Resource Allocation and Precoding in Cognitive Radio and Relay Networks

by

Hua Mu

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 8, 2012

Keywords: Cognitive radio, Relay network, Precoding, Interference alignment, Multiple input multiple output (MIMO)

Approved by

Jitendra K. Tugnait, James B. Davis Professor of Electrical and Computer Engineering
Soo-Young Lee, Professor of Electrical and Computer Engineering
Shiwen Mao, McWane Associate Professor of Electrical and Computer Engineering

Abstract

In this dissertation, two promising technologies which improve the spectrum efficiency in wireless networks are investigated: cognitive radio (CR) and relay technology. Cognitive radio, as a promising technology to dynamically access spectrum resources, has drawn great research attention recently. It provides a way to further improve the spectrum efficiency by allowing unlicensed radio devices to learn from the radio environment and adapt their transmit and receive parameters. This dissertation addresses the design of unlicensed networks, including their transmission scheme, resource allocation and precoding/beamforming design when multiple antennas are deployed.

Another topic in this dissertation is that of relay networks. By introducing relay stations into the network, multiple benefits can be obtained, such as extended network coverage, improved throughput and higher spectrum efficiency. Also, a relay network makes it possible to enable two-way (even multi-way) transmission among multiple users within the network. In this dissertation, precoding design in multiuser relay networks is discussed. Also, networks based on combined cognitive radio and relay technologies are considered to leverage higher performance, in terms of spectrum efficiency and network throughput.

This dissertation is organized into six chapters. Chapter 1 introduces the background of cognitive radio and cooperative relay networks. In Chapter 2, a soft-decision spectrum sensing concept is proposed and based on which, a joint spectrum sensing, access and power allocation problem is addressed for multi-band cognitive radio networks by means of convex optimization. Chapters 3 and 4 deal with the precoding design in multi-user cognitive relay networks. Chapter 3 considers the multi-way transmission among multiple users and adopts minimum mean square error (MMSE) as the design objective while Chapter 4 considers a two-way relay network and a joint signal and interference alignment algorithm is proposed. In

Chapter 5, a signal group based interference alignment algorithm is proposed for a generalized MIMO Y channel where each user sends messages to all the other users. In Chapter 6, conclusions are drawn and possible future research avenues are highlighted.

Some interesting ideas about how to improve the spectrum efficiency in wireless networks have been proposed and analyzed in this dissertation. The proposed soft-decision spectrum sensing method allows more flexibility in designing the radio access strategies in cognitive radio systems and achieves significantly higher throughput compared with a traditional hard decision spectrum sensing based algorithm. Furthermore, the proposed precoding/beamforming algorithms in the latter part of this dissertation enable concurrent transmission of multiple users within the same frequency band, which can significantly reduce or completely remove the inter-user interference. These technologies make it possible to utilize the limited radio resources more efficiently and therefore can support the ever-increasing demand arising from various wireless devices and applications.

Acknowledgments

First of all, my sincere appreciation goes to my advisor Prof. Jitendra K. Tugnait for his persistent support and inspiring guidance through my doctoral study. His broad and deep expertise in the area of communication and signal processing and professional attitude towards research has been very enlightening and encouraging for me. Without his kind help and support, the fulfilment of this dissertation won't be possible.

I would like to thank my dissertation committee members, Prof. Shiwen Mao and Prof. Soo-Young Lee for their valuable advice on my research work and dissertation. Special thanks go to Prof. Chris Rodger for being the university reader of my dissertation.

I also want to thank all my friends. Their friendship has made the past three year a pleasant journey for me. I have been very luck to have them.

Last but not least, I would like to express my appreciation to my parents and grandmothers for their financial investment in my education as well as their never-ending love, caring and support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cognitive radio is a concept aimed at tackling the spectrum scarcity problem by permitting unlicensed users (secondary users) to access the licensed spectrum as long as interference to the licensed users (primary users) is “tolerable”. This dissertation addresses some design and analysis problems in cognitive radio, with emphasis on spectrum sensing, joint resource allocation and interference management in multi-user cognitive radio networks. The aim is to improve the performance of secondary networks, such as throughput, bit error rate (BER) and mean square error (MSE) of received signal, while keeping the interference to primary users (PU) within a tolerable range.

The concept of cooperative communication and the deployment of relays in the network can significantly increase throughput and extend network coverage. Various network topologies, transmission schemes and techniques have been proposed recently which explore the benefits of relay networks. Among these, two-way relay networks have attracted great research interest for its ability of facilitating the information exchange of two partners. The two-way relaying concept has also been extended to support multi-pair users and multi-way transmission. This dissertation address some key problems in two-way relay networks. Also, this work extends to multi-user two-way relay networks and also supports multi-way transmission of multiple users. A focus in this dissertation is on precoding/beamforming design at transmitter (both relay and end users) when multiple antennas are deployed. Precoding/beamforming is performed from multiple aspects including mean square error (MSE), zero-forcing (ZF) and interference alignment. Also, the effect of imperfect channel state information (CSI), mainly due to the time varying nature of wireless channel, is considered and a robust design is proposed to deal with this problem.

This chapter provides background for the problems investigated in this dissertation and also introduces the motivations for our work.

1.1 Cognitive Radio

With the demand for more bandwidth from widespread wireless applications growing rapidly, spectrum scarcity becomes a major problem which makes further improvement of wireless network's performance difficult. In order to overcome the limitation of traditional spectrum regulation principle which assigns wireless spectrum to licensed users on a long-term basis, the cognitive radio concept has been proposed which allows more flexibility in spectrum usage and therefore increases spectrum efficiency [31], [32], [33]. This concept is motivated by the fact that licensed users access their spectrum non-continuously and thus leave spectrum holes (non occupied spectrum within a certain time) available to other users/applications. Federal Communications Commission (FCC) defines CR as [33]: "Cognitive Radio: A radio or system that senses its operational electromagnetic environment and can dynamically and autonomously adjust its radio operating parameters to modify system operation, such as maximize throughput, mitigate interference, facilitate interoperability, access secondary markets." As an intelligent wireless communication system, cognitive radio should be be aware of its environment, such as the spectrum occupancy by licensed users and should adapt its operating parameters including access probability, transmit power, modulation strategies, etc to make sure that licensed users' network (primary network) can still function well.

Generally, cognitive radio system's deployment doesn't need to modify the primary network and can operate transparently to the licensed users. In such cognitive radio system, unlicensed users can access the spectrum when it is not occupied by licensed users (underlay scheme) or the unlicensed users can dynamically adjust their transmit power such that the interference to the primary users does not exceed a certain level (overlay scheme) [30]. However some recent research work indicates that by allowing some cooperation and/or information exchange between primary and secondary users (SU), secondary network's performance can be further improved without increasing the interference to licensed users. Such system brings together heterogeneous operators and services, various network topologies and radio access technologies and includes the joint management of several networks and reconfiguration of wireless devices such as base stations and terminals. For example, by allowing the secondary network to access the primary channel state information (CSI), secondary precoding/beamforming can be performed at secondary transmitters to completely remove the interference to the primary network by using the secondary-to-primary CSI.

During the past decade, many efforts have been made to design and develop effective cognitive radio technologies. For *obtaining knowledge of cognitive radio's operational and geographical environment*, spectrum sensing, geolocation, etc. have received significant research attention, while for *dynamically and autonomously adjusting its operational parameters and protocols*, many resource management techniques such as dynamically optimizing transmit power and access probabilities of multiple frequency bands as well as beamforming/precoding schemes have been proposed.

1.1.1 Spectrum Sensing

In cognitive radio, spectrum sensing is the essential task to determine the "spectrum holes" which can be used by unlicensed users. The term "spectrum holes" refers to subbands of radio spectrum which are not occupied by licensed users within a certain time and location. Therefore spectrum sensing involves three dimensions: frequency, time and location.

Spectrum sensing can be categorized into three types: energy detection [3], [4] matched filter detection [5] and cyclostationary feature detection [5], [6]. While the latter two can provide more sensing accuracy, energy detection is the most widely used method for its computational simplicity and least requirement on prior knowledge.

Frequently, the spectrum sensing problem is formulated as a binary hypothesis testing problem and secondary users make hard decisions about primary users' behavior (PU is present or not). Then a spectrum access strategy is implemented based on the sensing decision. However, no spectrum sensing technique can detect spectrum holes perfectly. So secondary users' transmissions will always cause collisions with primary users with some nonzero probability. In the binary test based spectrum sensing, information loss is introduced when transforming the continuous sensing statistics into a binary detection decision. When this decision is made separately from other cognitive radio operating parameters, such as spectrum access probabilities (or binary access decisions), transmit power, etc, the secondary users' performance is usually not optimized. Also, in the scenario where secondary users access the channels with some probabilities, it is not reasonable to make the access decisions based on a binary sensing result. Motivated by these, we propose a soft-decision spectrum sensing concept which allows more design flexibility and significantly improves the system throughput.

The accuracy of signal user spectrum sensing is always compromised by multipath fading and shadowing. Deep fading and shadowing can make the primary users' signal at the sensing node very weak and therefore cause mis-detection of PU's presence. Cooperative spectrum sensing is proposed to overcome this problem and also to increase sensing speed. Cooperative sensing can be performed in either centralized or distributed manner [5].

In centralized sensing, a central controller collects sensing information from all or some of the secondary users, makes detection decision and then broadcasts this decision to all secondary users to regulate the secondary traffic. A control channel usually exists between central controller and secondary users to allow information reporting and decision distribution. The subset of secondary users which will report their sensing information can be selected based on their credibility which is usually determined by their channel conditions from primary transmitter. Many decision fusion rules have been developed, including hard information combining, such as AND, OR and M-out-of-N methods, and soft information combining like equal gain combining (EGC), selection combining (SC) and switch and stay combining (SSC). Usually soft information combining provides lower mis-detection and false alarm probabilities but it requires more information shared by participating secondary users.

In distributed sensing, secondary users can share their sensing information with each other but they make spectrum detection and access decisions independently. Therefore the network structure is simplified since no central controller or control channel is needed. In this case, coordinating protocol for information sharing decision distribution among secondary users needs to be designed wisely to avoid collision and reduce protocol overhead and complexity.

1.1.2 Resource Allocation in Cognitive Radio

Cognitive radio introduces several challenges to network resource allocation algorithms. Wireless spectrum is now shared between primary and secondary networks. To secondary network, this resource may only be partially available during certain time interval. Resource allocation of secondary network, generally in terms of channel and power allocation, should consider the interference they cause to the primary network.

When resource allocation is performed based upon the hard decisions of channel availabilities from the spectrum sensing stage, it is similar to traditional resource allocation optimization problems with more constraints on interference to primary users. Some of the recent works extend the resource optimization problem to include both spectrum sensing, access and power allocation. Ref. [26] jointly optimizes the sensing thresholds over a wideband spectrum in both single cognitive radio and cooperative cognitive radio scenarios. Ref. [17] jointly optimizes sensing thresholds and transmit power in a multiband system, and Ref. [22] considers a similar problem in orthogonal frequency division multiplexing (OFDM) scenario. However, given the fact that the statistics collected in the spectrum sensing stage only provides a probability of channel's occupancy, it is preferable to optimize resource allocation problems based on these channel availability probabilities. Therefore soft sensing concept is incorporated into resource allocation stage. Ref. [19] optimized spectrum access probabilities based on quantized sensing statistics aiming at maximizing secondary user's service rate while maintaining primary user's queue stable. In [16] an adaptive power and rate allocation algorithm is developed for a single band cognitive radio system. Ref. [27] considers the optimal power control problem in single band cognitive radio based on soft decision spectrum sensing. In this dissertation, a soft-decision spectrum sensing based resource allocation algorithm is proposed to jointly optimize channel access and transmit power.

1.1.3 Beamforming/Precoding in Cognitive Radio

In underlay cognitive radio networks, primary users and secondary users transmit concurrently over the same spectrum band. One way to avoid interference to licensed PUs is to employ multiple antennas at SU transmitters and design proper precoding matrix. When secondary-to-primary CSI can be perfectly obtained, interference to PU can be completely removed by choosing beamforming vectors that align in the null space of the secondary-to-primary channel matrix. This also restricts the number of antennas at the secondary transmitters. When CSI can not be perfectly obtained due to the time-varying nature of wireless channels or quantization errors, or when number of transmit antennas at SUs are not large enough, a compromise can be made by allowing some residual interference at the PU receivers.

1.2 Cooperative Communication and Relay Networks

In cooperative communication, users exploit the broadcast nature of wireless channels to help others' transmission [12], [13]. Sometimes, dedicated relay stations are deployed to help forward users' information in order to increase multiplexing and/or diversity order, extend network coverage and achieve higher throughput, etc.

In a typical two-hop relay network operating in a half-duplex mode (this is a widely used case since it is practical to assume that each user can not transmit and receive simultaneously), the transmission is completed within two time slots. In the first time slot relay receives signals from all transmitting users while in the secondary time slot the relay processes and forwards its received signal to users. Two commonly employed relay techniques are decode-and-forward (DF) and amply-and-forward (AF). These two techniques refer to different signal processing methods at the relay node. In the DF scheme the relay decodes its received information from the first time slot, re-encodes it and transmits it to desired users in the second time slot. While in the AF scheme, the relay simply amplifies its received signal according to its power budget in the single antenna case or designs a precoding matrix to be multiplied with its received signal vector and then transmits this new signal vector to receivers in the multiple antennas case. The key difference between these two schemes is whether relay decodes the information it receives or not. The advantage of the AF scheme is that it is simpler for the relay since no decoding is required. Also, from the security point of view, the relay will not gain access to users' information since no decoding is performed, therefore the privacy of users' data is preserved.

Relay deployment also enables two-way and multi-way transmission in two-user and multi-user networks since users can exchange information via the help of the relay node. Network coding concept [58] is usually combined with the relay techniques to achieve multiplexing gain due to the fact that each user knows its own signal and can perform self-interference cancellation. For each user, its desired signal and its "self" signal can be network coded together at the relay station which reduces the number of independent datastreams in the network and in turn increases the multiplexing gain.

1.2.1 Multi-user MIMO Relay Network

MIMO is known as an advanced technique to improve data throughput without increasing bandwidth or power [9–11]. In a multi-user relay network, concurrent transmission of multiple users can be supported by deploying multiple antennas at the relay station or by deploying multiple relays which form a virtual multiple antenna node. A multi-user network is usually interference limited. However a proper design of precoding/decoding matrices at transmitters/receivers with multiple antennas can mitigate the negative effects of inter-user interference and therefore enable concurrent transmission of multiple users [14, 48]. Next we discuss three types of beamforming/precoding methods in a multi-user MIMO relay network: minimum mean square error (MMSE), zero-forcing (ZF) and interference alignment.

One way to design precoding/decoding matrices is using mean square error (MSE) of estimated signal as the design criterion. This method considers both inter-user interference and noise as *errors* and tries to minimize them. MSE has been widely utilized in MIMO relay networks. When both relay and end users employ multiple antennas, joint optimization of precoding/decoding matrices at all nodes seems impossible since the problem is non-convex. So suboptimal iterative methods are often utilized. In a two-user two-way relay network, Ref. [14] designs optimal beamforming matrix at the relay node and derives capacity region. Ref. [13] proposes an iterative source and relay precoding algorithm for two-user two-way MIMO relay system. [8], [10] considers precoder design in a system where one pair of users are assisted by multiple relays. While in a two-user two-way relay system inter-user interference can be completely removed, this kind of interference can significantly degrade the performance of a multi-user two-way relay system. For a multi-user two-way relay system, Ref. [6] designs MIMO relay beamforming matrix based on zero-forcing and minimum MSE criteria.

Zero-forcing is another method to eliminate the harmful effect of inter-user interference. By performing ZF, interference is removed completely by forcing interference and desired signal orthogonal subspaces at each receiver. For zero-forcing relay systems, Ref. [55] proposes distributed beamforming for a multi-pair two-way relay system where each node has a single antenna. The inter-pair interference is canceled out by designing relay weights of multiple signal antenna relays. In [1], MIMO technique is employed so that inter-pair interference is nulled out. In [57], a coordinated eigen beamforming method is proposed to optimize the throughput given that the inter-pair interference has already been eliminated.

Interference alignment is a recently emerged technique for the interference-limited network which can dramatically increase the system's degree of freedom (DOF) and therefore supports multi-user transmission [8]. This technique shows that interference is not a prohibitive factor in multi-user networks anymore. The basic idea is to align all interference from different interferers into the same subspace therefore more dimensions are left to useful signals. Many alignment schemes have been proposed for various networks including multi-user networks, cellular networks, MIMO X-channel, relay networks and Y channel, etc. For interference alignment in MIMO Y channel, feasible schemes have been proposed for both 3-user [63] and multi-user scenarios [7]. By incorporating the network coding concept, signals that can be network coded together at the relay node are aligned into the same direction. Therefore a higher DOF is achieved and equivalently, the number of required antennas are reduced. In this dissertation a signal group based alignment method is proposed for a generalized MIMO Y channel, which supports the concurrent transmission of K users to convey $K(K-1)$ datastreams for each other. Our scheme significantly reduces the number at users at the cost that more antennas are deployed at the relay node.

1.3 Overview of this Dissertation

The remainder of this dissertation is organized as follows.

In Chapter 2, a soft-decision spectrum sensing concept is introduced. Instead of making binary hard decisions regarding the channel state, the channel availability probabilities are directly utilized in the optimization problem including spectrum access and power allocation. The benefits of this proposed algorithm is illustrated via simulations. Also some practical implementation issues are discussed to leverage design tradeoffs for practical implementations.

Chapter 3 through Chapter 5 address the precoding/beamforming design problems for two-way and multi-way transmission in relay networks. The aim of these precoding/beamforming methods are to alleviate the harmful effects of inter-user interference and therefore support concurrent transmissions of multiple users utilizing MMSE criterion and interference alignment methods. Chapters 3 and 4 deal with a cognitive radio system where a multi-user relay network is the secondary network which coexists with a primary network. In this kind of system setup, the design of secondary precoding schemes address not only the inter-user interference problem but also the interference avoidance from the secondary network to the primary network. More specially, Chapter 3 investigates the precoding design in a multi-way multi-user MIMO relay secondary network based on the MMSE criterion. Both iterative and low-complexity non-iterative algorithms are presented. Also, the effects of channel uncertainty is taken into consideration and a robust non-iterative algorithm is proposed. In Chapter 4, a joint signal and interference alignment precoding method is proposed for multi-user two-way relay systems. This method can be viewed as a generalization of zero-forcing precoding and provides better performance in the low-to-medium SNR regime. This method incurs low computational complexity while provides high design flexibility.

Chapter 5 considers a generalized MIMO Y channel. This channel includes multiple users and each user intends to transmit to all the other users. By the help of a relay node and the deployment of multiple antennas, the concurrent transmissions of multiple users is supported with no inter-user interference. A novel signal group based interference alignment method is proposed which only requires a minimum number of antennas at each user and a relatively small number of antennas at the relay. Although this chapter considers a non cognitive radio system model, the extension to cognitive radio network is straightforward provided that the channel state information between primary and secondary networks is available.

In Chapter 6, conclusions are provided and future research directions are presented.

Chapter 2

Resource Allocation in Cognitive Radio Network

The proliferation of wireless applications is requiring more and more radio resources. The cognitive radio concept has been proposed to address the spectrum scarcity by relaxing the traditional fixed spectrum regulation and allowing unlicensed users to access spectrum provided they do not degrade licensed users's quality of service. Various techniques have been proposed to ensure that licensed and unlicensed users can share the radio resource. The design of cognitive radio networks is mainly focused on the design of the secondary network since the goal is to increase spectrum efficiency without modifying the operation of the primary network. The most important design criteria for cognitive radio networks can be summarized into two aspects. Firstly, interference to primary network should be avoided. This is usually achieved by spectrum sensing, power control and beamforming, etc. Secondly, the utilization of resources should be optimized to achieve a better performance for the CR network. This requires an optimized design of spectrum access strategies, power allocation and network scheduling protocol, etc. In this chapter, the design of a multiband cognitive radio network is presented in which a soft-decision spectrum sensing method is proposed and spectrum access and power allocation are jointly optimized.

This chapter is organized as follows. In Sec. 2.1, an introduction to the background and related work of resource allocation in CR networks is presented. In Sec. 2.2 we describe the system model, underlying assumptions, optimization objective function and constraints in detail. In Sec. 2.3 we provide a solution to the proposed convex optimization problem via a dual convex optimization formulation. The interference constraint used in this formulation is the total interference to the primary network. An alternative formulation is presented in Sec. 2.4 where additionally interference to individual PUs is also constrained. In order to reduce computational complexity of the optimal solution, two heuristic algorithms are proposed in Sec. 2.5 for the primary network interference formulation. In all our problem formulations knowledge of the CSI (channel state information) for all links (SUs to secondary network base station, and PUs to SUs) is assumed to be available. In Sec. 2.6 some practical/implementation issues involving CSI acquisition and imperfect CSI are briefly discussed. In Sec. 2.7 a hard decision spectrum sensing scheme is presented for comparison purpose following [17]. Numerical results are offered in Sec. 2.8 to illustrate the proposed soft sensing based algorithms and to compare them with the hard sensing approach of Sec. 2.7. Concluding remarks are given in Sec. 2.9 and some technical details for the solution in Sec. 2.3 are provided in the Appendix.

2.1 Introduction

Spectrum sharing is central to cognitive radio which permits unlicensed users (secondary users) to access the licensed spectrum as long as the interference (instantaneous power) to licensed users (primary users) is tolerable. There are two popular spectrum sharing schemes. The first one (spectrum underlay) is to allow primary and secondary users access to the same channel simultaneously while constraining the transmitted power of secondary users so that it can be treated as background noise at primary users without exceeding the primary users noise floor. In the second scheme (spectrum overlay) secondary users need to detect spectrum holes and then access spectrum white space in an nonintrusive way [30]. In this chapter, we will focus on the second scheme.

In this chapter we propose a joint spectrum sensing, access and power allocation algorithm for spectrum overlay schemes where spectrum sensing is of great importance since this is how secondary users find available spectrum holes which could be used. Improving secondary network's performance while keeping the interference to primary network acceptable (tolerable) is an important tradeoff in cognitive radio networks. In traditional spectrum sensing, secondary users make hard decisions about primary users' behavior (PU is present or not). As discussed in [24] and [25] (also [16] and [27]), significant performance improvement can be expected if one employs a soft-decision sensing technique where one uses the continuous-valued sensing statistics directly (explicitly) in joint optimization of secondary network's parameters. On the other hand, a well-designed power allocation strategy is another effective way to further improve secondary users' throughput.

Past work on joint power allocation and spectrum access using hard-decision spectrum sensing includes [19], [26], [17], [22] and [29], and that using soft-sensing includes [16] and [27]. In [16] an adaptive power and rate allocation algorithm is developed for a single band cognitive radio system. Ref. [27] considers optimal power control in single band cognitive radios based on soft decision spectrum sensing. Both [16] and [27] allow secondary users to transmit simultaneously with primary user, which means they are dealing with spectrum underlay system and access probability is not considered. Note that [16] and [27] define interference to primary users to be the average received power at primary users from the secondary users' transmissions. This is different from the "overlay" scenario where the interference is usually defined to be the probability of collisions or percentage of bandwidth corrupted by collisions [26]. Few papers consider power adaptation in overlay systems because under the same spectrum access decision, adapting power does not change the collision probability. Therefore, in such a case, spectrum sense-access can be separated from power allocation. However, joint optimization of spectrum sensing, spectrum access and power adaptation can achieve better sum throughput of secondary users while keeping the interference to primary users under certain tolerable level.

Notation: We will use the variable h with various subscripts to denote the flat-fading channel gain from various transmitters (primary or secondary users) to various receivers (PUs or SUs).

h_k gain of subchannel k from the secondary BS to the SU using subchannel k during the transmission phase (after sensing has been done).

h_{pk} gain of subchannel k from the PU using subchannel k to the SU also using subchannel k during the transmission phase (after sensing has been done).

$h_{pi,k}$ gain of subchannel k from the PU using subchannel k to the SU sensing subchannel k during the sensing phase.

In the following the CSI for all links (i.e. h_k, h_{pk} and $h_{pi,k}$) is assumed to be available to all nodes of the secondary network (if needed). We will use $1_{\{*\}}$ to denote an indicator function which takes value 1 when $\{*\}$ is true and 0 when $\{*\}$ is false. $(*)^+ := \max\{0, *\}$. If random variable X is Gaussian with mean μ and variance σ^2, we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

2.2 System Model and Problem Formulation

2.2.1 System Model

Consider a cognitive radio scenario where secondary users can potentially access K orthogonal subchannels. One of the secondary users is designated as cognitive radio base station (BS); we will refer to this node as BS. [Alternatively, the BS instead of being just one of many users in the network, could be a single secondary user which transmits to multiple users.] There are also N secondary users in the system. The secondary BS determines spectrum access probabilities and allocates resources to other secondary users while both secondary BS and users participate in spectrum sensing. We consider a downlink scenario for the secondary network where the BS transmits to the secondary users and the association between secondary users and subchannels are preassigned. Assume that the subchannels experience block fading, which means the channel gains of these subchannels remain constant during one time slot. Also assume the aggregated primary users' behavior over each channel follows a block static pattern, i.e. during each time slot, primary users will access subchannel k, $(k = 1, 2, ..., K)$, with probability $P_{H_{1k}}$ and remain idle with probability $P_{H_{0k}} = 1 - P_{H_{1k}}$ where H_1 denotes the fact that primary user is transmitting while H_0 indicates the alternative. Primary users' behavior over the K subchannels are assumed to be independent.

2.2.2 Spectrum Sensing

Spectrum sensing is performed at the beginning of each time slot. Secondary BS and secondary users (or a subset of secondary users) sense the K subchannels and collect sufficient sensing metrics for each channel. Any sensing technique can be used, such as the commonly used energy detector [20, 23]. Let $\mathcal{A}_k,\ k = 1, 2, ..., K$, denote the sensing user subset for subchannel k where $\mathcal{A}_k \subseteq \{0, 1, ..., N\}$ with $m = 1, 2, \cdots, N$ indexing one of N secondary users and $m = 0$ indexing secondary BS. Let $Q_{ik}, i \in \mathcal{A}_k, k = 1, 2, ..., K$, be the sensing statistic for subchannel k from secondary user i. The conditional probability density function of Q_{ik} conditioned on primary users' behavior (transmitting or silent) are assumed to be known and denoted as $f(Q_{ik}|H_{0k})$ and $f(Q_{ik}|H_{1k})$. Let $\mathbf{Q}_k = [Q_{ik}, i \in \mathcal{A}_k]$ be the sensing metric vector for subchannel k. With the knowledge of $\mathbf{Q}_k$'s distribution and primary users' access/idle probabilities $P_{H_{1k}}$ / $P_{H_{0k}}$, the probabilities of H_0 and H_1 for subchannel k conditioned on $\mathbf{Q}_k$ is given by

$$P(H_{jk}|\mathbf{Q}_k) = \frac{f(\mathbf{Q}_k|H_{jk})P_{H_{jk}}}{f(\mathbf{Q}_k|H_{0k})P_{H_{0k}} + f(\mathbf{Q}_k|H_{1k})P_{H_{1k}}}, \quad j \in \{0, 1\}, \quad k = 1, 2, ..., K, \tag{2.1}$$

where $f(\mathbf{Q}_k|H_{jk}) = \prod_{i \in \mathcal{A}_k} f(Q_{ik}|H_{jk}),\ j \in \{0, 1\}$. An example of Q_{ik}'s for energy detectors and resultant probability densities may be found in Sec. 2.8.

2.2.3 Problem Formulation

After the sensing phase, the secondary BS will jointly optimize spectrum access and power allocation. Assume that during each time slot, channel conditions (flat fading gains) of secondary links (assumed to be flat fading), denoted as $\mathbf{h} = [h_1, h_2, ..., h_K]$, can be obtained using channel estimation techniques. Let $\mathbf{s} = [s_1, s_2, ..., s_K]$ and $\mathbf{p} = [p_1, p_2, ..., p_K]$ denote the access probabilities and allocated powers, respectively, over the K subchannels, that is, s_k is the probability with which the secondary BS accesses the k-th subchannel (recall we are considering a downlink scenario for the secondary network). We allow $s_k \in [0, 1]$ because we believe that this will increase secondary network's throughput, otherwise the optimal value of s_k will end up being either 0 or 1. Our goal is to maximize the expectation of sum throughput (instantaneous capacity) of secondary network while keeping the interference to primary users under some bound ε, under a transmit power constraint.

Let W_k represent the bandwidth of the k-th subchannel and N_0 represent the (one-sided) power spectrum density of white noise at secondary receivers. The transmission rate of secondary BS over subchannel k consists of two parts. One part is the transmission rate when primary user is not transmitting over this channel and the other part is the transmission rate when primary user is also transmitting. Let h_{pk} denote the gain of subchannel k from PU transmitter to secondary receiver and let p_{puk} be the PU's transmitted power in subchannel k. Then the maximum transmission rate (instantaneous capacity) over subchannel k is given by

$$R(p_k, s_k) = s_k P(H_{0k}|\mathbf{Q}_k) W_k \log_2 \left(1 + \frac{p_k |h_k|^2}{N_0 W_k}\right)$$

$$+ s_k P(H_{1k}|\mathbf{Q}_k) W_k \log_2 \left(1 + \frac{p_k |h_k|^2}{N_0 W_k + p_{puk}|h_{pk}|^2}\right) \tag{2.2}$$

where the first term on the right-side of (2.2) is the instantaneous capacity when PU is not transmitting while the second term is the instantaneous capacity when PU is transmitting, $p_{puk}|h_{pk}|^2$ representing the interference power at the secondary receiver from PU transmitter over subchannel k, and $P(H_{0k}|\mathbf{Q}_k)$ and $P(H_{1k}|\mathbf{Q}_k)$ are the posterior probability of PU not transmitting and transmitting, respectively, in the k-th subchannel conditioned on the subchannel sensing statistic $\mathbf{Q}_k$. Interference over subchannel k occurs when the secondary BS's decision is to transmit over this subchannel and the PU is also transmitting over the same subchannel. Defining the overall interference to be the average fraction of bandwidth corrupted by secondary BS's transmissions, we have the interference expression for subchannel

k as

$$I(s_k) = s_k P(H_{1k}|\mathbf{Q}_k) W_k. \tag{2.3}$$

Introduce variables

$$t_k = s_k p_k, \quad u_k = |h_k|^2/(N_0 W_k), \quad v_k = |h_k|^2/(N_0 W_k + p_{puk}|h_{pk}|^2), \tag{2.4}$$

and let $\mathbf{t} = [t_1, t_2, ..., t_K]$. Then we have $p_k = t_k/s_k$. Then the optimization problem under consideration can be formulated as a convex optimization problem [18, 21]:

$$\begin{aligned} \text{P1}: \quad \max_{\mathbf{t},\mathbf{s}} \quad & \sum_{k=1}^{K} R(s_k, t_k) && (2.5)\\ \text{s.t.} \quad & \sum_{k=1}^{K} I(s_k) \le \varepsilon && (2.6)\\ & \sum_{k=1}^{K} t_k \le P_{tot} && (2.7)\\ & 0 \le s_k \le 1, \quad t_k \ge 0, \quad k = 1, 2, \cdots, K, && (2.8) \end{aligned}$$

where

$$R(s_k, t_k) = s_k P(H_{0k}|\mathbf{Q}_k) W_k \log_2\left(1 + \frac{t_k u_k}{s_k}\right) + s_k P(H_{1k}|\mathbf{Q}_k) W_k \log_2\left(1 + \frac{t_k v_k}{s_k}\right). \tag{2.9}$$

In the above problem, (2.6) is the total interference constraint with a pre-specified bound ε, and (2.7) is the total transmit power constraint with upper limit P_{tot}. As discussed in [24] and [25], direct maximization of (2.5) w.r.t. $\mathbf{p}$ and $\mathbf{s}$ does not lead to a convex problem since the constraint $\sum_{k=1}^{K} t_k = \sum_{k=1}^{K} s_k p_k \le P_{tot}$ is not convex in $\mathbf{p}$ and $\mathbf{s}$. The objective function (2.5) in problem P1 is a concave function of variables $\mathbf{t}, \mathbf{s}$ (a proof is given in Appendix A) and all the constraints are affine in the variables $\mathbf{t}$ and $\mathbf{s}$. Also, it is obvious that the feasible set of $\mathbf{t}$ and $\mathbf{s}$ are non-empty since we can always find non negative $\mathbf{t}$ and $\mathbf{s}$ such that all the constraints are satisfied. Therefore this is a convex problem [18, Sec. 4.2.1] and the global optimal solution can be obtained. This solution is discussed next.

2.3 Optimization Solution

We now discuss solution to the convex problem P1. We will solve it via a dual formulation. Since objective function (2.5) is concave, all constraints are affine, and the feasible set is non-empty, Slater's condition is satisfied and there is no duality gap [18, Sec. 5.2.3].

Let λ and μ be the dual variables associated with constraints (2.6) and (2.7). The Lagrangian of problem P2 can be expressed as

$$\begin{aligned} J(\lambda,\mu,\mathbf{s},\mathbf{t}) &= \sum_{k=1}^{K} s_k P(H_{0k}|\mathbf{Q}_k) W_k \log_2\left(1+\frac{t_k u_k}{s_k}\right) + s_k P(H_{1k}|\mathbf{Q}_k) W_k \log_2\left(1+\frac{t_k v_k}{s_k}\right) \\ &\quad +\lambda\left(\varepsilon - \sum_{k=1}^{K} s_k P(H_{1k}|\mathbf{Q}_k) W_k\right) + \mu\left(P_{tot} - \sum_{k=1}^{K} t_k\right) \qquad (2.10) \\ &= \varepsilon\lambda + \mu P_{tot} + \sum_{k=1}^{K} J_k(\lambda,\mu,s_k,t_k) \qquad (2.11) \end{aligned}$$

where

$$\begin{aligned} J_k(\lambda,\mu,s_k,t_k) &= s_k W_k P(H_{0k}|\mathbf{Q}_k)\log_2\left(1+\frac{t_k u_k}{s_k}\right) + s_k W_k P(H_{1k}|\mathbf{Q}_k)\log_2\left(1+\frac{t_k v_k}{s_k}\right) \\ &\quad -\lambda s_k P(H_{1k}|\mathbf{Q}_k) W_k - \mu t_k. \qquad (2.12) \end{aligned}$$

Let χ denote the domain of problem P1, therefore, $\chi = \cap_{k=1}^{K}\chi_k$ where $\chi_k = \{t_k \geq 0,\ 0 \leq s_k \leq 1\}$. Then the Lagrange dual function can be expressed as

$$g(\lambda,\mu) = \max_{(\mathbf{s},\mathbf{t})\in\chi} J(\lambda,\mu,\mathbf{s},\mathbf{t}) = \varepsilon\lambda + \mu P_{tot} + \sum_{k=1}^{K} \max_{(s_k,t_k)\in\chi_k} J_k(\lambda,\mu,s_k,t_k) \qquad (2.13)$$

and the dual optimization problem is

$$\text{P2}: \quad \min_{\lambda,\mu} g(\lambda,\mu) \tag{2.14}$$

$$\text{s.t.} \quad \lambda \geq 0, \quad \mu \geq 0. \tag{2.15}$$

From (4.26) we have

$$g(\lambda,\mu) \;=\; \varepsilon\lambda + \mu P_{tot} + \sum_{k=1}^{K} g_k(\lambda,\mu) \tag{2.16}$$

where

$$g_k(\lambda,\mu) = \max_{s_k,t_k} J_k(\lambda,\mu,s_k,t_k). \tag{2.17}$$

First we calculate $g(\lambda,\mu)$ for fixed $\lambda \geq 0$ and $\mu \geq 0$ and then solve problem P2. From (2.16), we observe that calculation of $g(\lambda,\mu)$ can be decomposed into solving for $g_k(\lambda,\mu)$ for each subchannel k. Note that $J_k(.)$ is concave in t_k for $t_k \geq \max\{-s_k/u_k, -s_k/v_k\} = -s_k/u_k$ and second order differentiable. Also, we have $\frac{\partial J_k(\lambda,\mu,s_k,t_k)}{\partial t_k}|_{t_k=-s_k/u_k} = \infty > 0$ and $\frac{\partial J_k(\lambda,\mu,s_k,t_k)}{\partial t_k}|_{t_k=\infty} = -\mu < 0$. So there exists exactly one $t_k^* \geq -s_k/u_k$ with $\frac{\partial J_k(\lambda,\mu,s_k,t_k)}{\partial t_k}|_{t_k=t_k^*} = 0$ which maximizes $J_k(\lambda,\mu,s_k,t_k)$. To solve for $g_k(\lambda,\mu)$, we take partial derivative of $J_k(\lambda,\mu,s_k,t_k)$ with respect to t_k and set $\frac{\partial J_k(\lambda,\mu,s_k,t_k)}{\partial t_k} = 0$ to obtain two solutions

$$\tilde{t}_{k_1} = s_k \left(\frac{-b_k - \sqrt{b_k^2 - 4u_k v_k c_k}}{2u_k v_k} \right) \tag{2.18}$$

and

$$\tilde{t}_{k_2} = s_k \left(\frac{d_k}{2u_k v_k} \right) \tag{2.19}$$

where

$$b_k := u_k + v_k - \frac{W_k u_k v_k}{\mu \ln 2}, \tag{2.20}$$

$$c_k := 1 - \frac{(P(H_{0k}|\mathbf{Q}_k)u_k + P(H_{1k}|\mathbf{Q}_k)v_k)\,W_k}{\mu \ln 2}, \quad k = 1, 2, ..., K, \tag{2.21}$$

$$d_k := -b_k + \sqrt{b_k^2 - 4u_k v_k c_k}. \tag{2.22}$$

If $t_k^* = \tilde{t}_{k_1}$, then since $\tilde{t}_{k_2} \geq \tilde{t}_{k_1} \geq -s_k/u_k$ and $\frac{\partial J_k(.)}{\partial t_k}|_{t_k=\tilde{t}_{k_2}} = 0$, it follows that $\tilde{t}_{k_2}$ is another point which maximizes $J_k(.)$. This contradicts the fact that exactly one $t_k^* \geq -s_k/u_k$ maximizes $J_k(.)$. Hence, $\tilde{t}_{k_2}$ is the only possible value to maximize $J_k(.)$; therefore, the value of $t_k \in \chi_k$ that maximizes J_k for a fixed $s_k \in [0, 1]$ is given by

$$t_k^* = s_k \left(\frac{d_k}{2u_k v_k}\right)^+ \tag{2.23}$$

where $(*)^+ := \max\{0, *\}$. Since $t_k = s_k p_k$, (2.23) allows us to obtain the optimal power allocation for subchannel k as

$$p_k^* = \left(\frac{d_k}{2u_k v_k}\right)^+ \tag{2.24}$$

even when $s_k = 0$. Then we have

$$\begin{aligned} J_k(\lambda, \mu, s_k, t_k^*) = \; & s_k \Bigg\{ P(H_{0k}|\mathbf{Q}_k) W_k \left[\log_2\left(1 + \frac{d_k}{2v_k}\right)\right]^+ + P(H_{1k}|\mathbf{Q}_k) W_k \left[\log_2\left(1 + \frac{d_k}{2u_k}\right)\right]^+ \\ & -\mu \left(\frac{d_k}{2u_k v_k}\right)^+ - \lambda P(H_{1k}|\mathbf{Q}_k) W_k \Bigg\}. \end{aligned} \tag{2.25}$$

Define

$$\begin{aligned} \beta_k := \; & P(H_{0k}|\mathbf{Q}_k) W_k \left[\log_2(1 + \frac{d_k}{2v_k})\right]^+ + P(H_{1k}|\mathbf{Q}_k) W_k \left[\log_2(1 + \frac{d_k}{2u_k})\right]^+ \\ & -\mu \left(\frac{d_k}{2u_k v_k}\right)^+ \end{aligned} \tag{2.26}$$

Then maximizing $J_k(\lambda, \mu, s_k, t_k^*)$ w.r.t. $s_k \in \chi_k$ leads to

$$s_k^* = \begin{cases} 1 & \text{if} \quad \beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k > 0 \\ 0 & \text{if} \quad \beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k < 0 \\ \in [0,1] & \text{if} \quad \beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k = 0. \end{cases} \tag{2.27}$$

Therefore,

$$g_k(\lambda, \mu) = J_k(\lambda, \mu, s_k^*, t_k^*) = (\beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k)^+ . \tag{2.28}$$

We optimize problem P2 (2.14)-(2.15) with respect to λ first. Define

$$a_k := P(H_{1k}|\mathbf{Q}_k)W_k, \tag{2.29}$$

which leads to $g_k(\lambda, \mu) = (\beta_k - \lambda a_k)^+$. Let

$$g(\mu) = \min_{\lambda \geq 0} g(\lambda, \mu) = \min_{\lambda \geq 0} \left\{ \sum_{k=1}^{K} g_k(\lambda, \mu) + \varepsilon\lambda \right\} + \mu P_{tot} \tag{2.30}$$

$$= \min_{\lambda \geq 0} \varepsilon \left\{ \sum_{k=1}^{K} \frac{a_k}{\varepsilon} \left(\frac{\beta_k}{a_k} - \lambda \right)^+ + \lambda \right\} + \mu P_{tot}. \tag{2.31}$$

Proposition 2.1 : Sort β_k / a_k in a non-increasing order of magnitude and denote the sorted sequence as $\{\tilde{d}_1, \tilde{d}_2, ..., \tilde{d}_K\}$, and also re-order a_k in the same order and denote the re-ordered sequence as $\{\tilde{a}_k, \ \ k = 1, 2, \cdots, K\}$. Then (2.31) is minimized w.r.t. $\lambda \geq 0$ for

$$\lambda^* = \begin{cases} \tilde{d}_{l^*} & \text{if } 1 - \sum_{k=1}^{l^*-1} \frac{\tilde{a}_k}{\varepsilon} \geq 0 \text{ and } 1 - \sum_{k=1}^{l^*} \frac{\tilde{a}_k}{\varepsilon} < 0 \text{ for some } l^* \leq K \\ 0 & \text{if } 1 - \sum_{k=1}^{K} \frac{\tilde{a}_k}{\varepsilon} \geq 0 \end{cases} \tag{2.32}$$

where, by definition, $\sum_{k=1}^{0} \frac{\tilde{a}_k}{\epsilon} = 0$.

Proof : See Appendix B. □

Then we have

$$g(\mu) = g(\lambda^*, \mu) = \mu P_{tot} + \varepsilon\lambda^* + \sum_{k=1}^{l^*}(\tilde{a}_k\tilde{d}_k - \tilde{a}_k\lambda^*)^+ \tag{2.33}$$

where $l^* = K$ for the second case in (2.32). Since $g(\lambda, \mu)$ is a convex function, $g(\mu) = \min_{\lambda\geq 0} g(\lambda, \mu)$ is also convex. Therefore, $\min_{\mu\geq 0} g(\mu)$ can be obtained via a line search. When we search for μ, the lower bound can be set as $\mu_l = 0$. As to the upperbound, we need to make sure that BS will allocate a positive power to at least one subchannel. From (2.24), we choose $\mu_u = \max\{(P(H_{0k}|\mathbf{Q}_k)u_k + P(H_{1k}|\mathbf{Q}_k)v_k)W_k/\ln 2, k = 1, 2, ..., K\}$.

After we obtain the optimal dual variables μ^* and λ^*, we can recover the optimal primary variables according to (2.24) and (2.27). Note that the optimal solution s_k will take a value between 0 and 1 only when $\beta_k - \lambda^* P(H_{1k}|Q_k)W_k = 0$. From Proposition 2.1, it follows that $\beta_{i^*} - \lambda P(H_{1i^*}\,|\,Q_{i^*})W_{i^*} = 0$ where i^* corresponds to l^* pertaining to re-ordered β_k/a_k in (2.32). However, we may have $\beta_i/a_i = \beta_{i^*}/a_{i^*}$ for some $i \neq i^*$. Define the sets

$$\mathcal{M} = \{i\,|\,\frac{\beta_i}{a_i} = \frac{\beta_{i^*}}{a_{i^*}},\ i = 1, 2, ...K\} \tag{2.34}$$

and

$$\mathcal{M}^c = \{i\,|\,\frac{\beta_i}{a_i} \neq \frac{\beta_{i^*}}{a_{i^*}},\ i = 1, 2, ...K\}. \tag{2.35}$$

Then for any $k \in \mathcal{M}$, we have $s_k^* \in [0, 1]$. If $|\mathcal{M}| = 1$, then as in [24] (see discussion therein after Proposition 2.1), we have only one sub-channel, i.e., the i^*-th sub-channel whose access probability $s_{i^*}^* \in [0, 1]$, and we pick the largest s_{i^*} satisfying all the constraints and then set $p_k^* = \left(\frac{d_k}{2u_k v_k}\right)^+$ for $s_k^* > 0$, and $p_k^* = 0$ for $s_k^* = 0$. When $|\mathcal{M}| > 1$, the optimal solution should satisfy the complementary slackness conditions [24, (29),(30)] since strong duality holds. If $\lambda^* = 0$, then constraint (2.6) is never active. Therefore, we can access all subchannels with probability 1. Also, from (2.20)-(2.22) and (2.24), we notice that $\mu^* > 0$, otherwise as long as we access some subchannel, the total transmitted power will tend to infinity. For $\lambda^* > 0$

and $\mu^* > 0$, constraints (2.6) and (2.7) hold true with equality (this is [24, (31)]). For $i \in \mathcal{M}^c$, s_i^* are determined using (2.27). We now address how to compute the optimal value for $s_i \in [0,1]$, $i \in \mathcal{M}$, which satisfy (2.6) and (2.7) with equality. Define the set

$$\mathcal{D} = \left\{ \tilde{\mathbf{s}} \;\middle|\; (\tilde{\mathbf{s}}, \tilde{\mathbf{t}}) = \arg\max_{(\mathbf{s},\mathbf{t})\in\chi} J(\lambda^*, \mu^*, \mathbf{s}, \mathbf{t}),\ \sum_{k=1}^{K} \tilde{s}_k P(H_{1k}|\mathbf{Q}_k) W_k = \varepsilon \right\}; \tag{2.36}$$

recall also from (4.26) that $\max_{(\mathbf{s},\mathbf{t})\in\chi} J(\lambda^*, \mu^*, \mathbf{s}, \mathbf{t}) = g(\lambda^*, \mu^*)$. Let $\tilde{p}_k = p_k^*$ as in (2.24) with $\lambda = \lambda^*$ and $\mu = \mu^*$. Then for $\tilde{\mathbf{s}} \in \mathcal{D}$, the corresponding $\tilde{\mathbf{t}}$ are determined as $\tilde{t}_k = \tilde{s}_k \tilde{p}_k$, $k = 1, 2, ..., K$, and $\partial_\mu g(\lambda^*, \mu^*)|_{\tilde{\mathbf{s}}} = P_{tot} - \sum_{k=1}^{K} \tilde{s}_k \tilde{p}_k$ is a subgradient of $g(\lambda, \mu)$ w.r.t. μ evaluated at (λ^*, μ^*). We must have one of these subgradients w.r.t. μ to be 0 since (λ^*, μ^*) denotes the optimal solution which minimizes $g(\lambda, \mu)$. Let

$$\mathbf{s}_1 = \arg\max_{\mathbf{s}\in\mathcal{D}} \left\{ P_{tot} - \sum_{k=1}^{K} s_k \tilde{p}_k \right\} \tag{2.37}$$

$$\mathbf{s}_2 = \arg\min_{\mathbf{s}\in\mathcal{D}} \left\{ P_{tot} - \sum_{k=1}^{K} s_k \tilde{p}_k \right\}. \tag{2.38}$$

Then we have $\partial_\mu g(\lambda^*, \mu^*)|_{\mathbf{s}_1} \geq 0$ and $\partial_\mu g(\lambda^*, \mu^*)|_{\mathbf{s}_2} \leq 0$. Therefore, there exists a scalar γ, $0 \leq \gamma \leq 1$, such that for $\mathbf{s}' = \mathbf{s}_1 + \gamma(\mathbf{s}_2 - \mathbf{s}_1)$, we have $\mathbf{s}' \in \mathcal{D}$ and $\partial_\mu g(\lambda^*, \mu^*)|_{\mathbf{s}'} = 0$. Let $t_k^* = s_k' \tilde{p}_k$, $k = 1, 2, ..., K$. Since $(\mathbf{s}', \mathbf{t}^*)$ is a solution to $\arg\max_{(\mathbf{s},\mathbf{t})\in\chi} J(\lambda^*, \mu^*, \mathbf{s}, \mathbf{t})$, primary feasible, and satisfies (2.6) and (2.7) with equality, it is an optimal solution of problem P2. Let $p_k^* = \tilde{p}_k$ for $s_k' > 0$ and let $p_k^* = 0$ for $s_k' = 0$. This yields optimal primary variables $\mathbf{s}^* = \mathbf{s}'$ and $\mathbf{p}^*$.

Remark 1. In the preceding developments, we have provided an algorithm to find an optimal solution for $\boldsymbol{p}$ and $\mathbf{s}$. This optimal solution is not necessarily unique even though the solution for $\mathbf{t}$ is unique as stated in (2.23). Note that t_k^* in (2.23) is a function of yet to be determined s_k. For example, consider a special case in which all secondary links have the same channel gain and the channel available probabilities over all sub-channels are the same. By (2.23), this would lead to identical t_k^*/s_k (for $s_k \neq 0$), $k = 1, 2, \cdots, K$. After the optimal dual variables μ^* and λ^* are obtained, the optimal power allocation for each sub-channel will be as in (2.24), identical for $k = 1, 2, \cdots, K$. Then the remaining task is to determine the optimal channel access probabilities s_k^*. If $\lambda^* = 0$, then it implies that the interference constraint is not active, therefore the secondary BS can access all sub-channels with probability 1. For the case where $\lambda^* > 0$, (2.6) should be satisfied with equality. Since p_k^* are the same for all sub-channels, all $\mathbf{s} \in \mathcal{D}$ ($\mathcal{D}$ is defined in (2.36)) should satisfy $\sum_{k=1}^{K} s_k p_k^* = P_{tot}$. Therefore, any $\mathbf{s} \in \mathcal{D}$ is an optimal solution. Thus, p_k^* is unique whereas s_k^* may not be. □

Actual throughput (instantaneous capacity) of the secondary network is then given by

$$\sum_{k=1}^{K} 1_{H_{0k}} s_k^* W_k \log_2(1 + p_k^* u_k) + 1_{H_{1k}} s_k^* W_k \log_2(1 + p_k^* v_k) \tag{2.39}$$

where $1_{\{*\}}$ is an indicator function which takes value 1 when $\{*\}$ is true and 0 when $\{*\}$ is false. Actual interference to primary users caused by secondary BS's transmission is given by

$$\sum_{k=1}^{K} 1_{H_{1k}} s_k^* W_k. \tag{2.40}$$

Remark 2. We notice from (2.27) that most channel access probabilities turn out to be 0 or 1. However, this is still different from hard decision spectrum sensing where one makes decisions solely based on spectrum sensing statistics. In our proposed algorithm, channel access decisions are make jointly with power allocation based on channel availabilities as well as secondary channel conditions. Eqn. (2.27) shows that even when two channels have exactly the same busy and idle probabilities $P(H_{0k}|\mathbf{Q}_k)$ and $P(H_{1k}|\mathbf{Q}_k)$, the one with better channel condition (u_k and v_k defined in (2.4)) is more likely to be accessed by the secondary BS. Also, from (2.24) we observe that the optimal power allocation is a function of the continuous-valued sensing statistics $\mathbf{Q}_k$ for $k = 1, 2, ..., K$, which is different from hard spectrum sensing where all that matters is whether the sufficient sensing metric is larger than a threshold or not. □

This optimal algorithm is summarized as follows:

ALGORITHM OPT

1) Initialize μ_l, μ_u and set $\mu = (\mu_l + \mu_u)/2$.

2) While $(\partial_\mu g(\lambda, \mu) \neq 0)$

   Calculate λ^* given the current μ according to Proposition 2.1.

   if $|\mathcal{M}| = 1$, calculated p_k^* and s_k^* as in (2.24) and (2.27).

   if $|\mathcal{M}| > 1$, find $\mathbf{s}_1$ and $\mathbf{s}_2$ as in (2.37) and (2.38) and let $\mathbf{s}^* = \mathbf{s}' = \mathbf{s}_1 + \gamma(\mathbf{s}_2 - \mathbf{s}_1)$,

   where γ is chosen such that (2.6) and (2.7) are satisfied with equality.

   If $(\partial_\mu g(\lambda, \mu) = P_{tot} - \sum_{k=1}^{K} s_k^* p_k^*) > 0$, set $\mu_u = \mu$; otherwise set $\mu_l = \mu$.

   End while

3) Set $p_k^* = \left(\frac{d_k}{2u_k v_k}\right)^+$ for $s_k^* > 0$, and $p_k^* = 0$ for $s_k^* = 0$.

2.4 Alternative Formulation with Individual Interference Constraints

In Sec. 2.2.3 we define the interference to primary network (see (2.6)) to be the total fraction of bandwidth that is corrupted by secondary network's transmission. Here we consider the case where the interference over each sub-channel is also required to be bounded. This is realistic when a primary user may occupy no more than one sub-channel simultaneously. In this case we need to guarantee that no single primary user is interfered with too much. Let ε_k be the interference bound for sub-channel k. Then the soft-decision spectrum sensing based optimization problem with an individual interference constraint is formulated as follows:

$$\begin{aligned}
\text{P3}: \quad \max_{\mathbf{t},\mathbf{s}} \quad & \sum_{k=1}^{K} R(s_k, t_k) && (2.41)\\
\text{s.t.} \quad & \sum_{k=1}^{K} I(s_k) \leq \varepsilon && (2.42)\\
& I(s_k) \leq \varepsilon_k, \quad k = 1, 2, ..., K && (2.43)\\
& \sum_{k=1}^{K} t_k \leq P_{tot} && (2.44)\\
& 0 \leq s_k \leq 1, \quad t_k \geq 0, \quad k = 1, 2, \cdots, K. && (2.45)
\end{aligned}$$

In order for (2.42) and (2.43) to make sense, we assume that $\sum_{k=1}^{K} \varepsilon_k > \varepsilon$ else (2.42) is redundant.

From (2.3), we notice that the individual interference constraint (2.43) is actually an upper bound of each channel access probability. Let $\tilde{\varepsilon}_k = \min\{1, \frac{\varepsilon_k}{P(H_{1k}|\mathbf{Q}_k)W_k}\}$. Then problem $P3$ becomes

$$\begin{aligned}
\text{P4}: \quad \max_{\mathbf{t},\mathbf{s}} \quad & \sum_{k=1}^{K} R(s_k, t_k) && (2.46)\\
\text{s.t.} \quad & \sum_{k=1}^{K} I(s_k) \leq \varepsilon && (2.47)\\
& \sum_{k=1}^{K} t_k \leq P_{tot} && (2.48)\\
& 0 \leq s_k \leq \tilde{\varepsilon}_k, \quad t_k \geq 0, \quad k = 1, 2, \cdots, K. && (2.49)
\end{aligned}$$

Similar to problem $P1$, this is also a convex problem which can be solved by a Langrangian dual formulation. Let μ and λ be the dual variables associated with constraints (2.47) and (2.48) respectively. Then for a given μ and λ the optimal power allocation has a same expression as in (2.24). The optimal channel access probabilities become

$$s_k^* = \begin{cases} \tilde{\varepsilon}_k & \text{if} \quad \beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k > 0 \\ 0 & \text{if} \quad \beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k < 0 \\ \in [0, \tilde{\varepsilon}_k] & \text{if} \quad \beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k = 0, \end{cases} \tag{2.50}$$

where β_k is as in (2.26). Then corresponding to (2.28), we have

$$g_k(\lambda, \mu) = J_k(\lambda, \mu, s_k^*, t_k^*) = \tilde{\varepsilon}_k \left(\beta_k - \lambda P(H_{1k}|\mathbf{Q}_k)W_k\right)^+ . \tag{2.51}$$

Define $\hat{\beta}_k = \tilde{\varepsilon}_k \beta_k$ and $\hat{a}_k = \tilde{\varepsilon}_k P(H_{1k}|\mathbf{Q}_k)W_k$. Then (2.51) becomes $g_k(\lambda, \mu) = (\hat{\beta}_k - \hat{a}_k)^+$. The solution for λ^* follows from Proposition 2.1 by substituting β_k and a_k with $\hat{\beta}_k$ and $\hat{a}_k$, respectively. And μ^* can be found by a line search as in the non-individual interference constraint case discussed earlier in this section.

2.5 Heuristic Algorithms

We notice that in Sec. 2.3 the computational complexity of the optimal solution is relatively high because in order to find the optimal dual variable μ, a line search is employed and within each iteration of this search, sorting metric β_k, $k = 1, 2, ..., K$ is required to obtain subchannel access probabilities $\mathbf{s}$, which has a complexity of $\mathcal{O}(K \log K)$. In order to reduce the computational complexity, we propose two heuristic algorithms. Both first solve for the access probabilities s_k, $k = 1, 2, ..., K$, and then allocate power. The details of these algorithms are provided next.

ALGORITHM 1

1) Allocate equal power $p_k = P_{tot}/K$ to each subchannel.

2) Calculate the rate of each subchannel as $R(p_k, s_k)$. Sort the rate vector in non-increasing order of magnitude and re-order $P(H_{1k}|\mathbf{Q}_k)W_k$ accordingly.

3) For $k = 1 : K$

   Set $s_k^* = 1$ if $\sum_{i=1}^{k} P(H_{1i}|\mathbf{Q}_i)W_i \leq \varepsilon$,

   otherwise set $s_k^* = \frac{\varepsilon - \sum_{i=1}^{k-1} P(H_{1i}|\mathbf{Q}_i)W_i}{P(H_{1k}|\mathbf{Q}_k)W_k}$, and if $k < K$, break after setting

   $s_j^* = 0$ for $j = k+1, ..., K$.

   end for.

4) Using s_k^*, $k = 1, 2, ...K$, obtained from Step 3 and following a similar way as in Sec. 2.3, optimal power allocation is given as follows: $p_k^* = \left(\frac{d_k}{2u_k v_k}\right)^+$ if $s_k^* > 0$, and $p_k^* = 0$ otherwise. The parameter μ is selected to satisfy $\sum_{k=1}^{K} s_k^* p_k^* = P_{tot}$; from (2.21)-(2.22), d_k is a function of c_k which, in turn, is a function of μ.

The sum rate of Algorithm 1 is

$$\sum_{k=1}^{K} s_k^* W_k \left[P(H_{0k}|\mathbf{Q}_k)\log_2(1 + p_k^* u_k) + P(H_{1k}|\mathbf{Q}_k)\log_2(1 + p_k^* v_k)\right]$$

and the actual rate is calculated as in (2.39). In Algorithm 1, the access probabilities s_k^*, $k = 1, 2, ..., K$, are determined based on equal power allocation. The subchannels with larger expected transmission rate are more likely to used by the secondary BS.

In Algorithm 1, when the secondary BS chooses subchannels, each subchannel's probability of interference is not taken into consideration. To remedy this drawback, we propose Algorithm 2.

ALGORITHM 2

1) Allocate power $p_k = P_{tot}/K$ to each subchannel.

2) Calculate the *rate-to-interference ratio* for each subchannel as $R(p_k, s_k)/I(s_k)$. Sort the rate-to-interference ratio vector in non-increasing order of magnitude, and re-order $P(H_{1k}|\mathbf{Q}_k)W_k$ accordingly.

3) Repeat Step 3 of Algorithm 1.

4) Repeat Step 4 of Algorithm 1 using the results of the Step 3 of Algorithm 2.

The sum rate of Algorithm 2 is

$$\sum_{k=1}^{K} s_k^* W_k \left[P(H_{0k}|\mathbf{Q}_k)\log_2(1+p_k^* u_k) + P(H_{1k}|\mathbf{Q}_k)\log_2(1+p_k^* v_k)\right]$$

and the actual rate is calculated as in (2.39).

The complexity of these two heuristic algorithms is comprised of two parts. The first part comes from sorting *rate* or *rate-to-interference ratio*, which is $\mathcal{O}(K\log K)$, and the second part is from allocating power using the pre-determined channel access probabilities. This can be solved using a line search over μ to satisfy the transmit power constraint. A (crude) comparison based on computer simulation run time of these two heuristic algorithms as well as the optimal algorithm is presented in Sec. 2.8.

2.6 Implementation issues

In this section we discuss some implementation issues that arise in our algorithms. In the proposed soft-decision cooperative sensing based algorithms one requires knowledge of conditional distribution of sensing statistics $f(Q_{ik}|H_{jk})$, $i \in \mathcal{A}_k$, $k = 1, 2, ..., K$, $j \in \{0, 1\}$, interference power from primary users' transmission at secondary users $p_{puk}|h_{pk}|^2$, $k = 1, 2, ..., K$ and the CSI h_k between secondary BS and users. In this section we briefly discuss possible solutions to how this knowledge can be obtained. We assume that the CSI among the secondary links in the secondary network can be acquired as in [17] (and others) "by a feedback link from receiver to transmitter, or just exploiting channel reciprocity when transmitter and receiver transmit over the same band." [Note that a feedback link is also needed to transmit the sensing metric to the secondary BS.] Knowledge of this CSI is also critical to the hard decision approaches of [26] and [17], as it is needed to calculate the needed instantaneous capacity. The CSI information among various nodes is also required in [16, 27]. The nature of prior knowledge needed regarding the CSI between SUs and PU (h_{pk} and $h_{pi,k}$ in our notation) depends upon the nature of the spectrum sensing metric. This aspect is discussed next.

2.6.1 Obtaining primary user related prior knowledge

Although our algorithms are open to the selection of sensing techniques, suppose that we use the energy detector (sole choice in [26] and [17], and also used in our simulations in Sec. 2.8). For energy detector, the sensing metric Q_{ik} follows the following normal distributions under both hypotheses [26]:

$$f(Q_{ik}|H_{0k}) \sim \mathcal{N}\left(M\sigma_0^2, 2M\sigma_0^4\right) \tag{2.52}$$

$$f(Q_{ik}|H_{1k}) \sim \mathcal{N}\left(M(\sigma_0^2 + p_{puk}|h_{pi,k}|^2), 2M\sigma_0^2(\sigma_0^2 + 2p_{puk}|h_{pi,k}|^2)\right) \tag{2.53}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2, M is the number of signal samples collected in the sensing phase, σ_0^2 is the noise power and $h_{pi,k}$ is the channel gain between the primary user and the secondary user i over subchannel k. In this case, as noted in [26] (see also [20]), one only needs estimates of σ_0^2 and $\sigma_0^2 + p_{puk}|h_{pi,k}|^2$, representing received signal power under no PU transmission and PU transmission, respectively. These entities can be estimated *a priori* during the (confirmed) periods the PU is known to be inactive or active which would allow computation of the probabilities in (2.52) and (2.53).

For more general sensing metric Q_{ik}, one may need knowledge of $h_{pi,k}$ which would require cooperation between the primary and secondary networks in the sense that PU's may need to transmit training sequences to enable CSI acquisition.

2.6.2 Effect of Imperfect Channel Estimation

Due to the time-varying nature of the wireless channel and noise, error is always present in the CSI estimation results. This will result in performance degradation for both primary and secondary network. An analysis of this aspect is outside the scope of this chapter but we have investigated this aspect to some extent via simulations in Sec. 2.8 where we have evaluated the imperfect CSI effect on the system performance by modeling the channel estimation errors as zero-mean Gaussian random variables. For a single source-destination pair, let h and $\hat{h}$ denote the true and estimated channel coefficient (scalar, flat fading), respectively, and let e be the channel estimation error. Then we have ($\mathcal{N}_c$ denotes complex Gaussian distribution)

$$\hat{h} = h + e, \quad e \sim \mathcal{N}_c(0, \sigma_h^2) \tag{2.54}$$

(In Sec. 2.8 we have used $\sigma_h^2 = a|h|^2$ for $a = 0.02$ or 0.10.) Due to these channel estimation errors, secondary network's throughput will be affected. Also interference to primary users may exceed the design interference bound since the calculation of channel a posteriori probabilities is related to the channel gain between primary transmitter and secondary spectrum sensor. The simulation results show that the throughput of secondary network is not sensitive to channel estimation error, however, the interference to primary network can exceed the interference bound.

Another observation is when the cost of sending CSI of secondary links and sensing metrics of sub-channels back to BS is too expensive and becomes a major implementation obstacle, we can quantize the CSI between secondary-links into fewer bits since our algorithms are more robust to this kind of error than to the errors in the sensing metric. In general, the cost of feeding back the required variables is likely to be "small." In the feedback link from the i-th SU to the secondary BS, one needs to send the estimated CSI h_k and the sensing metric Q_{ik} (with its associated distribution parameters) for the k-th subchannel, once every time slot. For the energy detector, the sensing metric Q_{ik} is the received energy (a real number) and the associated parameters are the two means and two variances under hypotheses H_{0k} and H_{1k} (all real numbers), whereas the CSI h_k is complex-valued. Thus six scalars (one complex and five reals) are required to be sent over the feedback link once every time slot. This is to be contrasted with the number of data samples (symbols) per slot in each subchannel during direct transmission. In our simulations we used $M = 50$ signal samples (symbols) per time slot for sensing. Assuming that we have $\geq$150 symbols in the information transmission phase (total $\geq$200 symbols per slot), it is seen that the feedback overhead is just 6 samples per slot versus $\geq$200 samples in direct transmission, leading to commensurate bandwidth requirements. A detailed comparison would depend upon actual implementation including number of bits per sample.

2.7 Comparison with Hard Decision based Spectrum Sensing

For comparison, we also consider hard decision spectrum sensing and power adaptation. An approach is available in [17], which in turn extends the approach of [26]. Both [17] and [26] consider energy detectors for spectrum sensing; the case of more general detectors is yet unsolved. Therefore, we will assume that energy detector is used and let Q_{ik} denote the sensing statistic (received energy) at secondary user i over subchannel k. Then for each subchannel k, the sensing metric is given by

$$T_k = \sum_{i \in \mathcal{A}_k} Q_{ik}. \tag{2.55}$$

Assume Q_{ik} follows a normal distribution $\mathcal{N}(\mu_{i,k}^{(j)}, \sigma_{i,k}^{(j)2})$ under H_{jk}, $j \in \{0, 1\}$. Then we have $T_k \sim \mathcal{N}(\sum_{i\in\mathcal{A}_k} \mu_{i,k}^{(j)}, \sum_{i\in\mathcal{A}_k} \sigma_{i,k}^{(j)2})$ under hypothesis H_{jk}, $j \in \{0, 1\}$.

Let $\tilde{s}_k$ be the access action to subchannel k. Then in hard spectrum sensing based on the test threshold γ, we have the hard decision

$$\tilde{s}_k^* = \begin{cases} 1 & \text{if} \quad T_k \geq \gamma \\ 0 & \text{if} \quad T_k < \gamma; \end{cases} \tag{2.56}$$

therefore, optimizing $\tilde{s}_k$, $k = 1, 2, ..., K$, is actually optimizing γ. [Note that under the null hypothesis of only noise at any sensing receiver, the sensing statistic Q_{ik} is identically distributed over every sensor i and every subchannel k if the number of samples used at each sensor are identical and the noise variance is identical in each subchannel and receiver. This allows for consideration of identical threshold γ for each subchannel. This is the case for the numerical example considered in Sec. VI.] Hence joint optimization of spectrum access parameter $\tilde{s}_k$, $k = 1, 2, ..., K$ and power allocation p_k, $k = 1, 2, ..., K$, can be formulated as follows:

$$\text{P5}: \quad \max_{\mathbf{p},\gamma} \quad \tilde{R}(\gamma, \mathbf{p}) \tag{2.57}$$

$$\text{s.}t. \quad \sum_{k=1}^{K} P_{H_{1k}} \left(1 - P_{dk}(\gamma)\right) W_k \leq \varepsilon \tag{2.58}$$

$$\sum_{k=1}^{K} P_{H_{0k}} \left(1 - P_{fa}(\gamma)\right) p_k + P_{H_{1k}} \left(1 - P_{dk}(\gamma)\right) p_k \leq P_{tot} \tag{2.59}$$

$$p_k \geq 0, \quad k = 1, 2, ..., K, \tag{2.60}$$

where

$$\tilde{R}(\gamma, \mathbf{p}) = \sum_{k=1}^{K} P_{H_{0k}} W_k \left(1 - P_{fa}(\gamma)\right) \log_2(1 + p_k u_k) + P_{H_{1k}} W_k \left(1 - P_{dk}(\gamma)\right) \log_2(1 + p_k v_k) \tag{2.61}$$

and P_{fa} (the same for all subchannels) and P_{dk} denote the probability of false alarm and detection, respectively, for subchannel k. Notice that P_{fa} and P_{dk} are functions of γ rather than sensing statistics $\mathbf{Q}_k$. While $\tilde{s}_k$ will still be determined by $\mathbf{Q}_k$ from (2.56), the optimal power p_k^* will be a function of only the subchannel state and γ. This problem set-up is patterned after [17] (modified to accommodate our interference bound constraint), and the optimal solution to this optimization problem can be obtained by following [17]. [Note that [17] extends the results of [26] to include power allocation.]

For a fixed threshold γ, the value of $P_{fa}(\gamma)$ and $P_{dk}(\gamma)$ are fixed, where $\mu_{i,k}^{(j)}$, $\sigma_{i,k}^{(j)}$, $i \in \mathcal{A}_k$, $k = 1, 2, ..., K$, $j \in \{0, 1\}$ are known. Define the feasible region of γ as $\mathcal{F} = \left\{\gamma \middle| \sum_{k=1}^{K} P_{H_{1k}}(1 - P_{dk}(\gamma))W_k \leq \varepsilon\right\}$. Then for any $\gamma_0 \in \mathscr{F}$, optimal power allocation can be obtained by solving

$$\begin{aligned} \text{P6}: \quad \max_{\mathbf{p}} \quad & \tilde{R}(\gamma, \mathbf{p}) && (2.62)\\ \text{s.}t. \quad & \sum_{k=1}^{K} P_{H_{0k}} \left(1 - P_{fa}(\gamma_0)\right) p_k + P_{H_{1k}} \left(1 - P_{dk}(\gamma_0)\right) p_k \leq P_{tot} && (2.63)\\ & p_k \geq 0, \quad k = 1, 2, ..., K. && (2.64) \end{aligned}$$

Similar to Sec. 2.3, the optimal solution is as follows:

$$p_k^* = \left[\frac{-\tilde{b}_k + \sqrt{\tilde{b}^2 - 4\tilde{e}_k\tilde{c}_k}}{2\tilde{e}_k}\right]^+, \quad (2.65)$$

where

$$\begin{aligned} \tilde{e}_k &= u_k v_k \tilde{\lambda}(P_{H_{0k}} \left(1 - P_{fa}(\gamma_0)\right) + P_{H_{1k}} \left(1 - P_{dk}(\gamma_0)\right)) \ln 2 && (2.66)\\ \tilde{b}_k &= \tilde{\lambda} \left[P_{H_{0k}} \left(1 - P_{fa}(\gamma_0)\right) + P_{H_{1k}} \left(1 - P_{dk}(\gamma_0)\right)\right] \left(u_k + v_k\right) \ln 2 \\ &\quad -[P_{H_{0k}} W_k \left(1 - P_{fa}(\gamma_0)\right) + P_{H_{1k}} W_k \left(1 - P_{dk}(\gamma_0)\right)] u_k v_k && (2.67)\\ \tilde{c}_k &= \tilde{\lambda} \left[P_{H_{0k}} \left(1 - P_{fa}(\gamma_0)\right) + P_{H_{1k}} \left(1 - P_{dk}(\gamma_0)\right)\right] \ln 2 \\ &\quad -P_{H_{0k}} W_k \left(1 - P_{fa}(\gamma_0)\right) u_k - P_{H_{1k}} W_k \left(1 - P_{dk}(\gamma_0)\right) v_k, && (2.68) \end{aligned}$$

and $\tilde{\lambda}$ is chosen to satisfy $\sum_{k=1}^{K} P_{H_{0k}} \left(1 - P_{fa}(\gamma_0)\right) p_k^* + P_{H_{1k}} \left(1 - P_{dk}(\gamma_0)\right) p_k^* = P_{tot}$. The optimal sensing threshold γ^* can be obtained by solving

$$\text{P7}: \quad \max_{\gamma} \quad \tilde{R}(\gamma, \mathbf{p}^*) \tag{2.69}$$

$$\text{s.t.} \quad \gamma \in \mathcal{F}. \tag{2.70}$$

Problem P7 can be solved by a line search over γ.

Thus, in each time slot, the optimal spectrum access actions are given by $s_k^* = 1$ if $T_k \geq \gamma^*$ and $s_k^* = 0$ if $T_k < \gamma^*$, for $k = 1, 2, ..., K$. The optimal power allocation is given by (2.65). The actual transmission rate (bound) we obtain in each time slot using this hard spectrum sensing method is given by

$$\tilde{R}_a = \sum_{k=1}^{K} 1_{H_{0k}} s_k^* W_k \log_2(1 + p_k^* u_k) + 1_{H_{1k}} s_k^* W_k \log_2(1 + p_k^* v_k). \tag{2.71}$$

The interference to primary user is

$$\sum_{k=1}^{K} 1_{H_{1k}} s_k^* W_k. \tag{2.72}$$

2.8 Simulation Examples

Now we provide numerical results using the algorithms proposed in Secs. 2.3 and 2.5. and compare them with hard sensing algorithm in Sec. 2.7. Secondary users associated with $K = 16$ subchannels are located within a circular ring area with radii between 200 m to 1 km, while the secondary BS is located at the center. The path-loss exponent for large-scale fading is set to be 3.5. The secondary BS transmits to secondary users through K subchannels; recall that, in this chapter, we are considering a downlink scenario for the secondary network. We assume that the subchannels between the secondary BS and secondary users experience flat Rayleigh fading. The channel fading coefficients remain constant during each time slot and change independently from slot to slot. The "sum" received SNR of K secondary receivers is defined to be

$$\text{sum received SNR } = P_{tot}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{E\{|h_k|^2\}}{N_0W_k}\right) \tag{2.73}$$

where $E\{|h_k|^2\}$ is calculated according to the kth subchannel's path-loss. The randomness of h_k, $k = 1, 2, ..., K$ is caused by Rayleigh fading.

We assume that the secondary BS employs energy detector as in [26]; (note that our algorithms are open to the selection of sensing techniques). Assume the primary users' transmit power p_{puk} is normalized to 1 in all K subchannels. Then in each time slot, the sensing statistics Q_{ik} for $i \in \mathcal{A}_k$ and $k = 1, 2, ..., K$, given primary users' behavior on the k-th subchannel, follows a normal distribution given by (2.52) under H_{0k}, and under H_{1k} we have

$$f(Q_{ik}|H_{1k}) \sim \mathcal{N}\left(M(\sigma_0^2 + |h_{pi,k}|^2),\ 2M\sigma_0^2(\sigma_0^2 + 2|h_{pi,k}|^2)\right) \tag{2.74}$$

Considering that sharing each secondary user's sensing statistics Q_{ik} with BS can be expensive, we assume only 4 out of 16 users will transmit their Q_{ik}, $k = 1, 2..., K$, to the BS.

In simulations, we use $M = 50$, and ("normalized") $W_k = 1$ and $N_0 = 1$ leading to $\sigma_0^2 = W_kN_0 = 1$. The subchannel power gain $g_{pi,k} := |h_{pi,k}|^2$ from transmitting primary user to secondary user i over subchannel k follows an exponential distribution with mean $\bar{g}_{pi,k}$. The received SNR of PU's signals, averaged over all secondary users' and BS's receivers, is given by $\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{A}_k|}\sum_{i\in\mathcal{A}_k} p_{puk}\bar{g}_{pi,k}/\sigma_0^2$. Finally, we assume that the primary users occupy each subchannel with probability 0.5 during each time slot. All simulation results are based on 1000 runs.

Figs. 2.1 and 2.2 show the sum transmission rate vs sum received SNR ($\propto P_{tot}$) for the secondary network of 16 subchannels/secondary users for two different PU SNR's when the (total) PU interference bound ε is set to be 3% (by which we mean that $\varepsilon = 0.03\sum_{k=1}^{K} W_k$ in (2.6)); we will also refer to this as 0.03 fraction of bandwidth). Also, to better evaluate the performance of our soft-decision sensing based algorithm, the actual throughput using both soft and hard spectrum sensing is also presented in the following figures. Here the actual throughput is calculated according to (2.39). It is seen from Figs. 2.1 and 2.2 that the actual rate of the optimal algorithm agrees well with the theoretical value. Also, the soft decision algorithm significantly outperforms the hard decision algorithm in both high PU SNR (Fig. 2.2) and low PU SNR (Fig. 2.1) scenarios, and the heuristic Algorithm 2 (with power control) has a near optimal performance. Even when the PU's SNR is low (-20dB), which implies that sensing is less accurate, the soft-decision sensing based optimal algorithm still outperforms the hard sensing scheme.

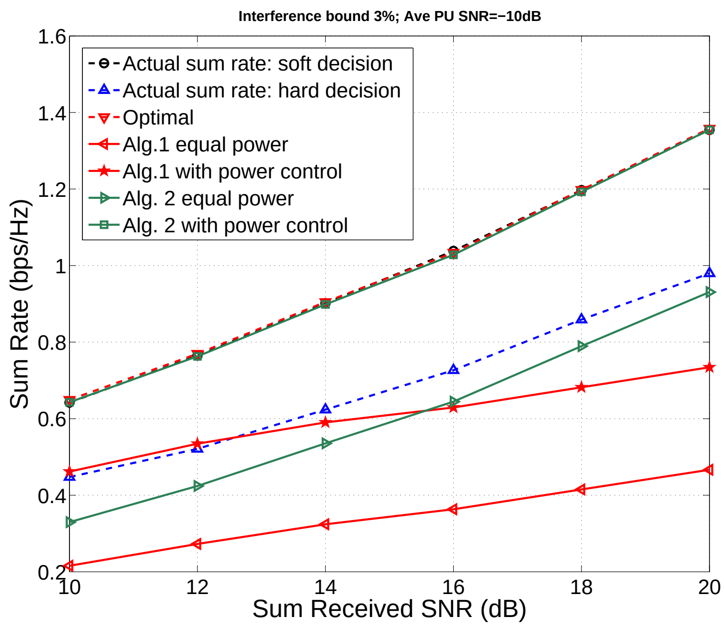

Figure 2.1: Sum transmission rate of secondary network versus sum received SNR (2.73) at secondary receivers when 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -10dB. [The curves labeled "optimal," "Actual sum rate: soft decision," and "Alg. 2 with power control" are all quite close to each other toward the top of the fig.]

Figs. 2.3 and 2.4 show the actual fraction of bandwidth which is interfered with by secondary transmissions, when the bound ε is set to be 3% (0.03). Both soft and hard-sensing algorithms have interferences close to the bound of 3%. We observe that the soft decision algorithm causes a slightly higher interference than hard decision algorithm. However, given

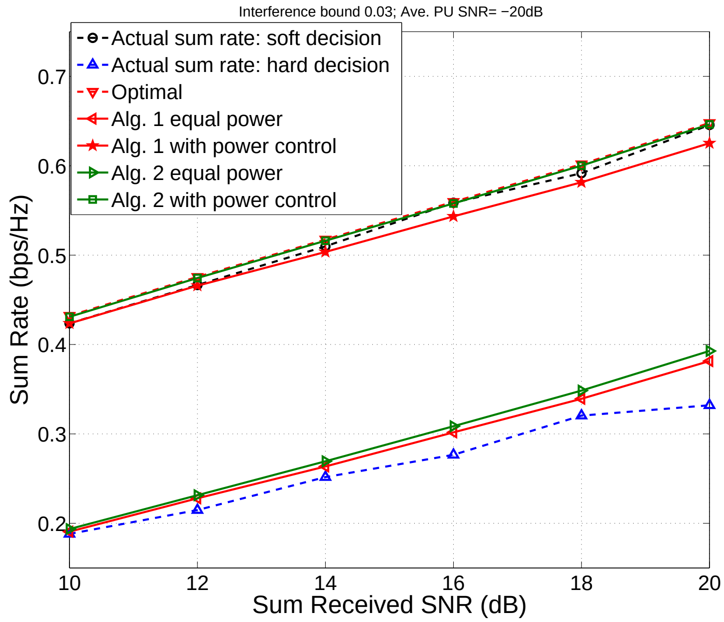

Figure 2.2: Sum transmission rate of secondary network versus sum received SNR (2.73) at secondary receivers when 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -20dB. [The curves labeled "optimal" and "Alg. 2 with power control" are all quite close to eachother in the top half of the figure.]

the fact that soft decision algorithm has a much higher feasible transmission rate than hard decision algorithm, we can still conclude that using soft decision spectrum sensing can significantly improve the performance of the secondary user network.

Fig. 2.5 shows that the the actual interference is close to the interference bound when the SNR of PU's signal is -10dB and the total received SNR of secondary users is 15dB. From Fig. 2.6 we can observe that the actual rate of soft decision optimal algorithm is close to the the optimal value and is significantly higher than hard decision algorithm. Also, the heuristic Algorithm 2 has a near optimal performance.

Figs. 2.7 and 2.8 show the imperfect CSI effect where the channel estimation error variance is set to be either 2% or 10% of the true channel power gain of its corresponding channel (see (2.54) and discussion after it) and the PU SNR at secondary sensor is $-10dB$. It is observed that the secondary network throughput is not sensitive to channel estimation error. However, the interference to primary users exceeds the 3% bound, marginally for

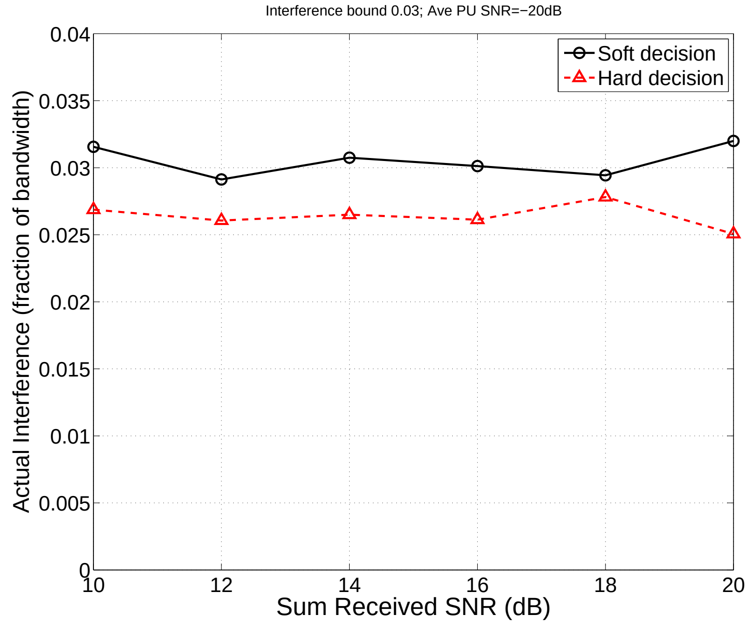

Figure 2.3: Actual total interference versus sum received SNR (2.73) at secondary receivers when 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -20dB

$\sigma_h^2 = 0.02|h|^2$ but more significantly for $\sigma_h^2 = 0.10|h|^2$. This is because the primary signal power at the spectrum sensor of secondary network is used in calculation of the *a posteriori* channel probabilities. Figs. 2.9 and 2.10 show the results corresponding to Figs. 2.7 and 2.8 except that now the PU SNR at the secondary sensor is $-20dB$. Compared to Figs. 2.7 and 2.8, now the effect of imperfect CSI on interference to PU network is much less severe. [We have also investigated the case where with noise variance normalized to one, we used $\sigma_h^2 = 0.02$ or 0.1 (i.e. it is not a function of $|h|^2$); there was no significant difference from the results shown in Figs. 2.7-2.10, and the results are not shown.]

Figs. 2.11 and 2.12 show the results when we use individual interference constraints as discussed in Sec. 2.4. The total PU interference bound ε was set to be 3% while the individual PU interference bound ε_k was set to 5% (i.e. $\varepsilon_k = 0.05W_k$ in (2.43)) for each of the 16 subchannels ($k = 1, 2, \cdots, 16$). It is seen that under additional constraints, the performance (sum rate) is poorer but not by much while the total interference is reduced.

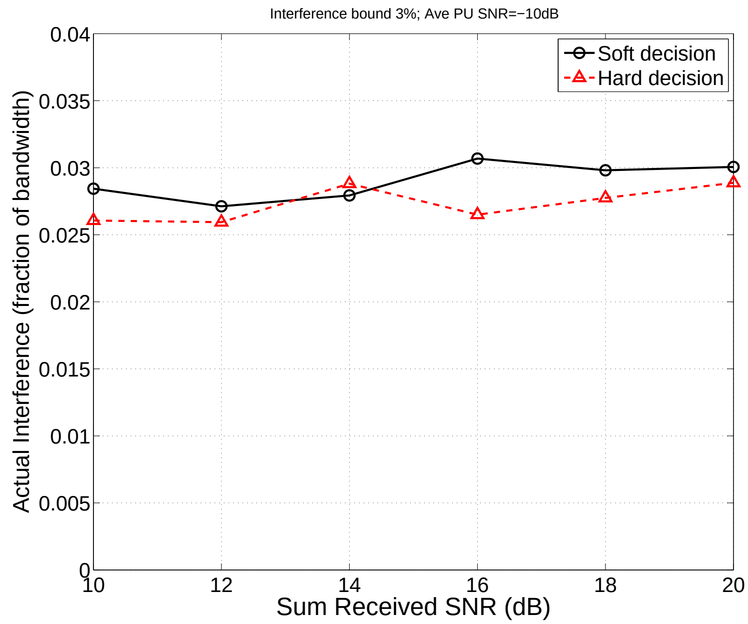

Figure 2.4: Actual total interference versus sum received SNR (2.73) at secondary receivers when 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -10dB.

Finally, the CPU times (secs) for 1000 runs on a computer with Intel Pentium(R) Dual-Core CPU E5200@2.5GHz processor running under Windows 7 (professional) were 2.9267, 0.2083 and 0.1606 for the optimal algorithm, heuristic Algorithms 1 and 2, respectively, for the results shown in Fig. 2.1.

2.9 Conclusions

We investigated joint optimization of cooperative spectrum sensing, channel access and power allocation in an overlay multi-band cognitive radio network. Instead of making traditional hard binary decisions, a soft-decision cooperative spectrum sensing concept using the continuous-valued sensing test statistics was considered to maximize the secondary users' sum throughput while keeping the interference to primary users under a specified threshold. The problem was shown to be a convex optimization problem and the Lagrangian dual method was employed to obtain the optimal solution. We also provided an alternative formulation where additionally interference to individual PUs was also constrained. Two

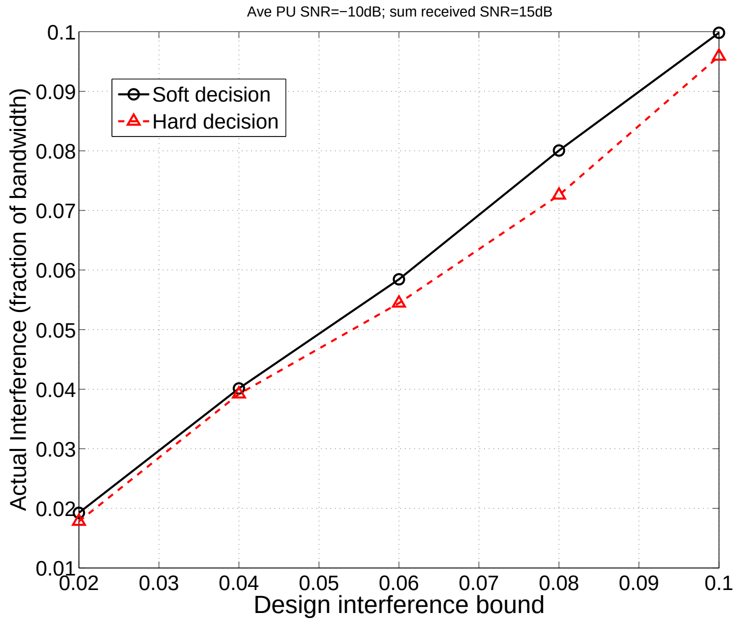

Figure 2.5: Actual total interference versus design bound on total interference to primary network when sum received SNR (2.73) at secondary receivers is 15dB and SNR of PU's signal is -10dB.

heuristic algorithms were also proposed to reduce the computational complexity. Simulation results showed that our soft sensing based algorithm significantly outperforms traditional hard decision sensing algorithms and one of the proposed heuristic algorithms (Algorithm 2 with power control) achieves a near optimal performance. Practical implementation issues are also discussed, regarding obtaining channel state information of both secondary links and primary to secondary links. The performance of our proposed optimal algorithm under imperfect CSI was illustrated via simulations.

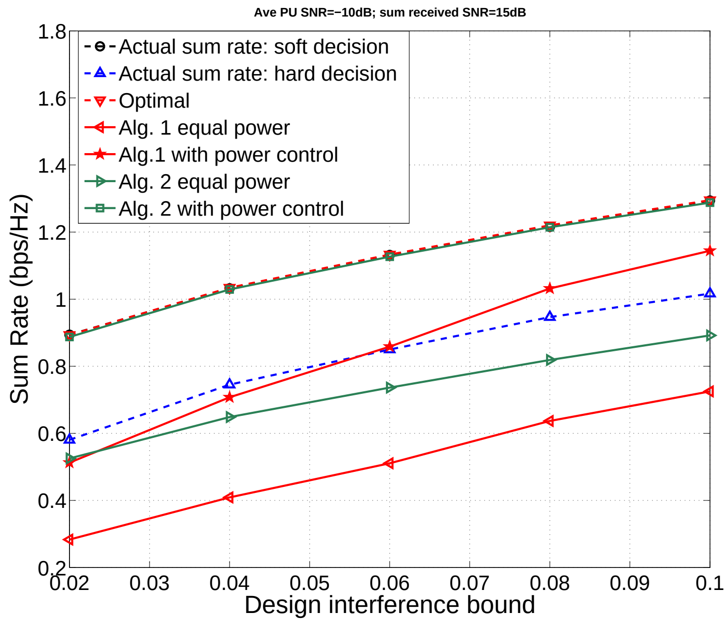

Figure 2.6: Sum transmission rate of secondary network versus design bound on total interference to primary network when sum received SNR (2.73) at secondary receivers is 15dB and SNR of PU's signal is -10dB.

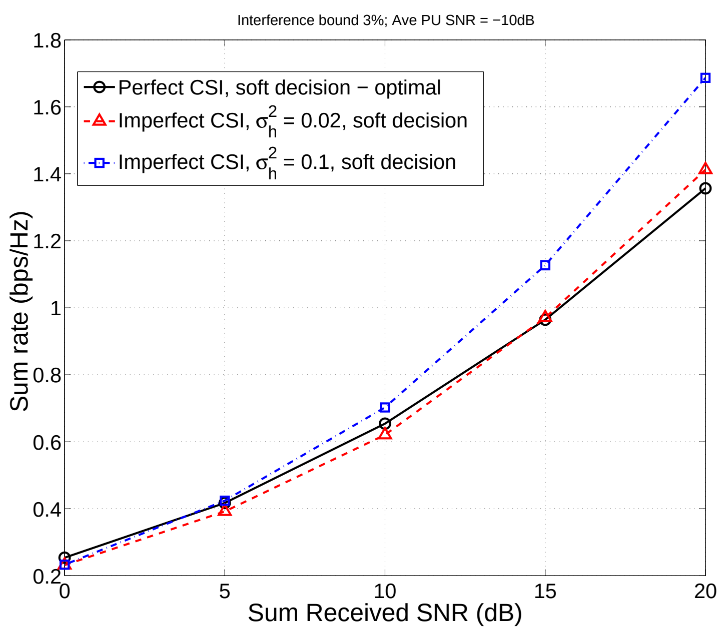

Figure 2.7: Imperfect CSI: Sum transmission rate of secondary network versus sum received SNR (2.73) at secondary receivers when up to 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -10dB. Channel estimation error variance σ_h^2 is set to be either 2% or 10% of the true channel power gain of its corresponding channel (see (2.54)).

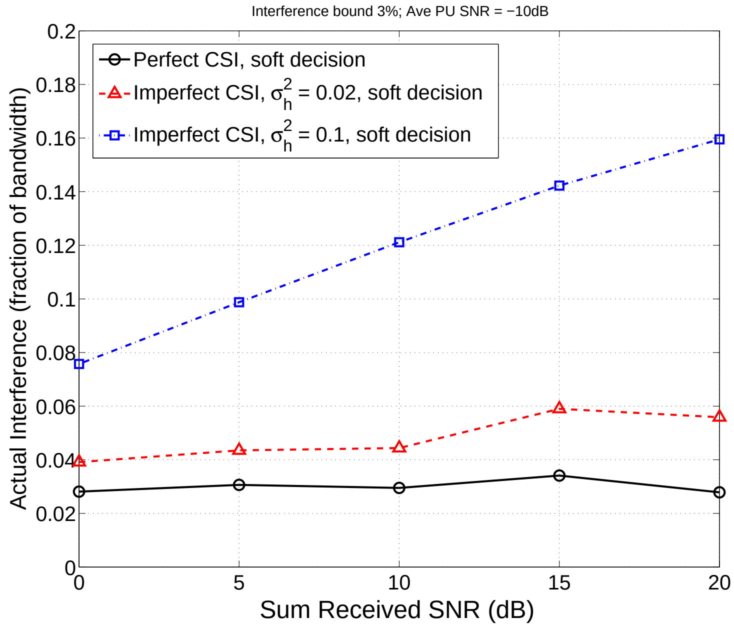

Figure 2.8: Imperfect CSI: Actual total interference versus sum received SNR (2.73) at secondary receivers when up to 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -10dB. Channel estimation error variance σ_h^2 is set to be either 2% or 10% of the true channel power gain of its corresponding channel (see (2.54)).

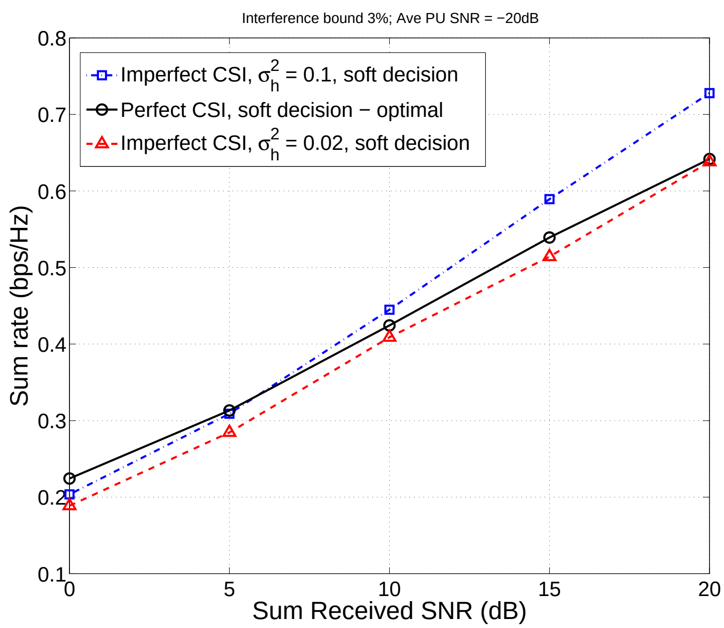

Figure 2.9: Imperfect CSI: Sum transmission rate of secondary network versus sum received SNR (2.73) at secondary receivers when up to 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -20dB. Channel estimation error variance σ_h^2 is set to be either 2% or 10% of the true channel power gain of its corresponding channel (see (2.54)).

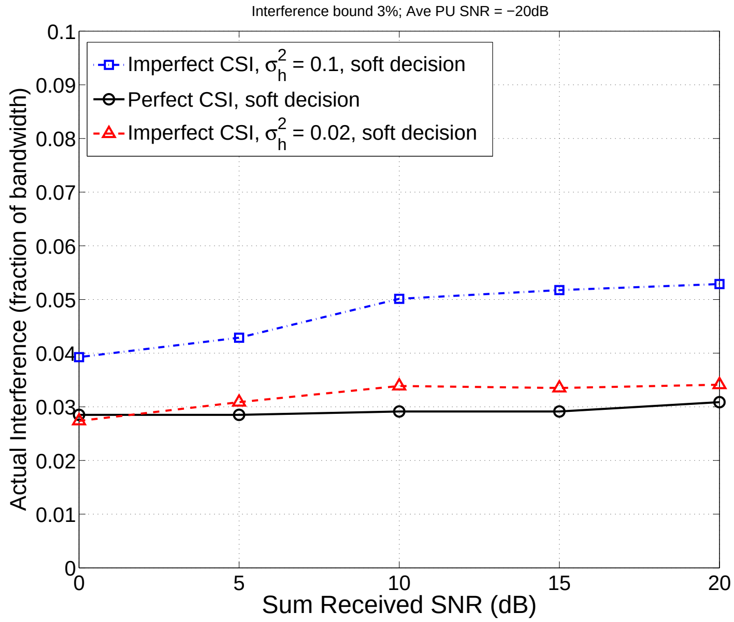

Figure 2.10: Imperfect CSI: Actual total interference versus sum received SNR (2.73) at secondary receivers when up to 3% of the primary network's bandwidth is interfered with by secondary BS's transmission and the received SNR of PU's signal is -20dB. Channel estimation error variance σ_h^2 is set to be either 2% or 10% of the true channel power gain of its corresponding channel (see (2.54)).

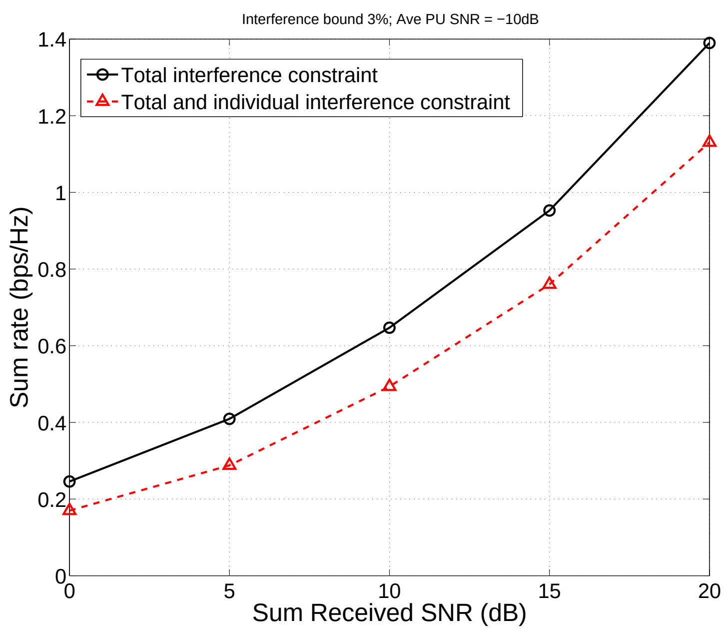

Figure 2.11: Individual PU interference constraint: Sum transmission rate of secondary network versus sum received SNR (2.73) at secondary receivers when up to 3% of the primary network's total bandwidth and up to 5% of individual subchannel bandwidths (total 16 subchannels) are interfered with by secondary BS's transmission and the received SNR of PU's signal is -10dB.

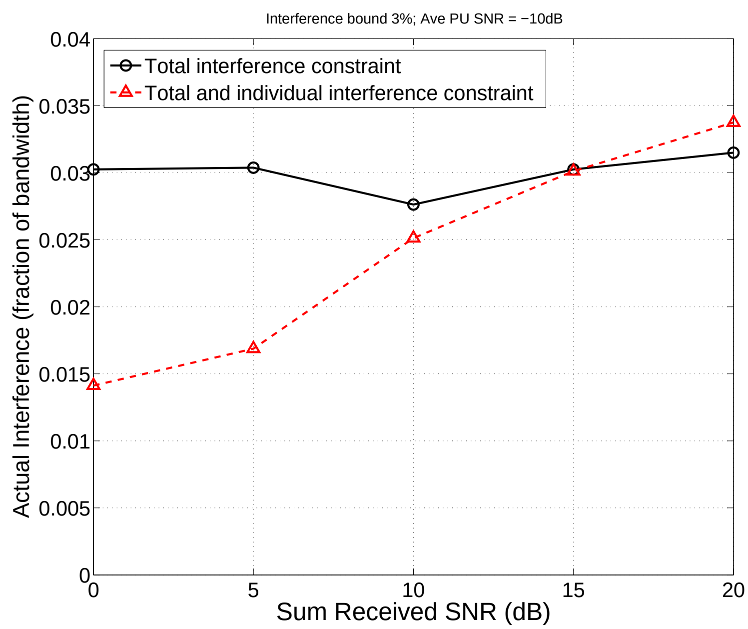

Figure 2.12: Individual PU interference constraint: Actual total interference versus sum received SNR (2.73) at secondary receivers when up to 3% of the primary network's total bandwidth and up to 5% of individual subchannel bandwidths (total 16 subchannels) are interfered with by secondary BS's transmission and the received SNR of PU's signal is -10dB.

Chapter 3

Precoding Design in Multi-User Cognitive Relay Network: a MMSE Approach

In this chapter, the cognitive radio network employs a relay station to facilitate the transmissions of multiple users. These secondary users intend to transmit data streams to multiple partners, therefore form a multi-way transmission, which is a generalization of multi-user two-way relay network. In designing this cognitive radio network with multiple users, how to manage interference is of significant importance since interference not only exists between primary and cognitive radio networks, but also among multiple cognitive users. With the help of multiple antennas at transmitters and receivers, proper precoding/beamforming design can cancel inter-network (between primary and secondary network) interference and inter-user interference. In this chapter, an MSE-based joint source and relay precoding algorithm is proposed for a multi-user multi-way cognitive radio network.

This chapter is organized as follows. Sec. 3.1 introduces the background of precoding design in relay and cognitive radio networks. In Sec. 3.2 we introduce the system model and formulate the optimization problem. Then this problem is decomposed into convex subproblems and solved iteratively in Sec. 3.3. To reduce the complexity, we also propose an non-iterative algorithm in Sec. 3.4. In Sec. 3.5, a robust design is proposed to consider the situation when only imperfect CSI is available, based on our non-iterative algorithm. Simulation results are presented in Sec. 3.6 to illustrate the effectiveness of our algorithms. Finally, conclusions are drawn in Sec. 3.7 and some technical details are provided in the Appendix.

3.1 Introduction

In cognitive radio network, when SUs concurrently access the spectrum with licensed primary users, how to mitigate interference to PUs becomes a crucial design issue of secondary network. By employing multiple antennas at the secondary transmitters, interference to PUs can be potentially nulled out by designing precoders provided the channel state information (CSI) between the relevant source-destination pair is known. On the other hand, multiple input multiple output (MIMO) technology can further improve spectrum efficiency by providing multiplexing gain. However in multi-user MIMO systems, the performance is usually limited by inter-user interference which can be mitigated by system precoding/decoding designs. In traditional multi-user MIMO systems, many interference alignment (IA) algorithms have been proposed which focus on eliminating (or minimizing) the interference. For example, in [34] inter-user interference is minimized by adjusting transmitting and receiving "directions". A similar transmitting and receiving directions based iterative algorithm has been proposed in [35] which also considers the received signal's directions. In this chapter we adopt mean square error (MSE) of the decoded signal as the design criterion rather than residual interference because by minimizing the MSE, desired signal's strength is also taken into consideration. In [36], a duality based robust transceiver design method is proposed for multi-user MIMO system using MMSE (minimum MSE) as the design criterion, while in [37] optimal receivers are designed using the MMSE criterion for a downlink multi-user MIMO system to maximize the weighted sum rate.

Considerable research has been done regarding the transceiver design in CR MIMO systems. A rate balanced transceiver design for multiuser cognitive system is studied in [38] where fairness among users is considered. In [39], a linear precoding method was proposed for CR multiuser downlink MIMO system based on MMSE. For systems with knowledge of the CSI at the transmitter, a nonlinear transceiver design problem in a multi-tier MIMO cognitive radio network has been investigated in [53] while a linear transceiver design in a downlink cognitive MIMO system has been proposed in [54]. All these previous works deal with one-way transmission. For a network where multiple users have multiple data streams for each other, a relay station could be employed to enable multi-user two-way transmission. Moreover, introduction of the relay node can further improve spectrum efficiency. Therefore, in this chapter, we employ an amplify-and-forward relay node in the secondary network to support multi-user transmission. For non-cognitive radio relay systems, optimal beamforming matrix at the relay node has been designed in [40] where capacity region for a two-user two-way system has also been derived. An iterative source and relay precoding algorithm for two-user two-way MIMO relay system has been proposed in [41] using the MMSE criterion. Precoder design in a system where one pair of users are assisted by multiple relays has been considered in [42, 43]. While in two-user two-way relay system inter-user interference can be completely removed, this kind of interference can significantly degrade the performance of multi-user two-way relay system. For multi-user two-way relay systems, design of MIMO relay beamforming matrix based on zero-forcing (ZF) and MMSE criteria has been investigated in [44].

In this chapter, we develop a precoding and decoding matrices design algorithm in a multi-user multi-way relay MIMO system for cognitive radio using MSE as our design criteria. Due to the non-convexity of this problem, an iterative algorithm is proposed to alternately optimize precoding matrices at secondary transmitters and the relay station, and decoding matrices at secondary receivers. Our object is to minimize the sum MSE of all users under transmit power constraints at each secondary transmitter as well as the relay station, while interference to the primary user is nulled out by proper precoding matrices design. A matrix distance based non-iterative algorithm is also proposed to reduce the computational complexity.

We also consider the effect of imperfect CSI. In practical system, exact CSI may be difficult to obtain due to the time-varying nature of channels and the cost of information feedback from the receiver to transmitter. Therefore designing a robust precoder which is insensitive to CSI errors maybe a more practical choice. A robust relay precoder design of a two-hop amplify-and-forward(AF) MIMO system with multiple relays under imperfect CSI has been proposed in [45]. Joint relay precoder and destination receive filters designs in a nonregenerative MIMO relay network have been investigated in [46] under two models of CSI errors: stochastic error and norm-bounded error. For two-way relay systems, transceiver design for a three node two-way relay system under imperfect CSI has been considered in [47]. In this chapter, we model the errors in channel coefficients as additive Gaussian random variables which have zero mean and known variance. With this uncertainty in the CSI, the expected sum MSE over channel coefficient errors becomes our objective function. Also, interference to the PUs can not be eliminated completely. Therefore the constraint on interference to PU now becomes keeping the mean interference power less than a threshold.

Notation: We use bold lower-case and upper-case letters to denote vectors and matrices, respectively, $\mathbf{I}$ is the identity matrix, vec$\{.\}$ is the vectorization operator, $\otimes$ is the Kronecker product, and $(.)^T$, $(.)^\dagger$, $(.)^\ddagger$ and $(.)^{-1}$ denote the transpose, conjugate transpose, conjugate and inverse of a matrix, respectively. We use $\|.\|$ and $\|.\|_F$ to denote the 2-norm and Frobenius norm, respectively, tr(.) and rank(.) are the trace and rank of a matrix, $\mathbb{R}\{.\}$ indicates the real part of a complex number, $\mathbb{E}(.)$ is the expectation operation, $\{\mathbf{A}\}_{mn}$ denotes the mn-th element of matrix $\mathbf{A}$, and $\mathbb{C}^{m\times n}$ is the space of m by n complex matrices. "Subject to", "with respect to" and "quality of service" are abbreviated as s.t., w.r.t. and QoS, respectively.

3.2 System Model and Problem Formulation

Consider a cognitive radio network consisting of K secondary users and one non-regenerative relay station. The secondary users intend to send information to each other through the help of the relay node. Each secondary user is equipped with M antennas while the relay station has N antennas. A two time slot half-duplex transmission scheme is employed here. In the first time slot all of the K users who have information to send transmit to the relay node while in the second time slot relay sends a linear combination of its received signal from the first time slot, and K secondary users listen and decode their desired signals, as shown in Fig. 3.1. This secondary network coexists with a primary source-destination pair within a single band. Assume primary transmitter and receiver are equipped with J_p and M_p antennas respectively. We assume that the network operates in a flat-fading environment. We assume that the signals transmitted from different SUs and PUs, and noise at all receivers, are all mutually statistically independent. Suppose that the i-th SU transmits

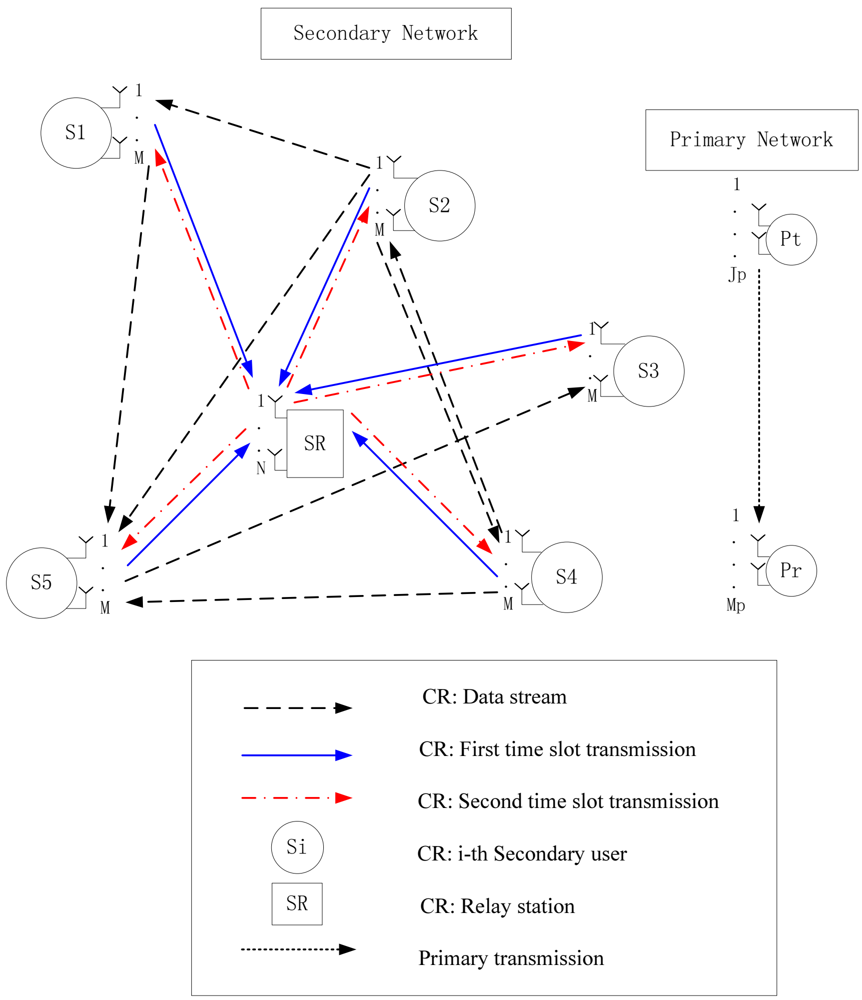

Figure 3.1: System set-up

d_i independent data streams. Then we will denote the $i-$th SU's transmitted signal vector as $\mathbf{s}_i = [s_{i1}, s_{i2}, ..., s_{id_i}]^T$, $i = 1, 2, ...K$. Let $\tilde{\mathbf{s}}_i = [\tilde{s}_{i1}, \tilde{s}_{i2}, ..., \tilde{s}_{i\tilde{d}_i}]^T$, $i = 1, 2, ...K$, denote the desired signal vector at i-th SU (acting as receiver), where $\tilde{d}_i$ is the number of desired data streams of user i. [Note that $\tilde{d}_i$ is not necessarily equal to d_i but $\sum_{i=1}^{K}\tilde{d}_i = \sum_{i=1}^{K} d_i$.] For each secondary user i, the data streams in $\tilde{\mathbf{s}}_i$ can be from multiple secondary transmitters. Therefore this is a multi-user multi-way transmission scheme [49–51]. Let

$$\tilde{\mathbf{s}} = [\tilde{\mathbf{s}}_1^T, \tilde{\mathbf{s}}_2^T, ..., \tilde{\mathbf{s}}_K^T]^T \text{ and } \mathbf{s} = [\mathbf{s}_1^T, \mathbf{s}_2^T, ..., \mathbf{s}_K^T]^T. \tag{3.1}$$

Then we have the relationship between transmitted signal and desired signal as

$$\tilde{\mathbf{s}} = \mathbf{E}\mathbf{s} \tag{3.2}$$

where $\mathbf{E}$ is a permutation matrix.

Assume that the transmitted signal $\mathbf{s}$ satisfies $\mathbb{E}\{\mathbf{s}\mathbf{s}^\dagger\} = \sigma_s^2\mathbf{I}$. Let $\mathbf{H}_{ir} \in \mathbb{C}^{N\times M}$ and $\mathbf{H}_{ri} \in \mathbb{C}^{M\times N}$, $i = 1, 2, ..., K$ denote the channel coefficient matrices from user i to relay station and from relay to user i. Also let $\mathbf{H}_{ip} \in \mathbb{C}^{M_p\times M}$, $\mathbf{H}_{pr} \in \mathbb{C}^{N\times J_p}$, $\mathbf{H}_{rp} \in \mathbb{C}^{M_p\times N}$ and $\mathbf{H}_{pi} \in \mathbb{C}^{M\times J_p}$, $i = 1, 2, ..., K$ denote the channel coefficient matrices from secondary user i to primary receiver, from primary transmitter to relay, from relay to primary receiver and from primary transmitter to secondary user i, respectively. Assume that these channel coefficients remain constant during one time slot and are known to the secondary network.

In the first time slot, let $\mathbf{T}_i \in \mathbb{C}^{M\times d_i}$ be the precoding matrix at the i-th secondary user. Then the transmitted signal from the i-th user is $\mathbf{x}_i = \mathbf{T}_i\mathbf{s}_i$. Secondary users' transmission will cause interference to primary receiver, which is given by

$$\mathbf{i}_1 = \sum_{i=1}^{K} \mathbf{H}_{ip}\mathbf{T}_i\mathbf{s}_i\,. \tag{3.3}$$

Let $\mathbf{x}_p \in \mathbb{C}^{J_p\times 1}$ denote the signal vector sent from primary transmitter during the first time slot with $\mathbb{E}\{\mathbf{x}_p\mathbf{x}_p^\dagger\} = \sigma_p^2\mathbf{I}$, and let $\mathbf{n}_r$ denote complex white Gaussian noise at the relay station with zero mean and covariance $\mathbb{E}\{\mathbf{n}_r\mathbf{n}_r^\dagger\} = \sigma_n^2\mathbf{I}$. Then the received signal at the relay station is given by

$$\mathbf{y}_r = \sum_{i=1}^{K} \mathbf{H}_{ir}\mathbf{T}_i\mathbf{s}_i + \mathbf{H}_{pr}\mathbf{x}_p + \mathbf{n}_r. \tag{3.4}$$

In the second time slot, the relay station transmits a linear combination of its received signal $\mathbf{y}_r$, using a precoding matrix $\mathbf{T}_r \in \mathbb{C}^{N\times N}$:

$$\mathbf{x}_r = \mathbf{T}_r\mathbf{y}_r = \sum_{i=1}^{K} \mathbf{T}_r\mathbf{H}_{ir}\mathbf{T}_i\mathbf{s}_i + \mathbf{T}_r\mathbf{H}_{pr}\mathbf{x}_p + \mathbf{T}_r\mathbf{n}_r. \tag{3.5}$$

The interference to primary receiver caused by relay's transmission is

$$\mathbf{i}_2 = \mathbf{H}_{rp}\mathbf{x}_r = \mathbf{H}_{rp}\mathbf{T}_r\left(\sum_{i=1}^{K} \mathbf{H}_{ir}\mathbf{T}_i\mathbf{s}_i + \mathbf{H}_{pr}\mathbf{x}_p + \mathbf{n}_r\right). \tag{3.6}$$

Let $\mathbf{n}_i$ denote the complex white Gaussian noise at secondary user i with zero mean and covariance $\mathbb{E}\{\mathbf{n}_i\mathbf{n}_i^\dagger\} = \sigma_n^2\mathbf{I}$ and let $\tilde{\mathbf{x}}_p \in \mathbb{C}^{J_p\times 1}$ denote the signal vector sent by the primary transmitter during the second time slot with $\mathbb{E}\{\tilde{\mathbf{x}}_p\tilde{\mathbf{x}}_p^\dagger\} = \sigma_p^2\mathbf{I}$. Then the received signal at i-th secondary user is given by

$$\mathbf{y}_i = \mathbf{H}_{ri}\left(\sum_{j=1}^{K} \mathbf{T}_r\mathbf{H}_{jr}\mathbf{T}_j\mathbf{s}_j + \mathbf{T}_r\mathbf{H}_{pr}\mathbf{x}_p + \mathbf{T}_r\mathbf{n}_r\right) + \mathbf{H}_{pi}\tilde{\mathbf{x}}_p + \mathbf{n}_i \tag{3.7}$$

Each secondary user i can subtract its own signal from $\mathbf{y}_i$. Then the self-interference free received signal at user i is

$$\tilde{\mathbf{y}}_i = \mathbf{H}_{ri}\left(\sum_{j=1, j\neq i}^{K} \mathbf{T}_r\mathbf{H}_{jr}\mathbf{T}_j\mathbf{s}_j + \mathbf{T}_r\mathbf{H}_{pr}\mathbf{x}_p + \mathbf{T}_r\mathbf{n}_r\right) + \mathbf{H}_{pi}\tilde{\mathbf{x}}_p + \mathbf{n}_i. \tag{3.8}$$

To ensure that the secondary network's transmission does not interfere with the primary network, we need $\mathbf{i}_1 = \mathbf{i}_2 = \mathbf{0}$ for every possible $\mathbf{s}_i$, $i = 1, 2, .., K$. This can be satisfied by letting

$$\mathbf{H}_{ip}\mathbf{T}_i = \mathbf{0}, \quad i = 1, 2, ..., K, \quad \text{and} \quad \mathbf{H}_{rp}\ \mathbf{T}_r = \mathbf{0}. \tag{3.9}$$

Denote the null space matrices of $\mathbf{H}_{ip}$, $i = 1, 2, ..., K$ and $\mathbf{H}_{rp}$ as $\mathbf{H}_{ip}^{\perp}$, $i = 1, 2, ..., K$ and $\mathbf{H}_{rp}^{\perp}$, respectively, i.e. the columns of $\mathbf{H}_{ip}^{\perp}$ span the null space of $\mathbf{H}_{ip}$, and similarly for $\mathbf{H}_{rp}^{\perp}$. Since $\mathbf{H}_{ip}$, $i = 1, 2, ..., K$ and $\mathbf{H}_{rp}$ are random channel coefficient matrices, they have full rank almost surely. Therefore $\text{rank}\{\mathbf{H}_{ip}^{\perp}\} = M - M_p$ and $\text{rank}\{\mathbf{H}_{rp}^{\perp}\} = N - M_p$. With matrices $\mathbf{P}_i, i = 1, 2, ..., K$ and $\mathbf{P}_r$ denoting arbitrary $(M - M_p) \times d_i$ and $(N - M_p) \times N$ matrices, to satisfy (3.9) we choose

$$\mathbf{T}_i = \mathbf{H}_{ip}^{\perp}\mathbf{P}_i, \quad i = 1, 2, ..., K, \quad \text{and} \quad \mathbf{T}_r = \mathbf{H}_{rp}^{\perp}\mathbf{P}_r. \tag{3.10}$$

We can obtain $\text{rank}\{\mathbf{T}_i\} = \min\{M - M_p, d_i\}$, $i = 1, 2, ..., K$ by choosing full rank $\mathbf{P}_i$; similarly $\text{rank}\{\mathbf{T}_r\} = N - M_p$ for a full rank $\mathbf{P}_r$. At user i, the number of antennas should at least satisfy $M > M_p$ in order to null interference to the primary network (otherwise $\mathbf{T}_i$ is null). Similarly, the number of relay antennas should satisfy $N > M_p$. As long as these two conditions are meet, $\mathbf{H}_{ip}^{\perp}, i = 1, 2, ..., K$ and $\mathbf{H}_{rp}^{\perp}$ are non-empty and the secondary user and relay precoders can be designed over $\mathbf{P}_i, i = 1, 2, ..., K$ and $\mathbf{P}_r$. Thus our design parameters become $\mathbf{P}_i$, $i = 1, 2, ..., K$ and $\mathbf{P}_r$ instead of $\mathbf{T}_i$, $i = 1, 2, ..., K$ and $\mathbf{T}_r$, respectively.

Note that since user i wants to transmit d_i independent datastreams, to mitigate inter-datastream interference (particularly when transmit power is high), intuitively it is desirable to choose $\text{rank}\{\mathbf{T}_i\} \geq d_i$ to achieve d_i linearly independent (preferably orthogonal) directions, which leads to $M - M_p \geq d_i$. Similarly, since the relay receives and transmits all datastreams from K secondary users, to mitigate the inter-datastream interference it is desirable to have number of relay antennas N such that $N - M_p \geq \sum_{i=1}^{K} d_i$. We do note that the conditions $M - M_p \geq d_i$ and $N - M_p \geq \sum_{i=1}^{K} d_i$ are preferred but not essential in the iterative algorithms presented in this chapter. As noted in Sec. 3.4.1, one needs $M - M_p \geq d_i$ for the non-iterative algorithm discussed therein.

Let $\hat{\mathbf{s}}_i$ denote the estimated signal at SU i, given by the decoding matrix $\mathbf{R}_i$ at secondary user i operating upon the self-interference free received signal $\tilde{\mathbf{y}}_i$ at SU i, as specified in (3.8). Then we have

$$\hat{\mathbf{s}}_i = \mathbf{R}_i\tilde{\mathbf{y}}_i = \mathbf{R}_i\mathbf{H}_{ri}\left(\sum_{j=1,j\neq i}^{K}\mathbf{T}_r\mathbf{H}_{jr}\mathbf{T}_j\mathbf{s}_j + \mathbf{T}_r\mathbf{H}_{pr}\mathbf{x}_p + \mathbf{T}_r\mathbf{n}_r\right) + \mathbf{R}_i\mathbf{H}_{pi}\tilde{\mathbf{x}}_p + \mathbf{R}_i\mathbf{n}_i. \quad (3.11)$$

Then the estimated signal at all K users is given by

$$\hat{\mathbf{s}} = [\hat{\mathbf{s}}_1^T, \hat{\mathbf{s}}_2^T, ..., \hat{\mathbf{s}}_K^T]^T. \quad (3.12)$$

We assume that all the signals from the primary/secondary users and the noise are all independent from each other. Let $\mathbf{E}_i$ denote a submatrix of the permutation matrix $\mathbf{E}$ such that $\mathbf{E}_i\mathbf{s}$ is the desired signal at receiver i. The sum MSE at all K secondary users can be expressed as

$$\mathbb{E}\{\| \hat{\mathbf{s}} - \tilde{\mathbf{s}} \|_2^2\} = \mathbb{E}\{\| \hat{\mathbf{s}} - \mathbf{E}\mathbf{s} \|_2^2\} = \sum_{i=1}^{K}\mathbb{E}\{\| \hat{\mathbf{s}}_i - \mathbf{E}_i\mathbf{s} \|_2^2\}. \quad (3.13)$$

Our goal is to minimize the sum MSE under a transmit power constraint for all transmitters while ensuring that the interference to primary receiver is zero.

Let $P_{tot,i}, i = 1, 2, ..., K$ and $P_{tot,r}$ be the maximum transmit power at secondary user i and relay node, respectively. Then using (3.13), the optimization problem under consideration can be expressed as

$$\text{P1}: \quad \min_{\mathbf{P}_r,\mathbf{P}_i,\mathbf{R}_i,i=1,2,...,K} \textstyle\sum_{i=1}^{K}\mathbb{E}\{\| \hat{\mathbf{s}}_i - \mathbf{E}_i\mathbf{s} \|_2^2\} \quad (3.14)$$

$$\text{s.t.} \quad \mathbb{E}\{\text{tr}\{\mathbf{T}_i\mathbf{s}_i\mathbf{s}_i^\dagger\mathbf{T}_i^\dagger\}\} \leq P_{tot,i}, \; i = 1, 2, ..., K, \quad (3.15)$$

$$\mathbb{E}\{\text{tr}\{\mathbf{x}_r\mathbf{x}_r^\dagger\}\} \leq P_{tot,r}. \quad (3.16)$$

The objective function is not jointly convex in $\mathbf{P}_r, \mathbf{P}_i$ and $\mathbf{R}_i$. therefore, problem P1 is not a convex optimization problem.

3.3 Iterative Algorithm

Due to the non-convexity of problem P1, optimizing $\mathbf{P}_r$, $\mathbf{P}_i$ and $\mathbf{R}_i$, $i = 1, 2, ..., K$ simultaneously to get the global optimal may require exhaustive numerical search whose complexity is likely to be too high. Therefore an iterative method is proposed to alternately optimize precoding matrices at secondary users, precoding matrix at relay station, and decoding matrices at the secondary receivers; each of these three subproblems is convex (shown in the Appendix).

3.3.1 Update Decoding Matrices

For given $\mathbf{T}_i$, $i = 1, 2, ..., K$ and $\mathbf{T}_r$, the optimization problem P1 with respect to receiver matrices $\mathbf{R}_i$, $i = 1, 2, ..., K$ can be decomposed into K independent optimization problems as

$$\text{P2}: \quad \min_{\mathbf{R}_i} \ \mathbb{E}\{\| \hat{\mathbf{s}}_i - \mathbf{E}_i \mathbf{s} \|_2^2\}, \quad i = 1, 2, ..., K. \tag{3.17}$$

Define $\mathbf{A}_i = \mathbf{H}_{ri}\mathbf{T}_r[\mathbf{H}_{1r}\mathbf{T}_1, ..., \mathbf{H}_{(i-1)r}\mathbf{T}_{i-1}, \mathbf{0}, \mathbf{H}_{(i+1)r}\mathbf{T}_{i+1}, ..., \mathbf{H}_{Kr}\mathbf{T}_K]$, $\mathbf{B}_i = [\mathbf{H}_{ri}\mathbf{T}_r\mathbf{H}_{pr}, \mathbf{H}_{pi}]$, $\mathbf{C}_i = [\mathbf{H}_{ri}\mathbf{T}_r, \mathbf{I}]$ $\check{\mathbf{n}}_i = [\mathbf{n}_r^T, \mathbf{n}_i^T]^T$ and $\check{\mathbf{x}}_p = [\mathbf{x}_p^T, \tilde{\mathbf{x}}_p^T]^T$. Then using (3.11) the estimated signal $\hat{\mathbf{s}}_i$ can be expressed as

$$\begin{aligned} \hat{\mathbf{s}}_i &= \mathbf{R}_i\mathbf{H}_{ri}\mathbf{T}_r \left(\textstyle\sum_{j=1, j\neq i}^{K} \mathbf{H}_{jr}\mathbf{T}_j\mathbf{s}_j + \mathbf{H}_{pr}\mathbf{x}_p + \mathbf{n}_r\right) + \mathbf{R}_i\mathbf{H}_{pi}\tilde{\mathbf{x}}_p + \mathbf{R}_i\mathbf{n}_i \\ &= \mathbf{R}_i \left(\mathbf{A}_i\mathbf{s} + \mathbf{B}_i\check{\mathbf{x}}_p + \mathbf{C}_i\check{\mathbf{n}}_i\right). \end{aligned} \tag{3.18}$$

Problem P2 is an unconstrained least-squares problem which is obviously convex having a well-known solution. Therefore the minimum MSE decoder at the i-th SU which solves problem P2 can be expressed as (details are provided in Appendix C):

$$\mathbf{R}_i^* = \sigma_s^2 \mathbf{E}_i \mathbf{A}_i^\dagger \left(\sigma_s^2 \mathbf{A}_i \mathbf{A}_i^\dagger + \sigma_p^2 \mathbf{B}_i \mathbf{B}_i^\dagger + \sigma_n^2 \mathbf{C}_i \mathbf{C}_i^\dagger \right)^{-1} . \tag{3.19}$$

3.3.2 Update Precoding Matrices at the Secondary Transmitters

Let $\tilde{\mathbf{E}}_i$ denote the submatrix of the permutation matrix $\mathbf{E}$ comprising its columns $\sum_{j=1}^{i-1} d_j + 1$ through $\sum_{j=1}^{i} d_j$. Then $\tilde{\mathbf{E}}_i \mathbf{s}_i$ denotes the desired signal at all K secondary users from the i-th secondary user's transmission. Consequently we can rewrite $\mathbf{Es}$ as

$$\mathbf{Es} = [\tilde{\mathbf{E}}_1, \tilde{\mathbf{E}}_2, ..., \tilde{\mathbf{E}}_K][\mathbf{s}_1^T, \mathbf{s}_2^T, ..., \mathbf{s}_K^T]^T. \tag{3.20}$$

With the optimal decoder $\mathbf{R}_i^*$ obtained from Sec. 3.3.1, set $\mathbf{R}_i = \mathbf{R}_i^*$ and define

$$\mathbf{D}_i = \left[(\mathbf{R}_1 \mathbf{H}_{r1})^T, ..., (\mathbf{R}_{i-1} \mathbf{H}_{r(i-1)})^T, \mathbf{0}, (\mathbf{R}_{i+1} \mathbf{H}_{r(i+1)})^T, ..., (\mathbf{R}_K \mathbf{H}_{rK})^T \right]^T \mathbf{T}_r \mathbf{H}_{ir} \mathbf{H}_{ip}^\perp . \tag{3.21}$$

Let $\hat{\mathbf{s}}^{(i)}$ denote the estimated signal at all K secondary users from i-th secondary user's transmission, after self-interference cancellation. Then using (3.21), we have $\hat{\mathbf{s}}^{(i)} = \mathbf{D}_i \mathbf{P}_i \mathbf{s}_i$. Since $\mathbf{s}_i, i = 1, 2, ..., K$ are independent, optimizing (3.14) w.r.t. $\mathbf{P}_i$, $i = 1, 2, ..., K$, reduces to the problem

$$\text{P3}: \qquad \min_{\mathbf{P}_i, i=1,2,...,K} \mathbb{E} \sum_{i=1}^{K} \| (\mathbf{D}_i \mathbf{P}_i - \tilde{\mathbf{E}}_i) \mathbf{s}_i \|_2^2 \tag{3.22}$$

$$\text{s.t.} \quad \mathbb{E}\{\text{tr}\{\mathbf{H}_{ip}^\perp \mathbf{P}_i \mathbf{s}_i \mathbf{s}_i^\dagger \mathbf{P}_i^\dagger \mathbf{H}_{ip}^{\perp\dagger}\}\} \le P_{tot,i}, \quad i = 1, 2, ..., K, \tag{3.23}$$

$$\mathbb{E}\{\text{tr}\{\mathbf{x}_r \mathbf{x}_r^\dagger\}\} \le P_{tot,r} \ . \tag{3.24}$$

This is a convex problem, as shown in Appendix B. Let $\mathbf{P} = \Big[\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_K\Big]$. The Lagrangian for problem P3 is given by

$$\begin{aligned} L(\boldsymbol{\lambda}, \mu, \mathbf{P}) \quad &= \textstyle\sum_{i=1}^{K} \Big\{ L_i(\lambda_i, \mu, \mathbf{P}_i) - \lambda_i P_{tot,i} \Big\} \\ &+ \mu \, \mathrm{tr}\{\mathbf{T}_r(\mathbf{H}_{pr}\mathbf{H}_{pr}^{\dagger}\sigma_p^2 \; + \sigma_n^2\mathbf{I})\mathbf{T}_r^{\dagger}\} - \mu P_{tot,r} \end{aligned} \tag{3.25}$$

where λ_i and μ are the dual variables for constraints (3.23) and (3.24), respectively, $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_K]$ and

$$\begin{aligned} L_i(\lambda_i, \mu, \mathbf{P}_i) \quad &= \mathrm{tr}\Big\{(\mathbf{D}_i\mathbf{P}_i - \tilde{\mathbf{E}}_i)(\mathbf{D}_i\mathbf{P}_i - \tilde{\mathbf{E}}_i)^{\dagger}\sigma_s^2 + \lambda_i \mathbf{H}_{ip}^{\perp}\mathbf{P}_i\mathbf{P}_i^{\dagger}\mathbf{H}_{ip}^{\perp\dagger}\sigma_s^2 \\ &+ \mu \mathbf{T}_r\mathbf{H}_{ir}\mathbf{H}_{ip}^{\perp}\mathbf{P}_i\mathbf{P}_i^{\dagger}\mathbf{H}_{ip}^{\perp\dagger}\mathbf{H}_{ir}^{\dagger}\mathbf{T}_r^{\dagger}\sigma_s^2\Big\}. \end{aligned} \tag{3.26}$$

The Lagrangian dual function is expressed as

$$g(\boldsymbol{\lambda}, \mu) = \min_{\mathbf{P}} L(\lambda, \mu, \mathbf{P}) = \sum_{i=1}^{L} g_i(\lambda_i, \mu) + \mu \, \mathrm{tr}\{\mathbf{T}_r(\mathbf{H}_{pr}\mathbf{H}_{pr}^{\dagger}\sigma_p^2 \; + \sigma_n^2\mathbf{I})\mathbf{T}_r^{\dagger}\} - \mu P_{tot,r}, \tag{3.27}$$

where $g_i(\lambda_i, \mu) = \min_{\mathbf{P}_i} \{L_i(\lambda_i, \mu, \mathbf{P}_i) - \lambda_i P_{tot,i}\}$. The dual optimization problem is given by

$$\text{P4}: \qquad \max_{\boldsymbol{\lambda},\mu} g(\boldsymbol{\lambda}, \mu) \tag{3.28}$$

$$\text{s.t. } \boldsymbol{\lambda} \succeq \mathbf{0}, \; \mu \geq 0. \tag{3.29}$$

To calculate $g_i(\lambda_i, \mu)$, take derivative of $L_i(\lambda_i, \mu, \mathbf{P}_i)$ with respect to $\mathbf{P}_i^{\dagger}$ and set it to zero to obtain the optimal $\mathbf{P}_i$ given λ_i and μ, as

$$\mathbf{P}_i^{*}(\lambda_i, \mu) = \Big(\mathbf{D}_i^{\dagger}\mathbf{D}_i + \lambda_i\mathbf{H}_{ip}^{\perp\dagger}\mathbf{H}_{ip}^{\perp} + \mu\mathbf{H}_{ip}^{\perp\dagger}\mathbf{H}_{ir}^{\dagger}\mathbf{T}_r^{\dagger}\mathbf{T}_r\mathbf{H}_{ir}\mathbf{H}_{ip}^{\perp}\Big)^{-1}\mathbf{D}_i^{\dagger}\tilde{\mathbf{E}}_i\,.$$

In solving the dual problem P4 over the two dual variables, we employ a two-loop algorithm. In the inner loop, optimal $\boldsymbol{\lambda}$ is solved for a fixed value of μ, so we denote the optimal $\boldsymbol{\lambda}$ as

$\boldsymbol{\lambda}^*(\mu)$. In the outer loop, μ^* is found using the value of $\boldsymbol{\lambda}^*(\mu)$. The inner loop problem can be expressed as $g_{\lambda^*}(\mu) = \max_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}, \mu),$ s.t. $\boldsymbol{\lambda} \succeq 0$ and the outer loop problem can then be expressed as $\max_\mu g_{\boldsymbol{\lambda}^*}(\mu)$ s.t. $\mu \geq 0$. The dual problem is convex, therefore both the outer and inner loop problems are convex.

Define

$$\mathbf{V}_{1i}(\lambda_i^*(\mu), \mu) = \sigma_s^2 \mathbf{H}_{ip}^{\perp} \mathbf{P}_i^*(\lambda_i^*(\mu), \mu) \mathbf{P}_i^{*\dagger}(\lambda_i^*(\mu), \mu) \mathbf{H}_{ip}^{\perp\dagger}. \tag{3.30}$$

Then for the inner loop, for a fixed value of μ, the optimal value $\lambda_i^*(\mu)$ of $\lambda_i(\mu), i = 1, 2, ..., K$ should satisfy (complementary-slackness condition [18]) $\lambda_i^*(\mu) \left(\mathrm{tr}\left\{\mathbf{V}_{1i}(\lambda_i^*(\mu), \mu)\right\} - P_{tot,i}\right) = 0$. Therefore we pick $\lambda_i^*(\mu) = 0$ if $\mathrm{tr}\left\{\mathbf{V}_{1i}(0, \mu)\right\} \leq P_{tot,i}$ and $\lambda_i^*(\mu) > 0$ if $\mathrm{tr}\left\{\mathbf{V}_{1i}(0, \mu)\right\} > P_{tot,i}$. For the $\lambda_i^*(\mu) > 0$ case, we can find $\lambda_i^*(\mu)$ by solving the following problem

$$\max_{\lambda_i} g(\lambda_i, \mu) = \max_{\lambda_i} \left\{L_i(\lambda_i, \mu, \mathbf{P}_i^*(\lambda_i, \mu)) - \lambda_i P_{tot,i}\right\} \tag{3.31}$$

$$\text{s.t. } \lambda_i > 0, \quad i = 1, 2, ..., K \tag{3.32}$$

using a line search over λ_i. We want to find optimal $\lambda_i^*(\mu)$ such that $\mathrm{tr}\left\{\mathbf{V}_{1i}(\lambda_i^*(\mu), \mu)\right\} = P_{tot,i}$. Set the lower bound of λ_i as $\lambda_{i,l} = 0$ since by complementary-slackness $\mathrm{tr}\left\{\mathbf{V}_{1i}(0, \mu)\right\} > P_{tot,i}$ for the $\lambda_i^*(\mu) > 0$ case. Then we increase the value of λ_i to find its upper bound $\lambda_{i,u}$ such that $\mathrm{tr}\left\{\mathbf{V}_{1i}(\lambda_{i,u}(\mu), \mu)\right\} < P_{tot,i}$. A bisection method can then be employed to find the optimal $\lambda_i^*(\mu)$. Finally we solve for optimal μ^* in the outer loop. Define

$$\mathbf{Z} = \mathbf{T}_r \Big(\sum_{i=1}^{K} \mathbf{H}_{ir} \mathbf{V}_{1i}(\lambda_i^*(0), 0) \mathbf{H}_{ir}^\dagger + \sigma_p^2 \mathbf{H}_{pr} \mathbf{H}_{pr}^\dagger + \sigma_n^2 \mathbf{I} \Big) \mathbf{T}_r^\dagger.$$

The optimal μ^* should satisfy (complementary-slackness condition [18]) $\mu^* = 0$ if $\mathrm{tr}\left\{\mathbf{Z}\right\} < P_{tot,r}$ and $\mu^* > 0$ if $\mathrm{tr}\left\{\mathbf{Z}\right\} \geq P_{tot,r}$ For the $\mu^* > 0$ case, μ^* can be found by solving the following problem using a line search (a bisection method can be employed similar to the one we used to solve the inner dual problem):

$$\max_{\mu} g_{\lambda^*}(\mu) \quad \text{s.t. } \mu > 0$$

.

3.3.3 Updating Precoding Matrix at Relay Station

Define the matrices $\check{\mathbf{H}}_1$, $\check{\mathbf{H}}_2$, $\check{\mathbf{H}}_p$, $\mathbf{E}_l^{(i)}$, $\mathbf{E}_r^{(i)}$ and $\mathbf{R}$ and vector $\mathbf{n}$

$$\check{\mathbf{H}}_1 = [(\mathbf{R}_1\mathbf{H}_{r1})^T, (\mathbf{R}_2\mathbf{H}_{r2})^T, ..., (\mathbf{R}_K\mathbf{H}_{rK})^T]^T, \tag{3.33}$$

$$\check{\mathbf{H}}_2 = [\mathbf{H}_{1r}\mathbf{T}_1, \mathbf{H}_{2r}\mathbf{T}_2, ..., \mathbf{H}_{Kr}\mathbf{T}_K], \tag{3.34}$$

$$\check{\mathbf{H}}_p = [(\mathbf{R}_1\mathbf{H}_{p1})^T, (\mathbf{R}_2\mathbf{H}_{p2})^T, ..., (\mathbf{R}_K\mathbf{H}_{pK})^T]^T, \tag{3.35}$$

$$\mathbf{E}_l^{(i)} = \text{diag}\{\mathbf{0}_{\tilde{d}_1\times\tilde{d}_1}, ..., \mathbf{0}_{\tilde{d}_{i-1}\times\tilde{d}_{i-1}}, \mathbf{I}_{\tilde{d}_i\times\tilde{d}_i}\mathbf{0}_{\tilde{d}_{i+1}\times\tilde{d}_{i+1}}, ..., \mathbf{0}_{\tilde{d}_K\times\tilde{d}_K}\}, \tag{3.36}$$

$$\mathbf{E}_r^{(i)} = \text{diag}\{\mathbf{0}_{d_1\times d_1}, ..., \mathbf{0}_{d_{i-1}\times d_{i-1}}, \mathbf{I}_{d_i\times d_i}, \mathbf{0}_{d_{i+1}\times d_{i+1}}, ..., \mathbf{0}_{d_K\times d_K}\}, \tag{3.37}$$

$$\mathbf{R} = \text{diag}\{\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_K\}, \quad \mathbf{n} = [\mathbf{n}_1^T, \mathbf{n}_2^T, ..., \mathbf{n}_K^T]^T. \tag{3.38}$$

Then from (3.11), (3.12) and (3.33)-(3.37), the estimated signal vector at secondary receivers after self-interference cancellation can be expressed as

$$\hat{\mathbf{s}} = \check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\mathbf{s} - \sum_{i=1}^{K}\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\mathbf{s} + \check{\mathbf{H}}_1\mathbf{T}_r\mathbf{H}_{pr}\mathbf{x}_p + \check{\mathbf{H}}_1\mathbf{T}_r\mathbf{n}_r + \check{\mathbf{H}}_p\tilde{\mathbf{x}}_p + \mathbf{R}\mathbf{n} \tag{3.39}$$

where $\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\mathbf{s}$ is the received signal part at secondary receiver i from itself (user i). Define

$$\mathbf{V}_3 = \check{\mathbf{H}}_2\check{\mathbf{H}}_2^\dagger\sigma_s^2 + \mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger\sigma_p^2 + \sigma_n^2\mathbf{I} \text{ and } \mathbf{V}_2(\mathbf{P}_r) = \mathbf{H}_{rp}^\perp\mathbf{P}_r\mathbf{V}_3\mathbf{P}_r^\dagger\mathbf{H}_{rp}^{\perp\dagger}. \tag{3.40}$$

For given $\mathbf{T}_i$ and $\mathbf{R}_i$, $i = 1, 2, ..., K$, minimizing the sum MSE w.r.t. $\mathbf{P}_r$ can then be formulated as

$$\text{P5:} \qquad \min_{\mathbf{P}_r} \mathbb{E}\{\| \hat{\mathbf{s}} - \mathbf{E}\mathbf{s} \|_2^2\} \tag{3.41}$$

$$\text{s.t.} \;\; \text{tr}\{\mathbf{V}_2(\mathbf{P}_r)\} \leq P_{tot,r}. \tag{3.42}$$

This is a convex problem as shown in Appendix C. The Lagrangian of problem P5 is

$$L(\lambda, \mathbf{P}_r) = \mathbb{E}\,\text{tr}\{(\hat{\mathbf{s}} - \mathbf{E}\mathbf{s})(\hat{\mathbf{s}} - \mathbf{E}\mathbf{s})^\dagger\} + \text{tr}\{\lambda \mathbf{V}_2(\mathbf{P}_r)\} - \lambda P_{tot,r}. \tag{3.43}$$

The Lagrangian dual function is $g(\lambda) = \min_{\mathbf{P}_r} L(\lambda, \mathbf{P}_r)$. To calculate the dual function, set derivative of $L(\lambda, \mathbf{P}_r)$ w.r.t. $\mathbf{P}_r^\dagger$ to zero. Then we can obtain the optimal value of $\mathbf{P}_r$ for a fixed value of λ as

$$\text{vec}\{\mathbf{P}_r^*(\lambda)\} = \Big\{\mathbf{V}_3^T \otimes [\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}] - \sum_{i=1}^{K}[\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\check{\mathbf{H}}_2^\dagger]^T \otimes [\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}]\sigma_s^2$$

$$+\lambda\mathbf{V}_3^T \otimes [\mathbf{H}_{rp}^{\perp\dagger}\mathbf{H}_{rp}^{\perp}]\Big\}^{-1} \times \text{vec}\left\{\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\mathbf{E}\check{\mathbf{H}}_2^\dagger\sigma_s^2 - \sum_{i=1}^{K}\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\mathbf{E}\mathbf{E}_r^{(i)}\check{\mathbf{H}}_2^\dagger\sigma_s^2\right\}. \tag{3.44}$$

The optimal value λ^* of the dual variable λ satisfies the complementary-slackness condition

$\lambda^*\left(\text{tr}\left\{\mathbf{V}_2(\mathbf{P}_r^*(\lambda^*))\right\} - P_{tot,r}\right) = 0$ and we should have $\lambda^* = 0$ if $\text{tr}\left\{\mathbf{V}_2(\mathbf{P}_r^*(0))\right\} < P_{tot,r}$ and $\lambda^* > 0$ if $\text{tr}\left\{\mathbf{V}_2(\mathbf{P}_r^*(0))\right\} \geq P_{tot,r}$. The dual problem is always convex, so the optimal dual variable λ^* in the case $\lambda^* > 0$ can be obtained by solving the following dual problem using a line search (a bisection method can be used here):

$$\max_{\lambda} g(\lambda) \quad \text{s.t. } \lambda > 0,$$

where $g(\lambda) = L(\lambda, \mathbf{P}_r(\lambda))$.

3.3.4 Algorithm Summary

The iterative algorithm is summarized as follows:

1) Initialize $\mathbf{P}_i$ $(i = 1, 2, ..., K)$ and $\mathbf{P}_r$ such that $\mathrm{tr}\{\mathbf{H}_{ip}^{\perp}\mathbf{P}_i\mathbf{P}_i^{\dagger}\mathbf{H}_{ip}^{\perp\dagger}\} = P_{tot,i}/\sigma_s^2$ and $\mathbb{E}\,\mathrm{tr}\{\mathbf{H}_{rp}^{\perp}\mathbf{P}_r\mathbf{y}_r\mathbf{y}_r^{\dagger}\mathbf{P}_r^{\dagger}\mathbf{H}_{rp}^{\perp\dagger}\} = P_{tot,r}$. [One way to accomplish this is to randomly and independently choose each element in $\mathbf{P}_i$ and $\mathbf{P}_r$, and then scale the respective matrices to satisfy the aforementioned power constraints. An alternative (used in our simulations) is to use the results of the non-iterative algorithm proposed in Sec. 3.4 for initialization; this algorithm needs no initialization.]

2) Repeat (iterate):

   – Update $\mathbf{R}_i$, $i = 1, 2, ..., K$ using (3.19) for fixed $\mathbf{P}_i$, $i = 1, 2, ..., K$ and $\mathbf{P}_r$. We use the latest available estimates of $\mathbf{P}_i$ and $\mathbf{P}_r$.

   – Update $\mathbf{P}_i$, $i = 1, 2, ..., K$ by solving problem P3 for fixed $\mathbf{R}_i$, $i = 1, 2, ..., K$ and $\mathbf{P}_r$. We use the latest available estimates of $\mathbf{R}_i$ and $\mathbf{P}_r$.

   – Update $\mathbf{P}_r$ by solving problem P5 for fixed $\mathbf{R}_i$, $i = 1, 2, ..., K$ and $\mathbf{P}_i$, $i = 1, 2, ..., K$. We use the latest available estimates of $\mathbf{R}_i$ and $\mathbf{P}_i$.

   Continue until sum MSE converges.

Since all subproblems have exactly the same objective function (sum MSE) and each subproblem is convex, the objective function is decreasing in each subproblem, hence in each iteration. The sum MSE is lower bounded by zero. As a decreasing sequence that is lower bounded always converges, the proposed iterative algorithm is convergent to a local optimum.

3.3.5 Effects of Allowing Interference to Primary Users

In this section we investigate the case when the interference from secondary users to primary network is not completely removed (nulled). This may improve the secondary network's performance because the constraints on interference to PU are relaxed from zero to some nonzero upperbound, namely I_{tot} (the same for the two time slots). The optimization problem becomes

$$\begin{aligned} \text{P6}: \quad & \min_{\mathbf{T}_r,\mathbf{T}_i,\mathbf{R}_i,i=1,2,...,K} \mathbb{E}\{\| \hat{\mathbf{s}} - \mathbf{E}\mathbf{s} \|_2^2\} && (3.45)\\ \text{s.t.} \quad & \mathbb{E}\{\text{tr}\{\mathbf{T}_i\mathbf{s}_i\mathbf{s}_i^\dagger\mathbf{T}_i^\dagger\}\} \le P_{tot,i},\ i=1,2,...,K, && (3.46)\\ & \mathbb{E}\{\text{tr}\{\mathbf{x}_r\mathbf{x}_r^\dagger\}\} \le P_{tot,r}, && (3.47)\\ & \textstyle\sum_{i=1}^K \mathbb{E}\{\text{tr}\{\mathbf{H}_{ip}\mathbf{T}_i\mathbf{s}_i\mathbf{s}_i^\dagger\mathbf{T}_i^\dagger\mathbf{H}_{ip}^\dagger\}\} \le I_{tot}, && (3.48)\\ & \mathbb{E}\{\text{tr}\{\mathbf{H}_{rp}\mathbf{x}_r\mathbf{x}_r^\dagger\mathbf{H}_{rp}^\dagger\}\} \le I_{tot}. && (3.49) \end{aligned}$$

As in the previous sections, P6 is not a convex problem, therefore we decompose it into three convex subproblems and solve them iteratively.

First consider the design of decoding matrices $\mathbf{R}_i$ given $\mathbf{T}_i, 1 \le i \le K$ and $\mathbf{T}_r$. Since the design of $\mathbf{R}_i$ does not affect interference to the primary network, it is exactly the same as in Sec. 3.3.1.

Similar to (3.21), define

$$\tilde{\mathbf{D}}_i = \left[(\mathbf{R}_1\mathbf{H}_{r1})^T, ..., (\mathbf{R}_{i-1}\mathbf{H}_{ri-1})^T, \mathbf{0}, (\mathbf{R}_{i+1}\mathbf{H}_{ri+1})^T, ..., (\mathbf{R}_K\mathbf{H}_{rK})^T\right]^T \mathbf{T}_r\mathbf{H}_{ir}.$$

For user precoding design, given $\mathbf{R}_i$, $1 \le i \le K$, and $\mathbf{T}_r$, the problem is formulated as follows:

$$\begin{aligned} \text{P7}: \quad & \min_{\mathbf{T}_i,i=1,2,...,K} \mathbb{E} \textstyle\sum_{i=k}^K \| (\tilde{\mathbf{D}}_i\mathbf{T}_i - \tilde{\mathbf{E}}_i)\mathbf{s}_i \|_2^2 && (3.50)\\ & \text{s.t. } (3.46) - (3.49) \text{ are satisfied.} && (3.51) \end{aligned}$$

Following a similar argument as in Appendix B, we can show that problem P7 is a convex optimization problem and the optimal solution for $\mathbf{T}_i, 1 \le i \le K$ can be obtained. Define

$$\mathbf{V}_4 = \mathbf{T}_r(\mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger\sigma_p^2 + \sigma_n^2\mathbf{I})\mathbf{T}_r^\dagger \text{ and } \mathbf{V}_5 = \mathbf{H}_{rp}\mathbf{V}_4\mathbf{H}_{rp}^\dagger. \quad (3.52)$$

With $\mathbf{P} = [\mathbf{P}_1,\ \mathbf{P}_2,\ \cdots, \mathbf{P}_K,]$, the Lagrangian for problem P7 can be expressed as

$$\begin{aligned} L(\boldsymbol{\lambda}, \mu, \beta, \theta, \mathbf{P}) &= \sum_{i=1}^{K} \Big\{ L_i(\lambda_i, \mu, \beta, \theta, \mathbf{P}_i) - \lambda_i P_{tot,i} \Big\} \\ &\quad + \mu\, \mathrm{tr}\{\mathbf{V}_4\} + \theta\, \mathrm{tr}\{\mathbf{V}_5\} - \mu P_{tot,r} - \theta I_{tot} - \beta I_{tot} \end{aligned} \tag{3.53}$$

where λ_i, β, μ and θ are the dual variables, $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_K]$ and

$$\begin{aligned} L_i(\lambda_i, \beta, \mu, \theta, \mathbf{T}_i) &= \mathrm{tr}\Big\{ (\tilde{\mathbf{D}}_i \mathbf{T}_i - \tilde{\mathbf{E}}_i)(\tilde{\mathbf{D}}_i \mathbf{T}_i - \tilde{\mathbf{E}}_i)^\dagger \sigma_s^2 + \lambda_i \mathbf{T}_i \mathbf{T}_i^\dagger \sigma_s^2 + \beta \mathbf{H}_{ip} \mathbf{T}_i \mathbf{T}_i^\dagger \mathbf{H}_{ip}^\dagger \sigma_s^2 \\ &\quad + \mu \mathbf{T}_r \mathbf{H}_{ir} \mathbf{T}_i \mathbf{T}_i^\dagger \mathbf{H}_{ir}^\dagger \mathbf{T}_r^\dagger \sigma_s^2 + \theta \mathbf{H}_{rp} \mathbf{T}_r \mathbf{H}_{ir} \mathbf{T}_i \mathbf{T}_i^\dagger \mathbf{H}_{ir}^\dagger \mathbf{T}_r^\dagger \mathbf{H}_{rp}^\dagger \sigma_s^2 \Big\}. \end{aligned} \tag{3.54}$$

With $\mathbf{T} = [\mathbf{T}_1,\ \mathbf{T}_2,\ \cdots, \mathbf{T}_K,]$, the Lagrangian dual function $g(\boldsymbol{\lambda}, \mu, \beta, \theta) = \min_{\mathbf{T}} L(\boldsymbol{\lambda}, \mu, \beta, \theta, \mathbf{T})$ is expressed as

$$g(\boldsymbol{\lambda}, \mu, \beta, \theta) = \sum_{i=1}^{L} g_i(\lambda_i, \mu, \beta, \theta) + \mu\, \mathrm{tr}\{\mathbf{V}_4\} + \theta\, \mathrm{tr}\{\mathbf{V}_5\} - \mu P_{tot,r} - \theta I_{tot} - \beta I_{tot}, \tag{3.55}$$

where $g_i(\lambda_i, \mu, \beta, \theta) = \min_{\mathbf{T}_i} \{L_i(\lambda_i, \mu, \beta, \theta \mathbf{T}_i) - \lambda_i P_{tot,i}\}$. The dual optimization problem is

$$\max_{\boldsymbol{\lambda}, \mu, \beta, \theta} g(\boldsymbol{\lambda}, \mu, \beta, \theta)$$

$$\text{s.t. } \boldsymbol{\lambda} \succeq \mathbf{0},\ \mu \geq 0,\ \beta \geq 0,\ \theta \geq 0.$$

To calculate $g_i(\lambda_i, \mu, \beta, \theta)$, take derivative of $L_i(\lambda_i, \mu, \beta, \theta, \mathbf{T}_i)$ with respect to $\mathbf{T}_i^\dagger$ and set it to zero to obtain the optimal $\mathbf{T}_i$ given λ_i, μ, β and θ, as

$$\mathbf{T}_i^*(\lambda_i, \mu, \beta, \theta) = \left(\tilde{\mathbf{D}}_i^\dagger \tilde{\mathbf{D}}_i + \lambda_i \mathbf{I} + \mu \mathbf{H}_{ir}^\dagger \mathbf{T}_r^\dagger \mathbf{T}_r \mathbf{H}_{ir} + \beta \mathbf{H}_{ip}^\dagger \mathbf{H}_{ip} + \theta \mathbf{H}_{ir}^\dagger \mathbf{T}_r^\dagger \mathbf{H}_{rp}^\dagger \mathbf{H}_{rp} \mathbf{T}_r \mathbf{H}_{ir} \right)^{-1} \tilde{\mathbf{D}}_i^\dagger \tilde{\mathbf{E}}_i.$$

In solving the dual problem, there are four dual variables. A two-loop algorithm, similar to the one that solves P4 is employed here. In the inner loop, optimal $\boldsymbol{\lambda}$ is solved for fixed values of μ, β and θ, and the optimal $\boldsymbol{\lambda}$ is denoted by $\boldsymbol{\lambda}^*(\mu, \beta, \theta)$. In the outer loop, μ^*,

β^* and θ^* are jointly found using the value of $\boldsymbol{\lambda}^*(\mu,\beta,\theta)$ by a subgradient method. The inner loop problem can be expressed as $g_{\boldsymbol{\lambda}^*}(\mu,\beta,\theta) = \max_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda},\mu,\beta,\theta)$, s.t. $\boldsymbol{\lambda} \succeq 0$ and the outer loop problem can then be expressed as $\max_{\mu,\beta,\theta} g_{\boldsymbol{\lambda}^*}(\mu,\beta,\theta)$ s.t. $\mu,\beta,\theta \geq 0$. The dual problem is convex, therefore both the outer and inner loop problems are convex. The inner loop problem is similar as the one in problem P4, therefore the steps are omitted due to space limitations. For the outer loop, a subgradient method is employed to solve for the optimal μ^*, β^* and θ^*. The subgradient of $g_{\boldsymbol{\lambda}^*}(\mu,\beta,\theta)$ w.r.t. these three dual variables are given by

$$\partial_\mu g_{\boldsymbol{\lambda}^*}(\mu,\beta,\theta) = \mathrm{tr}\{\mathbf{V}_4 + \textstyle\sum_{i=1}^K \mathbf{T}_r\mathbf{H}_{ir}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}_{ir}^\dagger\mathbf{T}_r^\dagger\sigma_s^2\} - P_{tot,r} \tag{3.56}$$

$$\partial_\beta g_{\boldsymbol{\lambda}^*}(\mu,\beta,\theta) = \textstyle\sum_{i=1}^K \mathrm{tr}\{\mathbf{H}_{ip}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}_{ip}^\dagger\sigma_s^2\} - I_{tot} \tag{3.57}$$

$$\partial_\theta g_{\boldsymbol{\lambda}^*}(\mu,\beta,\theta) = \mathrm{tr}\{\mathbf{V}_5 + \textstyle\sum_{i=1}^K \mathbf{H}_{rp}\mathbf{T}_r\mathbf{H}_{ir}\mathbf{T}_i\mathbf{T}_i^\dagger\mathbf{H}_{ir}^\dagger\mathbf{T}_r^\dagger\mathbf{H}_{rp}^\dagger\sigma_s^2\} - I_{tot}. \tag{3.58}$$

For designing the precoder at the relay for given $\mathbf{R}_i$ and $\mathbf{T}_i$, $1 \leq i \leq K$, the optimization problem P6 can be reformulated as follows where $\check{\mathbf{H}}_2$ and $\mathbf{V}_3$ are as in (3.34) and (3.40):

$$\text{P8:} \qquad \min_{\mathbf{T}_r} \mathbb{E}\|\hat{\mathbf{s}} - \mathbf{E}\mathbf{s}\|^2 \tag{3.59}$$

$$\text{s.t. } \mathrm{tr}\left\{\mathbf{T}_r\mathbf{V}_3\mathbf{T}_r^\dagger\right\} \leq P_{tot,r} \tag{3.60}$$

$$\mathrm{tr}\left\{\mathbf{H}_{rp}\mathbf{T}_r\mathbf{V}_3\mathbf{T}_r^\dagger\mathbf{H}_{rp}^\dagger\right\} \leq I_{tot}\ . \tag{3.61}$$

The Lagrangian of problem P8 can be expressed as

$$\begin{aligned} L(\lambda,\mu,\mathbf{T}_r) = \; & \mathbb{E}\,\mathrm{tr}\{(\hat{\mathbf{s}} - \mathbf{E}\mathbf{s})(\hat{\mathbf{s}} - \mathbf{E}\mathbf{s})^\dagger\} + \lambda\,\mathrm{tr}\{\mathbf{T}_r\mathbf{V}_3\mathbf{T}_r^\dagger\} \\ & + \mu\,\mathrm{tr}\left\{\mathbf{H}_{rp}\mathbf{T}_r\mathbf{V}_3\mathbf{T}_r^\dagger\mathbf{H}_{rp}^\dagger\right\} - \lambda P_{tot,r} - \mu I_{tot}. \end{aligned} \tag{3.62}$$

By setting $\frac{\partial L(\lambda,\mu,\mathbf{T}_r,\gamma)}{\partial \mathbf{T}_r^\dagger} = 0$ and using (3.33)-(3.37), we obtain the optimal $\mathbf{T}_r^*$ satisfying

$$\begin{aligned} \mathrm{vec}\{\mathbf{T}_r^*\} = \Big\{ & \mathbf{V}_3^T \otimes [\check{\mathbf{H}}_1^\dagger \check{\mathbf{H}}_1] - \textstyle\sum_{i=1}^K [\check{\mathbf{H}}_2 \mathbf{E}_r^{(i)} \mathbf{H}_2^\dagger]^T \otimes [\check{\mathbf{H}}_1^\dagger \mathbf{E}_l^{(i)} \check{\mathbf{H}}_1]\sigma_s^2 + \lambda \mathbf{V}_3^T \otimes \mathbf{I} \\ & + \mu \mathbf{V}_3 \otimes \Big(\mathbf{H}_{rp}^\dagger \mathbf{H}_{rp}\Big)\Big\}^{-1} \times \mathrm{vec}\Big\{\check{\mathbf{H}}_1^\dagger \mathbf{E} \check{\mathbf{H}}_2^\dagger - \textstyle\sum_{i=1}^K \check{\mathbf{H}}_1^\dagger \mathbf{E}_l^{(i)} \mathbf{E} \mathbf{E}_r^{(i)} \check{\mathbf{H}}_2^\dagger\Big\}\sigma_s^2. \end{aligned} \tag{3.63}$$

To obtain the optimal dual variables λ^* and μ^*, a subgradient method can be employed. The subgradients of the Lagrangian dual function $g(\lambda, \mu) = L(\lambda, \mu, \mathbf{T}_r^*, \gamma)$ w.r.t. these two dual variables are given by

$$\partial_\lambda L(\lambda, \mu, \mathbf{T}_r^*) = \mathrm{tr}\Big\{\mathbf{T}_r^* \mathbf{V}_3 \mathbf{T}_r^{*\dagger}\Big\} - P_{tot,r} \tag{3.64}$$

$$\partial_\mu L(\lambda, \mu, \mathbf{T}_r^*) = \mathrm{tr}\Big\{\mathbf{H}_{rp} \mathbf{T}_r^* \mathbf{V}_3 \mathbf{T}_r^{*\dagger} \mathbf{H}_{rp}^\dagger\Big\} - I_{tot}. \tag{3.65}$$

Compared to the zero PU interference approach of Secs. 3.3.1-3.3.3, in the current case we have higher computational complexity since in solving for $\mathbf{T}_i$ (compared with solving for $\mathbf{P}_i$ in Sec. 3.3.2), the outer loop has three dual variables instead of one variable in Sec. 3.3.2, therefore, a subgradient method is employed in the current case instead of a line search as in Sec. 3.3.2. For solving $\mathbf{T}_r$ (compared with solving for $\mathbf{P}_r$), there are two dual variables instead of one, hence again, a subgradient method is used instead of a line search.

3.4 Non-iterative Algorithm

In this section we propose a non-iterative algorithm in order to reduce the computational complexity. We want to design an algorithm such that each secondary user can design its precoding and decoding matrices based on local channel matrices. Then the precoding matrix at the relay station is designed to minimize the sum MSE of all secondary users. This algorithm is based on a matrix (pseudo-)distance defined following [34] (see also [35]). Let $\mathbf{U}$ be an $n \times m$ ($m \leq n$) (complex) matrix with orthonormal columns and let $\mathbf{A}$ be an arbitrary $n \times p$ complex matrix. The orthogonal projection of $\mathbf{A}$ onto the subspace spanned by the columns of $\mathbf{U}$ is given by $\tilde{\mathbf{A}} = \mathbf{U}\mathbf{U}^\dagger\mathbf{A}$. Define the distance between $\mathbf{A}$ and its orthogonal projection $\tilde{\mathbf{A}}$ as

$$\|\mathbf{A}, \mathbf{U}\|_{\text{dis}} = \|\mathbf{A} - \mathbf{U}\mathbf{U}^\dagger\mathbf{A}\|_F. \tag{3.66}$$

If the columns of $\mathbf{A}$ lie in the subspace spanned by $\mathbf{U}$, then $\|\mathbf{A}, \mathbf{U}\|_{\text{dis}} = 0$. If none of the columns of $\mathbf{A}$ lie in the subspace spanned by $\mathbf{U}$, then $\|\mathbf{A}, \mathbf{U}\|_{\text{dis}} = \|\mathbf{A}\|_F$.

3.4.1 Matrix Distance based Secondary User Precoding/Decoding Matrices Design

Precoding matrix $\mathbf{T}_i$ at the i-th secondary user is designed to minimize its distance to $\mathbf{H}_{ir}^\dagger$ so that "more information" can be transmitted to the relay station. (If a symbol b is transmitted along a (column) direction $\mathbf{p}_1$, then by the Cauchy-Schwartz inequality, the received bit $\hat{b} = \mathbf{r}_1^T\mathbf{p}_1$ has maximum magnitude if and only if $\mathbf{p}_1$ and $\mathbf{r}_1^\ddagger$ are in the same one-dimensional subspace. This motivates picking columns of the precoding matrix so that they span the same subspace as the columns of $\mathbf{H}_{ir}^\dagger$.) We also need to make sure that secondary users' transmission will not degrade primary user's QoS; therefore, $\mathbf{T}_i = \mathbf{H}_{ip}^\perp\mathbf{P}_i$ should be satisfied. We assume *a priori* that $\mathbf{P}_i$ has orthonormal columns, i.e. $\mathbf{P}_i^\dagger\mathbf{P}_i = \mathbf{I}$; this is suboptimal but leads to analytical tractability. Another reason for this assumption is that each column corresponds to one datastream's transmit beamforming vector. For a user that is transmitting multiple datastreams, it is reasonable to make these beamforming vectors orthogonal to each other to mitigate inter-datastream interference. The requirement $\mathbf{P}_i^\dagger\mathbf{P}_i = \mathbf{I}$ also implies that the power scaling of each datastream is set to be the same; power allocation and fairness among multiple datastreams is outside the scope of this chapter. A scaler β_i is introduced to satisfy the transmit power constraint of user i. Let $\mathbf{T}_i = \beta_i\mathbf{H}_{ip}^\perp\mathbf{P}_i$. Then the relay receives the signal $\mathbf{H}_{ir}\mathbf{T}_i\mathbf{s}_i = \beta_i\mathbf{H}_{ir}\mathbf{H}_{ip}^\perp\mathbf{P}_i$ from the i-th secondary user. Since $\mathbf{H}_{ip}^\perp$ is "fixed," we make it a part of $\mathbf{H}_{ir}$ for the design of $\mathbf{P}_i$ and define an equivalent channel matrix as $\mathbf{G}_{ir} = \mathbf{H}_{ir}\mathbf{H}_{ip}^{\perp}$. Then the precoding design at user i is formulated as

$$\text{P9:} \quad \mathbf{P}_i^* = \arg\min_{\mathbf{P}_i^\dagger \mathbf{P}_i = \mathbf{I}} \|\mathbf{G}_{ir}^\dagger, \mathbf{P}_i\|_{\text{dis}}^2 = \arg\min_{\mathbf{P}_i^\dagger \mathbf{P}_i = \mathbf{I}} \|\mathbf{G}_{ir}^\dagger - \mathbf{P}_i\mathbf{P}_i^\dagger\mathbf{G}_{ir}^\dagger\|_F^2. \tag{3.67}$$

In problem P9, the objective function represents the signal leakage that falls out of the subspace spanned by the rows of equivalent channel $\mathbf{G}_{ir}$. By simplifying (4.7), we obtain $\mathbf{P}_i^* = \arg\max_{\mathbf{P}_i^\dagger \mathbf{P}_i = \mathbf{I}} \text{tr}\{\mathbf{P}_i^\dagger \mathbf{G}_{ir}^\dagger \mathbf{G}_{ir} \mathbf{P}_i\}$. Since $\mathbf{P}_i \in \mathbb{C}^{(M-M_p)\times d_i}$ has orthonormal columns, the columns of $\mathbf{P}_i^*$ should be the d_i dominant eigenvectors of matrix $\mathbf{G}_{ir}^\dagger \mathbf{G}_{ir}$ and each column's 2-norm is normalized to 1. This is possible only if $M - M_P \geq d_i$; thus, this condition is essential for the non-iterative algorithm whereas for the iterative algorithm (as discussed in Sec. 3.2) we require only $M - M_p > 0$. To satisfy the transmit power constraint, the optimal value of β_i is chosen such that $\mathbb{E}\text{tr}\{\mathbf{T}_i\mathbf{s}_i\mathbf{s}_i^\dagger\mathbf{T}_i^\dagger\} = P_{tot,i}$, leading to

$$\beta_i^* = \sqrt{P_{tot,i}/\text{tr}\left\{\sigma_s^2 \mathbf{H}_{ip}^{\perp}\mathbf{P}_i^*\mathbf{P}_i^{*\dagger}\mathbf{H}_{ip}^{\perp\dagger}\right\}}, \quad i = 1, 2, ..., K. \tag{3.68}$$

For the secondary receiver i, the decoding matrix design could follow the same method as for the precoder design, i.e. minimizing the matrix distance of decoding matrix $\mathbf{R}_i$ and corresponding channel matrix $\mathbf{H}_{ri}^\dagger$ (equivalently the distance between $\mathbf{R}_i^\dagger$ and $\mathbf{H}_{ri}$). However, interference from the primary user should also be taken into consideration, which means that we should also minimize the matrix distance between $\mathbf{H}_{pi}$ and $\mathbf{R}_i^{\perp\dagger}$ where $\mathbf{R}_i^{\perp}$ denotes a matrix whose columns are orthonormal basis for the complement of the subspace spanned by $\mathbf{R}_i$. As for the design of $\mathbf{P}_i$, for analytical tractability we assume *a priori* that $\mathbf{R}_i\mathbf{R}_i^\dagger = \mathbf{I}$. Then the decoder design of i-th secondary user is formulated as

$$\begin{aligned} \text{P10:} \quad \mathbf{R}_i^* = & \arg\min_{\mathbf{R}_i\mathbf{R}_i^\dagger=\mathbf{I}} \|\mathbf{H}_{ri}, \mathbf{R}_i^\dagger\|_{\text{dis}}^2 + \omega\|\mathbf{H}_{pi}, \mathbf{R}_i^{\perp\dagger}\|_{\text{dis}}^2 \\ = & \arg\min_{\mathbf{R}_i\mathbf{R}_i^\dagger=\mathbf{I}} \|\mathbf{H}_{ri} - \mathbf{R}_i^\dagger\mathbf{R}_i\mathbf{H}_{ri}\|_F^2 + \omega\|\mathbf{H}_{pi} - \mathbf{R}_i^{\perp\dagger}\mathbf{R}_i^{\perp}\mathbf{H}_{pi}\|_F^2 \ , \end{aligned} \tag{3.69}$$

where ω is a non-negative scalar weight. In problem P10, the first term in the objective function is the signal leakage that falls out of the signal subspace spanned by rows of $\mathbf{R}_i$ while the second term represents the interference leakage which falls out the interference subspace spanned by rows of $\mathbf{R}_i^{\perp}$. The weight ω is chosen empirically to strike a balance between minimizing the interference leakage and the minimizing of signal leakage. By simplifying (3.69) we have $\mathbf{R}_i^* = \arg\min_{\mathbf{R}_i\mathbf{R}_i^\dagger=\mathbf{I}} \mathrm{tr}\{\mathbf{R}_i(\omega\mathbf{H}_{pi}\mathbf{H}_{pi}^\dagger - \mathbf{H}_{ri}\mathbf{H}_{ri}^\dagger)\mathbf{R}_i^\dagger\}$. Since $\mathbf{R}_i \in \mathbb{C}^{\tilde{d}_i\times M}$, the columns of $\mathbf{R}_i^{*\dagger}$ should be the $\tilde{d}_i$ least dominant eigenvectors of $(\omega\mathbf{H}_{pi}\mathbf{H}_{pi}^\dagger - \mathbf{H}_{ri}\mathbf{H}_{ri}^\dagger)$ whose 2-norms are normalized to 1.

3.4.2 MSE based Relay Precoding Matrix Design

Given $\mathbf{R}_i$ and $\mathbf{T}_i$ for $i = 1, 2, ..., K$, the precoding matrix design at the relay node is based on an MSE criterion. Let the estimated received signal at secondary receivers be $\check{\mathbf{s}} = \gamma\hat{\mathbf{s}}$. We introduce the scaling parameter γ because the decoding matrices at secondary users are designed to consist of orthonormal rows. Therefore a scalar γ is necessary to adjust the scaling of the received signal. Then using (3.40), the MSE based relay precoding design problem is formulated as

$$\text{P11:} \quad \min_{\mathbf{P}_r,\gamma} \mathbb{E} \parallel \gamma\hat{\mathbf{s}} - \mathbf{E}\mathbf{s} \parallel^2 \quad \text{s.t.} \;\; \mathrm{tr}\{\mathbf{V}_2(\mathbf{P}_r)\} \le P_{tot,r}. \tag{3.70}$$

We now show that the optimal solution of P11 always satisfies the constraint with equality. Assume that $\tilde{\mathbf{P}}_r, \tilde{\gamma}$ are the optimal solution of P11 and $\mathrm{tr}\left\{\mathbf{V}_2(\tilde{\mathbf{P}}_r)\right\} < P_{tot,r}$. We introduce a positive scalar α, let $\bar{\mathbf{P}}_r = \alpha\tilde{\mathbf{P}}_r$, and choose α such that $\mathrm{tr}\left\{\mathbf{V}_2(\bar{\mathbf{P}}_r)\right\} = P_{tot,r}$ is satisfied. It is obvious that $\alpha > 1$. Let $\bar{\gamma} = \tilde{\gamma}/\alpha$; therefore we have $\bar{\gamma}\bar{\mathbf{P}}_r = \tilde{\gamma}\tilde{\mathbf{P}}_r$. Use $\bar{\mathbf{P}}_r, \bar{\gamma}$ and $\tilde{\mathbf{P}}_r, \tilde{\gamma}$ to evaluate the objective function values of P11 and denote them by $\bar{O}$ and $\tilde{O}$, respectively. Using (3.39) it can be shown that

$$\tilde{O} - \bar{O} = (|\tilde{\gamma}|^2 - |\bar{\gamma}|^2)\mathrm{t}r\left\{(\check{\mathbf{H}}_p\check{\mathbf{H}}_p^\dagger\sigma_p^2 + \mathbf{R}\mathbf{R}^\dagger\sigma_n^2)\right\} > 0.$$

This shows that $\tilde{\mathbf{P}}_r, \tilde{\gamma}$ could not be the optimal solution of P11. Therefore, we will use the equality constraint $\text{tr}\{\mathbf{V}_2(\mathbf{P}_r)\} = P_{tot,r}$ instead of the inequality constraint in P11. The Lagrangian of problem P11 is

$$L(\lambda, \mathbf{P}_r, \gamma) = \mathbb{E}\text{tr}\{(\gamma\hat{\mathbf{s}} - \mathbf{E}\mathbf{s})(\gamma\hat{\mathbf{s}} - \mathbf{E}\mathbf{s})^\dagger\} + \lambda\,\text{tr}\{\mathbf{V}_2(\mathbf{P}_r)\} - \lambda P_{tot,r}\,. \tag{3.71}$$

By setting $\frac{\partial L(\lambda,\mathbf{P}_r,\gamma)}{\partial \mathbf{P}_r} = 0$ and $\frac{\partial L(\lambda,\mathbf{P}_r,\gamma)}{\partial \gamma} = 0$, and using constraint (3.70), we have

$$\frac{\lambda^*}{\gamma^{*2}} = \frac{\sigma_n^2 \sum_{i=1}^K \tilde{d}_i + \sigma_p^2 \text{tr}\{\check{\mathbf{H}}_p \check{\mathbf{H}}_p^\dagger\}}{P_{tot,r}}, \tag{3.72}$$

$$\begin{aligned}\text{vec}\{\mathbf{P}_r^*\} = &\Big\{\mathbf{V}_3^T \otimes [\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}] - \textstyle\sum_{i=1}^K [\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\check{\mathbf{H}}_2^\dagger]^T \otimes [\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}]\sigma_s^2 \\ &+ \tfrac{\lambda^*}{\gamma^{*2}}\mathbf{V}_3^T \otimes [\mathbf{H}_{rp}^{\perp\dagger}\mathbf{H}_{rp}^{\perp}]\Big\}^{-1} \times \text{vec}\Big\{\mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\mathbf{E}\check{\mathbf{H}}_2^\dagger - \textstyle\sum_{i=1}^K \mathbf{H}_{rp}^{\perp\dagger}\check{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\mathbf{E}\mathbf{E}_r^{(i)}\check{\mathbf{H}}_2^\dagger\Big\}\sigma_s^2\gamma^{*-1}.\end{aligned} \tag{3.73}$$

Let $\mathbf{P}_r^* = \tilde{\mathbf{P}}_r^*\gamma^{*-1}$. Then $\gamma^* = \sqrt{\text{tr}\Big\{\mathbf{V}_2(\tilde{\mathbf{P}}_r^*)\Big\}/P_{tot,r}}$. Thus, first solve (3.72) and use it in (3.73) to obtain $\tilde{\mathbf{P}}_r^*$. Next using (3.40), the calculated value of $\tilde{\mathbf{P}}_r^*$ and the expression $\gamma^* = \sqrt{\text{tr}\Big\{\mathbf{V}_2(\tilde{\mathbf{P}}_r^*)\Big\}/P_{tot,r}}$, find γ^* to complete the solution.

3.4.3 Distributed Implementation

We now present a distributed implementation of the non-iterative algorithm. Firstly, assume that only local channel coefficient matrices are available at each secondary user. For example, the i-th secondary user only has the knowledge of $\mathbf{H}_{ir}$, $\mathbf{H}_{ri}$, $\mathbf{H}_{ip}$ and $\mathbf{H}_{pi}$, and the relay node only has the knowledge of $\mathbf{H}_{ir}$, $\mathbf{H}_{ri}$, $i = 1, 2, ..., K$, $\mathbf{H}_{rp}$ and $\mathbf{H}_{pr}$. Also assume that there exist control channels between relay and secondary users in order to facilitate exchanges of design precoding and decoding matrices. The proposed distributed algorithm is as follows:

1) Each secondary user i designs its own precoding and decoding matrices $\mathbf{P}_i$ and $\mathbf{R}_i$ by solving problems P9 and P10, using only the local channel information. Then it feeds back $\mathbf{R}_i^*$ and $\mathbf{T}_i^*$ to the relay station.

2) The relay station designs its precoding matrix by solving problem P11, using its local channel information and $\mathbf{R}_i^*$, $\mathbf{T}_i^*$, $i = 1, 2, ..., K$.

3) The relay station sends its precoding matrix to the secondary users which will be used by them to perform self-interference cancellation.

The benefits of this distributed implementation are: firstly, there is no need to assign a central node and collect global channel information. Secondly, computation is broken down into small parts and performed at all secondary nodes.

3.4.4 Effects of Allowing Interference to Primary Users

When a small amount of interference to the primary network is allowed instead of complete interference cancellation, the problem formulation and optimization solution will be similar to the robust algorithm in the next section while the only modification required will be setting all the variances in CSI to be zero. Due to space limitations, the problem formulation and solution is omitted.

3.5 Robust Algorithm Design with Imperfect CSI

In this section, we will take into account some practical design concerns and try to design a more practical and robust precoding algorithm. Due to the time-varying nature of wireless channels, the channel state information may already be out-dated when used in the precoder design. The feedback of channel information may also introduce some errors to the channel estimation results. Therefore, we propose a robust precoder design in this section by modeling the channel estimation errors as random errors with known variance, and considering its effect in the optimization problem. On the other hand, computational complexity is also another practical concern, which means that a non-iterative algorithm may be more cost-efficient for implementation. Also we want to design an algorithm that can be implemented in a distributed manner so that no central controller is needed and the computation can be distributed to all secondary users as well as the relay. Therefore, we will propose a robust precoder using the matrix-distance based non-iterative algorithm proposed in Sec. 3.4.

Assume that the CSI of all links is subject to stochastic errors. Channel matrices $\mathbf{H}_{ir}$, $\mathbf{H}_{ri}$, $\mathbf{H}_{ip}$, $\mathbf{H}_{pi}$, $\mathbf{H}_{pr}$ and $\mathbf{H}_{rp}$ are used to denote actual channel coefficient matrices among various source-destination pair nodes, as in the earlier sections. Then the estimated channel coefficients is the actual channel plus some random errors given by

$$\tilde{\mathbf{H}}_* = \mathbf{H}_* + \mathbf{\Lambda}_*, \tag{3.74}$$

where $* \in \{ir, ri, ip, pi, pr, rp\}$ and the matrix $\mathbf{\Lambda}_*$ models the estimation error in channel coefficient matrix $\mathbf{H}_*$. The various components of $\mathbf{\Lambda}_*$ are mutually independent and we have

$$\mathbb{E}\{\mathbf{\Lambda}_*\} = \mathbf{0}, \quad \text{and} \quad \mathbb{E}\{[\text{vec}\mathbf{\Lambda}_*]\,[\text{vec}\mathbf{\Lambda}_*]^\dagger\} = \sigma_*^2\mathbf{I}. \tag{3.75}$$

3.5.1 Robust SU Precoding/Decoding Matrices Design

Due to imperfect CSI of link from the i-th secondary transmitter to the primary receiver, interference to the primary receiver can not be canceled completely. Therefore the design of $\mathbf{T}_i$ should not only consider its distance to $\mathbf{H}_{ir}^\dagger$ but also its distance to $\mathbf{H}_{ip}^\dagger$. Let $\mathbf{T}_i = \beta_i\tilde{\mathbf{T}}_i$, where β_i is introduced to scale $\mathbf{T}_i$ so that the transmit power constraint can be satisfied and the interfering power to primary receivers can be bounded. We first state the MSE and all constraints in terms of unknown (true) $\mathbf{H}_*$ and then use (3.74) to express them in terms of

$\tilde{\mathbf{H}}_*$ and σ_*^2. The precoding design at secondary user i is formulated as:

$$\begin{aligned}\text{P12}: \quad \tilde{\mathbf{T}}_i^* &= \arg\min_{\tilde{\mathbf{T}}_i^\dagger \tilde{\mathbf{T}}_i = \mathbf{I}} \mathbb{E}\Big\{ \|\mathbf{H}_{ir}^\dagger, \tilde{\mathbf{T}}_i\|_{\text{dis}}^2 + \omega \|\mathbf{H}_{ip}^\dagger, \tilde{\mathbf{T}}_i^\perp\|_{\text{dis}}^2 \Big\} \\ &= \arg\min_{\tilde{\mathbf{T}}_i^\dagger \tilde{\mathbf{T}}_i = \mathbf{I}} \operatorname{tr}\Big\{ \tilde{\mathbf{T}}_i^\dagger \Big(\tilde{\mathbf{H}}_{ip}^\dagger \tilde{\mathbf{H}}_{ip} \omega + \sigma_{ip}^2 M_p \omega \mathbf{I} - \tilde{\mathbf{H}}_{ir}^\dagger \tilde{\mathbf{H}}_{ir} - \sigma_{ir}^2 N \mathbf{I} \Big) \tilde{\mathbf{T}}_i \Big\} \end{aligned} \tag{3.76}$$

where ω is a non-negative scalar weight which can be empirically chosen and as in the design of $\mathbf{P}_i$ in Sec. IV, $\tilde{\mathbf{T}}_i \in \mathbb{C}^{M \times d_i}$ is assumed to have orthonormal columns. Therefore, the columns of $\tilde{\mathbf{T}}_i$ should be the d_i least dominant eigenvectors of matrix $\Big(\tilde{\mathbf{H}}_{ip}^\dagger \tilde{\mathbf{H}}_{ip} \omega + \sigma_{ip}^2 M_p \omega \mathbf{I} - \tilde{\mathbf{H}}_{ir}^\dagger \tilde{\mathbf{H}}_{ir}^\dagger - \sigma_{ir}^2 N \mathbf{I} \Big)$ whose 2-norms are normalized to 1. After $\tilde{\mathbf{T}}_i^*$ is determined, we need to make sure that the total transmit power constraint at secondary user i is satisfied. Besides, we also need to make sure the interference caused by the i-th secondary user to primary receiver is strictly less than a threshold. In order to design a distributed algorithm we choose this interference threshold as I_{tot}/K for each secondary user so that they can determine their own transmitting parameters $\tilde{\mathbf{T}}_i^*$ and β_i^*. Therefore the following two constraints should be satisfied.

$$\operatorname{tr}\Big\{ \mathbf{T}_i^* \mathbf{T}_i^{*\dagger} \sigma_s^2 \Big\} \le P_{tot,i} \text{ and } \operatorname{tr}\Big\{ \Big(\tilde{\mathbf{H}}_{ip}^\dagger \tilde{\mathbf{H}}_{ip} + \sigma_{ip}^2 M_p \mathbf{I} \Big) \mathbf{T}_i^* \mathbf{T}_i^{*\dagger} \sigma_s^2 \Big\} \le I_{tot}/K. \tag{3.77}$$

This leads to

$$\beta_i^* = \min \left\{ \sqrt{P_{tot,i} / \operatorname{tr}\Big\{ \tilde{\mathbf{T}}_i^* \tilde{\mathbf{T}}_i^{*\dagger} \sigma_s^2 \Big\}}, \ \sqrt{(I_{tot}/K) / \operatorname{tr}\Big\{ \Big(\tilde{\mathbf{H}}_{ip}^\dagger \tilde{\mathbf{H}}_{ip} + \sigma_{ip}^2 M_p \mathbf{I} \Big) \tilde{\mathbf{T}}_i^* \tilde{\mathbf{T}}_i^{*\dagger} \sigma_s^2 \Big\}} \right\}. \tag{3.78}$$

The robust design of $\mathbf{R}_i$ is similar to that for problem P10: $\mathbf{R}_i$ is assumed to have orthonormal rows. However, uncertainty in $\mathbf{H}_{ri}$ and $\mathbf{H}_{pi}$ also needs to be taken into account. This leads to the optimization problem

$$\begin{aligned}\text{P13:} \quad \mathbf{R}_i^* &= \arg\min_{\mathbf{R}_i \mathbf{R}_i^\dagger = \mathbf{I}} \mathbb{E}\Big\{ \|\mathbf{H}_{ri}, \mathbf{R}_i^\dagger\|_{\text{dis}}^2 + \omega \|\mathbf{H}_{pi}, \mathbf{R}_i^{\perp\dagger}\|_{\text{dis}}^2 \Big\} \\ &= \arg\min_{\mathbf{R}_i \mathbf{R}_i^\dagger = \mathbf{I}} \Big\{ \mathbf{R}_i \Big(\omega \tilde{\mathbf{H}}_{pi} \tilde{\mathbf{H}}_{pi}^\dagger + \omega \sigma_{pi}^2 J_p \mathbf{I} - \tilde{\mathbf{H}}_{ri} \tilde{\mathbf{H}}_{ri}^\dagger - \sigma_{ri}^2 N \mathbf{I} \Big) \mathbf{R}_i^\dagger \Big\}. \end{aligned} \tag{3.79}$$

Since $\mathbf{R}_i \in \mathbb{C}^{\tilde{d}_i \times M}$, the columns of $\mathbf{R}_i^{*\dagger}$ should be the $\tilde{d}_i$ least dominant eigenvectors of $\left(\omega \tilde{\mathbf{H}}_{pi} \tilde{\mathbf{H}}_{pi}^{\dagger} + \omega \sigma_{pi}^2 J_p \mathbf{I} - \tilde{\mathbf{H}}_{ri} \tilde{\mathbf{H}}_{ri}^{\dagger} - \sigma_{ri}^2 N \mathbf{I}\right)$ whose 2-norms are normalized to 1.

3.5.2 Robust Relay Precoding Design

In this section we design a robust precoder at the relay station using $\mathbf{R}_i^*$ and $\mathbf{T}_i^*$ obtained in Sec. V-A, based on the MMSE criterion. Define the estimated channels as

$$
\begin{aligned}
\tilde{\mathbf{H}}_1 &= [(\mathbf{R}_1^* \tilde{\mathbf{H}}_{r1})^T, (\mathbf{R}_2^* \tilde{\mathbf{H}}_{r2})^T, ..., (\mathbf{R}_K^* \tilde{\mathbf{H}}_{rK})^T]^T, && (3.80) \\
\tilde{\mathbf{H}}_2 &= [\tilde{\mathbf{H}}_{1r} \mathbf{T}_1^*, \tilde{\mathbf{H}}_{2r} \mathbf{T}_2^*, ..., \tilde{\mathbf{H}}_{Kr} \mathbf{T}_K^*], && (3.81) \\
\tilde{\mathbf{H}}_p &= [(\mathbf{R}_1^* \tilde{\mathbf{H}}_{p1})^T, (\mathbf{R}_2^* \tilde{\mathbf{H}}_{p2})^T, ..., (\mathbf{R}_K^* \tilde{\mathbf{H}}_{pK})^T]^T. && (3.82)
\end{aligned}
$$

Since interference to the primary users can not be eliminated completely due to the channel uncertainty, we now impose a constraint on the interfering power at the primary receiver. Define

$$
\mathbf{V}_6 = \mathbf{T}_r \left(\check{\mathbf{H}}_2 \check{\mathbf{H}}_2^{\dagger} \sigma_s^2 + \mathbf{H}_{pr} \mathbf{H}_{pr}^{\dagger} \sigma_p^2 + \sigma_n^2 \mathbf{I} \right) \mathbf{T}_r^{\dagger}. \qquad (3.83)
$$

Then this optimization problem is formulated as follows:

$$
\begin{aligned}
\text{P14:} \quad & \min_{\mathbf{T}_r, \gamma} \mathbb{E} \| \gamma \hat{\mathbf{s}} - \mathbf{E} \mathbf{s} \|^2 && (3.84) \\
& \text{s.t. } \ \mathbb{E} \operatorname{tr} \{ \mathbf{V}_6 \} \leq P_{tot,r} \text{ and } \mathbb{E} \operatorname{tr} \left\{ \mathbf{H}_{rp} \mathbf{V}_6 \mathbf{H}_{rp}^{\dagger} \right\} \leq I_{tot}. && (3.85)
\end{aligned}
$$

The Lagrangian of problem P14 can be expressed as:

$$
\begin{aligned}
L(\lambda, \mu, \mathbf{T}_r, \gamma) &= \mathbb{E} \operatorname{tr} \{ (\gamma \hat{\mathbf{s}} - \mathbf{E} \mathbf{s})(\gamma \hat{\mathbf{s}} - \mathbf{E} \mathbf{s})^{\dagger} \} + \lambda \mathbb{E} \operatorname{tr} \{ \mathbf{V}_6 \} \\
&\quad + \mu \mathbb{E} \operatorname{tr} \left\{ \mathbf{H}_{rp} \mathbf{V}_6 \mathbf{H}_{rp}^{\dagger} \right\} - \lambda P_{tot,r} - \mu I_{tot}. && (3.86)
\end{aligned}
$$

By setting $\frac{\partial L(\lambda,\mu,\mathbf{T}_r,\gamma)}{\partial \mathbf{T}_r^\dagger} = 0$ and substituting true channels in terms of estimated channels and estimation errors, we obtain

$$\begin{gathered}\text{vec}\{\mathbf{T}_r^*\} = \text{vec}\{\tilde{\mathbf{T}}_r\gamma^{-1}\} = \Big\{[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger + e_2\mathbf{I}]^T \otimes [\tilde{\mathbf{H}}_1^\dagger\tilde{\mathbf{H}}_1 + e_1\mathbf{I}]\sigma_s^2 \\ - \textstyle\sum_{i=1}^K [\tilde{\mathbf{H}}_2\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger + \sigma_{ir}^2 \text{tr}\{\mathbf{T}_i^*\mathbf{T}_i^{*\dagger}\}\mathbf{I}]^T \otimes [\tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\tilde{\mathbf{H}}_1 + \sigma_{ri}^2\text{tr}\{\mathbf{R}_i^{*\dagger}\mathbf{R}_i^*\}\mathbf{I}]\sigma_s^2 \\ +[\tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger + \sigma_{pr}^2 J_p\mathbf{I}]^T \otimes [\tilde{\mathbf{H}}_1^\dagger\tilde{\mathbf{H}}_1 + e_1\mathbf{I}]\sigma_p^2 + \mathbf{I}\otimes[\tilde{\mathbf{H}}_1^\dagger\tilde{\mathbf{H}}_1 + e_1\mathbf{I}]\sigma_n^2 \\ +\tfrac{\lambda}{\gamma^2}[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger\sigma_s^2 + e_2\sigma_s^2\mathbf{I} + \tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger\sigma_p^2 + \sigma_{pr}^2 J_p\sigma_p^2\mathbf{I} + \sigma_n^2\mathbf{I}]^T \otimes \mathbf{I}\tfrac{\mu}{\gamma^2}[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger\sigma_s^2 + e_2\sigma_s^2\mathbf{I} + \tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger\sigma_p^2 \\ +\sigma_{pr}^2 J_p\sigma_p^2\mathbf{I} + \sigma_n^2\mathbf{I}] \otimes \Big(\tilde{\mathbf{H}}_{rp}^\dagger\tilde{\mathbf{H}}_{rp} + \sigma_{rp}^2 N\mathbf{I}\Big)\Big\}^{-1} \\ \text{vec}\{\tilde{\mathbf{H}}_1^\dagger\mathbf{E}\tilde{\mathbf{H}}_2^\dagger - \textstyle\sum_{i=1}^K \tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\mathbf{E}\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger\}\sigma_s^2\gamma^{*-1}, \qquad (3.87)\end{gathered}$$

where

$$e_1 = \sum_{i=1}^K \sigma_{ri}^2\text{tr}\{\mathbf{R}_i^{*\dagger}\mathbf{R}_i^*\}, \quad e_2 = \sum_{i=1}^K \sigma_{ir}^2\text{tr}\{\mathbf{T}_i^*\mathbf{T}_i^{*\dagger}\}. \qquad (3.88)$$

Now we need the optimal value of μ, λ and γ, or equivalently $\frac{\mu}{\gamma^2}, \frac{\lambda}{\gamma^2}$ and γ to obtain the optimal precoder $\mathbf{T}_r^*$. We have the following Proposition.

Proposition 3.1 : The optimal value of $\frac{\lambda}{\gamma^2}$ and $\frac{\mu}{\gamma^2}$ should satisfy

$$\frac{\lambda}{\gamma^2}P_{tot,r} + \frac{\mu}{\gamma^2}I_{tot} = \text{tr}\Big\{\tilde{\mathbf{H}}_p\tilde{\mathbf{H}}_p^\dagger\sigma_p^2\Big\} + \sigma_p^2 J_p\sum_{i=1}^K \tilde{d}_i\sigma_{ip}^2 + \sum_{i=1}^K \tilde{d}_i\sigma_n^2. \qquad (3.89)$$

Proof : See Appendix F. □

Variables λ, μ and γ^2 should be nonnegative. Therefore finding the optimal value of $\frac{\lambda}{\gamma^2}$ and $\frac{\mu}{\gamma^2}$ can be performed using a line search over one variable, say, $\frac{\lambda}{\gamma^2}$, since for a given value of $\frac{\lambda}{\gamma^2}$, the other one $\frac{\mu}{\gamma^2}$ is determined according to Proposition 1. A bisection method over $\frac{\lambda}{\gamma^2}$ in the region $\frac{\lambda}{\gamma^2} > 0$ can be performed to find the optimal value of $\frac{\lambda}{\gamma^2}$. Within this bisection method, for a fixed value of $\frac{\lambda}{\gamma^2}$, $\frac{\mu}{\gamma^2}$ is obtained using Proposition 1, $\mathbf{T}_r$ is given by (3.87) and

γ is chosen to satisfy the two constraints of P11:

$$\gamma^* = \arg\max_{\gamma} \Bigg\{ \mathrm{tr}\left\{ \tilde{\mathbf{T}}_r^* \left(\tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_2^\dagger \sigma_s^2 + e_2 \sigma_s^2 \mathbf{I} + \tilde{\mathbf{H}}_{pr} \tilde{\mathbf{H}}_{pr}^\dagger \sigma_p^2 + \sigma_{pr}^2 J_p \sigma_p^2 \mathbf{I} + \sigma_n^2 \mathbf{I} \right) \tilde{\mathbf{T}}_r^{*\dagger} \right\} / P_{tot,r},$$

$$\mathrm{tr}\Bigg\{ \left(\tilde{\mathbf{H}}_{rp}^\dagger \tilde{\mathbf{H}}_{rp} + \sigma_{rp}^2 N \mathbf{I} \right) \tilde{\mathbf{T}}_r^* \left(\tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_2^\dagger \sigma_s^2 + e_2 \sigma_s^2 \mathbf{I} + \tilde{\mathbf{H}}_{pr} \tilde{\mathbf{H}}_{pr}^\dagger \sigma_p^2 + \sigma_{pr}^2 J_p \sigma_p^2 \mathbf{I} + \sigma_n^2 \mathbf{I} \right) \tilde{\mathbf{T}}_r^{*\dagger} \Bigg\} / I_{tot} \Bigg\}. \quad (3.90)$$

(Note however that, from (3.87), when the values of $\frac{\lambda}{\gamma^2}$ and $\frac{\mu}{\gamma^2}$ are given, the value of γ will not change the value of the objective function or of the Lagrangian of P14). This robust algorithm can also be implemented in a distributed manner, as in Sec. 3.4.3.

3.6 Simulation Results

We consider a secondary network with $K = 4$ secondary users and one relay station. There is one primary transmitter-receiver pair within the same channel band. Assume that all channel links experience flat Rayleigh fading. The noise power σ_n^2 at every receiver, the PU transmit power σ_p^2 and the mean channel power gain of all links are all normalized to one. Furthermore we also normalize the SU information sequence power σ_s^2 to one, with desired transmit power, hence the desired receiver signal-to-noise ratio (SNR), achieved by scaling the precoder $\mathbf{P}_i$ at individual SU; for the case where there is no precoder, σ_s^2 is scaled to achieve the desired transmit power. Each secondary user has M antennas and relay is equipped with N antennas. The primary transmitter has J_p antennas and the primary receiver has M_p antennas. Each secondary user has d data streams to transmit and it wants to receive d data streams. All simulation results are based on 100 Monte Carlo runs. For illustrating the performance of various algorithms, in addition to the sum MSE, we also use sum rate (instantaneous capacity) of the secondary network as a performance metric. Let $D = \sum_{k=1}^{K} d_k$. Then the sum rate is defined as

$$\mathrm{Rate} = \sum_{i=1}^{D} \log_2 \left(1 + \frac{\left| \left\{ \check{\mathbf{H}}_1 \mathbf{T}_r \check{\mathbf{H}}_2 \right\}_{i,j(i)} \right|^2}{\mathrm{i}_{Ni}} \right) \quad (3.91)$$

where i_{Ni} is noise and interference power given by

$$\mathrm{i}_{Ni} = \Bigg|\Bigg\{ \left(\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2 - \sum_{i=1}^{K}\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\right)\left(\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2 - \sum_{i=1}^{K}\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\right)^{\dagger}\sigma_s^2$$

$$\check{\mathbf{H}}_1\mathbf{T}_r\mathbf{H}_{pr}\mathbf{H}_{pr}^{\dagger}\mathbf{T}_r^{\dagger}\check{\mathbf{H}}_1^{\dagger}\sigma_p^2+\check{\mathbf{H}}_1\mathbf{T}_r\mathbf{T}_r^{\dagger}\check{\mathbf{H}}_1^{\dagger}\sigma_n^2+\mathbf{R}\mathbf{R}^{\dagger}\sigma_n^2+\mathbf{H}_p\mathbf{H}_p^{\dagger}\sigma_p^2\Bigg\}_{i,i}\Bigg|^2 - \left|\left\{\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\right\}_{i,j(i)}\right|^2 \quad (3.92)$$

and $\{\mathbf{A}\}_{m,n}$ denotes the $mn-$th element of matrix $\mathbf{A}$ and $j(i) := \{l \,|\, \{\mathbf{E}\}_{i,l} = 1\}$.

Figs. 3.2 and 3.3 show the sum rate (3.91) and the sum MSE, respectively, of our proposed iterative and non-iterative algorithms together with that of three other cases, as a function of secondary user SNR. [The parameter ω in the non-iterative algorithm was chosen to be 10.] The results of the non-iterative algorithm were used for initializing the proposed iterative algorithm. It is seen from Figs. 3.2 and 3.3 that the sum rates and the sum MSEs, respectively, of our proposed algorithms are significantly higher and lower, respectively, than that of the schemes with no precoding. In the non-iterative algorithm, performing self-information cancellation provides noticeable improvement in the sum rate and the sum MSE. However, this comes at a cost that the precoding matrix of relay node needs to be forwarded to all secondary users. In the no source and relay precoding scheme, only decoding matrices $\mathbf{R}_i, i = 1, 2, ..., K$ are designed according to problem P2 while in the no source precoding algorithm, $\mathbf{R}_i, i = 1, 2, ..., K$ and $\mathbf{T}_r$ are designed according to problems P2 and P5 and no iteration is performed. It is also seen from Figs. 3.2 and 3.3 that allowing some interference to the PU networks improves the performance of the proposed iterative schemes but the improvement is not "substantial"; in the remaining simulation examples we omit this case. Fig. 3.4 shows the effect of varying values of the scalar weight ω used in the proposed non-iterative algorithm (see (3.69)). It is seen that over a wide range of values (ω ranging from 10 through 100) the sum-MSE performance is essentially unchanged.

Fig. 3.5 shows the sum MSE vs the SNR per secondary user (relay's transmit power is K times the secondary user's power) under several different secondary network setups. It is observed that increasing the number M of SU antennas for the same number of datastreams

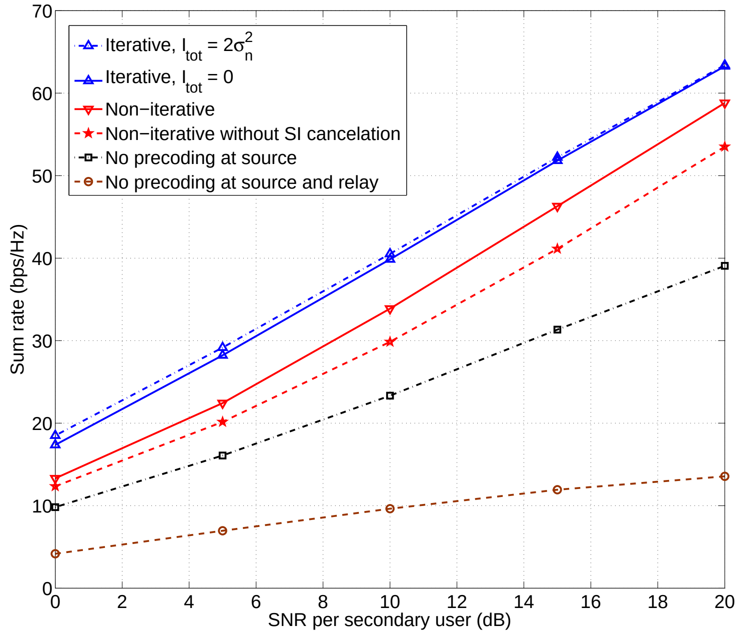

Figure 3.2: Sum rate (3.91) vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 4$, $N = 10$, $d = 2$, $M_p = J_p = 1$, $P_{tot,r} = KP_{tot,i}$, $\omega = 10$. The label "Iterative, $I_{tot} = 2\sigma_n^2$" refers to the approach of Sec. 3.3.5 and the label "Iterative, $I_{tot} = 0$" refers to the approach summarized in Sec. 3.3.4.

d significantly improves the sum MSE. It was noted in Sec. 3.2 that one needs $M - M_p > 0$ in order to null interference to the primary network, and $M - M_p \geq d$ was desirable to mitigate inter-datastream interference. In Fig. 3.5 with $M_p = 2$, the case $M = 3$, $d = 2$ represents $M - M_p = 1 > 0$ but $< d = 2$ whereas all other cases satisfy $M - M_p \geq d$. It is seen that the performance for the case $M = 3$, $d = 2$ is much worse than all other cases shown in Fig. 3.5. In Fig. 3.6 it is observed that increasing the number of relay antennas can decrease the sum MSE for a fixed number of SU antennas and datastreams per secondary user, and the gap between the iterative and non-iterative algorithms tends to be smaller as the number of antennas at the relay station increases. Fig. 3.7 shows the convergence of sum MSE in our iterative algorithms for a single run.

Figs. 3.8 and 3.9 compare the performance of our proposed robust algorithm with the "non-robust" non-iterative algorithm of Sec. 3.4. In each Monte Carlo run various channel gains generated according to Rayleigh fading were further perturbed with zero-mean complex

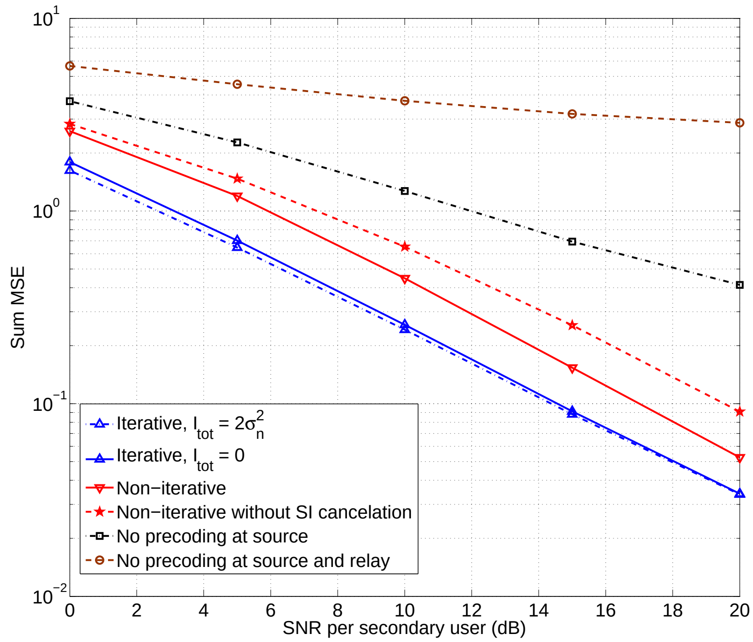

Figure 3.3: Sum MSE vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 4$, $N = 10$, $d = 2$, $M_p = J_p = 1$, $P_{tot,r} = KP_{tot,i}$, $\omega = 10$. The label "Iterative, $I_{tot} = 2\sigma_n^2$" refers to the approach of Sec. 3.3.5 and the label "Iterative, $I_{tot} = 0$" refers to the approach summarized in Sec. 3.3.4.

Gaussian noise (error). The curves labeled "non-iterative $\cdots$ perfect CSI" are based on perfect knowledge of CSI (error-free unperturbed CSI) whereas in the other cases the noisy channel CSI was used in the algorithm. We assume all the channel links have the same level of estimation errors, which means $\sigma_*^2 = \sigma_e^2$, $\forall * \in \{ir, ri, ip, pi, pr, rp\}$ (see (3.74) and (3.75)). In the interference constraints (see (3.77) and (3.85)), we set $I_{tot} = 2\sigma_n^2 = 2$. It is seen in Fig. 3.8 that when the channel estimation errors are relatively small ($\sigma_e^2 = 0.01$), the robust design achieves a performance improvement (lower sum MSE) over the non-robust non-iterative algorithm in the low transmit power (equivalently low SNR) regime. At higher SNRs, the channel errors are dominant over the noise power leading to "plateauing" of the sum MSE curve with increasing SNR. When the channel estimation errors are high ($\sigma_e^2 = 0.1$), the robust design achieves a performance improvement (lower sum MSE) over the non-robust non-iterative algorithm over the entire transmit power range. Now the channel errors seem to dominate the noise power at all SNRs, showing little improvement with increasing SNR.

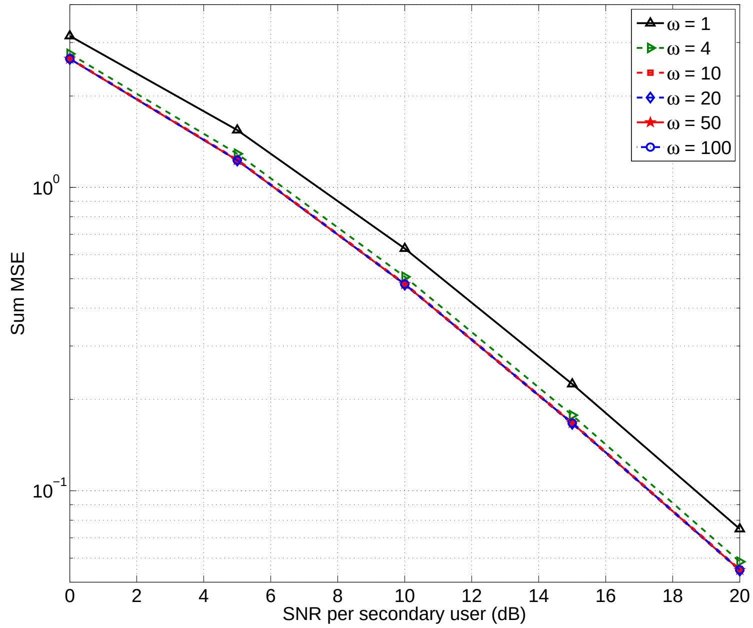

Figure 3.4: Sum MSE using non-iterative approach vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 4$, $N = 10$, $d = 2$, $M_p = J_p = 1$, $P_{tot,r} = KP_{tot,i}$, variable ω

From Fig. 3.9 we also see that although the sum MSE of the robust algorithm in Fig. 3.8 in the high SNR regime and lower channel estimation errors case of $\sigma_e^2 = 0.01$ is slightly higher than that for the non-iterative algorithm, the interference power to the primary network is well constrained (see Fig. 3.9), while the interference power of the non-robust algorithm has increased drastically and significantly exceeds the bound $I_{tot} = 2$. By design, when perfect CSI is available, the non-robust algorithm can null out the interference to the primary network; this is seen in Fig. 3.9.

3.7 Conclusions

We investigated joint design of precoders and decoders in a multiuser multi-way relay system in cognitive radio networks, which operates concurrently with a primary network within the same frequency band. The design objective was to minimize the sum MSE of all secondary users under a transmit power constraint for each transmitting node while keeping the interference to primary receiver to be zero, assuming complete knowledge of the

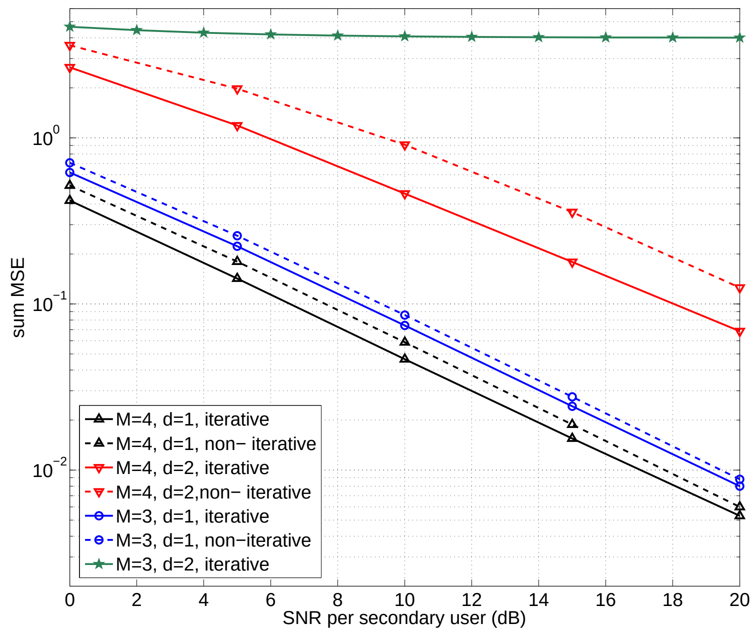

Figure 3.5: Sum MSE vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 3$ or 4, $N = 10$, $d = 1$ or 2, $M_p = J_p = 2$, $P_{tot,r} = KP_{tot,i}$, $\omega = 10$

channel state information (CSI) was available. We considered iterative optimization as well as a non-iterative approach, which can be implemented in a distributed manner. A channel error-aware robust matrix distance based algorithm was also proposed to address the case of imperfect CSI knowledge. The efficacy of our proposed algorithms was illustrated via computer simulations. Significant gains in the sum rate of the secondary network can be obtained by proper design of precoders and decoders in the multiuser multi-way relay system compared to the case of no precoders, while mitigating interference to the primary system.

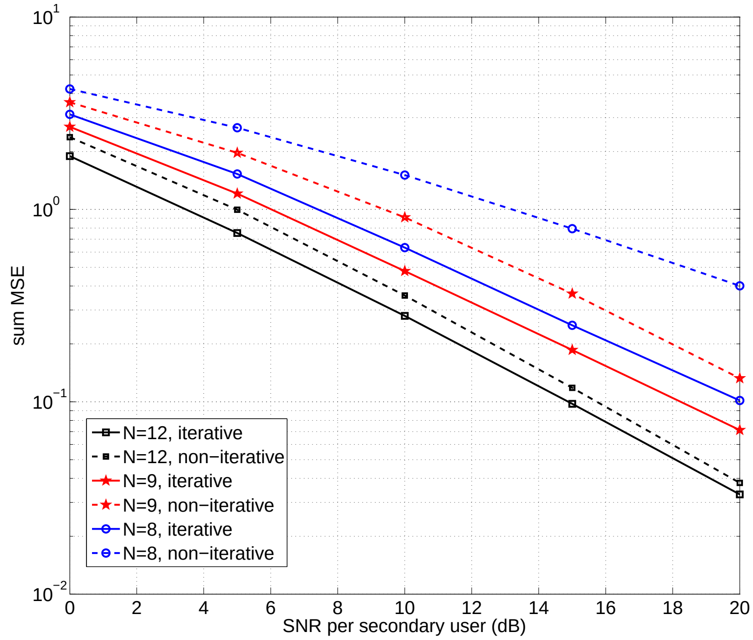

Figure 3.6: Sum MSE vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 3$, $N = 8$, 9 or 12, $d = 2$, $M_p = J_p = 1$, $P_{tot,r} = KP_{tot,i}$, $\omega = 10$.

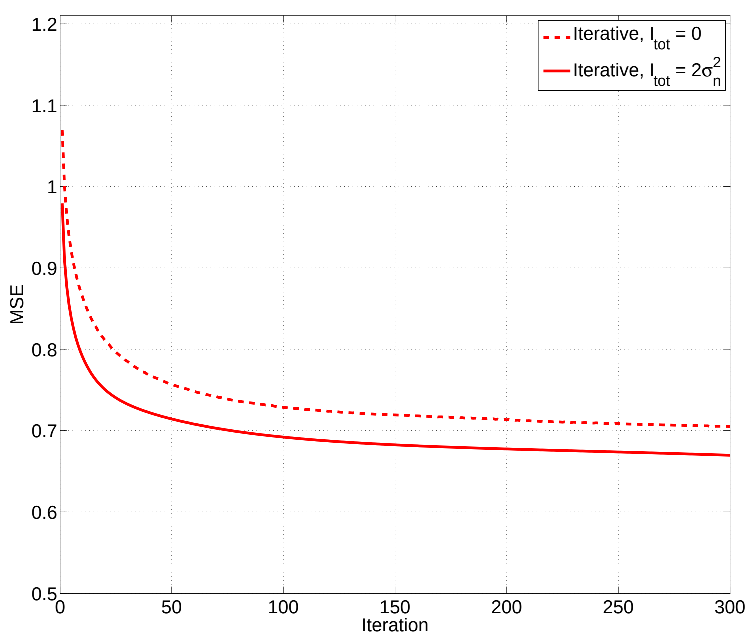

Figure 3.7: Sum MSE vs iteration (for a single run): $10\log_{10}(P_{tot,i}/\sigma_n^2) = 5$dB, $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 4$, $N = 10$, $d = 2$, $M_p = J_p = 1$, $P_{tot,r} = KP_{tot,i}$. Recall that noise variances have been normalized to one.

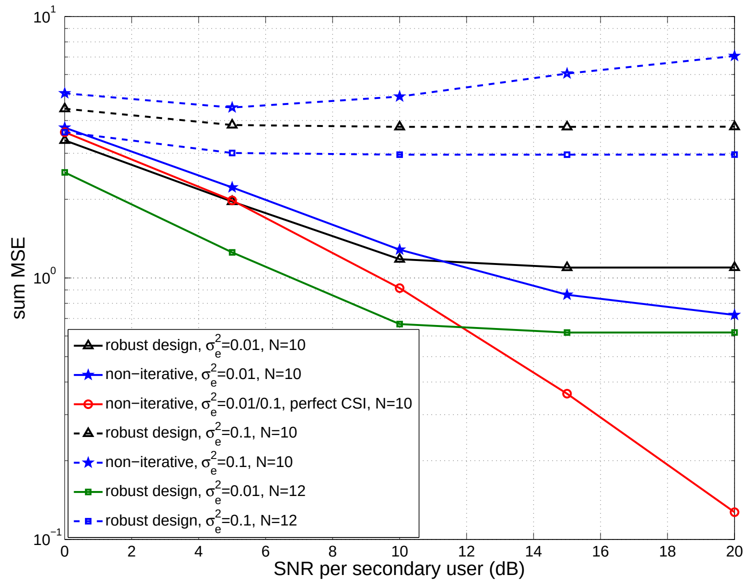

Figure 3.8: Sum MSE vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 4$, $N = 10$, $d = 2$, $M_p = J_p = 2$, $P_{tot,r} = KP_{tot,i}$, $\omega = 10$, receiver noise variance $\sigma_n^2 = 1$, design interference bound $I_{tot} = 2\sigma_n^2$.

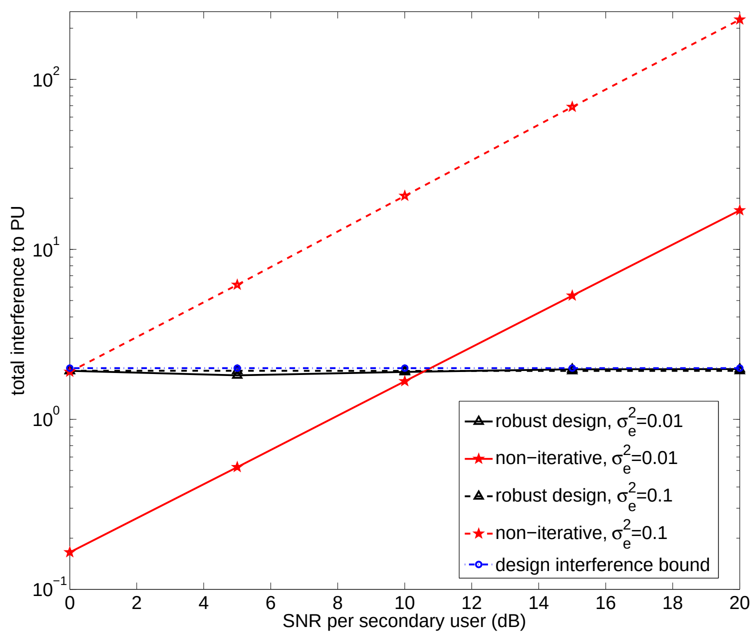

Figure 3.9: Average interference power to primary receivers vs SNR per secondary user $10\log_{10}(P_{tot,i}/\sigma_n^2)$: $K = 4$, $P_{tot,i}$ is the same for $i = 1, \cdots, K$, $M = 4$, $N = 10$, $d = 2$, $M_p = J_p = 2$, $P_{tot,r} = KP_{tot,i}$, $\omega = 10$, receiver noise variance $\sigma_n^2 = 1$, design interference bound $I_{tot} = 2\sigma_n^2$.

Chapter 4

Precoding Design in Multi-Pair Two-Way Cognitive Relay Networks

Two-way relay network is an effective way to support the data exchange of two users within the same user pair. Traditional two-way relay system contains two users who want to share information and one relay node. Given the fact that each user has perfect knowledge of its own transmitted signal, the information from one user can be coded with its desired signal, called network coding, and the recovery of its desired signal is achieved by subtracting its own signal, which is usually called self-interference cancellation. Suppose both the end nodes and the relay operate in a half-duplex mode, the information exchange of two users can be completed by a two-time slot transmission scheme. In the first time slot, also called as multiple access (MAC) phase, two end users transmit to the relay simultaneously, while in the second time slot, also called as broadcasting (BC) phase, the relay broadcasts its received and coded signals to the two end users. Similar to one-way relaying, the relay node can either perform decode-and-forward or amplify-and-forward [15, 52]. The advantage of AF is its simplicity in relay design and implementation since the relay only amplifies its received signal instead of decoding it as in DF scheme.

With the deployment of multiple antennas, one relay node may support the data exchange of multiple users pairs since multiple antennas provide extra spacial dimensions which can accommodate multiple network coded datastreams. The MIMO techniques can significantly improve the system throughput. However in multi-pair two-way relay networks, interference between user pairs can dramatically degrade the system's performance. Transmit and relay precoding is an efficient way to explore the multiplexing gain provided by multiple antennas and diminish the harmful effects of inter-pair interference.

In this chapter, we propose a joint source and relay precoding method for a multi-pair two-way relay cognitive network. In this network, multiple secondary user pairs exchange information with the help of a relay node. They coexist with a primary source-destination pair within the same frequency band and both the secondary nodes and the primary node transmit concurrently. With the perfect CSI available, a matrix-distance based precoding/decoding scheme is proposed which align both the signal and inter-pair interference so that an improved performance under certain condition is achieved, compared with the MSE-based method as in Chapter 3. The iteration between precoding/decoding at the users and at the relay can be performed to further improve the system performance at a cost of higher complexity.

This chapter is organized as follows. Background of precoding in cognitive radio and two-way relay networks and related works are presented in Sec. 4.1. System model is introduced in Sec. 4.2. A matrix distance based precoding method is proposed in Sec. 4.3. In Sec. 4.4 simulation results show the performance of the proposed precoding design. Conclusions are drawn in Sec. 4.5.

4.1 Introduction

Cognitive radio is a promising technology to increase spectrum efficiency. The unlicensed users can access licensed spectrum either opportunistically or concurrently. When secondary users concurrently access the spectrum with licensed primary users, how to mitigate interference to primary users becomes a crucial design issue of secondary network. The employment of multiple antennas at secondary transmitters makes it possible to cancel out the interference to primary network by aligning the transmit signal direction given complete channel state information. Therefore in multi-user cognitive radio system, both the inter-secondary user interference and the secondary-primary interference should be mitigated or at lease minimized by proper transceiver design. In [38] a rate balanced transceiver was proposed in multiuser cognitive radio network. In [39], a linear precoding method was proposed for CR multiuser downlink MIMO system based on minimum mean square error (MMSE) criterion. For systems with imperfect CSI, [53] studied a nonlinear transceiver design problem in a multi-tier MIMO cognitive radio network while [54] proposed a linear transceiver design in a downlink cognitive MIMO system.

Relay based two-way transmission schemes provide significant performance improvement such as higher power and spectrum efficiency and extended coverage (compared to the case when there is no relay). By performing self-interference cancellation, each user can obtain its desired information. However, when there are multiple user pairs, the inter-pair interference can dramatically reduce the system performance. Therefore how to eliminate interference or alleviate the negative effects of this interference is crucial in system design. Two common methods to perform interference cancellation are based on MMSE and zero-forcing (ZF) criteria. By adopting the MMSE criterion, in [48], the precoders at relay and sources are designed to minimize the sum MSE of received signals at all users. For zero-forcing system design, [55] proposed a distributed beamforming for a multi-pair two-way relay system where each node has a single antenna. The inter-pair interference is canceled out by designing relay weights of multiple signal antenna relays. In [44] the relay beamforming matrix was designed based on ZF and MMSE criteria for a multi-user MIMO relay system. By employing MIMO techniques, many interference alignment methods have been proposed to null out the inter-pair interference therefore enabling multi-pair transmission [1, 34]. In [35], a joint signal and interference alignment algorithm is proposed. In [57], a coordinated eigen beamforming method is proposed to optimize the throughput given that the inter-pair interference has already been eliminated.

In this chapter, a multi-pair two-way relay system with a single relay in a cognitive radio network is considered. Secondary end users and the relay are equipped with multiple antennas. Unlike Chapter 3 in which the relay precoder is designed using a MMSE criterion for both iterative and non-iterative algorithms, in this chapter we propose an interference alignment (IA)-like relay precoder to align both the interference and desired signal directions while ensuring no interference is caused to the primary network. By allowing some residual inter-pair interference, the system performance can be further improved compared with ZF-based design in which the inter-pair interference is completely removed in the low SNR regime. Also this is a more general algorithm which can be employed whether the number of antennas at the relay station is large enough to permit zero-forcing or not.

Notation: We use bold lower-case and upper-case letters to denote vectors and matrices respectively. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. $(.)^T$ and $(.)^\dagger$ denote the transpose and conjugate transpose of a matrix respectively. $\|.\|$ and $\|.\|_F$ are 2-norm and Frobenius norm respectively. $\mathbb{C}^{n\times m}$ denotes the space of $n\times m$ complex matrices. $\{\mathbf{A}\}_{m,n}$ denotes the $mn-$th element of matrix $\mathbf{A}$.

4.2 System Model and Problem Formulation

Consider a cognitive radio network consisting of $2K$ secondary users and one relay station. This cognitive radio network coexists with a primary source-destination pair within a single band. Suppose secondary user k and $k+1$, ($k < 2K$ and k is odd) form a user pair who want to send information to each other through the help of the relay node. Each secondary user is equipped with M antennas while the relay station has N antennas and the primary transmitter and receiver have J antennas. A two-time-slot half-duplex transmission scheme is used to support the two-way transmission of K secondary user pairs. In the first time slot all of the $2K$ users transmit to the relay while in the second time slot relay sends a linear combination of its received signal from the first time slot, and $2K$ secondary users listen and decode their desired signals, showed in Fig. 4.1. Let $\mathbf{H}_{ir} \in \mathbb{C}^{N\times M}$ and $\mathbf{H}_{ri} \in \mathbb{C}^{M\times N}$, $i = 1, 2, ..., K$ denote the channel coefficient matrices from user i to relay station and from relay to user i. Also $\mathbf{H}_{ip} \in \mathbb{C}^{J\times M}$, $\mathbf{H}_{pr} \in \mathbb{C}^{N\times J}$, $\mathbf{H}_{rp} \in \mathbb{C}^{J\times N}$ and $\mathbf{H}_{pi} \in \mathbb{C}^{M\times J}$, $i = 1, 2, ..., K$ denote the channel coefficient matrices from secondary user i to primary receiver, from primary transmitter to relay, from relay to primary receiver and from primary transmitter to secondary user i, respectively. It is assumed that these channel matrices remain constant during the two-time slot transmission and are available (known) to the secondary network.

Assume each secondary user has d independent data streams to transmit. Let $\mathbf{s}_i$ be the $d \times 1$ information vector from i-th secondary user with zero mean and $\mathbb{E}\{\mathbf{s}_i\mathbf{s}_i^\dagger\} = \sigma_s^2\mathbf{I}_d$. $\mathbf{T}_i \in \mathbb{C}^{M\times d}$ is the precoding matrix of user i. Then in the first time slot, transmitted signal from user i is $\mathbf{x}_i = \mathbf{T}_i\mathbf{s}_i$. During the first time slot, the received signal $\mathbf{y}_r$ at secondary relay node and interference $\mathbf{i}_1$ to primary receiver are given by

$$\begin{aligned}\mathbf{i}_1 &= \sum_{i=1}^{2K}\mathbf{H}_{ip}\mathbf{T}_i\mathbf{s}_i \\ \mathbf{y}_r &= \sum_{i=1}^{2K}\mathbf{H}_{ir}\mathbf{T}_i\mathbf{s}_i + \mathbf{H}_{pr}\mathbf{x}_p + \mathbf{n}_r,\end{aligned}$$

where $\mathbf{x}_p$ is the $J\times 1$ signal vector transmitted by primary user with zero mean and $\mathbb{E}\{\mathbf{x}_p\mathbf{x}_p^\dagger\} = \sigma_p^2\mathbf{I}_J$ and $\mathbf{n}_r$ is the $N\times 1$ complex white Gaussian noise vector at relay with $\mathbf{n}_r \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I}_N)$. In the second time slot, relay transmits a linear combination of its received signal, using a precoding matrix $\mathbf{T}_r$. The transmitted signal is denoted as $\mathbf{x}_r = \mathbf{T}_r\mathbf{y}_r$

The interference $\mathbf{i}_2$ to primary receiver and received signal $\mathbf{y}_i$ at $2K$ secondary users during the second time slot are expressed as follows:

$$\mathbf{i}_2 = \mathbf{H}_{rp}\mathbf{T}_r\Big(\sum_{i=1}^{2K}\mathbf{H}_{ir}\mathbf{T}_i\mathbf{s}_i + \mathbf{H}_{pr}\mathbf{x}_p + \mathbf{n}_r\Big) \tag{4.1}$$

$$\begin{aligned}\mathbf{y}_i &= \mathbf{H}_{ri}\Big(\sum_{j=1}^{2K}\mathbf{T}_r\mathbf{H}_{jr}\mathbf{T}_j\mathbf{s}_j + \mathbf{T}_r\mathbf{H}_{pr}\mathbf{x}_p + \mathbf{T}_r\mathbf{n}_r\Big) \\ &\quad + \mathbf{H}_{pi}\tilde{\mathbf{x}}_p + \mathbf{n}_i,\ \ i = 1, 2, ..., 2K,\end{aligned} \tag{4.2}$$

$\tilde{\mathbf{x}}_p$ is the $J \times 1$ signal vector transmitted by primary user in second time slot, with zero mean and $\mathbb{E}\{\tilde{\mathbf{x}}_p\tilde{\mathbf{x}}_p^\dagger\} = \sigma_p^2\mathbf{I}_J$ and $\mathbf{n}_i$ is the $M \times 1$complex white Gaussian noise at secondary user i which follows $\mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I}_M)$.

Suppose each user i uses a decoding matrix $\mathbf{R}_i$ to estimate its received signal $\hat{\mathbf{s}}_i$. Then we have

$$\hat{\mathbf{s}}_i = \mathbf{R}_i \mathbf{y}_i = \mathbf{R}_i \Big\{ \mathbf{H}_{ri} \Big(\sum_{j=1}^{2K} \mathbf{T}_r \mathbf{H}_{jr} \mathbf{T}_j \mathbf{s}_j + \mathbf{T}_r \mathbf{H}_{pr} \mathbf{x}_p + \mathbf{T}_r \mathbf{n}_r \Big) + \mathbf{H}_{pi} \tilde{\mathbf{x}}_p + \mathbf{n}_i \Big\}. \tag{4.3}$$

As in other interference alignment schemes, we assume noise is negligible compared to interference from other users. So we rewrite (4.3) as

$$\hat{\mathbf{s}}_i = \mathbf{R}_i \Big\{ \mathbf{H}_{ri} \Big(\sum_{j=1}^{2K} \mathbf{T}_r \mathbf{H}_{jr} \mathbf{T}_j \mathbf{s}_j + \mathbf{T}_r \mathbf{H}_{pr} \mathbf{x}_p \Big) + \mathbf{H}_{pi} \tilde{\mathbf{x}}_p \Big\}.$$

One crucial requirement for cognitive radio design is to cause no harm to primary network. This means $\mathbf{i}_1 = \mathbf{i}_2 = \mathbf{0}$ should be satisfied for any possible transmit signal vectors $\mathbf{s}_i, i = 1, 2, ..., K$. Therefore we enforce the following constraints on precoders $\mathbf{T}_i$ and $\mathbf{T}_r$:

$$\mathbf{H}_{ip} \mathbf{T}_i = \mathbf{0}, \quad i = 1, 2, ..., 2K, \tag{4.4}$$

$$\mathbf{H}_{rp} \mathbf{T}_r = \mathbf{0}. \tag{4.5}$$

Let $\mathbf{H}_{ip}^{\perp}$ and $\mathbf{H}_{rp}^{\perp}$ be the null space matrices of $\mathbf{H}_{ip}$ and $\mathbf{H}_{rp}$ respectively. Then letting $\mathbf{T}_i = \mathbf{H}_{ip}^{\perp} \mathbf{P}_i$ and $\mathbf{T}_r = \mathbf{H}_{rp}^{\perp} \mathbf{P}_r$ satisfies (4.4) and (4.5). The precoding parameters that are to be designed now become $\mathbf{P}_i$, $i = 1, 2, ..., K$ and $\mathbf{P}_r$.

Lemma 1 : In order to cause no interference to primary network, the number of antennas at secondary users and relay should satisfy $M > J$, $N > J$. The maximum multiplexing order that each SU can employ should not exceed $M - J$, i.e. $d \leq M - J$.

Proof: This follows directly from the fact that channel matrices $\mathbf{H}_{ip}, \mathbf{H}_{rp}$ have full rank almost surely. For the channel between secondary user i and primary receiver, $\text{rank}\{\mathbf{H}_{ip}\} = \min\{M, J\}$. If $M \leq J$, then the dimension of $\mathbf{H}_{ip}$'s null space will be empty, which makes it impossible to cancel the interference to primary network completely. So it is required that $M > J$, which leads to $\text{rank}\{\mathbf{H}_{ip}\} = J$, $\text{rank}\{\mathbf{H}_{ip}^{\perp}\} = M - J$ and $\mathbf{H}_{ip}^{\perp}$ is a $M \times (M - J)$

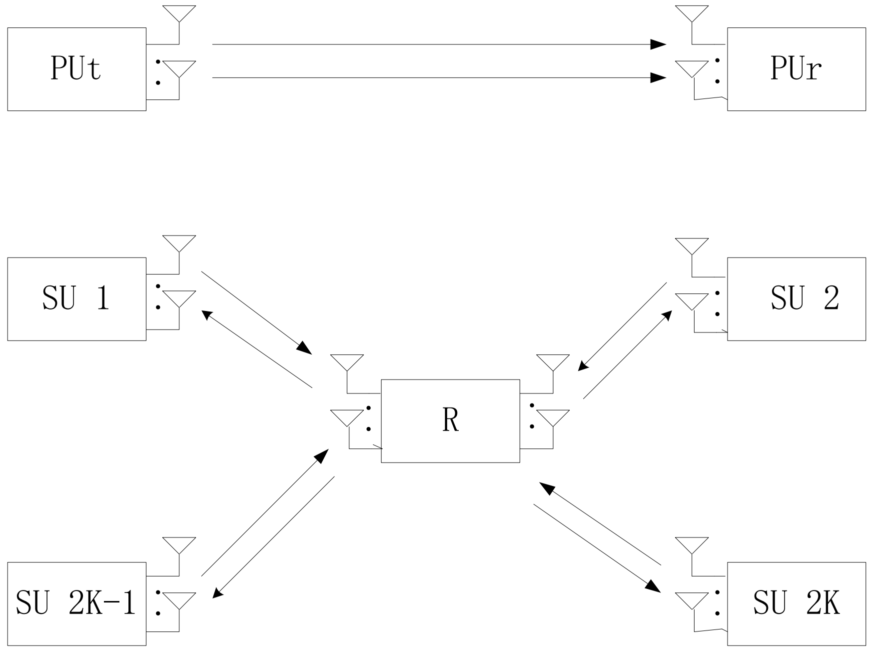

Figure 4.1: System model with 2K SUs, one secondary relay and one primary source-destination pair

matrix. Then rank$\{\mathbf{T}_i\} \leq M - J$ which means the multiplexing order user i can employ should not exceed $M - J$. For a similar reason, the requirement on relay antenna number is $N > J$. $\mathbf{H}_{rp}^{\perp}$ is an $N \times (N - J)$ matrix and its rank is $N - J$. $\square$

In this chapter, we assume secondary users employ the maximum multiplexing order, i.e. $d = M - J$.

For notational simplicity, define the equivalent channel matrices $\tilde{\mathbf{H}}_{ir} = \mathbf{H}_{ir}\mathbf{H}_{ip}^{\perp}$, $\tilde{\mathbf{H}}_{ri} = \mathbf{H}_{ri}\mathbf{H}_{rp}^{\perp}$, for $i = 1, 2, ..., K$. Since it is guaranteed that secondary users' transmission won't cause any interference to primary network, we can concentrate on designing precoding and decoding matrices to alleviate the negative effects of inter-user interference and improve secondary network's performance. From the above discussion, the precoding matrices design at the relay and SU i now becomes the task of designing $\mathbf{P}_i \in \mathbb{C}^{(M-J)\times d}$ and $\mathbf{P}_r \in \mathbb{C}^{(N-J)\times N}$.

4.3 Matrix Distance based Precoding Design

In this section, we present a matrix distance based precoding design at secondary users as well as the relay station. Unlike most IA schemes aiming to remove inter-user interference completely, we propose a matrices distance based algorithm in which the direction of both the desired signal and inter-user interference are jointly considered. A scalar weight is introduced to strike a balance between the "desired signal" and "interference". It will be shown that by allowing a small residual inter-user interference, it is possible to increase the strength of desired signal; therefore better performance can be achieved.

The distance between two matrices A and B, each with orthonormal columns, is defined as [35] (the motivation of this definition was discussed in detail in Sec. 3.4)

$$\|\mathbf{A}, \mathbf{B}\|_{\text{dis}} = \|\mathbf{A} - \mathbf{B}\mathbf{B}^{\dagger}\mathbf{A}\|_F. \tag{4.6}$$

The design goal is to align the signal direction as close to its desired receiver as possible while trying to keep it away from the direction to other receivers so that inter-user interference is mitigated. Firstly we initialize the precoding and decoding matrices $\mathbf{T}_i$ and $\mathbf{R}_i$ at each user i using a matrix distance criterion. Then the relay precoding matrix is expressed as a cascade of 3 parts including a receive combining matrix, power amplifying matrix and transmit beamforming matrix. The two relay beamforming matrices are then designed by jointly considering the inter-user interference and desired signal strength. Finally another SU decoding matrix $\tilde{\mathbf{R}}_i$ at user i is cascaded with $\mathbf{R}_i$ in order to decode each datastream for user i, i.e. removing the inter-datastream interference. The transmit precoder design at SUs and relay guarantees that no interference is caused to the primary network.

4.3.1 Secondary User Precoding/Decoding Matrices Initialization

Before designing precoding matrix $\mathbf{P}_r$ at relay node, precoding matrix $\mathbf{T}_i$ and decoding matrix $\mathbf{R}_i$ at i-th SU's are initialized. $\mathbf{T}_i$ is designed to minimize its distance to the channel matrix $\mathbf{H}_{ir}$ so "more information" can be transmitted to relay station. Besides, to make sure that primary user's transmission is not interfered with by secondary network, $\mathbf{T}_i = \mathbf{H}_{ip}^{\perp}\mathbf{P}_i$ is imposed. Without loss of generality, we assume $\mathbf{P}_i$ has orthonormal columns. A scaler β_i is introduced to satisfy the transmit power constraint of user i; therefore $\mathbf{T}_i$ is of the form $\beta_i\mathbf{H}_{ip}^{\perp}\mathbf{P}_i$. The equivalent channel matrix for $\mathbf{P}_i$ is $\tilde{\mathbf{H}}_{ir}$. Then the precoding design at user i is formulated as:

$$\begin{aligned}\mathbf{P}_i \quad &= \arg\min_{\mathbf{P}_i^{\dagger}\mathbf{P}_i=\mathbf{I}_d} \|\tilde{\mathbf{H}}_{ir}^{\dagger}, \mathbf{P}_i\|_{\text{dis}}^2 \\ &= \arg\min_{\mathbf{P}_i^{\dagger}\mathbf{P}_i=\mathbf{I}_d} \|\tilde{\mathbf{H}}_{ir}^{\dagger} - \mathbf{P}_i\mathbf{P}_i^{\dagger}\tilde{\mathbf{H}}_{ir}^{\dagger}\|_F^2.\end{aligned} \tag{4.7}$$

By simplifying (4.7), we obtain

$$\mathbf{P}_i \;=\; \arg\max_{\mathbf{P}_i^{\dagger}\mathbf{P}_i=\mathbf{I}_d} \text{tr}\{\mathbf{P}_i^{\dagger}\tilde{\mathbf{H}}_{ir}^{\dagger}\tilde{\mathbf{H}}_{ir}\mathbf{P}_i\}. \tag{4.8}$$

Since $\mathbf{P}_i \in \mathbb{C}^{(M-J)\times d}$ has orthonormal columns, the columns of $\mathbf{P}_i$ should be the d dominant eigenvectors of matrix $\tilde{\mathbf{H}}_{ir}^{\dagger}\tilde{\mathbf{H}}_{ir}$ and each column's 2-norm is normalized to 1. Let $\mathbf{P}_i^{*}$ be the optimized solution of 4.7. β_i is chosen to satisfy the total transmit power constraint $P_{tot,i}$ at user i. So $\beta_i = \sqrt{P_{tot,i}/\text{tr}\{\sigma_s^2\mathbf{H}_{ip}^{\perp}\mathbf{P}_i^{*}\mathbf{P}_i^{*\dagger}\mathbf{H}_{ip}^{\perp\dagger}\}}$, $i = 1, 2, ..., K$.

The decoding matrices initialization at SU i is to minimize the distance of decoding matrix $\mathbf{R}_i$ and corresponding channel matrix $\mathbf{H}_{ri}$. The interference from primary user should also be considered, which means that we should let the signal from primary user lie closer to the orthogonal complement of the subspace spanned by $\mathbf{R}_i$. Let $\mathbf{R}_i^{\perp}$ be a matrix whose rows are orthonormal basis for the complement of the subspace spanned by $\mathbf{R}_i$'s rows. It is easy to show that $\mathbf{R}_i^{\perp\dagger}\mathbf{R}_i^{\perp} = \mathbf{I}_d - \mathbf{R}_i^{\dagger}\mathbf{R}_i$. Then the design of decoder $\mathbf{R}_i$ is formulated as

$$\begin{aligned}\mathbf{R}_i \quad &= \arg\min_{\mathbf{R}_i\mathbf{R}_i^{\dagger}=\mathbf{I}_d} \left\{\omega\|\mathbf{H}_{ri}, \mathbf{R}_i^{\dagger}\|_{\text{dis}}^2 + \|\mathbf{H}_{pi}, \mathbf{R}_i^{\perp\dagger}\|_{\text{dis}}^2\right\} \\ &= \arg\min_{\mathbf{R}_i\mathbf{R}_i^{\dagger}=\mathbf{I}_d} \left\{\omega\|\mathbf{H}_{ri} - \mathbf{R}_i^{\dagger}\mathbf{R}_i\mathbf{H}_{ri}\|_F^2 + \|\mathbf{H}_{pi} - \mathbf{R}_i^{\perp\dagger}\mathbf{R}_i^{\perp}\mathbf{H}_{pi}\|_F^2\right\},\end{aligned} \tag{4.9}$$

where ω is a scalar weight to strike a balance between signal from secondary relay and interference from primary user's transmission. By simplifying (4.9) we have

$$\mathbf{R}_i = \arg \min_{\mathbf{R}_i\mathbf{R}_i^\dagger=\mathbf{I}_d} \mathrm{tr}\{\mathbf{R}_i(\mathbf{H}_{pi}\mathbf{H}_{pi}^\dagger - \omega\mathbf{H}_{ri}\mathbf{H}_{ri}^\dagger)\mathbf{R}_i^\dagger\}. \tag{4.10}$$

Since $\mathbf{R}_i \in \mathbb{C}^{d\times M}$, the rows of $\mathbf{R}_i$ should be the d least dominant eigenvectors of $(\mathbf{H}_{pi}\mathbf{H}_{pi}^\dagger - \omega\mathbf{H}_{ri}\mathbf{H}_{ri}^\dagger)$ whose 2-norms are normalized to 1. $\mathbf{R}_i^*$ is used to denote the solution of (4.9).

4.3.2 Iterative Relay Precoding and Secondary User Precoding/Decoding Design

After $\mathbf{T}_i$ and $\mathbf{R}_i$ are determined, we present a joint transmit and receive beamforming design at the relay station. Its precoding matrix $\mathbf{P}_r$ is designed to have the following structure:

$$\mathbf{P}_r = \mathbf{F}\mathbf{\Lambda}\mathbf{G}, \tag{4.11}$$

where $\mathbf{G} \in \mathbb{C}^{Kd\times N}$ is the receiving combining matrix, $\mathbf{F} \in \mathbb{C}^{(N-J)\times Kd}$ is the transmitting beamforming matrix, and $\mathbf{\Lambda}$ is a $Kd \times Kd$ diagonal matrix. It is the amplifying matrix of Kd datastreams after network coding. Let $\tilde{\mathbf{s}}_i$ be the desired signal vector at secondary user i and $\tilde{\mathbf{s}} = [\tilde{\mathbf{s}}_1^T \quad \tilde{\mathbf{s}}_2^T \quad ... \quad \tilde{\mathbf{s}}_{2K}^T]^T$ be the desired signal at all $2K$ users. Partition $\mathbf{G} = [\mathbf{G}_1^T \quad \mathbf{G}_2^T \quad ... \quad \mathbf{G}_K^T]^T$ and $\mathbf{F} = [\mathbf{F}_1 \quad \mathbf{F}_2 \quad ... \quad \mathbf{F}_K]$. Also denote $\mathbf{\Lambda} = \mathrm{diag}\{\mathbf{\Lambda}_1 \quad \mathbf{\Lambda}_2, ..., \mathbf{\Lambda}_K\}$. Then $\mathbf{G}_i \in \mathbb{C}^{d\times N}$, $\mathbf{F}_i \in \mathbb{C}^{(N-J)\times d}$ and $\mathbf{\Lambda}_i \in \mathbb{C}^{d\times d}$ are the receive beamforming matrix, transmit beamforming matrix and amplifying matrix of data-streams of the i-th SU pair i.e. users $2i-1$ and $2i$.

Let $\mathbf{s}$ and $\hat{\mathbf{s}}$ be the transmitted and received signal vector from all $2K$ secondary users. Also define $\mathbf{x} = [\mathbf{s}^T \quad \mathbf{x}_p^T]^T$ to be the transmitted signal in the first time slot including both signals from primary and secondary network. Then $\mathbf{x}$ is a $(2Kd + J) \times 1$ vector and the received signal of all $2K$ secondary users can be expressed as

$$\hat{\mathbf{s}} = \mathbf{H}_2\mathbf{F}\mathbf{\Lambda}\mathbf{G}\mathbf{H}_1\mathbf{x} + \mathbf{R}\tilde{\mathbf{x}}_p, \tag{4.12}$$

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are the equivalent channel matrices of the two hops, defined as

$$\mathbf{H}_1 = [\tilde{\mathbf{H}}_{1r}\mathbf{P}_1, \quad \tilde{\mathbf{H}}_{2r}\mathbf{P}_2, \ ..., \quad \tilde{\mathbf{H}}_{2Kr}\mathbf{P}_{2K}, \quad \mathbf{H}_{pr}], \tag{4.13}$$

$$\mathbf{H}_2 = [(\mathbf{R}_1\mathbf{H}_{r1})^T, \quad (\mathbf{R}_2\mathbf{H}_{r2})^T, \ ..., \quad (\mathbf{R}_{2K}\mathbf{H}_{r2K})^T]^T\mathbf{H}_{rp}^{\perp}. \tag{4.14}$$

Only the first term in (4.12) is related to relay precoding matrix $\mathbf{P}_r$. The second term is the received interference from primary network's transmission in the second time slot. It is up to the receive filters $\mathbf{R}_i$, $i = 1, 2, ..., 2K$ to mitigate this interference. Therefore in this subsection, we only focus on the first term. Let $\mathbf{s}^{(R)} = \mathbf{H}_2\mathbf{F}\mathbf{\Lambda}\mathbf{G}\mathbf{H}_1\mathbf{x}$ be the *relay-related* signal part. Since we need to cancel the interference to primary users, the equivalent channels from SU i to relay and from relay to SU i do not have reciprocity anymore. Therefore the receive and transmit beamformers of relay should be designed differently. Firstly we design the i-th SU pair's receive beamforming vector $\mathbf{G}_i$.

For the i-th SU pair, only the signal from users $2i-1$ and $2i$ are desired or considered self-interference which can be completely removed by self-interference cancellation. Signals transmitted by all other users are considered interference and should be mitigated by precoding design at the relay node. Define $\tilde{\mathbf{H}}_{1i}$ as the interference channel matrix associated with i-th user pair. Then $\tilde{\mathbf{H}}_{1i}$ is formed by removing $(2i-2)d+1 : 2id$-th columns from $\mathbf{H}_1$. Let $\mathbf{B}_{1i}$ be the $(2i-2)d+1 : 2id$-th column of matrix $\mathbf{H}_1$, then $\mathbf{B}_{1i}$ is considered channel matrix associated with useful signals. The goal is to receive "more" of the useful signal and remove as much as possible the harmful interference. Then receive beamforming matrix $\mathbf{G}_i$

is designed according to the following matrix-distance criterion:

$$
\begin{aligned}
\mathbf{G}_i &= \arg\min_{\mathbf{G}_i\mathbf{G}_i^\dagger=\mathbf{I}_d} \left\{\omega\|\mathbf{B}_{1i}, \mathbf{G}_i^\dagger\|_{dis}^2 + \|\tilde{\mathbf{H}}_{1i}, \mathbf{G}_i^{\perp\dagger}\|_{dis}^2\right\} \\
&= \arg\min_{\mathbf{G}_i\mathbf{G}_i^\dagger=\mathbf{I}_d} \left\{\omega\|\mathbf{B}_{1i} - \mathbf{G}_i^\dagger\mathbf{G}_i\mathbf{B}_{1i}\|_F^2 + \|\tilde{\mathbf{H}}_{1i} - \mathbf{G}_i^{\perp\dagger}\mathbf{G}_i^{\perp}\tilde{\mathbf{H}}_{1i}\|_F^2\right\} \qquad (4.15)
\end{aligned}
$$

where $\mathbf{G}_i^{\perp}$ is a matrix with its rows being the orthonormal basis for the orthogonal complement of the subspace spanned by $\mathbf{G}_i$'s rows and ω is a scalar weight to strike a balance between the desired signal and interference. It will be shown later that by setting $\omega = 0$, this matrix-distance based beamforming design becomes zero-forcing beamforming; therefore zero-forcing is a special case of our method. By simplifying (4.15) and using the fact that $\mathbf{G}_i^{\perp\dagger}\mathbf{G}_i^{\perp} = \mathbf{I}_d - \mathbf{G}_i^\dagger\mathbf{G}_i$, we get

$$
\mathbf{G}_i = \arg\min_{\mathbf{G}_i\mathbf{G}_i^\dagger=\mathbf{I}_d} \operatorname{tr}\left\{\mathbf{G}_i[\tilde{\mathbf{H}}_{1i}\tilde{\mathbf{H}}_{1i}^\dagger - \omega\mathbf{B}_{1i}\mathbf{B}_{1i}^\dagger]\mathbf{G}_i^\dagger\right\} . \qquad (4.16)
$$

Since $\mathbf{G}_i$ has d rows, its rows should be the d least eigenvectors of the matrix $\tilde{\mathbf{H}}_{1i}\tilde{\mathbf{H}}_{1i}^\dagger - \omega\mathbf{B}_{1i}\mathbf{B}_{1i}^\dagger$, with its 2-norm normalized to 1. The solution of (4.15) is denoted as $\mathbf{G}_i^*$.

The design of transmit beamforming matrix $\mathbf{F}_i$ of i-th user pair's data-stream follows in a similar way. Let $\tilde{\mathbf{H}}_{2i}$ be the interference channel matrix for i-th user pair. It is formed by deleting $(2i-2)d+1 : 2id$-th rows from $\mathbf{H}_2$. The channel matrix associated with the useful datastream of user pair i is the $(2i-2)d+1 : 2id$-th rows of $\mathbf{H}_2$, which is denoted as $\mathbf{B}_{2i}$. Using ω as a scalar weight to strike a balance between aligning the desired signal and mitigating inter-user interference, the relay transmit beamforming matrix for i-th SU pair is chosen to be

$$
\begin{aligned}
\mathbf{F}_i &= \arg\min_{\mathbf{F}_i^\dagger\mathbf{F}_i=\mathbf{I}_d} \left\{\omega\|\mathbf{B}_{2i}^\dagger, \mathbf{F}_i\|_{dis}^2 + \|\tilde{\mathbf{H}}_{2i}^\dagger, \mathbf{F}_i^{\perp}\|_{dis}^2\right\} \\
&= \arg\min_{\mathbf{F}_i^\dagger\mathbf{F}_i=\mathbf{I}_d} \left\{\omega\|\mathbf{B}_{2i}^\dagger - \mathbf{F}_i\mathbf{F}_i^\dagger\mathbf{B}_{2i}^\dagger\|_F^2 + \|\tilde{\mathbf{H}}_{2i}^\dagger - \mathbf{F}_i^{\perp}\mathbf{F}_i^{\perp\dagger}\tilde{\mathbf{H}}_{2i}^{\perp\dagger}\|_F^2\right\}, \qquad (4.17)
\end{aligned}
$$

Here $\mathbf{F}_i^{\perp}$ is a matrix whose columns are orthonormal basis of the orthogonal complement of the subspace spanned by $\mathbf{F}_i$'s columns. By simplifying (4.17) and using the fact $\mathbf{F}_i^{\perp}\mathbf{F}_i^{\perp\dagger} = \mathbf{I}_d - \mathbf{F}_i\mathbf{F}_i^{\dagger}$, the transmit beamforming matrix $\mathbf{F}_i$ is

$$\mathbf{F}_i = \arg\min_{\mathbf{F}_i^{\dagger}\mathbf{F}_i=\mathbf{I}_d} \operatorname{tr}\left\{\mathbf{F}_i^{\dagger}\left[\tilde{\mathbf{H}}_{2i}^{\dagger}\tilde{\mathbf{H}}_{2i} - \omega\mathbf{B}_{2i}^{\dagger}\mathbf{B}_{2i}\right]\mathbf{F}_i\right\}. \tag{4.18}$$

$\mathbf{F}_i$'s columns are chosen to be the d least dominant eigenvectors of matrix $\tilde{\mathbf{H}}_{2i}^{\dagger}\tilde{\mathbf{H}}_{2i} - \omega\mathbf{B}_{2i}^{\dagger}\mathbf{B}_{2i}$ with each column's 2-norm normalized to 1. The solution is denoted as $\mathbf{F}_i^{*}$. The amplifying matrix Λ is chosen to be an identity matrix since the power amplifying to different datastreams is not the focus in this chapter. A scalar β_r is introduced to satisfy the transmit power constraint for relay node. Let $P_{tot,r}$ be the maximum transmit power for relay node, then $\beta_r = \sqrt{P_{tot,r}/\operatorname{tr}\{\sigma_s^2\mathbf{H}_{rp}^{\perp}\mathbf{F}^{*}\Lambda\mathbf{G}^{*}\mathbf{H}_1\operatorname{diag}\{\sigma_s^2\mathbf{I}_{2Kd} \quad \sigma_p^2\mathbf{I}_J\}\mathbf{H}_1^{\dagger}\mathbf{G}^{*\dagger}\Lambda^{\dagger}\mathbf{F}^{*\dagger}\mathbf{H}_{rp}^{\perp\dagger} + \sigma_n^2\mathbf{H}_{rp}^{\perp}\mathbf{F}^{*}\Lambda\mathbf{G}^{*}\mathbf{G}^{*\dagger}\Lambda^{\dagger}\mathbf{F}^{*\dagger}\mathbf{H}_{rp}^{\perp\dagger}\}}$ and the precoding matrix at relay station is $\beta_r\mathbf{H}_{rp}^{\perp}\mathbf{F}\mathbf{\Lambda}\mathbf{G}$.

Now we show that zero-forcing beamforming is a special case of our proposed matrix-distance based IA-like precoding design. By setting $\omega = 0$ in (4.15), rows of receive beamforming matrix $\mathbf{G}_i$ should be chosen as the least dominant eigenvector of matrix $\tilde{\mathbf{H}}_{1i}\tilde{\mathbf{H}}_{1i}^{\dagger}$. In the case when $\tilde{\mathbf{H}}_{1i}$'s row null space has at least d dimension, $\mathbf{G}_i$ should satisfy $\mathbf{G}_i\tilde{\mathbf{H}}_{1i} = \mathbf{0}$. For transmit beamforming vector $\mathbf{F}_i$, by setting $\omega = 0$ in (4.17), columns of $\mathbf{F}_i$ are the d least dominant eigenvectors of $\tilde{\mathbf{H}}_{2i}^{\dagger}\tilde{\mathbf{H}}_{2i}$. Therefore when $\tilde{\mathbf{H}}_{2i}$'s null space has at least dimension d, $\mathbf{F}_i$ should satisfy $\tilde{\mathbf{H}}_{2i}\mathbf{F}_i = \mathbf{0}$. Therefore, when zero-forcing is possible, by setting $\omega = 0$, our proposed precoder becomes the zero-forcing precoder.

Lemma 2 : By performing relay precoding, the number of antennas at relay station should satisfy $N \geq (2K-1)d + J$ to perform zero-forcing and cancel out all inter-user interference (including the interference between primary and secondary network).

Proof: We already know that for each $\mathbf{G}_i$ and $\mathbf{F}_i$, to satisfy the zero-forcing criteria $\mathbf{G}_i^{\dagger}\tilde{\mathbf{H}}_{1i} = \mathbf{0}$ and $\tilde{\mathbf{H}}_{2i}\mathbf{F}_i = \mathbf{0}$, we need the dimension of $\tilde{\mathbf{H}}_{1i}$'s column null space and $\tilde{\mathbf{H}}_{2i}$'s row null space to be at least d. Since $\tilde{\mathbf{H}}_{1i}$ is an $N \times (2K-2)d + J$ matrix who has full rank almost surely,

$\text{rank}\{\tilde{\mathbf{H}}_{1i}\} = \min\{N, (2K-2)d+J\} \leq N-d$ should be satisfied such that its row null space's dimension $N - \text{rank}\{\tilde{\mathbf{H}}_{1i}\} \geq d$. Therefore, we need $N \geq (2K-1)d + J$. Similarly, for $\tilde{\mathbf{H}}_{2i}$, we need $\text{rank}\{\tilde{\mathbf{H}}_{2i}\} = \min\{N-J, (2K-2)d\} \leq N-J-d$. Therefore, $(2K-2)d \leq N-J-d$ should be satisfied, which also leads to $N \geq (2K-1)d + J$. $\square$

After the relay precoding matrix is designed, end users can adjust their precoding and decodicng matrices to further improve the performance. Since the relay precoding matrix is decomposed into three parts including receive combining matrix $\mathbf{G}$, transmit beamforming matrix $\mathbf{F}$, and amplifying matrix $\mathbf{\Lambda}$ for Kd network coded datastreams, the two-way transmission can be decomposed into two one-way transmission, namely SU-to-relay and relay-to-SU. Therefore, the secondary user precoding can be designed only based on $\mathbf{G}$ while the secondary user decoding matrix is designed based on $\mathbf{F}$.

For the transmit precoding design at secondary user $k, k \in \{1, 2, ..., 2K\}$, let $\mathbf{A}_k$ be the equivalent channel matrix of its signal, and $\tilde{\mathbf{A}}_k$ be the channel matrix associated with interference. Since at relay, the signal of two users' are network coded together, then k-th user's desired signal is indexed as $\lceil k/2 \rceil$-th datastream at the relay station. $\mathbf{A}_k$ and $\tilde{\mathbf{A}}_k$ are defined as the follows

$$\mathbf{A}_k = \mathbf{G}_{\lceil k/2 \rceil}\mathbf{H}_{kr}\mathbf{H}_{kp}^{\perp}\ , \quad \tilde{\mathbf{A}}_k = \tilde{\mathbf{G}}_{\lceil k/2 \rceil}\mathbf{H}_{kr}\mathbf{H}_{kp}^{\perp}, \tag{4.19}$$

where

$$\tilde{\mathbf{G}}_{\lceil k/2 \rceil} = [\mathbf{G}_1^T, ..., \mathbf{G}_{\lceil k/2 \rceil - 1}^T, \mathbf{G}_{\lceil k/2 \rceil + 1}^T, ..., \mathbf{G}_K^T]^T.$$

The design of k-th user's precoding matrix $\mathbf{P}_k$ follows a similar way as the design of relay transmit precoding matrix $\mathbf{F}_i$.

$$\begin{aligned} \mathbf{P}_k &= \arg\min_{\mathbf{P}_k^{\dagger}\mathbf{P}_k = \mathbf{I}_d} \left\{\omega \|\mathbf{A}_k, \mathbf{P}_k\|_{dis}^2 + \|\tilde{\mathbf{A}}_k, \mathbf{P}_k^{\perp}\|_{dis}^2\right\} \\ &= \arg\min_{\mathbf{P}_k^{\dagger}\mathbf{P}_k = \mathbf{I}_d} \text{tr}\left\{\mathbf{P}_k^{\dagger}\left[\tilde{\mathbf{A}}_k^{\dagger}\tilde{\mathbf{A}}_k - \omega\mathbf{A}_k^{\dagger}\mathbf{A}_k\right]\mathbf{P}_k\right\} \end{aligned} \tag{4.20}$$

Then the columns of $\mathbf{P}_k$ should be the d least dominant eigenvectors of matrix $\tilde{\mathbf{A}}_k^\dagger\tilde{\mathbf{A}}_k - \omega\mathbf{A}_k^\dagger\mathbf{A}_k$.

For k-th user's receive decoding matrix $\mathbf{R}_k$ design, we define

$$\mathbf{C}_k = \mathbf{H}_{rk}\mathbf{H}_{rp}^\perp\mathbf{F}_{\lceil k/2\rceil}\ , \tilde{\mathbf{C}}_k = [\mathbf{H}_{rk}\mathbf{H}_{rp}^\perp\tilde{\mathbf{F}}_{\lceil k/2\rceil}, \mathbf{H}_{pi}]. \tag{4.21}$$

where

$$\tilde{\mathbf{F}}_{\lceil k/2\rceil} = [\tilde{\mathbf{F}}_1, ..., \tilde{\mathbf{F}}_{\lceil k/2\rceil-1}, \tilde{\mathbf{F}}_{\lceil k/2\rceil+1}, ..., \tilde{\mathbf{F}}_K].$$

The design of $\mathbf{R}_i$ is formulated as

$$\begin{aligned}\mathbf{R}_k &= \arg\min_{\mathbf{R}_k\mathbf{R}_k^\dagger=\mathbf{I}_d} \left\{\omega\|\mathbf{C}_k, \mathbf{R}_k^\dagger\|_{dis}^2 + \|\tilde{\mathbf{C}}_k, \mathbf{R}_k^{\dagger\perp}\|_{dis}^2\right\} \\ &= \arg\min_{\mathbf{R}_k\mathbf{R}_k^\dagger=\mathbf{I}_d} \mathrm{tr}\left\{\mathbf{R}_k[\tilde{\mathbf{C}}_k\tilde{\mathbf{C}}_k^\dagger - \omega\mathbf{C}_k\mathbf{C}_k^\dagger]\mathbf{R}_k^\dagger\right\}.\end{aligned} \tag{4.22}$$

Therefore the rows of $\mathbf{R}_k$ should be the d least dominant eigenvectors of matrix $\tilde{\mathbf{C}}_k\tilde{\mathbf{C}}_k^\dagger - \omega\mathbf{C}_k\mathbf{C}_k^\dagger$. The detailed procedure of how to perform iteration is summarized in Algo. 1.

4.3.3 Convergence of Proposed Iterative Method

In this two time slot transmission scheme, the design of precoding/decoding matrices is divided into two parts. The first part is the precoding matrices design at the users and the decoding matrix $\mathbf{G}$ design at the relay in the first time slot. The second part is to design the relay precoding matrix $\mathbf{F}$ and user decoding matrices in the second time slot. The iteration is performed for these two parts in parallel. In the next we show that for each part, our algorithm converges.

For the design in the first time slot. The two objectives of the two problems (4.16) and (4.20) are denoted as O_i and $\tilde{O}_k$ respectively. We have

$$O_i = \sum_{k=1,k\neq 2i-1,k\neq 2i}^{K} \mathrm{tr}\left\{\mathbf{G}_i(\tilde{\mathbf{H}}_{kr}\mathbf{P}_k)(\tilde{\mathbf{H}}_{kr}\mathbf{P}_k)^\dagger\mathbf{G}_i^\dagger\right\} + \mathrm{tr}\left\{\mathbf{G}_i\mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger\mathbf{G}_i^\dagger\right\}$$
$$-\omega\mathrm{tr}\left\{\mathbf{G}_i(\tilde{\mathbf{H}}_{2i-1,r}\mathbf{P}_{2i-1})(\tilde{\mathbf{H}}_{2i-1,r}\mathbf{P}_{2i-1})^\dagger\mathbf{G}_i^\dagger\right\} - \omega\mathrm{tr}\left\{\mathbf{G}_i(\tilde{\mathbf{H}}_{2i,r}\mathbf{P}_{2i})(\tilde{\mathbf{H}}_{2i,r}\mathbf{P}_{2i})^\dagger\mathbf{G}_i^\dagger\right\}$$
$$\tilde{O}_k = \sum_{i=1,i\neq\lceil k/2\rceil}^{K} \mathrm{tr}\left\{\mathbf{P}_k^\dagger(\mathbf{G}_i\tilde{\mathbf{H}}_{kr})^\dagger(\mathbf{G}_i\tilde{\mathbf{H}}_{kr})\mathbf{P}_k\right\} - \omega\mathrm{tr}\left\{\mathbf{P}_k^\dagger(\mathbf{G}_{\lceil k/2\rceil}\tilde{\mathbf{H}}_{kr})^\dagger(\mathbf{G}_{\lceil k/2\rceil}\tilde{\mathbf{H}}_{kr})\mathbf{P}_k\right\} \quad (4.23)$$

It can be shown that

$$\sum_{i=1}^{K} O_i = \sum_{k=1}^{2K} \tilde{O}_k + \sum_{i=1}^{K} \mathrm{tr}\left\{\mathbf{G}_i\mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger\mathbf{G}_i^\dagger\right\}$$

The user precoding design in the first time slot can be expressed as

$$\min_{\mathbf{P}_k^\dagger\mathbf{P}_k=\mathbf{I},k=1,2\ldots2K} \tilde{O}_k \Leftrightarrow \min_{\mathbf{P}_k^\dagger\mathbf{P}_k=\mathbf{I},k=1,2\ldots2K} \tilde{O}_k + \sum_{i=1}^{K} \mathrm{tr}\left\{\mathbf{G}_i\mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger\mathbf{G}_i^\dagger\right\} \quad (4.24)$$

The equality holds because $\sum_{i=1}^{K} \mathrm{tr}\left\{\mathbf{G}_i\mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger\mathbf{G}_i^\dagger\right\}$ is not dependent on value of $\mathbf{P}_k$ therefore the two problems in (4.24) will lead to the same solution. This problem can be decomposed into $2K$ subproblems as in (4.20) and these subproblems can be solved in parallel and independent of each other.

On the other hand, the relay decoding matrix $\mathbf{G}$ is designed by solving the following problem

$$\min_{\mathbf{G}\mathbf{G}^\dagger=\mathbf{I}} \sum_{i=1}^{K} O_i, \quad (4.25)$$

which can be decomposed into K subproblems as in (4.16). These subproblems can also be solved in parallel and independent of each other.

The iteration for the first time slot precoding/decoding design is performed by alternately solving problem (4.24) and (4.25). We have shown that both problems are to minimize the same objective. Since this objective is lower bounded by zero, we can conclude that the iteration converges.

The convergence of the iteration in the second time slot, i.e. the iteration between the optimization of relay precoding matrix $\mathbf{F}$ and user receive decoding matrices $\mathbf{R}_i$, can be proved in a similar way. They try to minimize the same objective and the objective is lower bounded by zero.

4.3.4 Refining Decoding Matrices to Decode Each Datastream

In the previous subsection, we designed relay precoding which tries to mitigate inter-pair interference. After performing self-interference cancellation, the multiple data-streams for each SU are still "mixed up" together which leaves to the decoding filter at user i to solve. Since $\mathbf{R}_i$ has already been designed, a secondary decoding matrix $\tilde{\mathbf{R}}_i \in \mathbb{C}^{d\times d}$ is designed to be cascaded with $\mathbf{R}_i$ to distinguish among the d different data-streams for user i. Therefore the decoding matrix at user i is $\tilde{\mathbf{R}}_i\mathbf{R}_i$. Express the received signal $\hat{\mathbf{s}}_i$ as

$$\hat{\mathbf{s}}_i = \tilde{\mathbf{R}}_i\check{\mathbf{H}}_i\mathbf{x}_t,$$

where $\check{\mathbf{H}}_i$ is the equivalent channel matrix from both primary and secondary transmitters to secondary user i, $\check{\mathbf{H}}_i = \mathbf{R}_i[\mathbf{H}_{ri}\mathbf{T}_r\mathbf{H}_1 \quad \mathbf{H}_{pi}]$, and $\mathbf{x}_t = [\mathbf{s} \quad \mathbf{x}_p \quad \tilde{\mathbf{x}}_p]$ is the transmitted signal from all secondary users in the first time slot and from primary user in both time slots. Denote j-th row of $\tilde{\mathbf{R}}_i$ by $\tilde{\mathbf{r}}_{ij}$ which is the receive combining vector of j-th datastream at user i. It wants to receive the j-th datastream sent by user $i-1$ if i is even and user $i+1$ if i is odd. Let $l(i,j) = j + id$ if i is odd while $l(i,j) = j + (i-2)d$ if i is even. Let $\mathbf{h}_{ij}$ be the channel vector associated with j-th desired signal at user i; then $\mathbf{h}_{ij}$ is the $l(i,j)$-th column of $\check{\mathbf{H}}_i$. Also, by performing self-interference cancellation, the signal sent by i-th secondary user can be completely removed from j-th estimated signal at user i. Let $\tilde{\mathbf{H}}_{ij}$, which is formed by deleting $l(i,j)$-th and $(i-1)d+1 : id$-th columns from $\check{\mathbf{H}}_i$, be the interference channel matrix. $\tilde{\mathbf{r}}_{il}$ can be chosen using the matrix-distance criterion:

$$\tilde{\mathbf{r}}_{ij} \quad = \arg\min_{\tilde{\mathbf{r}}_{ij}\tilde{\mathbf{r}}_{ij}^{\dagger}=1} \left\{ \omega \|\tilde{\mathbf{h}}_{ij}, \tilde{\mathbf{r}}_{ij}^{\dagger}\|_{dis}^2 + \|\tilde{\mathbf{H}}_{ij}, \tilde{\mathbf{r}}_{ij}^{\perp\dagger}\|_{dis}^2 \right\}. \tag{4.26}$$

So $\tilde{\mathbf{r}}_{ij}$ should be chosen as the least dominant eigenvector of $\tilde{\mathbf{H}}_{1i}\tilde{\mathbf{H}}_{1i}^{\dagger} - \omega\tilde{\mathbf{h}}_{il}\tilde{\mathbf{h}}_{1i}^{\dagger}$, with its 2-norm normalized to 1.

It can be shown that when $N \geq (2K-1)d + J$, the inter-user interference can be completely removed. In $\tilde{\mathbf{H}}_{ij}$ only the columns associated with i-th user's desired datastreams (not including j-th) are non zero. So we have $\text{rank}\{\tilde{\mathbf{H}}_{ij}\} = d - 1$ if zero-forcing precoder at relay is used. The rank of $\tilde{\mathbf{H}}_{ij}$'s column null space is 1 and by setting $\omega = 0$ in (4.26), $\tilde{\mathbf{r}}_{ij}\tilde{\mathbf{H}}_{ij} = \mathbf{0}$, which means the interference among different datastreams of the same user can also be removed completely.

The proposed iterative algorithm which optimize the precoding/decoding matrices at the secondary users and the precoding matrix at the relay node is summarized as follows.

ALGORITHM 4.1

---

1) Each user initialize $\mathbf{P}_i$ and $\mathbf{R}_i$ as described in Sec. 4.3.1.

2) Repeat (iterate):

   - Design $\mathbf{G}, \mathbf{F}$ and $\mathbf{\Lambda}$ based on most current $\mathbf{P}_i$ and $\mathbf{R}_i, i = 1, 2, ..., K$ as described in Sec. 4.3.2.

   - Update $\mathbf{P}_i$ and $\mathbf{R}_i$ based on $\mathbf{G}$ and $\mathbf{F}$ as described in Sec. 4.3.2.

   Continue until ending iteration criterion is meet.

3) Refine each $\mathbf{R}_i$ to recover d datastreams wanted by each user as in Sec. 4.3.4.

One simple way to end the iteration process is to set a maximum iteration number and algorithm ends when this max iteration number is reached.

4.4 Simulation Results

In this section we provide simulation results to show the effectiveness of our algorithm. Consider a secondary network consisting of $K = 2$ SU pairs (total 4 SUs) and one non-regenerative relay node, coexisting with one source-destination primary user pair. All channels experience flat Rayleigh fading and the channel gains remain constant during the two time slots transmission. Assume the expectation of all channel power gains is 1 and $\sigma_s^2 = \sigma_p^2 = \sigma_n^2 = 1$. Also assume the number of antennas at secondary users is $M = 3$ and relay has N antennas. The number of data-streams d=2. The primary user is equipped with a single antenna $J = 1$. We evaluate the sum rate of K secondary users of our proposed algorithm. Also by comparing it with MSE-based algorithm in Chapter 3 and zero-forcing, we illustrate the performance of this proposed matrix distance based alignment method in both high and low SNR regime with different number of antennas at the relay.

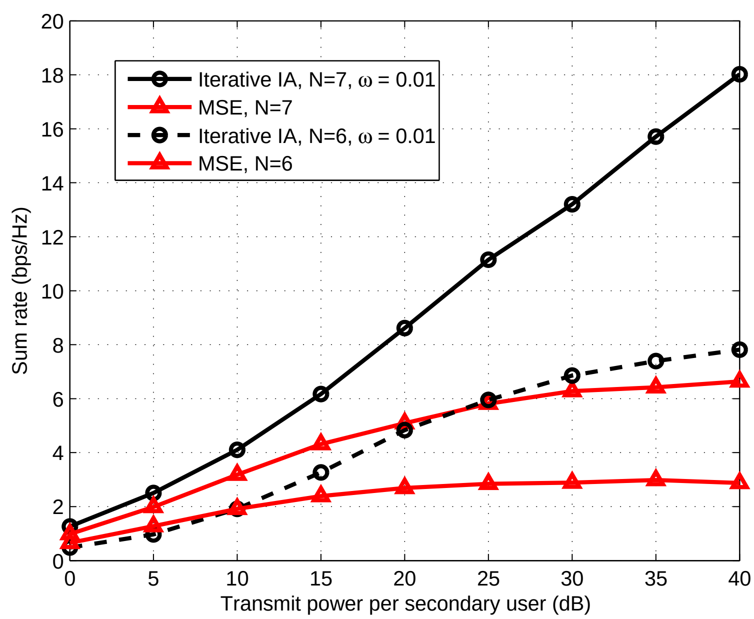

Figure 4.2: Sum rate (4.27) vs. transmit power per secondary user

Define the sum rate of $2K$ SUs as

$$\text{Rate} = \sum_{i=1}^{2K} \sum_{j=1}^{d} \log_2 \left(1 + \frac{\mid \{\tilde{\mathbf{R}}_i \check{\mathbf{H}}_i\}_{j,l(i,j)} \mid^2 \sigma_s^2}{\mathrm{i}_{Nij}} \right) \tag{4.27}$$

where i_{Nij} is noise and interference power for j-th datastream at user i given by

$$\mathrm{i}_{Nij} = \tilde{\mathbf{r}}_{ij} \tilde{\mathbf{H}}_{ij} \mathrm{diag}\{\sigma_s^2 \mathbf{I}_{(2K-1)d-1}, \sigma_p^2 \mathbf{I}_{2J}\} \tilde{\mathbf{H}}_{ij}^\dagger \tilde{\mathbf{r}}_{ij}^\dagger$$

$$+ \tilde{\mathbf{r}}_{ij} \mathbf{R}_i \mathbf{H}_{ri} \mathbf{T}_r \mathbf{T}_r^\dagger \mathbf{H}_{ri}^\dagger \mathbf{R}_i^\dagger \tilde{\mathbf{r}}_{ij}^\dagger \sigma_n^2 + \tilde{\mathbf{r}}_{ij} \mathbf{R}_i \mathbf{R}_i^\dagger \tilde{\mathbf{r}}_{ij}^\dagger \sigma_n^2$$

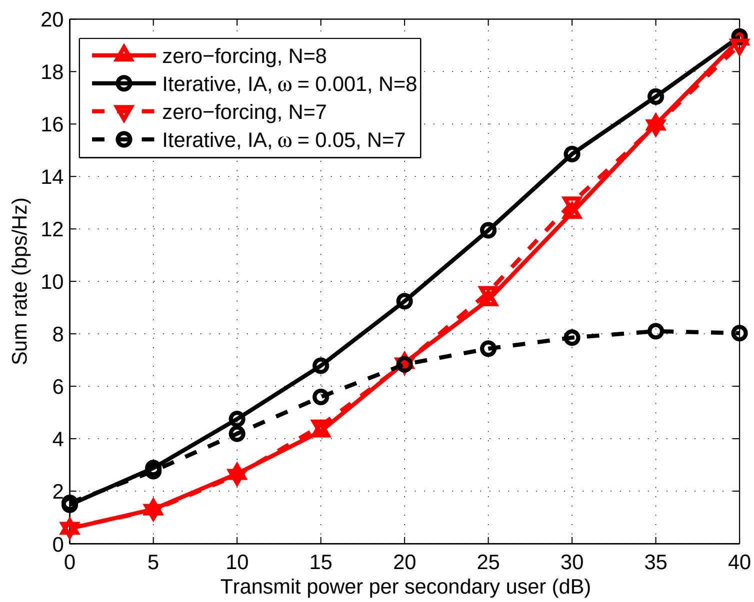

Figure 4.3: Sum rate (4.27) vs. transmit power per secondary user

In Fig. 4.2 the sum rate of secondary network using matrix distance alignment algorithm is compared with MSE-based algorithm (non-iterative) in Chapter 3 for $N \geq (2K-1)d + J$ and $N < (2K-1)d + J$ cases. It is observed that IA-like algorithm significantly outperforms MSE-based algorithm in medium to high SNR regime, because the IA-like algorithm focuses on inter-pair interference and primary-to-secondary interference, which are the most important problems when SNR is high and the MSE-based algorithm deals with both interference and noise. Fig. 4.3 compares the proposed algorithm with zero-forcing, which is

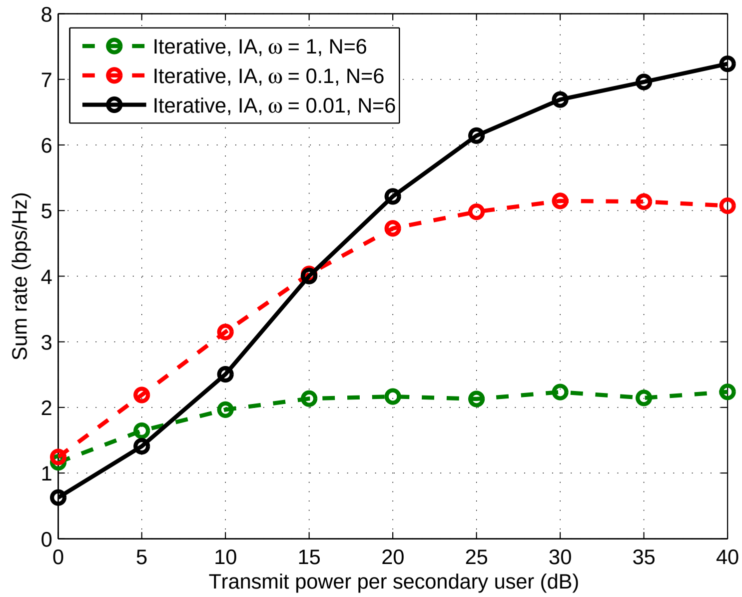

Figure 4.4: Sum rate (4.27) vs. transmit power per secondary user

achieved by setting $\omega = 0$ for $N \geq (2K-1)d + J$ case. When $N = (2K-1)d + J$, i.e. $N = 7$, the IA-like algorithm outperforms zero-forcing in low to medium SNR regime while the zero-forcing achieves better performance in high SNR regime because interference among users is a more important factor compared to desired signal strength when SNR is high. For $N > (2K-1)d + J$, e.g. $N = 8$ case, the proposed algorithm achieves a higher sum rate for the entire SNR regime because more degrees of freedom are provided in this case and the proposed algorithm not only tries to cancel the interference but also employ the extra degrees of freedom to maximize the desired signal strength. In Fig. 4.4 we show the effect of different choices of the scaler weight ω on the system performance. A larger value of ω means more weight on the signal strength over interference. It is observed that in the low SNR regime, it is better to choose larger ω while in the higher SNR regime the negative effect of interference is more severe, therefore smaller value of ω is more desirable. In all simulations, the proposed matrix distance based alignment algorithm performs 10 iterations.

4.5 Conclusion

In this chapter we proposed a interference and signal alignment precoding algorithm based on matrices distance for multi-pair two-way relay system in a cognitive radio scenario where secondary users and relay are equipped with multiple antennas. This proposed algorithm not only considers the negative effect of inter-pair interference but also takes into consideration the desired signal strength. We showed that our algorithm is a generalization of zero-forcing algorithm and can also work when zero-forcing is not possible. The performance of our proposed algorithm was compared with zero-forcing and MSE-based algorithms, and discussed for different SNR regimes and for both $N \geq (2K-1)d+J$ and $N < (2K-1)d+J$ cases.

Chapter 5

Beamforming for Generalized MIMO Y Channels

Relaying is a powerful technique to support the concurrent transmission of multiple users, including single/multiple user pair two-way transmission, multi-way transmission, etc. In this chapter, we consider a newly proposed MIMO Y channel. A typical Y channel consists of 3 users transmitting to each other through the help of a relay node. In this system model, each user transmits two independent messages to the other two users. A generalized Y channel refers to the network consisting of $K, K \geq 3$ users and a relay. Each user has $K-1$ independent messages for all the other $K-1$ users. With the deployment of multiple antennas at both end users and the relay, $K(K-1)$ messages can be conveyed to their desired receivers within two time slot (assume this network works in a half-duplex mode). The beamforming design of such a network is usually based on signal space alignment and network coding concept in recent literatures. However, this requires that the end users have abundant antennas and it is up to the end users to align their signals so the two signals that should be network coded together fall into the same direction at the relay. In this chapter, we propose an alternative method where the end users just try to align their signals into a smaller subspace at the relay but it is eventually the relay's beamforming design which eliminates the harmful effect of inter-pair interference.

This chapter is organized as follows. In Sec. 5.1, the background and related work of MIMO Y channel is introduced. System model and main results are presented in Sec. 5.2. A signal group based alignment method is proposed in Sec. 5.3. This method significantly reduces the required antenna number at end users with a higher antenna number at the relay, compared with the signal space alignment method. Simulation results are given in Sec. 5.4 to show the achieved DOF of our proposed method. Conclusions are drawn in Sec. 5.5.

5.1 Introduction

Bi-directional communications between two users can be facilitated by a relay station. Using the fact that each user is fully aware of its own sent signal, the network coding concept [58] is utilized in bi-directional communication systems to allow information exchange of two users within two time slots, assuming all nodes in this system operate in a half-duplex manner. In a typical two user bidirectional communication system, data transmission is completed in two phases: multiple access channel (MAC) phase and broadcast channel (BC) phase. During the MAC phase, two users transmit to the relay node simultaneously. In the BC phase, the relay retransmits its received signal during the first time slot (with or without processing) to the two users and by performing self-interference cancellation, the two users can decode their desired signals. This bi-direction relay channel has been extensively investigated by researchers during the recent years. Many extensions have also been considered, such as multi-pair two-way relaying, multi-user multi-way transmission [61] and the recently proposed Y channel [62], [63]. In these multi-user networks, system performance is interference limited and hence how to avoid inter-user interference becomes a crucial design issue. Many zero-forcing based beamforming designs have been proposed in recent literature for multi-user two-way or multi-way transmission. In [44], multiple users which are equipped with single antenna conduct two-way transmission via a MIMO relay. The relay transceiver is optimized based on MMSE and zero-forcing criteria. Both co-channel interference (CCI) and self-interference (SI) are eliminated by the proposed methods. In [65], a projection based separation of multiple operators (ProBaSeMo) relay transmit strategy is proposed for a mulit-pair two-way relaying channel, which is inspired by block diagonalization (BD) and regularized block diagonalization (RBD). This algorithm provides significant sharing gain [65].

The three user Y channel, as a multi-user multi-way transmission scheme has been drawing increasing research attention [63, 64]. In a typical Y channel, three users communicate with each other with the help of one relay node. Each user has two independent messages, each message is meant to be sent to one of the other two users. A signal space alignment method is proposed in [63, 64] for MIMO Y channel. In these papers, a K, $K = 3$, user relay channel is considered where each user intends to send two independent messages to the other two users via the help of an intermediate relay. This system model is a generalization of two-way transmission and by the concept of network coding, two messages that are to be exchanged by two users can be network coded together given the fact that each user is aware of its own sent messages and can perform self-interference cancellation. The proposed signal space alignment method requires that the two messages to be network coded together be aligned into the same direction at the relay station in the multiple access (MAC) phase. Therefore, the $K(K-1)$ messages only occupy $K(K-1)/2$ dimensions at the relay, which reduces the minimum required number of antennas at the relay node. This novel method is extended to more general system models in [62] such as multi-user Y channel (also referred to as generalized Y channel) in which there are K, $K > 3$, users and each user conveys $K-1$ independent messages to all other $K-1$ users. In order to support the transmission of $K(K-1)$ messages in this system, the number of antennas M_i at the i-th user and the number of antennas N at the relay should satisfy $M_i \geq K-1$, $N \geq \frac{K(K-1)}{2}$, and $\min_{i,j,i\neq j}\{M_i + M_j\} > N$ [62].

In this chapter, a signal group based alignment is proposed which requires fewer antennas at the end users compared to [62]. This method is suitable for the scenarios where the relay serves as a central station while the end users are small mobile devices such as in a cellular network. The key idea is to assign signals from all users into several groups and the beamforming vectors and receive combining vectors for the signals in each group are jointly designed to align these signals into a smaller subspace. This algorithm uses network coding concept and signal group alignment to support the multi-user MIMO Y channel free from inter-user and inter-message interference.

Notations: In this chapter, bold upper and lower case letters are used to denote matrices and vectors respectively. Superscripts $(.)^*$, $(.)^T$ and $(.)^H$ represent complex conjugate, transpose and complex conjugate transpose (Hermitian) operation, respectively, on a vector/matrix. $\mathbb{E}\{.\}$ denotes the expectation operation and $\text{tr}\{.\}$ denotes the trace of a matrix. The null space of a matrix is denoted as $\text{null}\{.\}$ and the subspace spanned by the columns of a matrix is denoted by $\text{span}\{.\}$.

5.2 System Model

Consider a generalized MIMO Y channel with K end users and one relay node. Each user is equipped with M antennas while the relay node has N antennas. Each user has $K-1$ independent messages for the other $K-1$ users. This system model is illustrated in Fig. 5.1. Let s_{ij} denote the data signal from user i to user j and $\mathbf{v}_{ij} \in \mathbb{C}^{M\times 1}$ denote the beamforming vector for signal s_{ij}. Then the precoding matrix at user i can be expressed as $\mathbf{T}_i = \beta_i \mathbf{V}_i$ where $\mathbf{V}_i = [\mathbf{v}_{i1}, ..., \mathbf{v}_{i(i-1)}, \mathbf{v}_{i(i+1)}, ..., \mathbf{v}_{iK}] \in \mathbb{C}^{M\times(K-1)}$ and β_i is a scalar to scale the transmit power. The signal vector sent by user i is $\beta_i \mathbf{T}_i \mathbf{s}_i$, where $\mathbf{s}_i = [s_{i1}, ..., s_{i(i-1)}, s_{i(i+1)}, ..., s_{iK}]^T$ is the information vector from user i. Assume the signals s_{ij}, $i \neq j$, are independent from each other with zero mean and variance σ_s^2.

A half-duplex transmission scheme is considered in this chapter. In the first time slot, all K users send signals to the relay while in the second time slot, relay sends a linear combination of its received signal to K users. Let $\mathbf{H}_{ir} \in \mathbb{C}^{N\times M}$ and $\mathbf{H}_{ri} \in \mathbb{C}^{M\times N}$ denote the channel matrices from user i to the relay and from the relay to user i and assume the channel state information (CSI) is completely known. Assume each element in the channel matrix is drawn from i.i.d. (independent and identically distributed) complex Gaussian distribution with zero mean and unit variance. The received signal at the relay node during the first time slot (MAC phase) is expressed as

$$\mathbf{y}_r = \sum_{i=1}^{K} \beta_i \mathbf{H}_{ir} \mathbf{V}_i \mathbf{s}_i + \mathbf{n}_r, \tag{5.1}$$

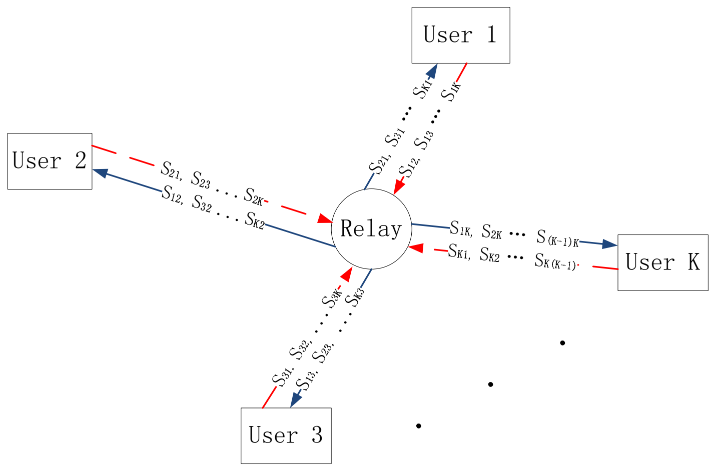

Figure 5.1: System model

where $\mathbf{n}_r \in \mathbb{C}^{N\times 1}$ is the additive white complex Gaussian noise at the relay with zero mean and $\mathbb{E}\left\{\mathbf{n}_r\mathbf{n}_r^H\right\} = \sigma_n^2\mathbf{I}$. In the second time slot (BC phase), the relay precodes its received signal $\mathbf{y}_r$ with a precoding matrix $\mathbf{T}_r$. So the received signal at each user during the second time slot is

$$\mathbf{y}_i = \mathbf{H}_{ri}\mathbf{T}_r\sum_{k=1}^{K}\beta_k\mathbf{H}_{kr}\mathbf{V}_k\mathbf{s}_k + \mathbf{H}_{ri}\mathbf{T}_r\mathbf{n}_r + \mathbf{n}_i, \tag{5.2}$$

where $\mathbf{n}_i \in \mathbb{C}^{N\times 1}$ is the additive white complex Gaussian noise at user i with zero mean and $\mathbb{E}\left\{\mathbf{n}_i\mathbf{n}_i^H\right\} = \sigma_n^2\mathbf{I}$. Each user i can perform self-interference calcellation upon the receive of signal vector $\mathbf{y}_i$. Let $\tilde{\mathbf{y}}_i$ be the signal vector at user i after self-interference cancellation. Then we have

$$\tilde{\mathbf{y}}_i = \mathbf{H}_{ri}\mathbf{T}_r\sum_{k=1,k\neq i}^{K}\beta_k\mathbf{H}_{kr}\mathbf{V}_k\mathbf{s}_k + \mathbf{H}_{ri}\mathbf{T}_r\mathbf{n}_r + \mathbf{n}_i.$$

Let $\mathbf{R}_i = [\mathbf{r}_{i,1}, \cdots, \mathbf{r}_{i,i-1}, \mathbf{r}_{i,i+1}, ..., \mathbf{r}_{i,K}]^T \in \mathbb{C}^{(K-1)\times M}$ be the receiver combining matrix at user i while $\mathbf{r}_{i,j}^T \in \mathbb{C}^{1\times M}$, $j \neq i$, is the combining vector for s_{ij}. Then the estimated signal at user i is

$$\hat{\mathbf{s}}_i = \mathbf{R}_i\tilde{\mathbf{y}}_i = [\hat{s}_1, \hat{s}_{2,i}, ..., \hat{s}_{i-1,i}, \hat{s}_{i+1,i}, ..., \hat{s}_{K,i}]^T, \tag{5.3}$$

with $\hat{s}_{j,i} = \mathbf{r}_{i,j}^T\tilde{\mathbf{y}}_i$, $i \neq j$, being the estimated signal at user i from user j. By the concept of network coding, signal s_{ij} and s_{ji} can be network coded together at the relay since each user can perform self-interference cancellation to decode its desired signal. Assume the inter-user and inter-message interference are treated as noise. Define

$$a_{ji} = \mathbf{r}_{ij}^T\mathbf{H}_{ri}\mathbf{T}_r \sum_{k=1,k\neq i,k\neq j}^{K} \sum_{j'=1,j'\neq k}^{K} \beta_k\mathbf{H}_{kr}\mathbf{v}_{kj'}s_{kj'}$$

$$+\mathbf{r}_{ij}^T\mathbf{H}_{ri}\mathbf{T}_r \sum_{j'=1,j'\neq i,j'\neq j}^{K} \beta_j\mathbf{H}_{jr}\mathbf{v}_{jj'}s_{jj'},$$

$$b_{ji} = \mathbf{r}_{ij}^T\mathbf{H}_{ri}\mathbf{T}_r\mathbf{n}_r + \mathbf{r}_{ij}^T\mathbf{n}_i,$$

$$c_{ji} = \beta_j\mathbf{r}_{ij}^T\mathbf{H}_{ri}\mathbf{T}_r\mathbf{H}_{jr}\mathbf{v}_{ji}s_{ji},$$

where a_{ji} represent the inter-user and inter-message interference, b_{ji} is the overall noise and c_{ji} is the desired signal at user i from user j. Then $\hat{s}_{ji} = a_{ji} + b_{ji} + c_{ji}$. Let P denote the user transmit power constraint, i.e.

$$\beta_i^2\mathbb{E}\,\mathrm{tr}\Big\{(\mathbf{V}_i\mathbf{s}_i)(\mathbf{V}_i\mathbf{s}_i)^H\Big\} \leq P, \;\; 1 \leq i \leq K, \tag{5.4}$$

$$\mathbb{E}\,\mathrm{tr}\Big\{(\mathbf{T}_r\mathbf{y}_r)(\mathbf{T}_r\mathbf{y}_r)^H\Big\} \leq P. \tag{5.5}$$

Then the achievable rate of message s_{ij} is calculated as [28]

$$R_{ji}(P) = \frac{1}{2}\left\{\log_2\left(1 + \frac{\mathbb{E}\{|c_{ji}|^2\}}{\mathbb{E}\{|a_{ji}|^2\} + \mathbb{E}\{|b_{ji}|^2\}}\right)\right\} \tag{5.6}$$

The $\frac{1}{2}$ factor is due to the fact that this system operates in a half-duplex mode and K users' transmission is completed in two time slots. The sum achievable degree of freedom (DOF) of this K-user MIMO Y channel is [62]

$$\eta(K) = \lim_{P\to\infty} \frac{\sum_{i=1}^{K}\sum_{j=1,j\neq i}^{K} R_{ji}(P)}{\log_2 P} \tag{5.7}$$

Also, define the normalized DOF as:

$$\bar{\eta}(K) = \frac{\eta(K)}{\text{average antenna number per node}}. \tag{5.8}$$

In recent research works, the beamforming vectors $\mathbf{v}_{ij}$ and $\mathbf{v}_{ji}$ are usually designed to satisfy $\text{span}\{\mathbf{H}_{ir}\mathbf{v}_{ij}\} \doteq \text{span}\{\mathbf{H}_{jr}\mathbf{v}_{ji}\}$, i.e., the subspaces spanned by $\mathbf{H}_{ir}\mathbf{v}_{ij}$ and $\mathbf{H}_{jr}\mathbf{v}_{ji}$ are the same. This means the signal vectors associated with s_{ij} and s_{ji} are aligned into the same direction at the relay node such that $K(K-1)$ signals only occupy $K(K-1)/2$ dimensions. The requirement on the number of antennas is $2M > N$, $i \neq j$, $M \geq K-1$ and $N \geq K(K-1)/2$. This requires the end users to be equipped with a large number of antennas especially when K is large. However, under some circumstances, it is more realistic that the relay node can have more antennas while the number of antennas of end users should be made as small as possible. One example is cellular network where the base station (BS) serves as the relay and the end users are mobile users. Therefore we propose a signal group based beamforming design. The requirement on the number of antennas is summarized in the following theorem:

Theorem 1 : To achieve DOF $\frac{1}{2}K(K-1)$, when K is even, the number of antennas at the relay and the users should satisfy $N \geq (K-1)^2$ and $KM \geq N + (K-1)$, and when K is odd, the number of antennas should satisfy $N \geq K(K-2)$ and $(K-1)M \geq N+1$. •

In the next section, we will show how to construct the beamforming vectors such that Theorem 1 is achieved. From Theorem 1, it is easy to deduce the following minimum antenna requirement.

Corollary 1 : The minimum number of antennas to achieve DOF $\frac{1}{2}K(K-1)$ is $M = K-1$ and $N = (K-1)^2$ when K is even, and $M = K-1$ and $N = K(K-2)$ when K is odd. •

5.3 Signal Group Based Alignment Algorithm

In this section, a beamforming method is proposed to establish Theorem 1. Firstly, we present the a lemma which leads to the constructions of this beamforming method.

Lemma 1 : Let $\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_l \in \mathbb{C}^{N\times M}$ be l matrices such that their elements are i.i.d. and generated from some continuous random distribution. Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_l \in \mathbb{C}^{M\times 1}$ be l beamforming vectors for signal $d_1, ..., d_l$. Let $\mathcal{S} = \text{span}\{\mathbf{H}_1\mathbf{x}_1, ..., \mathbf{H}_l\mathbf{x}_l\}$. If $\dim\{S\} = l-1$, then a weighted sum of any two scalar signals d_i and d_j can be obtained from $\sum_{k=1}^{l} \mathbf{H}_k\mathbf{x}_k d_k$. •

Proof : Without loss of generality, suppose we want to recover a weighted sum of scalar signals d_1 and d_2. Since $\dim\{\mathcal{S}\} = l-1$, then $N \geq l-1$. Let $\tilde{\mathbf{X}} = [\mathbf{H}_3\mathbf{x}_3, \mathbf{H}_4\mathbf{x}_4, ..., \mathbf{H}_l\mathbf{x}_l] \in \mathbb{C}^{N\times(l-2)}$. Then $\dim\{\tilde{\mathbf{X}}\} \leq l-2$. Therefore, $N > \dim\{\tilde{\mathbf{X}}\}$ and there exists a vector $\mathbf{y}^* \in \text{null}\{\tilde{\mathbf{X}}^H\}$, i.e. $\mathbf{y}^T\tilde{\mathbf{X}} = 0$. This leads to $\mathbf{y}^T \sum_{k=1}^{l} \mathbf{H}_k\mathbf{x}_k d_k = \mathbf{y}^T\mathbf{H}_1\mathbf{x}_1 d_1 + \mathbf{y}^T\mathbf{H}_2\mathbf{x}_2 d_2 = w_1 d_1 + w_2 d_2$. Additionally, since elements of $\mathbf{H}_i$, $0 \leq i \leq l$, are generated independently from some continuous distribution, the probability that $w_1 = 0$ or $w_2 = 0$ is zero. □

5.3.1 User Beamforming in MAC Phase

In this section, the beamforming vectors design for each end user in the MAC phase is presented. Based on Lemma 1, a beamforming method is proposed to align signals from l users into a $l-1$ subspace at the relay, such that the dimensions occupied by these signals is reduced from l to $l-1$ while the network coded signal, i.e. weighted sum of two signals can still be recovered. The beamforming vectors construction for both MAC and BC phase will help prove theorem 1.

Firstly, we consider the case when the number of users K is even. Assume the number of antennas at end users satisfy $M \geq K-1$ so that $K-1$ independent datastreams can be sent simultaneously. The goal is to align K signals from K users (one signal from each user) into a $K-1$ subspace at the relay node. There are $K(K-1)$ signals from all the end users. Therefore, these signals need to be assigned into $K-1$ groups. We propose that the grouping of all $K(K-1)$ signals should satisfy two criteria. Firstly, each group contains K signals from K users and these signals are meant to be sent to K users. Secondly, the signals in each group can form $K/2$ pairs and the two signals within each pair can be network coded together, i.e. each pair contains $s_{i,j}$ and $s_{j,i}$, $i \neq j$. Then at the relay, $K(K-1)$ signals only occupy $(K-1)^2$ dimensional subspace. Since the relay has N antennas, $N \geq (K-1)^2$ should be satisfied so that the relay can accommodate these signals. The relay combining matrix is designed to recover the network coded signals, which will be presented later.

To group signals from all end users into $K-1$ groups, we employ a graph edge-coloring method. The K user Y channel can be represented as a K-complete graph, which is a graph with K vertices and any two vertices are connected by an edge. In this K-complete graph, each vertex represents a user and the edge connecting any two users represents two signals that these two users want to send to each other. The vertices are indexed as $A_1, A_2, ..., A_K$ and the edge are indexed as $e_{12}, e_{13}, ..., e_{1K}, e_{23}, ..., e_{K(K-1)}$ with e_{ij}, $i < j$, being the edge connecting A_i and A_j. Then e_{ij} represents two signals s_{ij} and s_{ji} that will be network coded together at the relay.

When K is even, there exists a $K-1$ proper edge coloring for this K-complete graph, which is also a 1-factorization of this graph. Each 1-factor contains $K/2$ edges and the edges belonging to each 1-factor are colored by the same color. By the definition of 1-factor, the $K/2$ edges in each 1-factor connect all K users, which means the signals in each group are from all K users and are meant to be sent to all K users. Also, the $K/2$ edges are associated with K signals which can be paired into $K/2$ pairs. There are exactly $K-1$ such 1-factors for this graph. Therefore, the signals associated with each 1-factor form a group. There are $K-1$ such groups.

After the grouping of $K(K-1)$ signals is obtained, the beamforming vectors should be designed such that the K signals in each group are aligned into a $K-1$ subspace at the relay. Denote the groups as $G_1, G_2, ..., G_{K-1}$. Let $\pi_i(.)$ denote the pairing method of group G_i, i.e. if signals $s_{j'j}$ and $s_{jj'}$ form a pair in group i, then $\pi_i(j) = j'$ and $\pi_i(j') = j$. Denote the signals in group i as $G_i = \{s_{1\pi_i(1)}, s_{2\pi_i(2)}, ..., s_{K\pi_i(K)}\}$. The design of beamforming vectors for signals in each group i should satisfy

$$\dim\left\{\text{span}\{\mathbf{H}_{1r}\mathbf{v}_{1,\pi_i(1)}, \mathbf{H}_{2r}\mathbf{v}_{2,\pi_i(2)}, ..., \mathbf{H}_{Kr}\mathbf{v}_{K,\pi_i(K)}\}\right\} = K-1, \quad 1 \leq i \leq K-1.$$

This can be achieved by aligning vectors $[\mathbf{v}_{1,\pi_i(1)}^T, \mathbf{v}_{2,\pi_i(2)}^T, ..., \mathbf{v}_{K,\pi_i(K)}^T]^T$ into the null space of matric $[\mathbf{H}_{1r}, \mathbf{H}_{2r}, ..., \mathbf{H}_{Kr}]$, i.e.

$$[\mathbf{v}_{1,\pi_i(1)}^T, \mathbf{v}_{2,\pi_i(2)}^T, ..., \mathbf{v}_{K,\pi_i(K)}^T]^T \in \text{null}\{[\mathbf{H}_{1r}, \mathbf{H}_{2r}, ..., \mathbf{H}_{Kr}]\}.$$

Scalar $\beta_i, i = 1, 2, ..., K$ is chosen such that the power constraint in (5.4) is satisfied with equality. Since we have $\mathbb{E}\{\mathbf{s}_i\mathbf{s}_i^H\} = \sigma_s^2\mathbf{I}$, β_i can be expressed as:

$$\beta_i = \sqrt{\frac{P}{\text{tr}\{\mathbf{V}_i\mathbf{V}_i^H\}\sigma_s^2}}.$$

Since there are $K-1$ groups, the null space of $[\mathbf{H}_{1r}, \mathbf{H}_{2r}, ..., \mathbf{H}_{Kr}]$ should have dimension of at least $K-1$. Also $[\mathbf{H}_{1r}, \mathbf{H}_{2r}, ..., \mathbf{H}_{Kr}]$ is a $N \times MK$ matrix with full rank almost surely. So the number of antennas should satisfy $MK - (K-1) \geq N$. Recall that the number of relay antennas should satisfy $N \geq (K-1)^2$, so $M \geq K-1$ is implicitly satisfied. This is the requirement in Theorem 1.

When K is odd, it is not reasonable to group K signals into one group since these K signals can not be paired (simply because of the fact that K is odd). However, $K-1$ signals can be assigned into one group and $(K-1)/2$ pairs can be possibly obtained. Similarly as the case when K is even, the network can be represented as a K-complete graph with each vertex representing a user and each edge connecting two vertices representing these two signals the two users want to send to each other. For a K-complete graph with K being odd, there exists a K proper edge coloring. Let c_i, $i = 1, 2, ..., K$, be the number of edges colored by color i. We have $c_i \leq (K-1)/2$ since there are K vertices in the graph and K is odd. So from the fact that $\sum_{i=1}^{K} c_i = K(K-1)/2$, we have $c_i = (K-1)/2$, $1 \leq i \leq K$. Therefore, each color colors exactly $(K-1)/2$ edges, which connects $K-1$ vertices. We assign the signals associated with the edges colored by the same color into one group, because the signal in each group can be paired into $(K-1)/2$ pairs. Each group contains signals from/to only $K-1$ users. Since each vertex is incident to $K-1$ edges, these edges are colored with $K-1$ different colors. Therefore, each user's signals belong to $K-1$ groups, which means each group is missing 1 user's signals. Since there are K groups and K users, each group is missing a different user's signals. Let i-th group be the group that is missing i-th user's signals. Then the i-th group can be defined as $G_i = \{s_{1,\pi_i(1)}, s_{2,\pi_i(2)}, ..., s_{i-1,\pi_i(i-1)}, s_{i+1,\pi_i(i+1)}, s_{K,\pi_i(K)}\}$.

After the grouping of signals, user beamforming vectors are designed such that the $K-1$ signals in each group are aligned into a $K-2$ subspace at the relay. Since there are K signal groups in total and the signals in each group are aligned into a $K-2$ dimensional subspace at the relay, the relay needs at least $K(K-2)$ antennas to accommodate all these signals from end users, i.e. $N \geq K(K-2)$. For G_i, the beamforming vectors should satisfy

$$\dim\Big\{\text{span} \quad \{\mathbf{H}_{1r}\mathbf{v}_{1,\pi_i(1)}, \mathbf{H}_{2r}\mathbf{v}_{2,\pi_i(2)}, ..., \mathbf{H}_{i-1,r}\mathbf{v}_{i-1,\pi_i(i-1)}, \mathbf{H}_{i+1,r}\mathbf{v}_{i+1,\pi_i(i+1)}, ..., \mathbf{H}_{Kr}\mathbf{v}_{K,\pi_i(K)}\}\Big\}$$
$$= K-2, \quad 1 \leq i \leq K-1. \tag{5.9}$$

This requires that vectors $[\mathbf{v}_{1,\pi_i(1)}^T, \mathbf{v}_{2,\pi_i(2)}^T, ..., \mathbf{v}_{i-1,\pi_i(i-1)}^T, \mathbf{v}_{i+1,\pi_i(i+1)}^T, ..., \mathbf{v}_{K,\pi_i(K)}^T]^T$ are aligned into the null space of matric $\tilde{\mathbf{H}}_i = [\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_{i-1}, \mathbf{H}_{i+1}, ..., \mathbf{H}_K]$, i.e.

$$[\mathbf{v}_{1,\pi_i(1)}^T, \mathbf{v}_{2,\pi_i(2)}^T, ..., \mathbf{v}_{i-1,\pi_i(i-1)}^T, \mathbf{v}_{i+1,\pi_i(i+1)}^T, ..., \mathbf{v}_{K,\pi_i(K)}^T] \in \text{null}\Big\{\tilde{\mathbf{H}}_i\Big\}, \quad i = 1, 2, ..., K. \tag{5.10}$$

The scalar β_i is chosen as

$$\beta_i = \sqrt{\frac{P}{\text{tr}\{\mathbf{V}_i\mathbf{V}_i^H\}\sigma_s^2}}$$

so that power constraint (5.4) is satisfied with equality. Since all groups are missing different users' signals, $\tilde{\mathbf{H}}_i$, $1 \leq i \leq K$, are all different. So the dimension of the null space of $\tilde{\mathbf{H}}_i, 1 \leq i \leq K$ should be at least one. The number of antennas should satisfy $M(K-1) \geq N+1$. This condition plus the requirement $N \geq K(K-2)$ imply that $M \geq K-1$ is also satisfied. This is the requirement in Theorem 1.

5.3.2 Relay Receive Combining Vectors Design in MAC Phase

Let the relay precoding matrix be $\mathbf{T}_r = \gamma\mathbf{P}\mathbf{U}$ where $\mathbf{U} \in \mathbb{C}^{\frac{K(K-1)}{2}\times N}$ and $\mathbf{P} \in \mathbb{C}^{N\times\frac{K(K-1)}{2}}$ are the receive combining and beamforming matrices, respectively, of the $K(K-1)/2$ signal pairs, and γ is a scalar to satisfy the power constraint at the relay as in (5.5). In this section we consider the design of receive combining matrix $\mathbf{U}$ for the MAC phase. During the MAC phase, relay receives $K(K-1)$ signals and needs to recover $K(K-1)/2$ signal pairs, i.e. $s_{ij} + s_{ji}$, $1 \leq i, j, \leq K$, $i \neq j$. Let $\mathbf{u}_{ij}^T \in \mathbb{C}^{1\times N}, i < j$ be the receive combining vector at the relay for signal pair s_{ij} and s_{ji}. Then $\mathbf{U} = [\mathbf{u}_{12}, ..., \mathbf{u}_{ij}, ..., \mathbf{u}_{K-1,K}]^T$. In order to recover signal pair s_{ij} and s_{ji} without inter-user or inter-message interference, $\mathbf{u}_{ij}^*$ should fall into the null space of the *interference matrix* $\tilde{\mathbf{U}}_{ij}^H$. The columns of $\tilde{\mathbf{U}}_{ij}$ are $\mathbf{H}_{i'r}\mathbf{v}_{i'j'}$ with $\{i', j'\} \neq \{i, j\}$ and $\{i', j'\} \neq \{j, i\}$. So $\tilde{\mathbf{U}}_{ij}$ is a $N \times (K(K-1)-2)$ complex matrix. In the next Lemma, it is proved that such non-zero vectors $\mathbf{u}_{ij}$, $i < j$, $1 \leq i, j \leq K$ exist when beamforming vectors $\mathbf{v}_{i,j}$, $i < j$, $1 \leq i, j \leq K$ are designed as in Sec. 5.3.1.

Lemma 2: If the beamforming vectors $\mathbf{v}_{i,j}$, $i \neq j$, $1 \leq i, j \leq K$ are designed as in Sec. 5.3.1, then for even K, $\text{rank}\{\tilde{\mathbf{U}}_{ij}\} \leq (K-1)^2 - 1$ for all signal pairs i, j, and for odd K, $\text{rank}\{\tilde{\mathbf{U}}_{ij}\} \leq K(K-2) - 1$ for all signal pairs i, j. •

Proof : $\tilde{\mathbf{U}}_{ij}$ is a matrix whose columns are associated with all signals from K users except for s_{ij} and s_{ji}. So the rank of $\tilde{\mathbf{U}}_{ij}$ is the dimensions that all signals other than s_{ij}, s_{ji} occupy at the relay.

When K is even, there are $K-1$ signal groups and in each group, K signals are aligned into a $K-1$ subspace. Since s_{ij} and s_{ji} form a signal pair, they belong to the same group. We name this group as G_t and use $\tilde{G}_t$ to denote the group formed by removing s_{ij} and s_{ji} from G_t. Then there are $K-2$ signals in $\tilde{G}_t$ and they can occupy at most $K-2$ dimensions at the relay. Therefore, the signals in group $\tilde{G}_t$ and signals in all other groups can occupy at most $(K-1)^2-1$ dimensions at the relay.

When K is odd, there are K groups while each group contains $K-1$ signals. Let G_t be the group that signals s_{ij} and s_{ji} belong to and $\tilde{G}_t$ be a group formed by removing s_{ij} and s_{ji} from group G_t. There are $K-3$ signals in group $\tilde{G}_t$ so they can occupy at most $K-3$ dimensions at the relay. Therefore, the signals in $\tilde{G}_t$ and all other groups can occupy at most $K(K-2)-1$ dimensions at the relay. $\square$

When the number of antennas at the relay satisfy $N \geq (K-1)^2$ for even K and $N \geq K(K-2)$ for odd K, from Lemma 2 we always have $\text{rank}\{\tilde{\mathbf{U}}_{ij}\} < N$. So there always exists a non-zero vector $\mathbf{u}_{i,j}$ such that $\mathbf{u}_{ij}^* \in \text{null}\{\tilde{\mathbf{U}}_{ij}^H\}$.

5.3.3 User Receive Combining Design in BC Phase

In the BC phase, relay transmits the signals from all K users using a linear precoding matrix $\mathbf{T}_r$. All users receive signals from the relay and try to recover $K-1$ desired messages sent from all other $K-1$ users. Similar as in the MAC phase, we consider the user receive combining design in two cases: K is even and K is odd.

When user number K is even, the $K(K-1)$ total desired signals are assigned to $K-1$ groups while each group contains K signals from/to K users. Furthermore, the K signals in each group form $K/2$ pairs and the two signals in each pair are the signals two users want to send to each other, i.e. s_{ij} and s_{ji}, $i \neq j$. The grouping of signals is the same as in MAC phase, i.e. signals from all K users are assigned to groups $G_1, G_2, ..., G_{K-1}$ with $G_i = \{s_{1\pi_i(1)}, s_{2\pi_i(2)}, ..., s_{K\pi_i(K)}\}$. The receive combining vector at user i to recover signal s_{ji} is denoted as $\mathbf{r}_{ij}^T \in \mathbb{C}^{1\times M}$. If these receive combining vectors are chosen randomly, in BC phase, the signals in each group will be received along K directions. We propose to design the receive combining vectors such that the signals in each group are received along $K-1$ directions, which leads to

$$\dim\Big\{\text{span}\left\{[\mathbf{H}_{r1}^T\mathbf{r}_{1,\pi_i(1)}, \mathbf{H}_{r2}^T\mathbf{r}_{2,\pi_i(2)}, ..., \mathbf{H}_{rK}^T\mathbf{r}_{K,\pi_i(K)}]\Big\}\right\} = K-1, \;\; 1 \leq i \leq K. \tag{5.11}$$

When the channels in MAC and BC phases have reciprocity, $\mathbf{H}_{rk} = \mathbf{H}_{kr}^T$. In order to make equation (5.11) hold, with $\mathbf{z}_i = [\mathbf{r}_{1\pi_i(1)}^T, \mathbf{r}_{2\pi_i(2)}^T, ..., \mathbf{r}_{K\pi_i(K)}^T] \in \mathbb{C}^{1\times MK}$, vector $\mathbf{z}_i^T$ should fall into the null space of the stack of the corresponding channel matrices as follows:

$$\mathbf{z}_i^T \in \text{null}\left\{[\mathbf{H}_{r1}^T, \mathbf{H}_{r2}^T, ..., \mathbf{H}_{rK}^T]\right\}, \quad 1 \leq i \leq K-1. \tag{5.12}$$

Since there are $K-1$ signal groups, there are $K-1$ stacked receive combining vectors $\mathbf{z}_1^T, ..., \mathbf{z}_{K-1}^T$ that should fall into the null space of matrix $[\mathbf{H}_{r1}^T, \mathbf{H}_{r2}^T, ..., \mathbf{H}_{rK}^T]$. The null space of matrix $[\mathbf{H}_{r1}^T, \mathbf{H}_{r2}^T, ..., \mathbf{H}_{rK}^T]$ should at least have $K-1$ dimension, which puts a constraint on the number of antennas: $KM \geq N+(K-1)$. Considering that $N \geq (K-1)^2$, $M \geq K-1$ is implicitly satisfied.

When K is odd, the $K(K-1)$ signals are assigned to K groups with each group containing $K-1$ signals using the same method as in the user beamforming design for the MAC phase. The i-th group is the group that is missing the signals to and from i-th user, denoted as $G_i = \{s_{1,\pi_i(1)}, s_{2,\pi_i(2)}, ..., s_{i-1,\pi_i(i-1)}, s_{i+1,\pi_i(i+1)}, s_{K,\pi_i(K)}\}$. The design of receive combining vectors allows to receive $K-1$ signals in each group along $K-2$ dimensions. This is achieved by designing $\mathbf{r}_{ij}$, $i \neq j$, $1 \leq i, j, \leq K$ to satisfy the following condition:

$$\begin{aligned}\dim\Big\{\text{span}\Big\{[\mathbf{H}_{r1}^T\mathbf{r}_{1,\pi_i(1)}, ..., \mathbf{H}_{r,i-1}^T\mathbf{r}_{i-1,\pi_i(i-1)}, \mathbf{H}_{r,i+1}^T\mathbf{r}_{i+1,\pi_i(i+1)}, ..., \mathbf{H}_{rK}^T\mathbf{r}_{K,\pi_i(K)}]\Big\}\Big\} \\ = K-2, \quad 1 \leq i \leq K.\end{aligned} \tag{5.13}$$

Let $\mathbf{z}_i = [\mathbf{r}^T_{1,\pi_i(1)}, \mathbf{r}^T_{2,\pi_i(2)}, ..., \mathbf{r}^T_{i-1,\pi_i(i-1)}, \mathbf{r}^T_{i+1,\pi_i(i+1)}, ..., \mathbf{r}^T_{K,\pi_i(K)}] \in \mathbb{C}^{1\times M(K-1)}$. The above condition is meet by letting $\mathbf{z}_i^T$ fall into the null space of its corresponding channel matrix as follows:

$$\mathbf{z}_i^T \in \text{null}\Big\{\text{span}\left\{[\mathbf{H}_{r1}, \mathbf{H}_{r2}, ..., \mathbf{H}_{r,i-1}, \mathbf{H}_{r,i+1}, ..., \mathbf{H}_{rK}]\right\}\Big\}, 1 \leq i \leq K. \tag{5.14}$$

Since for each group, the channel matrix as in the right-side of the above equation is different, it is required that the null space of the channel matrix should have dimension ≥ 1. Therefore, a restriction on the number of antennas for the users and the relay is $M(K-1) \geq N+1$. Considering $N \geq K(K-2)$ for odd K, $M \geq K-1$ is implicitly satisfied.

5.3.4 Relay Beamforming Design in BC Phase

During the MAC phase, the relay recovers $K(K-1)/2$ network coded signals. In the BC phase, they are sent to K end users. Let $\mathbf{P} = [\mathbf{p}_{12}, ..., \mathbf{p}_{ij}, ..., \mathbf{p}_{K-1,K}]$, where $\mathbf{p}_{ij} \in \mathbb{C}^{N\times 1}$, $i < j$ is the beamforming vector for signal pair s_{ij} and s_{ji}. This signal pair should be sent to user i and user j and $\mathbf{r}_{ij}$ and $\mathbf{r}_{ji}$ are the associated receive combining vectors at the desired receivers. To avoid inter-user and inter-message interference, the beamforming vectors at the relay should be designed such that each signal pair can only be recovered by desired receive combining vectors at desired users. Define the *interference matrix* for i, j signal pair as $\tilde{\mathbf{V}}_{ij}, i < j$. The rows of $\tilde{\mathbf{V}}_{ij}$ are $\mathbf{r}^T_{i'j'}\mathbf{H}_{ri'}$ with $\{i', j'\} \neq \{i, j\}$, $\{i', j'\} \neq \{j, i\}$. Therefore, $\tilde{\mathbf{V}}_{ij}$ is a $(K(K-1)-2) \times N$ complex matrix.

Lemma 3 : When K is even, $\text{rank}\{\tilde{\mathbf{V}}_{ij}\} \leq (K-1)^2 - 1$ for all signal pairs i, j. When K is odd, $\text{rank}\{\tilde{\mathbf{V}}_{ij}\} \leq K(K-2) - 1$ for all signal pairs i, j. •

The proof of Lemma 3 is similar to Lemma 2. Let G_t be the signal group that contains signals s_{ij} and s_{ji}, and $\tilde{G}_t$ is the group formed by removing s_{ij} and s_{ji} from G_t. When K is even, the *receive-channel* vectors $\mathbf{r}^T_{i'j'}\mathbf{H}_{ri'}$ associated with signals in group $\tilde{G}_t$ span at most $K-2$ dimensions. Since the *receive-channel* vectors associated with signals in each group will occupy $K-1$ dimensions, the rank of $\tilde{\mathbf{V}}_{ij}$ is at most $(K-1)^2-1$. When K is odd, the *receive-channel* vectors $\mathbf{r}_{i'j'}\bar{\mathbf{H}}_{i'}$ associated with signals in group $\tilde{G}_t$ span at most $K-3$ dimensions, and the *receive-channel* vectors associated with signals in each other group occupy $K-2$ dimensions. Therefore, the rank of $\tilde{\mathbf{V}}_{ij}$ is at most $K(K-2)-1$.

The design of beamforming vector $\mathbf{p}_{ij}$, $i<j$, $1 \le i,j \le K$ should make sure that it falls into the null space of $\tilde{\mathbf{V}}_{ij}$.

$$\mathbf{p}_{ij} \in \text{null}\left\{\tilde{\mathbf{V}}_{ij}\right\}, i<j, 1 \le i,j, \le K.$$

In order for the non-zero vector $\mathbf{p}_{ij}$ to exist, we need $N > \text{rank}\{\tilde{\mathbf{V}}_{ij}\}$. Then the number of antennas at the relay should satisfy $N \ge (K-1)^2$ when K is even, and $N \ge K(K-2)$ when K is odd. This is the same requirement as in designing the receive combining matrix $\mathbf{U}$.

Finally, the scalar γ is chosen as

$$\gamma = \sqrt{\frac{P}{\text{tr}\left\{\sum_{i=1}^{K}\mathbf{PUH}_{ir}\mathbf{H}_{ir}^{H}\mathbf{U}^{H}\mathbf{P}^{H}\sigma_s^2 + \mathbf{PUU}^{H}\mathbf{P}^{H}\sigma_n^2\right\}}}$$

to satisfy the relay power constraint (5.5) with equality. So far, a beamforming method based on signal group alignment has been proposed. This method completes the transmission of $K(K-1)$ datastreams in two time slots and requires the numbers of antennas to satisfy $MK \ge N+K-1$ and $N \ge (K-1)^2$ for even K and to satisfy $(K-1)M \ge N+1$ and $N \ge K(K-2)$ for odd K. Thus, Theorem 1 is proved.

Remark 1: In the proposed algorithm, $\eta(K) = K(K-1)/2$ is achieved for this K user MIMO Y channel in a half-duplex mode. The averaged DOF is given by

$$\bar{\eta}(K) = \begin{cases} \frac{K(K-1)/2}{(MK+N)/(K+1)} \le \frac{K^2+K}{4K-2} & \text{for even } K \\ \frac{K(K-1)/2}{(MK+K(K-2))/(K+1)} \le \frac{K^2-1}{4K-6} & \text{for odd } K \end{cases} \tag{5.15}$$

K	Sum DOF	M	N	total number of antenna
3	3	2	3	9
4	6	9	3	21
5	10	15	4	35
6	15	25	5	55

Table 5.1: Required minimum number of antennas for the signal group based alignment algorithm

We can see that the averaged DOF increases in the order of $K/4$. The table 5.1 shows the minimum required antenna numbers as well as the achieved DOF of the proposed signal group alignment.

5.4 Simulation Results

In this section, we show that our proposed signal group based alignment algorithm can achieve the DOF $K(K-1)/2$ when the antenna number conditions are met as stated in Theorem 1. In Fig. 5.2, the sum rate of all K users' datastreams are evaluated versus SNR. Two system set-ups are considered to show the effectiveness of the proposed beamforming algorithm for both odd and even number of users. In Fig. 5.2, K is the number of users, M is the number of user antennas and N is the number of relay antennas. All K users and the relay transmit with the same power. With all channel coefficients assumed to be i.i.d. Gaussian random variable with zero mean and unit variance, SNR is defined as

$$SNR = \log_2 \left(\frac{P}{\sigma_n^2} \right).$$

It is seen from Fig. 2 that as the SNR increases, the slope of the $K = 4$ curve is 6 while the slope of the $K = 5$ curve is 10. Both slopes equal $K(K-1)/2$, thus verifying our theoretical results.

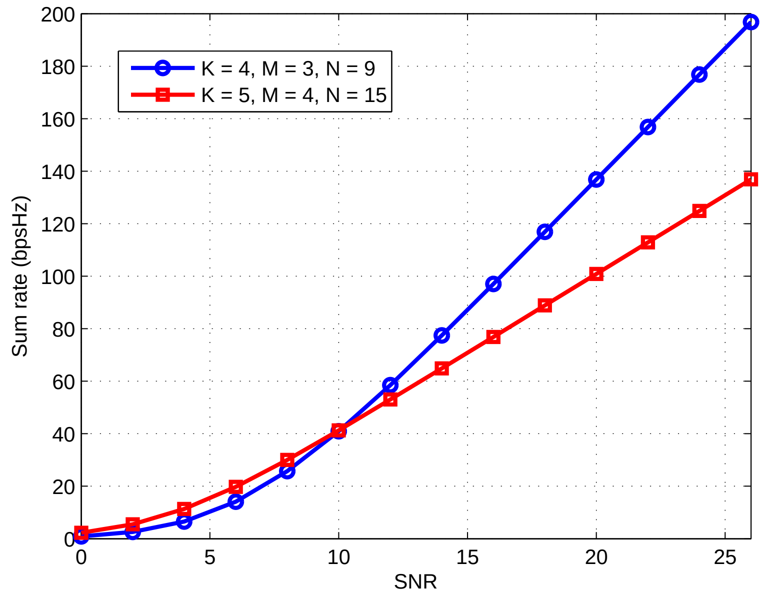

Figure 5.2: Sum rate vs SNR $\log_2\left(\frac{P}{\sigma_n^2}\right)$

5.5 Conclusion

In this chapter, we proposed a signal group based alignment method for a generalized K user MIMO Y channel in which each user has $K-1$ independent datastreams for the other $K-1$ users. This multi-way transmission is facilitated by a relay station and operate in a two time-slot half-duplex mode. The $K(K-1)$ datastreams are assigned to g groups ($g = K$ when K is odd and $g = K-1$ when K is even) using a graph theory method. Let l denote the number of signals in each group. The beamforming design is to align these signals in each group in each group into a $l-1$ dimensional subspace at the relay node during the MAC phase and let the receive combing vectors associated with these signals receive along a $l-1$ dimensional subspace. To achieve $K(K-1)/2$ DOF, minimum number of antennas (M at each user and N at the relay) are specified in Theorem 1. In our proposed approach we significantly decrease the minimum M at the expense of higher N for a given number of users K., compared to an existing approach. For instance, for $K = 8$ and the same of number of antennas at each user, [62] would require $N \geq K(K-1)/2 = 28$, and $M \geq K-1 = 7$ as well as $2M > N \geq 28$, leading to $M > 14$ or $M \geq 15$. On the other hand, by our Corollary 1, we require $M \geq K - 1 = 7$ and $N \geq (K-1)^2 = 49$.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this dissertation, two kind of problems were addressed: 1) resource allocation in cognitive radios, 2) precoding design in cognitive radio and relay networks. They are two effective ways to improve spectrum efficiency and to allow coexistence of multiple mobile devices which are share the same radio resources.

A joint spectrum sensing, access and power allocation algorithm was proposed in Chapter 2 in a multi-channel cognitive radio environment. By introducing the soft-decision spectrum sensing concept into the optimization problem and utilizing channel availability probabilities instead of binary channel state decisions, it was shown that the system performance in terms of sum throughput is significantly improved while interference to the primary network is kept under a specified threshold. The optimization problem was transformed into a convex optimization problem and optimal solution was obtained. To reduce the computational complexity, two heuristic algorithms were also proposed which solve for the access strategy first and then allocate power. Simulation results showed these soft-decision spectrum sensing based algorithms outperform hard-decision counterpart and one of the heuristic algorithm achieves a near optimal performance with a much smaller complexity.

In Chapter 3, a multi-user multi-way relay network acting as a secondary network in a cognitive radio environment was considered. With the deployment of a relay node and multiple antennas equipped at both users and relays, the multi-way transmission was made possible via proper predcoding/decoding design at users and relays while interference to the primary users was completely or partially eliminated. A two time slot half-duplex transmission scheme was adopted in this network. An iterative algorithm based on the MMSE criterion was first proposed which optimized the precoding matrices at transmit users, precoding matrix at the relay and decoding matrices design at receive users iteratively. It was shown all the three sub-problems are convex and optimal solutions to the sub-problems were presented. Then a non-iterative algorithm was also proposed to provide a low complexity solution. This algorithm optimized the user precoding and decoding matrices based on a matrix distance criterion and the relay precoding was designed to minimize sum MSE. Simulation results showed that this non-iterative algorithm with reduced complexity only results in a small performance degradation compared with the iterative algorithm. Furthermore, considering the face that perfect CSI is sometimes hard to obtain due to the time-varying nature of wireless channel, a robust precoding method was proposed based on the non-iterative algorithm. In this robust algorithm, the channel uncertainty was modeled into some stochastic error which follows a zero mean normal distribution. Simulation results showed that this robust algorithm can keep the interference to primary network under control and improve the performance of secondary network compared with the non-robust algorithm.

In Chapter 4, a joint signal and interference alignment precoding was proposed for a multi-pair two-way relay network as a secondary network in a cognitive radio system. We considered a system in which multiple pairs of users wish to transmit multiple datastreams to each other and a relay station is deployed to facilitate the multi-pair two-way transmission. In this system model, three types of interference are present: inter-user interference, inter-datastream interference and primary-secondary network interference. To avoid the harmful effects of the interference, precoding and decoding matrices at users and relay are designed to avoid or completely remove all the interference. A matrix distance based algorithm was proposed to jointly consider the desired signal strength and avoidance of interference. A non-negative scalar weight was used to balance the signal alignment and interference alignment. It was shown that the zero forcing design is a special case of this proposed algorithm. Also, by appropriately changing the value of the scalar weight, this algorithm can adapt to both high SNR and low SNR regimes.

A generalized Y channel consisting of $K, K > 3$, users and each user having $K-1$ messages to be sent to all other $K-1$ users was considered in Chapter 5. A relay node was used to enable this multi-way transmission of multiple users. Based on the network coding concept, a signal group based alignment algorithm was proposed which aligns the signals from all users into a smaller subspace at the relay such that fewer antennas are required for the relay as well as the users. The proposed signal group alignment is suitable for a system in which it is more practical to equip the relay node with more antennas and the end users are small or mobile devices in which only a small number of antennas can be installed.

6.2 Future Work

The research work reported in this dissertation about resource allocation and precoding design in cognitive radio and relay networks also suggests the following future research ideas.

Robust precoding design

In most of this dissertation we assumed perfect CSI, which may be difficult to obtain in practical systems due to time varying nature of the wireless channel and the overhead of transmitting CSI back to the transmitter or central control node. A promising research topic is to develop algorithms that do not require the knowledge of CSI or are insensitive to the CSI errors.

Precoding/beamforming design with security concerns

The precoding and beamforming algorithms proposed in this dissertation can dramatically improve the system performance by mitigating the interference effects. The beamforming and receive combining design can also be an effective way to provide secure data transmission from a physical layer point of view. The broadcast nature of wireless signals makes the security issue a major concern in wireless network design. Therefore incorporating security concerns into the precoding and beamforming design will be a promising research topic.

Bibliography

[1] C. Leow, Z. Ding, K. Leung and D. Goeckel "On the Study of Analogue Network Coding for Multi-Pair, Bidirectional Relay Channels," *IEEE Trans. Wireless Comm.*, vol. 10, pp. 670-681, Feb. 2011.

[2] ITU-R SM. 2152, "Definitions of Software Defined Radio (SDR) and cognitive Radio System (CRS)", Sept. 2009.

[3] D. Cabric, S. M. Mishra and R. W. Brodersen, "Implementation Issues in Spectrum Sensing for Cognitive Radios," in *Proc. 38th Asilomar Conf. Signals, Systems, Computers,* Pacific Groe, CA, USA, Nov. 2004.

[4] D. Cabric, A. Tkachenko and R. W. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation," *Proceeding TAPAS '06 Proceedings of the first international workshop on Technology and policy for accessing spectrum,* Boston, MA, USA, Aug. 2006.

[5] T. Yucek, H. Arslan, "A survey of spectrum sensing algorithms for cogntivive radio applications," *Communications Surveys & Tutorials, IEEE,* vol. 11, pp. 116-130, First Quater, 2009.

[6] P. D. Sutton, K. E. Nolan and L. E. Doyle, "Cyclostationary signatures in practical cognitive radio applications," in *Communications Surveys & Tutorials, IEEE,* vol. 26, pp. 13-24, Jan, 2008.

[7] K. Lee, N. Lee and I. Lee, "Feasibility Conditions of Signal Space Alignment for Network Coding on K-user MIMO Y channels," in *Proc. 2011 IEEE IEEE International Conference on Communications (ICC 2011)*, Kyoto, Japan, June 2011.

[8] S. Jafar, "Interference Alignment: A New Look at Signal Dimensions in a Communication Network", *Foundations and Trends in Communications and Information Theory,* Vol. 7, No. 1, pages: 1-136.

[9] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment, " *IEEE J. Sel. Areas Commun.* vol. SAC-5, no. 6, pp. 871-878, Jun. 1987.

[10] G. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas, " *Bell Labs Tech. J.* pp. 4159, 1996.

[11] I. Telatar, "Capacity of multi-antenna Gaussian channels" *AT&T Bell Labs. Tech. Rep.* BL0112170-950615-07TM, Jun. 1995

[12] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversitypart I: System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 19271938, Nov. 2003.

[13] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part II: Implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 19391948, Nov. 2003.

[14] H. Mu and J. Tugnait, "Interference Alignment-Like Precoder Design in Multi-Pair Two-Way Relay Cognitive Radio Networks", *Proc. 2012 IEEE Global Telecom. Conf. (GLOBECOM)*, Anaheim, CA, Dec. 2012.

[15] D. Hu and S. Mao "Cooperative relay in cognitive radio networks: Decode-and-forward or amplify-and-forward?", *Proc. 2010 IEEE Global Telecom. Conf. (GLOBECOM)*, Miami, FL, Dec. 2010.

[16] V. Asghari and S. Aissa, "Adaptive rate and power transmission in spectrum-sensing system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3272-3280, Oct. 2010.

[17] S. Barbarossa, S. Sardellitti and G. Scutari, "Joint optimization of detection thresholds and power allocation for opportunistic access in multicarrier cognitive radio networks," in *Proc. 2009 3rd IEEE Intern. Workshop Comp. Adv. Multi-sensor Adaptive Proc. (CAMSAP)*, Aruba, Dutch Antilles, Dec. 2009.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[19] A. El-Sherif, A. Sultan and K. Seddik, "Soft sensing-based multiple access for cognitive radio networks," in *Proc. 2010 IEEE Global Telecom. Conf. (GLOBECOM)*, Miami, FL, Dec. 2010.

[20] R. Fantacci and A. Tani, "Performance evaluation of a spectrum-sensing technique for cognitive radio applications in B-VHF communication systems," *IEEE Trans. Vehicular Tech.*, vol. 58, no. 4, pp. 1722-1730, May 2009.

[21] J. Huang, V.G. Subramanian, R. Agrawal and R.A. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 288-296, Jan. 2009.

[22] X. Huang and B. Beferull-Lozano, "Joint optimization of detection and power allocation for OFDM-based cognitive radios," in *Proc. 2010 IEEE Global Telecom. Conf. (GLOBECOM)*, Miami, FL, Dec. 2010.

[23] J. Ma, G.Y. Li and B.H. Juang, "Signal processing in cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 805-823, May 2009.

[24] H. Mu and J.K. Tugnait, "Soft spectrum sensing and power adaptation in multiband cognitive radios," in *Proc. 2011 IEEE Global Telecom. Conf. (GLOBECOM)*, Houston, TX, Dec. 5-9, 2011.

[25] H. Mu and J.K. Tugnait, "Joint soft-decision cooperative spectrum sensing and power control in multiband cognitive radios," in *Proc. IEEE 2012 Inter Conf. Commun. (ICC)*, Ottawa, Canada, June 10-15, 2012.

[26] Z. Quan, S. Cui, A.H. Sayed and H.V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Proc.*, vol. 57, no. 3, pp. 1128-1139, March 2009.

[27] S. Srinivasa and S.A. Jafar, "Soft sensing and optimal power control for cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3638-3645, Dec. 2010.

[28] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication," *Cambridge University Press*, 2005.

[29] X. Wang, "Joint sensing-channel selection and power control for cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 958-967, March 2011.

[30] Q. Zhao and B.M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Proc. Mag.*, vol. 24, no. 3, pp. 79-89, May 2007.

[31] F. Fitzek and M. Katz (Editors). "Cognitive Wireless Networks", *Springer*, Dordrecht, The Netherlands, 2007.

[32] E. Hossain, D. Niyato and Z. Han. " Dynamic Spectrum Access and Management in Cognitive Radio Networks", *Cambridge Univ. Press*, Cambridge, UK, 2009.

[33] Federal Communications Commission, "Notice of proposed rule making and order: Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies," ET Docket No. 03-108, Feb. 2005.

[34] S. Peters and R. Heath, Jr, "Interference alignment via alternating minimization," in *Proc. 2009 IEEE Intern. Conf. Acoustics, Speech, Signal Proc. (ICASSP 2009)*, Taipei, Taiwan, Apr. 2009.

[35] K. Kumar and F. Xue, "An iterative algorithm for joint signal and interference alignment," in *Proc. 2010 IEEE Intern. Symp. Info. Theory (ISIT 2010)*, Austin, TX, USA, Jun. 2010.

[36] T. Bogale and L. Vandendorpe, "Robust sum MSE optimization for downlink multiuser MIMO systems with arbitrary power constraint: generalized duality approach," *IEEE Trans. Signal Proc.*, vol. 60, pp. 1862-1875, April 2012.

[37] T. Bogale and L. Vandendorpe, "Weighted sum rate optimization for downlink multiuser MIMO coordinated base station systems: Centralized and distributed algorithms," *IEEE Trans. Signal Proc.*, vol. 60, pp. 1876-1889, April 2012.

[38] K. Cumanan, J. Tang and S. Lambotharan, "Rate balancing based linear transceiver design for multiuser MIMO system with multiple linear transmit covariance constraints," in *Proc. 2011 IEEE Intern. Conf. Commun. (ICC2011)*, Kyoto, Japan, June 2011.

[39] K. Lee, H. Sung and I. Lee, "Linear precoder designs for cognitive radio multiuser MIMO downlink systems," in *Proc. 2011 IEEE Intern. Conf. Commun (ICC 2011)*, Kyoto, Japan, June 2011.

[40] R. Zhang, Y. Liang, C. Chai and S. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 699-712, Jun. 2009.

[41] R. Wang and M. Tao, "Joint source and relay precoding designs for MIMO two-way relaying based on MSE criterion," *IEEE Trans. Signal Proc.*, vol. 60, pp. 1352-1365, Mar. 2012.

[42] K. Lee, H. Sung, E. Park and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3671-3681, Dec. 2010.

[43] C. Li, L. Yang and W. Zhu, "Two-way MIMO relay precoder design with channel state information," *IEEE Trans. Commun.*, vol. 58, pp. 3358-3363, Dec. 2010.

[44] J. Joung and A.H. Sayed, "Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems," *IEEE Trans. Signal Proc.*, vol. 58, pp. 1833-1846, Mar. 2010.

[45] Y. Fu, L. Yang, W. Zhu and C. Liu, "Optimum linear design of two-hop MIMO relay networks with QoS requirements," *IEEE Trans. Signal Proc.*, vol. 59, pp. 5473-5484, May 2011.

[46] P. Ubaidulla and A. Chockalingam, "Relay precoder optimization in MIMO-relay networks with imperfect CSI," *IEEE Trans. Signal Proc.*, vol. 59, pp. 5473-5484, Nov. 2011.

[47] J. Zou, W. Liu, M. Ding, H. Luo and H. Yu, "Transceiver design for AF MIMO two-way relay systems with imperfect channel estimation," in *Proc. 2011 IEEE Global Telecom. Conf. (GLOBECOM)*, Houston, TX, Dec. 2011.

[48] H. Mu and J.K. Tugnait, "MSE-based source and relay precoder design for cognitive radio multiuser two-way relay systems," in *Proc. 2012 IEEE Wireless Commun. Networking Conf. (WCNC 2012)*, pp. 742-747, Paris, France, April 1-4, 2012.

[49] A. Amah and A. Klein, "Non-regenerative multi-antenna multi-group multi-way relaying," *EURASIP J. Wireless Commun. Networking*, 2011, 2011:29.

[50] D. Gunduz, A. Yener, A. Goldsmith and H.V. Poor,, "Multi-way relay channel," in *Proc. 2011 IEEE Intern. Symp. Info. Theory*, Seoul, Korea, 2009.

[51] W. Long, T. Lv, H. Gao and Y. Lu, "Interference alignment for multi-user multi-way relaying X networks," in *Proc. 2012 IEEE 75th Veh. Tech. Conf. (VTC Spring)*, Yokohama, Japan, May 6-9, 2012.

[52] B. Rankov and A. Wittneben "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. on Selected Areas in Comm.* , Vol. 25, pp. 379-389, Feb. 2007.

[53] E. Gharavol, Y. Liang and K. Mouthaan, "Collaborative Nonlinear Transceiver Optimization in Multi-tier MIMO Cognitive Radio Networks with Deterministically Imperfect CSI," in *Proc. 2010 IEEE Global Telecom. Conf. (GLOBECOM)*, Miami, FL, Dec. 2010.

[54] X. Gong, A. Ishaque, G. Dartmann and G. Ascheid, "MSE-based Stochastic Transceiver Optimization in Downlink Cognitive Radio Networks," in *Proc. 2011 IEEE Wireless Comm. and Networking Conf. (WCNC)*, Quintana-roo, Mexico, Mar. 2011.

[55] C. Wang, H. Chen, Q. Yin, A. Feng and A. Molisch,"Multi-User Two-Way Relay Networks with Distributed Beamforming," *IEEE Trans. Wireless Comm.*, vol. 10, pp. 3460-3470, Oct. 2011.

[56] A. Amah and A. Klein, "Pair-Aware Transceive Beamforming for Non-Regenative Multi-User Two-Way Relaying," in *Proc. 2010 IEEE Intern. Conf. Acoustics, Speech, Signal Proc. (ICASSP 2010)*, Dallas, TX, Mar. 2010.

[57] H. Xin, Y. Peng, C. Wang, Y. Yang and W. Wang, "Coordinated Eigen Beamforming for Multi-Pair MIMO Two-Way Relay Network," in *Proc. 2011 IEEE Global Telecom. Conf. (GLOBECOM2011)*, Houston, TX, Dec. 2011.

[58] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Nenwork Information Flow, " *IEEE Trans. Information Theory*, Vol. 46, pp 1204-1216, July, 2000.

[59] H. Du and T. Ratnarajah, "Robust joint signal and interference alignment for MIMO cognitive radio network," in *Proc. 2012 IEEE Wireless Commun. Netw. Conf. (WCNC 2012)*, Paris, France, April 2012.

[60] Z. Zhou and B. Vucetic, "An iterative beamforming optimization algorithm for generalized MIMO Y channels," in *Proc. 2012 IEEE Intern. Conf. Commun (ICC 2012)*, Ottawa, Canada, Jun. 10-15, 2012.

[61] N. Lee and J. Chun, "Signal space alignment for an encryption message and successive network code decoding on the MIMO K-way relay channel," in *Proc. 2011 IEEE Intern. Conf. Commun (ICC 2011)*, Kyoto, Japan, Jun. 5-9, 2012.

[62] K. Lee, N. Lee, and I. Lee, "Achievable degrees of freedom on K-user Y channel," *IEEE Trans. Wireless Comm.*, pp. 1210-1218, March 2012.

[63] N. Lee, J. Lim and J. Chun, "Degrees of freedom on the MIMO Y channel: Signal space alignment for network coding," *IEEE Trans. Information Theory*, vol. 56, pp. 3332-3342, July 2010.

[64] N. Lee and J. Lim, "A novel signaling for communication on MIMO Y Channel: Signal space alignment for network coding" in *Proc. IEEE Intern. Symp. Info. Thy. (ISIT 09)*, Seoul, Korea, June 28 - July 3, 2009.

[65] J. Zhuang, F. Roemer and M. Haardt, "Relay assisted physical resource sharing: projection based separation of multiple operators (ProBaSeMO) for two-way relaying with MIMO amplify and forward relays," *IEEE Trans. Signal Processing*, vol. 60, pp. 4834-4836, Sept. 2012.

Appendices

Appendix A

Convexity of (2.5)

Here we will provide a proof of our claim that the objective function (2.5) in problem P1 is a concave function of variables $\mathbf{t}, \mathbf{s}$. Since a positive weighted sum of concave functions is concave, the desired result follows if we can show that $R_0(s,t)$ defined in (A.1) is a concave function of scalar variables s and t ($0 \leq s \leq 1$ and $t \geq 0$: compare with (4.17)) where

$$R_0(s,t) = as \ln\left(1 + \frac{bt}{s}\right) + cs \ln\left(1 + \frac{dt}{s}\right) \tag{A.1}$$

and a, b, c and d are nonnegative real scalars. A necessary and sufficient condition for $R_0(s,t)$ to be concave is that its Hessian matrix is negative semi-definite. We have the following:

$$\frac{\partial R_0(s,t)}{\partial s} = a \ln\left(1 + \frac{bt}{s}\right) + c \ln\left(1 + \frac{dt}{s}\right) - \frac{abt}{s+bt} - \frac{cdt}{s+dt}, \tag{A.2}$$

$$\frac{\partial R_0(s,t)}{\partial t} = \frac{abs}{s+bt} + \frac{cds}{s+dt}, \tag{A.3}$$

$$\frac{\partial^2 R_0(s,t)}{\partial^2 s} = -\frac{ab^2t^2/s}{(s+bt)^2} - \frac{cd^2t^2/s}{(s+dt)^2}, \tag{A.4}$$

$$\frac{\partial^2 R_0(s,t)}{\partial^2 t} = -\frac{ab^2 s}{(s+bt)^2} - \frac{cd^2 s}{(s+dt)^2} \tag{A.5}$$

and

$$\frac{\partial^2 R_0(s,t)}{\partial t\, \partial s} = \frac{ab^2 t}{(s+bt)^2} + \frac{cd^2 t}{(s+dt)^2}. \tag{A.6}$$

Thus the Hessian of $R_0(s,t)$ is given by

$$\nabla^2 R_0(s,t) = \begin{bmatrix} -\frac{t^2}{s}\left(\frac{ab^2}{(s+bt)^2} + \frac{cd^2}{(s+dt)^2}\right) & t\left(\frac{ab^2}{(s+bt)^2} + \frac{cd^2}{(s+dt)^2}\right) \\ t\left(\frac{ab^2}{(s+bt)^2} + \frac{cd^2}{(s+dt)^2}\right) & -s\left(\frac{ab^2}{(s+bt)^2} + \frac{cd^2}{(s+dt)^2}\right) \end{bmatrix}. \tag{A.7}$$

For any real scalars x_1 and x_2, we have

$$[x_1 \;\; x_2]\nabla^2 R_0(s,t)[x_1 \;\; x_2]^T = -\left(\frac{ab^2}{(s+bt)^2} + \frac{cd^2}{(s+dt)^2}\right)\left(\frac{t}{\sqrt{s}}x_1 - \sqrt{s}x_2\right)^2 \leq 0, \qquad \text{(A.8)}$$

and therefore, $R_0(s,t)$ is concave.

Appendix B

Proof of Proposition 2.1

Here we provide a proof of Proposition 2.1. Firstly β_k can be rewritten as

$$\beta_k = P(H_{0k}|\mathbf{Q}_k)W_k \left\{[\log(z)^+ - (1 - \frac{1}{z})^+]\right\}$$

where

$$z = \frac{P(H_{0k}|\mathbf{Q}_k)W_k\alpha_k}{\mu}.$$

It turns out that $[\log(z)]^+ - (1 - \frac{1}{z}) \geq 0$ for any $z > 0$ so that $\beta_k \geq 0$ (hence $\tilde{d}_k \geq 0$) $\forall k$. Define

$$f(\lambda) = \sum_{k=1}^{K} \frac{\tilde{a}_k}{\varepsilon} \left(\tilde{d}_k - \lambda\right)^+ + \lambda. \tag{B.1}$$

Then

$$\min_{\lambda} g(\lambda, \mu) \Leftrightarrow \min_{\lambda} f(\lambda). \tag{B.2}$$

The derivative of $f(\lambda)$ with respect to λ is given by

$$\frac{df(\lambda)}{d\lambda} = 1 - \sum_{k=1}^{i} \frac{\tilde{a}_k}{\varepsilon} \quad \text{if} \quad \tilde{d}_{i+1} \leq \lambda \leq \tilde{d}_i. \tag{B.3}$$

Since $\tilde{a}_k > 0$ $\forall k$, $1 - \sum_{k=1}^{l} \frac{\tilde{a}_k}{\varepsilon}$ is decreasing in l; we take $1 - \sum_{k=1}^{0} \frac{\tilde{a}_k}{\varepsilon} := 1$. We have

$$\frac{df(\lambda)}{d\lambda}\Big|_{\lambda=0} = 1 - \frac{\sum_{k=1}^{K} \tilde{a}_k}{\varepsilon} \tag{B.4}$$

$$\frac{df(\lambda)}{d\lambda}\Big|_{\lambda=\infty} = 1. \tag{B.5}$$

If $1 - \sum_{k=1}^{K} \frac{\tilde{a}_k}{\varepsilon} > 0$, then we have $\frac{df(\lambda)}{d\lambda} > 0$ for $\lambda \geq 0$, so $f(\lambda)$ is an increasing function of λ. Minimizing $f(\lambda)$ with respect to λ gives us the optimal solution $\lambda^* = 0$. If $1 - \sum_{k=1}^{K} \frac{\tilde{a}_k}{\varepsilon} = 0$, we will have $\frac{df(\lambda)}{d\lambda} = 0$ for some $0 < \lambda \leq \tilde{d}_{min} = \arg\min_{1 \leq k \leq K} \{\tilde{d}_k \mid \tilde{d}_k > 0\}$ and $\frac{df(\lambda)}{d\lambda} > 0$ for $\lambda > \tilde{d}_{min}$, so $\lambda^* = 0$ is still the optimal solution.

If $1 - \sum_{k=1}^{K} \frac{\tilde{a}_k}{\varepsilon} < 0$, then we have $\left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=0} < 0$ and $\left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=\infty} > 0$. Then the optimal solution λ^* should be such that either $\left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=\lambda^*} = 0$, or $\left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=\lambda^*-\delta} < 0$ and $\left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=\lambda^*+\delta} > 0$ for some $\delta > 0$. This leads to $\lambda^* = \tilde{d}_{l^*}$ if $1 - \sum_{k=1}^{l^*-1} \frac{\tilde{a}_k}{\varepsilon} \geq 0$ and $1 - \sum_{k=1}^{l^*} \frac{\tilde{a}_k}{\varepsilon} < 0$.

Appendix C

Proof of (3.19)

Here we derive (3.19). For the i-th user, the objective function of problem P2 becomes

$$\mathbb{E} \parallel \hat{\mathbf{s}}_i - \mathbf{E}_i\mathbf{s} \parallel_2^2 = \mathbb{E} \parallel (\mathbf{R}_i\mathbf{A}_i - \mathbf{E}_i)\,\mathbf{s} + \mathbf{R}_i\mathbf{B}_i\check{\mathbf{x}}_p + \mathbf{R}_i\mathbf{C}_i\check{\mathbf{n}}_i \parallel_2^2$$
$$= \mathrm{tr}\Big\{\mathbf{R}_i[\mathbf{A}_i\mathbf{A}_i^\dagger\sigma_s^2 + \mathbf{B}_i\mathbf{B}_i^\dagger\sigma_p^2 + \mathbf{C}_i\mathbf{C}_i^\dagger\sigma_n^2]\mathbf{R}_i^\dagger - \mathbf{E}_i\mathbf{A}_i^\dagger\mathbf{R}_i^\dagger\sigma_s^2 - \mathbf{R}_i\mathbf{A}_i\mathbf{E}_i^\dagger\sigma_s^2\Big\}. \quad \text{(C.1)}$$

Since $\mathbf{A}_i\mathbf{A}_i^\dagger\sigma_s^2+\mathbf{B}_i\mathbf{B}_i^\dagger\sigma_p^2+\mathbf{C}_i\mathbf{C}_i^\dagger\sigma_n^2$ is a positive semi-definite matrix, (C.1) is a convex function in $\mathbf{R}_i$. Take derivative of (C.1) with respect to $\mathbf{R}_i^\dagger$ and set it to zero to obtain the optimal decoding matrix of i-th secondary user as $\mathbf{R}_i^* = \sigma_s^2\mathbf{E}_i\mathbf{A}_i^\dagger\left(\sigma_s^2\mathbf{A}_i\mathbf{A}_i^\dagger + \sigma_p^2\mathbf{B}_i\mathbf{B}_i^\dagger + \sigma_n^2\mathbf{C}_i\mathbf{C}_i^\dagger\right)^{-1}$.

Appendix D

Proof: Problem P3 in Chapter 3 is convex

The objective function of P3 can be written as

$$\sum_{i=1}^{K} \mathrm{tr}\left\{\mathbf{P}_i^\dagger \mathbf{D}_i^\dagger \mathbf{D}_i \mathbf{P}_i - \mathbf{E}_i^\dagger \mathbf{D}_i \mathbf{P}_i - \mathbf{P}_i^\dagger \mathbf{D}_i^\dagger \mathbf{E}_i + \mathbf{E}_i \mathbf{E}_i^\dagger\right\} \sigma_s^2.$$

It it obvious that $\mathbf{D}_i^\dagger \mathbf{D}_i$ is a positive semi-definite matrix. Therefore the objective function of P3 is convex in $\mathbf{P}_i$. Using the fact that $\mathbf{tr}\{\mathbf{AB}\} = \mathbf{tr}\{\mathbf{BA}\}$, it is easy to show that the left side of the power constraint (3.23) of P3 is also convex in $\mathbf{P}_i$. For the power constraint (3.24), it can be shown that $\mathbb{E}\mathrm{tr}\{\mathbf{x}_r \mathbf{x}_r^\dagger\} = \mathrm{tr}\{\mathbf{P}_i^\dagger \mathbf{H}_{ip}^{\perp\dagger} \mathbf{H}_{ir}^\dagger \mathbf{T}_r^\dagger \mathbf{T}_r \mathbf{H}_{ir} \mathbf{H}_{ip}^\perp \mathbf{P}_i \sigma_s^2 + \mathbf{T}_r (\mathbf{H}_{pr} \mathbf{H}_{pr}^\dagger \sigma_p^2 + \sigma_n^2 \mathbf{I}) \mathbf{T}_r^\dagger\}$. This is a convex function of $\mathbf{P}_i$ because $\mathbf{H}_{ip}^{\perp\dagger} \mathbf{H}_{ir}^\dagger \mathbf{T}_r^\dagger \mathbf{T}_r \mathbf{H}_{ir} \mathbf{H}_{ip}^\perp$ is positive semi-definite. Hence problem P3 is convex. □

Appendix E

Proof: Problem P5 in Chapter 3 is convex

Define $\mathbf{G} = \check{\mathbf{H}}_2^T \otimes (\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}) - \sum_{i=1}^{K}(\check{\mathbf{H}}_2\mathbf{E}_r^{(i)})^T \otimes (\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp})$ and $\mathbf{L} = \mathbf{H}_{pr}^T \otimes (\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp})$. Using (3.39) and after considerable manipulations, the objective function of problem P5 can be expressed as

$$\| \check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}\mathbf{P}_r\check{\mathbf{H}}_2\mathbf{s} - \sum_{i=1}^{K}\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}\mathbf{P}_r\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\mathbf{s} + \check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}\mathbf{P}_r\mathbf{H}_{pr}\mathbf{x}_p + \check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}\mathbf{P}_r\mathbf{n}_r + \check{\mathbf{H}}_p\tilde{\mathbf{x}}_p + \mathbf{R}\mathbf{n} - \mathbf{E}\mathbf{s} \|_2^2$$

$$= \mathrm{tr}\left\{\mathrm{vec}\{\mathbf{P}_r\}^{\dagger}\mathbf{G}^{\dagger}\mathbf{G}\mathrm{vec}\{\mathbf{P}_r\}\sigma_s^2\right\} - \mathrm{tr}\Bigg\{\mathbf{E}^{\dagger}\left(\check{\mathbf{H}}_2^{\dagger}\mathbf{H}_{rp}^{\perp}\mathbf{P}_r^{\dagger}\mathbf{H}_1^{\dagger} - \sum_{i=1}^{K}\mathbf{E}_r^{(i)\dagger}\check{\mathbf{H}}_2^{\dagger}\mathbf{H}_{rp}^{\perp}\mathbf{P}_r^{\dagger}\check{\mathbf{H}}_1^{\dagger}\mathbf{E}_l^{(i)\dagger}\right)\sigma_s^2$$

$$-\left(\check{\mathbf{H}}_1\mathbf{H}_{rp}^{\perp}\mathbf{P}_r\check{\mathbf{H}}_2 - \sum_{i=1}^{K}\mathbf{E}_l^{(i)}\check{\mathbf{H}}_1\mathbf{T}_r\check{\mathbf{H}}_2\mathbf{E}_r^{(i)}\right)\mathbf{E}^* + \mathbf{R}\mathbf{R}^{\dagger}\sigma_n^2\Bigg\} + \mathrm{tr}\{\mathrm{vec}\{\mathbf{P}_r\}^{\dagger}\mathbf{L}^{\dagger}\mathbf{L}\mathrm{vec}\{\mathbf{P}_r\}\sigma_p^2\}. \quad \text{(E.1)}$$

In (E.1) the first term is convex in $\mathbf{P}_r$ since $\mathbf{G}^{\dagger}\mathbf{G}$ is positive semi-definite and the the second term is linear in $\mathbf{P}_r$. The third term in (E.1) is also convex in $\mathbf{P}_r$ since $\mathbf{L}^{\dagger}\mathbf{L}$ is a positive semi-definite matrix. Therefore the objective function, which is the sum of these three convex terms, is also a convex function in $\mathbf{P}_r$. Now we consider the constraint of P5. Since $(\check{\mathbf{H}}_2\check{\mathbf{H}}_2^{\dagger}\sigma_s^2 + \mathbf{H}_{pr}\mathbf{H}_{pr}^{\dagger}\sigma_p^2 + \sigma_n^2\mathbf{I})$ is Hermitian, it can be expressed as $\mathbf{Q}\mathbf{Q}^{\dagger}$. So the constraint of P5 can be expressed as $\mathrm{tr}\left\{\mathrm{vec}\{\mathbf{P}_r\}^{\dagger}(\mathbf{Q}^T \otimes \mathbf{H}_{rp}^{\perp})^{\dagger}(\mathbf{Q}^T \otimes \mathbf{H}_{rp}^{\perp})\mathrm{vec}\{\mathbf{P}_r\}\right\} \leq P_{tot,r}$. Since $(\mathbf{Q}^T \otimes \mathbf{H}_{rp}^{\perp})^{\dagger}(\mathbf{Q}^T \otimes \mathbf{H}_{rp}^{\perp})$ is positive semi-definite, this constraint is convex. Therefore problem P5 is convex. □

Appendix F

Proof of Proposition 3.1

By setting $\frac{\partial L(\lambda,\mu,\mathbf{T}_r,\gamma)}{\partial \mathbf{T}_r^\dagger}=0$ and $\frac{\partial L(\lambda,\mu,\mathbf{T}_r,\gamma)}{\partial \gamma}=0$, we obtain respectively

$$\left[\tilde{\mathbf{H}}_1^\dagger\tilde{\mathbf{H}}_1+e_1\mathbf{I}\right]\mathbf{T}_r\left[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger+e_2\mathbf{I}+\tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger\sigma_p^2+\sigma_{pr}^2J_p\sigma_p^2\mathbf{I}+\sigma_n^2\mathbf{I}\right]\sigma_s^2$$

$$-\sum_{i=1}^{K}\left[\tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\tilde{\mathbf{H}}_1+\sigma_{ri}^2\mathrm{tr}\{\mathbf{R}_i^{*\dagger}\mathbf{R}_i^*\}\mathbf{I}\right]\mathbf{T}_r\left[\tilde{\mathbf{H}}_2\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger+\sigma_{ir}^2\mathrm{tr}\{\mathbf{T}_i^*\mathbf{T}_i^{*\dagger}\mathbf{I}\}\right]\sigma_s^2$$

$$+\left[\frac{\mu}{\gamma^2}\left(\tilde{\mathbf{H}}_{rp}^\dagger\tilde{\mathbf{H}}_{rp}+\sigma_{rp}^2N\mathbf{I}\right)+\frac{\lambda}{\gamma^2}\mathbf{I}\right]\mathbf{T}_r\left[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger\sigma_s^2+e_2\sigma_s^2\mathbf{I}+\tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger\sigma_p^2+\sigma_{pr}^2J_p\sigma_p^2\mathbf{I}+\sigma_n^2\mathbf{I}\right]$$

$$=\left(\tilde{\mathbf{H}}_1^\dagger\mathbf{E}\tilde{\mathbf{H}}_2^\dagger-\sum_{i=1}^{K}\tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\mathbf{E}\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger\right)\sigma_s^2\gamma^{*-1} \tag{F.1}$$

and

$$\gamma\Big\{\left[\tilde{\mathbf{H}}_1^\dagger\tilde{\mathbf{H}}_1+e_1\mathbf{I}\right]\mathbf{T}_r\left[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger+e_2\mathbf{I}+\tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger\sigma_p^2+\sigma_{pr}^2J_p\sigma_p^2\mathbf{I}+\sigma_n^2\mathbf{I}\right]\mathbf{T}_r^\dagger\sigma_s^2$$

$$-\sum_{i=1}^{K}\left[\tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\tilde{\mathbf{H}}_1+\sigma_{ri}^2\mathrm{tr}\{\mathbf{R}_i^{*\dagger}\mathbf{R}_i^*\}\mathbf{I}\right]\mathbf{T}_r\left[\tilde{\mathbf{H}}_2\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger+\sigma_{ir}^2\mathrm{tr}\{\mathbf{T}_i^*\mathbf{T}_i^{*\dagger}\mathbf{I}\}\right]\mathbf{T}_r^\dagger\sigma_s^2\Big\}$$

$$-\gamma\mathrm{tr}\Big\{\tilde{\mathbf{H}}_p\tilde{\mathbf{H}}_p^\dagger\sigma_p^2+\sigma_p^2J_p\sum_{i=1}^{K}\tilde{d}_i\sigma_{pi}^2+\mathbf{R}\mathbf{R}^\dagger\sigma_n^2\Big\}=\mathbb{R}\Big\{\mathrm{tr}\Big\{\Big[\tilde{\mathbf{H}}_1^\dagger\mathbf{E}\tilde{\mathbf{H}}_2^\dagger-\sum_{i=1}^{K}\tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\mathbf{E}\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger\Big]\sigma_s^2\mathbf{T}_r^\dagger\Big\}\Big\}. \tag{F.2}$$

Right multiply (F.1) by $\mathbf{T}_r^\dagger\gamma$. We observe that $\left\{\left[\tilde{\mathbf{H}}_1^\dagger\mathbf{E}\tilde{\mathbf{H}}_2^\dagger-\sum_{i=1}^{K}\tilde{\mathbf{H}}_1^\dagger\mathbf{E}_l^{(i)}\mathbf{E}\mathbf{E}_r^{(i)}\tilde{\mathbf{H}}_2^\dagger\right]\sigma_s^2\mathbf{T}_r^\dagger\right\}$ is Hermitian. Then we take the trace of equation (F.1) right multiplied by $\mathbf{T}_r^\dagger\gamma$. We observe that its right-side is exactly the same as the right-side of (F.2). Then we have

$$\mathrm{tr}\left\{\left[\tfrac{\mu}{\gamma^2}\left(\tilde{\mathbf{H}}_{rp}^\dagger\tilde{\mathbf{H}}_{rp}+\sigma_{rp}^2N\mathbf{I}\right)+\tfrac{\lambda}{\gamma^2}\mathbf{I}\right]\mathbf{T}_r\left[\tilde{\mathbf{H}}_2\tilde{\mathbf{H}}_2^\dagger\sigma_s^2+e_2\sigma_s^2\mathbf{I}+\tilde{\mathbf{H}}_{pr}\tilde{\mathbf{H}}_{pr}^\dagger\sigma_p^2+\sigma_{pr}^2J_p\sigma_p^2\mathbf{I}+\sigma_n^2\mathbf{I}\right]\mathbf{T}_r^\dagger\right\}$$

$$=\mathrm{tr}\Big\{\tilde{\mathbf{H}}_p\tilde{\mathbf{H}}_p^\dagger\sigma_p^2\Big\}+\sigma_p^2J_p\textstyle\sum_{i=1}^{K}\tilde{d}_i\sigma_{pi}^2+\sum_{i=1}^{K}\tilde{d}_i\sigma_n^2. \tag{F.3}$$

Let I_c and P_c denote the interfering power to primary users and "consumed" power, respectively, given by

$$P_c = \mathbf{T}_r \left[\tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_2^\dagger \sigma_s^2 + e_2 \sigma_s^2 \mathbf{I} + \tilde{\mathbf{H}}_{pr} \tilde{\mathbf{H}}_{pr}^\dagger \sigma_p^2 + \sigma_{pr}^2 J_p \sigma_p^2 \mathbf{I} + \sigma_n^2 \mathbf{I} \right] \mathbf{T}_r^\dagger, \tag{F.4}$$

$$I_c = \left(\tilde{\mathbf{H}}_{rp}^\dagger \tilde{\mathbf{H}}_{rp} + \sigma_{rp}^2 N \mathbf{I} \right) \mathbf{T}_r \left[\tilde{\mathbf{H}}_2 \tilde{\mathbf{H}}_2^\dagger \sigma_s^2 + e_2 \sigma_s^2 \mathbf{I} + \tilde{\mathbf{H}}_{pr} \tilde{\mathbf{H}}_{pr}^\dagger \sigma_p^2 + \sigma_{pr}^2 J_p \sigma_p^2 \mathbf{I} + \sigma_n^2 \mathbf{I} \right] \mathbf{T}_r^\dagger. \tag{F.5}$$

Then we can observe that the left-side of (F.3) is actually $\frac{\lambda}{\gamma^2} P_c + \frac{\mu}{\gamma^2} I_c$. When the optimal values of the primal and dual variables are achieved, the following conditions (complementary-slackness) should be satisfied:

$$\frac{\mu^*}{\gamma^2} \left\{ I_c - I_{tot} \right\} = 0, \quad \frac{\lambda^*}{\gamma^2} \left\{ P_c - P_{tot,r} \right\} = 0. \tag{F.6}$$

Therefore, whether the values of λ^* and μ^* are positive or zero, we will always have (3.89). □