \boldsymbol{A}_1 &= \boldsymbol{I}_8, \\ \boldsymbol{A}_2 &= \mathrm{diag}[1, \mathrm{e}^{j\frac{2\pi}{8}5}, \mathrm{e}^{j\frac{2\pi}{8}2}, \mathrm{e}^{j\frac{2\pi}{8}7}, \mathrm{e}^{j\frac{2\pi}{8}4}, \mathrm{e}^{j\frac{2\pi}{8}1}, \mathrm{e}^{j\frac{2\pi}{8}6}, \mathrm{e}^{j\frac{2\pi}{8}3}], \\ \boldsymbol{A}_3 &= \mathrm{diag}[1, \mathrm{e}^{j\frac{2\pi}{8}6}, \mathrm{e}^{j\frac{2\pi}{8}4}, \mathrm{e}^{j\frac{2\pi}{8}2}, 1, \mathrm{e}^{j\frac{2\pi}{8}6}, \mathrm{e}^{j\frac{2\pi}{8}4}, \mathrm{e}^{j\frac{2\pi}{8}2}], \end{aligned}$$

where the normalization factor $1/\sqrt{8}$ is dropped to make $\boldsymbol{A}_i^{\mathcal{H}}\boldsymbol{A}_i = \boldsymbol{I}_8$. With this design, the codeword $\boldsymbol{X}_k$ in (2.4) is effectively in the form of (2.12). Therefore, the noncoherent detection of the codeword vector $\boldsymbol{s}_k$ can be carried out similarly as in (2.11), where the received signal matrix $\boldsymbol{Y}$ in (2.11) is substituted by the received signal vector $\boldsymbol{y}$ in (2.4).

2.4.1 ML Receiver for DUSTM over the Partially Coherent Relay Network

In the partially coherent relay network considered in this part, the destination has perfect knowledge of all the channels from the relays, but not the channels from the source to the relays. This means that $\{g_i\}$ is known, while $\{f_i\}$ is unknown, for $i = 1, \ldots, R$.

Conditioned on $\{g_i\}$ and the transmitted codeword $\boldsymbol{X}_k$, the received vector $\boldsymbol{y}$ is a circularly symmetric Gaussian vector with covariance matrix

$$\boldsymbol{\Omega} = \gamma \boldsymbol{I}_T + \boldsymbol{X}_k \boldsymbol{G} \boldsymbol{X}_k^{\mathcal{H}}, \tag{2.14}$$

where $\boldsymbol{G} = \text{diag}\,(\beta_1|g_1|^2, \ldots, \beta_R|g_R|^2)$, and $\beta_i = \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2$ is the ith diagonal element of $\boldsymbol{\Lambda}^2$. The received signal vector has the following conditional pdf:

$$p(\boldsymbol{y}|\boldsymbol{X}_k, \{g_i\}) = \frac{\exp(-\boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Omega}^{-1} \boldsymbol{y})}{\pi^T \det(\boldsymbol{\Omega})}. \tag{2.15}$$

Using property $\det(\boldsymbol{I} + \boldsymbol{AB}) = \det(\boldsymbol{I} + \boldsymbol{BA})$ [28], the determinant of $\boldsymbol{\Omega}$ can be found as

$$\begin{aligned} \det(\boldsymbol{\Omega}) &= \gamma^T \det\left(\boldsymbol{I}_T + \boldsymbol{X}_l \overline{\boldsymbol{G}} \boldsymbol{X}_l^{\mathcal{H}}\right) \\ &= \gamma^T \det(\boldsymbol{I}_R + \overline{\boldsymbol{G}}) \\ &= \gamma^T \prod_{i=1}^{R} \left(1 + \frac{\beta_i}{\gamma}|g_i|^2\right), \end{aligned}$$

where $\overline{\boldsymbol{G}} = \boldsymbol{G}/\gamma$. Likewise, using the matrix inverse formula [28]

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{D}\boldsymbol{A}^{-1},$$

the inverse of $\boldsymbol{\Omega}$ can be calculated as

$$\begin{aligned} \boldsymbol{\Omega}^{-1} &= \frac{1}{\gamma}\left(\boldsymbol{I}_T + \boldsymbol{X}_k \overline{\boldsymbol{G}} \boldsymbol{X}_k^{\mathcal{H}}\right)^{-1} \\ &= \frac{1}{\gamma}\left(\boldsymbol{I}_T - \boldsymbol{X}_k (\overline{\boldsymbol{G}}^{-1} + \boldsymbol{X}_k^{\mathcal{H}} \boldsymbol{I}_T \boldsymbol{X}_k)^{-1} \boldsymbol{X}_k^{\mathcal{H}}\right) \\ &= \frac{1}{\gamma}\left(\boldsymbol{I}_T - \boldsymbol{X}_k \boldsymbol{C} \boldsymbol{X}_k^{\mathcal{H}}\right), \end{aligned}$$

where

$$\boldsymbol{C} = \mathrm{diag}\left(\frac{\beta_1|g_1|^2}{\gamma+\beta_1|g_1|^2}, \ldots, \frac{\beta_R|g_R|^2}{\gamma+\beta_R|g_R|^2}\right).$$

The partially coherent ML receiver then becomes

$$\begin{aligned}\hat{\boldsymbol{X}}_{ML} &= \arg \max_{\boldsymbol{X}_k=\boldsymbol{X}_1,\ldots,\boldsymbol{X}_L} \mathbb{P}(\boldsymbol{y}|\boldsymbol{X}_k, \{g_i\}) \\ &= \arg \max_{\boldsymbol{X}_k=\boldsymbol{X}_1,\ldots,\boldsymbol{X}_L} -\frac{1}{\gamma}\boldsymbol{y}^{\mathcal{H}}\left(\boldsymbol{I}_T - \boldsymbol{X}_k\boldsymbol{C}\boldsymbol{X}_k^{\mathcal{H}}\right)\boldsymbol{y} \\ &= \arg \max_{\boldsymbol{X}_k=\boldsymbol{X}_1,\ldots,\boldsymbol{X}_L} \boldsymbol{y}^{\mathcal{H}}\boldsymbol{X}_k\boldsymbol{C}\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{y}. \end{aligned} \tag{2.16}$$

2.4.2 GLRT Receiver for DUSTM over the Noncoherent Relay Network

For a (fully) noncoherent relay network, neither the CSI of the relay-to-destination channels nor the CSI of the source-to-relay channels is known at the destination. Since each element of $\boldsymbol{h}$ in (2.4) is a product of two complex Gaussian random variables, the source-relay-destination link is represented by *cascaded* fading. Furthermore, the conditional pdf $p(\boldsymbol{y}|\boldsymbol{X}_k)$ does not appear to have a closed-form expression. Thus, it is not trivial to derive the optimal ML receiver for the network. Instead, the suboptimal GLRT receiver [27, 29] shall be considered.

With GLRT, the receiver first estimates the channel $\boldsymbol{h}$ under the hypothesis that the codeword $\boldsymbol{X}_k$ was sent. From (2.4), conditioned on the transmitted codeword $\boldsymbol{X}_k$, $\{f_i\}$, and $\{g_i\}$, the received vector $\boldsymbol{y}$ is a Gaussian random vector with mean $\boldsymbol{X}_k\boldsymbol{\Lambda}\boldsymbol{h}$ and covariance matrix $\gamma\boldsymbol{I}_T$, the same covariance matrix of the noise vector. Thus, the ML estimation of $\boldsymbol{h}$ is

$$\begin{aligned}\hat{\boldsymbol{h}}_k &= \arg\max_{\boldsymbol{h}} p(\boldsymbol{y}|\boldsymbol{X}_k, \boldsymbol{h}) \\ &= \arg\min_{\boldsymbol{h}} \|\boldsymbol{y} - \boldsymbol{X}_k\boldsymbol{\Lambda}\boldsymbol{h}\|^2. \end{aligned} \tag{2.17}$$

It then follows that $\hat{\boldsymbol{h}}_k$ is given by

$$\hat{\boldsymbol{h}}_k = \boldsymbol{\Lambda}^{-1}\left(\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{X}_k\right)^{-1}\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{y} = \boldsymbol{\Lambda}^{-1}\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{y}, \tag{2.18}$$

where the last equality follows from the property that $\boldsymbol{X}_k$ is unitary. Substitute $\hat{\boldsymbol{h}}_k$ into (2.4), the GRLT receiver is expressed as

$$\begin{aligned}\hat{\boldsymbol{X}}_{GLRT} &= \arg \max_{\boldsymbol{X}_k=\boldsymbol{X}_1,\ldots,\boldsymbol{X}_L} \left\{-\|\boldsymbol{y}-\boldsymbol{X}_k\boldsymbol{\Lambda}\hat{\boldsymbol{h}}_k\|^2\right\} \\ &= \arg \max_{\boldsymbol{X}_k=\boldsymbol{X}_1,\ldots,\boldsymbol{X}_L} \boldsymbol{y}^{\mathcal{H}}\boldsymbol{X}_k\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{y}. \end{aligned} \tag{2.19}$$

Observe that the GLRT receiver in (2.19) operates in the same way as the GLRT receiver for co-located multiple transmit antennas in [27], and the receiver for DUSTM with DF relaying protocol in [24,25]. Comparing the receiver in (2.16) for the partially coherent network and the one in (2.19) for the noncoherent network, it can be seen that the difference is in the existence of the matrix $\boldsymbol{C}$ in the former one. The matrix $\boldsymbol{C}$ contains the CSI of the relay-to-destination channels. However, as the signal power becomes large enough, the matrix $\boldsymbol{C}$ comes closer to an identity matrix and, therefore, the two receivers are basically the same. This observation is reconfirmed in the next sections with the pairwise error probability (PEP) analysis.

2.4.3 PEP of DUSTM over the Partially Coherent Relay Network

Here, the objective is to evaluate the pairwise error probability (PEP) performance of partially coherent DUSTM and relates it to the constellation design of DUSTM. Suppose that all the relays are active, which means that the matrix $\boldsymbol{G}$ is full rank. In [22], the authors derive the PEP for the partially coherent DUSTM, its Chernoff bound, as well as an approximation for the average symbol error rate (SER) at high SNR. The result is summarized as followed. Suppose that $\boldsymbol{X}_k$ and $\boldsymbol{X}_l$ are two codewords and $\boldsymbol{Z}_{kl} = [\boldsymbol{X}_k,\ \boldsymbol{X}_l]$ is full rank. Define $\boldsymbol{R}_{kl} = \boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{X}_l$ and let $\boldsymbol{R}_{kk} = \boldsymbol{R}_{ll} = \boldsymbol{K}$. The error probability of decoding to $\boldsymbol{X}_l$ for large η, given that $\boldsymbol{X}_k$ was transmitted, is approximated as

$$\mathbb{P}_{k,l|\{g_i\}} \approx \frac{\gamma^R}{\det(\mu(1-\mu)\boldsymbol{G})}\frac{1}{\det(\boldsymbol{K}-\boldsymbol{R}_{lk}\boldsymbol{K}^{-1}\boldsymbol{R}_{kl})}, \tag{2.20}$$

which is minimized with $\mu = 1/2$.

On the other hand, since $\boldsymbol{y}$ is Gaussian distributed, and $\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{C}\boldsymbol{X}_k$ is Hermitian for $k = 1, \ldots, L$, the ML receiver in (2.16) can be interpreted as a quadratic receiver [30]. The asymptotic PEP performance of the quadratic receiver is readily given as (cf. (28) in [30])

$$\mathbb{P}_{k,l|\{g_i\}} = \frac{\gamma^R}{\det(\boldsymbol{G})} \frac{\binom{2R}{R}}{\det(\boldsymbol{K} - \boldsymbol{R}_{lk}\boldsymbol{K}^{-1}\boldsymbol{R}_{kl})}. \tag{2.21}$$

The two PEP expressions above differ only in a scaling factor and they clearly indicate that the effect of channel coefficients $\{g_i\}$, which are in $\boldsymbol{G}$, can be separated from the effect of the distributed code [22]. With the implementation of DUSTM over partially coherent networks, one has $\boldsymbol{K} = \boldsymbol{I}_R$, $\det(\boldsymbol{K} - \boldsymbol{R}_{lk}\boldsymbol{K}^{-1}\boldsymbol{R}_{kl}) = \prod_{r=1}^{R}(1 - d_r^2)$, where d_r, $r = 1, \ldots, R$, are the singular values of the correlation matrix $\boldsymbol{X}_k^{\mathcal{H}}\boldsymbol{X}_l$. Therefore, the PEP will be minimized when this product is maximized. This is the same condition on the constellation design of USTM for co-located transmit antennas in [27]. This means that the best Fourier-based constellation design in [27] is also the best Fourier-based constellation design for DUSTM. As pointed out in [22], the system achieves the full diversity order, which is equal to the number of relays, if $\boldsymbol{Z}_{kl}$ is full rank for any pair of $\boldsymbol{X}_k$ and $\boldsymbol{X}_l$. The necessary condition for this is $T \geq 2R$, which is similar to the condition imposed in USTM, namely $T \geq 2M$.

In order to calculate the symbol error probability, the conditional PEP expression in (2.20) or (2.21) has to be averaged over the distribution of $\{g_i\}$. This is analyzed in [22] by performing similar derivation steps for the coherent distributed space-time coding in [8]. An important remark from such PEP analysis is that the SER is proportional to $(\log P/P)^R$ as P becomes large enough. Thus, DUSTM over a partially coherent relay network is able to achieve the full diversity for large P, which is the same result obtained for the coherent space-time coding in [8].

The above discussion is only concerned with the diversity order of DUSTM. It should be pointed out that the exact PEP still depends on the relay network's power allocation, which affects the coding gain of DUSTM. The next chapter establishes the optimal power allocation to asymptotically minimize the upper-bound of the PEP of

the partially coherent DUSTM.

2.4.4 PEP of DUSTM over the Noncoherent Relay Network

This section considers the PEP analysis of DUSTM over the noncoherent relay network. Intuitively, the performance of DUSTM over such a network is poorer than that over a partially coherent relay network. However, as the total transmit power P becomes very large, it is shown that the two performances are asymptotically the same.

Recall the GLRT receiver in (2.19). Suppose that the codeword $\boldsymbol{X}_k$ was sent, the PEP of decoding to the wrong codeword $\boldsymbol{X}_l$ is given by

$$\mathbb{P}_{k,l} = \mathbb{P}\left(\boldsymbol{y}^{\mathcal{H}}(\boldsymbol{X}_l\boldsymbol{X}_l^{\mathcal{H}} - \boldsymbol{X}_k\boldsymbol{X}_k^{\mathcal{H}})\boldsymbol{y} > 0|\boldsymbol{X}_k\right). \tag{2.22}$$

Since $\boldsymbol{y}$ is not Gaussian distributed, (2.19) cannot be interpreted as a quadratic receiver [30]. To calculate the PEP as well as its asymptotic behavior, reintroduce $\{g_i\}$ into the above equation as follows:

$$\mathbb{P}_{k,l} = \mathop{\mathbb{E}}_{\{g_i\}} \underbrace{\left[\mathbb{P}\left(\boldsymbol{y}^{\mathcal{H}}(\boldsymbol{X}_l\boldsymbol{X}_l^{\mathcal{H}} - \boldsymbol{X}_k\boldsymbol{X}_k^{\mathcal{H}})\boldsymbol{y} > 0|\boldsymbol{X}_k, \{g_i\}\right)\right]}_{\tilde{\mathbb{P}}_{k,l|\{g_i\}}}. \tag{2.23}$$

In other words, $\mathbb{P}_{k,l}$ can be obtained by taking the expectation of $\tilde{\mathbb{P}}_{k,l|\{g_i\}}$ over $\{g_i\}$ [31].

Conditioned on a specific realization of $\{g_i\}$, $\boldsymbol{y}$ is now Gaussian distributed with zero mean and covariance matrix given in (2.14). Since $\boldsymbol{X}_l\boldsymbol{X}_l^{\mathcal{H}}$ is Hermitian for $l = 1, \ldots, L$, one can interpret the GLRT receiver as a quadratic receiver. Thus its asymptotic performance is given as (cf. (36) in [30])

$$\begin{aligned}\tilde{\mathbb{P}}_{k,l|\{g_i\}} &= \frac{\gamma^R}{\det(\boldsymbol{G})} \frac{\binom{2R-1}{R}\left(1 + \frac{\det(\boldsymbol{K})}{\det(\boldsymbol{K})}\right)}{\det(\boldsymbol{K} - \boldsymbol{R}_{lk}\boldsymbol{K}^{-1}\boldsymbol{R}_{kl})} \\ &= \frac{\gamma^R}{\det(\boldsymbol{G})} \frac{2\binom{2R-1}{R}}{\det(\boldsymbol{K} - \boldsymbol{R}_{lk}\boldsymbol{K}^{-1}\boldsymbol{R}_{kl})}.\end{aligned} \tag{2.24}$$

It can be seen that $\tilde{\mathbb{P}}_{k,l|\{g_i\}}$ in (2.24) and $\mathbb{P}_{k,l|\{g_i\}}$ in (2.21) are essentially the same since $2\binom{2R-1}{R} = \binom{2R}{R}$. Thus, conditioned on $\{g_i\}$, the performance of GLRT

receiver for the noncoherent relay network is asymptotically the same as that of the ML receiver for the partially relay network. Interestingly, this fact also makes the constellation design of DUSTM for noncoherent system analogous to the design of the partially coherent DUSTM. Moreover, the optimal power allocation scheme that minimizes the PEP's upper-bound of the partially coherent DUSTM also minimizes the PEP's upper-bound of the noncoherent DUSTM at high P. To obtain the average PEP $\mathbb{P}_{k,l}$ at high P, $\tilde{\mathbb{P}}_{k,l|\{g_i\}}$ has to be averaged over the distribution of $\{g_i\}$, which is similar to the process discussed in the previous section.

2.4.5 Impact of Non-functioning Relays

In designing a good DSTC, reference [8] points out that the code should be "*scale-free*" in the sense that it should have a large diversity product when one or more of the relays are not functioning. This section investigates the impact of node failures to the performance of the proposed DUSTM, and whether the decoding rules for DUSTM are still valid in such situations.

Without loss of generality, it is assumed that the last d relays out of total R relays are not working. As the destination knows the channels from the relays in the partially coherent networks, the destination also knows which relay(s) is not working. Recall the decoding rule for partially coherent networks in (2.16), it can be seen that the last d diagonal elements of matrix $\boldsymbol{C}$ are now zero. Let

$$\boldsymbol{C}' = \mathrm{diag}\left(\frac{\beta_1|g_1|^2}{\gamma+\beta_1|g_1|^2}, \ldots, \frac{\beta_{R-d}|g_{R-d}|^2}{\gamma+\beta_{R-d}|g_{R-d}|^2}\right),$$

which is full rank. Define $\boldsymbol{X}'_k$ as the $T \times (R-d)$ matrix that contains the first $R-d$ columns of $\boldsymbol{X}_k$. It is easy to see that $\boldsymbol{X}_k\boldsymbol{C}\boldsymbol{X}_k^{\mathcal{H}} = \boldsymbol{X}'_k\boldsymbol{C}'\boldsymbol{X}_k'^{\mathcal{H}}$. The decoding rule in (2.16) is then equivalent to

$$\boldsymbol{X}'_{ML} = \arg\max_{\boldsymbol{X}'_1,\ldots,\boldsymbol{X}'_L} \boldsymbol{y}^{\mathcal{H}}\boldsymbol{X}'_k\boldsymbol{C}'\boldsymbol{X}_k'^{\mathcal{H}}\boldsymbol{y},$$

which is the same as the decoding rule for the system with $R-d$ relays. Hence, it is concluded that the relay network is able to achieve the diversity order of $R-d$ at very high P.

In noncoherent relay networks, it might not be possible that the destination knows which relay(s) is not working. Since the decoding rule for noncoherent detection in (2.19) is similar to the decoding rule for the DUSTM using selection relay (SR) protocol in [24], the analysis on the impact of node failures in the proposed DUSTM scheme is also similar to that of incorrect decoding and non-forwarding relays with the SR protocol. It was shown in [24] (cf. (9)) that, when d out of R relays are not functioning, the PEP derived from the decoding rule (2.19) decays at the order of $R-d$ for very large SNR. A similar conclusion could be drawn for the proposed noncoherent AF DUSTM.

Note that in both the partially coherent and noncoherent receptions, the optimal *frequencies* $u_1, \ldots, u_T$ specifically designed for the network with R relays are no longer optimal for the network with $R-d$ relays. Finally, it should be pointed out that in order to achieve the best SER performance of the distributed code from the remaining $R-d$ functioning relays, the optimal power allocation scheme needs to be adjusted. Such an adjustment can be readily made from the analysis in Chapter 3 applied to $R-d$ relays.

2.5 Simulation Results

This section presents the simulation results to illustrate the performance of DUSTM for some specific configurations of relay networks that have 2 or 3 relays. The data rate is set at 1 bit/channel use. All the channels are assumed to remain constant for $T = 8$. Thus, the codeword $\boldsymbol{X}_k$ is a 8×3 matrix for 3 relays, or a 8×2 matrix for 2 relays. The DUSTM constellations are chosen from the optimal designs for 2 and 3 transmit antennas [27]. Specifically, the sets of frequencies are $\{1, 7, 60, 79, 187, 125, 198, 154\}$ and $\{220, 191, 6, 87, 219, 236, 173, 170\}$ for 2 and 3 relays, respectively. In all simulations, σ_R^2 and σ_D^2 are normalized to 1.

2.5.1 Performance Comparison of DUSTM over Partially Coherent, Noncoherent, and Coherent Networks

This part compares the performances of DUSTM over different types of relay networks, including partially coherent, noncoherent and coherent networks. Emphasis shall be placed on the comparison between partially coherent and noncoherent networks in order to illustrate the analysis in Sections 2.4.3 and 2.4.4. At each value of the total power P, the equal power allocation scheme is applied, i.e., $P_0 = P/2$, $P_i = P/(2R), i = 1, \ldots, R$.

Three balanced network configurations are considered. Let σ_F^2 and σ_G^2 denote the common variances of $S \to R$, and $R \to D$ links after normalized by the noise variances σ_R^2 and σ_D^2, respectively. In the first configuration, the relays are located in the midway between the source and destination. The variances of the channel coefficients f_i and g_i are set to be unity. As can be seen from Figure 2.2, the partially coherent DUSTM outperforms the noncoherent DUSTM in both two-relay and three-relay configurations. However, the performance difference is only noticeable at small values of P, i.e., over the low SNR region. As P increases, the performance curves merge together, which agrees with the asymptotic analysis in previous sections. Another observation is that when P surpasses 18 dB, the DUSTM for the system with 3 relays starts to perform better than the one with 2 relays and clearly shows its higher diversity order.

The second simulation investigates the case when the relays are closer to the source than to the destination. This implies that the source-to-relay channels experience a much better condition than the relay-to-destination channels. Specifically σ_F^2 is to be 10, and σ_G^2 is set to 1. The performance of DUSTM in this network configuration is shown in Figure 2.3.

The next network configuration considered is the opposite of the second one. The relays are assumed to be closer to the destination than to the source, with $\sigma_F^2 = 1$ and $\sigma_G^2 = 10$. The relay-to-destination channels are relatively much better than

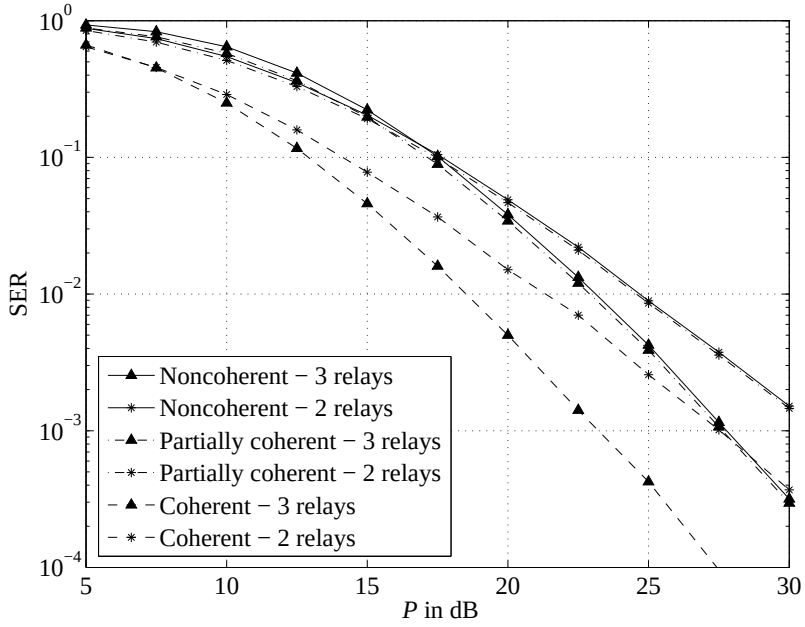

Figure 2.2 Symbol error performance of DUSTM with $\sigma_F^2 = 1$ and $\sigma_G^2 = 1$.

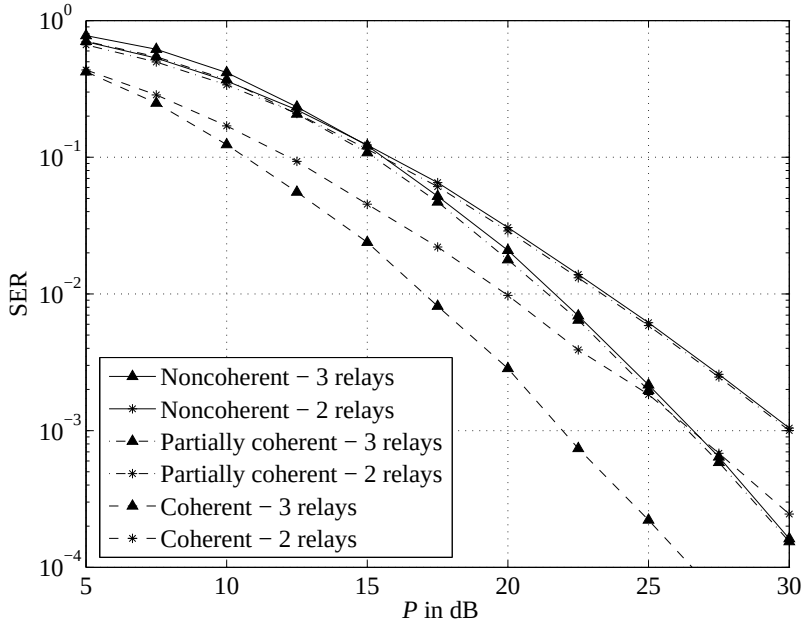

Figure 2.3 Symbol error performance of DUSTM with $\sigma_F^2 = 10$ and $\sigma_G^2 = 1$.

the source-to-relay channels. Similar to the first simulation, the performances of DUSTM in Figures 2.3 and 2.4 for the two types of relay networks eventually merge as P becomes large enough. Therefore, it can be concluded that at high SNR, the

knowledge of the relay-to-destination channels have no apparent effect on the system performance, no matter how good or poor these channels are compared relatively to the source-to-relay channels. Note that since either σ_F^2 or σ_G^2 is increased to 10 in the last two simulations, it requires less total transmit power, P to realize the full diversity order of the network.

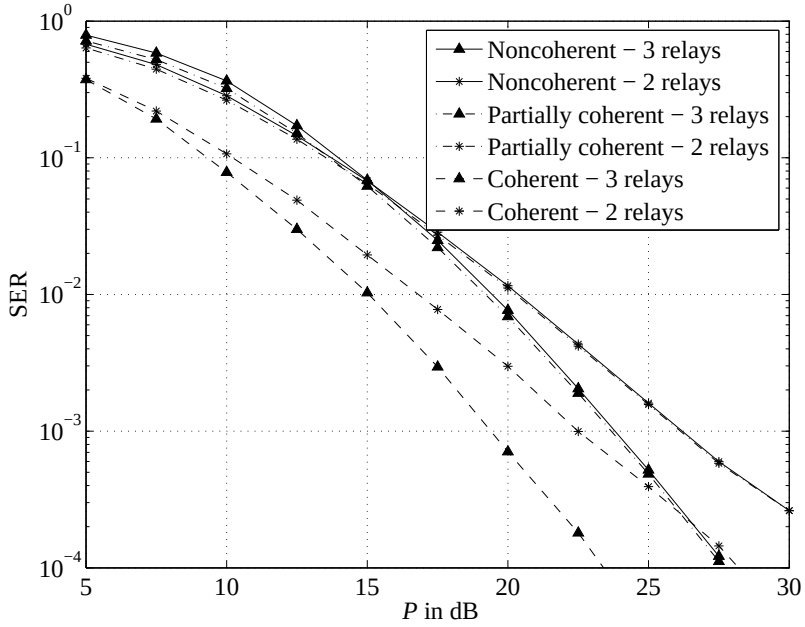

Figure 2.4 Symbol error performance of DUSTM with $\sigma_F^2 = 1$ and $\sigma_G^2 = 10$.

Also presented in Figures 2.2, 2.3, and 2.4 is the performance of the coherent decoder (where the channel state information of both source-to-relay and relay-to-destination links are available at the destination). As can be seen from the figures, with the same network configuration, three types of decoders achieve the same diversity order. It is noted that the coherent decoder outperforms the other two decoders by approximately 3 to 4 dB. This observation is similar to that noted for the performance of USTM in point-to-point communications where the unitary constellations perform about 2 to 4 dB better when the channel is known, as compared to the case of unknown channel [27].

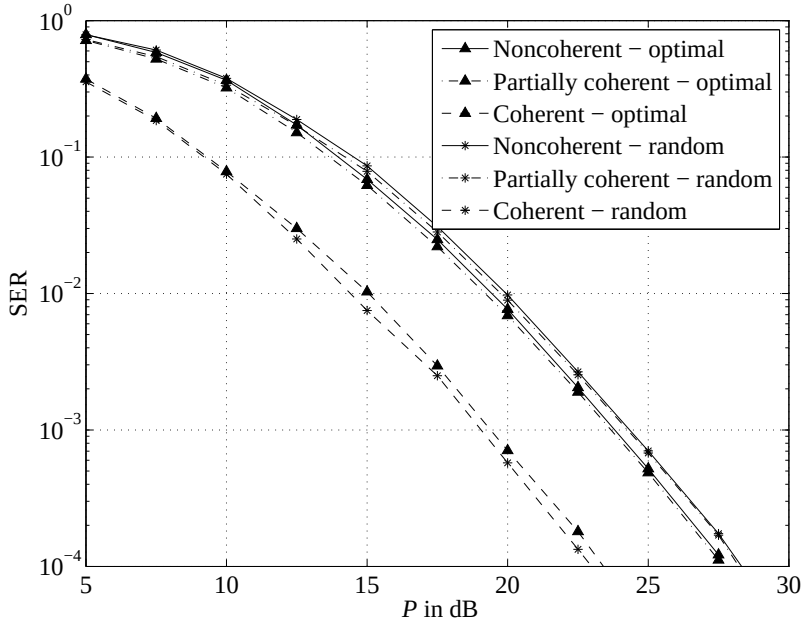

Figure 2.5 Symbol error performances of the proposed DUSTM and a random code with $\sigma_F^2 = 1$ and $\sigma_G^2 = 10$.

2.5.2 Comparison Between DUSTM and a Random Code

This section compares the performance achieved by the proposed DUSTM with that of a random code, where the third network configuration ($\sigma_F^2 = 1$ and $\sigma_G^2 = 10$) is used as an illustrative example. Instead of using the optimally found *frequencies* $u_1, \ldots, u_T$ for 3 transmit antennas [27] as in the proposed DUSTM, a random code is formed by randomly generating the *frequencies* with a uniform distribution between 0 and L. The obtained frequencies used in all the simulations are $\{114, 239, 119, 107, 217, 134, 52, 172\}$. Depending on the information bits, the source signals for the random code are formed in a similar process discussed in Section 2.4. The unitary relay matrices $\boldsymbol{A}_i$ are kept the same as in the proposed DUSTM. As can be seen from Figure 2.5, under the partially coherent and noncoherent decoders, the proposed DUSTM outperforms the random code by about 0.5 dB. This performance advantage is due to the use of the optimal *frequencies* designed for the noncoherent USTM. Note, however, that under the coherent decoding rule, the proposed DUSTM is performed by the random code. This is not unexpected since the optimality of the

frequencies found in [27] and applied in this work only holds for partially coherent or noncoherent networks.

Though desirable, it is noted that the optimal *frequencies* in USTM can only be found by an exhaustive search [27], which is not practically attractive when the number of relays is large and when the transmission interval T in DUSTM is varied. On the other hand, the relatively good performance of random codes makes them a viable option in designing DUSTM. More specifically, the source signal is simply formed from randomly chosen *frequencies*, while the relay matrices $\boldsymbol{A}_i$ are formed by selecting columns of DFT matrices of size $T \times T$.

2.6 Summary

This chapter has considered the design of DSTC in wireless multi-relay networks with different amounts of CSI at the destination. In the first part, the chapter reviewed the DSTC design criteria in coherent relay networks. It was shown that the space-time code that achieves the full diversity and optimum coding gain in MIMO networks also achieves the maximum performance in DSTC. The chapter then focused on the design and the performance of DSTC over the partially coherent and the noncoherent networks. Proposed was a new distributed code, which relied on the Fourier-based USTM, for the two types of relay networks. In addition, the new code allowed noncoherent reception at the destination. The ML receiver for the DUSTM over partially coherent relay networks and the GLRT receiver for the noncoherent relay networks were then developed. The performances of the DUSTM over the two types of relay networks were thoroughly studied. Interestingly, the analysis showed a very small impact by the knowledge of $R \to D$ channels on the performances of the distributed code. Full diversity order equal to the number of relays is achievable by the proposed code in both networks if the coherence time is larger than twice the number of relays.

3. Optimal Power Allocation in DSTC

3.1 Introduction

Power control at the transmitter is a key consideration in designing any wireless network. First, power control helps to maintain a good connectivity to the receiver. Due to the random nature of wireless channels, it is important to maintain a minimum signal power reception by adjusting the transmitted power. Second, power control helps to manage the power usage and maximize the lifetime of the transmitting node and even the network. This is due to the fact that most of operating nodes in wireless communications are mobile nodes with limited battery power. Finally, due to the broadcast nature of the wireless channel, transmitted signals usually interfere with each other. Power control helps to manage the interference to adjacent networks, and facilitate efficient spectral reuse [33].

In the wireless relay networks, given a limited power budget, it is necessary to optimize the power consumption at all the transmitting nodes, including the source and relays with the aim of achieving the optimal performance at the destination. While most of the works on DSTC in the literature consider the optimal power allocation (PA) problem with a fixed value of the channel strength of each $S \rightarrow R$, $R \rightarrow D$ link (see [8, 9, 20, 22]) and thus imply a fixed location for all the relays, the optimal power allocation for DSTC [8, 20–22] in an arbitrary relay network has not been fully studied. Reference [34] analyzed such a problem, with the assumption of either full channel state information (CSI) or partial CSI available at the source. The authors [34] also proposed an optimal PA scheme, which was obtained via an iterative

[0]The contributions in this chapter are also presented in [16, 32].

method.

This chapter studies an optimal PA scheme with the assumption that the CSI of all the links is not available at either the source or the relays, but their statistics are known. The aim is to optimize the PA between the source and the relays to maximize the long-term effective average SNR at the destination. It is first shown that with the optimal solution for such PA, some of the relays might not be active.[1] As a result, this scheme may degrade the system performance as it reduces the maximum achievable diversity order. The concept of amount of fading is then introduced for the relay network to leverage the fading statistics of each $S \to R \to D$ link, and used as a constraint to derive a novel optimal PA scheme. This new scheme is presented in a closed-form solution, and proved to be capable of obtaining the maximum diversity order in all coherent, partially coherent, and noncoherent DSTC systems. Simulations confirm the analysis and show a significant performance advantage of the new scheme.

3.2 Problem Formulation

Recall the system model described in Equation (2.4) in Chapter 2. All the fading coefficients $\{f_i\}_{i=1}^R$, $\{g_i\}_{i=1}^R$, the noise vectors $\{\boldsymbol{z}_{R_i}\}_{i=1}^R$, and $\boldsymbol{z}_D$ are independent of each other. Since $\boldsymbol{z}_{R_i}$ is circularly symmetric and $\boldsymbol{A}_i$ acts as a rotation matrix, the rotated noise vector $\boldsymbol{A}_i\boldsymbol{z}_{R_i}^{(*)}$ has the same distribution as that of $\boldsymbol{z}_{R_i}$. Conditioned on g_i, the average noise power is

$$\gamma = \mathop{\mathbb{E}}_{\{\boldsymbol{z}_{R_i}\}_{i=1}^R, \boldsymbol{z}_D} \left[\|\boldsymbol{z}\|^2\right] = \frac{1}{P_0 T}\left(1 + \sum_{i=1}^{R} \varepsilon_i \sigma_{G_i}^2 |g_i|^2\right). \tag{3.1}$$

On the other hand, the *instantaneous* signal power is

$$\rho = \frac{1}{R}\sum_{i=1}^{R} \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2 |f_i|^2 |g_i|^2. \tag{3.2}$$

Accordingly, the *instantaneous* SNR at the destination is ρ/γ. Thus, the *exact* average SNR can be calculated by averaging ρ/γ over random variables $\{g_i\}_{i=1}^R$ and $\{f_i\}_{i=1}^R$

[1]Hereafter, a relay is said to be "active" if it is distributed a non-zero power by the PA scheme. Otherwise, the relay is said to be "inactive".

as follows:

$$\begin{aligned}\eta &= \underset{\{g_i\}_{i=1}^R,\{f_i\}_{i=1}^R}{\mathbb{E}}\left[\frac{\rho}{\gamma}\right] \\ &= \frac{P_0 T}{R}\underset{\{g_i\}_{i=1}^R}{\mathbb{E}}\left[\frac{\sum_{i=1}^R \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2 |g_i|^2}{1+\sum_{i=1}^R \varepsilon_i \sigma_{G_i}^2 |g_i|^2}\right] \\ &= \frac{P_0 T}{R}\int_0^\infty \cdots \int_0^\infty \frac{\sum_{i=1}^R \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2 \lambda_i}{1+\sum_{i=1}^R \varepsilon_i \sigma_{G_i}^2 \lambda_i} \mathrm{e}^{-\lambda_1} \cdots \mathrm{e}^{-\lambda_R} \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_R. \end{aligned} \tag{3.3}$$

Note that the integral follows from the fact that $\lambda_i = |g_i|^2$ is exponentially distributed with mean 1.

Though the exact average SNR can be evaluated numerically by the Gauss-Laguerre integration method [35], a closed-form expression of η in (3.3) is hard to obtain. To provide insight and perform the optimization of the average SNR, the following approximation of η, resulted by taking the first term of the Taylor's series expansion of the expectation in (3.3) [36], shall be considered:

$$\begin{aligned}\eta \approx \frac{\mathbb{E}_{\{g_i\}_{i=1}^R,\{f_i\}_{i=1}^R}[\rho]}{\mathbb{E}_{\{g_i\}_{i=1}^R,\{f_i\}_{i=1}^R}[\gamma]} = \frac{P_0 T}{R}\frac{\sum_{i=1}^R \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2}{1+\sum_{i=1}^R \varepsilon_i \sigma_{G_i}^2} &= \frac{P_0 T}{R}\frac{\sum_{i=1}^R \frac{P_i \sigma_{F_i}^2 \sigma_{G_i}^2}{P_0 \sigma_{F_i}^2 + 1}}{1+\sum_{i=1}^R \frac{P_i \sigma_{G_i}^2}{P_0 \sigma_{F_i}^2 + 1}} \\ &= \frac{P_0 T}{R}\frac{\sum_{i=1}^R \frac{P_i \sigma_{F_i}^2 \sigma_{G_i}^2}{P_0 \sigma_{F_i}^2 + 1}}{1+\sum_{i=1}^R \frac{P_i \sigma_{G_i}^2}{P_0 \sigma_{F_i}^2 + 1}}, \end{aligned} \tag{3.4}$$

Let $P_{\text{total}} = \sum_{i=0}^R P_i$ be the total transmitted power spent in the network. The aim of PA between the source and the relays is to maximize the average SNR at the destination, given that the total power consumption P_{total} is not greater than the power budget P. It is observed that increasing P makes the average SNR at the destination higher. However, it is not clear how to split the total power between the source and the relays in order to maximize η. Such a PA problem is analyzed next where the approximation of η in (3.4) is used as the objective function. Numerical results presented in Section 3.7 confirm the accuracy of the approximate SNR in (3.4) under all the PA schemes proposed and examined in this chapter.

From (3.4), ignore the constant factor T/R, the optimization problem is stated as

$$\begin{aligned} \underset{P_0,\ldots,P_R}{\text{minimize}} \quad & f(P_0,\ldots,P_R) = \frac{1+\sum_{i=1}^{R}\frac{P_i\sigma_{G_i}^2}{P_0\sigma_{F_i}^2+1}}{P_0\sum_{i=1}^{R}\frac{P_i\sigma_{F_i}^2\sigma_{G_i}^2}{P_0\sigma_{F_i}^2+1}} \\ \text{subject to} \quad & \sum_{i=0}^{R} P_i \le P, \quad P_i \ge 0, \quad i=0,\ldots,R. \end{aligned} \tag{3.5}$$

Note that the objective function of the above optimization problem is nonconvex. Thus, it is difficult to perform the optimization directly over the variables $P_0,\ldots,P_R$. However, if P_0 is fixed, the objective function is linear-fractional. Thus, the optimization problem is a linear-fractional program, which is a subclass of convex programming [37] and can be solved effectively. Before presenting the solution to Problem (3.5), Lemma 1 is first presented as it is useful to establish the analytically closed-form expression of the optimal PA scheme.

Lemma 1. *At the optimal solution to the optimization problem* (3.5)*, the inequality* $\sum_{i=0}^{R} P_i \le P$ *is met with equality.*

Proof. If $P_i^\star$ is the optimal solution with $\sum_{i=0}^{R} P_i^\star < P$, increasing $P_0^\star$ up to $P - \sum_{i=1}^{R} P_i^\star$ makes the objective function strictly smaller, which contradicts to the optimality of $P_i^\star$. This is because the objective function is a monotonically decreasing in P_0, which follows from the fact that its numerator, $1+\sum_{i=1}^{R}\frac{P_i\sigma_{G_i}^2}{P_0\sigma_{F_i}^2+1}$, is monotonically decreasing in P_0 while its denominator, $P_0\sum_{i=1}^{R}\frac{P_i\sigma_{F_i}^2\sigma_{G_i}^2}{P_0\sigma_{F_i}^2+1} = \sum_{i=1}^{R}\frac{P_i\sigma_{F_i}^2\sigma_{G_i}^2}{\sigma_{F_i}^2+1/P_0}$, is monotonically increasing in P_0. □

In [8], the optimal PA for the case of $\{\tilde{\sigma}_{F_i}^2\}_{i=1}^{R} = \{\tilde{\sigma}_{G_i}^2\}_{i=1}^{R} = 1$, $\sigma_R^2 = \sigma_D^2 = 1$ was given. The allocation, referred to as "equal PA" in this thesis, assigns half of the total power to the source and equally divide the other half to all the relays. However, the condition of $\tilde{\sigma}_{F_i}^2 = \tilde{\sigma}_{G_i}^2 = 1$ loosely implies that all the relays are in the midway between the source and the destination. More general cases of network topology, as summarized in Table 3.1, are now considered. For convenience, hereafter a relay network is classified as *balanced* if $\tilde{\sigma}_{F_1}^2 = \ldots = \tilde{\sigma}_{F_R}^2$, and $\tilde{\sigma}_{G_1}^2 = \ldots = \tilde{\sigma}_{G_R}^2$,

Table 3.1 Network configurations and optimization problems

Network Configuration	**Optimization Problem**
Balanced network	Equal power sharing at relays, joint optimization between the source and relays.
Unbalanced network	Fix source power P_0, relay power allocated to the best relay(s).
Unbalanced network with the amount of fading constraint	Balance the network by the amount of fading constraint, joint optimization between the source and the relays.

i.e., $\sigma_{F_1}^2 = \ldots = \sigma_{F_R}^2$, and $\sigma_{G_1}^2 = \ldots = \sigma_{G_R}^2$. Otherwise, it is classified as *unbalanced*. Section 3.3 considers a balanced network, where equal power sharing at the relays is assumed and joint power optimization between the source and the relays is performed. For the unbalanced network topology examined in Section 3.4, we first fix the source power P_0, and prove that the remaining power is allocated only to the best relay(s). In Section (3.5), the "amount of fading" concept is introduced for the relay networks to balance $S \rightarrow R \rightarrow D$ links. Taking into account the amount of fading constraint, a closed-form expression for the joint power optimization between the source and the relays is established.

3.3 SNR-Maximized PA in Balanced Networks

For a balanced network, the "average qualities" of $S \rightarrow R \rightarrow D$ links are the same. Thus, the relay power is equally shared between the relays, and the optimization problem reduces to how to distribute the power budget between the source and the relays. Substitute $P_1 = \ldots = P_R = (P - P_0)/R$ into (3.4), the optimal power

allocation problem is equivalent to

$$\begin{aligned} \underset{P_0}{\text{minimize}} \quad & f(P_0) = \frac{(P-P_0)\sigma_G^2 + P_0\sigma_F^2 + 1}{P_0(P-P_0)\sigma_F^2\sigma_G^2}, \\ \text{subject to} \quad & 0 < P_0 < P. \end{aligned} \tag{3.6}$$

In the special case of $\sigma_F^2 = \sigma_G^2$, (3.6) returns to the same problem in [8] and $P_0 = P/2$, $P_1 = \ldots = P_R = \frac{P}{2R}$ constitute the optimal allocation scheme. If $\sigma_F^2 \neq \sigma_G^2$, it is simple to verify that the second-order derivative of $f(P_0)$ is always positive. Thus, $f(P_0)$ is a convex function for $0 < P_0 < P$. It is then straightforward to obtain the optimal value of P_0 from the first-order derivative of $f(P_0)$. The solution is

$$\begin{aligned} P_0 &= \begin{cases} \frac{\sqrt{(P\sigma_F^2+1)(P\sigma_G^2+1)}-(P\sigma_G^2+1)}{\sigma_F^2-\sigma_G^2}, & \text{if } \sigma_F^2 \neq \sigma_G^2 \\ P/2, & \text{if } \sigma_F^2 = \sigma_G^2 \end{cases} \\ P_1 &= \ldots = P_R = (P-P_0)/R, \end{aligned} \tag{3.7}$$

where one can easily verify that the condition $0 < P_0 < P$ is met.

3.4 SNR-Maximized PA in Unbalanced Networks

Considered now is the PA problem in an even more general setting than the cases previously examined. Here $\sigma_{F_i}^2$ and $\sigma_{G_i}^2$, $i = 1, \ldots, R$, can take on any values. This consideration is more practical than that of balanced networks since there are no restrictions on the relay locations as well as the fading profiles of $S \to R$ and $R \to D$ channels.

Since the objective function in (3.5) is nonconvex, it may not be straightforward to solve the optimization problem directly. The approach here is to fix P_0 and optimize the function over $P_1, \ldots, P_R$ first. This is equivalent to fixing the source's transmitted power and optimizing the PA between the relays. Then, the objective function is optimized over P_0 to yield the optimal solution. The optimization problem (3.5) can

be rewritten as

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f_1(\boldsymbol{x}) = \frac{1}{\boldsymbol{p}^{\mathcal{T}}\boldsymbol{x}} + \frac{\boldsymbol{q}^{\mathcal{T}}\boldsymbol{x}}{\boldsymbol{p}^{\mathcal{T}}\boldsymbol{x}} \\ \text{subject to} \quad & \mathbf{1}^{\mathcal{T}}\boldsymbol{x} = P - P_0 \\ & \boldsymbol{x} \succeq 0, \end{aligned} \tag{3.8}$$

where

$$\begin{aligned} \boldsymbol{p} &= P_0 \left[\frac{\sigma_{F_1}^2 \sigma_{G_1}^2}{P_0 \sigma_{F_1}^2 + 1}, \ldots, \frac{\sigma_{F_R}^2 \sigma_{G_R}^2}{P_0 \sigma_{F_R}^2 + 1} \right]^{\mathcal{T}} \\ \boldsymbol{q} &= \left[\frac{\sigma_{G_1}^2}{P_0 \sigma_{F_1}^2 + 1}, \ldots, \frac{\sigma_{G_R}^2}{P_0 \sigma_{F_R}^2 + 1} \right]^{\mathcal{T}} \\ \boldsymbol{x} &= [P_1, \ldots, P_R]^{\mathcal{T}} \\ \mathbf{1} &= [1, \ldots, 1]^{\mathcal{T}} \in \mathbb{R}^R. \end{aligned}$$

Note that the equality constraint follows from Lemma 1.

With the variable change $\boldsymbol{z} = \frac{\boldsymbol{x}}{\boldsymbol{p}^{\mathcal{T}}\boldsymbol{x}}$, which implies that $\boldsymbol{p}^{\mathcal{T}}\boldsymbol{z} = 1$ and $\frac{1}{\boldsymbol{p}^{\mathcal{T}}\boldsymbol{x}} = \frac{\mathbf{1}^{\mathcal{T}}\boldsymbol{z}}{\mathbf{1}^{\mathcal{T}}\boldsymbol{x}} = \frac{\mathbf{1}^{\mathcal{T}}\boldsymbol{z}}{P - P_0}$ whenever $\mathbf{1}^{\mathcal{T}}\boldsymbol{x} = P - P_0$, the optimization problem in (3.8) becomes the following linear programming (LP):

$$\begin{aligned} \underset{\boldsymbol{z}}{\text{minimize}} \quad & \left(\frac{1}{P - P_0}\mathbf{1} + \boldsymbol{q} \right)^{\mathcal{T}} \boldsymbol{z} = \bar{\boldsymbol{q}}^{\mathcal{T}}\boldsymbol{z} \\ \text{subject to} \quad & \boldsymbol{z} \succeq 0 \\ & \boldsymbol{p}^{\mathcal{T}}\boldsymbol{z} = 1, \end{aligned} \tag{3.9}$$

where $\bar{\boldsymbol{q}} = \frac{1}{P - P_0}\mathbf{1} + \boldsymbol{q}$.

An obvious optimal solution of the above problem is

$$\boldsymbol{z} = [0, \ldots, 0, 1/\boldsymbol{p}_i, 0, \ldots, 0]^{\mathcal{T}}, \tag{3.10}$$

where the index i is such that $\frac{\bar{\boldsymbol{q}}_i}{\boldsymbol{p}_i}$ is minimum. Further, the corresponding optimal value of the objective function in (3.9) is $\frac{\bar{\boldsymbol{q}}_i}{\boldsymbol{p}_i}$. If there are more than one $\frac{\bar{\boldsymbol{q}}_i}{\boldsymbol{p}_i}$ attaining the minimum value, any distribution between the corresponding relays will result in the same optimal value of the objective function.

It can be further verified that when the relay network is balanced, $\frac{\bar{\boldsymbol{q}}_i}{\boldsymbol{p}_i}$ is equal for all $i = 1, \ldots, R$ and any distribution of the total relay power results in the same optimal value of the objective function, i.e., the same maximum value of the effective SNR. Nevertheless, the scheme that equally shares the power between the relays is most preferred, since it makes use of all the relays and minimizes the amount of fading experienced over the network as introduced and discussed in more detail in the next section. This equal-PA scheme is also consistent with the previous analysis of the balanced networks.

From the above analysis, it is clear that with a fixed value of P_0, not all the relays might be active in an unbalanced network. It is of interest to find which relay(s) are active and how the total power should be distributed between the source and the active relay(s). As shown above, for a fixed source power P_0, the active relay is identified by the index, denoted by $i_{\text{act}}(P_0)$, such that $\frac{\bar{\boldsymbol{q}}_{i_{\text{act}}(P_0)}}{\boldsymbol{p}_{i_{\text{act}}(P_0)}}$ is minimum. That is,

$$i_{\text{act}}(P_0) = \arg \min_{i=1,\ldots,R} \frac{\bar{\boldsymbol{q}}_i}{\boldsymbol{p}_i} = \arg \min_{i=1,\ldots,R} \frac{(P-P_0)\sigma_{G_i}^2 + P_0\sigma_{F_i}^2 + 1}{P_0(P-P_0)\sigma_{F_i}^2\sigma_{G_i}^2}. \tag{3.11}$$

Note that, if there are multiple minimizers in (3.11), those minimizers assign the corresponding multiple active relays.

Regardless of the number of active relays, the optimization of the source power can be solved by expressing the optimization problem in (3.5) as follows:

$$\min_{0<P_0<P} \min_{i=1,\ldots,R} \frac{(P-P_0)\sigma_{G_i}^2 + P_0\sigma_{F_i}^2 + 1}{P_0(P-P_0)\sigma_{F_i}^2\sigma_{G_i}^2} = \min_{i=1,\ldots,R} \min_{0<P_0<P} \frac{(P-P_0)\sigma_{G_i}^2 + P_0\sigma_{F_i}^2 + 1}{P_0(P-P_0)\sigma_{F_i}^2\sigma_{G_i}^2}. \tag{3.12}$$

For each $i = 1, \ldots, R$, let

$$\mu_i = \min_{0<P_0<P} \frac{(P-P_0)\sigma_{G_i}^2 + P_0\sigma_{F_i}^2 + 1}{P_0(P-P_0)\sigma_{F_i}^2\sigma_{G_i}^2}. \tag{3.13}$$

Observe that (3.13) is in the form of (3.6), where the common variances σ_F^2 and σ_G^2 for a balanced network are replaced by $\sigma_{F_i}^2$ and $\sigma_{G_i}^2$ for the specific $S \to R \to D$ link via the ith relay. Thus, for each i, the optimal source power, denoted by $P_{0,i}$, is given

by

$$P_{0,i} = \begin{cases} \frac{\sqrt{(P\sigma_{F_i}^2+1)(P\sigma_{G_i}^2+1)}-(P\sigma_{G_i}^2+1)}{\sigma_{F_i}^2-\sigma_{G_i}^2}, & \text{if } \sigma_{F_i}^2 \neq \sigma_{G_i}^2 \\ P/2, & \text{if } \sigma_{F_i}^2 = \sigma_{G_i}^2 \end{cases}. \tag{3.14}$$

The corresponding value of μ_i is

$$\mu_i = \begin{cases} \frac{\left(\sigma_{F_i}^2-\sigma_{G_i}^2\right)^2}{\left(\sqrt{P\sigma_{F_i}^2+1}-\sqrt{P\sigma_{G_i}^2+1}\right)^2\sigma_{F_i}^2\sigma_{G_i}^2}, & \text{if } \sigma_{F_i}^2 \neq \sigma_{G_i}^2 \\ \frac{4(P\sigma_{F_i}^2+1)}{P^2\sigma_{F_i}^2\sigma_{G_i}^2}, & \text{if } \sigma_{F_i}^2 = \sigma_{G_i}^2. \end{cases} \tag{3.15}$$

The following proposition summarizes the optimal PA for an unbalanced network.

Proposition 1. *The optimal PA scheme to maximize the effective SNR in an unbalanced relay network would spend all the relay power to the relays with the best overall $S \rightarrow R \rightarrow D$ channels. The indices of such active relays are found by $i_{\text{act}} = arg\min_{i=1,\ldots,R} \mu_i$, where μ_i is defined in* (3.15)*. Furthermore, given the index of an active relay, $i = i_{\text{act}}$, the optimal source power is found by* (3.14)*. Any distribution of the remaining power, i.e., the relay power, between the active relay(s) gives the same optimal SNR at the destination.*

It is clear from Proposition 1 that in unbalanced networks, while the optimal PA scheme maximizes the long term average SNR at the destination, it may distribute no power to some of the relays. This may affect the error performance of the distributed code as its diversity order is reduced (since the number of independent channels in the network is reduced). Therefore, some additional constraints should be introduced into the optimization problem to ensure that all the relays are active, i.e., the relay power is allocated to all the relays, hence providing the full diversity order. In the next section, the amount of fading concept is introduced for the relay networks and used as an additional constraint for the PA problem. The amount of fading gives a convenient measure of the fading severity experienced by the network and it should be kept as small as possible. It shall be shown that the amount of fading in a relay network can be reduced inversely to the number of relays. Then the condition on the

transmitted power at each relay is derived to obtain a lower bound on the amount of fading.

3.5 The Optimal PA Scheme under the Amount of Fading Constraint

The amount of fading is a common measure of fading severity in a fading channel model. In single-input single-output communication systems, the amount of fading is defined based on the instantaneous fading amplitude $\alpha = |h|$ of the complex fading coefficient h as [38]

$$\kappa = \frac{\mathrm{var}[\alpha^2]}{\left(\mathbb{E}[\alpha^2]\right)^2}. \tag{3.16}$$

In multiple-input multiple-output communication systems, α^2 is the summation of the instantaneous squared magnitudes of all the channel coefficients between each pair of transmit and receive antennas [39]. In the distributed space-time relay network considered in this thesis, by taking into account the fading coefficients in each $S \rightarrow R \rightarrow D$ link, α^2 can be loosely defined as

$$\alpha^2 = \sum_{i=1}^{R} \frac{P_i \sigma_{F_i}^2 \sigma_{G_i}^2 |f_i|^2 |g_i|^2}{P_0 \sigma_{F_i}^2 + 1}. \tag{3.17}$$

By disregarding common factors, the amount of fading of the considered relay network can be defined similarly to (3.16).

Since f_i and g_i are independent, for $i = 1, \ldots, R$, the mean value of α^2 is given as

$$\mathop{\mathbb{E}}_{\{g_i\},\{f_i\}}[\alpha^2] = \sum_{i=1}^{R} \frac{P_i \sigma_{F_i}^2 \sigma_{G_i}^2}{P_0 \sigma_{F_i}^2 + 1}. \tag{3.18}$$

The variance of α^2 is

$$\begin{aligned} \mathrm{var}[\alpha^2] &= \sum_{i=1}^{R} \frac{P_i^2 \sigma_{F_i}^4 \sigma_{G_i}^4}{(P_0 \sigma_{F_i}^2 + 1)^2} \mathrm{var}[|f_i|^2 |g_i|^2] \\ &= \sum_{i=1}^{R} \frac{3 P_i^2 \sigma_{F_i}^4 \sigma_{G_i}^4}{(P_0 \sigma_{F_i}^2 + 1)^2} \end{aligned} \tag{3.19}$$

which follows from the fact that $|f_i|^2$ and $|g_i|^2$ are exponentially distributed with mean 1, and $\mathrm{var}[|f_i|^2 |g_i|^2] = \mathbb{E}[|f_i|^4]\mathbb{E}[|g_i|^4] - (\mathbb{E}[|f_i|^2]\mathbb{E}[|g_i|^2])^2 = 2 \times 2 - 1 = 3$.

Next, observe that Hölder inequality (see [37], p. 48) can be used to obtain the following inequality

$$\sum_{i=1}^{R} |x_i|^2 \geq \frac{\left(\sum_{i=1}^{R} |x_i|\right)^2}{R} \tag{3.20}$$

where the equality is achieved when $|x_1| = \ldots = |x_R|$. Then applying (3.20) to (3.19), one obtains the following lower bound on the variance of α^2:

$$\text{var}[\alpha^2] \geq \frac{3}{R} \left(\sum_{i=1}^{R} \frac{P_i \sigma_{F_i}^2 \sigma_{G_i}^2}{P_0 \sigma_{F_i}^2 + 1} \right)^2. \tag{3.21}$$

Therefore, the amount of fading is lower bounded as

$$\kappa = \frac{\text{var}[\alpha^2]}{\left(\mathbb{E}[\alpha^2]\right)^2} \geq \frac{3}{R}, \tag{3.22}$$

and the lower bound is achieved with

$$\frac{P_1 \sigma_{F_1}^2 \sigma_{G_1}^2}{P_0 \sigma_{F_1}^2 + 1} = \ldots = \frac{P_R \sigma_{F_R}^2 \sigma_{G_R}^2}{P_0 \sigma_{F_R}^2 + 1}. \tag{3.23}$$

It is noted that the factor 3 in the amount of fading expression for the relay networks is due to the "cascaded" fading characteristic of the amplify-and-forward protocol employed. Furthermore, an amount of fading equal to 1 here does not mean that the system has the diversity order of 1 as in a typical point-to-point communication system over a Rayleigh fading channel.

In order to minimize the amount of fading of the distributed space-time coding in wireless relay networks, the PA scheme between the source and the relays needs to satisfy (3.23). Apparently, the condition in (3.23) makes the amount of fading in each $S \rightarrow R \rightarrow D$ link to be the same, an intuitively satisfying result. Therefore the optimization problem to maximize the average SNR is the same as in (3.5), but with the additional constraint for minimum amount of fading as in (3.23). Intuitively, the optimal power scheme would allocate more power to the relay with a weaker link, and less power for the relay with a stronger link. This is reasonable since only channel statistics are known at the relays and one would like to use all the relay channels reliably to achieve the full diversity order.

Define $a = \sum_{i=1}^{R} \frac{1}{\sigma_{F_i}^2}$, $b = \sum_{i=1}^{R} \frac{1}{\sigma_{G_i}^2}$, and $c = \sum_{i=1}^{R} \frac{1}{\sigma_{F_i}^2 \sigma_{G_i}^2}$. From the minimum amount of fading constraint, one has

$$\frac{P_1}{\frac{P_0\sigma_{F_1}^2+1}{\sigma_{F_1}^2\sigma_{G_1}^2}} = \ldots = \frac{P_R}{\frac{P_0\sigma_{F_R}^2+1}{\sigma_{F_R}^2\sigma_{G_R}^2}} = \frac{\sum_{i=1}^{R} P_i}{\sum_{i=1}^{R} \frac{P_0\sigma_{F_i}^2+1}{\sigma_{F_i}^2\sigma_{G_i}^2}} = \frac{P - P_0}{P_0 b + c}, \tag{3.24}$$

where the condition on optimality in Lemma 1 is used. It follows that

$$P_i = \frac{P - P_0}{P_0 b + c} \cdot \frac{P_0\sigma_{F_i}^2 + 1}{\sigma_{F_i}^2\sigma_{G_i}^2}, \quad i = 1, \ldots, R. \tag{3.25}$$

Substitute $P_1, \ldots, P_R$ from (3.25) to the objective function in (3.5), the new objective function can be written as

$$\begin{aligned} f(P_0) &= \frac{1 + \sum_{i=1}^{R} \frac{P-P_0}{(P_0 b+c)\sigma_{F_i}^2}}{P_0 R \frac{P-P_0}{P_0 b+c}} \\ &= \frac{1}{R}\left(\frac{P_0 b + c}{P_0(P - P_0)} + \frac{a}{P_0}\right). \end{aligned} \tag{3.26}$$

Ignore the constant factor $1/R$, the optimization problem can be simplified to

$$\begin{aligned} &\underset{P_0}{\text{minimize}} \;\; f(P_0) = \frac{P_0 b + c}{P_0(P - P_0)} + \frac{a}{P_0}, \\ &\text{subject to } 0 < P_0 < P. \end{aligned} \tag{3.27}$$

Since the second derivative of the objective function is always positive in the domain of P_0, the objective function is convex. This problem can be solved easily and the solution is given in in the following proposition.

Proposition 2. *The optimal PA scheme for a relay network that maximizes the effective SNR at the destination under the minimum amount of fading constraint is*

$$\begin{aligned} P_0 &= \begin{cases} \frac{\sqrt{(Pa+c)(Pb+c)} - (Pa+c)}{b-a}, & \text{if } b \neq a \\ P/2, & \text{if } b = a \end{cases} \\ P_i &= \frac{P - P_0}{P_0 b + c} \cdot \frac{P_0\sigma_{F_i}^2 + 1}{\sigma_{F_i}^2\sigma_{G_i}^2}, \quad i = 1, \ldots, R, \end{aligned} \tag{3.28}$$

where

$$a = \sum_{i=1}^{R} \frac{1}{\sigma_{F_i}^2}, \; b = \sum_{i=1}^{R} \frac{1}{\sigma_{G_i}^2}, \; c = \sum_{i=1}^{R} \frac{1}{\sigma_{F_i}^2\sigma_{G_i}^2}.$$

Proposition 2 provides simple closed-form expressions to optimally allocate the total transmitted power in the network to the source and relays. The scheme tries to maximize the effective average SNR at the destination while maintaining a balance between $S \to R \to D$ links in order to minimize the amount of fading. In Section 3.6, it is shown that such a scheme is also optimal in minimizing the SER performance of the coherent DSTC and the proposed DUSTM in Chapter 2. It is also noted that as the minimum amount of fading constraint is automatically met with the equal PA at the relays in balanced networks, the closed-form optimal power allocation scheme for such networks given in (3.7) is a special case of the expressions in Proposition 2.

3.6 Diversity Analysis of The Proposed Power Allocation Scheme

3.6.1 Coherent DSTC

The following lemmas are established to analyze the SER performance of the optimal PA scheme with minimum amount of fading constraint given in Proposition 2.

Lemma 2. *Asymptotically, the upper-bound of the coherent DSTC's PEP is minimized with the optimal PA scheme under the minimum amount of fading constraint. Maximum diversity order is also obtained by the optimal PA scheme.*

Proof. Suppose that $\boldsymbol{s}_k$ is the transmitted source signal. Following similar derivation steps of Theorem 1 in [8], the pairwise error probability (PEP) of mistaking $\boldsymbol{s}_k$ by $\boldsymbol{s}_l$ is upper-bounded as

$$\mathbb{P}_{k,l} \leq \frac{1}{2} \mathop{\mathbb{E}}_{\{g_i\}} \left[\det{}^{-1} \left(\boldsymbol{I}_R + \frac{\boldsymbol{\Delta}^{\mathcal{H}} \boldsymbol{\Delta} \boldsymbol{G}}{4\gamma} \right) \right], \tag{3.29}$$

where $\boldsymbol{\Delta} = \boldsymbol{X}_k - \boldsymbol{X}_l$, $\boldsymbol{G} = \mathrm{diag}\left(\beta_1 |g_1|^2, \ldots, \beta_R |g_R|^2\right)$ with $\beta_i = \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2$.

Since taking the expectation in (3.29) over $\{g_i\}$ is difficult, the approach as in [8] shall be followed to approximate the random variable γ by its mean value, i.e., $\gamma \approx \bar{\gamma} = (1/P_0 T)(1 + \sum_{i=1}^{R} \varepsilon_i \sigma_{G_i}^2)$. Note that by the law of large numbers, this

approximation holds almost surely as $R \to \infty$. Since the coherent code is assumed to be full rank, $\boldsymbol{\Delta}$ is also of full rank. Let σ^2 be the minimum singular value of $\boldsymbol{\Delta}^{\mathcal{H}}\boldsymbol{\Delta}$. Then the PEP $\mathbb{P}_{kl}$ is (approximately) upper-bounded as

$$\begin{aligned}
\mathbb{P}_{kl} &\lesssim \frac{1}{2}\mathop{\mathbb{E}}_{\{g_i\}}\left[\det{}^{-1}\left(\boldsymbol{I}_R + \frac{\sigma^2 P_0 T \boldsymbol{G}}{4\left(1+\sum_{i=1}^{R}\varepsilon_i\sigma_{G_i}^2\right)}\right)\right] \\
&= \frac{1}{2}\prod_{i=1}^{R}\mathop{\mathbb{E}}_{\{g_i\}}\left[\left(1+\frac{\beta_i\sigma^2 P_0 T|g_i|^2}{4\left(1+\sum_{i=1}^{R}\varepsilon_i\sigma_{G_i}^2\right)}\right)^{-1}\right] \\
&= \frac{1}{2}\prod_{i=1}^{R}\int_0^{\infty}\frac{\mathrm{e}^{-x}}{1+a_i x_i}\mathrm{d}x \\
&= \frac{1}{2}\prod_{i=1}^{R} a_i^{-1}\mathrm{e}^{1/a_i}\left[-\mathrm{Ei}\left(-a_i^{-1}\right)\right], \qquad (3.30)
\end{aligned}$$

where $a_i = \frac{\sigma^2\varepsilon_i P_0 T\sigma_{F_i}^2\sigma_{G_i}^2}{4\left(1+\sum_{j=1}^{R}\varepsilon_j\sigma_{G_j}^2\right)}$ and $\mathrm{Ei}(\chi) = \int_{-\infty}^{\chi}\frac{\mathrm{e}^t}{t}\mathrm{d}t$ is the exponential integral function [40]. When P_0T becomes large, a_i also becomes large, $\mathrm{e}^{1/a_i} = 1 + O(1/a_i) \approx 1$, and

$$-\mathrm{Ei}\left(-a_i^{-1}\right) = 1 + \log a_i \approx \log a_i.$$

Thus, the approximate upper-bound of the PEP becomes

$$\mathbb{P}_{kl} \lesssim \frac{1}{2}\prod_{i=1}^{R}\left(\frac{\log a_i}{a_i}\right)^R. \qquad (3.31)$$

It can be shown that $\log(\log(t)/t)$ is a convex function[2] for large t [37]. This means that $\log(t)/t$ is a log-convex function for large t. From the definition of a log-convex function [37], one has the following inequality:

$$\prod_{i=1}^{R}\frac{\log a_i}{a_i} \geq \left[\frac{R\log\left(\sum_{i=1}^{R}a_i/R\right)}{\sum_{i=1}^{R}a_i}\right]^R = \left[\frac{\log(\eta\sigma^2/4)}{\eta\sigma^2/4}\right]^R, \qquad (3.32)$$

as $\sum_{i=1}^{R}a_i = \eta\sigma^2/4$. The equality holds when $a_1 = \ldots = a_R$, i.e., the minimum amount of fading constraint is met. It is clear that the upper-bound of the coherent DSTC' PEP is minimized when the PA scheme meets the minimum amount of fading constraint. It is noted that $(\log\eta)/\eta$ is a decreasing function for large η. Maximizing the SNR, i.e., η, under the minimum amount of fading constraint will further minimize

[2]It can be easily verified that $\log(\log(t)/t)$ is convex when $t \geq \mathrm{e}^{\frac{1+\sqrt{5}}{2}}$ by its second-order derivative.

the upper-bound of PEP $\mathbb{P}_{kl}$. The expression in (3.32) also suggests that both the upper-bound on PEP and the PEP are proportional to $(\log\eta/\eta)^R$. Since $\log\eta = \eta^{\frac{\log\log\eta}{\log\eta}}$, one has $(\log\eta/\eta)^R = \eta^{-R\left(1-\frac{\log\log\eta}{\log\eta}\right)}$. It means that the diversity order is $R(1-\frac{\log\log\eta}{\log\eta})$, which is also the maximum achievable diversity order of DSTC [8]. □

3.6.2 Partially Coherent and Noncoherent DSTC

Lemma 3. *Asymptotically, the PEP of the partially coherent DUSTM is minimized with the optimal PA scheme under the minimum amount of fading constraint. Maximum diversity order R is also obtained by the scheme.*

Proof. The exact conditional PEP is given in [22] as

$$\mathbb{P}_{k,l|\{g_i\}} = \frac{\det\left(\boldsymbol{I}_R + \boldsymbol{G}/\gamma\right)}{2\,\det\left(\boldsymbol{I}_{2R} + \boldsymbol{\Psi}\boldsymbol{Z}_{kl}^{\mathcal{H}}\boldsymbol{Z}_{kl}/(4\gamma)\right)}, \tag{3.33}$$

where $\boldsymbol{\Psi} = \mathrm{diag}(\boldsymbol{G},\boldsymbol{G})$ and $\boldsymbol{Z}_{kl} = [\boldsymbol{X}_k,\ \boldsymbol{X}_l]$. In order to find the unconditional PEP $\mathbb{P}_{k,l}$, one needs to average (3.33) over the distribution of $\{g_i\}$. Since taking the expectation over $\{g_i\}$ is difficult, the same approach as in the previous section is taken to approximate the random variable γ by its mean value, i.e., $\gamma \approx \bar{\gamma} = (1/P_0T)\left(1+\sum_{i=1}^{R}\varepsilon_i\sigma_{G_i}^2\right)$. Since $\boldsymbol{Z}_{kl}$ is a full-rank matrix with the Fourier-based USTM design [27], the minimum singular value of $\boldsymbol{Z}_{kl}^{\mathcal{H}}\boldsymbol{Z}_{kl}$, denoted as ν^2, is non-zero. Then the asymptotic PEP $P_{k,l}$ is (approximately) upper-bounded as

$$\begin{aligned}
\mathbb{P}_{k,l} &\lesssim \mathop{\mathbb{E}}_{\{g_i\}} \frac{\det\left(\boldsymbol{I}_R + \boldsymbol{G}/\bar{\gamma}\right)}{2\,\det\left(\boldsymbol{I}_{2R} + \boldsymbol{\Psi}\nu^2/(4\bar{\gamma})\right)} \\
&= \frac{1}{2}\prod_{i=1}^{R}\mathop{\mathbb{E}}_{\{g_i\}}\left\{\frac{1+\beta_i|g_i|^2/\bar{\gamma}}{\left[1+\beta_i\nu^2|g_i|^2/(4\bar{\gamma})\right]^2}\right\} \\
&= \frac{1}{2}\prod_{i=1}^{R}\int_0^\infty \frac{(1+b_ix)\mathrm{e}^{-x}}{\left(1+\nu^2b_ix/4\right)^2}\mathrm{d}x,
\end{aligned} \tag{3.34}$$

where $b_i = \beta_i\sigma_{G_i}^2/\bar{\gamma}$. Let $1+\nu^2b_ix/4 = -\nu^2b_it/4$. Then (3.34) becomes

$$\mathbb{P}_{k,l} \lesssim \frac{1}{2}\left(\frac{16}{\nu^4}\right)^R\prod_{i=1}^{R}\frac{\mathrm{e}^{\frac{4}{b_i\nu^2}}}{b_i^2}\int_{-\infty}^{-\frac{4}{b_i\nu^2}}\frac{-b_it+1-4/\nu^2}{t^2}\mathrm{e}^t\mathrm{d}t. \tag{3.35}$$

When P_0T becomes large, b_i also becomes large and $\mathrm{e}^{4/(b_i\nu^2)} = 1 + O(4/(b_i\nu^2)) \approx 1$. Furthermore, one has the following approximations when b_i is large:

$$\int_{-\infty}^{-\frac{4}{b_i\nu^2}} \frac{-\mathrm{e}^t}{t}\mathrm{d}t = -\mathrm{Ei}\left(-\frac{4}{b_i\nu^2}\right) = 1 + \log\left(\frac{b_i\nu^2}{4}\right) \approx \log\left(\frac{b_i\nu^2}{4}\right),$$

and

$$\int_{-\infty}^{-\frac{4}{b_i\nu^2}} \frac{\mathrm{e}^t}{t^2}\mathrm{d}t = \frac{1}{2}\mathrm{Ei}\left(-\frac{4}{b_i\nu^2}\right) + \frac{b_i\nu^2\mathrm{e}^{-\frac{4}{b_i\nu^2}}}{4} \approx -\frac{1}{2}\log\left(\frac{b_i\nu^2}{4}\right) + \frac{b_i\nu^2}{4},$$

Clearly, when b_i becomes large, the dominant term of $\frac{\mathrm{e}^{\frac{4}{b_i\nu^2}}}{b_i^2}\int_{-\infty}^{-\frac{4}{b_i\nu^2}} \frac{-b_it+1-4/\nu^2}{t^2}\mathrm{e}^t\mathrm{d}t$ is $\frac{\log\left(b_i\nu^2/4\right)}{b_i}$. Thus, the PEP is upper-bounded as

$$\mathbb{P}_{k,l} \lesssim \frac{1}{2}\left(\frac{4}{\nu^2}\right)^R \prod_{i=1}^{R} \frac{\log\left(b_i\nu^2/4\right)}{b_i\nu^2/4}. \tag{3.36}$$

Similar to the argument in Lemma 3, maximizing the SNR under the amount of fading constraint will further minimize the PEP's upper-bound of the partially coherent DUSTM. Note that the PEP's upper-bound is then proportional to $(\log\eta/\eta)^R$. Thus, the maximum diversity order of DUSTM is also obtained by the PA in Proposition 2. □

3.7 Simulation Results

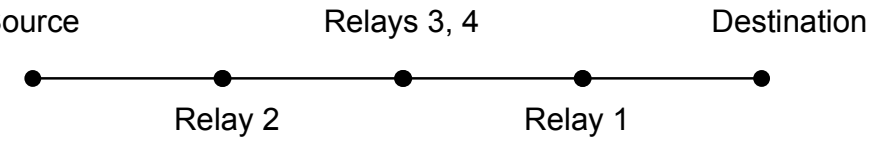

Figure 3.1 The relays' locations relatively to the source and destination.

Considered in the simulation are systems with 2, 3, or 4 relays whose locations are illustrated in Figure 3.1. Assume that the source and destination are located at (0,0) and (1,0), respectively. The first relay is located at (0.75,0), and the second relay is at (0.25,0). The location of the third and fourth relays, if deployed, is at midway between the source and destination, i.e., (0.5,0). The fading variances are assigned proportionally to the distance between the transmit and receive terminals, taking into

account the path loss exponent, which is set at 4. Thus, if $\tilde{\sigma}^2_{F_1}$ is normalized to 1, then $\tilde{\sigma}^2_{F_2} = 3^4, \tilde{\sigma}^2_{F_3} = \tilde{\sigma}^2_{F_4} = (3/2)^4$. Similarly, $\tilde{\sigma}^2_{G_1} = 3^4, \tilde{\sigma}^2_{G_2} = 1, \tilde{\sigma}^2_{G_3} = \tilde{\sigma}^2_{G_4} = (3/2)^4$[3]. Both the noise variances σ^2_R and σ^2_D are set at 1.

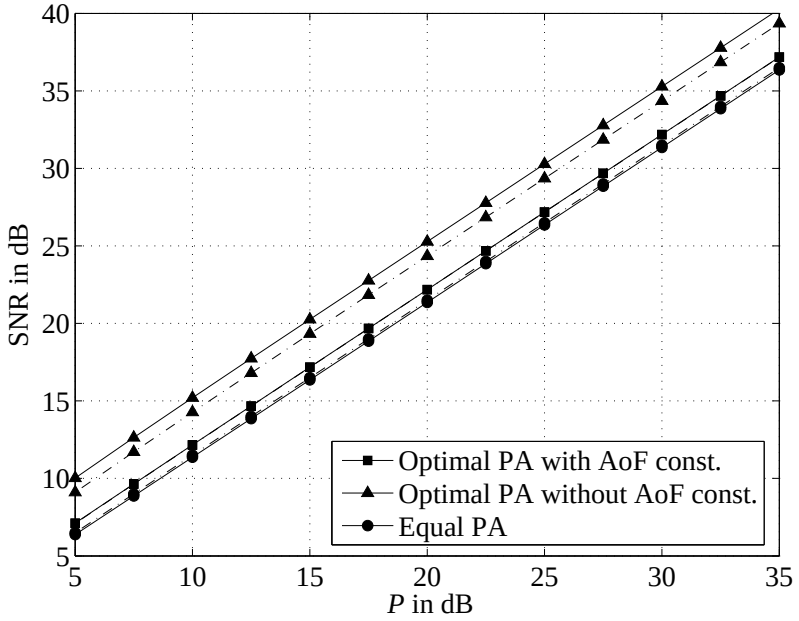

Figure 3.2 Average SNR at the destination with different PA schemes. "Dash-dot" lines are for the exact SNR evaluation, solid lines are for the SNR evaluated by (3.4).

Figure 3.2 illustrates the effective average SNR at the destination with three PA schemes: (i) optimal PA with the minimum amount of fading (AoF) constraint, (ii) optimal PA without the AoF constraint, and (iii) equal PA. The average SNR values are calculated by an exact evaluation and the approximation given in (3.4). A network with 3 relays is considered. As can be seen from the figure, given the total power budget P, the optimal PA without the AoF constraint scheme gives the highest SNR, followed by the optimal PA scheme with the minimum AoF constraint. However,

[3]The network configuration of Figure 3.1 is adopted for simplicity. An arbitrary network topology is not limited to the case that relays are on the same line and the relay locations are defined by two coordinates. For a general topology, all that need to be taken into account are distances from the source to each relay and from each relay to the destination so that the strengths of different channels are modeled appropriately.

there is only one active relay, the third relay, in the optimal PA without the AoF constraint. Therefore, although achieving the highest average SNR, having only one active relay leads to a decrease in the diversity order (as illustrated in Figure 3.3). Furthermore, in comparing the approximate average SNR to the exact average SNR, Figure 3.2 shows that the approximation taken in (3.4) is very accurate, especially in PA schemes where all the relays are active. In particular, the two evaluations of the average SNR for the optimal PA with the AoF constraint scheme are almost identical.

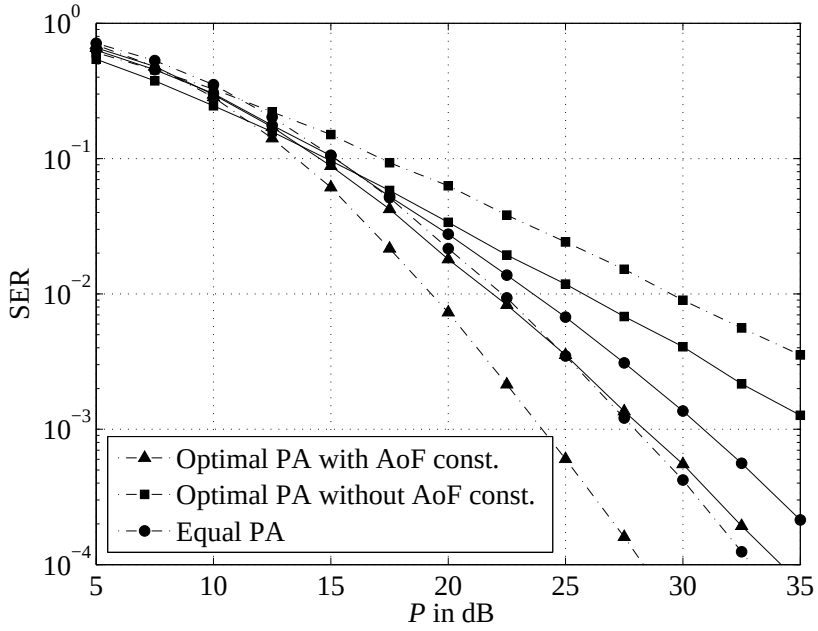

Figure 3.3 Performance of noncoherent DUSTM with different PA schemes. "Dash-dot" lines are for the three-relay system, solid lines are for the two-relay system.

In order to illustrate the performance of different PA schemes in noncoherent relay networks, we use the Fourier-based distributed unitary space-time modulation (DUSTM), stated in Chapter 2. Two-relay (relays 1 and 2) and three-relay (relays 1, 2, and 3) networks are considered. Figure 3.3 shows a significant improvement of the optimal PA scheme with AoF constraint over the equal PA scheme in terms of SER, where the former outperforms the latter by 4 dB and 2.5 dB in the three-relay, and two-relay networks, respectively. It is also observed that the optimal PA scheme

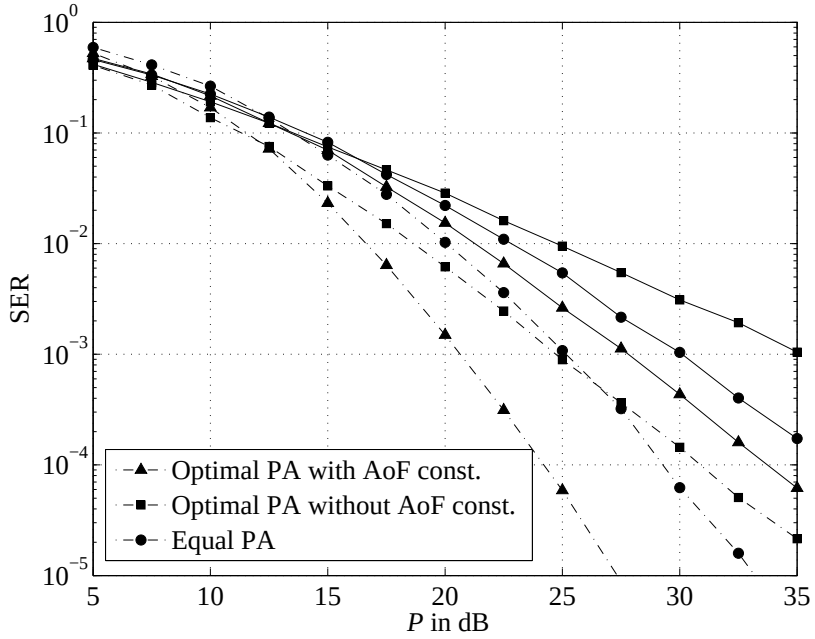

Figure 3.4 Performance of coherent DSTC with different PA schemes. "Dash-dot" lines are for the three-relay system, solid lines are for the two-relay system.

with AoF constraint is able to achieve the maximum diversity order at high SNR, while the equal PA scheme is not. Nevertheless, both schemes performs significantly better than the optimal PA scheme without AoF constraint. This is due to the fact that only one relay is active in this scheme, hence the DUSTM in both relay networks only offers a diversity order of 1. Figure 3.3 also shows the "*scale-free*" characteristic of the proposed DUSTM, as its performance does not collapse with one active relay.

The performances of each PA scheme in coherent relay networks are presented in Figure 3.4. The DSTC implemented are the distributed Alamouti code for two-relay network and the distributed quasi-orthogonal space-time block code for four-relay networks [9]. Similar to the noncoherent networks, the optimal PA with AoF constraint achieves the maximum diversity order of the codes in both two-relay and four-relay networks, while the equal PA scheme does not. It can be seen that the optimal PA with AoF constraint outperforms the equal PA scheme significantly; about 6 dB and 4 dB improvements are observed in the four-relay and two-relay networks,

respectively. The optimal PA scheme without AoF constraint clearly fails to deliver the maximum diversity in both networks. This is due to the fact that there are only one active relay in the two-relay network, and 2 active relays in the four-relay network (relays 3 and 4).

3.8 Summary

This chapter has studied the optimal PA scheme for DTSC, given a total power budget at the source and the relays. It was shown that maximizing the effective average SNR at the destination is not the most important thing to obtain the optimal performance of the DSTC. This is because some of the relays may become inactive in maximizing the average SNR, and hence the diversity order of the DTSC is reduced. To overcome this limitation, the concept of amount of fading was introduced for the relay networks to balance the fading statistics of each $S \rightarrow R \rightarrow D$ link, and to ensure that all the relays are active. A novel and simple PA scheme under the minimum amount of fading constraint was proposed in a closed-form solution. Furthermore, this PA scheme was proved to be capable of obtaining the maximum diversity order in coherent, partially coherent, and noncoherent DSTC networks.

4. Optimal Training and Mismatched Decoding in DSTC

4.1 Introduction

In most existing studies on DSTC, it is commonly assumed that the destination has a perfect knowledge of the CSI, namely the channel coefficient vector $\boldsymbol{h}$. With such an assumption, the coherent maximum-likelihood decoding of the distributed space-time codeword $\boldsymbol{X}$ is possible [8, 9]. However, in practical wireless relay networks, the channel vector $\boldsymbol{h}$ has to be estimated at the destination, typically via training signals. Unlike the DF protocol, where channel estimation can be performed separately at each stage of transmission ($S \rightarrow R$ and $R \rightarrow D$), separate channel estimation with the AF protocol has many drawbacks [15]. First, separate channel estimation requires additional training time. Second, the transmission of the $S \rightarrow R$ channel estimation to the destination may be prone to error. Note that the effective channel in AF transmission is a multiplication of the $S \rightarrow R$ and $R \rightarrow D$ channels. To estimate the effective channel, it is possible to estimate the overall effective $S \rightarrow R \rightarrow D$ channel directly to overcome these drawbacks. Thus, similar to [15], this chapter only considers the direct estimation of the effective channel $\boldsymbol{h}$.

This chapter first reviews the optimal training design in DSTC, proposed in [15]. Similar to the optimal training design in MIMO communications [42], orthogonal training is proved to be optimal in DSTC with regard to both maximum likelihood (ML) and minimum mean-square error (MMSE) estimation criteria. The chapter then investigates the optimal power allocation between the source and the relays

[0]The contributions in this chapter are also presented in [41].

to further minimize the total mean-square error (MSE) of the channel estimation. Finally, the effect of imperfect CSI estimation on system performance is thoroughly studied. It shall be shown that with orthogonal training signals, the DSTC with imperfect channel estimation is able to achieve the same diversity order as that with perfect CSI.

4.2 Optimal Training Design and Channel Estimation

Suppose that the source sends a training sequence $\boldsymbol{u}$ and the codeword $\boldsymbol{s}$ and both are affected by the same overall channel vector $\boldsymbol{h}$. Like the data transmission phase for $\boldsymbol{s}$, the transmission of the training sequence also needs to meet the power constraint of $\boldsymbol{u}^{\mathcal{H}}\boldsymbol{u} \leq 1$. In addition, since at least R independent measurements are needed to estimate the length-R channel vector $\boldsymbol{h}$, the training time should be no less than the number of relays,[1] R. In this work, in order to simplify the processing at each relay, it is assumed that the same relaying matrix is applied to both the training sequence and information codeword. Thus, the training time is also set at T, the same as the data transmission time. The channel $\boldsymbol{h}$ is therefore assumed to remain constant for a block of $T_C > 2T$ channel uses and change independently over the next block.

After possibly rearranging the order of the transmitted symbols, and normalizing the received signal during the training phase by factor $\sqrt{P_0 T \sigma_D^2}$, the destination observes the following:

$$\boldsymbol{y}_T = \boldsymbol{X}_T \boldsymbol{\Lambda} \boldsymbol{h} + \boldsymbol{z}_T, \tag{4.1}$$

where $\boldsymbol{X}_T = [\boldsymbol{A}_1 \boldsymbol{u}^{(*)}, \ldots, \boldsymbol{A}_R \boldsymbol{u}^{(*)}]$ is the training matrix formed at the relays and known at the destination. The noise vector $\boldsymbol{z}_T$, given in a form similar to (2.5), has the same distribution as that of $\boldsymbol{z}$ in the data transmission phase. Rewrite the channel model (4.1), one has the following equivalent input-output model for the

[1]This condition is similar to the requirement of having the training time at least equal to the number of transmit antennas in a MIMO system [42].

training phase:

$$\boldsymbol{y}_T = \boldsymbol{X}_T \boldsymbol{\Lambda} \boldsymbol{h} + \sqrt{\gamma} \bar{\boldsymbol{z}}_T, \tag{4.2}$$

where, conditioned on $\{g_i\}$, $\bar{\boldsymbol{z}}_T$ contains i.i.d. $\mathcal{CN}(0,1)$ random variables. Denote $\gamma_0 = 1/(P_0 T)$, then $\gamma = \gamma_0 \left(1 + \sum_{i=1}^{R} \sigma_{G_i}^2 |g_i|^2\right)$. Note that γ_0 would be the inverse of the signal-to-noise ratio of the $S \to D$ channel if there was no channel fading and the added noise's variance is unity. For convenience, γ_0^{-1} shall be generally referred to as the *channel* signal-to-noise ratio (CSNR). Meanwhile, the channel model (2.4) in Chapter 2 is also rewritten as:

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\Lambda} \boldsymbol{h} + \sqrt{\gamma} \bar{\boldsymbol{z}}, \tag{4.3}$$

where the noise vector $\bar{\boldsymbol{z}}$, conditioned on $\{g_i\}$, contains i.i.d. $\mathcal{CN}(0,1)$ random variables.

For the relay networks considered in this thesis, where f_i and g_i, $i = 1, \ldots, R$ are independent of each other, the optimal training matrix $\boldsymbol{X}_T$ will be shown to be orthogonal for both the ML and MMSE estimation criteria [15]. In particular, $\boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T$ is an identity matrix.

4.2.1 Maximum Likelihood (ML) Estimation

The ML estimation assumes a specific realization of the effective channels $\boldsymbol{h}$. From the channel model of the training phase, the optimal ML estimate of $\boldsymbol{h}$ is obtained as

$$\hat{\boldsymbol{h}}_{ML} = \boldsymbol{\Lambda}^{-1} \boldsymbol{X}_T^{\dagger} \boldsymbol{y}_T = \boldsymbol{h} + \sqrt{\gamma} \boldsymbol{\Lambda}^{-1} \boldsymbol{X}_T^{\dagger} \bar{\boldsymbol{z}}_T. \tag{4.4}$$

Given $\{g_i\}$, the covariance of the estimation error, $\boldsymbol{\Delta}_h = \hat{\boldsymbol{h}}_{ML} - \boldsymbol{h}$, can be shown to be

$$\mathrm{cov}\left(\boldsymbol{\Delta}_h | \{g_i\}\right) = \gamma \boldsymbol{\Lambda}^{-1} (\boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T)^{-1} \boldsymbol{\Lambda}^{-1}. \tag{4.5}$$

Then averaging over $\{g_i\}$, the mean-square error (MSE) in estimating $\boldsymbol{h}$ is

$$\mathrm{cov}\left(\boldsymbol{\Delta}_h\right) = \bar{\gamma} \boldsymbol{\Lambda}^{-1} (\boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T)^{-1} \boldsymbol{\Lambda}^{-1}, \tag{4.6}$$

where

$$\bar{\gamma} = \mathop{\mathbb{E}}_{\{g_i\}}[\gamma] = \gamma_0 \left(1 + \sum_{i=1}^{R} \varepsilon_i \sigma_{G_i}^2\right). \tag{4.7}$$

It should be emphasized at this point that the total MSE in ML estimation depends on both the power allocation factor in $\boldsymbol{\Lambda}$ and the training matrix in $\boldsymbol{X}_T$, which can be treated separately. To this end, the optimal conditions on the design of the training phase are established by the following 2 lemmas.

Lemma 4. *(From [15]) The optimal training matrix $\boldsymbol{X}_T$ must be orthogonal and $\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T = \boldsymbol{I}_R$ under the ML estimation criterion.*

Proof. Due to the constraint on source power $\boldsymbol{u}^{\mathcal{H}}\boldsymbol{u} \leq 1$, the diagonal elements of $\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T$ need to satisfy $[\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T]_{ii} \leq 1,\ \forall i$. Finding the optimal training matrix is to find the solution that minimizes the total MSE while meeting the above power constraint, i.e.,

$$\begin{aligned} &\underset{\boldsymbol{X}_T}{\text{minimize}}\ \ \text{tr}(\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T)^{-1}\boldsymbol{\Lambda}^{-1}) \\ &\text{subject to } [\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T]_{ii} \leq 1,\ i = 1, \ldots, R. \end{aligned} \tag{4.8}$$

The proof for this lemma follows from Theorem 1 in [15]. At first, $\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T$ is shown to be a diagonal matrix. This can be done by contradiction. Let $\boldsymbol{F}$ be an arbitrary $R \times R$ positive definite matrix, one has

$$\text{tr}(\boldsymbol{F}^{-1}) \geq \sum_{i=1}^{R} ([\boldsymbol{F}]_{ii})^{-1}, \tag{4.9}$$

where the equality hold if and only if $\boldsymbol{F}$ is a diagonal matrix. Suppose that $\boldsymbol{X}_T^{\star}$ is the optimal solution and $(\boldsymbol{X}_T^{\star})^{\mathcal{H}}\boldsymbol{X}_T^{\star}$ is not diagonal. Let $\boldsymbol{D}$ be a diagonal matrix, whose diagonal elements are the diagonal element of $(\boldsymbol{X}_T^{\star})^{\mathcal{H}}\boldsymbol{X}_T^{\star}$. Obviously, $\boldsymbol{D}^{1/2}$ is also a feasible solution to (4.8), as $[\boldsymbol{D}]_{ii} \leq 1,\ \forall i$. Also note that $\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{-1}\boldsymbol{\Lambda}^{-1}$ is now a diagonal matrix. Thus, apply the inequality in (4.9), one has:

$$\begin{aligned} &\text{tr}\left(\boldsymbol{\Lambda}^{-1}\left((\boldsymbol{X}_T^{\star})^{\mathcal{H}}\boldsymbol{X}_T^{\star}\right)^{-1}\boldsymbol{\Lambda}^{-1}\right) = \text{tr}\left(\left(\boldsymbol{\Lambda}(\boldsymbol{X}_T^{\star})^{\mathcal{H}}\boldsymbol{X}_T^{\star}\boldsymbol{\Lambda}\right)^{-1}\right) \\ &< \sum_{i=1}^{R}\left(\left[\boldsymbol{\Lambda}(\boldsymbol{X}_T^{\star})^{\mathcal{H}}\boldsymbol{X}_T^{\star}\boldsymbol{\Lambda}\right]_{ii}\right)^{-1} = \text{tr}\left(\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{-1}\boldsymbol{\Lambda}^{-1}\right), \end{aligned}$$

which contradicts the assumption that $\boldsymbol{X}_T^\star$ is optimal. Thus, the optimal $\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T$ must be diagonal. Applying the power constraint in (4.8), the optimal solution satisfies $\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T = \boldsymbol{I}_R$. The proof is complete. □

Given that the optimal training matrix is orthogonal under the ML estimation, it is of interests to find the optimal power allocation to further minimize the total MSE. Note that the total MSE now is $\mathrm{tr}(\bar{\gamma}\boldsymbol{\Lambda}^{-2})$. Before stating the solution to such a power allocation problem, Lemma 5 is presented next to establish a lower bound on the total MSE.

Lemma 5. *Given $x_i > 0,\ i = 1, \ldots, R$, one has the inequality*

$$\sum_{i=1}^{R} \frac{1}{x_i} \geq \frac{R^2}{\sum_{i=1}^{R} x_i}, \tag{4.10}$$

where the equality is achieved when $x_1 = \ldots = x_R$.

Proof. The proof follows directly by applying the Jensen's inequality [37] on the convex function $f(x) = 1/x$, $x > 0$. □

From Lemma 5, one has

$$\mathrm{tr}(\bar{\gamma}\boldsymbol{\Lambda}^{-2}) = \bar{\gamma}\sum_{i=1}^{R} \frac{1}{\varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2} \geq \bar{\gamma}\frac{R^2}{\sum_{i=1}^{R} \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2}, \tag{4.11}$$

and the equality is met when

$$\varepsilon_1 \sigma_{F_1}^2 \sigma_{G_1}^2 = \ldots = \varepsilon_R \sigma_{F_R}^2 \sigma_{G_R}^2. \tag{4.12}$$

Equivalently,

$$\frac{P_1 \sigma_{F_1}^2 \sigma_{G_1}^2}{P_0 \sigma_{F_1}^2 + 1} = \ldots = \frac{P_R \sigma_{F_R}^2 \sigma_{G_R}^2}{P_0 \sigma_{F_R}^2 + 1}, \tag{4.13}$$

which means that the amount of fading constraint is met.

Observe that $\eta = \sum_{i=1}^{R} \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2/(R\bar{\gamma})$ is the average effective SNR. Thus, the power allocation scheme that maximizes the SNR under the minimum amount of fading constraint will further minimize the total MSE. Such a power allocation scheme was already presented in Chapter 3 (see Proposition 2).

4.2.2 Minimum Mean-Square Error (MMSE) Estimation

The MMSE estimation requires the second-order statistics of the channel to be known at the destination. Let $\boldsymbol{\Sigma}_h$ be the covariance matrix of the channel vector $\boldsymbol{h}$. As $\{f_i\}$ and $\{g_i\}$ are i.i.d. $\mathcal{CN}(0,1)$ random variables, $\boldsymbol{\Sigma}_h = \boldsymbol{I}_R$. Based on the input-output model for the training phase in (4.2), the MMSE estimation yields [15]

$$\begin{aligned}\hat{\boldsymbol{h}}_{MMSE} &= \boldsymbol{\Sigma}_h \boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} \left(\boldsymbol{X}_T \boldsymbol{\Lambda} \boldsymbol{\Sigma}_h \boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} + \bar{\gamma} \boldsymbol{I}_T \right)^{-1} \boldsymbol{y}_T \\ &= \left(\boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T \boldsymbol{\Lambda} + \bar{\gamma} \boldsymbol{I}_R \right)^{-1} \boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{y}_T \\ &= \boldsymbol{B} \hat{\boldsymbol{h}}_{ML}, \end{aligned} \tag{4.14}$$

where $\boldsymbol{B} = \left(\boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T \boldsymbol{\Lambda} + \bar{\gamma} \boldsymbol{I}_R \right)^{-1} \boldsymbol{\Lambda}^2$ is considered as a biasing matrix to the unbiased estimator $\hat{\boldsymbol{h}}_{ML}$ of $\boldsymbol{h}$. The covariance of the estimation error $\boldsymbol{\Delta}_h = \hat{\boldsymbol{h}}_{MMSE} - \boldsymbol{h}$ under the MMSE estimation criterion can be found as

$$\mathrm{cov}(\boldsymbol{\Delta}_h) = \left(\boldsymbol{\Sigma}_h^{-1} + \frac{1}{\bar{\gamma}} \boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T \boldsymbol{\Lambda} \right)^{-1} = \left(\boldsymbol{I}_R + \frac{1}{\bar{\gamma}} \boldsymbol{\Lambda} \boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T \boldsymbol{\Lambda} \right)^{-1}. \tag{4.15}$$

Having known the total MSE as $\mathrm{tr}(\mathrm{cov}(\boldsymbol{\Delta}_h))$, it is now ready to derive the optimal training matrix to minimize the total MSE under the MMSE criterion.

Lemma 6. *(From [15]) The optimal training matrix $\boldsymbol{X}_T$ must be orthogonal and $\boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T = \boldsymbol{I}_R$ under the MMSE estimation criterion.*

Proof. The proof for this lemma is similar to that of Lemma 4, and is thus omitted. □

It is noted that the optimal orthogonal training design under the MMSE estimation criterion is only applicable when the covariance matrix $\boldsymbol{\Sigma}_h$ is a diagonal matrix, i.e., $\boldsymbol{h}$ contains independent elements. The optimal training design for an arbitrary covariance matrix $\boldsymbol{\Sigma}_h$ can be found as in [15].

Given that $\boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T = \boldsymbol{I}_R$, the total MSE under MMSE estimation is then given by

$$\mathrm{tr}\left(\mathrm{cov}(\boldsymbol{\Delta}_h)\right) = \sum_{i=1}^{R} \frac{1}{1 + \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2 / \bar{\gamma}} \geq \bar{\gamma} \frac{R^2}{\bar{\gamma} R + \left(\sum_{i=1}^{R} \varepsilon_i \sigma_{F_i}^2 \sigma_{G_i}^2 \right)}, \tag{4.16}$$

where Lemma 5 has been used again to arrive at the lower bound. Similar to the case of the ML estimation in the previous section, the lower bound in (4.16) is achieved when (4.12) is satisfied. Thus the optimal power allocation in MMSE estimation is also the same as that in ML estimation. Moreover, it is clear from (4.11) and (4.16) that the total MSE in MMSE estimation is always less than that in ML estimation if the same PA scheme is applied. This is an intuitively satisfying result since the MMSE estimation is optimum with regard to the mean-square error measure.

4.2.3 Examples of Training Design

The previous sections have shown that the optimal training matrix must be orthogonal under both ML and MMSE estimation criteria. The remaining task is to determine the training source $\boldsymbol{u}$, and the relaying matrix $\boldsymbol{A}_i$. There are more than one solution. As an example, the training sequence can be chosen as $\boldsymbol{u} = 1/\sqrt{T}[1, \ldots, 1]^{\mathcal{T}} \in \mathbb{R}^T$ to avoid the peak to average power ratio problem at the source. The relaying matrix $\boldsymbol{A}_i$ can be set as in Chapter 2. Such design allows the same relaying matrices at both the training phase and information transmission phase, and simplifies the implementation at the relays. Furthermore, it is possible to have $\boldsymbol{X}_T^{\mathcal{H}} \boldsymbol{X}_T = \boldsymbol{I}_R$.

For a two-relay network, the training time $T = 2$, $\boldsymbol{A}_i, i = 1, 2$ is chosen as in Chapter 2 to design the distributed Alamouti code, the resultant training matrix

$$\boldsymbol{X}_T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

is orthogonal.

For a four-relay network, the training time $T = 4$, $\boldsymbol{A}_i, i = 1, 2, 3, 4$ is chosen as in Chapter 2 to design the distributed QOSTBC. The resultant training matrix

$$\boldsymbol{X}_T = \frac{1}{2} \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

is orthogonal.

4.3 Performance of Mismatched Decoding

This section analyzes the diversity order of the mismatched decoding with imperfect channel estimation. The analysis follows a similar procedure in Section IV of [42] for point-to-point multiple-input multiple-output MIMO systems and it is performed in the large CSNR regime,[2] i.e., when $\gamma_0 \to 0$ (asymptotic analysis). It is important to point out that both γ in (3.1) and $\bar{\gamma}$ in (4.7) are proportional to γ_0, hence they are in the same order of γ_0. The analysis will show that the mismatched decoder of DSTC is able to achieve the same diversity order as that of the coherent decoder.

First, for the coherent decoding of DSTC, it is assumed that the channel $\boldsymbol{h}$ is perfectly known at the destination. The pairwise error probability (PEP) of mistaking the transmitted codeword $\boldsymbol{s}_k$ by $\boldsymbol{s}_l$, i.e., mistaking $\boldsymbol{X}_k$ by $\boldsymbol{X}_l$, is given in Equation (2.10) as

$$\begin{aligned}\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) &= \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}} \mathbb{P}\left(\boldsymbol{X}_k \to \boldsymbol{X}_l | \{f_i\}, \{g_i\}\right) \\ &= \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}} \left[Q\left(\sqrt{\frac{\|\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\|^2}{2\gamma}} \right) \right], \end{aligned} \tag{4.17}$$

where $\boldsymbol{\Delta} = \boldsymbol{X}_k - \boldsymbol{X}_l$.

Now, for the mismatched decoding, the destination uses the estimated CSI $\hat{\boldsymbol{h}}$ in the same way as with the perfect CSI $\boldsymbol{h}$. The decoder performs

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}_k} \left\| \boldsymbol{y} - \boldsymbol{X}_k \boldsymbol{\Lambda} \hat{\boldsymbol{h}} \right\|^2. \tag{4.18}$$

It is noted that under the MMSE channel estimation in (4.14), the Taylor series expansion of the biasing matrix is $\boldsymbol{B} = \left(\boldsymbol{I}_R + \gamma \boldsymbol{\Lambda}^{-2}\right)^{-1} = \boldsymbol{I}_R - O(\gamma_0 \boldsymbol{\Lambda}^{-2})$. This also implies that $\boldsymbol{B} \to \boldsymbol{I}_R$ when $\gamma_0 \to 0$.

Thus, under either ML or MMSE estimation, the channel estimate $\hat{\boldsymbol{h}}$ can be

[2] Recall that the CSNR refers to $P_0 T = 1/\gamma_0$. Since P_0 is allocated a portion of the total power P, the CSNR becomes large when the ratio P becomes large.

expressed as

$$\hat{\boldsymbol{h}} = \boldsymbol{h} + \sqrt{\gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{X}_T^{\mathcal{H}}\bar{\boldsymbol{z}}_T - O\left(\gamma_0\boldsymbol{\Lambda}^{-2}\right). \tag{4.19}$$

The mismatched metric for codeword $\boldsymbol{X}_l$ is

$$\begin{aligned}
\left\|\boldsymbol{y} - \boldsymbol{X}_l\boldsymbol{\Lambda}\hat{\boldsymbol{h}}\right\|^2 &= \left\|\boldsymbol{X}_k\boldsymbol{\Lambda}\boldsymbol{h} + \sqrt{\gamma}\bar{\boldsymbol{z}} - \boldsymbol{X}_l\boldsymbol{\Lambda}\left[\boldsymbol{h} + \sqrt{\gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{X}_T^{\mathcal{H}}\bar{\boldsymbol{z}}_T - O\left(\gamma_0\boldsymbol{\Lambda}^{-2}\right)\right]\right\|^2 \\
&= \left\|\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h} + \sqrt{\gamma}\left(\bar{\boldsymbol{z}} - \boldsymbol{X}_l\boldsymbol{X}_T^{\mathcal{H}}\bar{\boldsymbol{z}}_T\right) - O\left(\gamma_0\boldsymbol{\Lambda}^{-1}\right)\right\|^2 \\
&= \left\|\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\right\|^2 + 2\sqrt{\gamma}\mathrm{Re}\{\bar{\boldsymbol{z}}^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\} \\
&\quad -2\sqrt{\gamma}\mathrm{Re}\{\bar{\boldsymbol{z}}_T^{\mathcal{H}}\boldsymbol{X}_T\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\} + O\left(\gamma_0\right).
\end{aligned} \tag{4.20}$$

Note that, since $\bar{\boldsymbol{z}}$ and $\bar{\boldsymbol{z}}_T$ contain i.i.d. $\mathcal{CN}(0,1)$ random variables, the term $\left\|\sqrt{\gamma}(\bar{\boldsymbol{z}} - \boldsymbol{X}_l\boldsymbol{X}_T^{\mathcal{H}}\bar{\boldsymbol{z}}_T)\right\|^2 = \gamma\left\|\bar{\boldsymbol{z}} - \boldsymbol{X}_l\boldsymbol{X}_T^{\mathcal{H}}\bar{\boldsymbol{z}}_T\right\|^2 \to 0$ almost surely as $\gamma_0 \to 0$. Therefore this term can be subsumed in $O(\gamma_0)$ in (4.20).

The asymptotic PEP of the mismatched decoder can be calculated as

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) &= \mathbb{P}\left(\left\|\boldsymbol{y} - \boldsymbol{X}_l\boldsymbol{\Lambda}\hat{\boldsymbol{h}}\right\|^2 < \left\|\boldsymbol{y} - \boldsymbol{X}_k\boldsymbol{\Lambda}\hat{\boldsymbol{h}}\right\|^2\right) \\
&= \mathbb{P}\left(\left\|\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\right\|^2 + 2\sqrt{\gamma}\mathrm{Re}\{\bar{\boldsymbol{z}}^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\}\right. \\
&\quad \left. -2\sqrt{\gamma}\mathrm{Re}\{\bar{\boldsymbol{z}}_T^{\mathcal{H}}\boldsymbol{X}_T\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\} + O\left(\gamma_0\right) < 0\right) \\
&= \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}}\left[\mathbb{P}\left(\frac{\left\|\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\right\|^2}{\sqrt{\gamma}} + O\left(\sqrt{\gamma_0}\right) < \right.\right. \\
&\qquad \left.\left. 2\mathrm{Re}\{\bar{\boldsymbol{z}}_T^{\mathcal{H}}\boldsymbol{X}_T\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\} - 2\mathrm{Re}\{\bar{\boldsymbol{z}}^{\mathcal{H}}\boldsymbol{\Delta}_s\boldsymbol{\Lambda}\boldsymbol{h}\}\Big|\,\{f_i\},\{g_i\}\right)\right].
\end{aligned} \tag{4.21}$$

Note that, conditioned on $\{g_i\}$ and $\{f_i\}$, the elements of $\bar{\boldsymbol{z}}_T$ and $\bar{\boldsymbol{z}}$ are i.i.d. $\mathcal{CN}(0,1)$ random variables. Thus, $2\mathrm{Re}\{\bar{\boldsymbol{z}}_T^{\mathcal{H}}\boldsymbol{X}_T\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\} - 2\mathrm{Re}\{\bar{\boldsymbol{z}}^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\}$ is Gaussian distributed with zero mean and variance $2\|\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\|^2 + 2\|\boldsymbol{\Delta}\boldsymbol{\Lambda}\boldsymbol{h}\|^2$, since $\boldsymbol{X}_T^{\mathcal{H}}\boldsymbol{X}_T = \boldsymbol{I}_R$. The probability term in (4.21) is just the probability that a zero-mean Gaussian ran-

dom variable is bigger than some constant. Therefore (4.21) can be written as

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) &= \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}}\left[Q\left(\sqrt{\frac{\left[\frac{\left\|\boldsymbol{\Delta\Lambda h}\right\|^2}{\sqrt{\gamma}}+O\left(\sqrt{\gamma_0}\right)\right]^2}{2\|\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta\Lambda h}\|^2+2\|\boldsymbol{\Delta\Lambda h}\|^2}}\right)\right] \\
&= \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}}\left[Q\left(\sqrt{\frac{\frac{\left\|\boldsymbol{\Delta\Lambda h}\right\|^4}{\gamma}+2\frac{\left\|\boldsymbol{\Delta\Lambda h}\right\|^2}{\sqrt{\gamma}}O\left(\sqrt{\gamma_0}\right)}{2\|\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta\Lambda h}\|^2+2\|\boldsymbol{\Delta\Lambda h}\|^2}}\right)\right] \\
&= \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}}\left[Q\left(\sqrt{\frac{1}{2\gamma}\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{1+\frac{\|\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta\Lambda h}\|^2}{\|\boldsymbol{\Delta\Lambda h}\|^2}}+O\left(1\right)}\right)\right]. \qquad (4.22)
\end{aligned}$$

Similar to [42], apply the Cauchy-Schwarz inequality to the Frobenius norm [28], one has

$$1 \le 1+\frac{\|\boldsymbol{X}_l^{\mathcal{H}}\boldsymbol{\Delta\Lambda h}\|^2}{\|\boldsymbol{\Delta\Lambda h}\|^2} \le 1+\|\boldsymbol{X}_l\|^2. \qquad (4.23)$$

Thus, due to the monotonic decreasing of the Q function, the asymptotic PEP is bounded as

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\{f_i\},\{g_i\}}\left[Q\left(\sqrt{\frac{1}{2\gamma}\|\boldsymbol{\Delta\Lambda h}\|^2+O\left(1\right)}\right)\right] \\
&\quad\le \mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) \qquad (4.24) \\
&\quad\le \mathop{\mathbb{E}}_{\{f_i\},\{g_i\}}\left[Q\left(\sqrt{\frac{1}{2\gamma}\frac{\|\boldsymbol{\Delta\Lambda h}\|^2}{1+\|\boldsymbol{X}_l\|^2}+O\left(1\right)}\right)\right].
\end{aligned}$$

Note that at high CSNR, i.e., small γ, the constant term $O\left(1\right)$ becomes negligible compared to the term $\|\boldsymbol{\Delta\Lambda h}\|^2/(2\gamma)$ or the term $\|\boldsymbol{\Delta\Lambda h}\|^2/(2\gamma(1+\|\boldsymbol{X}_l\|^2))$. Therefore it can be neglected as far as the diversity order analysis is concerned. Comparing the PEP expression in (4.17) and the bounds in (4.24) clearly shows that the diversity order of the mismatched decoder is the same as that of the coherent decoder if the same PA scheme is applied. Moreover, it is of interest to find the PA scheme that achieves the maximum diversity order with the mismatched decoder. This is discussed next.

In Chapters 2-3, it was shown that if $\boldsymbol{\Delta} = \boldsymbol{X}_k - \boldsymbol{X}_l$ is full rank, the PEP in (4.17) decays in the order of $R(1 - \text{loglog}\eta/\text{log}\eta)$, as long as the optimal power allocation scheme under the minimum amount of fading is exercised. The analysis and simulation in Chapter 3 also indicate a significant performance improvement by the optimal power allocation scheme over the equal power allocation scheme. As a consequence of the analysis on the PEP of the mismatched decoding given in (4.24), under the optimal power scheme, the mismatched decoder is able to not only realize the maximum diversity order of the DSTC, but also outperform the equal power allocation scheme substantially.

4.4 Simulation Results

This section presents the simulation results on the total MSE of the ML and MMSE estimators, as well as the SER of the mismatched decoder in order to illustrate the superiority of the proposed optimal PA in both estimation and decoding performances. Considered are the relay networks with 2 or 4 relays as in Section 3.7 of Chapter 3. The second order statistics of each $S \rightarrow R$ and $R \rightarrow D$ channel are kept the same.

Figure 4.1 compares the total MSE achieved by the optimal and equal PA schemes in a four-relay network, for both the ML and MMSE estimations. It can be seen that, with either ML or MMSE estimation, the total MSE is significantly reduced with the optimal PA over the suboptimal equal PA scheme. The total MSE of the MMSE estimation is also smaller than that of the ML criterion with both PA schemes. However, at high CSNR, the difference is negligible, which validates the common representation of the two estimators in (4.19).

The SER performance of the mismatched decoder is compared to that of the coherent decoder under the two PA schemes in Figures 4.2 and 4.3. The distributed Alamouti code is implemented for the two-relay network, while the distributed quasi-orthogonal space-time block code (QOSTBC) is applied in the four-relay network [9]. As can be seen from the two figures, the diversity order achieved with the mismatched

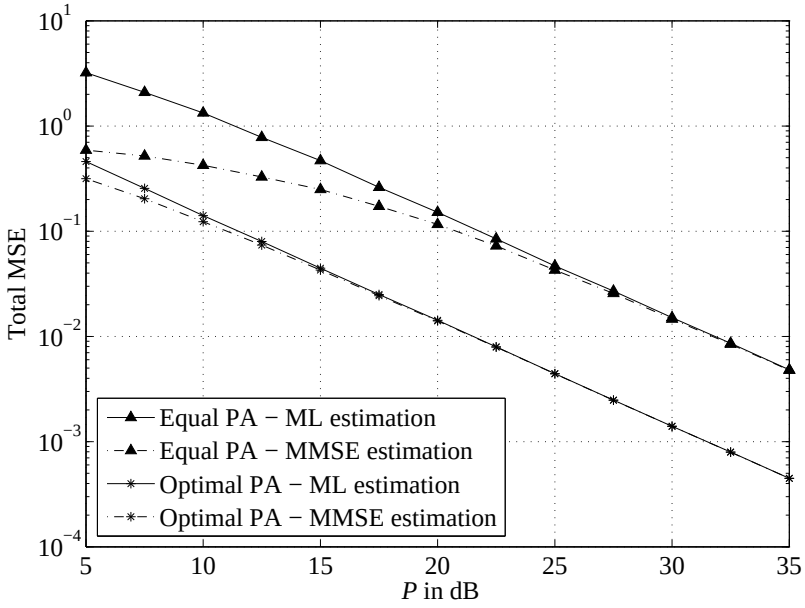

Figure 4.1 Total MSE achieved with ML and MMSE estimators, with the optimal and equal PA schemes in a four-relay network.

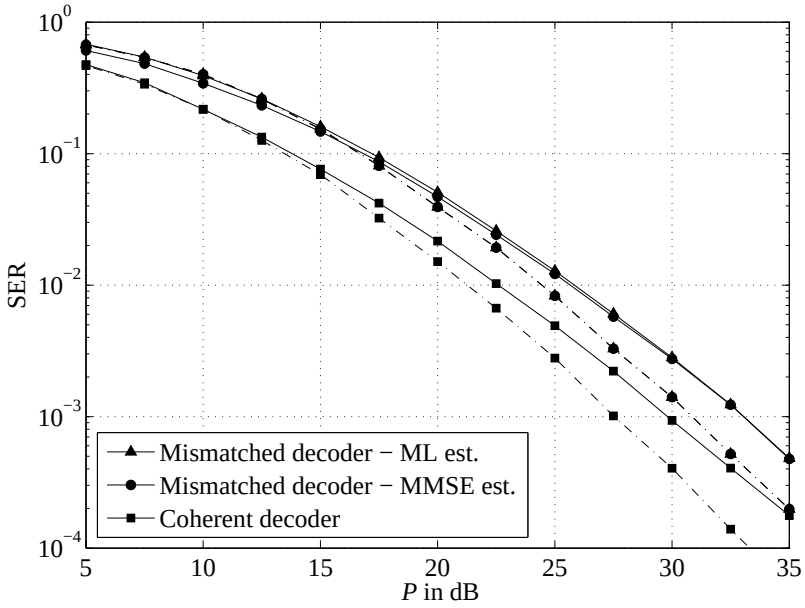

Figure 4.2 Error performance of DSTC with different types of detection in a two-relay network. "Dash-dot" lines are for the optimal PA scheme, solid lines are for the equal PA scheme.

decoding is the same as that of the coherent decoding, regardless of the PA scheme applied. This agrees with the analysis in Section 4.3. Moreover, under the optimal PA scheme, the mismatched decoder is able to achieve the maximum diversity order and significantly outperforms the decoder under the equal PA scheme, by about 2.5 dB and 5.0 dB in the two-relay and four-relay networks, respectively.

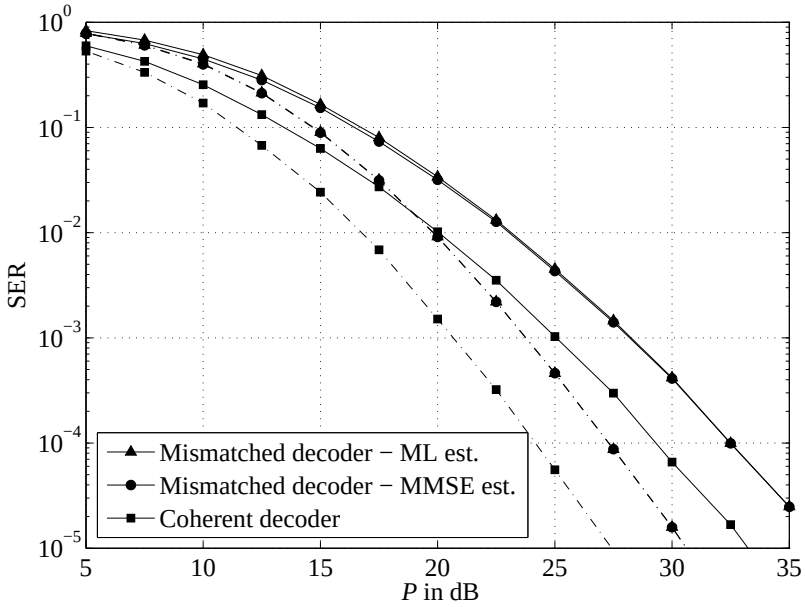

Figure 4.3 Error performance of DSTC with different types of detection in a four-relay network. "Dash-dot" lines are for the optimal PA scheme, solid lines are for the equal PA scheme.

Finally, the two figures show that the mismatched decoders perform almost the same with both the ML and MMSE channel estimations and under either the optimal or equal PA schemes, especially at high CSNR. In fact, their performances are identical for the case of distributed Alamouti code (orthogonal code) and optimal PA. This is because with the optimal PA, the channel estimate by the MMSE estimator is a scaled version of the channel estimate by the ML estimator, whereas the decoding performance of the orthogonal code is insensitive to the scaling factor in the channel estimate.

4.5 Summary

This section has studied optimal training design in wireless multi-relay networks. It was shown that orthogonal training scheme is optimal with both the ML and MMSE criteria. The optimal power allocation scheme to minimize the total MSE of both ML and MMSE channel estimations was obtained for distributed space-time coding (DSTC). The diversity order of the error performance of the mismatched decoder that works with the estimated channel information was also analyzed. It was shown that, with a given power allocation scheme in the data transmission phase, the mismatched decoder is able to achieve the *same* diversity order as the coherent decoder. In particular, if the optimal power allocation under the minimum amount of fading obtained in the training phase is also applied to the transmission phase, then the mismatched decoder achieves the same *maximum* diversity order as that of the coherent decoder.

Part II

Distributed Beamforming

5. Distributed Beamforming in a Multiuser Multi-relay Network with Guaranteed QoS

In the first part of this thesis, distributed space-time coding (DSTC) for a one-source one-destination network was studied. It was shown that DSTC can take advantage of cooperative diversity in the absence of CSI at the relays. However, should the relays know both the backward, i.e., source-to-relay ($S \rightarrow D$) and forward, i.e., relay-to-destination ($R \rightarrow D$), channels, they can beam their retransmitted signals so that the received signal at the destination is coherently constructed. The approach, referred to as distributed beamforming, is investigated in [11–14, 45]. In particular, reference [11] considers the problem of controlling the transmitted power at each relay in order to maximize the signal-to-noise ratio (SNR) at the destination. It shows that, depending on its own channels and other relays' channels, each relay may not use its maximum power to obtain the optimal SNR. The same problem is also considered in [13] by the technique of conic programming. Reference [12] studies a distributed relay strategy for wireless sensor networks to obtain a certain target SNR at the destination, whereas reference [14] investigates a similar problem with the objective of minimizing the sum of relay powers, referred to as "sum relay power" hereafter. More recently, distributed beamforming with second-order statistics is examined in [45].

It is noted that most of the early works in distributed beamforming focus on the system with one source and one destination. Multiuser multi-relay systems are first investigated in [46] and [47], where the relay strategies were proposed to minimize the mean-square error between the source signal and the received signal at the desti-

[0]The contributions in this chapter are also presented in [43, 44].

nation. In addition, such systems allow the relays to share their received signals from the sources, and thus require reliable links between the relays. In this work, those additional links are not needed as the relays *do not* share their received signals.

This part of thesis studies optimal distributed beamforming designs in a multiuser multi-relay network with multiple sources and multiple destinations. It is assumed that all the source-destination pairs operate in orthogonal channels to avoid inter-user interference at the destinations. The distributed beamforming designs are then optimized under one of the following two design criteria: (i) minimizing the sum relay power with guaranteed quality of service (QoS) in terms of SNR at the destinations, or (ii) maximizing the joint SNR margin subject to power constraints at the relays. These two design criteria are sequentially investigated in this chapter and Chapter 6 and shown to be closely related with each other.

In this chapter, the optimal distributed beamforming strategy under the design criterion (i) is examined. Considered are the optimization problems with and without per-relay power constraints. It will be shown that these two optimization problems can be transformed into convex second-order conic programs (SOCPs), and thus, can be solved effectively by any conic software package. In addition, this chapter also proposes simple and fast iterative algorithms to solve the optimization problems under consideration.

When no power constraints are applied, the problem can be recast as power minimization problems to determine the minimum relay power required for each user. The corresponding optimal distributed beamformer at the relays for each user is then presented in a closed-form expression, given its allocated relay power. A simple iterative fixed point algorithm is proposed and showed to converge to the optimal relay power value.

When per-relay power constraints are applied, the duality of the problem is established and studied. Through the interpretation via a virtual uplink channel, the dual problem could be solved quickly by the iterative fixed point algorithm. Finally, an

iterative algorithm based on fixed point iteration and Euclidean projection method is proposed to solve the original distributed beamforming problem.

5.1 System Model

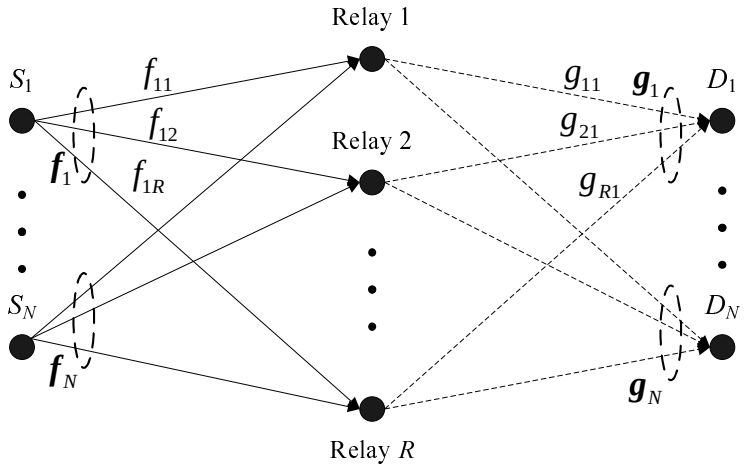

Figure 5.1 Block diagram of a distributed beamforming system with R relays and N users.

Consider a relay network with N pairs of source-destination users (S_n-D_n, $n = 1, \ldots, N$), and R relays,[1] as illustrated in Figure 5.1. All relays are assumed to work in a half-duplex mode, i.e., they cannot receive and transmit at the same time. Assume that there is no direct link between any source and destination and the communication between the two terminals of each user is assisted by all the relays, and implemented in two transmission stages. These assumptions are the same as that of the DSTC system presented in Chapter 2, except that the system model now consists of multiple $S - D$ users.

In the first stage, each user's source broadcasts its signals to all the relays. The transmissions from the sources are conducted over orthogonal channels, using time-division multiple access (TDMA) or frequency-division multiple access (FDMA), for instance, such that there is no inter-user interference at the relays. For the nth user,

[1]This system model is also applicable to a one-source one-destination OFDM system, where N is interpreted as the number of subcarriers.

given s_n as the source signal, the received signals at the relays are given as

$$\boldsymbol{r}_n = \boldsymbol{f}_n s_n + \boldsymbol{z}_{r_n} \in \mathbb{C}^{R\times 1}, \tag{5.1}$$

where $\boldsymbol{f}_n = [f_{n,1}, \ldots, f_{n,R}]^{\mathcal{T}}$, and $f_{n,i}$ is the channel from the nth source to the ith relay; $\boldsymbol{z}_{r_n}$ represents the AWGN at the relays, which contains i.i.d. $\mathcal{CN}(0, \sigma_R^2)$.

At the ith relay, the received signal for the nth user is amplified by a complex beamforming weight $w_{n,i}$, which is to be designed. Let $\boldsymbol{w}_n = [w_{n,1}, \ldots, w_{n,R}]^{\mathcal{T}}$ be the vector of the beamforming weights for the nth user. Also define $\boldsymbol{W}_n = \mathrm{diag}(\boldsymbol{w}_n)$. Accordingly, by applying the AF protocol [5], the retransmitted signals from the relays scheduled for the nth user are formed as

$$\boldsymbol{t}_n = \boldsymbol{W}_n \boldsymbol{r}_n = \boldsymbol{W}_n \boldsymbol{f}_n s_n + \boldsymbol{W}_n \boldsymbol{z}_{r_n}. \tag{5.2}$$

In the second stage of transmission, all the relays simultaneously transmit to the nth user's destination. Similar to the first stage, the transmission to each user's destination is carried out over orthogonal channels to avoid inter-user interference. Let $\boldsymbol{g}_n = [g_{1,n}, \ldots, g_{R,n}]^{\mathcal{T}}$ represent the channels from R relays to the nth destination. The received signal at the nth destination is written as

$$y_n = \boldsymbol{g}_n^{\mathcal{T}} \boldsymbol{t}_n + z_{d_n} = \boldsymbol{g}_n^{\mathcal{T}} \boldsymbol{W}_n \boldsymbol{f}_n s_n + \boldsymbol{g}_n^{\mathcal{T}} \boldsymbol{W}_n \boldsymbol{z}_{r_n} + z_{d_n}, \tag{5.3}$$

where $z_{d_n} \sim \mathcal{CN}(0, \sigma_D^2)$ is the AWGN at the destination. Obviously, $\boldsymbol{g}_n^{\mathcal{T}} \boldsymbol{W}_n \boldsymbol{f}_n s_n$ represents the signal part, whereas $\boldsymbol{g}_n^{\mathcal{T}} \boldsymbol{W}_n \boldsymbol{z}_{r_n} + z_{d_n}$ represents the noise part at the destination. Define $\boldsymbol{h}_n^* = [h_{n,1}^*, \ldots, h_{n,R}^*]^{\mathcal{T}} = \boldsymbol{f}_n \odot \boldsymbol{g}_n = [f_{n,1} g_{1,n}, \ldots, f_{n,R} g_{R,n}]^{\mathcal{T}}$, where $\odot$ represents the component-wise Hadamard product. As a result, $\boldsymbol{h}_n^*$ models the effective channel from source-n to destination-n through all the relays, excluding the beamforming factors. Then, one has $\boldsymbol{g}_n^{\mathcal{T}} \boldsymbol{W}_n \boldsymbol{f}_n = \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n$. Let $\sigma_{S_n}^2 = \mathbb{E}[|s_n|^2]$ be the average transmitted power of the nth source. Then, the SNR at the nth destination is given by

$$\mathrm{SNR}_n = \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2}, \tag{5.4}$$

where $\boldsymbol{G}_n = \mathrm{diag}(|g_{1,n}|^2, \ldots, |g_{R,n}|^2)$.

Let p_n be the total relay power allocated for the nth user, one has

$$p_n = \mathbb{E}\left[\|\boldsymbol{t}_n\|^2\right] = \sigma_{S_n}^2 \sum_{i=1}^{R} |w_{n,i}|^2 |f_{n,i}|^2 + \sigma_R^2 \sum_{i=1}^{R} |w_{n,i}|^2 = \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n, \tag{5.5}$$

where $\boldsymbol{D}_n$ is a $R \times R$ diagonal matrix, with the ith diagonal element $[\boldsymbol{D}_n]_{ii} = \sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2$. On the other hand, the transmitted power at the ith relay is

$$P_i = \sum_{n=1}^{N} \mathbb{E}\left[|t_{n,i}|^2\right] = \sum_{n=1}^{N} \sigma_{S_n}^2 |w_{n,i}|^2 |f_{n,i}|^2 + \sigma_R^2 \sum_{n=1}^{N} |w_{n,i}|^2 = \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n, \tag{5.6}$$

where $\boldsymbol{E}_i$ is a $R \times R$ matrix whose elements are zero, except the (i,i)-element, which is $[\boldsymbol{E}_i]_{ii} = 1$. The total transmitted power of all the relays (and for all the users) is therefore given by

$$P_{\text{relay}} = \sum_{i=1}^{R} P_i = \sum_{n=1}^{N} p_n = \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n. \tag{5.7}$$

5.2 Sum-Power Minimization

This section considers the optimal design of the beamforming vectors to minimize the sum power at the relays given the set of target SNRs at the destinations. This design provides a relaying strategy that can flexibly meet the quality of service (QoS) requirement at each user's destination. The optimization problem is formulated as follows:

$$\begin{aligned} \underset{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{minimize}} \quad & \sum_{n=1}^{N} p_n \\ \text{subject to} \quad & \text{SNR}_n \geq \gamma_n, \ n = 1, \ldots, N, \end{aligned} \tag{5.8}$$

where γ_n is the target SNR at the nth destination. Obviously, this optimization problem can be performed separately through N smaller optimization problems, each corresponds to one user. That is

$$\begin{aligned} \underset{\boldsymbol{w}_n}{\text{minimize}} \quad & \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n \\ \text{subject to} \quad & \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2} \geq \gamma_n. \end{aligned} \tag{5.9}$$

This optimization problem is illustrated in Figure 5.2. In particular, the effective channel coefficient from source-n to destination-n through relay-i is denoted as $h_{n,i}^*$.

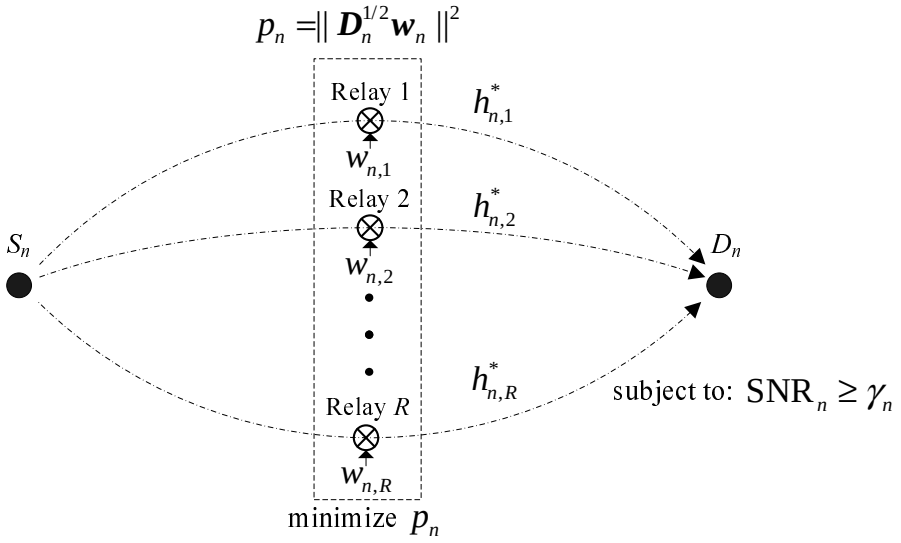

Figure 5.2 Sum relay power minimization for the nth user.

At relay-i, the retransmitted signal is amplified by the weight $w_{n,i}$. The relay power for user-n is given by $p_n = \|\boldsymbol{D}_n^{1/2}\boldsymbol{w}_n\|^2$. The optimization problem then attempts to minimize p_n subjected to the constraint on target SNR of user-n to be at least equal to γ_n.

While this problem is not readily in a convex form, it can be recast as an SOCP as follows. As a phase rotation to $\boldsymbol{w}_n$ has no impact on the SNR constraint and the objective function, without loss of generality, it is assumed that $\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n$ is real. Then the SNR constraint can now be recast as [13]

$$\sqrt{\frac{\sigma_{S_n}^2}{\gamma_n}}\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n \geq \left\| \begin{matrix} \sigma_R \boldsymbol{G}_n^{1/2}\boldsymbol{w}_n \\ \sigma_D \end{matrix} \right\|, \tag{5.10}$$

which is an SOC constraint. In addition, the objective function is quadratic. Thus, both the objective function and the constraint are convex, so is the optimization problem (5.9).[2] The solution to the problem can be obtained from any external conic solution package, such as `cvx` [48]. However, as the required conic package is not always readily available, the approach may not be suitable in real-time communications. To overcome this difficulty, this section considers an alternative approach to

[2]Detailed discussions on convex set, convex function, and convex optimization are in Appendix B.

optimize the relay power consumption p_n directly, instead of dealing with the beamforming vector. The new approach also motivates a simple and fast algorithm to solve the optimization problem.

First, Problem (5.9) can also be recast as

$$\begin{aligned} \underset{\boldsymbol{w}_n}{\text{minimize}} \quad & p_n \\ \text{subject to} \quad & \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2} \geq \gamma_n \\ & \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n = p_n. \end{aligned} \tag{5.11}$$

Second, the following lemma establishes the relation between the optimal beamforming vector $\boldsymbol{w}_n$ and the allocated relay power p_n.

Lemma 7. *Given p_n as the relay power allocated for user-n, the optimal beamforming weights at the relays to maximize the SNR of user-n are*

$$w_{n,i} = \frac{\sqrt{\delta_n} f_{n,i}^* g_{i,n}^*}{p_n \sigma_R^2 |g_{i,n}|^2 + \sigma_D^2 (\sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2)}, \tag{5.12}$$

where the normalization factor δ_n is

$$\delta_n = \frac{p_n}{\sum_{i=1}^{R} \frac{|f_{n,i}|^2 |g_{i,n}|^2 (\sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2)}{\left[p_n \sigma_R^2 |g_{i,n}|^2 + \sigma_D^2 (\sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2)\right]^2}}. \tag{5.13}$$

The corresponding maximum SNR is

$$\mathrm{SNR}_n(p_n) = \sum_{i=1}^{R} \frac{p_n \sigma_{S_n}^2 |f_{n,i}|^2 |g_{i,n}|^2}{p_n \sigma_R^2 |g_{i,n}|^2 + \sigma_D^2 (\sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2)}. \tag{5.14}$$

Proof. The derivation of the optimal beamforming weight vector $\boldsymbol{w}_n$ is similar to [49], and based on the Rayleigh-Ritz theorem [28].[3] Substitute $p_n = \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n$ into the SNR expression, one has

$$\mathrm{SNR}_n = \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \frac{\sigma_D^2}{p_n} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n} = \frac{p_n \sigma_{S_n}^2 \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n}{\boldsymbol{w}_n^{\mathcal{H}} (p_n \sigma_R^2 \boldsymbol{G}_n + \sigma_D^2 \boldsymbol{D}_n) \boldsymbol{w}_n}. \tag{5.15}$$

[3]Detailed discussion on the Rayleigh-Ritz theorem is in Appendix C.

Let $p_n\sigma_R^2\boldsymbol{G}_n + \sigma_D^2\boldsymbol{D}_n$ be decomposed into Cholesky factors as $\boldsymbol{L}_n\boldsymbol{L}_n^{\mathcal{H}}$. Apply Corollary 1 in Appendix C to the expression of SNR_n, the optimal beamforming vector $\boldsymbol{w}$ to maximize SNR_n is a scaled version of the vector $(p_n\sigma_R^2\boldsymbol{G}_n + \sigma_D^2\boldsymbol{D}_n)^{-1}\boldsymbol{h}_n$. It follows that the closed-form expression of each beamforming weight can be stated as in (5.12), while the normalization factor δ_n in (5.13) is to ensure $\boldsymbol{w}_n^{\mathcal{H}}\boldsymbol{D}_n\boldsymbol{w}_n = p_n$.

It is also noted that the largest eigenvalue of $\boldsymbol{L}_n^{-1}\boldsymbol{h}_n\boldsymbol{h}_n^{\mathcal{H}}(\boldsymbol{L}_n^{\mathcal{H}})^{-1}$ is its only non-zero eigenvalue, which is also its trace. From this fact, the maximum SNR value can be found in a closed-form expression as stated in (5.14). □

It should be pointed out that the result presented in Lemma 7 is different from the result in Section III.B of [49], where the optimal beamforming weight $\boldsymbol{w}_n$ is incorrectly stated as $\boldsymbol{L}_n^{-1}\boldsymbol{h}_n$. The normalization factor in [49] is also different from the correct value presented in Lemma 7.

By applying Lemma 7, one can optimize the power allocation requirement p_n for user-n, then determine the optimal beamforming vector accordingly. For notational simplicity, let

$$a_{n,i} = \frac{\sigma_{S_n}^2}{\sigma_R^2}|f_{n,i}|^2,\ b_{n,i} = \frac{\sigma_D^2(\sigma_{S_n}^2|f_{n,i}|^2 + \sigma_R^2)}{\sigma_R^2|g_{i,n}|^2}.$$

Then, the achievable SNR can be written as

$$\text{SNR}_n(p_n) = \sum_{i=1}^{R}\frac{a_{n,i}p_n}{b_{n,i}+p_n}, \tag{5.16}$$

which means that the achievable SNR at the nth destination solely depends on the relaying power p_n.

The optimization problem is now restated as

$$\mathcal{P}_n(\gamma_n) = \begin{cases} \underset{p_n}{\text{minimize}} & p_n \\ \text{subject to} & \sum_{i=1}^{R}\frac{a_{n,i}p_n}{b_{n,i}+p_n} \geq \gamma_n, \end{cases} \tag{5.17}$$

where $\mathcal{P}_n(\gamma_n)$ denotes the minimum relay power allocated for user-n in order to obtain its target SNR γ_n, and also denotes the optimization problem (5.17) itself. Like the original problem in (5.9), the problem in (5.17) is also convex, which then can be

solved efficiently. In addition, the structure of the restated problem also reveals several interesting properties of the problem, including its feasibility and solution. Since $p_n/(b_{n,i}+p_n) < 1$, one has $\text{SNR}_n = \sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n} < \sum_{i=1}^{R} a_{n,i}$. Thus, if the target SNR $\gamma_n \geq \sum_{i=1}^{R} a_{n,i}$, the problem will be infeasible.

Now, suppose that the target SNR is set such that the problem is feasible. Since $\sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n}$ is a monotonically increasing function,[4] the constraint $\sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n} \geq \gamma_n$ must be met with equality at optimum. Thus, the unique solution of

$$\sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n} = \gamma_n \tag{5.18}$$

is also the optimal solution to (5.17).

It is then of interest to find a simple and fast numerical algorithm to solve the R-th polynomial in (5.18). The monotonicity of $\sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n}$ makes the bisection method especially suitable to find the solution of (5.18). On the other hand, the structure in (5.18) also motivates a simple fixed point iteration algorithm to find the optimal $p_n^\star$. By rearranging (5.18), one has the following simple iteration.

$$p_n^{(t+1)} = \frac{\gamma_n}{\sum_{i=1}^{R} \frac{a_{n,i}}{b_{n,i}+p_n^{(t)}}}. \tag{5.19}$$

If (5.18) is feasible, then the above iteration will converge from any initial point $p_n^{(0)} \geq 0$. The convergence analysis of the fixed point iteration is based on the *standard* function approach introduced in [50]. Denote $f_n(p_n^{(t)}) = \frac{\gamma_n}{\sum_{i=1}^{R} \frac{a_{n,i}}{b_{n,i}+p_n^{(t)}}}$, then the fixed point iteration $p_n^{(t+1)} = f_n(p_n^{(t)})$ will converge to a unique fixed point $p_n^\star$ if the function $f_n(p_n)$ obeys the following properties [50]:

1. *Positivity:* $f_n(p_n) > 0$ for all $p_n > 0$.

2. *Monotonicity:* if $p_n > p_n'$, then $f_n(p_n) > f_n(p_n')$.

3. *Scalability:* if $\alpha > 1$, then $\alpha f_n(p_n) > f_n(\alpha p_n)$.

[4] It is easy to verify that $\frac{a_{n,i}p_n}{b_{n,i}+p_n}$, with $a_{n,i} > 0$, $b_{n,i} > 0$, and $p_n > 0$, is a monotonically increasing function. Then, the summation $\sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n}$ is also monotonically increasing.

It is easy to verify that all these three properties are satisfied by the function $f_n(p_n)$. Thus, the fixed point iteration (5.19) will surely converge if (5.18) is feasible.

In summary, this section has considered the problem of optimal distributed beamforming strategy to minimize the sum relay power with guaranteed QoS at each destination. It was shown that for a given sum relay power, the distributed beamforming can be determined in a closed-form solution. The optimization problem then can be recast as the power minimization problem at the relays. Also presented was a simple iterative fixed-point algorithm to determine the required sum power.

5.3 Sum-Power Minimization with Per-Relay Power Constraints

In the previous section, sum relay power minimization with guaranteed QoS at the destinations was considered. No restrictions on the individual power at each relay were imposed. However, in practical relay communications, each relay is equipped with its own amplifier and has its own power limit. Under the per relay power constraints, the relay strategy has to be modified accordingly while meeting the SNR requirement at each user's receiving end. This section considers the approach to uniformly minimize the margin $P_i/P_i^{\max}$ over all the relays, where $P_i^{\max}$ denotes the maximum transmitted power of the ith relay. The problem is stated as follows:

$$\begin{aligned} \underset{\alpha,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{minimize}} \quad & \alpha \\ \text{subject to} \quad & \text{SNR}_n \geq \gamma_n,\ n = 1,\ldots,N \\ & P_i \leq \alpha P_i^{\max},\ i = 1,\ldots,R. \end{aligned} \tag{5.20}$$

The problem is equivalent to

$$\begin{aligned} \underset{\alpha,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{minimize}} \quad & \alpha \sum_{i=1}^{R} P_i^{\max} \\ \text{subject to} \quad & \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2} \geq \gamma_n,\ n = 1,\ldots,N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n \leq \alpha P_i^{\max},\ i = 1,\ldots,R. \end{aligned} \tag{5.21}$$

Note that by multiplying the constant factor $\sum_{i=1}^{R} P_i^{\max}$ into the objective function, the optimization problem can be interpreted as a sum relay power minimization problem with per-relay power constraint awareness. This optimization problem is illustrated in Figure 5.3, where the objective function and the constraints are highlighted.

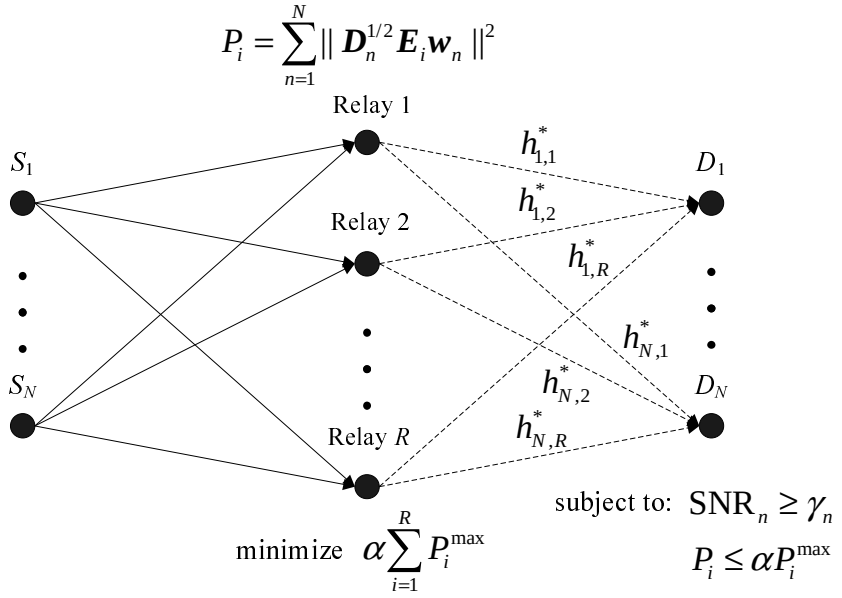

Figure 5.3 Sum relay power minimization with per-relay power constraints.

It is also noted that no explicit constraint on variable α is imposed in the optimization problem. Hence, it might happen that $\alpha > 1$ at optimum, i.e., the power consumption at one of the relays, say relay-i, exceeds its strict limit $P_i^{\max}$. As a result, it is not possible to find the beamforming vectors that meet both the QoS constraints and the strict per-relay power constraints. In such a case, an inverse problem, which tries to maximize the SNR under strict per-relay power constraints, may be of interest. This inverse problem shall be investigated in Chapter 6.

The approach of minimizing the power consumption margin was first investigated for the multiuser beamforming downlink problem in point-to-point communications [51]. When applied to relay networks, the idea is to serve all the users, while maintaining the balance in power consumption at the relays. Alternatively, one may formulate a problem to minimize the sum relay power with strict individual per-relay power constraints, i.e., no α is involved. Such a formulation would lead to

different dual problem and solution, although the analysis would be similar to what presented here.

Note that the optimization problem stated in (5.21) is not readily convex. However, as the SNR constraints can be recast as a SOC constraint as in (5.10), the problem can be transformed into a convex one. The following proposition establishes the relation between the Lagrangians of the convex and nonconvex forms of (5.21) and reveals the strong duality property of the optimization problem.

Proposition 3. *Strong duality holds for the optimization problem* (5.21).

Proof. First, the convex form of (5.21) is

$$\begin{aligned} \underset{\alpha,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{minimize}} \quad & \alpha \sum_{i=1}^{R} P_n^{\max} \\ \text{subject to} \quad & \sqrt{\frac{\sigma_{S_n}^2}{\gamma_n}} \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n \geq \left\| \begin{array}{c} \sigma_R \boldsymbol{G}_n^{1/2} \boldsymbol{w}_n \\ \sigma_D \end{array} \right\|, \ n = 1, \ldots, N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n \leq \alpha P_i^{\max}, \ i = 1, \ldots, R. \end{aligned} \tag{5.22}$$

Since the reformulated problem is convex, strong duality holds [37]. It is of interest to find whether the dual gap to the original nonconvex problem is also zero. The Lagrangian of the convex problem (5.22) is given by

$$\begin{aligned} \tilde{\mathcal{L}}(\alpha, \boldsymbol{w}_n, \boldsymbol{\nu}, \boldsymbol{\mu}) = \alpha \sum_{i=1}^{R} P_i^{\max} - \sum_{n=1}^{N} \nu_n \left(\sqrt{\frac{\sigma_{S_n}^2}{\gamma_n}} \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n - \left\| \begin{array}{c} \sigma_R \boldsymbol{G}_n^{1/2} \boldsymbol{w}_n \\ \sigma_D \end{array} \right\| \right) \\ + \sum_{i=1}^{R} \mu_i \left(\sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n - \alpha P_i^{\max} \right), \end{aligned} \tag{5.23}$$

where $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_N]^{\mathcal{T}}$, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_R]^{\mathcal{T}}$; ν_n's and μ_i's are the Lagrangian multipliers associated with the SNR and the per-relay power constraints, respectively. Let

$$\varepsilon_n = \sqrt{\frac{\sigma_{S_n}^2}{\gamma_n}} \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n + \left\| \begin{array}{c} \sigma_R \boldsymbol{G}_n^{1/2} \boldsymbol{w}_n \\ \sigma_D \end{array} \right\|,$$

then,

$$\nu_n\left(\sqrt{\frac{\sigma_{S_n}^2}{\gamma_n}}\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n-\left\|\begin{matrix}\sigma_R\boldsymbol{G}_n^{1/2}\boldsymbol{w}_n\\ \sigma_D\end{matrix}\right\|\right)=\frac{\nu_n}{\varepsilon_n}\left(\frac{\sigma_{S_n}^2}{\gamma_n}\left(\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n\right)^2-\left\|\begin{matrix}\sigma_R\boldsymbol{G}_n^{1/2}\boldsymbol{w}_n\\ \sigma_D\end{matrix}\right\|^2\right)$$
$$=\frac{\nu_n}{\varepsilon_n}\left(\frac{\sigma_{S_n}^2}{\gamma_n}\left(\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n\right)^2-\sigma_R^2\boldsymbol{w}_n^{\mathcal{H}}\boldsymbol{G}_n\boldsymbol{w}_n-\sigma_D^2\right).$$

Substitute to the Lagrangian in (5.23), one has

$$\tilde{\mathcal{L}}(\alpha,\boldsymbol{w}_n,\boldsymbol{\nu},\boldsymbol{\mu})=\alpha\sum_{i=1}^{R}P_i^{\max}-\sum_{n=1}^{N}\frac{\nu_n}{\varepsilon_n}\left(\frac{\sigma_{S_n}^2}{\gamma_n}\left(\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n\right)^2-\sigma_R^2\boldsymbol{w}_n^{\mathcal{H}}\boldsymbol{G}_n\boldsymbol{w}_n-\sigma_D^2\right)$$
$$+\sum_{i=1}^{R}\mu_i\left(\sum_{n=1}^{N}\boldsymbol{w}_n^{\mathcal{H}}\boldsymbol{D}_n\boldsymbol{E}_i\boldsymbol{w}_n-\alpha P_i^{\max}\right). \tag{5.24}$$

Now, consider the Lagrangian of the original nonconvex problem (5.21)

$$\mathcal{L}(\alpha,\boldsymbol{w}_n,\boldsymbol{\lambda},\boldsymbol{\mu})=\alpha\sum_{i=1}^{R}P_i^{\max}-\sum_{n=1}^{N}\lambda_n\left(\frac{\sigma_{S_n}^2}{\gamma_n}|\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n|^2-\sigma_R^2\|\boldsymbol{G}_n^{1/2}\boldsymbol{w}_n\|^2-\sigma_D^2\right)$$
$$+\sum_{i=1}^{R}\mu_i\left(\sum_{n=1}^{N}\boldsymbol{w}_n^{\mathcal{H}}\boldsymbol{D}_n\boldsymbol{E}_i\boldsymbol{w}_n-\alpha P_i^{\max}\right), \tag{5.25}$$

where $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_n]^{\mathcal{T}}$, and λ_n's are the Lagrangian multipliers associated with the SNR constraints at the destinations. Clearly, by replacing $\lambda_n=\nu_n/\varepsilon_n$, the Lagrangian of the convex form is the same as the Lagrangian of the nonconvex form in (5.25). Thus, strong duality also holds for original nonconvex problem (5.21). □

Because of strong duality as stated in Proposition 3, the optimal value of Problem (5.21) can be found by its dual problem. In the next section, the dual problem of (5.21) is investigated in detail. The solution of the dual problem will then reveal both the structure of the original problem's solution and the algorithm to solve it.

5.3.1 Beamforming Duality

Denote $\boldsymbol{Q}=\mathrm{diag}(\mu_1,\ldots,\mu_R)$ and $\boldsymbol{P}=\mathrm{diag}(P_1^{\max},\ldots,P_R^{\max})$. Rearrange the Lagrangian $\mathcal{L}(\alpha,\boldsymbol{w}_n,\boldsymbol{\lambda},\boldsymbol{\mu})$ in (5.25), one has

$$\mathcal{L}\left(\alpha,\boldsymbol{w}_n,\boldsymbol{\lambda},\boldsymbol{Q}\right)=\sum_{n=1}^{N}\lambda_n\sigma_D^2+\sum_{n=1}^{N}\mathcal{L}_n(\boldsymbol{w}_n,\lambda_n,\boldsymbol{Q})-\alpha\left[\mathrm{tr}(\boldsymbol{QP})-\mathrm{tr}(\boldsymbol{P})\right], \tag{5.26}$$

where

$$\mathcal{L}_n(\boldsymbol{w}_n, \lambda_n, \boldsymbol{Q}) = \boldsymbol{w}_n^{\mathcal{H}} \left(\boldsymbol{D}_n \boldsymbol{Q} - \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}} + \lambda_n \sigma_R^2 \boldsymbol{G}_n \right) \boldsymbol{w}_n,$$

which only depends on $\boldsymbol{w}_n$, λ_n, and $\boldsymbol{Q}$. By dual decomposition [52], via the Lagrangian, the sum relay power minimization with per-relay power constraints is effectively decoupled into a summation of N smaller problems.

The dual function of (5.26) is established as

$$\begin{aligned} g(\boldsymbol{Q}, \boldsymbol{\lambda}) &= \inf_{\alpha, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_N} \mathcal{L}\Big(\alpha, \boldsymbol{w}_n, \boldsymbol{\lambda}, \boldsymbol{Q}\Big) \\ &= \sum_{n=1}^{N} \lambda_n \sigma_D^2 - \inf_{\alpha} \Big\{ \alpha \left[\mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) - \mathrm{tr}(\boldsymbol{P}) \right] \Big\} + \sum_{n=1}^{N} \inf_{\boldsymbol{w}_n} \mathcal{L}_n(\boldsymbol{w}_n, \lambda_n, \boldsymbol{Q}). \end{aligned}$$

It is clear that if $\boldsymbol{D}_n \boldsymbol{Q} - \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}} + \lambda_n \sigma_R^2 \boldsymbol{G}_n$ is not a positive semidefinite matrix, there exists $\boldsymbol{w}_n$ to make $\mathcal{L}_n = \boldsymbol{w}_n^{\mathcal{H}} \Big(\boldsymbol{D}_n \boldsymbol{Q} - \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}} + \lambda_n \sigma_R^2 \boldsymbol{G}_n \Big) \boldsymbol{w}_n = -\infty$. Similarly, if $\mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) - \mathrm{tr}(\boldsymbol{P}) > 0$, it is possible to find $\alpha > 0$ to make $-\alpha[\mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) - \mathrm{tr}(\boldsymbol{P})] = -\infty$. Thus, the dual problem is stated as

$$\begin{aligned} \underset{\boldsymbol{Q}}{\text{maximize}} \quad & \max_{\boldsymbol{\lambda}} \ \sum_{n=1}^{N} \lambda_n \sigma_D^2 \\ \text{subject to} \quad & \boldsymbol{D}_n \boldsymbol{Q} + \lambda_n \sigma_R^2 \boldsymbol{G}_n \succeq \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}}, \ n = 1, \ldots, N \\ & \mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) \leq \mathrm{tr}(\boldsymbol{P}) \\ & \boldsymbol{Q} \text{ is diagonal}, \ \boldsymbol{Q} \succeq \boldsymbol{0}. \end{aligned} \tag{5.27}$$

Clearly, this problem is divided into a two levels of computation. For the outer problem, one has

$$\begin{aligned} \underset{\boldsymbol{Q}}{\text{maximize}} \quad & f(\boldsymbol{Q}) \\ \text{subject to} \quad & \mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) \leq \mathrm{tr}(\boldsymbol{P}) \\ & \boldsymbol{Q} \text{ is diagonal}, \ \boldsymbol{Q} \succeq \boldsymbol{0}. \end{aligned} \tag{5.28}$$

where $f(\boldsymbol{Q}) = \max_{\boldsymbol{\lambda}} \sum_{n=1}^{N} \lambda_n \sigma_D^2$. For the inner problem, one can decompose the computation of $f(\boldsymbol{Q})$ into multiple problems as $f(\boldsymbol{Q}) = \sum_{n=1}^{N} \lambda_n^{\star} \sigma_D^2$ and $\lambda_n^{\star} \sigma_D^2$ is

obtained from

$$\begin{aligned} \underset{\lambda_n}{\text{maximize}} \quad & \lambda_n \sigma_D^2 \\ \text{subject to} \quad & \boldsymbol{D}_n \boldsymbol{Q} + \lambda_n \sigma_R^2 \boldsymbol{G}_n \succeq \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}}. \end{aligned} \tag{5.29}$$

In the next section, an interpretation via a virtual single-input multiple-output (SIMO) uplink channel shows that the dual problem (5.27) is equivalent to the following minimax problem:

$$\begin{aligned} \underset{\boldsymbol{Q}}{\text{maximize}} \quad & \min_{\boldsymbol{\lambda}, \hat{\boldsymbol{w}}_n} \sum_{n=1}^{N} \lambda_n \sigma_D^2 \\ \text{subject to} \quad & \frac{\lambda_n \sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \hat{\boldsymbol{w}}_n|^2}{\hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q} \hat{\boldsymbol{w}}_n + \lambda_n \sigma_R^2 \hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{G}_n \hat{\boldsymbol{w}}_n} \geq \gamma_n, \ n = 1, \ldots, N \\ & \operatorname{tr}(\boldsymbol{Q}\boldsymbol{P}) \leq \operatorname{tr}(\boldsymbol{P}) \\ & \boldsymbol{Q} \text{ is diagonal}, \ \boldsymbol{Q} \succeq \boldsymbol{0}, \end{aligned} \tag{5.30}$$

where $\hat{\boldsymbol{w}}_n^{\mathcal{H}}$ is interpreted as the receive beamforming vector of the virtual uplink channel for user-n. More precisely, the subproblem in (5.29) is shown to be equivalent to

$$\begin{aligned} \underset{\lambda_n, \hat{\boldsymbol{w}}_n}{\text{minimize}} \quad & \lambda_n \sigma_D^2 \\ \text{subject to} \quad & \frac{\lambda_n \sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \hat{\boldsymbol{w}}_n|^2}{\hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q} \hat{\boldsymbol{w}}_n + \lambda_n \sigma_R^2 \hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{G}_n \hat{\boldsymbol{w}}_n} \geq \gamma_n. \end{aligned} \tag{5.31}$$

Since strong duality holds for the original optimization problem in (5.21), at optimum, one has

$$\alpha^\star \sum_{i=1}^{R} P_i^{\max} = \sum_{n=1}^{N} \lambda_n^\star \sigma_D^2, \tag{5.32}$$

where $\lambda_n^\star$ is obtained from subproblem (5.31) for the virtual uplink channel.

5.3.2 An Interpretation via a Virtual Uplink Channel

In point-to-point multiuser communications, it is widely known that the optimal beamforming design for the downlink MIMO channel can be found via its equivalent uplink channel, which is much easier to handle. This property is known as uplink-downlink duality [51, 53–55]. Inspired by the uplink-downlink duality property of

the MIMO channel, this section introduces the concept of a virtual uplink channel, and uses it to solve the optimal distributed beamforming design in the multi-relay network.

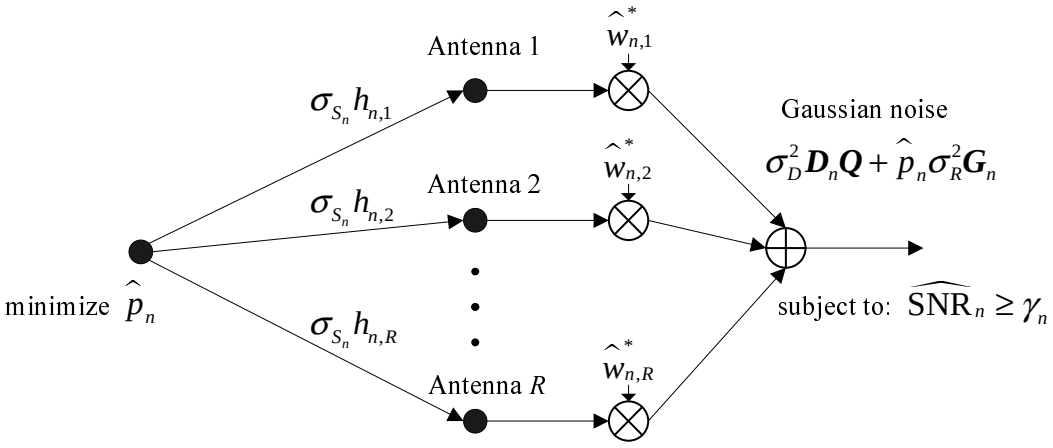

Figure 5.4 Block diagram of a virtual SIMO uplink channel.

Consider a virtual single-input multiple-output (SIMO) uplink channel as illustrated in Figure 5.4. Here a single-antenna transmitter with power $\hat{p}_n$ wants to communicate with an R-antenna receiver. The channel is modeled as $\sigma_{S_n}\boldsymbol{h}_n \in \mathbb{C}^{1\times R}$. The effective additive Gaussian noise at the receiver has the following covariance: $\sigma_D^2\boldsymbol{D}_n\boldsymbol{Q}+\hat{p}_n\sigma_R^2\boldsymbol{G}_n$. One can interpret $\sigma_D^2\boldsymbol{D}_n\boldsymbol{Q}$ as the added noise at the receiver and $\hat{p}_n\sigma_R^2\hat{\boldsymbol{G}}_n$ as the noise induced by the transmitter, which depends on the transmitted power $\hat{p}_n$. Now, it is of interest to find the optimal receive beamformer at the receiver and the minimal transmitted power $\hat{p}_n$ at the transmitter in order to obtain a certain target SNR of the virtual uplink channel's receiving end.

Let $\hat{\boldsymbol{w}}_n^{\mathcal{H}} = [\hat{w}_{n,1}^*, \ldots, \hat{w}_{n,R}^*]^{\mathcal{T}}$ be the receive beamforming vector. The SNR at the receiver can be expressed as

$$\widehat{\mathrm{SNR}}_n = \frac{\hat{p}_n\sigma_{S_n}^2|\boldsymbol{h}_n^{\mathcal{H}}\hat{\boldsymbol{w}}_n|^2}{\sigma_D^2\hat{\boldsymbol{w}}_n^{\mathcal{H}}\boldsymbol{D}_n\boldsymbol{Q}\hat{\boldsymbol{w}}_n + \hat{p}_n\sigma_R^2\hat{\boldsymbol{w}}_n^{\mathcal{H}}\boldsymbol{G}_n\hat{\boldsymbol{w}}_n}. \tag{5.33}$$

To maximize the above SNR, using the Rayleigh-Ritz theorem (see Appendix C), the optimal receive beamformer is

$$\hat{\boldsymbol{w}}_n = \left(\sigma_D^2\boldsymbol{D}_n\boldsymbol{Q} + \hat{p}_n\sigma_R^2\boldsymbol{G}_n\right)^{-1}\boldsymbol{h}_n. \tag{5.34}$$

Given a specific value of the transmit power $\hat{p}_n$, the weight of the optimal receive beamformer at the ith receive antenna only depends on the channel connected to itself, and is given by

$$\hat{w}_{n,i} = \frac{f_{n,i}^* g_{i,n}^*}{(\sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2)\sigma_D^2 \mu_n + \hat{p}_n \sigma_R^2 |g_{i,n}|^2}. \tag{5.35}$$

With the optimal receive beamformer, the constraint on the SNR at the receiver, $\widehat{\mathrm{SNR}}_n \geq \gamma_n$, is now equivalent to

$$\begin{aligned} & \frac{\hat{p}_n \sigma_{S_n}^2}{\gamma_n} |\boldsymbol{h}_n^{\mathcal{H}} \hat{\boldsymbol{w}}_n|^2 \geq \hat{\boldsymbol{w}}_n^{\mathcal{H}} \left(\sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q} + \hat{p}_n \sigma_R^2 \boldsymbol{G}_n \right) \hat{\boldsymbol{w}}_n \\ \Leftrightarrow \quad & \frac{\hat{p}_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n^{\mathcal{H}} \left(\sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q} + \hat{p}_n \sigma_R^2 \boldsymbol{G}_n \right)^{-1} \boldsymbol{h}_n \geq 1. \end{aligned} \tag{5.36}$$

Under the virtual uplink channel consideration with optimal receive beamformer in (5.34), the next task is to determine the minimal uplink transmitted power $\hat{p}_n$, given a target SNR at the receiver. This problem is stated as

$$\begin{aligned} \underset{\hat{p}_n}{\text{minimize}} \quad & \hat{p}_n \\ \text{subject to} \quad & \frac{\hat{p}_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n^{\mathcal{H}} \left(\sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q} + \hat{p}_n \sigma_R^2 \boldsymbol{G}_n \right)^{-1} \boldsymbol{h}_n \geq 1. \end{aligned} \tag{5.37}$$

Obviously, if the inequality is reversed, the optimization problem can be restated as a maximization problem as follows:

$$\begin{aligned} \underset{\hat{p}_n}{\text{maximize}} \quad & \hat{p}_n \\ \text{subject to} \quad & \frac{\hat{p}_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n^{\mathcal{H}} \left(\sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q} + \hat{p}_n \sigma_R^2 \boldsymbol{G}_n \right)^{-1} \boldsymbol{h}_n \leq 1. \end{aligned} \tag{5.38}$$

Apply the result in Lemma 1 of [51], one has

$$\begin{aligned} & \frac{\hat{p}_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n^{\mathcal{H}} \left(\sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q} + \hat{p}_n \sigma_R^2 \boldsymbol{G}_n \right)^{-1} \boldsymbol{h}_n \leq 1 \\ \Leftrightarrow \quad & \sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q} + \hat{p}_n \sigma_R^2 \boldsymbol{G}_n \succeq \frac{\hat{p}_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}}. \end{aligned}$$

Then, the optimization problem (5.38) will be exactly the same as the dual problem of the distributed beamforming problem in (5.29), if one sets $\hat{p}_n = \lambda_n \sigma_D^2$. Thus, the

solution of the dual problem in (5.29) given a fixed dual variable $\boldsymbol{Q}$ can be obtained from the virtual uplink channel problem (5.37). Note that (5.37) is equivalent to

$$\begin{aligned} \underset{\hat{p}_n}{\text{minimize}} \quad & \hat{p}_n \\ \text{subject to} \quad & \sum_{i=1}^{R} \frac{\hat{p}_n a_{n,i}}{b_{n,i}\mu_i + \hat{p}_n} \geq \gamma_n. \end{aligned} \tag{5.39}$$

Thus, the optimal value $\hat{p}_n^\star$ can be obtained from the simple fixed point iteration, as presented in Section 5.2. Moreover, with $\lambda_n = \hat{p}_n/\sigma_D^2$, the fixed point iteration

$$\lambda_n^{(t+1)} = \frac{\gamma_n}{\sigma_D^2 \sum_{i=1}^{R} \frac{a_{n,i}}{b_{n,i}\mu_i + \sigma_D^2 \lambda_n^{(t)}}}, \tag{5.40}$$

will surely converge to the optimal value $\lambda_n^\star$, which is the optimal value of Problem (5.29). Having known the optimal receive beamformer of the virtual uplink channel $\hat{\boldsymbol{w}}_n$, the optimal distributed beamformer $\boldsymbol{w}_n$ for the nth user can be determined by exploiting the relation between the two beamformers in the next lemma.

Lemma 8. *The optimal distributed beamforming vector $\boldsymbol{w}_n$ in the multiuser beamforming problem is a scaled version of $\hat{\boldsymbol{w}}_n$, i.e., $\boldsymbol{w}_n = \sqrt{\zeta_n}\hat{\boldsymbol{w}}_n$.*

Proof. From Karush-Kuhn-Tucker (KKT) condition [37], the gradient of Lagrangian $\mathcal{L}_n(\boldsymbol{Q}, \lambda_n, \boldsymbol{w}_n)$ vanishes at the optimum of $\boldsymbol{w}_n$, i.e.,

$$\frac{\partial \mathcal{L}_n(\boldsymbol{Q}, \lambda_n, \boldsymbol{w}_n)}{\partial \boldsymbol{w}_n^*} = \left(\boldsymbol{D}_n \boldsymbol{Q} - \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}} + \lambda_n \sigma_R^2 \boldsymbol{G}_n \right) \boldsymbol{w}_n = \boldsymbol{0}. \tag{5.41}$$

Thus,

$$\begin{aligned} \boldsymbol{w}_n &= \left(\boldsymbol{D}_n \boldsymbol{Q} + \lambda_n \sigma_R^2 \boldsymbol{G}_n \right)^{-1} \frac{\lambda_n \sigma_{S_n}^2}{\gamma_n} \boldsymbol{h}_n \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n \\ &= \frac{\lambda_n \sigma_D^2 \sigma_{S_n}^2 \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n}{\gamma_n} \hat{\boldsymbol{w}}_n, \end{aligned}$$

which suggests $\sqrt{\zeta_n} = (\lambda_n \sigma_D^2 \sigma_{S_n}^2/\gamma_n) \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n$. However, this expression of ζ_n still depends on $\boldsymbol{w}_n$. The next step is to determine the value ζ_n that is independent of $\boldsymbol{w}_n$. As the SNR constraints in (5.21) is met with equality at optimum, one has

$$\frac{\sigma_{S_n}^2}{\gamma_n} |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2 = \sigma_R^2 \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{G}_n \boldsymbol{w}_n + \sigma_D^2.$$

Substituting $\boldsymbol{w}_n = \sqrt{\zeta_n}\hat{\boldsymbol{w}}_n$ into the above expression yields

$$\frac{\zeta_n \sigma_{S_n}^2}{\gamma_n} |\boldsymbol{h}_n^{\mathcal{H}} \hat{\boldsymbol{w}}_n|^2 = \zeta_n \sigma_R^2 \hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{G}_n \hat{\boldsymbol{w}}_n + \sigma_D^2.$$

Therefore,

$$\zeta_n = \frac{\gamma_n \sigma_D^2}{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \hat{\boldsymbol{w}}_n|^2 - \gamma_n \sigma_R^2 \hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{G}_n \hat{\boldsymbol{w}}_n} \tag{5.42}$$

□

5.3.3 Numerical Algorithm

In the previous section, the dual problem of the optimal distributed beamforming design with per-relay power constraints (Problem (5.21)) was established. The motivation of presenting the dual problem is that it is generally easier to solve than the original problem, as well as it reveals the structure of the solution. This section continues with the dual problem in (5.27) and provides an efficient numerical algorithm to solve it.

It was shown in Section 5.3.1 that solving the dual problem involves two levels of computations: the outer maximization problem (5.28) and the inner maximization problem (5.29). In Section 5.3.2, via the proposal of a virtual uplink channel, the inner problem was solved by the simple fixed-point iteration algorithm. On the other hand, the outer maximization of $f(\boldsymbol{Q})$ can be solved by the subgradient projection method, as presented next.

Proposition 4. *The function $f(\boldsymbol{Q})$ is concave in $\boldsymbol{Q}$, and its subgradient is given by* $\text{diag}\left(\sum_{n=1}^{N} \boldsymbol{w}_n \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n\right)$*, where $\boldsymbol{w}_n$ is the optimal distributed beamforming vector obtained from Section 5.3.2.*

Proof. The proof of this proposition follows the proof of Proposition 3 in [51] for the point-to-point multiuser downlink beamforming problem. Since $f(\boldsymbol{Q})$ is the objective function of the dual problem, it is a concave function by nature [37].

Now look back at the Lagrangian of the distributed beamforming problem in

(5.25). For a fixed $\boldsymbol{Q}$, one has

$$\begin{aligned} f(\boldsymbol{Q}) &= \min_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_n} \min_{\alpha} \mathcal{L}\left(\alpha, \boldsymbol{w}_n, \boldsymbol{\lambda}, \boldsymbol{Q}\right) \\ &= \min_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_n} \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q} \boldsymbol{w}_n \\ &\text{subject to } \frac{\sigma_{S_n}^2}{\gamma_n} |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2 \geq \sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2. \end{aligned} \tag{5.43}$$

Suppose that $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ are the two diagonal and positive semidefinite matrices. By the definition applied to the first order condition of concave function $f(\boldsymbol{Q})$ [37], the diagonal matrix $\boldsymbol{H}$ is the subgradient of $f(\boldsymbol{Q})$ at $\boldsymbol{Q}_1$ if $f(\boldsymbol{Q}_2) \leq f(\boldsymbol{Q}_1) + \mathrm{tr}\left[\boldsymbol{H}(\boldsymbol{Q}_2 - \boldsymbol{Q}_1)\right]$. Denote $\boldsymbol{w}_{n,k}$ as the optimal distributed beamformers for $f(\boldsymbol{Q}_k)$, $k = 1, 2$. Thus, one has

$$\begin{aligned} f(\boldsymbol{Q}_2) - f(\boldsymbol{Q}_1) &= \sum_{n=1}^{N} \boldsymbol{w}_{n,2}^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}_2 \boldsymbol{w}_{n,2} - \sum_{n=1}^{N} \boldsymbol{w}_{n,1}^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}_1 \boldsymbol{w}_{n,1} \\ &\leq \sum_{n=1}^{N} \boldsymbol{w}_{n,1}^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}_2 \boldsymbol{w}_{n,1} - \sum_{n=1}^{N} \boldsymbol{w}_{n,1}^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}_1 \boldsymbol{w}_{n,1} \\ &= \mathrm{tr}\left(\mathrm{diag}\left(\sum_{n=1}^{N} \boldsymbol{w}_{n,1} \boldsymbol{w}_{n,1}^{\mathcal{H}} \boldsymbol{D}_n\right)(\boldsymbol{Q}_2 - \boldsymbol{Q}_1)\right), \end{aligned}$$

where the inequality follows from the fact that $\boldsymbol{w}_{n,2}$ is optimal for $f(\boldsymbol{Q}_2)$, then $\sum_{n=1}^{N} \boldsymbol{w}_{n,2}^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}_2 \boldsymbol{w}_{n,2} \leq \sum_{n=1}^{N} \boldsymbol{w}_{n,1}^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}_2 \boldsymbol{w}_{n,1}$, and the last equality follows because $\boldsymbol{D}_n$, $\boldsymbol{Q}_2$, and $\boldsymbol{Q}_1$ are all diagonal.

Therefore, the subgradient of $\boldsymbol{Q}$ is $\mathrm{diag}\left(\sum_{n=1}^{N} \boldsymbol{w}_n \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n\right)$. In particular, the subgradient of μ_i at the ith relay is $\sum_{n=1}^{N} |w_{n,i}|^2 [\boldsymbol{D}_n]_{ii}$. □

Having derived the subgradient of $f(\boldsymbol{Q})$, $\boldsymbol{Q}$ is then updated by applying the Euclidean projection $\mathcal{P}_{\mathcal{S}_Q}$ of the subgradient of $f(\boldsymbol{Q})$ on the constraint set $\mathcal{S}_Q = \{\boldsymbol{Q} : \mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) \leq \mathrm{tr}(\boldsymbol{P}), \boldsymbol{Q} \succeq \boldsymbol{0}\}$, i.e.,[5]

$$\boldsymbol{Q}^{(t+1)} = \mathcal{P}_{\mathcal{S}_Q}\left\{\boldsymbol{Q}^{(t)} + a_t \mathrm{diag}\left\{\sum_{n=1}^{N} \boldsymbol{w}_n \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n\right\}\right\}, \tag{5.44}$$

[5]Detailed discussion on the Euclidean projection is in Appendix B.

where a_t is an appropriate step size. The subgradient projection method is guaranteed to converge to the global optimum of $f(\boldsymbol{Q})$ [37]. The proposed algorithm with the property of distributed implementation is summarized as follows.

1. **Initialize** $\boldsymbol{Q}^{(t)}$. Set $t = 1$.

2. **Repeat**: fix $\boldsymbol{Q}^{(t)}$, then the relays transmit $\boldsymbol{Q}^{(t)}$ to every destination. Each destination then solves the fixed-point iteration in (5.40) to determine the required power $\lambda_n \sigma_D^2$ for its corresponding virtual uplink channel. The optimal receive beamformer $\hat{\boldsymbol{w}}_n$ and the scaling factor ζ_n are the determined by the nth destination.

3. The nth destination broadcasts λ_n and ζ_n back to relays. The ith relay calculates the beamforming coefficients $w_{1i}, w_{2i}, \ldots, w_{Ni}$ with local information pertaining to the relay as

$$w_{n,i} = \frac{\sqrt{\zeta_n} f_{n,i}^* g_{i,n}^*}{\sigma_D^2 \left[(\sigma_{S_n}^2 |f_{n,i}|^2 + \sigma_R^2)\mu_n + \lambda_n \sigma_R^2 |g_{i,n}|^2 \right]}. \tag{5.45}$$

4. The relays cooperate with each other to update $\boldsymbol{Q}^{(t)}$ as in (5.44).

5. Set $t = t + 1$ and **return** to Step 2 **until** convergence.

5.4 Simulation Results

This section presents the numerical results on the power consumptions at the relays of a multiuser relay-assisted network with and without per-relay power constraints. Also presented are the convergence plots of the proposed iterative algorithms. The network being considered is equipped with 4 relays. The number of users to be served by the network is 3 users. The source power is set at 10 for all the users' sources in all the simulations. The noise variances σ_R^2 and σ_D^2 are set to unity. Flat Rayleigh fading is assumed in all the channels, where each $S \to R$ and $R \to D$ channel coefficients are assumed to be i.i.d. $\mathcal{CN}(0, 1)$. When the per-relay power constraints

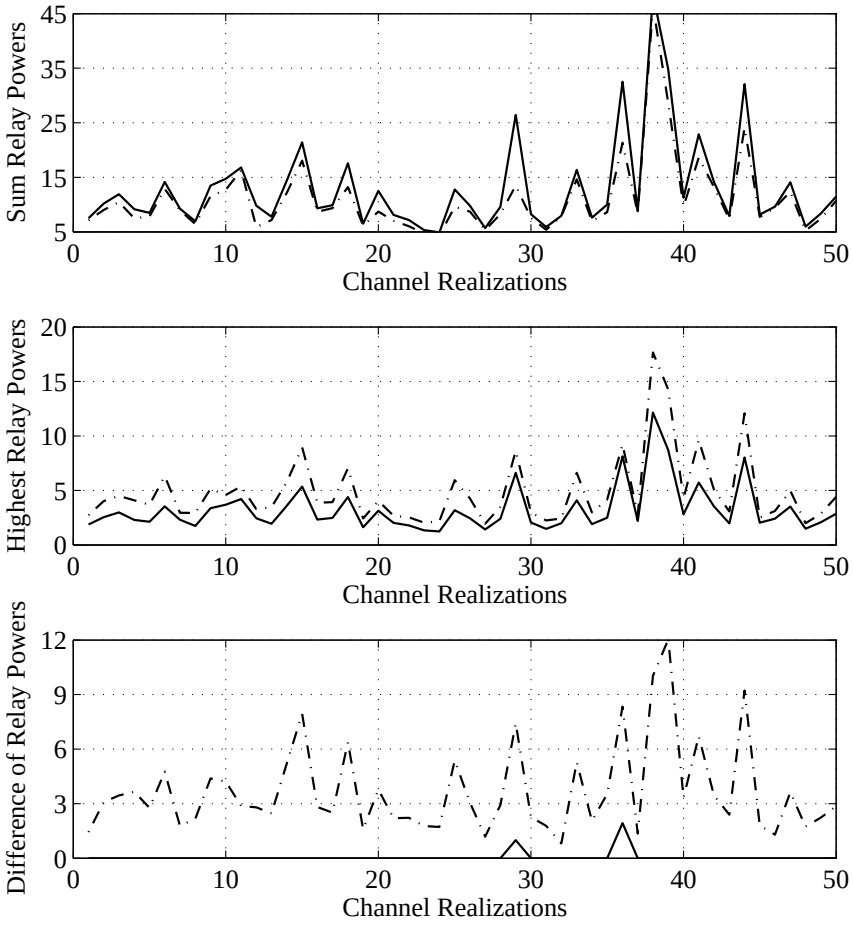

Figure 5.5 Power consumptions at the relays over 50 channel realizations with different power constraints: with per-relay power constraints (solid lines), without per-relay power constraints ("dash-dot" lines).

are imposed, the maximum per-relay power is set at 10. The target SNR γ_n is set at 5 (7 dB) for all the destinations.

Figure 5.5 illustrates the power consumptions at the relays for 50 different channel realizations. At each channel realization, the sum relay powers, the highest relay power level of the 4 relays, and the difference between the highest and lowest relay power levels of the 4 relays are plotted and compared between the two relaying strategies: with and without per-relay power constraints. As can be seen from the

figure, imposing the per-relay power constraints does increase the sum relay power, compared with the optimal strategy that does not impose the constraints. However, the main advantage of applying the per-relay power constraints is that it balances the power consumption at the relays and does not overuse any of them. Consequently, the highest relay power level of the 4 relays with the per-relay power constraints is always smaller than that without the constraints. In addition, all the relays transmit at the same power level almost all the time when the constraints are applied; whereas the difference between the highest and lowest power levels are quite significant without the constraints.

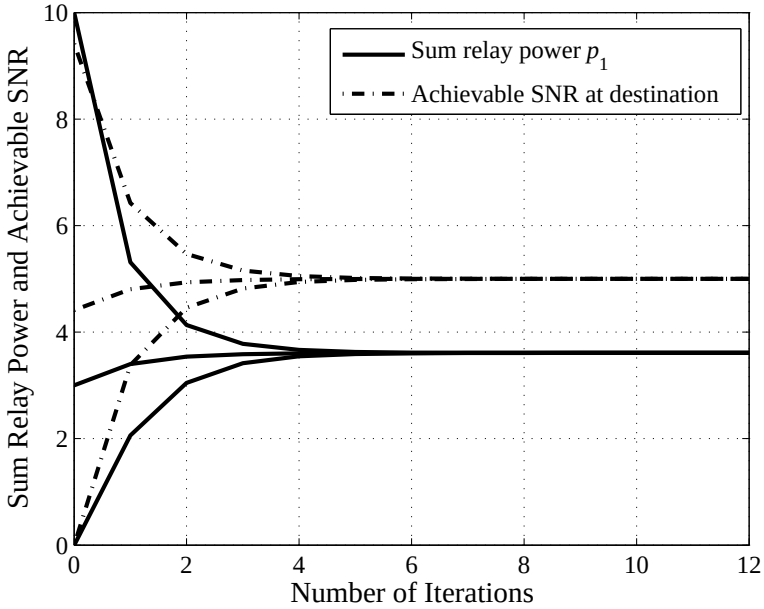

Figure 5.6 Convergence of the iterative fixed point algorithm (5.19) with different starting points and the achievable SNR at user-1's destination after each iteration.

The convergence of the proposed algorithms is illustrated in Figures 5.6 and 5.7. Figure 5.6 plots the evolution of the sum relay power allocated for user-1, p_1, and the corresponding SNR_1 after each iteration by the iterative fixed point algorithm (5.19). It can be seen that the algorithm converges very quickly after only a few iterations to the optimal $p_1^\star$ from various arbitrary starting points, while the corresponding SNR also converges to its target value $\gamma_1 = 5$. Figure 5.7 displays the convergence of

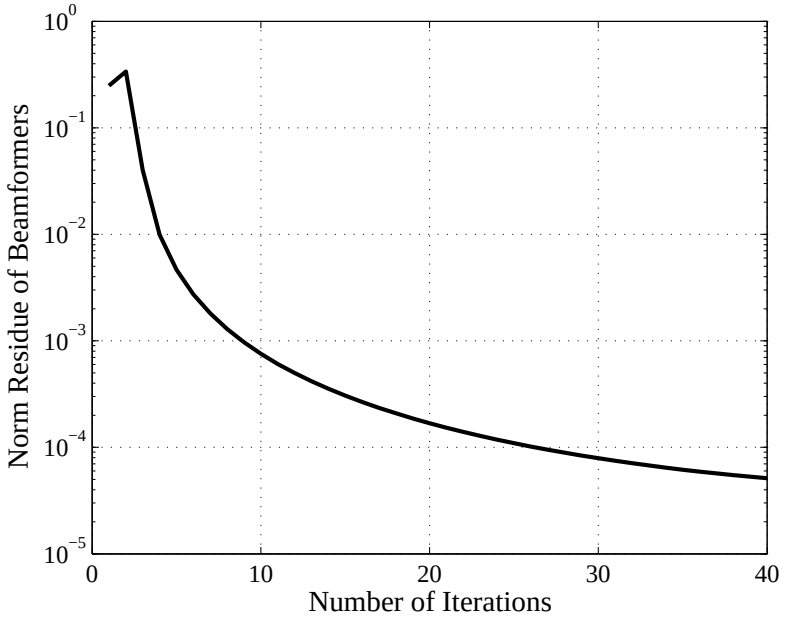

Figure 5.7 Convergence of the proposed algorithm in finding the optimal distributed beamformers with per-relay power constraints.

the proposed iterative algorithm in Section 5.3.3 in finding the optimal distributed beamformers $\boldsymbol{w}_n^\star$ with per-relay power constraints. The step-size $a_t = 1/t$ is used for the subgradient update of the iterative algorithm. The summation $\sum_{n=1}^{N} \|\boldsymbol{w}_n - \boldsymbol{w}_n^\star\|$, which is the norm residue of the beamformers, plotted after each iteration clearly shows the convergence of the proposed algorithm. Numerous simulations also show that the proposed algorithm converges in a small fraction of the running time required by the `cvx` package to converge.

5.5 Summary and Future Works

This chapter has studied the optimal distributed beamforming design to minimize the total relay power with guaranteed QoS in a wireless multiuser multi-relay network. It was assumed that each user $(S_n - D_n)$ communicates in orthogonal channels. Considered were the optimization problems with and without per-relay power constraints. While it was shown that these two optimization problems are convex SOCP and can be solved effectively by any conic software package, the chapter also

explored simple and fast iterative algorithms to solve the problems. Feasibility of the two optimization problems were also thoroughly studied in the chapter.

It should be emphasized that this chapter studies the relay network where $S-D$ pairs operate over orthogonal channels, i.e., no inter-user interference is incurred. When the orthogonality assumption is relaxed, several new problems in optimal distributed beamforming designs to minimize the sum relay power with guaranteed QoS may arise. In particular, there are 3 other scenarios that might be considered: (i) nonorthogonal $S \to R$ and nonorthogonal $R \to D$ channels, (ii) orthogonal $S \to R$ and nonorthogonal $R \to D$ channels, and (iii) nonorthogonal $S \to R$ and orthogonal $R \to D$ channels. Unlike the scenario considered in this chapter (orthogonal $S \to R$ and $R \to D$ channels), the received signal at each destination is subject to interference from other users in the 3 mentioned scenarios. The QoS design criterion is now defined as the target signal-to-interference-plus-noise ratios (SINRs) at the destinations.

With the network in scenario (i), the optimization problem of the optimal distributed beamforming designs can be formulated as a nonconvex quadratically constrained quadratic program (QCQP). Through relaxation techniques, including semidefinite programming (SDP) and duality relaxations, the problem can be efficiently solved. We report this optimization problem and its solution in a conference paper [44].

With the networks in scenarios (ii) and (iii), although no works have been carried out to date, initial derivations show that the corresponding optimization problems associated with the two scenarios can be formulated as convex SOCP. Hence, these problems might be effectively solved by convex techniques. The interpretation via virtual uplink channels may also be applicable in devising simple iterative algorithms to solve these two problems.

6. SNR Maximization and Distributed Beamforming in a Multiuser Multi-relay Network

In Chapter 5, the optimal designs of distributed beamforming in a multiuser multi-relay network with guaranteed QoS were studied. While guaranteed QoS is important to maintain the connectivity for the users, it is also critical to perform a strict power control at the relays due to the following reasons. First, the relays, which can be any nodes in the network, are often operating on a small power budget. Power conservation at the relays helps to extend the lifetime of the relays and the networks as well. Second, if one user's channel is in deep fades, i.e., the channel strength is low, the relays may consume an excessive amount of power to maintain its connectivity, and thus may induce significant interference to adjacent networks. As a result, the distributed beamforming strategy needs to adjust accordingly to the constrained power consumption of the relays. This chapter studies the optimal distributed beamforming designs to jointly maximize the SNR margin at the destinations subject to two different types of power constraints: sum relay power constraint and per-relay power constraints. Although the two problems can be effectively solved by the bisection method via SOCP feasibility problem, this chapter also proposes two simple and fast converging iterative algorithms to directly solve the problems.

[0]The contributions in this chapter are also presented in [56, 57].

6.1 Sum-Power Constraints

This section considers the distributed beamforming strategy to jointly maximize the minimal SNR margin subject to a sum relay power constraint. Under such a constraint, the relays are allowed to share a common power pool, although they do not necessarily share their received signals from the sources. The problem is stated as

$$\begin{aligned} \underset{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{maximize}} \quad & \min_n \frac{\text{SNR}_n}{\gamma_n} \\ \text{subject to} \quad & P_{\text{relay}} \leq P_{\text{relay}}^{\max}. \end{aligned} \tag{6.1}$$

Here, the sum relay power constraint is a firm system restriction. Even so, the problem is always feasible as it is always possible to scale $\boldsymbol{w}_n$ down to meet the sum relay power constraint. The parameter γ_n is now interpreted as the weight for user-n's SNR. Recast the problem by introducing the auxiliary variable τ, denoted as the SNR margin, as

$$\begin{aligned} \underset{\tau,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{maximize}} \quad & \tau \\ \text{subject to} \quad & \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2} \geq \tau \gamma_n, \ n = 1, \ldots, N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{w}_n \leq P_{\text{relay}}^{\max}. \end{aligned} \tag{6.2}$$

Under this reformulation, τ is also a variable, not a parameter. The SNR constraint is no longer convex [13]. As a result, Problem (6.2) is not a convex problem. However, with a fixed value of τ, the problem can be formulated as a convex feasibility problem and is readily solved by the bisection method [37], as presented in the next section.

6.1.1 Bisection Method

Define $\tau^\star$ as the maximum attained value of τ. For a specific target value of τ, the following SOCP feasibility problem is considered:

$$\begin{aligned} \text{find} \quad & \boldsymbol{w}_1, \ldots, \boldsymbol{w}_N \\ \text{subject to} \quad & \sqrt{\frac{\sigma_{S_n}^2}{\tau\gamma_n}}\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n \geq \left\| \begin{array}{c} \sigma_R \boldsymbol{G}_n^{1/2}\boldsymbol{w}_n \\ \sigma_D \end{array} \right\|, \ n = 1, \ldots, N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}}\boldsymbol{D}_n\boldsymbol{w}_n \leq P_{\text{relay}}^{\max}. \end{aligned} \tag{6.3}$$

If the problem is feasible, which means that $\tau < \tau^\star$, then it is possible to increase the target margin τ. Otherwise, if $\tau > \tau^\star$, the target margin should be reduced. The bisection method [37] is summarized as follows:

1. **Initialize** u and l as the upper and lower bounds of τ.
2. **Repeat**: $\tau = (u + l)/2$. Solve the feasibility problem (6.3).
3. **If** (6.3) is feasible, **then** set $l = \tau$, **else**, set $u = \tau$.
4. **Return** to Step 2 **until** $u - l < \epsilon$, where ϵ is a small positive value.

6.1.2 Convex Solution

In the previous section, an SOCP approach using bisection method is presented to solve the joint SNR margin maximization problem. However, such an approach is computationally expensive and unappealing, since it requires many iterations in the bisection method as well as a standard conic solution package. In this section, the result of Lemma 7 in Chapter 5 is applied to solve the problem in a convex fashion. In particular, problem (6.2) is recast with $\tau, p_1, \ldots, p_N$ as the variables, i.e.,

$$\mathcal{S}(P_{\text{relay}}^{\max}) = \begin{cases} \underset{p_1,\ldots,p_N,\tau}{\text{maximize}} & \tau \\ \text{subject to} & \tau\gamma_n - \sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n} \leq 0, \ n = 1, \ldots, N \\ & \sum_{n=1}^{N} p_n \leq P_{\text{relay}}^{\max}, \end{cases} \tag{6.4}$$

where $\mathcal{S}(P_{\text{relay}}^{\max})$ denotes the maximum achievable SNR margin τ for a given sum relay power $P_{\text{relay}}^{\max}$, and also denotes the optimization problem (6.4) itself. The problem in (6.4) is a strictly convex optimization problem due to the convexity of all the constraints and the objective function. It is then can be solved by any standard convex optimization algorithm. On the other hand, the connection between the power minimization problem $\mathcal{P}_n(\gamma_n)$ in (5.17) and the joint SNR margin maximization problem $\mathcal{S}(P_{\text{relay}}^{\max})$ in (6.4) can be exploited to solve the joint SNR margin maximization problem. In fact, the two problems are inverse of each other. Mathematically, this relation is given in the following lemma.

Lemma 9. *The joint SNR margin maximization problem* (6.4) *and the power minimization problem in* (5.17) *are inverse problems:*

$$\tau = \mathcal{S}\left(\sum_{n=1}^{N} \mathcal{P}_n(\tau\gamma_n)\right). \tag{6.5}$$

$$P_{\text{relay}}^{\max} = \sum_{n=1}^{N} \mathcal{P}_n\left(\gamma_n \mathcal{S}(P_{\text{relay}}^{\max})\right). \tag{6.6}$$

Proof. This lemma is proved by contradiction and by the monotonicity of the function $\sum_{i=1}^{R} \frac{a_{n,i}p_n}{b_{n,i}+p_n}$. Begin with (6.5), suppose that $p_n^\star$ is the optimal value and also the optimal argument of $\mathcal{P}_n(\tau\gamma_n)$, then $\sum_{i=1}^{R} \frac{a_{n,i}p_n^\star}{b_{n,i}+p_n^\star} = \tau\gamma_n$. Also, let $\tilde{\tau}^\star$ and $\tilde{p}_n^\star, n = 1,\ldots,N$ be the optimal value and arguments of $\mathcal{S}\left(\sum_{n=1}^{N} p_n^\star\right)$. If $\tilde{\tau}^\star < \tau$, there is a contradiction that $p_n^\star$'s are also feasible solution for $\mathcal{S}\left(\sum_{n=1}^{N} p_n^\star\right)$, and yet provide a higher objective value τ. On the other hand, if $\tilde{\tau}^\star > \tau$, there is a contradiction that $\tilde{p}_n^\star > p_n^\star$ to make $\tilde{\tau}^\star\gamma_n > \tau\gamma_n$, then the constraint $\sum_{n=1}^{N} \tilde{p}_n^\star \le \sum_{n=1}^{N} p_n^\star$ in $\mathcal{S}\left(\sum_{n=1}^{N} p_n^\star\right)$ cannot be true. The proof for (6.6) follows the same line. □

Using the results in Lemma 9, problem $\mathcal{S}(P_{\text{relay}}^{\max})$ in (6.4) can be solved by iteratively solving N problems $\mathcal{P}_n(\tau\gamma_n)$ for different values of τ until $\sum_{n=1}^{N} \mathcal{P}_n(\tau\gamma_n) = P_{\text{relay}}^{\max}$. Then the optimal arguments $p_n^\star$ of $\mathcal{P}_n(\tau^\star\gamma_n)$ are also optimal to $\mathcal{S}(P_{\text{relay}}^{\max})$. This process is summarized by the following bisection method:

1. **Initialize** u and l as the upper and lower bounds of τ.

2. **Repeat**: $\tau = (u + l)/2$. Solve N problems $\mathcal{P}_n(\tau\gamma_n)$ using the fixed point iteration (5.19), and take $P_{\text{relay}} = \sum_{n=1}^{N} \mathcal{P}_n(\tau\gamma_n)$.

3. **If** $P_{\text{relay}} < P_{\text{relay}}^{\max}$, **then** set $l = \tau$, **else**, set $u = \tau$.

4. **Return** to Step 2 **until** $P_{\text{relay}} = P_{\text{relay}}^{\max}$.

It should be pointed out that this algorithm still requires multiple iterations in the bisection methods, and multiple fixed point iterations in Step 2 to solve $\mathcal{P}_n(\tau\gamma_n)$ for each value of τ. In the next section, a modified fixed point iteration is proposed to directly find the optimal arguments $p_n^\star$'s of $\mathcal{S}(P_{\text{relay}}^{\max})$.

6.1.3 Modified Fixed Point Iteration for Finding $p_n^\star$

As presented in the previous section, the power minimization problem and the joint SNR margin maximization problem are inverse problems. Thus, the optimal solution $p_n^\star$ of the joint SNR margin maximization problem is also the optimal solution to the power minimization problem $\mathcal{P}_n(\tau^\star\gamma_n)$. Therefore, $p_n^\star$ must also satisfy the fixed point iteration (5.19) with γ_n replaced by $\tau^\star\gamma_n$. Unfortunately, $\tau^\star$ needs to be determined as well. However, the condition on optimality $\sum_{n=1}^{N} p_n^\star = P_{\text{relay}}^{\max}$ allows a modified fixed point iteration to overcome this difficulty, as follows. Let

$$\tilde{p}_n = \frac{\gamma_n}{\sum_{i=1}^{R} \frac{a_{n,i}}{b_{n,i}+p_n^{(t)}}}. \tag{6.7}$$

Then normalize the result such that the sum relay power is equal to the maximum allowable power:

$$p_n^{(t+1)} = \frac{P_{\text{relay}}^{\max}}{\sum_{l=1}^{N} \tilde{p}_l}. \tag{6.8}$$

The iteration gets back to step (6.7) until convergence. Numerous simulations show a rapid convergence rate of the modified fixed point iteration.

6.2 Per-Relay Power Constraints

With the same arguments as in Section 5.3, it may be desirable to have the power constraint at each relay. In this section, the joint SNR margin maximization problem

is investigated with a strict constraint on transmission power at each relay. This problem is stated as

$$\begin{aligned} \underset{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{maximize}} \quad & \min_n \frac{\text{SNR}_n}{\gamma_n} \\ \text{subject to} \quad & P_i \le P_i^{\max},\ i = 1,\ldots,R. \end{aligned} \tag{6.9}$$

By introducing the auxiliary variable τ, this problem is restated as

$$\begin{aligned} \underset{\tau,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{maximize}} \quad & \tau \\ \text{subject to} \quad & \frac{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n|^2}{\sigma_R^2 \|\boldsymbol{G}_n^{1/2} \boldsymbol{w}_n\|^2 + \sigma_D^2} \ge \tau\gamma_n,\ n = 1,\ldots,N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n \le P_i^{\max},\ i = 1,\ldots,R. \end{aligned} \tag{6.10}$$

6.2.1 Bisection Method

Like the optimization problem (6.2), problem (6.9) is not convex. However, with a fixed value of τ, the problem can be formulated as a convex feasibility problem. Thus, this problem can be solved by bisection method [37]. Define $\tau^\star$ as the maximum attained value of τ. For a specific value of τ, the following convex SOCP feasibility problem [37] is considered:

$$\begin{aligned} \text{find} \quad & \boldsymbol{w}_1,\ldots,\boldsymbol{w}_N \\ \text{subject to} \quad & \sqrt{\frac{\sigma_{S_n}^2}{\tau\gamma_n}} \boldsymbol{h}_n^{\mathcal{H}} \boldsymbol{w}_n \ge \left\| \begin{matrix} \sigma_R \boldsymbol{G}_n^{1/2} \boldsymbol{w}_n \\ \sigma_D \end{matrix} \right\|,\ n = 1,\ldots,N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n \le P_i^{\max},\ i = 1,\ldots,R. \end{aligned} \tag{6.11}$$

If the problem is feasible, then $\tau < \tau^\star$. Otherwise, $\tau > \tau^\star$. The bisection method is the same as the one in Section 6.1.1.

It should be pointed out that instead of solving the feasibility problem in (6.11), one can solve the total power minimization problem under per-relay power constraints

stated in Chapter 5:

$$\begin{aligned} \underset{\alpha,\boldsymbol{w}_1,\ldots,\boldsymbol{w}_N}{\text{minimize}} \quad & \alpha \sum_{i=1}^{R} P_i^{\max} \\ \text{subject to} \quad & \sqrt{\frac{\sigma_{S_n}^2}{\tau\gamma_n}}\boldsymbol{h}_n^{\mathcal{H}}\boldsymbol{w}_n \geq \left\| \begin{array}{c} \sigma_R \boldsymbol{G}_n^{1/2}\boldsymbol{w}_n \\ \sigma_D \end{array} \right\|, \ n = 1, \ldots, N \\ & \sum_{n=1}^{N} \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{E}_i \boldsymbol{w}_n \leq \alpha P_i^{\max}, \ i = 1, \ldots, R. \end{aligned} \tag{6.12}$$

The resultant optimal $\alpha^\star$ could be used to determine the feasibility of the distributed beamforming problem with the target SNR margin τ. More specifically, if the optimal $\alpha^\star > 1$, it means that at least one of the per-relay power constraints is violated, i.e., it is infeasible to meet the SNR constraints without compromising the per-relay power constraints. Thus, the target margin τ needs to be adjusted to a smaller value. Conversely, if $\alpha^\star < 1$, one can scale up the beamforming vector $\boldsymbol{w}_n$ to improve the target SNR margin τ without violating the per-relay power constraints. This suggests that at the optimal SNR margin target $\tau^\star$, $\alpha^\star = 1$. Thus, the aforementioned bisection method can be modified to solve Problem (6.12) until $\alpha^\star = 1$, instead of the feasibility problem in (6.11). The new method is equivalent to the bisection method proposed in Section 6.1.2, where one attempts to find the solution that makes $P_{\text{relay}} = P_{\text{relay}}^{\max}$.

6.2.2 An Iterative Algorithm for Finding $\boldsymbol{w}_n^\star$

It was shown in the previous section that the joint SNR margin maximization problem with per-relay power constraints can be solved by the bisection method. At each target SNR margin τ, one can solve either the SOCP feasibility problem (6.11) or the power minimization problem (6.12). Obviously, this method is very computationally expensive. In this section, by adapting the algorithm outlined in Section 5.3.3, a new iterative algorithm is proposed to directly solve the joint SNR margin maximization problem (6.9).

It should be emphasized that unlike the algorithm in Section 5.3.3 in finding the optimal beamforming design with a fixed target SNR at each destination, the SNR

at the nth destination $\tau\gamma_n$ is now a variable. More specifically, τ is an optimization variable, not a parameter, and has to be determined as well. At first, revisit the optimization problem (6.12). From Proposition 3, strong duality holds for the optimization problem (6.12), i.e.,

$$\alpha^\star \sum_{i=1}^{R} P_i^{\max} = \sum_{n=1}^{N} \lambda_n^\star \sigma_D^2. \tag{6.13}$$

Furthermore, the bisection method in Section 6.2.1 states that at the optimal value $\tau^\star$, $\alpha^\star = 1$, which also means that $\sum_{n=1}^{N} \lambda_n^\star \sigma_D^2 = \sum_{i=1}^{R} P_i^{\max}$ at optimum. This can be done by adjusting the fixed point iteration in step 2 of the algorithm in Section 5.3.3. In the algorithm presented here to jointly maximize the SNR margin in distributed beamforming with per-relay power constraints, the modified fixed point iteration is taken at Step 2:

1. **Initialize** $\boldsymbol{Q}^{(t)}$. Set $t = 1$.

2. **Repeat**: fix $\boldsymbol{Q}^{(t)}$, solve the fixed point λ_n by iterative function

$$\tilde{\lambda}_n = \frac{\gamma_n}{\sigma_D^2 \sum_{i=1}^{R} \frac{a_{n,i}}{b_{n,i}\mu_i + \sigma_D^2 \lambda_n}}, \tag{6.14}$$

   then normalize the result

$$\lambda_n = \frac{\tilde{\lambda}_n \sum_{i=1}^{R} P_i^{\max}}{\sigma_D^2 \sum_{l=1}^{N} \tilde{\lambda}_l} \tag{6.15}$$

   so that $\sum_{n=1}^{N} \lambda_n \sigma_D^2 = \sum_{i=1}^{R} P_i^{\max}$, then return to (6.14) until convergence.

3. Find the optimal receive beamformers of the virtual channels as

$$\hat{\boldsymbol{w}}_n = \left(\sigma_D^2 \boldsymbol{D}_n \boldsymbol{Q}^{(t)} + \lambda_n \sigma_D^2 \sigma_R^2 \boldsymbol{G}_n\right)^{-1} \boldsymbol{h}_n. \tag{6.16}$$

4. Determine the achievable SNR of the virtual uplink channel for each user:

$$\tilde{\gamma}_n = \frac{\lambda_n \sigma_{S_n}^2 |\hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{h}_n|^2}{\hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{D}_n \boldsymbol{Q}^{(t)} \hat{\boldsymbol{w}}_n + \lambda_n \sigma_R^2 \hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{G}_n \hat{\boldsymbol{w}}_n}. \tag{6.17}$$

5. Update the downlink beamformers $\boldsymbol{w}_n = \sqrt{\zeta_n} \hat{\boldsymbol{w}}_n$, where

$$\zeta_n = \frac{\tilde{\gamma}_n \sigma_D^2}{\sigma_{S_n}^2 |\boldsymbol{h}_n^{\mathcal{H}} \hat{\boldsymbol{w}}_n|^2 - \tilde{\gamma}_n \sigma_R^2 \hat{\boldsymbol{w}}_n^{\mathcal{H}} \boldsymbol{G}_n \hat{\boldsymbol{w}}_n}. \tag{6.18}$$

6. Update $\boldsymbol{Q}^{(t)}$ using subgradient projection method with step size a_t:

$$\boldsymbol{Q}^{(t+1)} = \mathcal{P}_{\mathcal{S}_Q}\left\{\boldsymbol{Q}^{(t)} + a_t \text{diag}\left(\sum_{n=1}^{N} \boldsymbol{w}_n \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n\right)\right\}. \tag{6.19}$$

7. Set $t \leftarrow t+1$ and **return** to step 2 **until** convergence.

The convergence of the algorithm can be shown by following the same analysis as in Chapter 5. Numerous simulations with random channel realizations show a much faster convergence time of the proposed algorithm as compared to the bisection methods in Section 6.2.1.

6.3 Simulation Results

This section presents the simulation results on the achievable SNR at the destinations of a multiuser relay-assisted network under one of the two relay power assumptions: sum relay power constraint or per-relay power constraints. Also presented are the convergence plots of the modified iterative fixed point algorithm in Section 6.1.3 and the iterative algorithm in Section 6.2.2. The same network configuration as in Chapter 5 is assumed. A sum relay power of 40 is assumed in the network with the sum relay power constraint. Meanwhile, each relay's transmitted power is constrained at 10 with the per-relay power constraints.

Figure 6.1 illustrates the achievable SNR margin at the destination, i.e., $\min_n \text{SNR}_n/\gamma_n$ for 50 different channel realizations. At each channel realization, lowest SNR margin τ's, the highest relay power levels of the 4 relays, and the difference between the highest and lowest relay power levels of the 4 relays are plotted and compared between two different constraints: sum relay power constraint and per-relay power constraints. As can be seen from the figure, imposing the per-relay power constraints does not decrease achievable SNR margin much in most of the simulations, compared with the optimal strategy that imposes the sum relay power constraint. However, the benefit of imposing the per-relay power constraints is very clear in terms of power consumption at each relay. While the highest transmitted power at

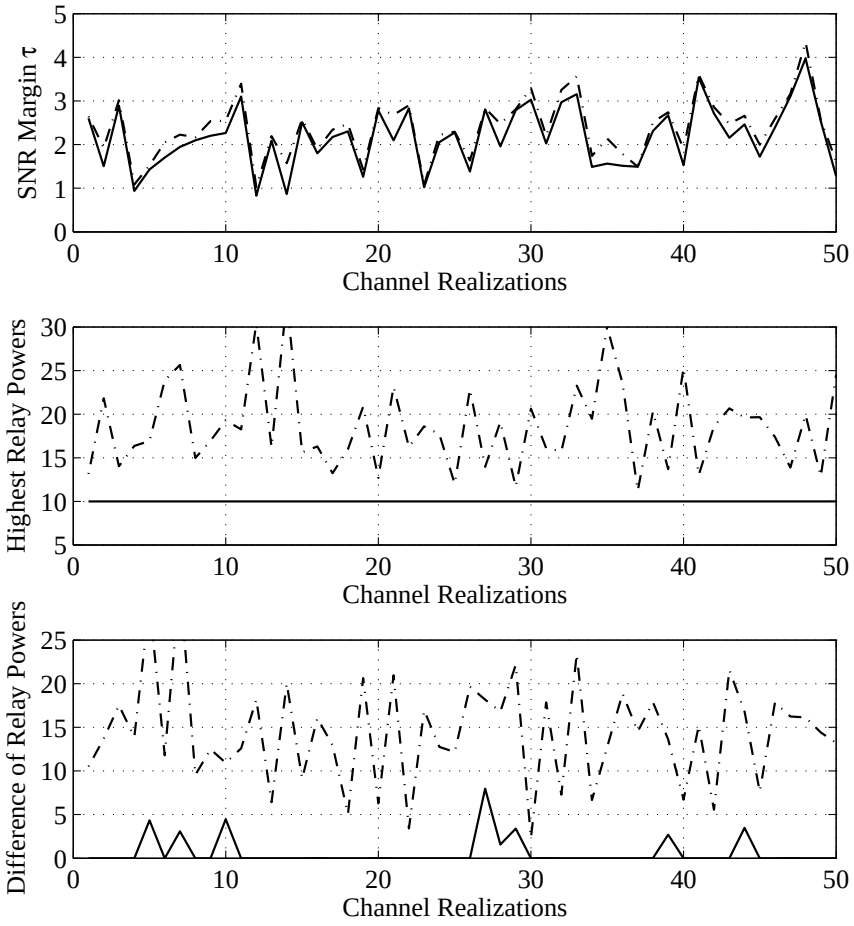

Figure 6.1 The achievable SNR margin τ and the power consumptions at the relays over 50 channel realizations with different power constraints: with per-relay power constraints (solid lines), and without per-relay power constraints ("dash-dot" lines).

each relay is strictly under or equal to 10 with the per-relay power constraints, it may reach to 25 with the sum power constraint. In addition, all the relays transmit at the same power level at 10 almost all the time with the per-relay power constraints; whereas the differences between the highest and lowest power levels between the relays are significant, as much as 20, with the sum relay power constraint.

Figures 6.2 illustrates the convergence of the modified fixed point algorithm in

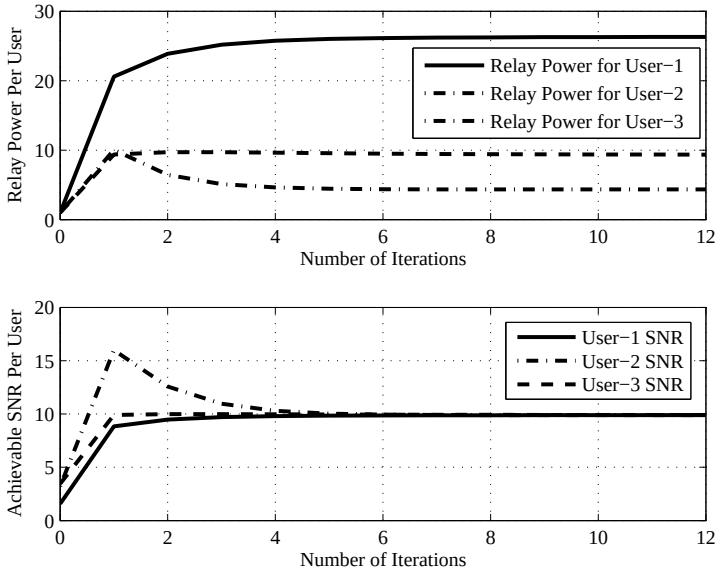

Figure 6.2 Convergence of the relay power for each user and the corresponding achievable SNR at each user's destination by the modified iterative fixed point algorithm.

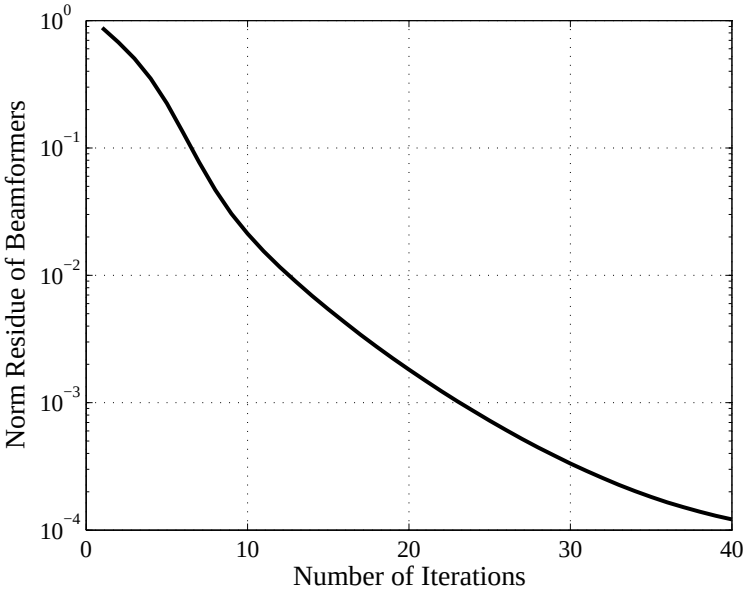

Figure 6.3 Convergence of the proposed algorithm in finding the optimal distributed beamformers with per-relay power constraints to jointly maximize the SNR margin.

Section 6.1.3. Plotted are the evolution of the allocated relay power p_n and the corresponding $\text{SNR}_n, n = 1, 2, 3$ for the 3 users after each iteration. It can be seen that the algorithm converges very quickly after only a few iterations, as the allocated relay power for each user converges to its optimal value. The corresponding SNR also converges to the same optimal value, as all users' SNRs are set at the same weight.

Finally, Figure 6.3 displays the convergence of the proposed iterative algorithm in Section 6.2.2 in finding the optimal distributed beamformers $\boldsymbol{w}_n^\star$ with per-relay power constraints to maximize the SNR margin. The summation $\sum_{n=1}^{N} \|\boldsymbol{w}_n - \boldsymbol{w}_n^\star\|$ plotted after each iteration clearly shows the convergence of the proposed algorithm. Numerous simulations also show that the proposed algorithm converges in a small fraction of the time required by the bisection method.

6.4 Summary and Future Works

This chapter was concerned with the inverse problems to the sum relay power minimization problems in Chapter 5. With the constraints on either the sum relay power or the per-relay power, optimal distributed beamforming designs were studied to jointly maximize the SNR margin at the destinations. It was shown in this chapter that the two optimization problems can be solved effectively by the bisection methods via SOCP feasibility problems. In addition, the chapter also proposed two simple and fast iterative algorithms to directly solve the two problems without the need of a standard conic solution package.

It should be pointed out that the works presented in chapter may also be regarded as per-user rate-balancing maximization problems. More specifically, should SNR_n be the SNR at user-n's destination, the achievable rate for the nth user is given by $\mathcal{R}_n = \log_2(1 + \text{SNR}_n)$. If the weights γ_n are set at the same value, the joint SNR margin maximization presented here tries to maintain the same transmission rate of each user. This strategy is also referred to as resource allocation with fairness. However, maintaining fairness in multiuser network is not always a good approach as it may limit the sum-rate of the network. As an example, when the channels of

one or more users are in deep fades, these users may take the majority of the relay power, and hence reduce the transmission rate of other users. One other consideration for resource allocation in multiuser networks is the sum-rate maximization problem. In particular, in a multiuser multi-relay network, one tries to design the optimal distributed beamformers at the relays to maximize $\sum_{n=1}^{N} \mathcal{R}_n$ subject to a sum relay power constraint or per-relay power constraints.

With a sum power constraint imposed, by applying the result in Lemma 7, the sum-rate maximization problem can be reformulated as a convex problem with p_n's as the variables. The optimization process to find the maximum sum-rate is then standard, and has been reported in a conference paper [57]. In addition, [57] also proposes a distributed algorithm to solve the problem.

With per-relay power constraints imposed, the sum-rate maximization problem is no longer a convex problem, and hence complicates the optimization process of solving the problem. Yet, no optimal solution to the problem has been reported to date. However, it is believed that via the interpretation of virtual uplink channels, it is possible to transform the sum-rate maximization problem to the sum-rate maximization of the virtual channels, which is solvable.

As mentioned in Chapter 5, this thesis only considers the networks with orthogonal $S \rightarrow R$ and orthogonal $S \rightarrow D$ channels. With the other 3 network scenarios, both the sum-rate maximization and rate-balancing optimization problems are still open.

7. Concluding Remarks

This thesis studied optimal designs in wireless multi-relay networks which take advantages of user cooperation. The thesis is divided into two main parts: "Distributed space-time coding" and "Distributed beamforming". These two parts are concerned with two main approaches in cooperative communications over multi-relay networks. While the first approach relies on space-time coding to take advantages of cooperative diversity without the need of CSI, the second approach requires full CSI at the relays to beam the retransmitted signals to the destinations.

Part I of the thesis addressed numerous aspects of distributed space-time coding (DSTC) including distributed code design, optimal power allocation (PA), channel estimation, and performance analysis of mismatched decoding. Regarding the code design, the thesis proposed a new fully-diverse distributed code, called distributed unitary space-time modulation (DUSTM). Not only does the proposed code allow noncoherent reception at the destination, it is also capable of obtaining the maximum diversity order provided by the relay network. In addition, it was unfolded in this part that the knowledge of $R \rightarrow D$ channel has a little or no effect on the code performance. With regard to PA, the thesis showed that maximizing the average effective SNR at the destination does not guarantee the best performance of DSTC. By appropriately of balancing the fading statistics of each $S \rightarrow R \rightarrow D$ link, a novel and simple PA scheme was proposed in a closed-form solution. The proposed PA scheme was shown to significantly outperform other suboptimal PA schemes in coherent, partially coherent, and noncoherent DSTC systems. Finally, from the perspectives of channel estimation and performance analysis of mismatched decoding, the thesis presented the

optimal PA scheme to minimize the mean-square error (MSE) in channel estimation during the training phase. The impact of imperfect channel estimation on the error performance of DSTC was also analyzed, where it was proved that the mismatched decoding of DSTC is able to achieve the same diversity order as the coherent decoding of DSTC.

Part II of the thesis considered a multiuser multi-relay network employing distributed beamforming. With different design criteria, this part attempted to find the optimal distributed beamforming designs at the relays. Two distributed beamforming problems were examined, including (i) minimizing the total relay power subject to guaranteed QoS in terms of SNR at the destinations, and (ii) jointly maximizing the SNR margin at the destinations subject to power constraints at the relays. By means of convex optimization techniques, it was shown in this part that these problems can be formulated and solved via second-order conic programming (SOCP). Thus, optimal solutions to the two problems can be obtained by any conic solution package. In addition, this part also proposed simple and fast iterative algorithms to directly solve the two problems. As the proposed algorithms work without the need of external software package, they can be easily implemented in real-time communications.

A. Space-Time Coding

Exploiting spatial diversity by implementing multiple antennas in transmission and reception can significantly improve the reliability of wireless communications. To take advantage of spatial diversity with multiple transmit antennas, space-time (ST) coding is required. A key characteristic that makes ST coding interesting is its capability of obtaining the full diversity order of the system at a high spectral efficiency without the need of channel state information (CSI) at the transmitter. In addition, as being shown in the first part of this thesis, ST coding can also be implemented in a distributed fashion to take advantage of cooperative diversity inherent in the relay network without the need of CSI at the relays.

The aim of this appendix is to describe the basic concepts of ST coding in traditional MIMO communications. Two examples of ST codes, one for coherent systems, and one for noncoherent systems are presented and discussed in detail. For simplicity of presentation, only systems equipped with one receive antenna are considered.

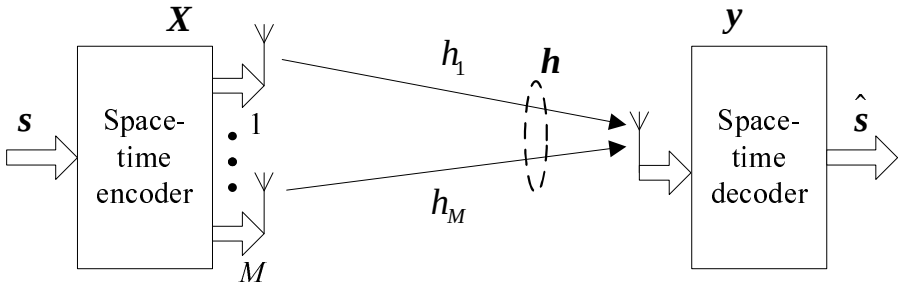

Figure A.1 Block diagram of a space-time coding system.

Figure A.1 illustrates an example of a ST coding system with M transmit antennas and 1 receive antenna. At the encoder (transmitter), an information symbol sequence

$\boldsymbol{s}$ is mapped into a ST codeword $\boldsymbol{X}$ of size $T \times M$. Here, the encoding is performed in space (across M antennas) and time (across T time slots). Suppose that the channel coefficient vector $\boldsymbol{h} = [h_1, \ldots, h_M]^{\mathcal{T}}$ remains constant over the transmission of one ST codeword. The received signal is given by

$$\boldsymbol{y} = \sqrt{\rho/M}\boldsymbol{X}\boldsymbol{h} + \boldsymbol{z}, \tag{A.1}$$

where the AWGN vector $\boldsymbol{z}$ at the receiver contains i.i.d. $\mathcal{CN}(0,1)$ random variables. The channel coefficient in $\boldsymbol{h}$ is also i.i.d. $\mathcal{CN}(0,1)$ random variables. The normalization in (A.1) is such that ρ is the average SNR at the receiver, regardless of the number of transmit antennas.

A.1 Coherent Space-Time Coding

When the CSI is fully known at the receiver, coherent decoding of the ST code is possible. Let $\boldsymbol{X}_k$ be the transmitted ST codeword. Conditioned on $\boldsymbol{X}_k$ and $\boldsymbol{h}$, $\boldsymbol{y}$ is Gaussian distributed with mean $\sqrt{\rho/M}\boldsymbol{X}_k\boldsymbol{h}$ and covariance matrix $\boldsymbol{I}_T$. Thus, the received signal vector $\boldsymbol{y}$ has the following probability density function (pdf)

$$p(\boldsymbol{y}|\boldsymbol{X}_k, \boldsymbol{h}) = \frac{\exp\left(-\left\|\boldsymbol{y} - \sqrt{\rho/M}\boldsymbol{X}_k\boldsymbol{h}\right\|^2\right)}{(\pi N_0)^T}. \tag{A.2}$$

In order to decode the original codeword with a minimum error probability, the coherent maximum likelihood (ML) decoder performs

$$\hat{\boldsymbol{X}} = \arg\max_{\boldsymbol{X}_k} p(\boldsymbol{y}|\boldsymbol{X}_k, \boldsymbol{h}) = \arg\min_{\boldsymbol{X}_k} \left\|\boldsymbol{y} - \sqrt{\rho/M}\boldsymbol{X}_k\boldsymbol{h}\right\|^2. \tag{A.3}$$

From the ML decoding rule of coherent ST, the probability of decoding the received signal as the codeword $\boldsymbol{X}_l$, conditioned on the channel vector $\boldsymbol{h}$ can be calculated as

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l|\boldsymbol{h}) &= \mathbb{P}\left(\left\|\boldsymbol{y} - \sqrt{\rho/M}\boldsymbol{X}_k\right\|^2 \geq \left\|\boldsymbol{y} - \sqrt{\rho/M}\boldsymbol{X}_l\boldsymbol{h}\right\|^2\right) \\
&= \mathbb{P}\left((\rho/M)\|\boldsymbol{\Delta h}\|^2 + 2\sqrt{\rho/M}\mathrm{Re}\{\boldsymbol{z}^{\mathcal{H}}\boldsymbol{\Delta h}\} \leq 0\right) \\
&= Q\left(\sqrt{\frac{\rho\|\boldsymbol{\Delta h}\|^2}{2M}}\right),
\end{aligned}$$

where $\boldsymbol{\Delta} = \boldsymbol{X}_k - \boldsymbol{X}_l$. To compute the (unconditional) pairwise error probability $\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l)$, $\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l|\boldsymbol{h})$ needs to be averaged over the distribution of $\boldsymbol{h}$:

$$\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) = \mathop{\mathbb{E}}_{\boldsymbol{h}} \left[Q\left(\sqrt{\frac{\rho\, \boldsymbol{h}^{\mathcal{H}} \boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathcal{H}} \boldsymbol{h}}{2M}} \right) \right]. \tag{A.4}$$

Since the matrix $\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathcal{H}}$ is Hermitian, it can be unitarily diagonalized as: $\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathcal{H}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathcal{H}}$, where $\boldsymbol{U}$ is unitary and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_M^2)$. Here, λ_i are the singular values of the codeword difference matrix $\boldsymbol{\Delta}$. Thus, the PEP in (A.4) is equivalent to

$$\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) = \mathop{\mathbb{E}}_{\tilde{\boldsymbol{h}}} \left[Q\left(\sqrt{\frac{M \sum_{i=1}^{M} |\tilde{h}_i|^2 \lambda_i^2}{2\rho}} \right) \right], \tag{A.5}$$

where $\tilde{\boldsymbol{h}}_l = \boldsymbol{U}\boldsymbol{h}$. Since h_i are i.i.d. $\mathcal{CN}(0,1)$ random variables and $\boldsymbol{U}$ is unitary, $\tilde{\boldsymbol{h}}$ has the same distribution as $\boldsymbol{h}$. Thus, the PEP can be upper-bounded as

$$\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) < \mathop{\mathbb{E}}_{\tilde{\boldsymbol{h}}} \left[\exp\left(-\frac{\rho \sum_{i=1}^{M} |\tilde{h}_i|^2 \lambda_i^2}{4M} \right) \right] = \prod_{i=1}^{M} \frac{1}{1 + (\rho/4M)\lambda_i^2}, \tag{A.6}$$

where the inequality follows from $Q(x) < \frac{1}{2}\exp\left(-\frac{x^2}{2}\right)$, and the equality comes from taking the expectation over the independent exponential random variables $|\tilde{h}_i|^2$. Since ρ is the average SNR at the receiver, the PEP expression in (A.6) reveals the following design criteria for coherent ST codes [4]:

- *Rank criterion:* If $(\boldsymbol{X}_k - \boldsymbol{X}_l)(\boldsymbol{X}_k - \boldsymbol{X}_l)^{\mathcal{H}}$ is full rank, for all $k \neq l$, the maximum diversity order of the ST code is guaranteed. This is because all the λ_i^2 are strictly positive. Thus, the PEP in (A.6) decays as ρ^{-M}, which indicates that the *maximum diversity order* of M is achieved.

- *Determinant criterion:* If the rank criterion is met, then at high SNR

  $$\mathbb{P}(\boldsymbol{X}_k \to \boldsymbol{X}_l) < \frac{4^M M^M}{\rho^M \prod_{i=1}^{M} \lambda_i^2} = \frac{4^M M^M}{\rho^M \det\left[(\boldsymbol{X}_k - \boldsymbol{X}_l)(\boldsymbol{X}_k - \boldsymbol{X}_l)^{\mathcal{H}}\right]},$$

  which indicates that the *coding gain* is determined by the minimum of the determinant, $\det\left[(\boldsymbol{X}_k - \boldsymbol{X}_l)(\boldsymbol{X}_k - \boldsymbol{X}_l)^{\mathcal{H}}\right]$.

Therefore, a good ST design must guarantee that every codeword difference matrix $\boldsymbol{X}_k - \boldsymbol{X}_l$ is full rank, and the minimum of the determinant of $(\boldsymbol{X}_k - \boldsymbol{X}_l)(\boldsymbol{X}_k - \boldsymbol{X}_l)^{\mathcal{H}}$ is maximized for all $k \neq l$.

We now consider the simplest, yet the most elegant ST block code design. It is the Alamouti code designed for 2 transmit antennas [3]. The Alamouti code possesses two important properties, namely full diversity order and symbol-wise ML decoding. Given the information symbol $\boldsymbol{s} = [s_1,\ s_2]^{\mathcal{T}}$, the Alamouti space-time codeword is

$$\boldsymbol{X} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix},$$

where the columns of $\boldsymbol{X}$ correspond to the transmit antennas, while the rows of $\boldsymbol{X}$ correspond to the time slots. Figure A.2 pictorially describe the encoding and transmission of an Alamouti codeword.

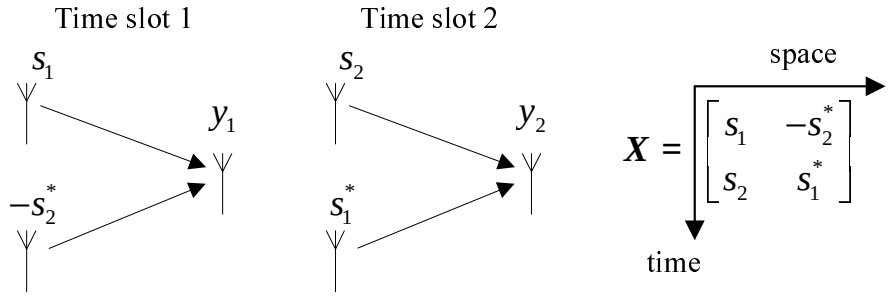

Figure A.2 The encoding and transmission of an Alamouti codeword.

Let $\boldsymbol{X}' = \begin{bmatrix} s_1' & -(s_2')^* \\ s_2' & (s_1')^* \end{bmatrix}$ be another Alamouti codeword that is different from $\boldsymbol{X}$. Then the determinant of the codeword difference matrix is

$$\begin{aligned} \det\left[(\boldsymbol{X} - \boldsymbol{X}')(\boldsymbol{X} - \boldsymbol{X}')^{\mathcal{H}}\right] &= \det\left(\begin{bmatrix} s_1 - s_1' & -s_2^* + (s_2')^* \\ s_2 - s_2' & s_1^* - (s_1')^* \end{bmatrix} \begin{bmatrix} s_1 - s_1' & -s_2^* + (s_2')^* \\ s_2 - s_2' & s_1^* - (s_1')^* \end{bmatrix}^{\mathcal{H}}\right) \\ &= \left(|s_1 - s_1'|^2 + |s_2 - s_2'|^2\right)^2. \end{aligned}$$

Obviously, $\det\left[(\boldsymbol{X} - \boldsymbol{X}')(\boldsymbol{X} - \boldsymbol{X}')^{\mathcal{H}}\right] \neq 0$, when $(s_1, s_2) \neq (s_1', s_2')$, i.e., $\boldsymbol{X} \neq \boldsymbol{X}'$. This clearly shows that the Alamouti code is fully-diverse.

The decoding of the Alamouti code can be done as follows. The system model (A.1) can be rewritten as:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{\boldsymbol{y}} = \sqrt{\frac{\rho}{2}} \underbrace{\begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}}_{\boldsymbol{h}} + \underbrace{\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}}_{\boldsymbol{z}}, \tag{A.7}$$

which is equivalent to

$$\begin{aligned} y_1 &= \sqrt{\rho/2}\,(h_1 s_1 - h_2 s_2^*) + z_1 \\ y_2 &= \sqrt{\rho/2}\,(h_1 s_2 + h_2 s_1^*) + z_2. \end{aligned}$$

In order to decode s_1, and s_2, perform the following:

$$\begin{aligned} h_1^* y_1 + h_2 y_2^* &= \sqrt{\rho/2}(|h_1|^2 + |h_2|^2) s_1 + h_1^* z_1 + h_2 z_2^* \\ -h_2^* y_1 + h_1 y_2^* &= \sqrt{\rho/2}(|h_1|^2 + |h_2|^2) s_2 - h_2^* z_1 + h_1 z_2^* \end{aligned}$$

Note that the above implementations are possible since h_1 and h_2 are assumed to be known at the receiver (coherent system). The above representations show that the decoding of the two information symbol s_1 and s_2 can be done separately, i.e., with low complexity:

$$\begin{aligned} \hat{s}_1 &= \arg\min_{s_1} \left| h_1^* y_1 + h_2 y_2^* - \sqrt{\rho/2}(|h_1|^2 + |h_2|^2) s_1 \right|^2, \\ \hat{s}_2 &= \arg\min_{s_2} \left| -h_2^* y_1 + h_1 y_2^* - \sqrt{\rho/2}(|h_1|^2 + |h_2|^2) s_2 \right|^2. \end{aligned}$$

Furthermore, based on the effective channel gain magnitude of $\frac{\sqrt{\rho/2}\left(|h_1|^2+|h_2|^2\right)}{\sqrt{\mathrm{var}\left[h_1^* z_1 + h_2 z_2^*\right]}} = \sqrt{\rho/2}\sqrt{|h_1|^2 + |h_2|^2}$, one concludes that the diversity order is 2, i.e., a full diversity order of the 2 transmit and 1 receive antenna system.

A.2 Noncoherent Space-Time Coding

When the CSI is not available at the receiver, the ML coherent decoding of the ST code is no longer possible. Let $\boldsymbol{X}_k$ be the transmitted codeword. As $\boldsymbol{h}$ is unknown, the received signal vector $\boldsymbol{y}$ is Gaussian distributed with zero mean and covariance

matrix: $\boldsymbol{\Omega} = (\rho/M)\boldsymbol{X}\boldsymbol{X}^{\mathcal{H}} + \boldsymbol{I}_T$. Thus, the received signal vector $\boldsymbol{y}$ has the following conditional pdf:

$$p(\boldsymbol{y}|\boldsymbol{X}_k) = \frac{\exp(-\boldsymbol{y}^{\mathcal{H}}\boldsymbol{\Omega}^{-1}\boldsymbol{y})}{\pi^T \det(\boldsymbol{\Omega})}.$$

To facilitate noncoherent decoding, unitary space-time constellations are commonly used. Suppose that there is a constellation of L unitary matrices $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_L$ (each of size $T \times M$), such that $\boldsymbol{\Phi}_k^{\mathcal{H}}\boldsymbol{\Phi}_k = \boldsymbol{I}_M, \forall k$. From this pool of unitary matrices, the transmitted signal matrix is formed as $\boldsymbol{X}_k = \sqrt{T}\boldsymbol{\Phi}_k$. Using property $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})$ [28], the determinant of $\boldsymbol{\Omega}$ is given by

$$\det(\boldsymbol{\Omega}) = \det\left[\boldsymbol{I}_T + (\rho T/M)\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathcal{H}}\right] = \det\left[\boldsymbol{I}_M + (\rho T/M)\boldsymbol{\Phi}_k^{\mathcal{H}}\boldsymbol{\Phi}_k\right] = \left[1 + (\rho T/M)\right]^M.$$

Whereas, using the matrix inverse formula [28]

$$(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{C}\boldsymbol{D})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{D}\boldsymbol{A}^{-1},$$

the inverse of $\boldsymbol{\Omega}$ is

$$\begin{aligned}
\boldsymbol{\Omega}^{-1} &= \left(\boldsymbol{I}_T + (\rho T/M)\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathcal{H}}\right)^{-1} \\
&= \boldsymbol{I}_T - \boldsymbol{\Phi}_k\left((M/\rho T)\boldsymbol{I}_M + \boldsymbol{\Phi}_k^{\mathcal{H}}\boldsymbol{\Phi}_k\right)^{-1}\boldsymbol{\Phi}_k^{\mathcal{H}} \\
&= \boldsymbol{I}_T - \frac{1}{1 + M/\rho T}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathcal{H}}.
\end{aligned}$$

Thus, the noncoherent ML decoder performs [58]

$$\hat{\boldsymbol{\Phi}} = \arg\max_{\boldsymbol{X}_k} p(\boldsymbol{y}|\boldsymbol{X}_k) = \arg\max_{\boldsymbol{X}_k} -\boldsymbol{y}^{\mathcal{H}}\boldsymbol{\Omega}^{-1}\boldsymbol{y} = \arg\max_{\boldsymbol{\Phi}_k} \boldsymbol{y}^{\mathcal{H}}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathcal{H}}\boldsymbol{y}. \tag{A.8}$$

From the ML decoding rule of noncoherent ST code, the probability of decoding the received signal as the codeword $\boldsymbol{\Phi}_l$ (the PEP) is

$$\mathbb{P}\left(\boldsymbol{\Phi}_k \to \boldsymbol{\Phi}_l\right) = \mathbb{P}\left(\boldsymbol{y}^{\mathcal{H}}\boldsymbol{\Phi}_k\boldsymbol{\Phi}_k^{\mathcal{H}}\boldsymbol{y} < \boldsymbol{y}^{\mathcal{H}}\boldsymbol{\Phi}_l\boldsymbol{\Phi}_l^{\mathcal{H}}\boldsymbol{y}\right). \tag{A.9}$$

This PEP, evaluated in [58], is upper-bounded as

$$\mathbb{P}\left(\boldsymbol{\Phi}_k \to \boldsymbol{\Phi}_l\right) \leq \frac{1}{2}\prod_{i=1}^{M}\frac{1}{1 + \frac{(\rho T/M)^2(1-d_i^2)}{4(1+\rho T/M)}}, \tag{A.10}$$

where $1 \geq d_1 \geq \cdots \geq d_M \geq 0$ are singular values of the $M \times M$ correlation matrix $\boldsymbol{\Phi}_k^{\mathcal{H}} \boldsymbol{\Phi}_l$. Clearly, the above upper-bound is minimized when $d_1 = \cdots = d_M = 0$. This means that the ideal constellation design is obtained when all the columns of $\boldsymbol{\Phi}_k$ are orthogonal to all the columns of $\boldsymbol{\Phi}_l$ for $k \neq l$ [58]. On the other hand, if any $d_i = 1$, the diversity order of the ST code will be reduced. Thus, in order to design a good noncoherent ST constellation, two design criteria are as follows: (i) no singular value of the correlation matrix $\boldsymbol{\Phi}_k^{\mathcal{H}} \boldsymbol{\Phi}_l, \forall k \neq l$ is 1 to achieve the maximum diversity order, and (ii) the singular values are minimized to maximize the achievable coding gain.

We now consider a simple orthogonal design of unitary constellation, which is based on the coherent Alamouti code [59]. Another design of unitary constellation is the Fourier-based design [27], which is presented in Section 2.4. For the Alamouti-based design, consider concatenating a known matrix $\boldsymbol{T} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ and the Alamouti matrix $\boldsymbol{A} = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}$. In order to guarantee the unitarity of $\boldsymbol{\Phi} = [\boldsymbol{T}^{\mathcal{T}} \ \boldsymbol{A}^{\mathcal{T}}]^{\mathcal{T}}$, one can choose s_1 and s_2 to be Q-PSK (phase shift keying) symbols. In particular, define

$$\boldsymbol{\Phi}_{m,n} = \frac{1}{2} \begin{bmatrix} 1 & 1 & \mathrm{e}^{j\frac{2\pi}{Q}m} & \mathrm{e}^{j\frac{2\pi}{Q}n} \\ -1 & 1 & -\mathrm{e}^{-j\frac{2\pi}{Q}n} & \mathrm{e}^{-j\frac{2\pi}{Q}m} \end{bmatrix}^{\mathcal{T}},$$

where $(m, n) \in S \times S$ and $S = \{0, 1, \ldots, Q-1\}$. As a result, a constellation of $L = Q^2$ unitary space-time matrices can be designed, where $\boldsymbol{\Phi}_{m,n}^{\mathcal{H}} \boldsymbol{\Phi}_{m,n} = \boldsymbol{I}_2$. Suppose that $\boldsymbol{\Phi}_{m,n}$ and $\boldsymbol{\Phi}_{k,l}$ are two distinguished codewords, the two singular values of $\boldsymbol{\Phi}_{m,n}^{\mathcal{H}} \boldsymbol{\Phi}_{k,l}$ can be calculated as [59]:

$$d_1 = d_2 = \frac{1}{2}\sqrt{2 + \cos\frac{2\pi}{Q}(m-k) + \cos\frac{2\pi}{Q}(n-l)}.$$

It is clear that $d_i < 1$. Thus, this orthogonal design of unitary constellation achieves the maximum diversity order of 2 with 1 receive antenna. Interestingly, the design also inherits the symbol-wise ML decoding from Alamouti code as follows. Since

$$\boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Phi}_{m,n} \boldsymbol{\Phi}_{m,n}^{\mathcal{H}} \boldsymbol{y} = 2\|\boldsymbol{y}\|^2 + \boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Phi}_{m,0} \boldsymbol{\Phi}_{m,0}^{\mathcal{H}} \boldsymbol{y} + \boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Phi}_{0,n} \boldsymbol{\Phi}_{0,n}^{\mathcal{H}} \boldsymbol{y},$$

the symbol-wise ML decoding is given by [59]

$$\hat{m} = \arg\max_{m \in S} \ \boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Phi}_{m,0} \boldsymbol{\Phi}_{m,0}^{\mathcal{H}} \boldsymbol{y},$$
$$\hat{n} = \arg\max_{n \in S} \ \boldsymbol{y}^{\mathcal{H}} \boldsymbol{\Phi}_{0,n} \boldsymbol{\Phi}_{0,n}^{\mathcal{H}} \boldsymbol{y}.$$

B. Convex Optimization Theory

The design and analysis of communication and signal processing systems have relied heavily on mathematical modeling tool. In that process, many communication problems can be naturally formulated or transformed into convex optimization problems, which facilitate their analytical and numerical solutions. There are several theoretical and conceptual advantages of formulating a problem as a convex optimization problem. First, a local optimum is also the global optimum in a convex problem. Second, the associated dual problem often reveals interesting interpretations to the original problem, such as the structure of the solution or an efficient method for solving it. Third, recent developments in convex optimization methods, such as the interior-point methods, allows fast and powerful numerical tools to solve several classes of convex problems.

This thesis, aims to find various optimal designs in multi-relay wireless networks, is no exception, as it also relies on convex optimization theory. Many aspects of convex optimization are mentioned throughout the thesis. The aim of this appendix is to present a brief overview of convex optimization theory, and review the theory behind optimization problems in the thesis. The appendix starts with the definitions of affine set, convex set, and cone, and summarizes the convex sets mentioned in the thesis. Convex optimization and duality are described in the second part. The appendix is then concluded by presenting algorithms to solve the problem of Euclidean projection on a convex set. It should be emphasized that a thorough discussion on convex optimization is well beyond the scope of this appendix. The interested reader is referred to complete and well-treated sources on the field, such as [37].

B.1 Convex Sets and Convex Functions

B.1.1 Convex Sets

A set $\mathcal{C} \subseteq \mathbb{R}^n$ is affine if the line through any two distinct points in $\mathcal{C}$ lies in $\mathcal{C}$, i.e., $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$ and $\theta \in \mathbb{R}$, we have $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in \mathcal{C}$. Figure B.1 gives an example of an affine set.

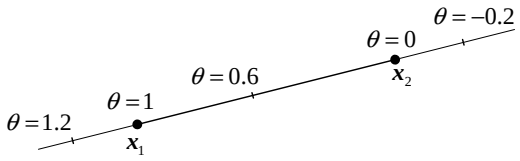

Figure B.1 Example of an affine set: a line passing through $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Any point, described by $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2$, where θ varies over $\mathbb{R}$, lies on the line.

A set $\mathcal{C}$ is convex if the line segment between any two points in $\mathcal{C}$ lies entirely in $\mathcal{C}$, i.e., $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$ and any $0 \leq \theta \leq 1$, we have $\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in \mathcal{C}$. Figure B.2 illustrates examples of convex and nonconvex sets.

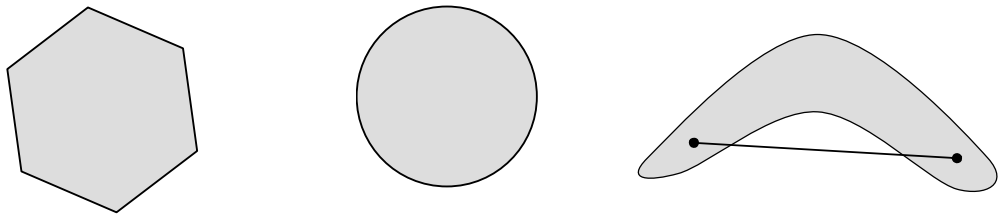

Figure B.2 Examples of convex and nonconvex sets: the hexagon and the circle are convex, whereas the the boomerang is not. The line segment between the two points in the "boomerang" set (shown as dots) is not contained in the set.

A set $\mathcal{C}$ is called a cone if for any $\boldsymbol{x} \in \mathcal{C}$, and $\theta \geq 0$, we have $\theta \boldsymbol{x} \in \mathcal{C}$. A set $\mathcal{C}$ is a convex cone if it is convex and a cone, i.e., for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$, and $\theta_1, \theta_2 \geq 0$, we have $\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 \in \mathcal{C}$. A cone example is shown in Figure B.3.

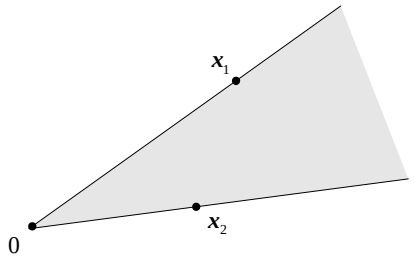

Figure B.3 Example of a cone: the pie slice shows all points of the form $\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2$, where $\theta_1, \theta_2 \geq 0$.

A second-order cone is a special class of convex cone. It is defined as the set

$$\mathcal{C} = \{(\boldsymbol{x}, t) |\ \|\boldsymbol{x}\| \leq t\} \subseteq \mathbb{R}^{n+1},$$

where the second-order norm is defined as $\|\boldsymbol{x}\| = \sqrt{x_1^2 + \ldots + x_n^2}$ with $\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathcal{T}}$. Figure B.4 shows the boundary of second-order cone in $\mathbb{R}^3$.

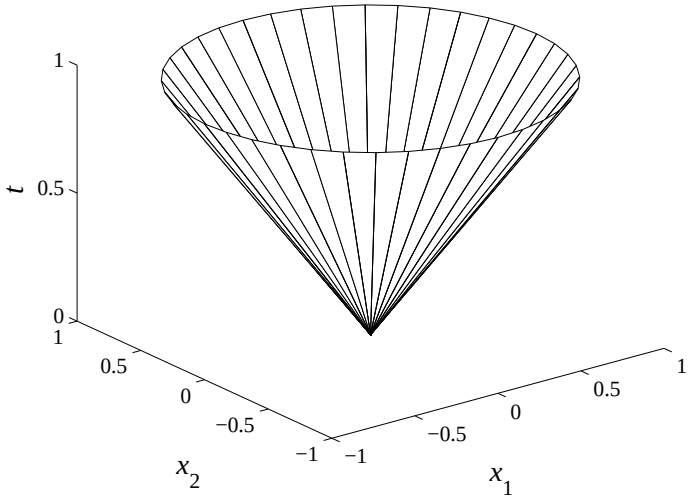

Figure B.4 Boundary of second-order cone in $\mathbb{R}^3 : \left\{(x_1, x_2, t) \,|\, (x_1^2 + x_2^2)^{1/2} \leq t\right\}$.

A positive semidefinite cone is a special case of convex cone, denoted as $\mathbf{S}_+^n = \{\boldsymbol{X} \in \mathbf{S}^n | \boldsymbol{X} \succeq \mathbf{0}\}$, where $\mathbf{S}^n$ is the set of symmetric $n \times n$ matrices. $\boldsymbol{X} \in \mathbf{S}_+^n \Leftrightarrow \boldsymbol{z}^{\mathcal{T}} \boldsymbol{X} \boldsymbol{z} \geq 0,\ \forall \boldsymbol{z}$.

Figure B.5 illustrates the boundary of positive semidefinite cone in $\mathbf{S}^2$. Here,

$$\boldsymbol{X} = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}_+^2 \Leftrightarrow x \geq 0,\, z \geq 0,\, xz \geq y^2.$$

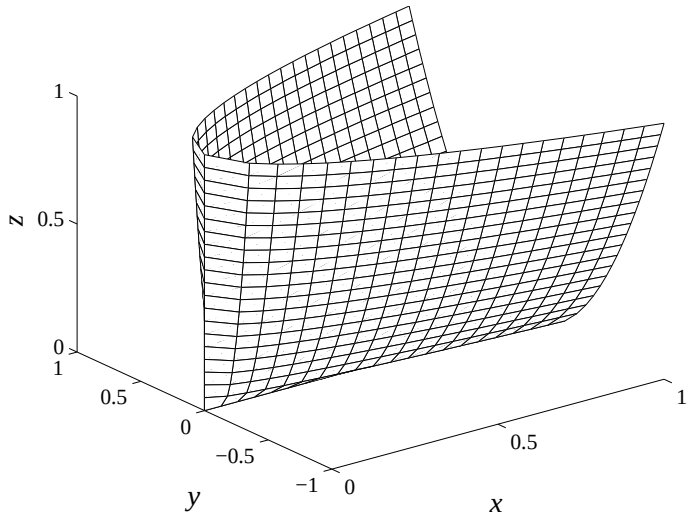

Figure B.5 Boundary of positive semidefinite cone in $\mathbf{S}^2$.

B.1.2 Convex Functions

Convex function: A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if the domain where f is defined, denoted as **dom** f, is a convex set and

$$f(\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2) \leq \theta f(\boldsymbol{x}_1) + (1-\theta) f(\boldsymbol{x}_2) \tag{B.1}$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in$ **dom** f, $0 \leq \theta \leq 1$. In geometric visualization, the inequality means that the graph segment of the function f between $(\boldsymbol{x}_1, f(\boldsymbol{x}_1))$ and $(\boldsymbol{x}_2, f(\boldsymbol{x}_2))$ is strictly below the line segment connecting the two points (see Figure B.6 for an illustration). A function f is said to be concave if $-f$ is a convex function.

Besides the definition of a convex function, there are several ways to verify the convexity of a function f:

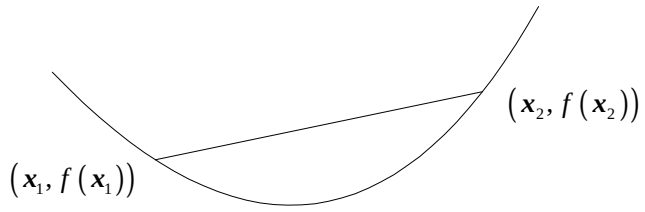

Figure B.6 Graph of a convex function. The line segment between any two points on the graph lies above the graph.

- *First order condition:* if f is differentiable, i.e., the gradient of f:

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

  exists at each $\boldsymbol{x} \in \mathbf{dom}\ f$, the first order condition states that f is convex if and only if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathcal{T}}(\boldsymbol{y} - \boldsymbol{x}), \quad \text{for all } \boldsymbol{y} \in \mathbf{dom}\ f. \tag{B.2}$$

- *Second order condition:* if f is twice-differentiable, i.e., the Hessian $\nabla^2 f(\boldsymbol{x})$:

$$\nabla^2 f(\boldsymbol{x})_{ij} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$$

  exists at each $\boldsymbol{x} \in \mathbf{dom}\ f$, the second order condition states that f is convex if and only if

$$\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0} \quad \text{for all } \boldsymbol{x} \in \mathbf{dom}\ f. \tag{B.3}$$

- Show that f is obtained from simpler convex functions by operations that preserves convexity, such as: nonnegative weighted sum, composition of affine function, pointwise maximum and supremum, minimization, and perspective.

Quasiconvex function: A function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if $\mathbf{dom}\ f$ is convex and the sub-level sets

$$\mathcal{S}_\alpha = \{\boldsymbol{x} \in \mathbf{dom}\ f | f(\boldsymbol{x}) \leq \alpha\} \tag{B.4}$$

is convex for all α. A convex function is also quasiconvex, but the inverse is not true. A function f is said to be quasiconcave if $-f$ is a quasiconvex function. If a function is both quasiconvex and quasiconcave, it is quasilinear. Figure B.7 illustrates a quasiconvex function on $\mathbb{R}$.

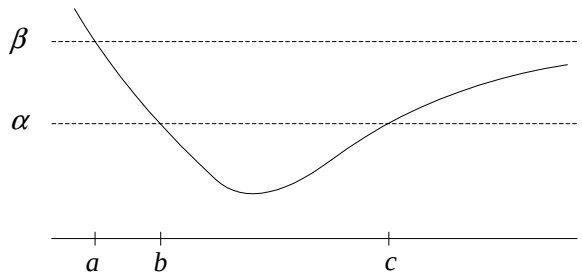

Figure B.7 Graph of a quasiconvex function on $\mathbb{R}$. The sublevel set $\mathcal{S}_\alpha$ is the interval $[b, c]$, which is convex. The sublevel set $\mathcal{S}_\beta$ is the interval $[a, \infty)$, which is also convex.

B.2 Convex Optimization

A convex optimization problem has the following standard form

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, m \\ & h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p \\ & \boldsymbol{x} \in \mathcal{S}, \end{array} \tag{B.5}$$

where $\boldsymbol{x}$ is the optimization variable, f_0 is the objective function, f_i and h_i are inequality and equality constraint functions, respectively.

A set of points for which the objective and all the constraints are defined,

$$\mathcal{S} = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\ h_i,$$

is called the domain of the optimization problem (B.5). Problem (B.5) is said to be *feasible* if there exists at least one solution in the domain $\mathcal{S}$, i.e., $\mathcal{S}$ is not empty. It is called *infeasible* otherwise.

B.2.1 Classes of Convex Optimization

Linear programming (LP): all the functions f_i and h_i are affine.

Linear fractional programming: is similar to LP, except the the objective function

is now in the form

$$f_0(\boldsymbol{x}) = \frac{\boldsymbol{c}^{\mathcal{T}}\boldsymbol{x} + d}{\boldsymbol{e}^{\mathcal{T}}\boldsymbol{x} + f},$$

and **dom** $f_0(\boldsymbol{x}) = \{\boldsymbol{x} | \boldsymbol{e}^{\mathcal{T}}\boldsymbol{x} + f \geq 0\}$. While a linear fractional problem is not readily in a convex form, a change in variables $\boldsymbol{y} = \frac{\boldsymbol{x}}{\boldsymbol{e}^{\mathcal{T}}\boldsymbol{x}+f}$ and $z = \frac{d}{\boldsymbol{e}^{\mathcal{T}}\boldsymbol{x}+f}$ can effectively transform it into a LP, which is convex.

Quadratic programming (QP): when the objective function is in quadratic form: $f_0(\boldsymbol{x}) = \boldsymbol{x}^{\mathcal{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathcal{T}}\boldsymbol{x} + c$, and $\boldsymbol{A}$ is positive semidefinite. If the constraints are also in quadratic form, the problem is called a quadratically constrained quadratic programming (QCQP).

Second order conic programming (SOCP): when the inequalities are SOC constraints:

$$\|\boldsymbol{A}_i\boldsymbol{x} + \boldsymbol{b}_i\| \leq \boldsymbol{c}_i^{\mathcal{T}}\boldsymbol{x} + d_i,$$

the optimization problem is called an SOC program. SOCP is more general than QCQP and LP.

Semidefinite programming (SDP): a semidefinite program has the following form

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \boldsymbol{c}^{\mathcal{T}}\boldsymbol{x} \\ \text{subject to} \quad & x_1\boldsymbol{F}_1 + x_2\boldsymbol{F}_2 + \ldots + x_n\boldsymbol{F}_n + \boldsymbol{G} \preceq \boldsymbol{0} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{B.6}$$

with $\boldsymbol{F}_i, \boldsymbol{G} \in \mathbf{S}_+^n$. Here, the inequality constraint is called linear matrix inequality (LMI). It is noted that multiple LMI constraints can be equivalently presented as a single LMI.

B.2.2 Lagrangian Duality

The basis of Lagrangian duality is to augment the objective function with a weighted-sum of the constraint functions. With this method, a constrained optimization problem can be expressed as a non-constrained problem. The Lagrangian

$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with Problem (B.5) is formed as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}), \tag{B.7}$$

where $\lambda_i \geq 0$ and $\nu_i \geq 0$ are the Lagrangian multipliers associated with $f_i(\boldsymbol{x}) \leq 0$ and $h_i(\boldsymbol{x}) = 0$ constraints, respectively. The parameters λ_i and ν_i are also called dual variables. The Lagrangian dual function $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ is defined as

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{S}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \mathcal{S}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \right). \tag{B.8}$$

Since the dual function is the pointwise infimum of a family of affine function of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$, it is always concave, regardless of the convexity of the original problem (B.5). Moreover, it is easy to verify that $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\boldsymbol{x})$ for every feasible point $\boldsymbol{x}$, and $\lambda_i \geq 0$, $\nu_i \geq 0$. This important property yields the lower bound on the optimal value $p^\star$ of Problem (B.5), i.e., $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$. Then, the *best* lower bound can be obtained from the Lagrangian dual function

$$\begin{aligned} &\underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} \quad g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &\text{subject to} \quad \boldsymbol{\lambda} \succeq \mathbf{0}. \end{aligned} \tag{B.9}$$

This optimization problem is called the Lagrangian dual problem of the primal problem (B.5). The reason to establish the dual problem is that the dual problem is usually easier to solve than the primal problem. Furthermore, not only the dual problem does reveal the result to the primal problem, it may also provide some insights and interesting interpretation to the structure of the solutions. Sometimes the dual problem leads to an efficient or distributed method to solve the original problem.

Suppose that $d^\star$ is the optimal value of (B.9). The difference $p^\star - d^\star$ is called the dual gap between the primal and dual problems. The dual gap is always nonnegative. Strong duality holds if the dual gap is 0. In general, if the primal problem is convex, the dual gap will be 0. If strong duality holds, and $\boldsymbol{x}^\star$, $\lambda_i^\star$, $\nu_i^\star$ are optimal solutions,

they must satisfy the KKT conditions:

$$
\begin{aligned}
&f_i(\boldsymbol{x}^\star) \leq 0,\ i = 1, \ldots, m,\ h_i(\boldsymbol{x}^\star) = 0,\ i = 1, \ldots, p,\\
&\boldsymbol{\lambda}^\star \succeq \mathbf{0},\\
&\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0,\ i = 1, \ldots, m,\\
&\nabla f_0(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\boldsymbol{x}^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(\boldsymbol{x}^\star) = 0.
\end{aligned}
$$

Likewise, if $\tilde{\boldsymbol{x}}$, $\tilde{\lambda}_i$, $\tilde{\nu}_i$ satisfy the KKT conditions for a convex problem, they are optimal. Solving the KKT conditions is equivalent to solving the primal and dual problems. Thus, the KKT conditions are necessary and sufficient for optimality in convex programming.

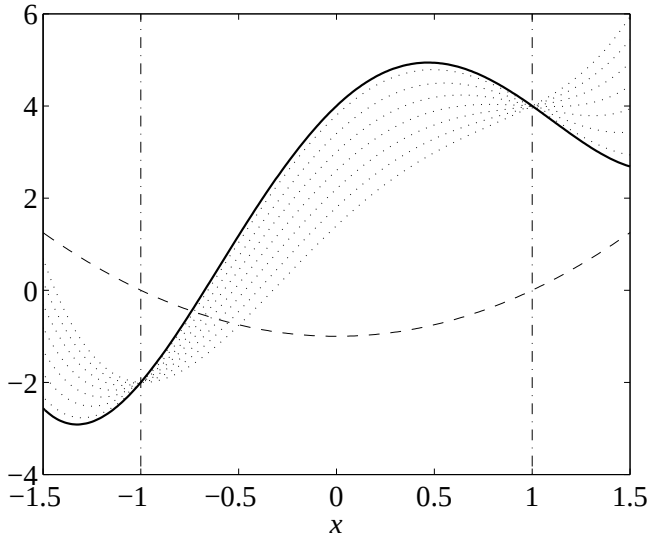

Figure B.8 Lower bound from the Lagrangian dual function. The solid curve is the objective function f_0, and the dashed curve is the constraint function f_1. The feasible set is the interval $[-1, 1]$, indicated by the two "dash-dot" vertical lines. The optimal point and value are $x^\star = -1$, $p^\star = -2$. The dot curves show $\mathcal{L}(x, \lambda)$ for $\lambda = 0.2, 0.6, 1.0, \ldots, 2.6$. Each of these has a minimum value smaller than $p^\star$, since on the feasible set and for $\lambda \geq 0$, we have $\mathcal{L}(x, \lambda) \leq f_0$.

Figure B.8 illustrates an example of an optimization problem of a nonconvex function with one inequality constraint. Also plotted in the figure are Lagrangian dual

functions associated with different values of the Lagrangian multiplier λ. The figure clearly shows the lower bound of the original function established by the Lagrangian dual functions. In Figure B.9, we plot the dual function $g(\lambda)$ for the problem in Figure B.8. Even though the primal function is not convex, the dual function is strictly concave. The figure also shows the strong duality in this problem as the maximum value of the dual function is the same as the minimum value of the primal function.

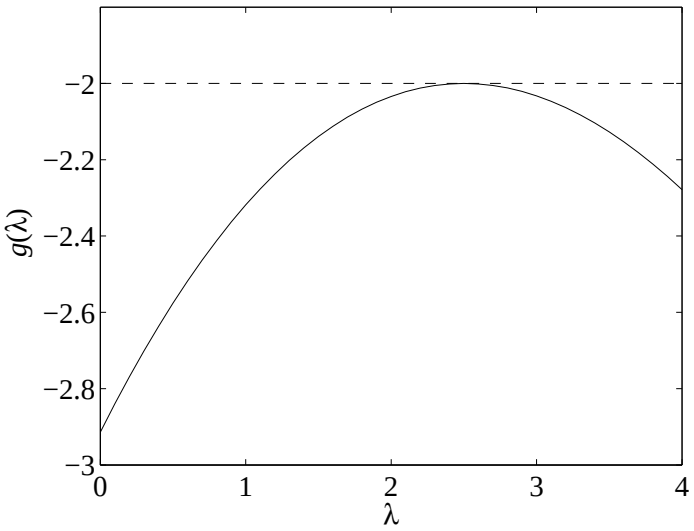

Figure B.9 Dual function $g(\lambda)$ for the problem in Figure B.8. f_0 is not convex, but the dual function is strictly concave. The horizontal dashed line shows $p^\star$, which is the upper-bound on $g(\lambda)$. Strong duality holds in this problem, as the maximum value $d^\star$ of $g(\lambda)$ satisfies $d^\star = p^\star$.

B.3 Projection on a Set

The projection $\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}_0)$ of a point $\boldsymbol{x}_0 \in \mathbb{R}^n$ to a closed set $\mathcal{C} \subseteq \mathbb{R}^n$ is equivalent to finding the point $\boldsymbol{x} \in \mathcal{C}$ that makes the Euclidean distance $\|\boldsymbol{x} - \boldsymbol{x}_0\|$ between $\boldsymbol{x}$ and $\boldsymbol{x}_0$ minimized. Mathematically, the projection can be stated as an optimization problem

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{x} - \boldsymbol{x}_0\| \\ &\text{subject to} \quad \boldsymbol{x} \in \mathcal{C}. \end{aligned} \tag{B.10}$$

In general, there can be more than one projection of $\boldsymbol{x}_0$ in $\mathcal{C}$. However, if $\mathcal{C}$ is a convex set, the projection is a convex QP and its solution is unique. If $\mathcal{C}$ is a halfspace, i.e., $\mathcal{C} = \{\boldsymbol{x} \mid \boldsymbol{a}^{\mathcal{T}}\boldsymbol{x} \leq b\}$, the projection of $\boldsymbol{x}_0$ on $\mathcal{C}$ is given by a closed-form solution [37]:

$$\mathcal{P}_{\mathcal{C}}(\boldsymbol{x}_0) = \begin{cases} \boldsymbol{x}_0 + (b - \boldsymbol{a}^{\mathcal{T}}\boldsymbol{x}_0)\boldsymbol{a}/\|\boldsymbol{a}\|^2 & \text{if } \boldsymbol{a}^{\mathcal{T}}\boldsymbol{x}_0 > b \\ \boldsymbol{x}_0 & \text{if } \boldsymbol{a}^{\mathcal{T}}\boldsymbol{x}_0 \leq b. \end{cases} \tag{B.11}$$

In Chapters 5 and 6, the update of $\boldsymbol{Q}^{(t)}$ requires the projection on the set $\mathcal{S}_Q = \{\boldsymbol{Q} : \mathrm{tr}(\boldsymbol{Q}\boldsymbol{P}) \leq \mathrm{tr}(\boldsymbol{P}), \boldsymbol{Q} \succeq \boldsymbol{0}\}$. That is $\boldsymbol{Q}^{(t+1)} = \mathcal{P}_{\mathcal{S}_Q}\left\{\boldsymbol{Q}^{(t)} + a_t \mathrm{diag}\left\{\sum_{n=1}^{N} \boldsymbol{w}_n \boldsymbol{w}_n^{\mathcal{H}} \boldsymbol{D}_n\right\}\right\} \triangleq \mathcal{P}_{\mathcal{S}_Q}\left\{\boldsymbol{Q}_0^{(t+1)}\right\}$. Since $\mathcal{S}_Q$ is convex, the projection can be found uniquely by convex QP. However, as the projected point $\boldsymbol{Q}_0^{(t+1)} \succeq \boldsymbol{0}$, the special structure of the set $\mathcal{S}_Q$ allows a simple and fast algorithm to find $\mathcal{P}_{\mathcal{S}_Q}\left\{\boldsymbol{Q}_0^{(t+1)}\right\}$ as explained next.

Note that $\boldsymbol{Q}$ is always a diagonal matrix of size $R \times R$, so the projection is performed in the R dimension vector space. Let $\boldsymbol{x} = \mathrm{diag}(\boldsymbol{Q})$. The projection of $\boldsymbol{Q}_0^{(t+1)}$ on the set $\mathcal{S}_Q$ is equivalent to the projection of $\boldsymbol{x}_0 = \mathrm{diag}(\boldsymbol{Q}_0^{(t+1)})$ on the set $\mathcal{C} = \{\boldsymbol{x} : \boldsymbol{a}^{\mathcal{T}}\boldsymbol{x} \leq b, \boldsymbol{x} \succeq \boldsymbol{0}\}$, where $\boldsymbol{a} = \mathrm{diag}(\boldsymbol{P})$, and $b = \mathrm{tr}(\boldsymbol{P})$.

Denote $\mathcal{C}_0 = \{\boldsymbol{x} : \boldsymbol{x} \succeq \boldsymbol{0}\}$, $\mathcal{C}_1 = \{\boldsymbol{x} : \boldsymbol{a}^{\mathcal{T}}\boldsymbol{x} \leq b\}$. Clearly, $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1$. If the set $\mathcal{C}$ were the half-space $\mathcal{C}_1$, the projection would be in the closed-form solution as in (B.11). Thus, by performing the simple projection $\boldsymbol{x}_1 = \mathcal{P}_{\mathcal{C}_1}(\boldsymbol{x}_0)$ first, and verifying if $\boldsymbol{x}_1 \in \mathcal{C}_0$ is true, then $\boldsymbol{x}_1$ will also be the projection of $\boldsymbol{x}_0$ on $\mathcal{C}$. On the other hand, if $\boldsymbol{x}_1 \notin \mathcal{C}_0$, the following iterative steps would ensure a fast projection to the true projection of $\boldsymbol{x}_0$:

1. **Set** the iteration index $t = 1$, and $\boldsymbol{a}_t = \boldsymbol{a}$.
2. **Repeat:**
   - Note down all the indices i's such that $[\boldsymbol{x}_t]_i < 0$.
   - Define $\tilde{\boldsymbol{a}}_t = \boldsymbol{a}_t$ except for the indices i's, set $[\tilde{\boldsymbol{a}}_t]_i = 0$. Similarly, define $\tilde{\boldsymbol{x}}_t = \boldsymbol{x}_t$ except for the indices i's, set $[\tilde{\boldsymbol{x}}_t]_i = 0$. Furthermore, define the set $\mathcal{C}_{t+1} = \{\boldsymbol{x} : \tilde{\boldsymbol{a}}_t^{\mathcal{T}}\boldsymbol{x} \leq b\}$.

- Perform the projection $\boldsymbol{x}_{t+1} = \mathcal{P}_{\mathcal{C}_{t+1}}(\tilde{\boldsymbol{x}}_t)$ by the closed-form solution in (B.11).

3. If $\boldsymbol{x}_{t+1} \in \mathcal{C}_0$, then $\boldsymbol{x}_{t+1}$ is the projection of $\boldsymbol{x}_0$ onto the set $\mathcal{C}$. Otherwise, **set** $t = t + 1$ and **return** to Step 2.

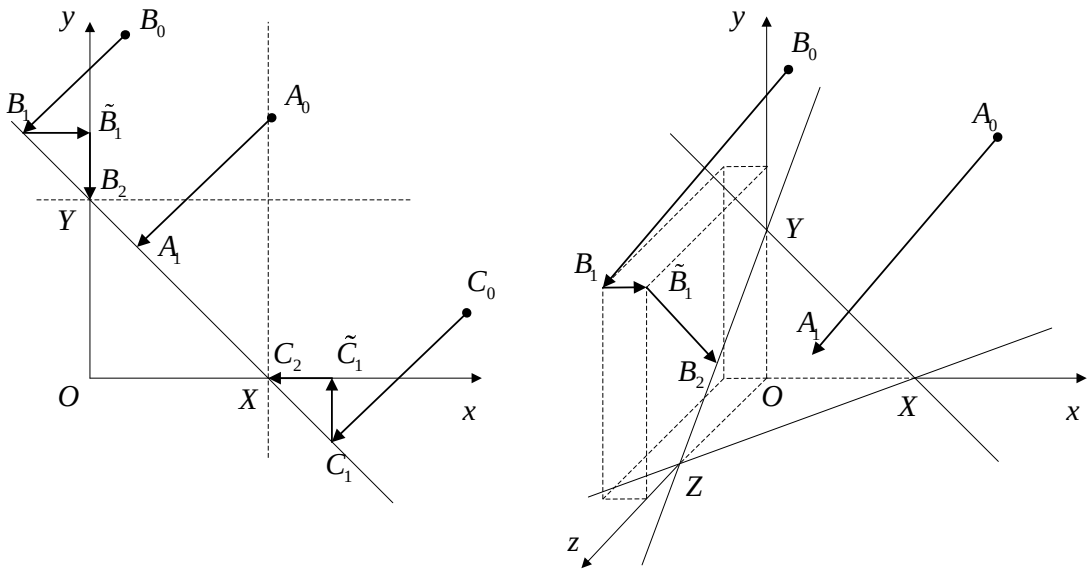

Figure B.10 Examples of projection in 2-D and 3-D vector spaces.

The process of projecting $\boldsymbol{x}_0$ on the set $\mathcal{C}$ is illustrated by examples in Figure B.10. It should be pointed out that all the elements of $\boldsymbol{x}_0$ are always nonnegative.

In a 2-D vector space, $\mathcal{C}$ is defined as the set bounded by the triangle OXY, where the Cartesian coordinates of X and Y are $(b/[\boldsymbol{a}]_1, 0)$ and $(0, b/[\boldsymbol{a}]_2)$, respectively. Now, investigate the projection of 3 points A_0, B_0, and C_0 on $\mathcal{C}$. At first, make the projection of A_0, B_0, and C_0 on $\mathcal{C}_1$, which is the line connecting X and Y, to get to the points A_1, B_1, and C_1. It can be seen that as the two coordinate values of A_1 are positive, i.e., $A_1 \in \mathcal{C}_0$, A_1 is also the true projection of A_0 in $\mathcal{C}$. On the other hand, one of two coordinate values of B_1 and C_1 is negative, B_1 and C_1 cannot be the true projections of B_0 and C_0 on $\mathcal{C}$, respectively.

By following the process described above, the first coordinate value of B_1 is set at 0 to get to the newly defined point $\tilde{B}_1$. Also defined is the half-space $\mathcal{C}_2 = \{\boldsymbol{x} : [\boldsymbol{x}]_2 \leq$

$b/[\boldsymbol{a}]_2$). The next step is to make the projection of $\tilde{B}_1$ on $\mathcal{C}_2$ to get to the point B_2, which is Y. Obviously, B_2 is the projection of B_0 on $\mathcal{C}$ as it is the closest point to B_0. Similarly, the projection of C_0 on $\mathcal{C}$ is done by making the projection of $\tilde{C}_1$ to the point C_2, which is X.

As examples in a 3-D vector space, A_1 is the true projection of A_0 on $\mathcal{C}$, whereas B_1 is not the true projection of B_0 on $\mathcal{C}$. By following the process described above, the first coordinate of B_1 is set at 0 to get to the newly defined point $\tilde{B}_1$. Then, the projection of $\tilde{B}_1$ to the set $\mathcal{C}_2$, which is the line YZ, is taken. The resultant projected point B_2 is also the projection of B_0 on $\mathcal{C}$. It can be verified that B_0B_2 is perpendicular to YZ. Thus, B_2 is the point in XY and also the point in the triangle XYZ that is closest to B_0.

In a higher dimension vector space, the process described above works in the same manner. Numerous simulations show that the proposed algorithm needs only one or two iterations to converge and achieves a much faster processing time compared to that by solving the QP with `cvx` package [48].

C. Rayleigh-Ritz Theorem

Theorem 1. *Rayleigh-Ritz (From [28]): Let* $\boldsymbol{A}$ *be an* $n \times n$ *Hermitian matrix, and let the eigenvalues of* $\boldsymbol{A}$ *be ordered as:*

$$\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq \lambda_n = \lambda_{\max}.$$

Then,

$$\lambda_1 \boldsymbol{x}^{\mathcal{H}} \boldsymbol{x} \leq \boldsymbol{x}^{\mathcal{H}} \boldsymbol{A} \boldsymbol{x} \leq \lambda_n \boldsymbol{x}^{\mathcal{H}} \boldsymbol{x} \quad \textit{for all } \boldsymbol{x} \in \mathbb{C}^n \tag{C.1}$$

$$\lambda_{\max} = \lambda_n = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathcal{H}} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\mathcal{H}} \boldsymbol{x}}. \tag{C.2}$$

$$\lambda_{\min} = \lambda_1 = \min_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathcal{H}} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\mathcal{H}} \boldsymbol{x}}. \tag{C.3}$$

Proof. Since $\boldsymbol{A}$ is Hermitian, by singular value decomposition [28], there exists a $n \times n$ unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{\mathcal{H}}$ with $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. For any $\boldsymbol{x} \in \mathbb{C}^n$, one has

$$\boldsymbol{x}^{\mathcal{H}} \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}^{\mathcal{H}} \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{\mathcal{H}} \boldsymbol{x} = \left(\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}\right)^{\mathcal{H}} \boldsymbol{\Lambda} \left(\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}\right) = \sum_{i=1}^{N} \lambda_i \left| [\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}]_i \right|^2.$$

Since each term $\left| [\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}]_i \right|^2$ is nonnegative, one has

$$\lambda_1 \sum_{i=1}^{N} \left| [\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}]_i \right|^2 \leq \boldsymbol{x}^{\mathcal{H}} \boldsymbol{A} \boldsymbol{x} = \sum_{i=1}^{N} \lambda_i \left| [\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}]_i \right|^2 \leq \lambda_n \sum_{i=1}^{N} \left| [\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}]_i \right|^2.$$

As $\boldsymbol{U}$ is unitary,

$$\sum_{i=1}^{N} \left| [\boldsymbol{U}^{\mathcal{H}} \boldsymbol{x}]_i \right|^2 = \left\| \boldsymbol{U}^{\mathcal{H}} \boldsymbol{x} \right\|^2 = \boldsymbol{x}^{\mathcal{H}} \boldsymbol{U} \boldsymbol{U}^{\mathcal{H}} \boldsymbol{x} = \boldsymbol{x}^{\mathcal{H}} \boldsymbol{x}.$$

Thus,

$$\lambda_1 \boldsymbol{x}^{\mathcal{H}} \boldsymbol{x} \leq \boldsymbol{x}^{\mathcal{H}} \boldsymbol{A} \boldsymbol{x} \leq \lambda_n \boldsymbol{x}^{\mathcal{H}} \boldsymbol{x}.$$

If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ corresponding to the eigenvalue λ_1, then $\boldsymbol{x}^{\mathcal{H}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathcal{H}}\lambda_1\boldsymbol{x} = \lambda_1\boldsymbol{x}^{\mathcal{H}}\boldsymbol{x}$. Thus,

$$\lambda_1 = \min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{x}}.$$

Similarly for λ_n,

$$\lambda_n = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{x}},$$

which is obtained when $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ corresponding to the largest eigenvalue λ_n. □

Corollary 1. *(From [60]) Let $\boldsymbol{A}$ be an $n \times n$ Hermitian matrix, and $\boldsymbol{B}$ be an $n \times n$ positive definite Hermitian. Furthermore, let $\boldsymbol{B}$ be decomposed into Cholesky factors as $\boldsymbol{B} = \boldsymbol{L}\boldsymbol{L}^{\mathcal{H}}$. Then,*

$$\frac{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{B}\boldsymbol{x}} \leq \lambda_{\max} \quad \textit{for all } \boldsymbol{x} \in \mathbb{C}^n, \tag{C.4}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{\mathcal{H}})^{-1}$. The equality holds if $\boldsymbol{x} = c(\boldsymbol{L}_n^{\mathcal{H}})^{-1}\boldsymbol{u}_{\max}$, where c is any non-zero constant and $\boldsymbol{u}_{\max}$ is the norm-1 eigenvector of $\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{\mathcal{H}})^{-1}$ corresponding to $\lambda_{\max}$.

Proof. Let $\tilde{\boldsymbol{x}} = \boldsymbol{L}^{\mathcal{H}}\boldsymbol{x}$, then apply the Ritz-Rayleigh theorem in (C.1), one has

$$\frac{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathcal{H}}\boldsymbol{B}\boldsymbol{x}} = \frac{\tilde{\boldsymbol{x}}^{\mathcal{H}}\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{\mathcal{H}})^{-1}\tilde{\boldsymbol{x}}}{\tilde{\boldsymbol{x}}^{\mathcal{H}}\tilde{\boldsymbol{x}}} \leq \lambda_{\max}.$$

The equality holds if $\tilde{\boldsymbol{x}}$ is an eigenvector of $\boldsymbol{L}^{-1}\boldsymbol{A}(\boldsymbol{L}^{\mathcal{H}})^{-1}$ corresponding to $\lambda_{\max}$. Thus, $\boldsymbol{x} = c(\boldsymbol{L}_n^{\mathcal{H}})^{-1}\boldsymbol{u}_{\max}$.

For a special case of $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}^{\mathcal{H}}$, then $\boldsymbol{L}^{-1}\boldsymbol{a}\boldsymbol{a}^{\mathcal{H}}(\boldsymbol{L}^{\mathcal{H}})^{-1}$ has rank 1. Its only eigenvector is $\boldsymbol{L}^{-1}\boldsymbol{a}$, and its only non-zero eigenvalue is its trace. As a result, the equality holds if $\boldsymbol{x} = c(\boldsymbol{L}^{\mathcal{H}})^{-1}\boldsymbol{L}^{-1}\boldsymbol{a} = c\boldsymbol{B}\boldsymbol{a}$. □

References

[1] D. Tse and P. Viswanath, *Fundamentals of wireless communications*. United Kingdom: Cambridge University Press, 2005.

[2] J. G. Proakis, *Digital communications*. New York: McGraw-Hill Higher Education, 4th ed., 2001.

[3] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Trans. Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[4] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[6] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[7] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity - Part I: System description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003.

[8] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3524–3536, Dec. 2006.

[9] Y. Jing and H. Jafarkhani, "Using orthogonal and quasi-orthogonal designs in wireless relay networks," *IEEE Trans. Inform. Theory*, vol. 53, pp. 4106–4118, Nov. 2007.

[10] K. G. Seddik, A. K. Sadek, A. S. Ibrahim, and K. J. R. Liu, "Design criteria and performance analysis for distributed sapce-time coding," *IEEE Trans. Veh. Technol.*, vol. 57, pp. 2280–2292, Jul. 2008.

[11] Y. Jing and H. Jafarkhani, "Network beamforming using relays with perfect channel information," in *Proc. IEEE Int. Conf. Accoustics, Speech, and Signal Process.*, vol. 3, pp. 473–476, Apr. 2007.

[12] N. Khajehnouri and A. H. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, pp. 3336–3348, Jul. 2007.

[13] T. Quek, M. Win, H. Shin, and M. Chiani, "Optimal power allocation for amplify-and-forward relay networks via conic programming," in *Proc. IEEE Int. Conf. Commun.*, pp. 5058–5063, Jun. 2007.

[14] T. Quek, H. Shin, and M. Win, "Robust wireless relay networks: Slow power allocation with guaranteed QoS," *IEEE J. Select. Topics in Signal Process.*, vol. 1, pp. 700–713, Dec. 2007.

[15] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1907–1916, May 2008.

[16] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "Power allocation and error performance of distributed unitary space-time modulation in wireless relay networks," *IEEE Trans. Veh. Technol., to appear*, 2009.

[17] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "Distributed unitary space-time modulation in partially coherent and noncoherent relay networks," in *Proc. Int. Conf. Signal Process. and Comm. Syst.*, (Gold Coast, Australia), Dec. 2008.

[18] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: Perfomance limits and space-time signal design," *IEEE J. Select. Areas in Commun.*, vol. 22, pp. 1099–1109, Aug. 2004.

[19] S. Yiu, R. Schober, and L. Lampe, "Distributed space-time block coding," *IEEE Trans. Commun.*, vol. 54, pp. 1019–1206, Jul. 2006.

[20] F. Oggier and B. Hassibi, "A coding scheme for wireless networks with no channel information," in *Allerton Conf. Commun. Contr. and Comp.*, pp. 113–117, Sep. 2006.

[21] F. Oggier and B. Hassibi, "Cyclic distributed space-time codes for wireless networks with no channel information," *submitted*, Mar. 2007.

[22] Kiran T. and B. S. Rajan, "Partially-coherent distributed space-time codes with differential encoder and decoder," *IEEE J. Select. Areas in Commun.*, vol. 25, pp. 426–433, Feb. 2007.

[23] H. Mheidat and M. Uysal, "Non-coherent and mismatched-cohernet receivers for distributed STBC with amplify-and-forward relaying," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 4060–4070, Nov. 2007.

[24] T. Wang, Y. Yao, and G. B. Giannakis, "Non-coherent distributed space-time processing for multiuser cooperative transmissions," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3339–3343, Dec. 2006.

[25] S. Yiu, R. Schober, and L. Lampe, "Noncoherent distributed space time block coding," *Proc. IEEE Veh. Technol. Conf.*, vol. 2, pp. 947–951, Sep. 2005.

[26] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Commun. Letters*, vol. 49, pp. 1–4, Jan. 2001.

[27] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1099–1109, Sep. 2000.

[28] R. A. Horn and C. R. Johnson, *Matrix analysis*. New York: Cambridge University Press, 1985.

[29] H. L. Van Trees, *Detection, estimation, and modulation theory, Part I.* New York: Wiley, 1968.

[30] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2383–2399, Sep. 2001.

[31] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes.* Boston: McGraw-Hill, 4th ed., 2002.

[32] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "A novel power allocation scheme for distributed space-time coding," in *Proc. IEEE Int. Conf. Commun.*, (Dresden, Germany), Jun. 2009.

[33] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Foundation and Trends in Networking*, vol. 2, pp. 381–533, Jul. 2008.

[34] M. Kobayashi and X. Mestre, "Impact of CSI on distributed space-time coding in wireless relay networks," *Assilomar Conf. Signals, Syst. and Comp.*, pp. 101–105, Nov. 2007.

[35] J. Stoer and R. Bulirsh, *Introduction to numerical analysis.* New York: Springer, 2002.

[36] P. Gregson, "Using angular dispersion of gradient direction for detecting edge ribbons," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 682–696, Jul. 1993.

[37] S. Boyd and L. Vandenberghe, *Convex optimization.* United Kingdom: Cambridge University Press, 2004.

[38] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels: A unified approach to performance analysis.* New York: Wiley, 2000.

[39] B. Holter and G. E. Oien, "On the amount of fading in MIMO diversity system," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2498–2507, Sep. 2005.

[40] I. S. Gradshteyn and I. M. Ryzhik, *Table of interals, series, and products*. Elsevier Academic Press, 7th ed., 2007.

[41] D. H. N. Nguyen and H. H. Nguyen, "Channel estimation and performance of mismatched decoding in wireless relay networks," in *Proc. IEEE Int. Conf. Commun.*, (Dresden, Germany), Jun. 2009.

[42] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1874–1888, Jul. 2005.

[43] D. H. N. Nguyen, H. H. Nguyen, and T. T. Pham, "Distributed beamforming in multiuser multi-relay networks with guaranteed QoS," in *submitted to Globecom 09*, 2009.

[44] D. H. N. Nguyen, H. H. Nguyen, and H. D. Tuan, "Distributed beamforming in relay-assisted multiuser communications," in *Proc. IEEE Int. Conf. Commun.*, (Dresden, Germany), Jun. 2009.

[45] V. H-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information," *IEEE Trans. Signal Process.*, vol. 56, pp. 4306–4316, Sep. 2008.

[46] S. Berger and A. Wittneben, "Cooperative distributed multiuser MMSE relaying in wireless ad-hoc networks," *Assilomar Conf. Signals, Syst. and Comp.*, pp. 1072–1076, Nov. 2005.

[47] R. Krishna, Z. Xiong, and S. Lambotharan, "A cooperative MMSE relay strategy for wireless sensor networks," *IEEE Signal Process. Letters*, vol. 15, pp. 549–552, 2008.

[48] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software). http://stanford.edu/~boyd/cvx," Aug. 2008.

[49] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 50, pp. 5062–5073, Oct. 2008.

[50] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas in Commun.*, vol. 7, pp. 1341–1347, Sep. 1995.

[51] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, pp. 2646–2660, Jun. 2007.

[52] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Select. Areas in Commun.*, vol. 24, pp. 1439–1451, Aug. 2006.

[53] F. Rashid-Farrokhi, L. Tassiulas, and K. J. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Commun.*, vol. 46, pp. 1313–1323, Oct. 1998.

[54] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, pp. 851–856, May 1999.

[55] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with indiviual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004.

[56] D. H. N. Nguyen and H. H. Nguyen, "SNR maximization and distributed beamforming in multiuser multi-relay networks," in *submitted to Globecom 09*, 2009.

[57] D. H. N. Nguyen and H. H. Nguyen, "Power allocation and distributed beamforming optimization in relay-assisted multiuser communications," in *Proc. Int. Wireless Commun. and Mobile Comp. Conf.*, (Leipzig, Germany), Jun. 2009.

[58] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[59] W. Zhao, G. Leus, and G. B. Giannakis, "Orthogonal design of unitary constellation for uncoded and trellis-coded noncoherent space-time systems," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1319–1327, Jun. 2004.

[60] Z. Yi and I.-M. Kim, "Joint optimization of relay-precoders and decoders with partial channel side information in cooperative networks," *IEEE J. Select. Areas in Commun.*, vol. 25, pp. 447–458, Feb. 2007.