Department of Electrical and Computer Engineering

Efficient Channel Estimation Algorithms for Cooperative Multiple-Input Multiple-Output (MIMO) Wireless Communication Networks

Choo Wee Raymond Chiong

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

September 2014

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: .....

Date: .....

Acknowledgements

I am grateful that God has given me the opportunity to pursue my doctoral studies. Through numerous people, which I am about to mention, He has generously provided me with the strength and wisdom to complete my PhD degree.

First and foremost, I would like to express my deepest appreciation and gratitude to my supervisor, A/Prof. Yue Rong for his invaluable guidance and support throughout the years of my PhD studies. Without his patience and enthusiasm in guiding me, I would not have gotten this far. A special thank you goes to Prof. Yong Xiang of Deakin University, Melbourne, who guides me with his technical expertise and suggestions. I would also like to thank my thesis committee members, Dr. Yee-Hong Leung and Prof. Sven Nordholm for their comments and suggestions.

My deep gratitude also goes to Curtin University for providing me with the financial and research supports. Further, I would like to express my sincere appreciation to my research peers at the Communication and Signal Processing (CSP) group of the Department of Electrical and Computer Engineering for their friendly help and supports.

Last but not least, I would like to thank my family and friends. To my lovely parents and my only sister, thank you for all the love and understanding. To my friends who share good moments with me, and also brothers and sisters in Immanuel Methodist Church (IMC), thank you for your companion, encouragements, and prayers.

Abstract

The technology of wireless communication has been playing a vital role in many aspects of the modern lifestyle. In recent years, the increasing number of wireless devices has resulted in a significant growth on the demand for reliable and high rate wireless communications. Multiple-input multiple-output (MIMO) relay communication systems have been identified to be one of the promising solutions to high rate wireless communication. Besides enhancing the capacity of the networks, MIMO technology is able to reduce the effect of fading in wireless communication networks by means of spatial diversity. The use of relaying scheme is able to increase the network coverage and system reliability.

However, challenging problems in implementing MIMO relay systems do exist and the solutions remain elusive. One of the challenges is to estimate the channel state information (CSI) of the MIMO relay systems, which is required for retrieving source signals and optimizing the design of the transceiver and relay matrices. However, the knowledge of CSI is unknown in practical MIMO relay networks, and thus needs to be estimated. This thesis focuses on the channel estimation issues for MIMO relay communication networks, considering the amplify-and-forward relaying scheme.

First, a robust channel estimation algorithm for one-way MIMO relay systems is proposed, which takes into account the CSI mismatch between the estimated and the true relay-destination channel matrices. The proposed algorithm always perform better than conventional algorithms in estimating the source-relay channel without the need of greater computational effort.

Then the channel estimation issues on two-way MIMO relay communication networks are discussed. Two channel estimation algorithms, namely the superimposed channel training scheme and the optimized two-stage channel estimation

algorithm, are presented and compared. The proposed algorithms are able to estimate the individual CSI for the first-hop and second-hop links. Through numerical examples, it can be concluded that both algorithms outperform the conventional two-stage channel estimation technique, with the optimized two-stage channel estimation algorithm performs better than the superimposed channel training scheme at the expense of higher computational complexity.

Next a more general situation is considered where two-way MIMO relay systems are operating in frequency-selective fading environments. The channel estimation problem is becoming more complicated in frequency-selective fading environments as there are multiple paths between each transmit-receive antenna pair. The superimposed channel training method is applied to derive the individual CSI of first-hop and second-hop links. A minimum MSE (MMSE)-based algorithm to retrieve first-hop channel matrices is also presented, considering the estimation error inherited from the estimation of the second-hop channel matrices.

Thus far, the channel estimation problems are addressed by sending known training sequences to the destination node to assist in the estimation of the channels. This thesis examines the blind approach for channel estimation, i.e., the channels are estimated based on the statistical properties of the source signals. Two blind source separation (BSS) methods are integrated to estimate the individual CSI of the source-relay and relay-destination links for one-way MIMO relay systems. In particular, a first-order Z-domain precoding technique is presented to blindly estimate the relay-destination links, while the source-relay links are estimated based on signal mutual information (MI) modified constant modulus algorithm. The proposed blind channel estimation algorithms always have a better bandwidth efficiency compared with training-based channel estimation algorithms as no bandwidth is used to send training sequences.

Author's Note

Parts of this thesis have been previously published in the following journals and/or conference papers.

Journal Papers

1. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Training Algorithms for Two-Way MIMO Relay Systems," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3988-3998, Aug. 2013.

2. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Estimation for Two-Way MIMO Relay Systems in Frequency-Selective Fading Environments," *IEEE Trans. Wireless Commun.*, to appear, 2014.

3. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Blind Channel Estimation and Signal Retrieving for MIMO Relay Systems," *IEEE Trans. Signal Process.*, revised, Sep. 2014.

Conference Papers

1. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Robust Channel Estimation Algorithm for Dual-Hop MIMO Relay Channels," *Proc. 23rd IEEE Int. Symposium Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sep. 9-12, 2012, pp. 2376-2381.

2. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Superimposed Channel Training for Two-Way MIMO Relay Systems," *Proc. 13th IEEE Int. Conf. Commun. Syst. (ICCS)*, Singapore, Nov. 21-23, 2012, pp. 21-25.

3. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Blind Estimation of MIMO Relay Channels," *Proc. IEEE Workshop on Statistical Signal Processing (SSP'2014)*, Gold Coast, Australia, Jun. 29-Jul. 2, 2014, pp. 400-403.

4. C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Estimation for Frequency-Selective Two-Way MIMO Relay Systems," *Proc. Int. Symposium Inf. Theory Its Applications (ISITA'2014)*, Melbourne, Australia, Oct. 26-29, 2014.

Contents

List of Figures

List of Tables

List of Acronyms

AWGN	additive white Gaussian noise
BER	bit-error-rate
BSS	blind source separation
CDMA	code-division multiple access
CSCG	circularly symmetric complex Gaussian
CSI	channel state information
DFT	discrete Fourier transform
DML	deterministic maximum likelihood
EM	expectation-maximization
EVD	eigenvalue decomposition
FFT	fast Fourier transform
GSM	Global System for Mobile
GSS	golden section search
HOS	higher-order statistics

i.i.d.	independent and identically distributed
ICA	independent component analysis
KKT	Karush-Kuhn-Tucker
LHS	left-hand-side
LMMSE	linear minimum mean-squared error
LMS	least mean squares
LS	least-square
LTE	long-term evolution
MAP	maximum a posteriori
MI	mutual information
MIMO	multiple-input multiple-output
MIRL	mean interference rejection level
ML	maximum likelihood
MMSE	minimum mean-squared error
MSE	mean-squared error
MVU	minimal variance unbiased
NMSE	normalized mean-squared error
OFDM	orthogonal frequency-division multiplexing

PLC	power line communications
PSD	positive semi-definite
PSK	phase-shift keying
QAM	quadrature amplitude modulation
RLS	recursive least squares
SISO	single-input single-output
SNR	signal-to-noise ratio
SOS	second-order statistics
SVD	singular value decomposition
WLAN	wireless local area network
WLS	weighted least squares

List of Notations

$(\cdot)^*$	complex conjugate
$(\cdot)^H$	matrix (vector) Hermitian transpose
$(\cdot)^T$	matrix (vector) transpose
$(\cdot)^\dagger$	matrix pseudo-inverse
$(\cdot)^{-1}$	matrix inversion
$E[\cdot]$	statistical expectation
$\otimes$	matrix Kronecker product
$\mathbf{0}_{m\times n}$	$m \times n$ zero matrix
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{a} * \mathbf{b}$	linear convolution between vectors $\mathbf{a}$ and $\mathbf{b}$
Bdiag[$\cdot$]	block diagonal matrix
diag[$\mathbf{x}$]	diagonal matrix taking $\mathbf{x}$ as the diagonal elements
tr($\cdot$)	matrix trace

$\text{vec}(\cdot)$	vectorization operator - stacks all column vectors of a matrix on top of each other
$\lvert \cdot \rvert$	absolute value
e.g., $\mathbf{A}, \mathbf{B}$	bold face upper case letters denote matrices
e.g., $\mathbf{a}, \mathbf{b}$	bold face lower case letters denote vectors
e.g., a, b	lower case letters denote scalars

Chapter 1

Introduction

Wireless communication networks have been a revolutionary part in the field of communication, which enable convenient multimedia communication between people and devices. With a tremendous demand for high speed and reliable wireless communications, multi-antenna and relaying techniques are expected to be included in the next generation of wireless communication networks. The main aim of this thesis is to develop efficient channel estimation algorithms for multiple-input multiple-output (MIMO) wireless relay communication networks. In this chapter, we introduce the background knowledge on MIMO relay networks and the estimation of channel state information. We also state the contributions and give an outline of the thesis.

1.1 Relaying and Cooperative Communication

Relaying and cooperative communications were first introduced by Van Der Meulen in [1], and were examined by Cover and El Gamal in [2]. The basic idea of cooperative communication is to provide multiple routes for data packets to be transmitted from the source to the destination. Fig. 1.1 shows a three-node two-hop relay network, where data packets from the source node can be transmitted to the destination node directly and/or through the relay node.

In recent years, growing interest in cooperative communications employing relay nodes has motivated great research efforts in this area such as in [3]-[10]. The major advantages of cooperative relay communication can be summarized

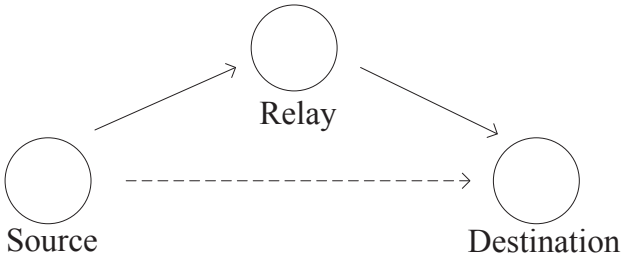

Figure 1.1: A three-node two-hop relay network

below [3]-[8], [11]-[14]:

- The wireless networks coverage are increased, and is particularly useful when the distance between the source and destination is long or the channel between the source and destination is strongly faded.

- The reliability and robustness of the wireless systems have been improved as multiple copies of source signals are transmitted, i.e., the diversity gain at the destination node is increased.

There are mainly two types of relaying strategies, namely the non-regenerative strategy and the regenerative strategy. The former technique only amplifies the received signals and forwards the signals to the receiver, while the later scheme decodes the received signals, then re-encodes them before retransmits the signals to the receiver. Compared with the regenerative relaying strategy, the non-regenerative relaying strategy delivers more noise to the destination as the noise is amplified at the relay node [10]. However, the non-regenerative relaying strategy does not require sophisticated signal processing at the relay node, thus is easier to implement than the regenerative relaying technique. In this thesis, the implementation of non-regenerative relaying strategy is considered. Recently, relaying and cooperative communications techniques have been considered in some industrial communication standards, such as IEEE 802.16j WiMAX standard [15] and 3GPP's long-term evolution (LTE)-advanced standard [16].

1.2 MIMO Relay Communication Networks

The emerging demands in multimedia applications such as broadcasting of high definition television have resulted in the need for wireless systems that are able to

support higher data rate compared with the existing systems. Many approaches have been suggested and developed to boost the performance of wireless systems, which include the use of multiple antennas at both the transmitter and receiver sides, i.e., MIMO technology [17]-[31].

The benefits of MIMO systems [18]-[23] are briefly explained below.

- *Spatial diversity gain.* MIMO systems are able to produce smaller error rates by transmitting and receiving the same source signals over multiple independent fading paths (when the antennas spacing is sufficiently large). This spatial diversity gain is able to combat the fading effects that usually occurred in wireless channel. Comparing with single-antenna transmission, MIMO systems are more reliable and robust when the same transmission rate is used. The spatial diversity has an advantage over time and frequency diversity as it does not consume additional bandwidth or require longer transmission time.

- *Spatial multiplexing gain.* In MIMO systems, independent data streams can be simultaneously transmitted from different antennas to increase the transmission rate without the need of extra bandwidth. The multiplexing gain achieved in MIMO networks is linear to the number of antennas at the transmitter/receiver, whichever is smaller.

- *Array gain.* signal-to-noise ratio (SNR) at the receiver can be improved using multi-antenna techniques. With the spatial filtering (beamingforming) technique, the directions of the transmitter and receiver antennas can be steered to favor the desirable signals, such as the information signals, while suppressed the interference signals. This technique is also known as smart antennas.

In general, tradeoff between the spatial diversity gain, spatial multiplexing gain, and array gain is required as it may not be possible to reap all the benefits of MIMO systems at the same time. The diversity-multiplexing tradeoff was investigated in [21]. Currently, MIMO technology has been applied in many commercial wireless products and systems, such as in routers, base stations, wireless local area network (WLAN) and 3G cellular network. It has been considered

in many industrial standards, for example, IEEE 802.11n WLAN standard [32] and IEEE 802.20 mobile broadband wireless access systems [33]. Recently, the application of MIMO systems for broadband power line communications (PLC) is being investigated in [34]-[35].

Considering the advantages of MIMO technology and cooperative communications, it is expected for the next generation of wireless systems to turn to MIMO relaying schemes. However, many challenging problems arise and research works have been actively carried out to maximize the utilization of MIMO relaying schemes [11],[36]-[46]. In [36] and [37], the optimal relay precoding matrix is derived to maximize the mutual information between the source and destination nodes for a three-node two-hop MIMO relay communication system. A unified framework has been developed in [11] to optimize the source and relay precoding matrices for two-hop MIMO relay systems with a broad class of commonly used objective functions. In [38], the capacity of MIMO relay networks for Gaussian channel and Rayleigh fading channel is studied. The joint transmit and relay precoding design problems were investigated for two-hop multicasting MIMO relay systems in [39]. A recent survey on transceiver design for amplify-and-forward MIMO relay systems is presented in [40]. Other optimization works on MIMO relay systems are investigated in [41]-[46].

1.3 Estimation of Channel State Information

Let us consider the simplest frequency-flat fading two-hop MIMO relay wireless network where a source node transmits information to a destination node through a relay node as shown in Fig. 1.2. The source, relay, and destination nodes are equipped with N_s, N_r, and N_d antennas, respectively. The amplify-and-forward strategy is considered at the relay node.

Figure 1.2: A two-hop MIMO relay wireless network

In the first time block, the source node transmits an $N_s \times L$ data matrix $\mathbf{S}_s$

to the relay node, where L is the length of the transmitted data sequence. The received signal at the relay node is given by

$$\mathbf{Y}_r = \mathbf{H}_1\mathbf{S}_s + \mathbf{V}_r \tag{1.1}$$

where $\mathbf{H}_1$ is the $N_r \times N_s$ source-relay channel matrix and $\mathbf{V}_r$ is the $N_r \times L$ noise matrix at the relay node. In the second time block, the relay node applies an $N_r \times N_r$ precoding matrix $\mathbf{F}$ on $\mathbf{Y}_r$ and retransmits the linear precoded signal matrix

$$\mathbf{X}_r = \mathbf{F}\mathbf{H}_1\mathbf{S}_s + \mathbf{F}\mathbf{V}_r \tag{1.2}$$

to the destination node. The received signal at the destination node can be written as

$$\mathbf{Y}_d = \mathbf{H}_2\mathbf{X}_r + \mathbf{V}_d = \mathbf{H}_2\mathbf{F}\mathbf{H}_1\mathbf{S}_s + \mathbf{H}_2\mathbf{F}\mathbf{V}_r + \mathbf{V}_d \tag{1.3}$$

where $\mathbf{H}_2$ is the $N_d \times N_r$ relay-destination channel matrix and $\mathbf{V}_d$ is the $N_d \times L$ noise matrix at the destination node. Let $\bar{\mathbf{H}} \triangleq \mathbf{H}_2\mathbf{F}\mathbf{H}_1$ be the compound channel matrix from the source node to the destination node and $\mathbf{V} = \mathbf{H}_2\mathbf{F}\mathbf{V}_r + \mathbf{V}_d$ be the equivalent noise matrix, we have

$$\mathbf{Y}_d = \bar{\mathbf{H}}\mathbf{S}_s + \mathbf{V}. \tag{1.4}$$

Using a linear receiver at the destination node, the estimated $\mathbf{S}_s$ is given by

$$\hat{\mathbf{S}}_s = \mathbf{W}^H\mathbf{Y}_d \tag{1.5}$$

where $\mathbf{W}$ is the $N_d \times N_s$ weight matrix of the linear receiver. The optimal $\mathbf{W}$ that minimizes the estimation error of $\mathbf{S}_s$ is the Wiener filter [43] given by

$$\mathbf{W} = \left(\bar{\mathbf{H}}\bar{\mathbf{H}}^H + \mathbf{R}_V\right)^{-1}\bar{\mathbf{H}}. \tag{1.6}$$

Here $\mathbf{R}_V = \mathrm{E}\left[\mathbf{V}\mathbf{V}^H\right]$ is the covariance matrix of $\mathbf{V}$.

Let us define the singular value decomposition (SVD) of $\mathbf{H}_1$ and $\mathbf{H}_2$ as

$$\mathbf{H}_1 = \mathbf{U}_{H_1}\mathbf{\Lambda}_{H_1}\mathbf{V}_{H_1}, \qquad \mathbf{H}_2 = \mathbf{U}_{H_2}\mathbf{\Lambda}_{H_2}\mathbf{V}_{H_2} \tag{1.7}$$

where $\mathbf{U}_{H_1}, \mathbf{V}_{H_1}, \mathbf{U}_{H_2}$, and $\mathbf{V}_{H_2}$ are the singular vector matrices, and $\mathbf{\Lambda}_{H_1}$ and $\mathbf{\Lambda}_{H_2}$ are the diagonal singular value matrix. It was shown in [36], [47]-[48] that the optimal relay precoding matrix for two-hop MIMO relay networks is given by

$$\mathbf{F} = \mathbf{V}_{H_2}\mathbf{\Lambda}_F\mathbf{U}_{H_1}^H \tag{1.8}$$

where $\mathbf{V}_{H_2}$ and $\mathbf{U}_{H_1}$ are unitary matrices of eigenvalue decompositions of the channel matrices $\mathbf{H}_2$ and $\mathbf{H}_1$, respectively, and $\mathbf{\Lambda}_F$ is a diagonal matrix. It is obvious from (1.5)-(1.6) and (1.8) that in MIMO relay networks, the knowledge of the channel characteristics, also known as instantaneous channel state information (CSI), is required for the purpose of

- retrieving the source information at the destination node
- optimization of the MIMO relay systems, for example, to derive the optimal source and relay matrices in linear amplify-and-forward relay networks.

However, in practical MIMO relay communication systems [11]-[13], [36]-[46], the instantaneous CSI is unavailable at both transmitter and receiver, and thus, it has to be estimated. Considering the importance of CSI in the overall performance of MIMO relay networks, it is crucial to address the channel estimation problems. Different from guided transmission medias (signal propagates along a solid medium), the channel estimation in wireless systems is much more difficult as wireless networks exhibit highly dynamic channels. Moreover, when multiple antennas are used, more channel parameters are needed to be estimated.

Generally, there are two types of channel estimation techniques, [18],[49]-[50].

- *Training-based channel estimation.* In this method, known training sequences are transmitted to the receiver for the estimation of the instantaneous CSI. The accuracy of the channel estimation can be improved by sending training sequences more frequently, especially for fast-fading channels. This method has the advantages of simplicity and reliability compared with blind channel estimation. However, training-based method has a lower bandwidth efficiency as part of the bandwidth is used to transmit the training sequences. For example, 26 bits of training sequences are placed in the middle of each packet for Global System for Mobile (GSM) systems as illustrated in Fig. 1.3. The training-based channel estimation technique is discussed in [51]-[70].

- *Blind channel estimation.* The blind channel estimation technique relies only on the received signals at the receiver during normal data transmission

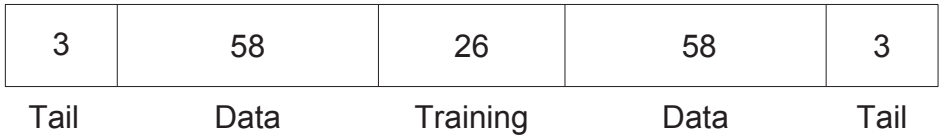

Figure 1.3: Structure of a data packet in GSM systems

to estimate the instantaneous CSI, without requiring the transmission of training sequences. This technique utilizes the known statistical properties of the transmitted signals and channels, such as cyclostationarity and constant modulus properties. Eventually, blind technique might need a large amount of data to achieve acceptable performance, which is not favorable for fast-fading channels. Many blind channel estimation algorithms also suffer from the complexity and stabilities issues. Nevertheless, this technique achieves higher bandwidth efficiency compared with training-based channel estimation, as no bandwidth is used for the transmission of training sequence [71]-[78].

1.4 Training-based Channel Estimation for Single-hop Networks

In the early stage of tackling the channel estimation issues, the training-based techniques have attracted the attention of the most researchers as the training-based techniques are more reliable and robust compared with the blind methods. The pilot-assisted channel estimation technique was discussed in [51]-[52] for single-antenna networks, and the performance of the technique was analyzed in [53]. Later, these works have been extended to wireless networks equipped with multiple antennas [54]-[65]. Let us consider the simplest frequency-flat fading single-hop MIMO wireless network where the source node transmits information to the destination node as shown in Fig. 1.4. The source and destination nodes are equipped with N_s and N_d antennas, respectively. The $N_d \times L$ received signal at the destination node can be written as

$$\mathbf{Y}_d = \mathbf{H}\mathbf{S}_s + \mathbf{V}_d \tag{1.9}$$

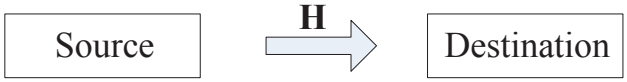

Figure 1.4: A single-hop MIMO wireless network

where $\mathbf{H}$ is the $N_d \times N_s$ channel matrix from the source node to the destination node, $\mathbf{S}_s$ is the $N_s \times L$ source signal matrix, and $\mathbf{V}_d$ is the $N_d \times L$ noise matrix at the destination node. Here, L is the length of the transmitted data sequence.

There are two popular training-based linear channel estimators, which are

- *Least-square (LS) estimator.* This estimator is used when the knowledge of the channel and noise parameters are unknown. From (1.9), the LS estimation of $\mathbf{H}$ is given by

$$\hat{\mathbf{H}}_{LS} = \mathbf{Y}_d \mathbf{S}_s^{\dagger}. \tag{1.10}$$

  Note that for a linear estimator, we have $L \geq N_s$, i.e., the matrix $\mathbf{S}_s$ is a fat matrix.

- *Minimum mean-squared error (MMSE) estimator.* When a priori information of the channel and noise distributions are available, an MMSE estimator can be used to reduce the estimation error. From (1.9), the MMSE estimation of $\mathbf{H}$ can be written as

$$\hat{\mathbf{H}}_{MMSE} = \mathbf{Y}_d \left(\mathbf{S}_s^H \mathbf{R}_H \mathbf{S}_s + \mathbf{R}_{V_d}\right)^{-1} \mathbf{S}_s^H \mathbf{R}_H \tag{1.11}$$

  where $\mathbf{R}_H$ and $\mathbf{R}_{V_d}$ are the channel and noise covariance matrices, respectively.

The channel estimation issues for single-hop multiple-antennas networks have been widely considered in the literature. In [54], the optimal design of the training sequences is studied by using the maximum likelihood (ML) method for block flat-fading MIMO channels. This work has been extended in [55] where a simple ML estimator is applied for continuous flat-fading MIMO channels. The tradeoff between the channel capacity and number of training symbols for MIMO networks is investigated in [56], and the optimal number of training symbols required for a meaningful channel estimation is discussed. A superimposed pilot sequence

technique for channel estimation is proposed in [57]-[58], where a known training sequence is linearly added into the unknown data sequence. In [59], optimal training sequences for several channel estimation methods are derived for networks with multiple transmit antennas and single receive antenna. Subsequently, the work in [59] was extended to the case of MIMO networks in [60]-[62]. Other training-based channel estimation methods for single-hop MIMO networks are discussed in [63]-[65].

1.5 Overview and Contributions of the Thesis

In near future, MIMO relay systems are expected to be deployed in the next generation wireless systems as they are capable of providing higher data rate, wider network coverage and better network reliability compared with the conventional wireless systems. As mentioned in Section 1.3, the estimation of instantaneous CSI is essential in practical MIMO relay networks. Several channel estimation algorithms are discussed in Section 1.4 for single-hop multiple antennas networks, but the extension to MIMO relay networks is not straightforward. This thesis focuses on the channel estimation issues for MIMO relay wireless communication systems. In this thesis, the amplify-and-forward strategy is considered at the relay node. This is because for MIMO relay networks, the complexity of amplify-and-forward scheme is much lower compared with the decode-and-forward scheme, since great computational efforts are required to decode and encode multiple data streams. Subsequently, the deployment of the amplify-and-forward relay scheme is easier, more cost effective, and have shorter end-to-end delay than that of the decode-and-forward scheme.

The major contribution of this thesis is the development of efficient channel estimation algorithms for MIMO relay networks. Both the training-based and blind channel estimation techniques are investigated. Chapter 2 introduces a robust training-based channel estimation algorithm for one-way MIMO relay systems. In Chapter 3, two training-based channel estimation methods for two-way MIMO relay systems operating in frequency-flat fading environment are proposed. Chapter 4 studies the channel estimation problem for frequency-selective two-way

MIMO relay systems. The blind channel estimation method for MIMO relay systems is presented in Chapter 5. The summary of the thesis and brief descriptions on the possible future works are given in Chapter 6.

Chapter 2: One-Way MIMO Relay Systems

In this chapter, a robust channel estimation algorithm for one-way MIMO relay systems is developed. For conventional two-phase channel estimation algorithms, the estimated relay-destination channel is used for the estimation of source-relay channel. However, there is always a mismatch between the estimated and the true relay-destination channel. This motivates us to investigate the impact of such CSI mismatch on the accuracy of the source-relay channel estimation. By explicitly taking into consideration the CSI mismatch, a robust algorithm to estimate the source-relay channel is presented. In the proposed algorithm, the stochastic channel mismatch model is considered. Compared with the conventional two-phase channel estimation techniques, the proposed robust channel estimation algorithm performs better in estimating the source-relay channel, without requiring greater computational effort.

The material in Chapter 2 is based on the conference publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Robust Channel Estimation Algorithm for Dual-Hop MIMO Relay Channels," *Proc. 23rd IEEE Int. Symposium Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sep. 9-12, 2012, pp. 2376-2381.

Chapter 3: Frequency-Flat Two-Way MIMO Relay Systems

It is known that two-way MIMO relay systems are able to provide higher spectral efficiency compared with conventional one-way MIMO relay systems. This motivates us to look into the channel estimation problem in two-way MIMO relay systems. With the increasing number of channel parameters to be estimated, the channel estimation problem in two-way MIMO relay systems is more difficult than that of one-way MIMO relay systems. In this chapter, we propose two channel estimation algorithms for two-way MIMO relay systems operating in frequency-flat fading environment. First, we develop the superimposed channel training scheme where a training matrix is inserted to the received signals at the relay node to

assist in the channel estimation, and the channel estimation is completed in one transmission cycle. Then, we present the two-stage channel estimation algorithm where the relay-users links are estimated in the first stage while the users-relay channels are estimated in the second stage.

For both algorithms, the optimal structure of the source and relay pilot matrices are derived to minimize the mean-squared error (MSE) of channel estimation. The power allocation issue is also discussed in this chapter, where the power allocation between the source and relay training sequences is optimized in the superimposed channel training scheme, and in the two-stage channel estimation algorithm, the optimal power allocation at the relay node between two stages is derived. Besides having a distinct advantage of capable to estimate the individual CSI for the users-relay and relay-users links, both proposed algorithms perform better in estimating the channel matrices than that of conventional channel estimation algorithms.

Chapter 3 is based on the journal publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Training Algorithms for Two-Way MIMO Relay Systems," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3988-3998, Aug. 2013.

and the conference publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Superimposed Channel Training for Two-Way MIMO Relay Systems," *Proc. 13th IEEE Int. Conf. Commun. Syst. (ICCS)*, Singapore, Nov. 21-23, 2012, pp. 21-25.

Chapter 4: Frequency-Selective Two-Way MIMO Relay Systems

In this chapter, we propose a channel estimation algorithm that generalizes the findings in Chapter 3 from frequency-flat fading channels to frequency-selective fading channels. This extension is non-trivial as the optimization of channel estimation problem in frequency-selective two-way MIMO relay systems is much more complicated compared with the frequency-flat systems. The method of superimposed channel training is proposed to estimate the individual CSI of the users-relay and relay-users links for frequency-selective two-way MIMO relay systems. To minimize the MSE of channel estimation, the optimal structure of

the source and relay training sequences, as well as the optimal power allocation between the source and relay training sequences, are derived. In addition, a novel method based on the MMSE criterion to retrieve the first-hop channel matrices is presented, which explicitly takes into account the estimation error inherited from the estimation of the second-hop channel matrices.

The material in Chapter 4 is based on the journal publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Estimation for Two-Way MIMO Relay Systems in Frequency-Selective Fading Environments," *IEEE Trans. Wireless Commun.*, to appear, 2014.

and the conference publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Estimation for Frequency-Selective Two-Way MIMO Relay Systems," *Proc. Int. Symposium Inf. Theory Its Applications (ISITA'2014)*, Melbourne, Australia, Oct. 26-29, 2014.

Chapter 5: Blind Channel Estimation for MIMO Relay Systems

Blind channel estimation technique is well known for better spectral efficiency compared with training-based scheme as all the bandwidth is used for the transmission of communication data. This motivates us to examine the channel estimation problem from the perspective of blind technique. In this chapter, two blind source separation (BSS) methods are integrated to estimate the individual CSI of source-relay and relay-destination links for two-hop MIMO relay systems. Specifically, a first-order Z-domain precoding technique is proposed to blindly estimate the relay-destination channel matrix. This technique exploits the Z-domain properties of the precoders applied at the relay node to derive the estimation criterion for the relay-destination channel matrix. Then, using the estimated signals at the relay node, an algorithm based on the constant modulus and signal mutual information properties is developed for the blind estimation of source-relay channel matrix.

Chapter 5 is based on the journal publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Blind Channel Estimation and Signal Retrieving for MIMO Relay Systems," *IEEE Trans. Signal Process.*, revised, Sep. 2014.

and the conference publication:

- C. W. R. Chiong, Y. Rong, and Y. Xiang, "Blind Estimation of MIMO Relay Channels," *Proc. IEEE Workshop on Statistical Signal Processing (SSP'2014)*, Gold Coast, Australia, Jun. 29-Jul. 2, 2014, pp. 400-403.

Chapter 2

One-Way MIMO Relay Systems

In conventional two-phase channel estimation algorithms for dual-hop MIMO relay systems, the relay-destination channel estimated in the first phase is used for the source-relay channel estimation in the second phase. For these algorithms, the mismatch between the estimated and the true relay-destination channel affects the accuracy of the source-relay channel estimation. In this chapter, the impact of such CSI mismatch on the performance of the two-phase channel estimation algorithm is investigated. The state of the art is given in Section 2.1 while the system model of an one-way MIMO relay communication system is introduced in Section 2.2. The impact of CSI mismatch on the performance of the two-phase channel estimation algorithm is investigated and by explicitly taking into account the CSI mismatch, a robust algorithm to estimate the source-relay channel is developed in Section 2.3. In Section 2.4, we show some numerical examples. Conclusions are drawn in Section 2.5.

2.1 State of the Art

An overview of the training-based channel estimation techniques for single-hop MIMO networks is given in Section 1.4. However, the techniques used to estimate the channel matrices for single-hop MIMO systems can not be directly applied in MIMO relay systems. In the following, we discuss some channel estimation algorithms for one-way MIMO relay networks that have been proposed in the literature.

A novel interim channel estimation technique has been proposed in [79] where the source-relay and relay-destination channels are estimated at the destination node with the help of a known pilot enhancement matrix inserted at the relay node. However, the algorithm in [79] is developed for a MIMO mimicking amplify-and-forward relay system, and it is proven in [80] that the relay system can never fully mimic a real MIMO relay system as the multiplexing gain is limited. Two algorithms have been proposed in [81], namely, Bayesian-based linear minimum mean-squared error (LMMSE) and expectation-maximization (EM)-based maximum a posteriori (MAP) channel estimation. In the LMMSE channel estimation algorithm, only a sub-optimal solution can be achieved due to the high complexity in the computational of the LMMSE estimator. Consequently, the authors of [81] suggested the EM-based MAP channel estimation algorithm, where the initial estimate of the EM algorithm depends on the LMMSE estimator proposed earlier. However, the training sequences and relay precoding matrix are not optimized in [81]. A parallel factor analysis-based MIMO channel estimator was proposed in [82].

In [83], an algorithm based on the LS method was developed to estimate the channel matrices of MIMO relay networks. In particular, both the source-relay and the relay-destination channel matrices are estimated from the observed composite source-relay-destination channel matrix. A drawback from channel estimation using [83] is the scalar ambiguity of the estimated channel matrices. The performance of [83] was further analyzed and improved in [84] by using the weighted least squares (WLS) fitting method. A superimposed channel training algorithm has been proposed in [85] for orthogonal frequency-division multiplexing (OFDM) modulated one-way relay systems.

A two-phase channel estimation scheme based on LMMSE was proposed in [86] for two-hop MIMO relay networks. In particular, in the first phase, the source node is silent while the relay node transmits a pilot matrix to the destination node to estimate the relay-destination channel matrix. In the second phase, the source transmits a source pilot matrix to the relay. The relay node linearly precodes its received signal and forward it to the destination node. Then the source-relay channel is estimated at the destination node making use of the relay-destination

channel matrix estimated at the first phase. Compared with the approach in [83], there is no scalar ambiguity in this approach.

However, in practical relay systems, there is always mismatch between the estimated and the true relay-destination channel. Such CSI mismatch affects the accuracy of the source-relay channel estimation in [86]. In this chapter, we investigate the impact of this CSI mismatch on the performance of the two-phase channel estimation algorithm [86]. By explicitly taking into account the CSI mismatch, we develop a robust algorithm to estimate the source-relay channel, without the need of greater computation effort.

2.2 One-Way MIMO Relay System Model

We consider a one-way two-hop MIMO relay system where the source node transmits information to the destination node through a relay node as shown in Fig. 2.1. The source, relay, and destination nodes are equipped with N_s, N_r, and N_d antennas, respectively. We focus on the case where the direct link between the source and destination nodes is sufficiently weak to be ignored [83], [86]. This scenario occurs when the direct link is blocked by an obstacle such as a mountain. In fact, a relay plays a much more important role when the direct link is weak than when it is strong.

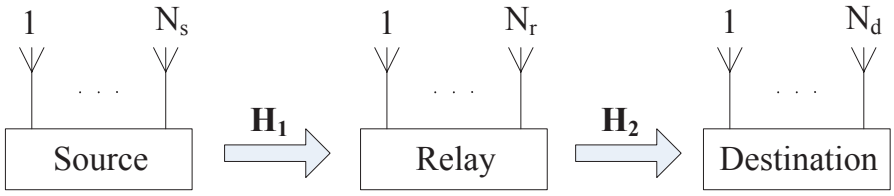

Figure 2.1: Block diagram of a one-way MIMO relay communication system.

Similar to [86], the channel matrices are estimated in two phases, where the relay-destination channel matrix $\mathbf{H}_2$ is estimated in phase one while the source-relay channel matrix $\mathbf{H}_1$ is estimated in phase two. In phase one, the signal received by the destination node is given by

$$\mathbf{Y}_d^{(1)} = \mathbf{H}_2\mathbf{S}_r + \mathbf{V}_d^{(1)} \tag{2.1}$$

where $\mathbf{S}_r$ is the $N_r \times N_r$ pilot matrix transmitted by the relay node to the destination node satisfying $\mathbf{S}_r^H \mathbf{S}_r = \mathbf{S}_r \mathbf{S}_r^H = \frac{p_r}{N_r} \mathbf{I}_{N_r}$ [61], and $\mathbf{V}_d^{(1)}$ is the $N_d \times N_r$ noise matrix at the destination node during phase one. Here p_r is the power budget available at the relay node. Note that we choose the length of $\mathbf{S}_r$ to be N_r to maximize the overall system spectral efficiency [87].

A minimal variance unbiased (MVU) estimation [88] of $\mathbf{H}_2$ can be obtained from (2.1) as

$$\hat{\mathbf{H}}_2 = \frac{N_r}{p_r} \mathbf{Y}_d^{(1)} \mathbf{S}_r^H = \mathbf{H}_2 + \frac{N_r}{p_r} \mathbf{V}_d^{(1)} \mathbf{S}_r^H. \tag{2.2}$$

It can be seen from (2.2) that due to the existence of the noise $\mathbf{V}_d^{(1)}$, there is a mismatch $\mathbf{\Delta}_2 \triangleq \frac{N_r}{p_r} \mathbf{V}_d^{(1)} \mathbf{S}_r^H$ between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$. Obviously, $\mathbf{\Delta}_2$ is a complex Gaussian random matrix with zero mean and the variance of its entries is N_r/p_r. Therefore, $\mathbf{H}_2$ is a complex Gaussian matrix with the following distribution

$$\mathbf{H}_2 \sim \mathcal{CN}(\hat{\mathbf{H}}_2, \beta \mathbf{I}_{N_r} \otimes \mathbf{I}_{N_d}) \tag{2.3}$$

where $\beta \triangleq N_r/p_r$. It can be seen from (2.3) that the variance of $\mathbf{H}_2$ decreases when p_r increases.

In phase two, the source node transmits an $N_s \times N_s$ pilot matrix $\mathbf{S}_s$ to the relay node. Here we choose the length of $\mathbf{S}_s$ to be N_s to maximize the overall system spectral efficiency. The relay node applies an $N_r \times N_r$ precoding matrix $\mathbf{F}$ and retransmits the linear precoded signal matrix

$$\mathbf{X}_r = \mathbf{F}\mathbf{H}_1 \mathbf{S}_s + \mathbf{F}\mathbf{V}_r \tag{2.4}$$

to the destination node, where $\mathbf{V}_r$ is the $N_r \times N_s$ noise matrix at the relay node. The signal received at the destination node can be written as

$$\mathbf{Y}_d = \mathbf{H}_2 \mathbf{F} \mathbf{H}_1 \mathbf{S}_s + \mathbf{H}_2 \mathbf{F} \mathbf{V}_r + \mathbf{V}_d \tag{2.5}$$

where $\mathbf{V}_d$ is the $N_d \times N_s$ noise matrix at the destination node during phase two.

By vectorizing both sides of (2.5), we obtain

$$\mathbf{y}_d = (\mathbf{S}_s^T \otimes \mathbf{H}_2 \mathbf{F}) \mathbf{h}_1 + (\mathbf{I}_{N_s} \otimes \mathbf{H}_2 \mathbf{F}) \mathbf{v}_r + \mathbf{v}_d \tag{2.6}$$

where $\mathbf{y}_d \triangleq \mathrm{vec}(\mathbf{Y}_d)$, $\mathbf{h}_1 \triangleq \mathrm{vec}(\mathbf{H}_1)$, $\mathbf{v}_r \triangleq \mathrm{vec}(\mathbf{V}_r)$, and $\mathbf{v}_d \triangleq \mathrm{vec}(\mathbf{V}_d)$. To obtain (2.6) from (2.5), we use the property of $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [89].

In this chapter, we assume that the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ satisfy the well-known Kronecker correlation model [90]

$$\mathbf{H}_i = \mathbf{C}_{r_i}^{\frac{1}{2}} \mathbf{H}_{w,i} \mathbf{C}_{t_i}^{\frac{T}{2}}, \qquad i = 1, 2 \tag{2.7}$$

where $\mathbf{C}_{t_i}$ and $\mathbf{C}_{r_i}$, $i = 1, 2$, are channel correlation matrices at the transmit side and the receive side of $\mathbf{H}_i$, respectively, and $\mathbf{H}_{w,i}$, $i = 1, 2$, are Gaussian random matrices with independent and identically distributed (i.i.d.) entries having zero mean and unit variance. We also assume that all noises are i.i.d. additive white Gaussian noise (AWGN) with zero mean and unit variance. The following lemma is useful in this chapter.

LEMMA 2.1 [91]: For $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Theta} \otimes \mathbf{\Phi})$, there is $\mathrm{E}[\mathbf{H}\mathbf{A}\mathbf{H}^H] = \mathrm{tr}(\mathbf{A}\mathbf{\Theta}^T)\mathbf{\Phi}$, and $\mathrm{E}[\mathbf{H}^H\mathbf{A}\mathbf{H}] = \mathrm{tr}(\mathbf{\Phi}\mathbf{A})\mathbf{\Theta}^T$.

2.3 Robust Channel Estimation Algorithm

In this section, we derive the optimal $\mathbf{S}_s$ and $\mathbf{F}$ that minimize the MSE of estimating $\mathbf{H}_1$. Using a linear estimator, the estimated $\mathbf{h}_1$ is given by

$$\hat{\mathbf{h}}_1 = \mathbf{W}\mathbf{y}_d \tag{2.8}$$

where $\mathbf{W}$ is the weight matrix of the linear estimator. Using (2.8), the MSE of estimating $\mathbf{h}_1$ can be written as

$$\begin{aligned} J_1 &= E\big[\mathrm{tr}\big((\mathbf{h}_1 - \hat{\mathbf{h}}_1)(\mathbf{h}_1 - \hat{\mathbf{h}}_1)^H\big)\big] \\ &= \mathrm{tr}\big(\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} - \mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H}\mathbf{W}^H - \mathbf{W}\mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H}^H + \mathbf{W}\mathbf{R}_{\mathbf{y}_d\mathbf{y}_d^H}\mathbf{W}^H\big). \end{aligned} \tag{2.9}$$

From (2.6) we have

$$\mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H} = E[\mathbf{h}_1\mathbf{y}_d^H] = (\mathbf{C}_{t_1}\mathbf{S}_s^*) \otimes (\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H) \tag{2.10}$$

$$\begin{aligned} \mathbf{R}_{\mathbf{y}_d\mathbf{y}_d^H} &= E[\mathbf{y}_d\mathbf{y}_d^H] \\ &= (\mathbf{S}_s^T\mathbf{C}_{t_1}\mathbf{S}_s^*) \otimes (\mathbf{H}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\mathbf{H}_2^H) + \mathbf{I}_{N_s} \otimes (\mathbf{H}_2\mathbf{F}\mathbf{F}^H\mathbf{H}_2^H) + \mathbf{I}_{N_sN_d} \end{aligned} \tag{2.11}$$

$$\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} = E[\mathbf{h}_1\mathbf{h}_1^H] = \mathbf{C}_{t_1} \otimes \mathbf{C}_{r_1}. \tag{2.12}$$

Here we use $\mathbf{h}_1 = (\mathbf{C}_{t_1}^{\frac{1}{2}} \otimes \mathbf{C}_{r_1}^{\frac{1}{2}})\mathbf{h}_{w,1}$ with $\mathbf{h}_{w,1} \triangleq \mathrm{vec}(\mathbf{H}_{w,1})$.

From (2.10)-(2.12), it can be seen that the CSI of $\mathbf{H}_2$ is needed in order to minimize J_1. However, the exact $\mathbf{H}_2$ is unknown in the second phase. In fact,

it is shown in (2.3) that $\mathbf{H}_2$ is a complex Gaussian random matrix with the mean matrix of $\hat{\mathbf{H}}_2$. Obviously, the mismatch between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$ affects the accuracy of the estimation of $\mathbf{H}_1$. To take such mismatch into account, we adopt a statistically robust objective function through averaging J_1 in (2.9) with respect to the distribution of $\mathbf{H}_2$ as

$$E_{\mathbf{H}_2}[J_1] = \mathrm{tr}\big(\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} - E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H}]\mathbf{W}^H - \mathbf{W}E_{\mathbf{H}_2}[\mathbf{R}^H_{\mathbf{h}_1\mathbf{y}_d^H}] + \mathbf{W}E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_d\mathbf{y}_d^H}]\mathbf{W}^H\big). \tag{2.13}$$

The estimator $\mathbf{W}$ which minimizes (2.13) is the linear MMSE estimator [88] given by

$$\mathbf{W} = E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H}]\big(E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_d\mathbf{y}_d^H}]\big)^{-1}. \tag{2.14}$$

Substituting (2.14) back into (2.13), we have

$$E_{\mathbf{H}_2}[J_1] = \mathrm{tr}\big(\mathbf{R}_{\mathbf{h}_1\mathbf{h}_1^H} - E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H}]\big(E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_d\mathbf{y}_d^H}]\big)^{-1} \times E_{\mathbf{H}_2}[\mathbf{R}^H_{\mathbf{h}_1\mathbf{y}_d^H}]\big). \tag{2.15}$$

It can be easily seen from (2.10) that

$$E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{h}_1\mathbf{y}_d^H}] = (\mathbf{C}_{t_1}\mathbf{S}_s^*) \otimes (\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H). \tag{2.16}$$

Using Lemma 2.1, we have from (2.3) that

$$\begin{aligned} E_{\mathbf{H}_2}[\mathbf{R}_{\mathbf{y}_d\mathbf{y}_d^H}] &= (\mathbf{S}_s^T\mathbf{C}_{t_1}\mathbf{S}_s^*) \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H)\mathbf{I}_{N_d}\big) \\ &\quad + \mathbf{I}_{N_s} \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{F}^H)\mathbf{I}_{N_d}\big) + \mathbf{I}_{N_sN_d}. \end{aligned} \tag{2.17}$$

Substituting (2.16) and (2.17) back into (2.15), we obtain that

$$\begin{aligned} E_{\mathbf{H}_2}[J_1] &= \mathrm{tr}\Big(\mathbf{C}_{t_1} \otimes \mathbf{C}_{r_1} - (\mathbf{S}_s^T\mathbf{C}_{t_1}^H\mathbf{C}_{t_1}\mathbf{S}_s^*) \otimes (\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}^H\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H) \\ &\quad \times\big[(\mathbf{S}_s^T\mathbf{C}_{t_1}\mathbf{S}_s^*) \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H)\mathbf{I}_{N_d}\big) \\ &\quad + \mathbf{I}_{N_s} \otimes \big(\hat{\mathbf{H}}_2\mathbf{F}\mathbf{F}^H\hat{\mathbf{H}}_2^H + \mathrm{tr}(\beta\mathbf{F}\mathbf{F}^H)\mathbf{I}_{N_d}\big) + \mathbf{I}_{N_sN_d}\big]^{-1}\Big). \end{aligned} \tag{2.18}$$

The transmission power consumed at the relay node during phase two can be calculated from (2.4) as

$$\begin{aligned} p_r &\triangleq E_{\mathbf{H}_1}\big[\mathrm{tr}\big(\mathbf{F}(\mathbf{H}_1\mathbf{S}_s\mathbf{S}_s^H\mathbf{H}_1^H + N_s\mathbf{I}_{N_r})\mathbf{F}^H\big)\big] \\ &= \mathrm{tr}(\mathbf{S}_s^T\mathbf{C}_{t_1}\mathbf{S}_s^*)\mathrm{tr}(\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H) + N_s\mathrm{tr}(\mathbf{F}\mathbf{F}^H). \end{aligned} \tag{2.19}$$

Using (2.18) and (2.19), the optimal robust $\mathbf{S}_s$ and $\mathbf{F}$ can be found as the solution to the following problem

$$\min_{\mathbf{S}_s,\mathbf{F}} \quad E_{\mathbf{H}_2}[J_1] \tag{2.20}$$

$$\text{s.t.} \quad \text{tr}(\mathbf{S}_s\mathbf{S}_s^H) \leq p_s \tag{2.21}$$

$$\text{tr}(\mathbf{S}_s^T\mathbf{C}_{t_1}\mathbf{S}_s^*)\text{tr}(\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H) + N_s\text{tr}(\mathbf{F}\mathbf{F}^H) \leq p_r \tag{2.22}$$

where (2.21) and (2.22) are the transmission power constraint at the source and the relay node, respectively, and p_s is the power budget available at the source node. The problem (2.20)-(2.22) is complicated with matrices variables. We first show the optimal structure of $\mathbf{S}_s$ and $\mathbf{F}$.

Let us define the following eigenvalue decompositions (EVDs)

$$\mathbf{S}_s^T\mathbf{C}_{t_1}\mathbf{S}_s^* = \mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^H \tag{2.23}$$

$$\hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}\mathbf{F}^H\hat{\mathbf{H}}_2^H = \mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H \tag{2.24}$$

$$\mathbf{C}_{t_1} = \mathbf{U}_{t_1}\mathbf{\Lambda}_{t_1}\mathbf{U}_{t_1}^H \tag{2.25}$$

$$\mathbf{C}_{r_1} = \mathbf{U}_{r_1}\mathbf{\Lambda}_{r_1}\mathbf{U}_{r_1}^H \tag{2.26}$$

where $\mathbf{U}_S$, $\mathbf{U}_F$, $\mathbf{U}_{t_1}$, and $\mathbf{U}_{r_1}$ are the unitary eigenvector matrices, and $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$, $\mathbf{\Lambda}_{t_1}$, and $\mathbf{\Lambda}_{r_1}$ are the diagonal eigenvalue matrices with descending diagonal elements. From (2.23)-(2.24), we can obtain that

$$\mathbf{S}_s^T\mathbf{C}_{t_1}^{\frac{1}{2}} = \mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S, \qquad \hat{\mathbf{H}}_2\mathbf{F}\mathbf{C}_{r_1}^{\frac{1}{2}} = \mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F \tag{2.27}$$

where $\mathbf{Q}_S$ and $\mathbf{Q}_F$ are unitary matrices. Here $\mathbf{C}_{t_1}^{\frac{1}{2}}$ and $\mathbf{C}_{r_1}^{\frac{1}{2}}$ are defined based on (2.25) and (2.26) as

$$\mathbf{C}_{t_1}^{\frac{1}{2}} = \mathbf{U}_{t_1}\mathbf{\Lambda}_{t_1}^{\frac{1}{2}}, \qquad \mathbf{C}_{r_1}^{\frac{1}{2}} = \mathbf{U}_{r_1}\mathbf{\Lambda}_{r_1}^{\frac{1}{2}} \tag{2.28}$$

Let us introduce the SVD of $\hat{\mathbf{H}}_2$ as

$$\hat{\mathbf{H}}_2 = \mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}\mathbf{V}_{H_2}^H \tag{2.29}$$

where $\mathbf{U}_{H_2}$ and $\mathbf{V}_{H_2}$ are the singular vector matrices and $\mathbf{\Sigma}_{H_2}$ is the singular value matrix with descending diagonal elements.

From (2.27) and (2.29) we have

$$\mathbf{S}_s^T = \mathbf{U}_S\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{C}_{t_1}^{-\frac{1}{2}} \tag{2.30}$$

$$\mathbf{F} = \mathbf{V}_{H_2}\mathbf{\Sigma}_{H_2}^{-1}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{C}_{r_1}^{-\frac{1}{2}}. \tag{2.31}$$

Using (2.23)-(2.31), $\bar{J}_1 \triangleq E_{\mathbf{H}_2}[J_1] - \mathrm{tr}(\mathbf{C}_{t_1} \otimes \mathbf{C}_{r_1})$ can be written as

$$\begin{aligned}\bar{J}_1 &= -\mathrm{tr}\big([\mathbf{\Lambda}_S \otimes (\mathbf{\Lambda}_F + a\mathbf{I}_{N_d}) + \mathbf{I}_{N_s} \otimes (\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}}) + b\mathbf{I}_{N_sN_d}]^{-1} \\ &\quad \times(\mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{Q}_S\mathbf{\Lambda}_{t_1}\mathbf{Q}_S^H\mathbf{\Lambda}_S^{\frac{1}{2}}) \otimes (\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}})\big) \end{aligned} \tag{2.32}$$

where

$$\begin{aligned} a &\triangleq \mathrm{tr}(\beta\mathbf{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F) \\ b &\triangleq \mathrm{tr}(\beta\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{U}_F^H\mathbf{U}_{H_2}\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H) + 1. \end{aligned}$$

The power constraints (2.21) and (2.22) can be rewritten as

$$\mathrm{tr}(\mathbf{\Lambda}_S\mathbf{Q}_S\mathbf{\Lambda}_{t_1}^{-1}\mathbf{Q}_S^H) \le p_s \tag{2.33}$$

$$\begin{aligned}&\mathrm{tr}(\mathbf{\Lambda}_S)\mathrm{tr}(\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F\mathbf{U}_F^H\mathbf{U}_{H_2}) \\ &\quad + N_s\mathrm{tr}(\mathbf{\Sigma}_{H_2}^{-2}\mathbf{U}_{H_2}^H\mathbf{U}_F\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{Q}_F\mathbf{\Lambda}_{r_1}^{-1}\mathbf{Q}_F^H\mathbf{\Lambda}_F^{\frac{1}{2}}\mathbf{U}_F^H\mathbf{U}_{H_2}) \le p_r. \end{aligned} \tag{2.34}$$

From (2.32), we see that the mismatch between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$ is considered by matrices $a\mathbf{I}_{N_d}$ and $b\mathbf{I}_{N_sN_d}$. In fact, the objective function in [86] can be viewed as a special case of (2.32) where $a = b = 0$. It can be proven similar to [86] that if $\mathbf{C}_{r_1} = \alpha\mathbf{I}_{N_r}$, then at the optimal $\mathbf{S}_s$, there is $\mathbf{Q}_S = \mathbf{I}_{N_s}$, $\mathbf{Q}_F = \mathbf{I}_{N_r}$, $\mathbf{U}_F = \mathbf{U}_{H_2}$, and $\mathbf{U}_S = \mathbf{I}_{N_s}$. Therefore, the optimal structure of $\mathbf{S}_s$ and $\mathbf{F}$ can be written as

$$\mathbf{S}_s^T = \mathbf{\Lambda}_S^{\frac{1}{2}}\mathbf{C}_{t_1}^{-\frac{1}{2}}, \qquad \mathbf{F} = \alpha^{-\frac{1}{2}}\mathbf{V}_{H_2}\mathbf{\Sigma}_{H_2}^{-1}\mathbf{\Lambda}_F^{\frac{1}{2}}. \tag{2.35}$$

Substituting (2.35) back into (2.32)-(2.34) and let $\lambda_{S,i}$, $\lambda_{F,i}$, $\lambda_{t_1,i}$, $\lambda_{r_1,i}$, and $\sigma_{H_2,i}$ be the ith diagonal element of $\mathbf{\Lambda}_S$, $\mathbf{\Lambda}_F$, $\mathbf{\Lambda}_{t_1}$, $\mathbf{\Lambda}_{r_1}$, and $\mathbf{\Sigma}_{H_2}$, respectively, the problem (2.20)-(2.22) is converted to the following problem with scalar variables

$$\min_{\{\lambda_{S,i}\},\{\lambda_{F,j}\}} \quad -\sum_{i=1}^{N_s}\sum_{j=1}^{N_d}\frac{c_{i,j}}{d_{i,j}} \tag{2.36}$$

$$\text{s.t.} \quad \sum_{i=1}^{N_s}\frac{\lambda_{S,i}}{\lambda_{t_1,i}} \le p_s \tag{2.37}$$

$$\sum_{i=1}^{N_s}\lambda_{S,i}\sum_{j=1}^{N_d}\frac{\lambda_{F,j}}{\sigma_{H_2,j}^2} + \sum_{j=1}^{N_d}\frac{N_s\lambda_{F,j}}{\sigma_{H_2,j}^2\lambda_{r_1,j}} \le p_r \tag{2.38}$$

$$\lambda_{S,i} \ge 0, \qquad i = 1, \cdots, N_s \tag{2.39}$$

$$\lambda_{F,j} \ge 0, \qquad j = 1, \cdots, N_d \tag{2.40}$$

where

$$
\begin{aligned}
c_{i,j} &\triangleq \lambda_{S,i}\lambda_{t_1,i}\lambda_{F,j}\lambda_{r_1,j} \\
d_{i,j} &\triangleq \lambda_{S,i}\lambda_{F,j} + \sum_{j=1}^{N_d} \frac{\beta\lambda_{S,i}\lambda_{F,j}}{\sigma^2_{H_2,j}} + \frac{\lambda_{F,j}}{\lambda_{r_1,j}} + \sum_{j=1}^{N_d} \frac{\beta\lambda_{F,j}}{\lambda_{r_1,j}\sigma^2_{H_2,j}} + 1 \\
\{\lambda_{S,i}\} &\triangleq \{\lambda_{S,i}, i = 1, \cdots, N_s\} \\
\{\lambda_{F,j}\} &\triangleq \{\lambda_{F,j}, j = 1, \cdots, N_d\}
\end{aligned}
$$

The problem (2.36)-(2.40) is non-convex. However, as the optimization of $\{\lambda_{F,j}\}$ is convex when $\{\lambda_{S,i}\}$ is fixed, and vice versa, (at least) a local optimum solution can be found by iteratively optimize $\{\lambda_{F,j}\}$ and $\{\lambda_{S,i}\}$. These two sub-optimizations problem are formulated as follows.

1. Optimizing $\{\lambda_{F,j}\}$ *with fixed* $\{\lambda_{S,i}\}$. The power constraint at the source node is irrelevant as $\{\lambda_{S,i}\}$ is fixed. Therefore, the Karush-Kuhn-Tucker (KKT) conditions of optimizing $\{\lambda_{F,j}\}$ can be written as

$$
\sum_{i=1}^{N_s} \frac{\lambda_{S,i}\lambda_{t_1,i}\lambda_{r_1,j}}{d^2_{i,j}} \left[\sum_{l=1,l\neq j}^{N_d} \frac{\beta\lambda_{F,l}}{\sigma^2_{H_2,l}} \Big(\lambda_{S,i} + \frac{1}{\lambda_{r_1,l}} \Big) + 1 \right] = \mu \left[\sum_{i=1}^{N_s} \frac{\lambda_{S,i}}{\sigma^2_{H_2,j}} + \frac{N_s}{\sigma^2_{H_2,j}\lambda_{r_1,j}} \right] \tag{2.41}
$$

$$
\mu \left(\sum_{i=1}^{N_s} \lambda_{S,i} \sum_{j=1}^{N_d} \frac{\lambda_{F,j}}{\sigma^2_{H_2,j}} + \sum_{j=1}^{N_d} \frac{N_s\lambda_{F,j}}{\sigma^2_{H_2,j}\lambda_{r_1,j}} - p_r \right) = 0 \tag{2.42}
$$

where $\mu \geq 0$ is the Lagrange multiplier such that equation (2.42) holds. With any fixed $\{\lambda_{S,i}\}$, μ, and $\lambda_{F,l}$, $l = 1, \cdots, N_d, l \neq j$, the non-negative $\lambda_{F,j}$ can be derived using the bi-section search, since the left-hand-side (LHS) of (2.41) is a monotonically decreasing function of $\lambda_{F,j}$. Note that (2.41) depends on $\lambda_{F,j}$, $j = 1, \cdots, N_d$, hence, the value of $\{\lambda_{F,j}\}$ needs to be updated each time a new $\lambda_{F,j}$ is obtained. To find the optimal value of μ, an outer bi-section loop is used as the LHS of (2.38) is an increasing function of $\lambda_{F,j}$, and $\lambda_{F,j}$ is a monotonically decreasing function of μ.

2. Optimizing $\{\lambda_{S,i}\}$ *with fixed* $\{\lambda_{F,j}\}$. The KKT conditions of this subproblem can be written as

$$\sum_{j=1}^{N_d} \frac{\lambda_{t_1,i}\lambda_{F,j}\lambda_{r_1,j}\left(\frac{\lambda_{F,j}}{\lambda_{r_1,j}} + \beta \sum_{l=1}^{N_d} \frac{\lambda_{F,l}}{\lambda_{r_1,l}\sigma^2_{H_2,l}} + 1\right)}{d^2_{i,j}} = \frac{\nu_1}{\lambda_{t_1,i}} + \nu_2 \sum_{j=1}^{N_d} \frac{\lambda_{F,j}}{\sigma^2_{H_2,j}} \quad (2.43)$$

$$\nu_1 \left(\sum_{i=1}^{N_s} \frac{\lambda_{S,i}}{\lambda_{t_1,i}} - p_s \right) = 0 \quad (2.44)$$

$$\nu_2 \left(\sum_{i=1}^{N_s} \lambda_{S,i} \sum_{j=1}^{N_d} \frac{\lambda_{F,j}}{\sigma^2_{H_2,j}} + \sum_{j=1}^{N_d} \frac{N_s \lambda_{F,j}}{\sigma^2_{H_2,j}\lambda_{r_1,j}} - p_r \right) = 0 \quad (2.45)$$

where $\nu_1 \geq 0$ and $\nu_2 \geq 0$ are the Lagrange multipliers. For any fixed $\{\lambda_{F,j}\}$, ν_1 and ν_2, the non-negative $\lambda_{S,i}$ can be found by a bi-section search for all i. This is because the LHS of (2.43) is a monotonically decreasing function of $\lambda_{S,i}$. Note that the LHS of both (2.37) and (2.38) are increasing function of $\lambda_{S,i}$, and $\lambda_{S,i}$ is a monotonically decreasing function of both ν_1 and ν_2. Generally, to find the optimal value of ν_1 and ν_2, a 2-D bi-section loop search is required. However, if only one of the constraints is active (i.e. only one of the constraints satisfies the equality), then only 1-D bisection loop search is required to find the corresponding multiplier for the constraint as the other multiplier is zero. If both constraints are inactive, then a 2-D bi-section loop is required to determine the optimal value of ν_1 and ν_2.

2.4 Numerical Examples

In this section, we study the performance of the proposed channel estimation algorithm through numerical simulations. We compare the proposed approach with the algorithm developed in [86] (denoted as "imperfect H_2") where $\hat{\mathbf{H}}_2$ is used in the second phase to estimate $\mathbf{H}_1$. As a benchmark, the performance of channel estimation algorithm with exactly known $\mathbf{H}_2$ is also studied.

In the simulations, for simplicity, we set $N_s = N_r = N_d = N$. The channel correlation matrices are modelled as $[\mathbf{C}_{t_i}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, $[\mathbf{C}_{r_2}]_{m,n} = \rho^{|m-n|}$, where ρ is the correlation coefficient, and $\mathbf{C}_{r_1} = \mathbf{I}_{N_r}$. For each channel realization, the normalized mean-squared error (NMSE) of channel estimation for all three algorithms is calculated as $\|\mathbf{H}_1 - \hat{\mathbf{H}}_1\|^2_F / N_s N_r$. All simulation results are averaged

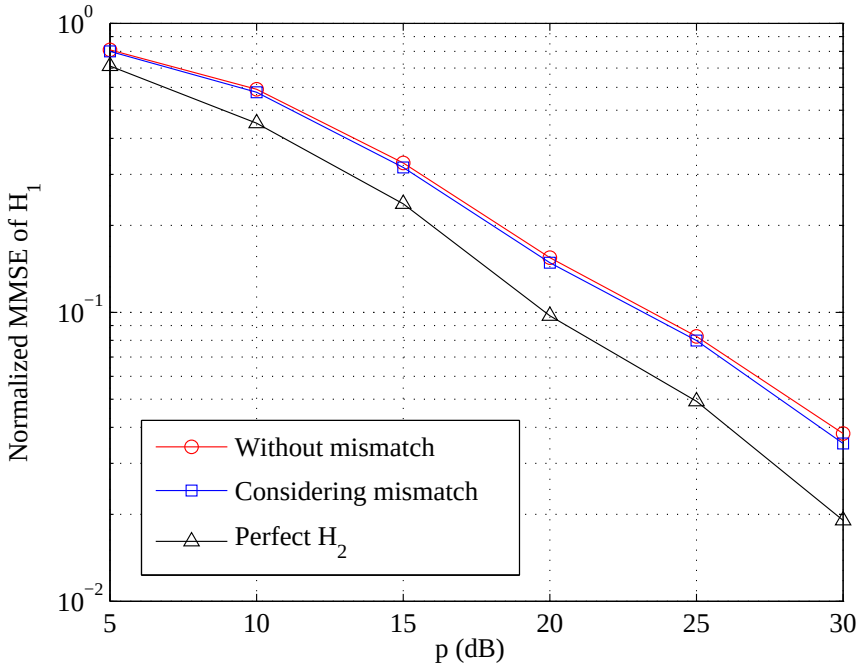

Figure 2.2: Example 2.1: Normalized MSE of $\mathbf{H}_1$ versus p for $N = 2$ and $\rho = 0.2$

over 100 random channel realizations.

In the first example, we show the normalized MSE of estimating $\mathbf{H}_1$ when $N = 2$ and $\rho = 0.2$ in Fig. 2.2. A different number of antennas $N = 4$ and normalized correlation coefficient $\rho = 0.8$ are used for the next scenario and the results are shown in Fig. 2.3. Note that for both scenarios, the power at the source node is assumed to be the same as the power at the relay node, i.e. $p_s = p_r = p$.

In the second example, we show the normalized MSE of estimating $\mathbf{H}_1$ in Fig. 2.4 when the power at the source node p_s is fixed at 20dB while the power at the relay node p_r is varied from 5dB to 30dB. The number of antennas and the normalized correlation coefficient are set to be $N = 2$ and $\rho = 0.8$ respectively.

From the simulation results, it is obvious that by considering the mismatch between $\hat{\mathbf{H}}_2$ and $\mathbf{H}_2$ in the algorithm, the performance of the algorithm has been improved without the need of greater computation effort. The simulations are executed with different parameters to examine the effectiveness of the algorithm, and all results show an improvement in the estimation of channel matrices.

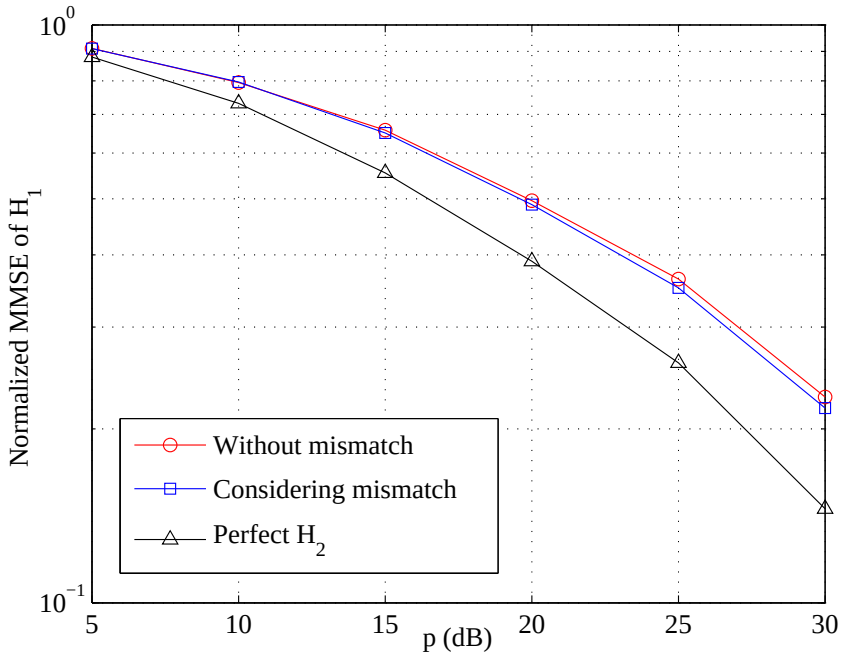

Figure 2.3: Example 2.1: Normalized MSE of $\mathbf{H}_1$ versus p for $N = 4$ and $\rho = 0.8$

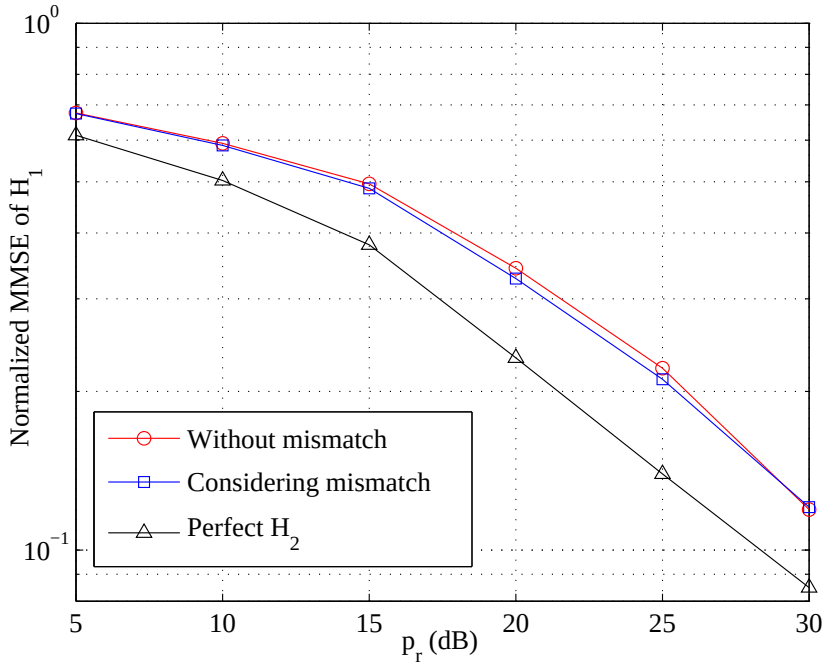

Figure 2.4: Example 2.2: Normalized MSE versus p_r. $N = 2$, $\rho = 0.8$.

2.5 Conclusions

The effect of the mismatch between the estimated and true relay-destination channel on the performance of the LMMSE-based MIMO relay channel estimation algorithm has been investigated in this chapter. It has been proven that the robust channel estimation algorithm performs better compared to the channel estimation algorithm proposed in [86] that does not take the mismatch into the consideration. Moreover, the robust channel estimation algorithm does not require greater computational effort.

Chapter 3

Frequency-Flat Two-Way MIMO Relay Systems

Two-way relay systems are known to be capable of providing higher spectral efficiency compared with conventional one-way relay systems. In this chapter, we focus on the channel estimation problem for two-way MIMO relay communication networks. Section 3.1 introduces background knowledge on channel estimation for two-way MIMO relay systems. In Section 3.2, the system model of a two-way MIMO relay network is presented. Two channel estimation algorithms, namely the superimposed channel training scheme and the two-stage channel estimation algorithm, are proposed and developed in Section 3.3 and Section 3.4, respectively. For both algorithms, we derive the optimal structure of the source and relay training sequences which minimize the MSE of channel estimation. In the superimposed channel training scheme, the power allocation between the source and relay training sequences is optimized. For the two-stage channel estimation algorithm, we optimize the power allocation at the relay node between two stages to improve the performance of the algorithm. Section 3.5 shows numerical examples to demonstrate the performance of the proposed algorithms. Conclusions are drawn in Section 3.6. This chapter ends with the proof of Theorem 3.1 in Section 3.A.

3.1 Background Information

The channel estimation algorithms in Chapter 2 are developed for one-way relay systems, where a source node sends signals to a destination node through relay node(s). In two-way relay systems, two source nodes exchange their information through assisting relay node(s). Initially studied by Shannon in [92], two-way relay systems are getting more attention recently as they have higher spectral efficiency compared with one-way relay systems. For two-way MIMO relay systems, the joint source and relay optimization is recently investigated in [93]-[95] assuming the channel matrices are known. Channel estimation issue is not discussed in [93]-[95].

The channel estimation problem becomes more complicated in two-way relay systems and several algorithms have been proposed in [96]-[98]. ML and linear maximum SNR channel estimation techniques have been introduced in [96], while block-based training and pilot-tone based training algorithms are presented in [97]. However, the algorithms in [96] and [97] are based on the assumption that each node is equipped with single antenna only, and extension to MIMO systems is not straightforward.

For two-way MIMO relay systems, cascaded channel estimation and individual channel estimation algorithms have been proposed in [98]. The cascaded channel estimation is easy to implement but does not provide the second-hop CSI, which is necessary for system optimization [95]. In the individual channel estimation algorithm, the first-hop CSI is first estimated at the relay node and then fed-forward to the receive nodes. However, this algorithm requires the relay node to be capable of performing advanced signal processing, and therefore, increases the cost and complexity at the relay node.

The major challenge in channel estimation for two-way MIMO relay systems is to obtain the instantaneous CSI of both the first-hop and second-hop links with a minimal amount of signal processing at the relay node. In this chapter, we address this challenge by proposing two algorithms: the superimposed channel training scheme and the two-stage channel estimation algorithm. Both algorithms are developed for two-way MIMO relay systems with frequency-flat fading channels. In the superimposed channel training algorithm, both source nodes transmit their

training sequence simultaneously to the relay node in the first time block. The relay node then amplifies the received signals and superimposes its own training sequence, before transmitting the superimposed signals to both receive nodes. By exploiting the training sequences from the source and relay nodes, the individual CSI of the first-hop and second-hop links can be successfully estimated.

In the two-stage channel estimation algorithm, both source nodes are silent at the first stage, while the relay node broadcasts a pilot matrix to both receive nodes for the estimation of the channel matrices from the relay node to the receive nodes (second-hop links). During the second stage, both source nodes transmit their training sequence simultaneously to the relay node, and the relay node amplifies the received signals and forwards them to the receive nodes. Then, the channel matrices from the source nodes to the relay node (first-hop links) are estimated by exploiting the second-hop channel matrices estimated at the first stage. We would like to mention that although the estimation of the second-hop channels at the first stage is similar to the problem in [61] and [62], an efficient estimation of the first-hop channels is a non-conventional problem.

For both algorithms, we derive the structure of the optimal training sequences that minimize the sum MSE of channel estimation. In particular, we show that the optimal training matrix for each hop matches the eigenvector matrix of the correlation matrix of the MIMO channel at that hop. Moreover, in the superimposed channel training scheme, the power allocation between the source and relay training sequences is optimized. For the two-stage channel estimation algorithm, we optimize the power allocation at the relay node between two stages to minimize the MSE of channel estimation. The performance of the superimposed channel training scheme and the two-stage channel estimation algorithm are demonstrated and compared through numerical examples.

3.2 Frequency-Flat Two-Way MIMO Relay System Model

We consider a three-node two-way frequency-flat MIMO communication system where node 1 and node 2 exchange information through a relay node as shown in

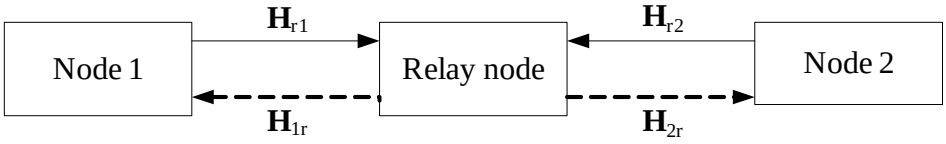

Figure 3.1: Block diagram of a two-way MIMO relay communication system.

Fig. 3.1. Nodes 1 and 2 are equipped with N_1 and N_2 antennas, respectively, while the relay node has N_r antennas. For $i = 1, 2$, $\mathbf{H}_{ir}$ is the $N_i \times N_r$ channel matrix from the relay node to node i, while $\mathbf{H}_{ri}$ denotes the $N_r \times N_i$ channel matrix from node i to the relay node. In this chapter, we consider that all nodes are operating in the half-duplex mode, i.e., one node cannot transmit and receive at the same time. Since in a two-way relay system, both source nodes transmit signals to the relay node at the first time slot, they cannot receive signals from each other. Therefore, there is no direct link between two source nodes. The half-duplex mode has been widely used in two-way relay communications [95]-[97].

In this chapter, we assume that the channel matrices $\mathbf{H}_{ri}$ and $\mathbf{H}_{ir}$ satisfy the well-known Gaussian-Kronecker model [90], where $\mathbf{H}_{ri}$ and $\mathbf{H}_{ir}$ are complex-valued Gaussian random matrices with

$$\mathbf{H}_{ri} \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}_{ri} \otimes \mathbf{R}_{ri}), \qquad \mathbf{H}_{ir} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_r \otimes \mathbf{R}_{ir}), \qquad i = 1, 2. \tag{3.1}$$

Here $\mathbf{T}_{ri}$ and $\mathbf{R}_{ri}$ denote the $N_i \times N_i$ and $N_r \times N_r$ covariance matrix at the transmit and receive side of $\mathbf{H}_{ri}$, respectively, while $\mathbf{C}_r$ and $\mathbf{R}_{ir}$ stand for the $N_r \times N_r$ and $N_i \times N_i$ covariance matrix at the transmit and receive side of $\mathbf{H}_{ir}$, respectively. In other words, from (3.1) we have

$$\mathbf{H}_{ri} = \mathbf{A}_{ri}\mathbf{H}_{ri,\mathrm{w}}\mathbf{B}_{ri}^H, \qquad \mathbf{H}_{ir} = \mathbf{A}_{ir}\mathbf{H}_{ir,\mathrm{w}}\mathbf{K}_r^H, \qquad i = 1, 2 \tag{3.2}$$

where $\mathbf{A}_{ri}\mathbf{A}_{ri}^H = \mathbf{R}_{ri}$, $\mathbf{B}_{ri}\mathbf{B}_{ri}^H = \mathbf{T}_{ri}^T$, $\mathbf{A}_{ir}\mathbf{A}_{ir}^H = \mathbf{R}_{ir}$, $\mathbf{K}_r\mathbf{K}_r^H = \mathbf{C}_r^T$, $i = 1, 2$, $\mathbf{H}_{ri,\mathrm{w}}$ and $\mathbf{H}_{ir,\mathrm{w}}$ are $N_r \times N_i$ and $N_i \times N_r$ Gaussian random matrices with i.i.d. zero mean and unit variance entries. We assume that $\mathbf{H}_{ri,\mathrm{w}}$ and $\mathbf{H}_{ir,\mathrm{w}}$, $i = 1, 2$, are statistically independent of each other.

3.3 Superimposed Channel Training Algorithm

In this section, we develop a superimposed channel training algorithm to estimate $\mathbf{H}_{ri}$ and $\mathbf{H}_{ir}$, $i = 1, 2$. This channel estimation scheme is completed in two time

blocks. In the first time block, the source node i transmits an $N_i \times L$ training signal matrix $\mathbf{S}_i$, where L is the length of the training sequence. The $N_r \times L$ received signal matrix $\mathbf{Y}_r$ at the relay node is given by

$$\mathbf{Y}_r = \sum_{i=1}^{2} \mathbf{H}_{ri}\mathbf{S}_i + \mathbf{V}_r \tag{3.3}$$

where $\mathbf{V}_r$ is an $N_r \times L$ noise matrix at the relay node.

In the second time block, the relay node amplifies $\mathbf{Y}_r$ and superimposes its own training matrix $\mathbf{S}_r$. Thus, the $N_r \times L$ signal matrix transmitted by the relay node can be written as

$$\mathbf{X}_r = \sqrt{\alpha}\mathbf{Y}_r + \mathbf{S}_r \tag{3.4}$$

where $\alpha > 0$ is the relay amplifying factor. From (3.3) and (3.4), the $N_i \times L$ received signal matrix at node i is given by

$$\begin{aligned}\mathbf{Y}_i &= \mathbf{H}_{ir}\mathbf{X}_r + \mathbf{V}_i \\ &= \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{H}_{ri}\mathbf{S}_i + \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{H}_{r\bar{i}}\mathbf{S}_{\bar{i}} + \mathbf{H}_{ir}\mathbf{S}_r + \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{V}_r + \mathbf{V}_i, \quad i = 1, 2\end{aligned} \tag{3.5}$$

where $\mathbf{V}_i$ is an $N_i \times L$ noise matrix at node i. Here, $\bar{i} = 2$ for $i = 1$, and $\bar{i} = 1$ for $i = 2$. The main idea of the superimposed channel training algorithm is to use $\mathbf{S}_r$ to estimate the second-hop channels $\mathbf{H}_{ir}$. Then the first-hop channels $\mathbf{H}_{rj}$, $j = i, \bar{i}$, can be estimated by exploiting $\mathbf{S}_j$ and the estimated $\mathbf{H}_{ir}$.

Let us introduce the EVD of $\mathbf{T}_{ri}^T$ as $\mathbf{U}_i\mathbf{\Lambda}_i\mathbf{U}_i^H$, $i = 1, 2$, and the EVD of $\mathbf{C}_r^T$ as $\mathbf{U}_r\mathbf{\Lambda}_r\mathbf{U}_r^H$. Then we have $\mathbf{B}_{ri}^H = \mathbf{\Pi}_i\mathbf{\Lambda}_i^{\frac{1}{2}}\mathbf{U}_i^H$, $i = 1, 2$, and $\mathbf{K}_r^H = \mathbf{\Pi}_r\mathbf{\Lambda}_r^{\frac{1}{2}}\mathbf{U}_r^H$, where $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_r$ are arbitrary $N_i \times N_i$ and $N_r \times N_r$ unitary matrix, respectively. Using (3.2), we can rewrite (3.5) as

$$\mathbf{Y}_i = \sqrt{\alpha}\mathbf{G}_{ii}\tilde{\mathbf{S}}_i + \sqrt{\alpha}\mathbf{G}_{i\bar{i}}\tilde{\mathbf{S}}_{\bar{i}} + \tilde{\mathbf{H}}_{ir}\tilde{\mathbf{S}}_r + \bar{\mathbf{V}}_i, \qquad i = 1, 2 \tag{3.6}$$

where $\tilde{\mathbf{S}}_r \triangleq \mathbf{U}_r^H\mathbf{S}_r$,

$$\begin{aligned}\mathbf{G}_{ij} \triangleq \mathbf{H}_{ir}\tilde{\mathbf{H}}_{rj}, \quad \tilde{\mathbf{S}}_j &\triangleq \mathbf{U}_j^H\mathbf{S}_j, \quad \tilde{\mathbf{H}}_{rj} = \mathbf{H}_{rj}\mathbf{U}_j, \quad j = i, \bar{i}, \\ \tilde{\mathbf{H}}_{ir} &\triangleq \mathbf{H}_{ir}\mathbf{U}_r, \quad i = 1, 2\end{aligned} \tag{3.7}$$

and

$$\bar{\mathbf{V}}_i \triangleq \sqrt{\alpha}\mathbf{H}_{ir}\mathbf{V}_r + \mathbf{V}_i, \qquad i = 1, 2 \tag{3.8}$$

is the equivalent noise matrix at node i. In the following, we develop an algorithm to estimate $\tilde{\mathbf{H}}_{ir}$ and $\mathbf{G}_{ij}$ in (3.6). Then an estimate of $\mathbf{H}_{ir}$ and $\mathbf{H}_{rj}$ can be obtained from (3.7) as $\hat{\mathbf{H}}_{ir} = \breve{\mathbf{H}}_{ir}\mathbf{U}_r^H$ and $\hat{\mathbf{H}}_{rj} = \hat{\mathbf{H}}_{ir}^{\dagger}\breve{\mathbf{G}}_{ij}\mathbf{U}_j^H$, $j = i, \bar{i}$, where $\breve{\mathbf{H}}_{ir}$ and $\breve{\mathbf{G}}_{ij}$ are the estimates of $\tilde{\mathbf{H}}_{ir}$ and $\mathbf{G}_{ij}$, respectively.

By vectorizing both sides of (3.6), we obtain

$$\begin{aligned} \mathbf{y}_i &= \left[\sqrt{\alpha}\tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i},\ \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i},\ \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}\right]\left[\mathbf{g}_{ii}^T,\ \mathbf{g}_{\bar{i}i}^T,\ \tilde{\mathbf{h}}_{ir}^T\right]^T + \bar{\mathbf{v}}_i \quad (3.9) \\ &\triangleq \mathbf{M}_i\boldsymbol{\gamma}_i + \bar{\mathbf{v}}_i, \qquad i = 1, 2 \quad (3.10) \end{aligned}$$

where for $i = 1, 2$, $\mathbf{y}_i \triangleq \mathrm{vec}(\mathbf{Y}_i)$, $\mathbf{g}_{ij} \triangleq \mathrm{vec}(\mathbf{G}_{ij})$, $j = i, \bar{i}$, $\tilde{\mathbf{h}}_{ir} \triangleq \mathrm{vec}(\tilde{\mathbf{H}}_{ir})$, $\bar{\mathbf{v}}_i \triangleq \mathrm{vec}(\bar{\mathbf{V}}_i)$. Here the identity of $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ [89] has been used to obtain (3.9) from (3.6). In (3.10), $\boldsymbol{\gamma}_i \triangleq \left[\mathbf{g}_{ii}^T,\ \mathbf{g}_{i\bar{i}}^T,\ \tilde{\mathbf{h}}_{ir}^T\right]^T$ is the vector of unknown variables at node i with a dimension of $Q_i \triangleq N_i(N_i + N_{\bar{i}} + N_r)$, and $\mathbf{M}_i \triangleq \left[\sqrt{\alpha}\tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i},\ \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i},\ \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}\right]$ has a dimension of $TN_i \times Q_i$.

Due to its simplicity, a linear MMSE estimator [88] is applied at node i to estimate $\boldsymbol{\gamma}_i$. We have

$$\hat{\boldsymbol{\gamma}}_i = \mathbf{W}_i^H\mathbf{y}_i, \qquad i = 1, 2 \quad (3.11)$$

where $\hat{\boldsymbol{\gamma}}_i$ stands for an estimation of $\boldsymbol{\gamma}_i$ and $\mathbf{W}_i$ is the weight matrix of the MMSE estimator. It can be seen from (3.11) that since a linear estimator is used, there is $L \geq N_1 + N_2 + N_r$, and the MSE of estimating $\boldsymbol{\gamma}_i$ can be written as

$$\begin{aligned} \mathrm{MSE}_i &= \mathrm{E}\left[\mathrm{tr}\left((\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)^H\right)\right] \\ &= \mathrm{tr}\Big(\left(\mathbf{W}_i^H\mathbf{M}_i - \mathbf{I}_{Q_i}\right)\mathbf{R}_{\gamma_i}\left(\mathbf{W}_i^H\mathbf{M}_i - \mathbf{I}_{Q_i}\right)^H + \mathbf{W}_i^H\mathbf{R}_{\bar{v}_i}\mathbf{W}_i\Big), \quad i = 1, 2 \end{aligned} \quad (3.12)$$

where $\mathbf{R}_{\gamma_i} \triangleq \mathrm{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^H]$ is the covariance matrix of $\boldsymbol{\gamma}_i$ and $\mathbf{R}_{\bar{v}_i} \triangleq \mathrm{E}[\bar{\mathbf{v}}_i\bar{\mathbf{v}}_i^H]$ is the noise covariance matrix. Using (3.2), (3.8), and Lemma 2.1, we obtain that

$$\mathbf{R}_{\bar{v}_i} = \mathbf{I}_L \otimes \left(\alpha\mathrm{tr}(\mathbf{K}_r^H\mathbf{K}_r)\mathbf{A}_{ir}\mathbf{A}_{ir}^H + \mathbf{I}_{N_i}\right) = \mathbf{I}_L \otimes \left(\alpha\mathrm{tr}(\mathbf{C}_r^T)\mathbf{R}_{ir} + \mathbf{I}_{N_i}\right), \qquad i = 1, 2. \quad (3.13)$$

Using Lemma 2.1, $\mathbf{R}_{\gamma_i}$ can be calculated as follows. First, the mth column of $\mathbf{G}_{ij}$ is given by $[\mathbf{G}_{ij}]_m = \lambda_{j,m}^{\frac{1}{2}}\mathbf{A}_{ir}\mathbf{H}_{ir,\mathrm{w}}\mathbf{K}_r^H\mathbf{A}_{rj}\mathbf{H}_{rj,\mathrm{w}}[\boldsymbol{\Pi}_j]_m$, $m = 1, \cdots, N_j$, where $\lambda_{j,m}$ is the mth diagonal element of $\boldsymbol{\Lambda}_j$, and $[\boldsymbol{\Pi}_j]_m$ is the mth column of $\boldsymbol{\Pi}_j$. Since $\mathbf{H}_{ir,\mathrm{w}}$ and $\mathbf{H}_{rj,\mathrm{w}}$ are independent, the covariance matrix of $[\mathbf{G}_{ij}]_m$ can be

calculated as

$$\begin{aligned}\mathrm{E}\big[[\mathbf{G}_{ij}]_m[\mathbf{G}_{ij}]_m^H\big] &= \lambda_{j,m}\mathrm{tr}(\mathbf{K}_r^H\mathbf{A}_{rj}\mathbf{A}_{rj}^H\mathbf{K}_r)\mathbf{A}_{ir}\mathbf{A}_{ir}^H \\ &= \lambda_{j,m}b_j\mathbf{R}_{ir}, \quad m=1,\cdots,N_j, \quad j=i,\bar{i} \end{aligned} \tag{3.14}$$

where $b_j \triangleq \mathrm{tr}(\mathbf{R}_{rj}\mathbf{C}_r^T)$. Second, the covariance matrix of the mth column of $\tilde{\mathbf{H}}_{ir}$, denoted as $[\tilde{\mathbf{H}}_{ir}]_m$, is given by

$$\mathrm{E}\big[[\tilde{\mathbf{H}}_{ir}]_m[\tilde{\mathbf{H}}_{ir}]_m^H\big] = \lambda_{r,m}\mathbf{R}_{ir}, \qquad m=1,\cdots,N_r \tag{3.15}$$

where $\lambda_{r,m}$ is the mth diagonal element of $\mathbf{\Lambda}_r$. From (3.14) and (3.15), $\mathbf{R}_{\gamma_i}$ can be written as

$$\mathbf{R}_{\gamma_i} = \mathrm{Bdiag}\big[\mathbf{\Lambda}_i \otimes b_i\mathbf{R}_{ir}, \;\; \mathbf{\Lambda}_{\bar{i}} \otimes b_{\bar{i}}\mathbf{R}_{ir}, \;\; \mathbf{\Lambda}_r \otimes \mathbf{R}_{ir}\big], \quad i=1,2. \tag{3.16}$$

The matrix $\mathbf{W}_i$ minimizing MSE_i in (3.12) is given by

$$\mathbf{W}_i = \big(\mathbf{M}_i\mathbf{R}_{\gamma_i}\mathbf{M}_i^H + \mathbf{R}_{\bar{v}_i}\big)^{-1}\mathbf{M}_i\mathbf{R}_{\gamma_i}, \qquad i=1,2. \tag{3.17}$$

Substituting (3.17) back into (3.12), and using the matrix inversion lemma of $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B}+\mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}$, the MSE of estimating $\boldsymbol{\gamma}_i$ can be obtained as

$$\mathrm{MSE}_i = \mathrm{tr}\Big(\big[\mathbf{R}_{\gamma_i}^{-1} + \mathbf{M}_i^H\mathbf{R}_{\bar{v}_i}^{-1}\mathbf{M}_i\big]^{-1}\Big), \quad i=1,2. \tag{3.18}$$

The transmission power consumed at nodes 1 and 2 is

$$\mathrm{tr}(\mathbf{S}_i\mathbf{S}_i^H) = \mathrm{tr}(\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H), \qquad i=1,2. \tag{3.19}$$

From (3.4), the power consumed at the relay node is given by

$$\begin{aligned}&\alpha\mathrm{E}\left[\mathrm{tr}\Big(\sum_{i=1}^{2}\mathbf{H}_{ri}\mathbf{S}_i\mathbf{S}_i^H\mathbf{H}_{ri}^H + \mathbf{I}_{N_r}\Big)\right] + \mathrm{tr}(\mathbf{S}_r\mathbf{S}_r^H) \\ &= \alpha N_r + \alpha\sum_{i=1}^{2}\mathrm{tr}(\mathbf{\Lambda}_i\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H)\mathrm{tr}(\mathbf{R}_{ri}) + \mathrm{tr}(\tilde{\mathbf{S}}_r\tilde{\mathbf{S}}_r^H). \end{aligned} \tag{3.20}$$

From (3.18)-(3.20), the optimal training matrices and the optimal α can be designed by solving the following optimization problem

$$\min_{\alpha,\tilde{\mathbf{S}}_1,\tilde{\mathbf{S}}_2,\tilde{\mathbf{S}}_r} \quad \sum_{i=1}^{2}\mathrm{tr}\Big(\big[\mathbf{R}_{\gamma_i}^{-1} + \mathbf{M}_i^H\mathbf{R}_{\bar{v}_i}^{-1}\mathbf{M}_i\big]^{-1}\Big) \tag{3.21}$$

$$\text{s.t.} \quad \mathrm{tr}(\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H) \le p_i, \qquad i=1,2 \tag{3.22}$$

$$\alpha\Big[N_r + \sum_{i=1}^{2}\mathrm{tr}(\mathbf{\Lambda}_i\tilde{\mathbf{S}}_i\tilde{\mathbf{S}}_i^H)\mathrm{tr}(\mathbf{R}_{ri})\Big] + \mathrm{tr}(\tilde{\mathbf{S}}_r\tilde{\mathbf{S}}_r^H) \le p_r \tag{3.23}$$

where p_i is the transmission power available at node i, $i = 1, 2$, and p_r is the transmission power available at the relay node. The following theorem establishes the optimal structure of $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_r$ as solution to the problem (3.21)-(3.23).

THEOREM 3.1: The optimal training sequences $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_r$ satisfy $\mathbf{S}_i\mathbf{S}_i^H = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{U}_i^H$, $i = 1, 2, r$, and $\mathbf{S}_i\mathbf{S}_j^H = \mathbf{0}$, $i, j = 1, 2, r$, $i \neq j$, where $\mathbf{\Sigma}_i$, $i = 1, 2, r$, is an $N_i \times N_i$ diagonal matrices.

PROOF: See Appendix 3.A. □

The optimal structure of $\mathbf{S}_i$, $i = 1, 2, r$, can be obtained from Theorem 3.1 as $\mathbf{S}_i = \mathbf{U}_i\mathbf{\Sigma}_i^{\frac{1}{2}}\mathbf{\Omega}_i$, where $\mathbf{\Omega}_i$ is an $N_i \times L$ semi-unitary matrix satisfying $\mathbf{\Omega}_i\mathbf{\Omega}_i^H = \mathbf{I}_{N_i}$, $i = 1, 2, r$, and $\mathbf{\Omega}_i\mathbf{\Omega}_j^H = \mathbf{0}$, $i, j = 1, 2, r$, $i \neq j$. Such $\mathbf{\Omega}_r$, $\mathbf{\Omega}_1$, and $\mathbf{\Omega}_2$ can be easily constructed, for example, from the normalized discrete Fourier transform (DFT) matrix when $L \geq N_1 + N_2 + N_r$.

Interestingly, it can be seen that the optimal training matrix at node i matches the eigenvector matrix of the transmitter correlation matrix of $\mathbf{H}_{ri}$, and the optimal training matrix at the relay node matches the eigenvector matrix of $\mathbf{C}_r^T$. Using Theorem 3.1 and (3.81) in Appendix 3.A, the problem (3.21)-(3.23) is equivalently converted to the following problem

$$\min_{\alpha, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_r} \quad \sum_{i=1}^{2} \mathrm{tr}\left(\sum_{j=1}^{2} \left[\mathbf{D}_{ij} + \alpha\mathbf{\Sigma}_j \otimes \mathbf{D}_{ri}\right]^{-1} + \left[\mathbf{D}_{si} + \mathbf{\Sigma}_r \otimes \mathbf{D}_{ri}\right]^{-1} \right) \tag{3.24}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{\Sigma}_i) \leq p_i, \qquad i = 1, 2 \tag{3.25}$$

$$\alpha N_r + \alpha \sum_{i=1}^{2} \mathrm{tr}(\mathbf{\Lambda}_i\mathbf{\Sigma}_i)\mathrm{tr}(\mathbf{R}_{ri}) + \mathrm{tr}(\mathbf{\Sigma}_r) \leq p_r \tag{3.26}$$

$$\alpha > 0, \qquad \mathbf{\Sigma}_i \geq 0, \quad i = 1, 2, r \tag{3.27}$$

where for a matrix $\mathbf{A}$, $\mathbf{A} \geq 0$ means that $\mathbf{A}$ is a positive semi-definite (PSD) matrix. Using the definition of $\mathbf{D}_{ij}$, $\mathbf{D}_{si}$, and $\mathbf{D}_{ri}$ in (3.78) in Appendix 3.A, the problem (3.24)-(3.27) can be equivalently rewritten as the following problem with

scalar variables

$$\min_{\alpha,\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_r} \quad \sum_{i=1}^{2}\left(\sum_{j=1}^{2}\sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\left[\frac{1}{b_j\lambda_{j,m}\delta_{i,n}}+\alpha d_{i,n}\sigma_{j,m}\right]^{-1} + \sum_{m=1}^{N_r}\sum_{n=1}^{N_i}\left[\frac{1}{\lambda_{r,m}\delta_{i,n}}+d_{i,n}\sigma_{r,m}\right]^{-1}\right) \tag{3.28}$$

$$\text{s.t.} \quad \sum_{m=1}^{N_i}\sigma_{i,m}\le p_i, \quad i=1,2 \tag{3.29}$$

$$\alpha N_r+\alpha\sum_{i=1}^{2}\left(\sum_{m=1}^{N_i}\lambda_{i,m}\sigma_{i,m}\mathrm{tr}(\mathbf{R}_{ri})\right)+\sum_{m=1}^{N_r}\sigma_{r,m}\le p_r \tag{3.30}$$

$$\alpha>0, \;\; \sigma_{i,m}\ge 0, \qquad m=1,\cdots,N_i, \qquad i=1,2,r \tag{3.31}$$

where $d_{i,n}\triangleq 1/(\alpha\mathrm{tr}(\mathbf{C}_r^T)\delta_{i,n}+1)$, $n=1,\cdots,N_i$, $i=1,2$, $\boldsymbol{\sigma}_i\triangleq\left[\sigma_{i,1},\cdots,\sigma_{i,N_i}\right]^T$, $i=1,2,r$, $\delta_{i,m}$, $i=1,2$, is the mth diagonal element of $\boldsymbol{\Delta}_i$, and $\lambda_{i,m}$, $\sigma_{i,m}$, $i=1,2,r$, are the mth diagonal element of $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Sigma}_i$, respectively.

Given that $b_j,\lambda_{j,m},\delta_{i,n},d_{i,n}$, and $\lambda_{r,m}$ are known variables with fixed α, the objective function (3.28) can be rewritten as

$$\min_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_r}\sum_{i=1}^{2}\left(\sum_{j=1}^{2}sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\frac{1}{a_{i,j,m,n}+c_{i,n}\sigma_{j,m}}+\sum_{m=1}^{N_r}\sum_{n=1}^{N_i}\frac{1}{g_{i,m,n}+d_{i,n}\sigma_{r,m}}\right)$$

where $a_{i,j,m,n}\triangleq 1/(b_j\lambda_{j,m}\delta_{i,n})$, $c_{i,n}\triangleq\alpha d_{i,n}$, and $g_{i,m,n}\triangleq 1/(\lambda_{r,m}\delta_{i,n})$ are known variables. It can be seen from the above equation that the triple summation terms and the double summation terms are monotonically decreasing and convex with respect to $\sigma_{j,m}$ and $\sigma_{r,m}$, respectively. Moreover, with fixed α, the constraints in (3.29) and (3.30) are linear inequality constraints which can be rewritten as $\mathbf{1}^T\boldsymbol{\sigma}_i\le p_i$, $i=1,2$, and $\bar{\mathbf{z}}_1^T\boldsymbol{\sigma}_1+\bar{\mathbf{z}}_2^T\boldsymbol{\sigma}_2+\mathbf{1}^T\boldsymbol{\sigma}_r\le p_r-\alpha N_r$, respectively, where $\bar{\mathbf{z}}_i\triangleq\alpha\mathrm{tr}(\mathbf{R}_{ri})[\lambda_{i,1},\cdots,\lambda_{i,N_i}]^T$, $i=1,2$, and $\mathbf{1}$ is a vector of all ones with a commensurate dimension. Therefore, the problem (3.28)-(3.31) with respect to $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_r$ is a convex optimization problem when α is fixed, where the optimal $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_r$ can be efficiently obtained through the KKT optimality conditions of the problem (3.28)-(3.31). In particular, the gradient conditions are

given by

$$\sum_{j=1}^{2}\sum_{n=1}^{N_j}\frac{\alpha d_{j,n}}{\left[(b_i\lambda_{i,m}\delta_{j,n})^{-1}+\alpha d_{j,n}\sigma_{i,m}\right]^2}=\mu_i+\mu_3 e_{i,m},\quad m=1,\cdots,N_i,\quad i=1,2 \tag{3.32}$$

$$\sum_{i=1}^{2}\sum_{n=1}^{N_i}\frac{d_{i,n}}{\left[(\lambda_{r,m}\delta_{i,n})^{-1}+d_{i,n}\sigma_{r,m}\right]^2}=\mu_3,\qquad m=1,\cdots,N_r \tag{3.33}$$

where $e_{i,m}\triangleq\alpha\mathrm{tr}(\mathbf{R}_{ri})\lambda_{i,m}$, $i=1,2$, and $\mu_i\geq 0$, $i=1,2,3$, are Lagrange multipliers such that the complementary slackness conditions [99] given by

$$\mu_i\left(p_i-\sum_{m=1}^{N_i}\sigma_{i,m}\right)=0,\qquad i=1,2 \tag{3.34}$$

$$\mu_3\left(p_r-\alpha N_r-\alpha\sum_{i=1}^{2}\sum_{m=1}^{N_i}\lambda_{i,m}\sigma_{i,m}\mathrm{tr}(\mathbf{R}_{ri})-\sum_{m=1}^{N_r}\sigma_{r,m}\right)=0 \tag{3.35}$$

are satisfied.

With fixed α and μ_i, $i=1,2,3$, for each m, the non-negative $\sigma_{1,m}$, $\sigma_{2,m}$, and $\sigma_{r,m}$ can be found by using the bi-section search, since the LHS of (3.32) and (3.33) are monotonically decreasing function of $\sigma_{i,m}$ and $\sigma_{r,m}$, respectively. To find the optimal μ_i, $i=1,2,3$, an outer bi-section search loop is used as the LHS of (3.29) is an increasing function of $\sigma_{i,m}$, and the LHS of (3.30) is an increasing function of $\sigma_{1,m}$, $\sigma_{2,m}$, and $\sigma_{r,m}$, while in (3.32), $\sigma_{i,m}$ is a monotonically decreasing function of μ_i and μ_3, and $\sigma_{r,m}$ is a monotonically decreasing function of μ_3 in (3.33).

When α is an optimization variable (not fixed), the problem (3.28)-(3.31) as a whole is not a convex optimization problem. However, we can show that (3.28) subjecting to (3.29)-(3.31) is a unimodal (quasi-convex) function with respect to α. Let us introduce $\chi_{i,m}\triangleq\alpha\sigma_{i,m}$, $m=1,\cdots,N_i, i=1,2$, the problem (3.28)-(3.31) can be rewritten as

$$\min_{\alpha,\boldsymbol{\chi}_1,\boldsymbol{\chi}_2,\boldsymbol{\sigma}_r}\ \sum_{i=1}^{2}\left(\sum_{j=1}^{2}\sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\frac{1}{a_{i,j,m,n}+d_{i,n}\chi_{j,m}}+\sum_{m=1}^{N_r}\sum_{n=1}^{N_i}\frac{1}{g_{i,m,n}+d_{i,n}\sigma_{r,m}}\right) \tag{3.36}$$

$$\text{s.t.}\ \ \mathbf{1}^T\boldsymbol{\chi}_i\leq\alpha p_i,\qquad i=1,2 \tag{3.37}$$

$$\mathbf{z}_1^T\boldsymbol{\chi}_1+\mathbf{z}_2^T\boldsymbol{\chi}_2+\mathbf{1}^T\boldsymbol{\sigma}_r\leq p_r-\alpha N_r \tag{3.38}$$

$$\alpha>0,\ \ \sigma_{r,m}\geq 0,\ \ \chi_{i,m}\geq 0,\qquad m=1,\cdots,N_i,\qquad i=1,2 \tag{3.39}$$

where $\boldsymbol{\chi}_i\triangleq[\chi_{i,1},\cdots,\chi_{i,N_i}]^T$, and $\mathbf{z}_i\triangleq\mathrm{tr}(\mathbf{R}_{ri})[\lambda_{i,1},\cdots,\lambda_{i,N_i}]^T$, $i=1,2$.

Let us first ignore the effect of all $d_{i,n}$ by treating them as known variables. Then the problem (3.36)-(3.39) is a convex optimization problem, since (3.36) is a convex function of $\boldsymbol{\chi}_1$, $\boldsymbol{\chi}_2$, $\boldsymbol{\sigma}_r$, and (3.37)-(3.39) are linear inequality constraints. In particular, with increasing α, the value of (3.36) first decreases and then increases based on the following reasons. For a significantly small α, the value of (3.36) is strongly governed by the constraints in (3.37), since constraint (3.38) is inactive compared with those in (3.37) when α is small. Once α increases from a small value, the feasible region specified by (3.37) expands, and thus, the value of (3.36) decreases. On the other hand, when α is large, the value of (3.36) is strongly governed by the constraint in (3.38), since constraints in (3.37) are inactive compared with that in (3.38) when α is large. Once α decreases from a large value, the feasible region specified by (3.38) expands, resulting in a deceasing of the value of (3.36). Now we consider the effect of $d_{i,n}$. Since $d_{i,n} = 1/(\alpha \mathrm{tr}(\mathbf{C}_r^T)\delta_{i,n} + 1)$, $d_{i,n}$ monotonically decreases with increasing α, and (3.36) increases when $d_{i,n}$ decreases.

Considering the two effects above, we can draw the following conclusion regarding the value of (3.36) with respect to α. When α increases from a significantly small positive number, the value of (3.36) starts to decrease since the potential decrease of (3.36) due to the expanded feasible region (3.37) dominates the potential increase of (3.36) caused by the decreasing $d_{i,n}$. The value of (3.36) keeps decreasing as α increases till a 'turning point' where the decreasing of $d_{i,n}$ starts to dominate the effect of relaxed feasible region (3.37). After such turning point, the value of (3.36) will monotonically increase with an increasing α.

To validate the analysis above, a plot of the MSE value (3.28) over a range of feasible values of α is generated in Fig. 3.2 for the case where all nodes have the same number of antennas, i.e., $N_i = N = 4$, $i = 1, 2, r$, and the channel matrices have i.i.d. entries, i.e., $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$, $i = 1, 2$. Fig. 3.2 shows the NMSE, which is (3.28) divided by $6N^2$, versus α for different $p_1 = p_2$, and p_r is set to be 20dB. It can be observed from Fig. 3.2 that (3.28) is a unimodal (quasi-convex) function of α. Thus, the optimal α for the problem (3.28)-(3.31) can be efficiently found by applying the golden section search (GSS) technique described in Table 3.1, where ε is a positive constant close to 0 and $\phi > 0$ is the reduction

factor. It is shown in [100] that the optimal $\phi = 1.618$, also known as the golden ratio. The GSS method can guarantee that the minimum of a unimodal function to be found by bracketing the minimum to an interval of 0.618 times the size of the preceding interval. Unlike the Fibonacci search, the GSS method is able to perform up to the desired accuracy and does not require the number of iterations as input. However, the GSS method may need more iterations compared with the Fibonacci search.

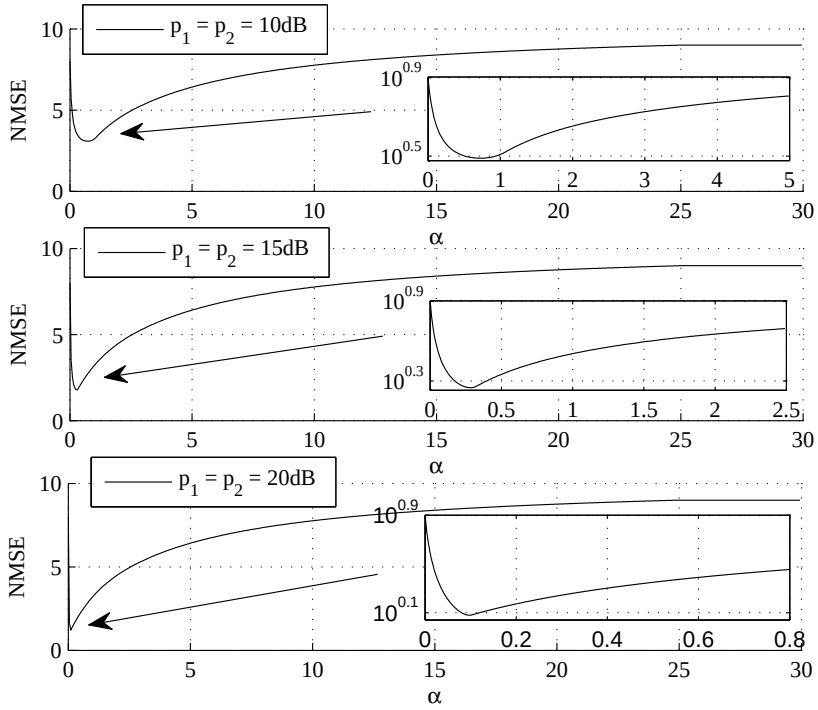

Figure 3.2: Superimposed channel training: NMSE versus α for different $p_1 = p_2$ with $N = 4$ and $p_r = 20$dB.

The complexity of the superimposed channel training algorithm can be estimated as $\mathcal{O}(c_\alpha c_\mu c_\sigma (N_1 + N_2)(N_1 + N_2 + N_r))$, where c_α is the number of GSS iterations required to obtain the optimal α, c_μ stands for the number of iterations in the outer bi-section loop to obtain the optimal μ_1, μ_2, and μ_3, and c_σ represents the number of bi-section operations required to obtain the optimal $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, and $\boldsymbol{\sigma}_r$.

Table 3.1: Procedure of applying the golden section search (GSS) to find the optimal α in the problem (3.28)-(3.31).

---

1. Set a feasible bound $[a,b]$ on α.
2. Define $c_1 = (\phi - 1)a + (2 - \phi)b$ and $c_2 = (2 - \phi)a + (\phi - 1)b$.
3. Solve the problem (3.28)-(3.31) for $\alpha = c_1$;
   Compute the MSE value defined in (3.28), $f_{MSE}(c_1)$ for $\alpha = c_1$.
4. Repeat Step 3 for $\alpha = c_2$.
5. If $f_{MSE}(c_1) < f_{MSE}(c_2)$, then assign $b = c_2$.
   Otherwise, assign $a = c_1$.
6. If $|b - a| \leq \varepsilon$, then end.
   Otherwise, go to step 2.

3.4 Two-Stage Channel Estimation Algorithm

There are two stages in this channel estimation scheme. In particular, the channel matrices $\mathbf{H}_{ir}$, $i = 1, 2$, from the relay node to the receive nodes are estimated in the first stage, while the channel matrices $\mathbf{H}_{ri}$, $i = 1, 2$, from the source nodes to the relay node are estimated in the second stage. The first stage requires one time block while the second stage requires two time blocks.

3.4.1 Stage One

At the first stage, the relay node transmits an $N_r \times L_1$ training signal matrix $\mathbf{P}_r$ to both receive nodes, where L_1 is the length of the training sequence and will be determined later. The $N_i \times L_1$ received signal matrix $\mathbf{Y}_{i,1}$ at node i is given by

$$\mathbf{Y}_{i,1} = \mathbf{H}_{ir}\mathbf{P}_r + \mathbf{V}_{i,1}, \qquad i = 1, 2 \tag{3.40}$$

where $\mathbf{V}_{i,1}$ is an $N_i \times L_1$ noise matrix at node i in stage one. By vectorizing both sides of (3.40), we obtain

$$\mathbf{y}_{i,1} = (\mathbf{P}_r^T \otimes \mathbf{I}_{N_i})\mathbf{h}_{ir} + \mathbf{v}_{i,1}, \qquad i = 1, 2 \tag{3.41}$$

where $\mathbf{y}_{i,1} \triangleq \text{vec}(\mathbf{Y}_{i,1})$, $\mathbf{h}_{ir} \triangleq \text{vec}(\mathbf{H}_{ir})$, and $\mathbf{v}_{i,1} \triangleq \text{vec}(\mathbf{V}_{i,1})$.

Using a linear MMSE estimator at node i to estimate $\mathbf{h}_{ir}$, we obtain

$$\hat{\mathbf{h}}_{ir} = \mathbf{W}_{i,1}^H \mathbf{y}_{i,1}, \qquad i = 1, 2 \tag{3.42}$$

where $\hat{\mathbf{h}}_{ir}$ denotes an estimation of $\mathbf{h}_{ir}$ and $\mathbf{W}_{i,1}$ is the weight matrix of the MMSE estimator given by

$$\mathbf{W}_{i,1} = \mathbf{R}_{i,1}^{-1} \mathbf{C}_{i,1}, \qquad i = 1, 2. \tag{3.43}$$

Here $\mathbf{R}_{i,1} \triangleq \mathrm{E}\big[\mathbf{y}_{i,1}\mathbf{y}_{i,1}^H\big] = (\mathbf{P}_r^T \mathbf{C}_r \mathbf{P}_r^*) \otimes \mathbf{R}_{ir} + \mathbf{I}_{N_i T_1}$ and $\mathbf{C}_{i,1} \triangleq \mathrm{E}\big[\mathbf{y}_{i,1}\mathbf{h}_{ir}^H\big] = (\mathbf{P}_r^T \mathbf{C}_r) \otimes \mathbf{R}_{ir}$, $i = 1, 2$. From (3.41) and (3.42), we find that since a linear estimator is used, there is $L_1 \geq N_r$. Using (3.41)-(3.43), the MSE of estimating $\mathbf{h}_{ir}$ can be written as

$$\begin{aligned} \mathrm{MSE}_{i,1} &= \mathrm{E}\big[\mathrm{tr}\big((\hat{\mathbf{h}}_{ir} - \mathbf{h}_{ir})(\hat{\mathbf{h}}_{ir} - \mathbf{h}_{ir})^H\big)\big] \\ &= \mathrm{tr}\big([(\mathbf{C}_r \otimes \mathbf{R}_{ir})^{-1} + (\mathbf{P}_r^* \mathbf{P}_r^T) \otimes \mathbf{I}_{N_i}]^{-1}\big), \quad i = 1, 2. \end{aligned}$$

Since the transmission power consumed by the relay node at stage one is $\mathrm{tr}(\mathbf{P}_r \mathbf{P}_r^H)$, the optimal $\mathbf{P}_r$ can be derived by solving the following optimization problem

$$\min_{\mathbf{P}_r} \quad \sum_{i=1}^{2} \mathrm{tr}\Big(\big[(\mathbf{C}_r \otimes \mathbf{R}_{ir})^{-1} + (\mathbf{P}_r^* \mathbf{P}_r^T) \otimes \mathbf{I}_{N_i}\big]^{-1}\Big) \tag{3.44}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{P}_r \mathbf{P}_r^H) \leq q_{r,1} \tag{3.45}$$

where $q_{r,1}$ is the power allocation at the relay node at the first stage. The following theorem establishes the optimal structure of $\mathbf{P}_r$ as the solution to the problem (3.44)-(3.45).

THEOREM 3.2: The optimal training sequence $\mathbf{P}_r$ satisfies $\mathbf{P}_r \mathbf{P}_r^H = \mathbf{U}_r \boldsymbol{\Xi}_r \mathbf{U}_r^H$, where $\boldsymbol{\Xi}_r$ is an $N_r \times N_r$ diagonal matrix.

PROOF: Similar to the proof of Theorem 3.1. □

The optimal structure of $\mathbf{P}_r$ can be obtained from Theorem 3.2 as $\mathbf{P}_r = \mathbf{U}_r \boldsymbol{\Xi}_r^{\frac{1}{2}} \boldsymbol{\Omega}_{r,1}$, where $\boldsymbol{\Omega}_{r,1}$ is an $N_r \times L_1$ semi-unitary matrix satisfying $\boldsymbol{\Omega}_{r,1} \boldsymbol{\Omega}_{r,1}^H = \mathbf{I}_{N_r}$ and can be easily constructed from the normalized DFT matrix when $L_1 \geq N_r$. Using Theorem 3.2, the problem (3.44)-(3.45) is equivalently converted to the following problem

$$\min_{\boldsymbol{\Xi}_r} \quad \sum_{i=1}^{2} \mathrm{tr}\Big(\big[(\boldsymbol{\Lambda}_r^{-1} \otimes \boldsymbol{\Lambda}_{ir}^{-1}) + (\boldsymbol{\Xi}_r \otimes \mathbf{I}_{N_i})\big]^{-1}\Big) \tag{3.46}$$

$$\text{s.t.} \quad \mathrm{tr}(\boldsymbol{\Xi}_r) \leq q_{r,1}, \qquad \boldsymbol{\Xi}_r \geq 0. \tag{3.47}$$

The problem (3.46)-(3.47) can be equivalently rewritten as the following problem with scalar variables

$$\min_{\boldsymbol{\xi}_r} \quad \sum_{i=1}^{2}\sum_{m=1}^{N_r}\sum_{n=1}^{N_i}\frac{1}{\lambda_{r,m}^{-1}\lambda_{ir,n}^{-1}+\xi_{r,m}} \tag{3.48}$$

$$\text{s.t.} \quad \sum_{m=1}^{N_r}\xi_{r,m}\le q_{r,1}, \qquad \xi_{r,m}\ge 0, \quad m=1,\cdots,N_r \tag{3.49}$$

where $\boldsymbol{\xi}_r \triangleq \left[\xi_{r,1},\cdots,\xi_{r,N_r}\right]^T$ and $\xi_{r,m}$ is the mth diagonal element of $\boldsymbol{\Xi}_r$.

The problem (3.48)-(3.49) is convex and thus can be efficiently solved through the KKT optimality conditions. The gradient condition is given by

$$\sum_{i=1}^{2}\sum_{n=1}^{N_i}\frac{1}{\left[\lambda_{r,m}^{-1}\lambda_{ir,n}^{-1}+\xi_{r,m}\right]^2}=\mu, \qquad m=1,\cdots,N_r \tag{3.50}$$

where $\mu\ge 0$ is the Lagrange multiplier such that the complementary slackness condition $\mu\big(q_{r,1}-\sum_{m=1}^{N_r}\xi_{r,m}\big)=0$ is satisfied. For each m, with fixed μ, the non-negative $\xi_{r,m}$ can be found using the bi-section search, since the LHS of (3.50) is a monotonically decreasing function of $\xi_{r,m}$. To find the optimal μ, an outer bi-section search is used as the LHS of (3.49) is an increasing function of $\xi_{r,m}$, while in (3.50), $\xi_{r,m}$ is a monotonically decreasing function of μ.

3.4.2 Stage Two

At the second stage, the source node i transmits an $N_i\times L_2$ training signal matrix $\mathbf{P}_i$ to the relay node. The $N_r\times L_2$ received signal matrix $\mathbf{Y}_{r,2}$ at the relay node is given by

$$\mathbf{Y}_{r,2}=\sum_{i=1}^{2}\mathbf{H}_{ri}\mathbf{P}_i+\mathbf{V}_{r,2}$$

where $\mathbf{V}_{r,2}$ is an $N_r\times L_2$ noise matrix at the relay node. Then the relay node amplifies $\mathbf{Y}_{r,2}$ and retransmits $\mathbf{X}_{r,2}=\sqrt{\eta}\mathbf{Y}_{r,2}$, where $\eta>0$ is the relay amplifying factor. The $N_i\times L_2$ received signal matrix at node i is given by

$$\begin{aligned}\mathbf{Y}_{i,2} &= \mathbf{H}_{ir}\mathbf{X}_{r,2}+\mathbf{V}_{i,2}\\ &= \sqrt{\eta}\mathbf{H}_{ir}\mathbf{H}_{ri}\mathbf{P}_i+\sqrt{\eta}\mathbf{H}_{ir}\mathbf{H}_{r\bar{i}}\mathbf{P}_{\bar{i}}+\sqrt{\eta}\mathbf{H}_{ir}\mathbf{V}_{r,2}+\mathbf{V}_{i,2}, \quad i=1,2\end{aligned} \tag{3.51}$$

where $\mathbf{V}_{i,2}$ is an $N_i\times L_2$ noise matrix at node i.

Introducing $\tilde{\mathbf{P}}_j \triangleq \mathbf{U}_j^H \mathbf{P}_j$, $j = i, \bar{i}$, $\bar{\mathbf{V}}_{i,2} \triangleq \sqrt{\eta}\mathbf{H}_{ir}\mathbf{V}_{r,2} + \mathbf{V}_{i,2}$, $i = 1, 2$, we can rewrite (3.51) as

$$\mathbf{Y}_{i,2} = \sqrt{\eta}\mathbf{G}_{ii}\tilde{\mathbf{P}}_i + \sqrt{\eta}\mathbf{G}_{i\bar{i}}\tilde{\mathbf{P}}_{\bar{i}} + \bar{\mathbf{V}}_{i,2}, \qquad i = 1, 2 \tag{3.52}$$

where $\mathbf{G}_{ij}$ is defined in (3.7). Similar to Section 3.3, we first estimate $\mathbf{G}_{ij}$. Then an estimation of $\mathbf{H}_{rj}$ is obtained as $\breve{\mathbf{H}}_{rj} = \breve{\mathbf{H}}_{ir}^{\dagger}\breve{\mathbf{G}}_{ij}\mathbf{U}_j^H$, $j = i, \bar{i}$, where $\breve{\mathbf{H}}_{ir}$ is the estimation of $\mathbf{H}_{ir}$ obtained from stage one and $\breve{\mathbf{G}}_{ij}$ is the estimation of $\mathbf{G}_{ij}$. By vectorizing both sides of (3.52), we obtain

$$\begin{aligned} \mathbf{y}_{i,2} &= \big[\sqrt{\eta}\tilde{\mathbf{P}}_i^T \otimes \mathbf{I}_{N_i}, \ \sqrt{\eta}\tilde{\mathbf{P}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}\big]\big[\mathbf{g}_{ii}^T, \ \mathbf{g}_{i\bar{i}}^T\big]^T + \bar{\mathbf{v}}_{i,2} \\ &\triangleq \mathbf{N}_i\boldsymbol{\theta}_i + \bar{\mathbf{v}}_{i,2}, \qquad i = 1, 2 \end{aligned} \tag{3.53}$$

where $\mathbf{y}_{i,2} \triangleq \mathrm{vec}(\mathbf{Y}_{i,2})$, $\bar{\mathbf{v}}_{i,2} \triangleq \mathrm{vec}(\bar{\mathbf{V}}_{i,2})$, $\mathbf{N}_i \triangleq \big[\sqrt{\eta}\tilde{\mathbf{P}}_i^T \otimes \mathbf{I}_{N_i}, \ \sqrt{\eta}\tilde{\mathbf{P}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}\big]$, and $\boldsymbol{\theta}_i \triangleq \big[\mathbf{g}_{ii}^T, \ \mathbf{g}_{i\bar{i}}^T\big]^T$ is the vector of unknown variables at node i.

Using a linear MMSE receiver to estimate $\boldsymbol{\theta}_i$, we have

$$\hat{\boldsymbol{\theta}}_i = \mathbf{W}_{i,2}^H \mathbf{y}_{i,2}, \qquad i = 1, 2 \tag{3.54}$$

where $\hat{\boldsymbol{\theta}}_i$ stands for an estimation of $\boldsymbol{\theta}_i$, $\mathbf{W}_{i,2}$ is the weight matrix of the MMSE estimator and given by

$$\mathbf{W}_{i,2} = \big(\mathbf{N}_i\mathbf{R}_{\theta_i}\mathbf{N}_i^H + \mathbf{R}_{\bar{v}_{i,2}}\big)^{-1}\mathbf{N}_i\mathbf{R}_{\theta_i}, \qquad i = 1, 2. \tag{3.55}$$

From (3.53) and (3.54), we find that since a linear estimator is used, there is $L_2 \geq N_1 + N_2$. In (3.55), $\mathbf{R}_{\theta_i} \triangleq \mathrm{E}[\boldsymbol{\theta}_i\boldsymbol{\theta}_i^H]$ is the covariance matrix of $\boldsymbol{\theta}_i$, which can be calculated similar to $\mathbf{R}_{\gamma_i}$ in (3.16) and written as

$$\mathbf{R}_{\theta_i} = \mathrm{Bdiag}\big[\boldsymbol{\Lambda}_i \otimes b_i\mathbf{R}_{ir}, \ \boldsymbol{\Lambda}_{\bar{i}} \otimes b_{\bar{i}}\mathbf{R}_{ir}\big], \qquad i = 1, 2.$$

In (3.55), $\mathbf{R}_{\bar{v}_{i,2}} \triangleq \mathrm{E}[\bar{\mathbf{v}}_{i,2}\bar{\mathbf{v}}_{i,2}^H]$ is the noise covariance matrix which can be calculated similar to $\mathbf{R}_{\bar{v}_i}$ (3.13) as

$$\mathbf{R}_{\bar{v}_{i,2}} = \mathbf{I}_{L_2} \otimes \big(\eta\mathrm{tr}(\mathbf{C}_r^T)\mathbf{R}_{ir} + \mathbf{I}_{N_i}\big), \qquad i = 1, 2.$$

Using (3.55), the MSE of estimating $\boldsymbol{\theta}_i$ can be obtained as

$$\begin{aligned} \mathrm{MSE}_{i,2} &= \mathrm{E}\big[\mathrm{tr}\big((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)^H\big)\big] \\ &= \mathrm{tr}\Big(\big[\mathbf{R}_{\theta_i}^{-1} + \mathbf{N}_i^H\mathbf{R}_{\bar{v}_{i,2}}^{-1}\mathbf{N}_i\big]^{-1}\Big), \quad i = 1, 2. \end{aligned} \tag{3.56}$$

The transmission power consumed at nodes 1 and 2 is

$$\mathrm{tr}(\mathbf{P}_i\mathbf{P}_i^H) = \mathrm{tr}(\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H), \qquad i = 1, 2. \tag{3.57}$$

And the power consumed at the relay node is given by

$$\eta \mathrm{E}\left[\mathrm{tr}\Big(\sum_{i=1}^{2}\mathbf{H}_{ri}\mathbf{P}_i\mathbf{P}_i^H\mathbf{H}_{ri}^H + \mathbf{I}_{N_r}\Big)\right] = \eta N_r + \eta\sum_{i=1}^{2}\mathrm{tr}(\mathbf{\Lambda}_i\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H)\mathrm{tr}(\mathbf{R}_{ri}). \tag{3.58}$$

From (3.56)-(3.58), the optimal training matrices $\mathbf{P}_i$, $i = 1, 2$, and the optimal η can be obtained through solving the following optimization problem

$$\min_{\eta,\tilde{\mathbf{P}}_1,\tilde{\mathbf{P}}_2} \quad \sum_{i=1}^{2}\mathrm{tr}\Big(\big[\mathbf{R}_{\theta_i}^{-1} + \mathbf{N}_i^H\mathbf{R}_{\bar{v}_{i,2}}^{-1}\mathbf{N}_i\big]^{-1}\Big) \tag{3.59}$$

$$\text{s.t.} \quad \mathrm{tr}(\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H) \le q_i, \qquad i = 1, 2 \tag{3.60}$$

$$\eta\Big[N_r + \sum_{i=1}^{2}\mathrm{tr}(\mathbf{\Lambda}_i\tilde{\mathbf{P}}_i\tilde{\mathbf{P}}_i^H)\mathrm{tr}(\mathbf{R}_{ri})\Big] \le q_{r,2} \tag{3.61}$$

where q_i is the transmission power available at node i, $i = 1, 2$, and $q_{r,2}$ is the transmission power available at the relay node at the second stage. Note that for a fair comparison with the superimposed channel training algorithm, the power at three nodes should satisfy

$$q_{r,1}L_1 + q_{r,2}L_2 = p_rL, \qquad q_iL_2 = p_iL, \qquad i = 1, 2. \tag{3.62}$$

The following theorem establishes the optimal structure of $\mathbf{P}_1$ and $\mathbf{P}_2$ as the solution to the problem (3.59)-(3.61).

THEOREM 3.3: The optimal training sequences $\mathbf{P}_1$ and $\mathbf{P}_2$ satisfy $\mathbf{P}_1\mathbf{P}_2^H = \mathbf{0}$ and $\mathbf{P}_i\mathbf{P}_i^H = \mathbf{U}_i\mathbf{\Xi}_i\mathbf{U}_i^H$, $i = 1, 2$, where $\mathbf{\Xi}_i$ is an $N_i \times N_i$ diagonal matrix.

PROOF: Similar to the proof of Theorem 3.1. □

The optimal structure of $\mathbf{P}_i$ can be obtained from Theorem 3.3 as $\mathbf{P}_i = \mathbf{U}_i\mathbf{\Xi}_i^{\frac{1}{2}}\mathbf{\Omega}_{i,2}$, where $\mathbf{\Omega}_{i,2}$ is an $N_i \times L_2$ semi-unitary matrix satisfying $\mathbf{\Omega}_{i,2}\mathbf{\Omega}_{i,2}^H = \mathbf{I}_{N_i}$, $i = 1, 2$, and $\mathbf{\Omega}_{1,2}\mathbf{\Omega}_{2,2}^H = \mathbf{0}$. Such $\mathbf{\Omega}_{1,2}$ and $\mathbf{\Omega}_{2,2}$ can be easily constructed from the normalized DFT matrix when $L_2 \ge N_1 + N_2$. Using Theorem 3.3, the problem (3.59)-(3.61) is equivalently converted to the following problem

$$\min_{\eta,\mathbf{\Xi}_1,\mathbf{\Xi}_2} \quad \sum_{i=1}^{2}\sum_{j=1}^{2}\mathrm{tr}\Big(\big[\mathbf{D}_{ij} + \eta\mathbf{\Xi}_j \otimes \mathbf{E}_{ri}\big]^{-1}\Big) \tag{3.63}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{\Xi}_i) \le q_i, \qquad i = 1, 2 \tag{3.64}$$

$$\eta N_r + \eta\sum_{i=1}^{2}\mathrm{tr}(\mathbf{\Lambda}_i\mathbf{\Xi}_i)\mathrm{tr}(\mathbf{R}_{ri}) \le q_{r,2} \tag{3.65}$$

$$\eta > 0, \quad \mathbf{\Xi}_i \ge 0, \quad i = 1, 2 \tag{3.66}$$

where $\mathbf{E}_{ri} \triangleq \left(\eta \mathrm{tr}(\mathbf{C}_r^T)\mathbf{\Delta}_i + \mathbf{I}_{N_i}\right)^{-1}$, $i = 1, 2$, are diagonal matrices. The problem (3.63)-(3.66) can be equivalently rewritten as the following problem with scalar variables

$$\min_{\eta, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2} \quad \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\left[\frac{1}{b_j\lambda_{j,m}\delta_{i,n}} + \eta f_{i,n}\xi_{j,m}\right]^{-1} \tag{3.67}$$

$$\text{s.t.} \quad \sum_{m=1}^{N_i}\xi_{i,m} \le q_i, \quad i = 1, 2 \tag{3.68}$$

$$\eta N_r + \eta\sum_{i=1}^{2}\left(\sum_{m=1}^{N_i}\lambda_{i,m}\xi_{i,m}\mathrm{tr}(\mathbf{R}_{ri})\right) \le q_{r,2} \tag{3.69}$$

$$\eta > 0, \;\; \xi_{i,m} \ge 0, \quad m = 1, \cdots, N_i, \quad i = 1, 2 \tag{3.70}$$

where $f_{i,n} \triangleq 1/(\eta \mathrm{tr}(\mathbf{C}_r^T)\delta_{i,n} + 1)$, $\boldsymbol{\xi}_i \triangleq \left[\xi_{i,1}, \cdots, \xi_{i,N_i}\right]^T$, $i = 1, 2$, and $\xi_{i,m}$ is the mth diagonal element of $\mathbf{\Xi}_i$.

With fixed η, the objective function (3.67) can be rewritten as

$$\min_{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2} \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{m=1}^{N_j}\sum_{n=1}^{N_i}\frac{1}{a_{i,j,m,n} + h_{i,n}\xi_{j,m}}$$

where $a_{i,j,m,n}$ and $h_{i,n} \triangleq \eta f_{i,n}$ are known variables. It can be seen from the above equation that the summation terms are monotonically decreasing and convex with respect to $\xi_{1,m}$ and $\xi_{2,m}$. Moreover, with fixed η, the constraints in (3.68)-(3.70) are linear inequality constraints. Therefore, the problem (3.67)-(3.70) is a convex optimization problem with respect to $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ when η is fixed. For a given η, the optimal $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ can be efficiently obtained through the KKT optimality conditions associated with the problem (3.67)-(3.70). The gradient conditions are given by

$$\sum_{j=1}^{2}\sum_{n=1}^{N_j}\frac{\eta f_{j,n}}{\left[(b_i\lambda_{i,m}\delta_{j,n})^{-1} + \eta f_{j,n}\xi_{i,m}\right]^2} = \nu_i + \nu_3 c_{i,m}, \qquad m = 1, \cdots, N_i, \quad i = 1, 2 \tag{3.71}$$

where $c_{i,m} \triangleq \eta \mathrm{tr}(\mathbf{R}_{ri})\lambda_{i,m}$, $i = 1, 2$, and $\nu_i \ge 0$, $i = 1, 2, 3$, are Lagrange multipliers such that the complementary slackness conditions given by

$$\nu_i\left(q_i - \sum_{m=1}^{N_i}\xi_{i,m}\right) = 0, \quad i = 1, 2 \tag{3.72}$$

$$\nu_3\left(q_{r,2} - \eta N_r - \eta\sum_{i=1}^{2}\sum_{m=1}^{N_i}\lambda_{i,m}\xi_{i,m}\mathrm{tr}(\mathbf{R}_{ri})\right) = 0 \tag{3.73}$$

are satisfied.

With fixed η and ν_i, $i = 1, 2, 3$, for each m, the non-negative $\xi_{1,m}$ and $\xi_{2,m}$ can be found by using the bi-section search, since the LHS of (3.71) is a monotonically decreasing function of $\xi_{1,m}$ and $\xi_{2,m}$. To find the optimal ν_i, $i = 1, 2, 3$, an outer bi-section search is used as the LHS of (3.68) and (3.69) are increasing functions of $\xi_{1,m}$ and $\xi_{2,m}$, while in (3.71), $\xi_{i,m}$ is monotonically decreasing with respect to ν_i and ν_3.

The problem (3.67)-(3.70) as a whole is is non-convex with respect to $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, η. However, based on a similar analysis used in the problem (3.28)-(3.31), it can be shown that (3.67) subjecting to (3.68)-(3.70) is a unimodal (quasi-convex) function with respect to η. To validate our analysis, a plot of the MSE value over a range of feasible values of η is generated in Fig. 3.3 for the case where all nodes have the same number of antennas, i.e., $N_i = N = 4$, $i = 1, 2, r$. The channel matrices have i.i.d. entries, i.e., $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$, $i = 1, 2$. Fig. 3.3 shows the NMSE value versus η for different $q_1 = q_2$ with $q_{r,2}$ set to be 20dB. It can be observed from Fig. 3.3 that (3.67) is a unimodal function of η. For a unimodal function, the minimum value can be efficiently found by the GSS algorithm [100]. Hence, the optimal η for the problem (3.67)-(3.70) can be obtained by applying the GSS technique similar to the procedure listed in Table 3.1.

Now let us investigate the optimal power allocation $q_{r,1}$ and $q_{r,2}$ at the relay node during two stages of channel training. Based on (3.62), we let $q_{r,1}L_1 = \beta p_r L$ and $q_{r,2}L_2 = (1-\beta)p_r L$, where $0 < \beta < 1$. The aim is to find the optimal β to minimize the overall MSE of channel estimation over two-stages which is given by the summation of (3.44) and (3.59), and can be written as

$$\sum_{i=1}^{2} \mathrm{tr}\Big(\big[(\mathbf{C}_r \otimes \mathbf{R}_{ir})^{-1} + (\mathbf{P}_r^* \mathbf{P}_r^T) \otimes \mathbf{I}_{N_i} \big]^{-1} + \big[\mathbf{R}_{\theta_i}^{-1} + \mathbf{N}_i^H \mathbf{R}_{\bar{v}_{i,2}}^{-1} \mathbf{N}_i \big]^{-1} \Big). \quad (3.74)$$

Fig. 3.4 shows the value of (3.74) over a range of feasible values of β for different $q_1 = q_2$ with $L_1 = N_r$, $L_2 = N_1 + N_2$, $L = N_1 + N_2 + N_r$, and $p_r = 20$dB. We assume that $N_i = N = 4$, $i = 1, 2, r$, and $\mathbf{T}_{ri} = \mathbf{R}_{ri} = \mathbf{R}_{ir} = \mathbf{C}_r = \mathbf{I}_N$, $i = 1, 2$. Here for each β, the problem (3.44)-(3.45) and the problem (3.59)-(3.61) are solved to obtain the optimal $\mathbf{P}_r$, $\mathbf{P}_1$, $\mathbf{P}_2$, and η. It can be seen from Fig. 3.4

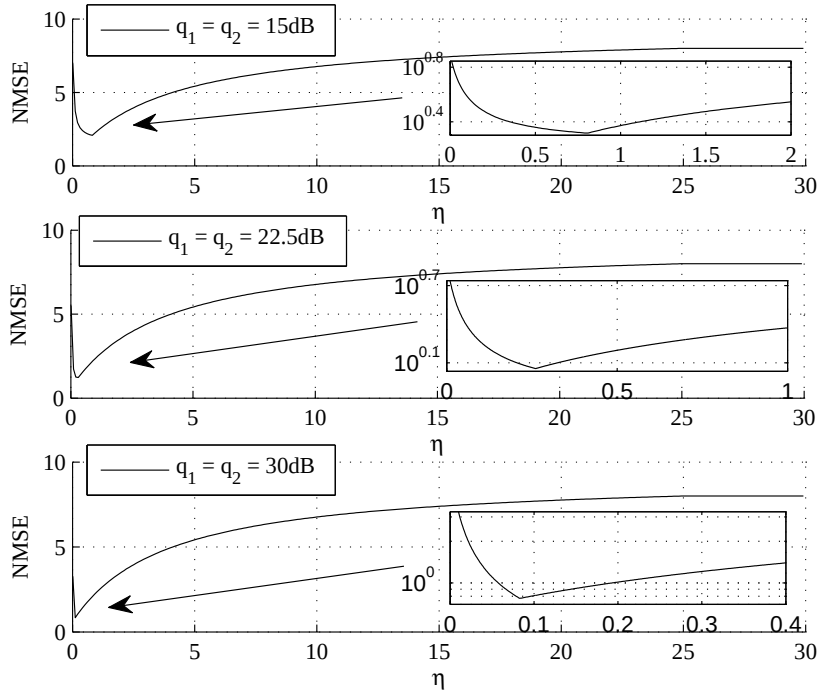

Figure 3.3: Two-stage channel estimation: NMSE versus η for different $q_1 = q_2$ with $N = 4$ and $q_{r,2} = 20$dB.

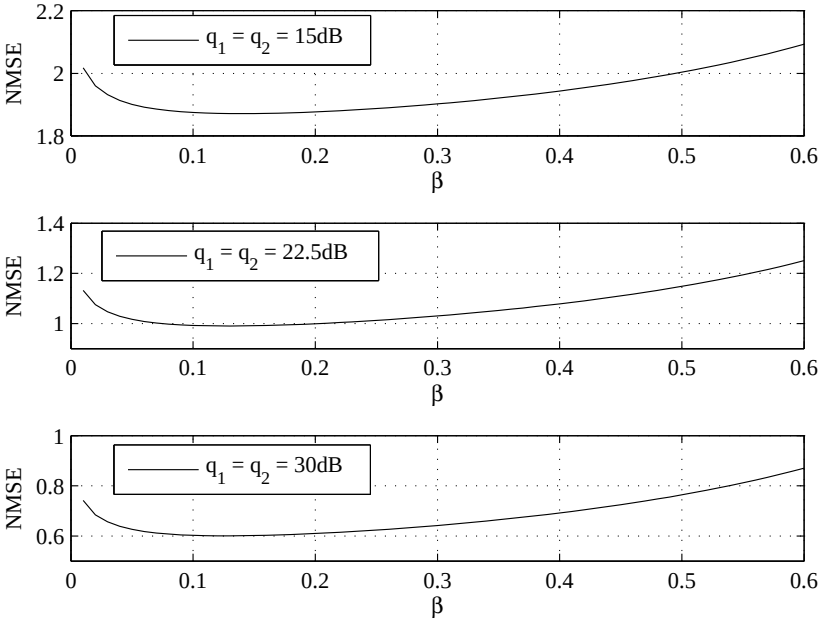

Figure 3.4: Two-stage channel estimation: NMSE versus β for different $q_1 = q_2$ with $N = 4$ and $p_r = 20$dB.

that (3.74) is a unimodal function of β. Hence, the GSS technique described in Table 3.1 can be applied to obtain the optimal β.

The complexity of the two-stage channel estimation algorithm can be estimated as $\mathcal{O}(d_\beta d_\mu d_\xi N_r(N_1+N_2) + d_\beta d_\eta d_\nu d_\xi (N_1+N_2)^2)$, where the first term is the complexity of stage one, and the second term represents the complexity involved in stage two. Here d_β, d_η, and d_μ stand for the numbers of iterations required to obtain the optimal β, η, and μ, respectively, d_ν is the number of iterations in the outer bi-section loop to obtain the optimal ν_1, ν_2, and μ_3, and d_ξ represents the number of bi-section operations required to obtain the optimal $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and $\boldsymbol{\xi}_r$.

3.5 Numerical Examples

In this section, we study the performance of the proposed superimposed channel training algorithm and two-stage channel estimation algorithm through numerical simulations. We consider a three-node two-way MIMO relay system where all nodes are equipped with the same number of antennas, i.e., $N_i = N$, $i = 1, 2, r$. We also assume that all nodes have the same transmission power $p_i = p$, $i = 1, 2, r$, and use the shortest training sequence possible with $L_1 = N$, $L_2 = 2N$, $L = 3N$. Thus, based on (3.62), there are $q_1 = q_2 = 1.5p$ and $(q_{r,1} + 2q_{r,2})/3 = p$ for the two-stage channel estimation algorithm. The channel covariance matrices have the commonly used exponential Toeplitz structure [90] such that $[\mathbf{T}_{ri}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, $[\mathbf{R}_{ri}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, $[\mathbf{R}_{ir}]_{m,n} = \rho^{|m-n|}$, $i = 1, 2$, and $[\mathbf{C}_r]_{m,n} = \rho^{|m-n|}$, where ρ is the correlation coefficient with magnitude $|\rho| < 1$. For all scenarios, the NMSE of channel estimation at nodes 1 and 2 are computed. The optimal training sequences for the superimposed channel training method and the two-stage channel estimation algorithm are generated by using Theorem 3.1 and Theorems 3.2 and 3.3, respectively. In particular, the semi-unitary matrices in the superimposed channel training method are set based on the normalized DFT matrix as $[\boldsymbol{\Omega}_1]_{m,n} = \frac{1}{\sqrt{3N}} e^{-j\frac{2\pi mn}{3N}}$, $[\boldsymbol{\Omega}_2]_{m,n} = \frac{1}{\sqrt{3N}} e^{-j\frac{2\pi(m+N)n}{3N}}$, $[\boldsymbol{\Omega}_r]_{m,n} = \frac{1}{\sqrt{3N}} e^{-j\frac{2\pi(m+2N)n}{3N}}$, $m = 1, \cdots, N$, $n = 1, \cdots, 3N$. Matrices $\boldsymbol{\Omega}_{r,1}$ and $\boldsymbol{\Omega}_{i,2}$, $i = 1, 2$, in the two-stage channel estimation algorithm are chosen as $[\boldsymbol{\Omega}_{r,1}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi mn}{N}}$, $m, n = 1, \cdots, N$, and $[\boldsymbol{\Omega}_{1,2}]_{m,n} = \frac{1}{\sqrt{2N}} e^{-j\frac{2\pi mn}{2N}}$,

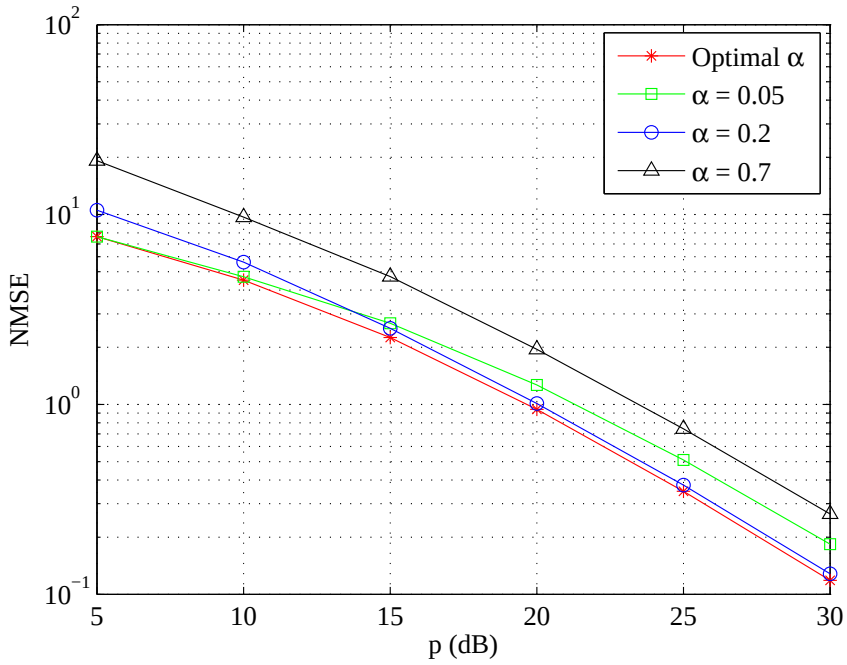

Figure 3.5: Example 3.1. Superimposed channel training: NMSE versus p for different α with $N = 4$ and $\rho = 0.8$.

$[\mathbf{\Omega}_{2,2}]_{m,n} = \frac{1}{\sqrt{2N}} e^{-j\frac{2\pi(m+N)n}{2N}}$, $m = 1, \cdots, N$, $n = 1, \cdots, 2N$.

In the first example, we study the performance of the superimposed channel training algorithm with respect to α. Fig. 3.5 shows the NMSE of this algorithm versus p with different α when $N = 4$ and $\rho = 0.8$. The curve associated with the optimal α is obtained by applying the GSS algorithm on the proposed superimposed channel training technique to find the optimal α for different p. It can be seen from Fig. 3.5 that the curve associated with the optimal α has the lowest MSE level. This justifies that the GSS algorithm can be applied to obtain the optimal α at different p efficiently. Interestingly, we observe from Fig. 3.5 that the optimal α vary with respect to p, indicating that using constant α is strictly suboptimal. In fact, the optimal α at low p level is smaller compared with the optimal α for large p. The reason is that the estimation of the first-hop channels $\mathbf{H}_{ri}$ is based on that of the second-hop channels $\mathbf{H}_{ir}$. When p is small, at the relay node, more power should be allocated for the estimation of $\mathbf{H}_{ir}$, which is also beneficial to the estimation of $\mathbf{H}_{ri}$. When a large amount of power p is available, the MSE of estimating $\mathbf{H}_{ir}$ is smaller compared with that of $\mathbf{H}_{ri}$. Therefore, more power should be allocated at the relay node to assist the estimation of $\mathbf{H}_{ri}$.

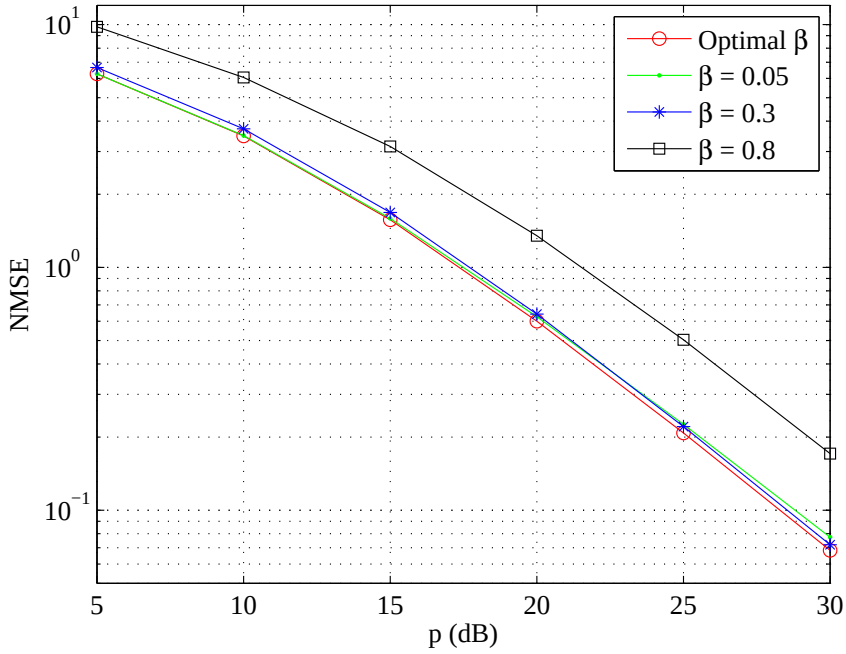

Figure 3.6: Example 3.2. Two-stage channel estimation: NMSE versus p for different β with $N = 4$ and $\rho = 0.8$.

In the second example, we investigate the performance of the two-stage channel estimation algorithm with respect to β. A plot of the NMSE of this algorithm for different β is shown in Fig. 3.6, where the curve with the optimal β is obtained from the GSS algorithm. Similar to Fig. 3.5, it can be seen from Fig. 3.6 that the curve associated with the optimal β has the lowest MSE level.

In the third example, we compare the performance of the superimposed and two-stage channel estimation algorithms when the optimal α and β are used. We also show the performance of the conventional two-stage channel estimator, where random orthogonal pilot sequences are used to estimate the channel matrices and the transmission power at the relay node is equally distributed between two stages. Fig. 3.7 demonstrates the MSE performance of all algorithms with $\rho = 0.2$ for different N, while Fig. 3.8 shows the MSE results at $\rho = 0.8$. It can be seen from Figs. 3.7 and 3.8 that the proposed algorithms yield much smaller estimation error compared with the conventional two-stage channel estimator, especially at high p level. It can also be observed from Figs. 3.7 and 3.8 that for both scenarios, the two-stage channel estimation algorithm yields smaller MSEs than the superimposed channel training scheme. This is mainly due to the fact that

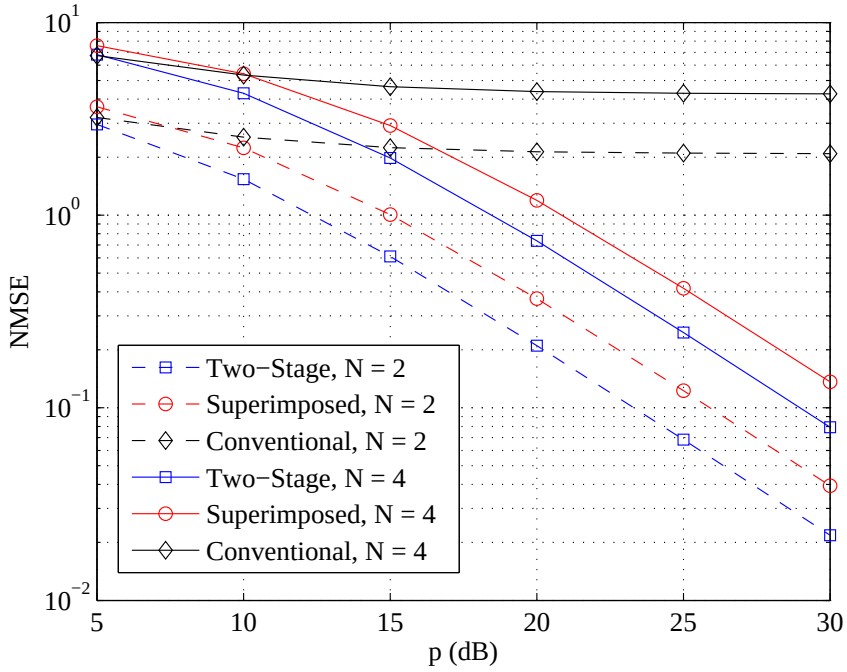

Figure 3.7: Example 3.3. NMSE versus p for $\rho = 0.2$ and different N.

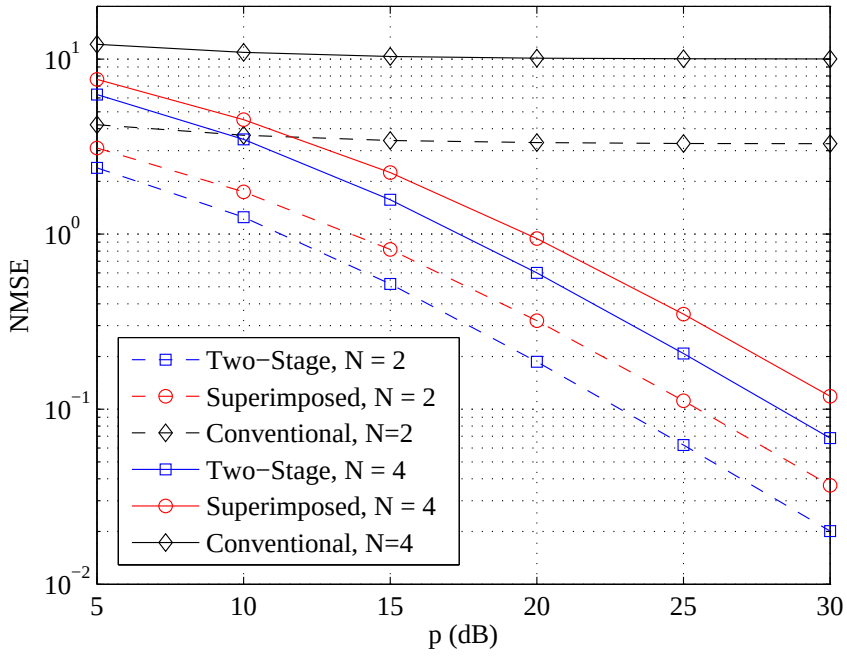

Figure 3.8: Example 3.3. NMSE versus p for $\rho = 0.8$ and different N.

in the superimposed channel training algorithm, the estimation of $\mathbf{H}_{ir}$ is affected by the noise at the relay node, which is not the case in the two-stage channel estimation scheme. However, the two-stage channel estimation algorithm has a higher computational complexity than that of the superimposed channel training scheme, since both β and η need to be optimized in the former algorithm. Such performance-complexity tradeoff can be exploited in practical two-way MIMO relay communication systems.

3.6 Conclusions

In this chapter, we have proposed and investigated the performance of two channel estimation algorithms, namely, the superimposed channel training and two-stage channel estimation schemes, for frequency-flat two-way MIMO relay communication systems. The proposed algorithms can efficiently estimate the individual CSI for two-way MIMO relay systems, with the two-stage channel estimation algorithm performs better than the superimposed channel training scheme at a higher computational complexity.

3.A Proof of Theorem 3.1

Let us introduce the EVD of $\mathbf{R}_{ir} = \mathbf{Q}_i \mathbf{\Delta}_i \mathbf{Q}_i^H$. We can equivalently rewrite (3.13) and (3.16) as

$$\mathbf{R}_{\bar{v}_i} = \mathbf{I}_L \otimes \big(\mathbf{Q}_i(\alpha \mathrm{tr}(\mathbf{C}_r^T)\mathbf{\Delta}_i + \mathbf{I}_{N_i})\mathbf{Q}_i^H\big), \quad i = 1, 2 \tag{3.75}$$

$$\mathbf{R}_{\gamma_i} = \mathbf{U}_{\gamma_i} \mathrm{Bdiag}\big[\mathbf{\Lambda}_i \otimes b_i \mathbf{\Delta}_i,\ \mathbf{\Lambda}_{\bar{i}} \otimes b_{\bar{i}} \mathbf{\Delta}_i,\ \mathbf{\Lambda}_r \otimes \mathbf{\Delta}_i\big] \mathbf{U}_{\gamma_i}^H, \quad i = 1, 2 \tag{3.76}$$

where $\mathbf{U}_{\gamma_i} \triangleq \text{Bdiag}\big[\mathbf{I}_{N_i} \otimes \mathbf{Q}_i,\ \mathbf{I}_{N_{\bar{i}}} \otimes \mathbf{Q}_i,\ \mathbf{I}_{N_r} \otimes \mathbf{Q}_i\big]$, $i = 1, 2$. Substituting (3.75) and (3.76) back into (3.18), MSE_i can be rewritten as

$$\text{MSE}_i = \text{tr}\left(\left[\left[\left[\begin{array}{ccc} \mathbf{D}_{ii} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{i\bar{i}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_{si} \end{array}\right] + \left(\begin{array}{c} \sqrt{\alpha}\tilde{\mathbf{S}}_i^* \otimes \mathbf{I}_{N_i} \\ \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^* \otimes \mathbf{I}_{N_i} \\ \tilde{\mathbf{S}}_r^* \otimes \mathbf{I}_{N_i} \end{array}\right)\right] \times \Big(\mathbf{I}_L \otimes \mathbf{D}_{ri}\Big)\Big(\sqrt{\alpha}\tilde{\mathbf{S}}_i^T \otimes \mathbf{I}_{N_i},\ \sqrt{\alpha}\tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i},\ \tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}\Big)\right]^{-1}\right) \quad (3.77)$$

where

$$\mathbf{D}_{ij} \triangleq \mathbf{\Lambda}_j^{-1} \otimes (b_j \mathbf{\Delta}_i)^{-1},\ \ j = i, \bar{i}, \qquad \mathbf{D}_{si} \triangleq \mathbf{\Lambda}_r^{-1} \otimes \mathbf{\Delta}_i^{-1},$$
$$\mathbf{D}_{ri} \triangleq \big(\alpha \text{tr}(\mathbf{C}_r^T)\mathbf{\Delta}_i + \mathbf{I}_{N_i}\big)^{-1} \quad (3.78)$$

are all diagonal matrices. It can be seen from (3.77) that the objective function (3.21) is minimized only if

$$(\tilde{\mathbf{S}}_i^* \otimes \mathbf{I}_{N_i})(\mathbf{I}_L \otimes \mathbf{D}_{ri})(\tilde{\mathbf{S}}_{\bar{i}}^T \otimes \mathbf{I}_{N_i}) = \big(\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_{\bar{i}}^T\big) \otimes \mathbf{D}_{ri} = \mathbf{0} \quad (3.79)$$

$$(\tilde{\mathbf{S}}_j^* \otimes \mathbf{I}_{N_i})(\mathbf{I}_L \otimes \mathbf{D}_{ri})(\tilde{\mathbf{S}}_r^T \otimes \mathbf{I}_{N_i}) = \big(\tilde{\mathbf{S}}_j^* \tilde{\mathbf{S}}_r^T\big) \otimes \mathbf{D}_{ri} = \mathbf{0} \quad (3.80)$$

for $i = 1, 2$, and $j = i, \bar{i}$. Equations (3.79) and (3.80) hold if and only if $\tilde{\mathbf{S}}_1^* \tilde{\mathbf{S}}_2^T = \mathbf{0}$ and $\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_r^T = \mathbf{0}$, $i = 1, 2$, or equivalently $\mathbf{S}_1 \mathbf{S}_2^H = \mathbf{0}$ and $\mathbf{S}_i \mathbf{S}_r^H = \mathbf{0}$, $i = 1, 2$. Then the objective function (3.21) can be written as

$$\sum_{i=1}^{2} \text{tr}\left(\sum_{j=1}^{2} \big[\mathbf{D}_{ij} + \alpha \tilde{\mathbf{S}}_j^* \tilde{\mathbf{S}}_j^T \otimes \mathbf{D}_{ri}\big]^{-1} + \big[\mathbf{D}_{si} + \tilde{\mathbf{S}}_r^* \tilde{\mathbf{S}}_r^T \otimes \mathbf{D}_{ri}\big]^{-1}\right). \quad (3.81)$$

Since $\mathbf{D}_{ij}$, $\mathbf{D}_{si}$, and $\mathbf{D}_{ri}$ are all diagonal, to minimize (3.81), $\tilde{\mathbf{S}}_i^* \tilde{\mathbf{S}}_i^T$, $i = 1, 2, r$, must be diagonal. Note that the diagonality of $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H$ does not change $\text{tr}(\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H)$, $i = 1, 2, r$, in the constraints (3.22) and (3.23). We would like to note that $\text{tr}(\mathbf{\Lambda}_i \tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H)$ in the constraints (3.23) is minimized if $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H$ is diagonal and its diagonal entries are in the inverse order of that of $\mathbf{\Lambda}_i$ [101]. Denoting $\tilde{\mathbf{S}}_i \tilde{\mathbf{S}}_i^H = \mathbf{\Sigma}_i$, $i = 1, 2, r$, then we have $\mathbf{S}_i \mathbf{S}_i^H = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{U}_i^H$, $i = 1, 2, r$. □

Chapter 4

Frequency-Selective Two-Way MIMO Relay Systems

In this chapter, the channel estimation problem for two-way MIMO relay communication systems in frequency-selective fading environments is investigated. After an overview of the existing techniques in Section 4.1, the system model of a two-way MIMO relay system in frequency-selective fading environments is presented in Section 4.2. The method of superimposed channel training is developed in Section 4.3 to estimate the CSI of the first-hop and second-hop links for two-way MIMO relay systems with frequency-selective fading channels. The optimal structure and power allocation of the source and relay training sequences are derived to minimize the MSE of channel estimation. Moreover, taking into account the estimation error inherited from the estimation of the second-hop channel matrices, a novel MMSE-based algorithm is developed to retrieve the first-hop channel matrices. Numerical examples are shown in Section 4.4 to demonstrate the performance of the proposed superimposed channel training algorithm for two-way MIMO relay systems in frequency-selective fading environments. In Section 4.5, summaries on this chapter are given. The proof of Theorem 4.1 and Theorem 4.2 are presented in Section 4.A and 4.B, respectively.

4.1 Overview of the existing works

Due to a larger number of unknowns, channel estimation problems are generally more challenging in two-way relay systems than those in one-way relay systems. In [96], two-way relay channel estimation algorithms based on the ML and linear maximum SNR criteria have been proposed. However, the algorithms in [96] were designed for single antenna relay systems, and the extension to MIMO relay systems is not straightforward. Two methods were presented in [98] for two-way MIMO relay systems, namely, cascaded channel estimation and individual channel estimation. In the first algorithm, the cascaded channel matrices are estimated at two source nodes. However, this approach cannot estimate the individual second-hop channel matrices, which are essential for the optimization of MIMO relay networks [95].

This problem has been addressed by the superimposed channel training algorithm for two-way MIMO relay systems in [68], where a training sequence is superimposed at the relay node. The purpose of superimposing a training matrix at the relay node is to estimate the CSI of *individual* first-hop and second-hop channel matrices at the destination nodes, which cannot be achieved by simply multiplying the received signals at the relay node with a relay precoding matrix. Individual CSI can also be obtained by first estimating the first-hop channel matrices at the relay node and then forwarding the estimated channel matrices to the destination nodes, as the individual channel estimation algorithm in [98]. Obviously, the approach in [98] increases the cost and complexity of the relay node.

The relay systems in [68], [96], and [98] are assumed to have frequency-flat fading channels, which is only valid for narrowband communication systems. In this chapter, we consider a more general situation where two-way MIMO relay systems are operating in frequency-selective fading environments, i.e., there are multiple paths between each transmit-receive antenna pair. We apply the method of superimposed channel training to estimate the individual channel matrices of the first-hop and second-hop links for two-way MIMO relay systems in frequency-selective fading environments. In particular, the channel training is completed in two time blocks. In the first time block, both source nodes transmit their

training sequences simultaneously to the relay node. The relay then amplifies the received signals and superimposes its own training sequences before broadcasting the superimposed signals to the destination nodes. The channel estimation processes are implemented at the destination nodes to minimize the amount of signal processing at the relay node.

Since the superimposed channel training approach does not require the relay node to be capable of performing the advanced signal processing of channel estimation, and hence, provides an easy and cost-effective implementation of two-way relay communication systems. Such advantage of the superimposed channel training approach is particularly important under frequency-selective channels, as the complexity at the relay node increases significantly compared with the frequency-flat fading environment when the approach in [98] is used. Thus, the superimposed channel estimation method is preferred from practical point of view.

We derive the optimal source and relay training sequences by minimizing the sum MSE of channel estimation. We also optimize the power allocation between the source and relay training sequences at the relay node. The algorithm developed in this chapter generalizes the results in Chapter 3 from frequency-flat fading channel to frequency-selective fading channels. We would like to note that such extension is non-trivial as the optimization problem for channel estimation in frequency-selective two-way MIMO relay systems is much more complicated than that of frequency-flat relay systems. Moreover, we develop a new MMSE-based algorithm to retrieve the first-hop channel matrices, which takes into account the estimation error inherited from the estimation of the second-hop channel matrices.

4.2 Frequency-Selective Two-Way MIMO Relay System Model

We consider a three-node two-way MIMO relay communication system operating in a frequency-selective fading environment, where two source nodes, node 1 and node 2, exchange information through a relay node as shown in Fig. 4.1. The source nodes and relay node are equipped with N_s and N_r antennas, respectively. In this chapter, we assume that the practical half-duplex mode is used at all

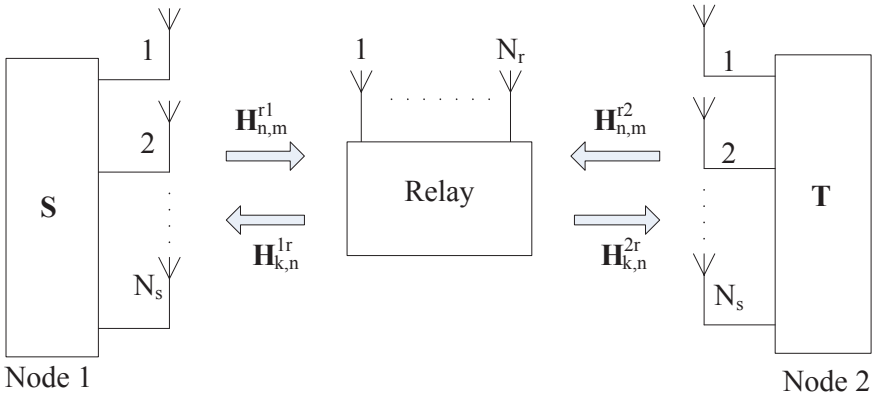

Figure 4.1: Block diagram of a two-way MIMO relay communication system.

nodes, i.e., each node is not able to send and receive signals at the same time. With this assumption, there is no direct link between two source nodes as both source nodes are transmitting signals at the first time block and cannot receive signals from each other. The implementation of half-duplex mode is common in two-way relay communications.

Let us denote $\mathbf{h}_{n,m}^{ri} = \left[h_{n,m,1}^{ri}, \cdots, h_{n,m,Q}^{ri}\right]^T$ as the $Q \times 1$ first-hop multipath channel vector from the mth antenna at node i to the nth antenna at the relay node, $i = 1, 2$, $m = 1, \cdots, N_s$, and $n = 1, \cdots, N_r$, where we assume that all channels have the same number of taps Q. The extension to systems with different number of channel taps between each transmit and receive antenna pair is straightforward. In a similar way, $\mathbf{h}_{n,m}^{ir} = \left[h_{n,m,1}^{ir}, \cdots, h_{n,m,Q}^{ir}\right]^T$ is used to denote the $Q \times 1$ second-hop multipath channel vector from the mth antenna at the relay node to the nth antenna at node i.

The channel estimation process is completed in two time blocks. In the first time block, source node 1 transmits an $N_s \times L$ training signal matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{N_s}]^T$ and node 2 transmits an $N_s \times L$ training matrix $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_{N_s}]^T$, respectively, where $L > Q$ is the length of the training sequence and will be determined later. Cyclic prefixes of length $L_{cp} \geq Q$ are inserted at $\mathbf{s}_m$ and $\mathbf{t}_n$, $m, n = 1, \cdots, N_s$, to prevent the inter-block interference at the relay node [85]. The received signal vectors at the relay node over L time slots after removing the

cyclic prefix can be written as

$$\begin{aligned}\mathbf{y}_{r,n} &= \sum_{m=1}^{N_s} \mathbf{H}_{n,m}^{r1}\mathbf{s}_m + \sum_{m=1}^{N_s} \mathbf{H}_{n,m}^{r2}\mathbf{t}_m + \mathbf{v}_{r,n} \\ &= \sum_{m=1}^{N_s} \mathbf{C}_Q(\mathbf{s}_m)\mathbf{h}_{n,m}^{r1} + \sum_{m=1}^{N_s} \mathbf{C}_Q(\mathbf{t}_m)\mathbf{h}_{n,m}^{r2} + \mathbf{v}_{r,n}, \quad n = 1, \cdots, N_r \quad (4.1)\end{aligned}$$

where $\mathbf{y}_{r,n}$ and $\mathbf{v}_{r,n}$ are the $L \times 1$ received signal vector and noise vector at the nth antenna of the relay node, respectively, $\mathbf{H}_{n,m}^{r1}$ and $\mathbf{H}_{n,m}^{r2}$ are $L \times L$ circulant channel matrices whose first columns are given by $\left[(\mathbf{h}_{n,m}^{r1})^T, \mathbf{0}_{1\times(L-Q)}\right]^T$ and $\left[(\mathbf{h}_{n,m}^{r2})^T, \mathbf{0}_{1\times(L-Q)}\right]^T$, respectively, and $\mathbf{C}_Q(\mathbf{s})$ represents an $L \times Q$ column-wise circulant matrix taking $\mathbf{s}$ as the first column.

In the second time block, the relay node amplifies $\mathbf{y}_{r,n}$, $n = 1, \cdots, N_r$, and superimposes its own training matrix $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{N_r}]^T$. Thus, the signal vector transmitted by the nth antenna of the relay node is given by

$$\mathbf{x}_{r,n} = \sqrt{\alpha}\,\mathbf{y}_{r,n} + \mathbf{r}_n, \quad n = 1, \cdots, N_r \quad (4.2)$$

where $\alpha > 0$ is the relay amplifying factor. Similarly, a cyclic prefix is inserted at $\mathbf{x}_{r,n}$ prior to the transmission. The received signal vectors at the source node i, $i = 1, 2$, after removing the cyclic prefix are given by[1]

$$\mathbf{y}_{i,k} = \sum_{n=1}^{N_r} \mathbf{H}_{k,n}^{ir}\mathbf{x}_{r,n} + \mathbf{v}_{i,k}, \quad k = 1, \cdots, N_s \quad (4.3)$$

where $\mathbf{y}_{i,k}$ and $\mathbf{v}_{i,k}$ are the $L \times 1$ received signal vector and noise vector at the kth antenna of node i, respectively, $\mathbf{H}_{k,n}^{ir}$ is an $L \times L$ circulant channel matrix whose first column is $\left[(\mathbf{h}_{k,n}^{ir})^T, \mathbf{0}_{1\times(L-Q)}\right]^T$.

The main idea of the superimposed channel training algorithm is to exploit $\mathbf{R}$ to estimate the second-hop channels $\{\mathbf{h}_{k,n}^{ir}\} \triangleq \{\mathbf{h}_{k,n}^{ir}, i = 1, 2, k = 1, \cdots, N_s, n = 1, \cdots, N_r\}$, and then estimate the first-hop channels $\{\mathbf{h}_{n,m}^{ri}\} \triangleq \{\mathbf{h}_{n,m}^{ri}, i = 1, 2, n = 1, \cdots, N_r, m = 1, \cdots, N_s\}$ using $\mathbf{S}$, $\mathbf{T}$, and the estimated $\{\mathbf{h}_{k,n}^{ir}\}$. In this chapter, we assume that

1. All channel taps are zero-mean circularly symmetric complex Gaussian (CSCG) random variables.

[1] In this chapter, cyclic prefix is removed to facilitate the superimposed channel training algorithm. It is an interesting topic to combine the proposed approach and the channel estimation using the cyclic prefix, which may improve the accuracy of channel estimation.

2. Channel taps associated with the same transmit-receive antenna pair, as well as different transmit-receive antenna pairs are independent from each other.

3. Channels are assumed to be quasi-static, i.e., channels do not change within one cycle of transmission.

4. All noises are i.i.d. AWGN with zero mean and unit variance.

4.3 MMSE-Based Optimal Training Matrices

In this section, we design the optimal training matrices $\mathbf{S}$, $\mathbf{T}$, $\mathbf{R}$, and the relay amplifying factor α to minimize the MSE of channel estimation. By substituting (4.1) and (4.2) into (4.3), we obtain

$$\begin{aligned}\mathbf{y}_{i,k} &= \sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\left(\sqrt{\alpha}\sum_{m=1}^{N_s}\mathbf{H}_{n,m}^{r1}\mathbf{s}_m+\sqrt{\alpha}\sum_{m=1}^{N_s}\mathbf{H}_{n,m}^{r2}\mathbf{t}_m+\mathbf{r}_n+\sqrt{\alpha}\mathbf{v}_{r,n}\right)+\mathbf{v}_{i,k}\\ &= \sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{H}_{n,m}^{r1}\mathbf{s}_m+\sqrt{\alpha}\sum_{m=1}^{N_s}\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{H}_{n,m}^{r2}\mathbf{t}_m+\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{r}_n+\bar{\mathbf{v}}_{i,k},\\ &\quad k=1,\cdots,N_s \end{aligned} \tag{4.4}$$

where

$$\bar{\mathbf{v}}_{i,k}\triangleq\sqrt{\alpha}\sum_{n=1}^{N_r}\mathbf{H}_{k,n}^{ir}\mathbf{v}_{r,n}+\mathbf{v}_{i,k},\quad k=1,\cdots,N_s \tag{4.5}$$

is the equivalent noise vector at the kth antenna of node i. Since both $\mathbf{H}_{k,n}^{ir}$ and $\mathbf{H}_{n,m}^{ri}$ are circulant matrices, (4.4) can be rewritten by exploiting the property of circulant matrix as

$$\begin{aligned}\mathbf{y}_{i,k} &= \sqrt{\alpha}\sum_{m=1}^{N_s}\left[\mathbf{C}_{2Q-1}(\mathbf{s}_m)\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir}*\mathbf{h}_{n,m}^{r1}\right]\\ &\quad+\sqrt{\alpha}\sum_{m=1}^{N_s}\left[\mathbf{C}_{2Q-1}(\mathbf{t}_m)\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir}*\mathbf{h}_{n,m}^{r2}\right]+\sum_{n=1}^{N_r}\mathbf{C}_Q(\mathbf{r}_n)\mathbf{h}_{k,n}^{ir}+\bar{\mathbf{v}}_{i,k}\\ &= \sqrt{\alpha}\mathbf{\Phi}(\mathbf{s})\mathbf{d}_k^{i1}+\sqrt{\alpha}\mathbf{\Phi}(\mathbf{t})\mathbf{d}_k^{i2}+\mathbf{\Phi}(\mathbf{r})\mathbf{d}_k^{ir}+\bar{\mathbf{v}}_{i,k},\quad k=1,\cdots,N_s \end{aligned} \tag{4.6}$$

where

$$\mathbf{\Phi}(\mathbf{s}) \triangleq [\mathbf{C}_{2Q-1}(\mathbf{s}_1), \mathbf{C}_{2Q-1}(\mathbf{s}_2), \cdots, \mathbf{C}_{2Q-1}(\mathbf{s}_{N_s})] \in \mathcal{C}^{L\times(2Q-1)N_s} \tag{4.7}$$

$$\mathbf{\Phi}(\mathbf{t}) \triangleq [\mathbf{C}_{2Q-1}(\mathbf{t}_1), \mathbf{C}_{2Q-1}(\mathbf{t}_2), \cdots, \mathbf{C}_{2Q-1}(\mathbf{t}_{N_s})] \in \mathcal{C}^{L\times(2Q-1)N_s} \tag{4.8}$$

$$\mathbf{\Phi}(\mathbf{r}) \triangleq [\mathbf{C}_{Q}(\mathbf{r}_1), \mathbf{C}_{Q}(\mathbf{r}_2), \cdots, \mathbf{C}_{Q}(\mathbf{r}_{N_r})] \in \mathcal{C}^{L\times QN_r} \tag{4.9}$$

$$\mathbf{d}_k^{i1} \triangleq \left[\left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,1}^{r1}\right)^T, \left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,2}^{r1}\right)^T, \cdots, \left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,N_s}^{r1}\right)^T\right]^T \tag{4.10}$$

$$\mathbf{d}_k^{i2} \triangleq \left[\left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,1}^{r2}\right)^T, \left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,2}^{r2}\right)^T, \cdots, \left(\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,N_s}^{r2}\right)^T\right]^T \tag{4.11}$$

$$\mathbf{d}_k^{ir} \triangleq \left[(\mathbf{h}_{k,1}^{ir})^T, (\mathbf{h}_{k,2}^{ir})^T, \cdots, (\mathbf{h}_{k,N_r}^{ir})^T\right]^T. \tag{4.12}$$

Here $\mathbf{d}_k^{i1}$ in (4.10) and $\mathbf{d}_k^{i2}$ in (4.11) can be viewed as the compound channel from all antennas of node 1 and node 2 to the kth antenna at node i, respectively, and $\mathbf{d}_k^{ir}$ in (4.12) is the channel from all antennas of the relay node to the kth antennas at node i.

By introducing

$$\mathbf{A} \triangleq [\sqrt{\alpha}\mathbf{\Phi}(\mathbf{s}), \sqrt{\alpha}\mathbf{\Phi}(\mathbf{t}), \mathbf{\Phi}(\mathbf{r})] \in \mathcal{C}^{L\times((4Q-2)N_s+QN_r)} \tag{4.13}$$

$$\boldsymbol{\theta}_{i,k} \triangleq \left[(\mathbf{d}_k^{i1})^T, (\mathbf{d}_k^{i2})^T, (\mathbf{d}_k^{ir})^T\right]^T, \quad k = 1, \cdots, N_s \tag{4.14}$$

we can rewrite (4.6) as

$$\mathbf{y}_{i,k} = \mathbf{A}\boldsymbol{\theta}_{i,k} + \bar{\mathbf{v}}_{i,k}, \quad k = 1, \cdots, N_s. \tag{4.15}$$

Here $\boldsymbol{\theta}_{i,k}$ in (4.14) is the vector of unknowns that need to be estimated at node i.

Due to its simplicity, a linear estimator is applied at node i to estimate $\boldsymbol{\theta}_{i,k}$ as

$$\hat{\boldsymbol{\theta}}_{i,k} = \mathbf{W}_{i,k}^H \mathbf{y}_{i,k}, \quad k = 1, \cdots, N_s, \quad i = 1, 2 \tag{4.16}$$

where $\hat{\boldsymbol{\theta}}_{i,k}$ denotes an estimation of $\boldsymbol{\theta}_{i,k}$ and $\mathbf{W}_{i,k}$ is the weight matrix of the linear receiver. As a linear estimator is used, we can see from (4.13) that the length of the training sequences should satisfy $L \geq (4Q-2)N_s + QN_r$. Based on (4.15) and (4.16), the sum MSE of channel estimation at two nodes can be

written as

$$\begin{aligned}\text{MSE} &= \sum_{i=1}^{2}\sum_{k=1}^{N_s}\text{tr}\big(\text{E}\big[(\hat{\boldsymbol{\theta}}_{i,k}-\boldsymbol{\theta}_{i,k})(\hat{\boldsymbol{\theta}}_{i,k}-\boldsymbol{\theta}_{i,k})^H\big]\big)\\ &= \sum_{i=1}^{2}\sum_{k=1}^{N_s}\text{tr}\big(\text{E}\big[(\mathbf{W}_{i,k}^H\mathbf{A}-\mathbf{I}_B)\mathbf{C}_x^{i,k}(\mathbf{W}_{i,k}^H\mathbf{A}-\mathbf{I}_B)^H+\mathbf{W}_{i,k}^H\mathbf{C}_{\bar{v}}^{i,k}\mathbf{W}_{i,k}\big]\big) \quad (4.17)\end{aligned}$$

where $B \triangleq (4Q-2)N_s+QN_r$, $\mathbf{C}_x^{i,k}=\text{E}\big[\boldsymbol{\theta}_{i,k}\boldsymbol{\theta}_{i,k}^H\big]$ is the covariance matrix of $\boldsymbol{\theta}_{i,k}$, and $\mathbf{C}_{\bar{v}}^{i,k}=\text{E}\big[\bar{\mathbf{v}}_{i,k}\bar{\mathbf{v}}_{i,k}^H\big]$ is the noise covariance matrix.

From (4.5), we have

$$\mathbf{C}_{\bar{v}}^{i,k}=\left(\alpha\sum_{n=1}^{N_r}\sum_{j=1}^{Q}\sigma_{k,n,j}^{ir}+1\right)\mathbf{I}_L,\quad i=1,2,\quad k=1,\cdots,N_s$$

where $\sigma_{k,n,j}^{ir}=\text{E}\big[h_{k,n,j}^{ir}\big(h_{k,n,j}^{ir}\big)^*\big]$ is the variance of $h_{k,n,j}^{ir}$, $j=1,\cdots,Q$. Based on (4.10)-(4.12) and (4.14), we obtain that $\mathbf{C}_x^{i,k}=\text{Bdiag}\big[\mathbf{C}_{i1}^k,\mathbf{C}_{i2}^k,\mathbf{C}_{ir}^k\big]$, where

$$\mathbf{C}_{ij}^k = \text{E}\big[\mathbf{d}_k^{ij}(\mathbf{d}_k^{ij})^H\big]=\text{Bdiag}\big[\mathbf{C}_{k,1}^{ij},\cdots,\mathbf{C}_{k,N_s}^{ij}\big],\quad j=1,2 \quad (4.18)$$

$$\mathbf{C}_{ir}^k = \text{E}\big[\mathbf{d}_k^{ir}(\mathbf{d}_k^{ir})^H\big]=\text{Bdiag}\big[\mathbf{C}_{k,1}^{ir},\cdots,\mathbf{C}_{k,N_r}^{ir}\big]. \quad (4.19)$$

By introducing $\boldsymbol{\sigma}_{k,n}^{ir}=\big[\sigma_{k,n,1}^{ir},\cdots,\sigma_{k,n,Q}^{ir}\big]^T$ and $\boldsymbol{\sigma}_{n,m}^{rj}=\big[\sigma_{n,m,1}^{rj},\cdots,\sigma_{n,m,Q}^{rj}\big]^T$, where $\sigma_{n,m,p}^{rj}=\text{E}\big[h_{n,m,p}^{rj}\big(h_{n,m,p}^{rj}\big)^*\big]$ is the variance of $h_{n,m,p}^{rj}$, $j=1,2$, $p=1,\cdots,Q$, we obtain that

$$\begin{aligned}\mathbf{C}_{k,m}^{ij} &= \text{E}\left[\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir}*\mathbf{h}_{n,m}^{rj}\right)\left(\sum_{n=1}^{N_r}\mathbf{h}_{k,n}^{ir}*\mathbf{h}_{n,m}^{rj}\right)^H\right]\\ &= \sum_{n=1}^{N_r}\text{diag}\big[\boldsymbol{\sigma}_{k,n}^{ir}*\boldsymbol{\sigma}_{n,m}^{rj}\big],\quad j=1,2,\quad m=1,\cdots,N_s\\ \mathbf{C}_{k,n}^{ir} &= \text{E}\big[\mathbf{h}_{k,n}^{ir}(\mathbf{h}_{k,n}^{ir})^H\big]\\ &= \text{diag}\big[\sigma_{k,n,1}^{ir},\cdots,\sigma_{k,n,Q}^{ir}\big],\quad n=1,\cdots,N_r.\end{aligned}$$

4.3.1 Structure of Optimal Training Sequences

The matrices $\mathbf{W}_{i,k}, i=1,2,k=1,\cdots,N_s$ that minimize MSE in (4.17) are given by

$$\mathbf{W}_{i,k}=\left(\mathbf{A}\mathbf{C}_x^{i,k}\mathbf{A}^H+\mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\mathbf{C}_x^{i,k},\quad i=1,2,\quad k=1,\cdots,N_s. \quad (4.20)$$

Substituting (4.20) back into (4.17), the MSE of channel estimation at both source nodes can be written as

$$\text{MSE} = \sum_{i=1}^{2}\sum_{k=1}^{N_s} \text{tr}\left(\left[\left(\mathbf{C}_x^{i,k}\right)^{-1} + \mathbf{A}^H\left(\mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\right]^{-1}\right). \tag{4.21}$$

The transmission power constraints at the source nodes are given by

$$\sum_{m=1}^{N_s} \mathbf{s}_m^H\mathbf{s}_m \le p_1, \quad \sum_{m=1}^{N_s} \mathbf{t}_m^H\mathbf{t}_m \le p_2 \tag{4.22}$$

where p_1 and p_2 are the transmission power available at source nodes 1 and 2, respectively. From (4.1) and (4.2), the transmission power constraint at the relay node is given by

$$\begin{aligned}&\sum_{n=1}^{N_r} \text{E}\Big[\text{tr}\big(\mathbf{x}_{r,n}\mathbf{x}_{r,n}^H\big)\Big] \\ &= \sum_{n=1}^{N_r}\left(\alpha\text{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m) + \mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right) + \mathbf{I}_L\right)\right. \\ &\qquad \left. + \mathbf{r}_n^H\mathbf{r}_n\right) \le p_r\end{aligned} \tag{4.23}$$

where $\mathbf{D}_{n,m}^{ri} \triangleq \text{diag}\big[\sigma_{n,m,1}^{ri}, \cdots, \sigma_{n,m,Q}^{ri}\big]$, $i = 1, 2$, and p_r is the transmission power available at the relay node. It can be seen from (4.23) that the feasible region of α depends on p_r as $0 < \alpha < (p_r - \sum_{n=1}^{N_r}\mathbf{r}_n^H\mathbf{r}_n)/\Omega$, where

$$\Omega \triangleq \sum_{n=1}^{N_r}\text{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m) + \mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right) + \mathbf{I}_L\right).$$

From (4.21)-(4.23), the optimal training sequences and the optimal α design problem can be written as

$$\min_{\mathbf{S},\mathbf{T},\mathbf{R},\alpha>0} \quad \sum_{i=1}^{2}\sum_{k=1}^{N_s}\text{tr}\left(\left[\left(\mathbf{C}_x^{i,k}\right)^{-1} + \mathbf{A}^H\left(\mathbf{C}_{\bar{v}}^{i,k}\right)^{-1}\mathbf{A}\right]^{-1}\right) \tag{4.24}$$

$$\text{s.t.} \quad \sum_{m=1}^{N_s}\mathbf{s}_m^H\mathbf{s}_m \le p_1 \tag{4.25}$$

$$\sum_{m=1}^{N_s}\mathbf{t}_m^H\mathbf{t}_m \le p_2 \tag{4.26}$$

$$\begin{aligned}&\sum_{n=1}^{N_r}\left(\alpha\text{tr}\left(\sum_{m=1}^{N_s}\left(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m) + \mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\right) + \mathbf{I}_L\right)\right. \\ &\qquad \left. + \mathbf{r}_n^H\mathbf{r}_n\right) \le p_r.\end{aligned} \tag{4.27}$$

The following theorem establishes the optimal structure of $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ as the solution to the problem (4.24)-(4.27).

THEOREM 4.1: The optimal training matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ satisfy the following equations for all $m, n = 1, \cdots, N_s$, and $p = 1, \cdots, N_r$

$$\begin{aligned}
&\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{s}_m) = \beta_m \mathbf{I}_{2Q-1}, \\
&\mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \gamma_n \mathbf{I}_{2Q-1}, \quad \mathbf{C}_Q^H(\mathbf{r}_p)\mathbf{C}_Q(\mathbf{r}_p) = \delta_p \mathbf{I}_Q \qquad (4.28) \\
&\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \mathbf{0}, \\
&\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0}, \quad \mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0} \qquad (4.29)
\end{aligned}$$

where $\beta_m = \mathbf{s}_m^H \mathbf{s}_m$, $\gamma_n = \mathbf{t}_n^H \mathbf{t}_n$, and $\delta_p = \mathbf{r}_p^H \mathbf{r}_p$.

PROOF: See Appendix 4.A. □

It is worth noting that the training matrices $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{R}$ satisfying (4.28) and (4.29) are not unique in general. Indeed, we are not particularly interested in a unique solution of the problem. The minimum MSE of channel estimation is achieved as long as the training matrices satisfy (4.28) and (4.29). One example of achieving (4.28) and (4.29) is given below

$$\begin{aligned}
\mathbf{s}_1 &= \mathbf{F}\tilde{\mathbf{s}}_1, \ |\tilde{s}_{1,i}| = \sqrt{\beta_1/L}, \quad i = 1, \cdots, L \\
\mathbf{s}_m &= \mathbf{F}\tilde{\mathbf{s}}_m, \ \tilde{s}_{m,i} = \sqrt{\beta_m/\beta_1}\tilde{s}_{1,i}\, e^{j2\pi(i-1)(2Q-1)(m-1)/L}, \\
&\qquad i = 1, \cdots, L, \quad m = 2, \cdots, N_s \\
\mathbf{t}_m &= \mathbf{F}\tilde{\mathbf{t}}_m, \ \tilde{t}_{m,i} = \sqrt{\gamma_m/\beta_1}\tilde{s}_{1,i}\, e^{j2\pi(i-1)(2Q-1)(N_s-1+m)/L}, \\
&\qquad i = 1, \cdots, L, \quad m = 1, \cdots, N_s \\
\mathbf{r}_n &= \mathbf{F}\tilde{\mathbf{r}}_n, \ \tilde{r}_{n,i} = \sqrt{\delta_n/\beta_1}\tilde{s}_{1,i}\, e^{j2\pi(i-1)[(2Q-1)(2N_s-1)+Q(n-1)]/L}, \\
&\qquad i = 1, \cdots, L, \quad n = 1, \cdots, N_r
\end{aligned}$$

where $j = \sqrt{-1}$ and $\mathbf{F}$ is an $L \times L$ normalized fast Fourier transform (FFT) matrix with $[\mathbf{F}]_{m,n} = (1/\sqrt{L})e^{-j2\pi(m-1)(n-1)/L}$. The training matrices shown above have the advantage that they are easy to implement as the elements of $\tilde{\mathbf{s}}_m$ (also $\tilde{\mathbf{t}}_m$ and $\tilde{\mathbf{r}}_n$) have a constant magnitude.

4.3.2 Optimal Power Loading

Applying Theorem 4.1, the MSE function in (4.21) can be written as

$$\begin{aligned}\text{MSE} \;=\; \sum_{i=1}^{2}\sum_{k=1}^{N_s}\text{tr}\Bigg(&\sum_{m=1}^{N_s}\left[\left(\mathbf{C}_{k,m}^{i1}\right)^{-1}+\alpha\beta_m\eta_{i,k}\mathbf{I}_{2Q-1}\right]^{-1}\\ &+\sum_{m=1}^{N_s}\left[\left(\mathbf{C}_{k,m}^{i2}\right)^{-1}+\alpha\gamma_m\eta_{i,k}\mathbf{I}_{2Q-1}\right]^{-1}\\ &+\sum_{n=1}^{N_r}\left[\left(\mathbf{C}_{k,n}^{ir}\right)^{-1}+\delta_n\eta_{i,k}\mathbf{I}_{Q}\right]^{-1}\Bigg)\end{aligned} \tag{4.30}$$

where $\eta_{i,k}$ is defined in (4.56). Let us denote $c_{k,m,q}^{ij} \triangleq \left[\left(\mathbf{C}_{k,m}^{ij}\right)^{-1}\right]_{q,q}$, $c_{k,n,p}^{ir} \triangleq \left[\left(\mathbf{C}_{k,n}^{ir}\right)^{-1}\right]_{p,p}$, and $\kappa_{i,m} \triangleq \sum_{n=1}^{N_r}\sum_{q=1}^{Q}\sigma_{n,m,q}^{ri}$, $i=1,2$. The problem (4.24)-(4.27) with matrix variables can be equivalently rewritten as the following problem in scalar variables

$$\begin{aligned}\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\alpha}\quad &\sum_{m=1}^{N_s}\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\left(\frac{1}{c_{k,m,q}^{i1}+\alpha\beta_m\eta_{i,k}}+\frac{1}{c_{k,m,q}^{i2}+\alpha\gamma_m\eta_{i,k}}\right)\\ &+\sum_{n=1}^{N_r}\sum_{k=1}^{N_s}\sum_{p=1}^{Q}\sum_{i=1}^{2}\frac{1}{c_{k,n,p}^{ir}+\delta_n\eta_{i,k}} &(4.31)\\ \text{s.t.}\quad &\sum_{m=1}^{N_s}\beta_m\le p_1 &(4.32)\\ &\sum_{m=1}^{N_s}\gamma_m\le p_2 &(4.33)\\ &\alpha\left(\sum_{m=1}^{N_s}\kappa_{1,m}\beta_m+\sum_{m=1}^{N_s}\kappa_{2,m}\gamma_m\right)+\sum_{n=1}^{N_r}\delta_n+\alpha LN_r\le p_r &(4.34)\\ &\alpha>0,\;\;\beta_m\ge 0,\gamma_m\ge 0,m=1,\cdots,N_s,\;\;\delta_n\ge 0,n=1,\cdots,N_r &(4.35)\end{aligned}$$

where $\boldsymbol{\beta} \triangleq [\beta_1,\cdots,\beta_{N_s}]^T$, $\boldsymbol{\gamma} \triangleq [\gamma_1,\cdots,\gamma_{N_s}]^T$, and $\boldsymbol{\delta} \triangleq [\delta_1,\cdots,\delta_{N_r}]^T$.

Given that $c_{k,m,q}^{i1}$, $c_{k,m,q}^{i2}$, $c_{k,m,q}^{ir}$, and $\eta_{i,k}$ are known variables with fixed α, it can be observed that the fractions in the objective function (4.31) are monotonically decreasing and convex functions with respect to β_m, γ_m, and δ_n. Moreover, when α is fixed, the constraints in (4.32)-(4.35) are linear inequality constraints. Therefore, with fixed α, the problem (4.31)-(4.35) with respect to β_m, γ_m, and δ_n is a convex optimization problem where the optimal β_m, γ_m, and δ_n can be efficiently obtained through the KKT optimality conditions [99] of the problem

(4.31)-(4.35). The gradient conditions are given by

$$\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\frac{\alpha\eta_{i,k}}{\left(c_{k,m,q}^{i1}+\alpha\beta_m\eta_{i,k}\right)^2} = \mu_1+\mu_3\alpha\kappa_{1,m},\quad m=1,\cdots,N_s \tag{4.36}$$

$$\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\frac{\alpha\eta_{i,k}}{\left(c_{k,m,q}^{i2}+\alpha\gamma_m\eta_{i,k}\right)^2} = \mu_2+\mu_3\alpha\kappa_{2,m},\quad m=1,\cdots,N_s \tag{4.37}$$

$$\sum_{k=1}^{N_s}\sum_{p=1}^{Q}\sum_{i=1}^{2}\frac{\eta_{i,k}}{\left(c_{k,n,p}^{ir}+\delta_n\eta_{i,k}\right)^2} = \mu_3,\quad n=1,\cdots,N_r \tag{4.38}$$

where $\mu_i \geq 0$, $i=1,2,3$, are Lagrange multipliers such that the complementary slackness conditions given by

$$\mu_1\left(p_1-\sum_{m=1}^{N_s}\beta_m\right)=0 \tag{4.39}$$

$$\mu_2\left(p_2-\sum_{m=1}^{N_s}\gamma_m\right)=0 \tag{4.40}$$

$$\mu_3\left(p_r-\alpha NL-\sum_{n=1}^{N_r}\delta_n-\alpha\sum_{m=1}^{N_s}\kappa_{1,m}\beta_m-\alpha\sum_{m=1}^{N_s}\kappa_{2,m}\gamma_m\right)=0 \tag{4.41}$$

are satisfied.

When α and μ_i, $i=1,2,3$, are fixed, the non-negative β_m, γ_m, $m=1,\cdots,N_s$, and δ_n, $n=1,\cdots,N_r$, can be found by using the bi-section search, as the LHS of (4.36), (4.37), and (4.38) are monotonically decreasing functions of β_m, γ_m, and δ_n, respectively. An outer bi-section search is applied to find the optimal μ_i, $i=1,2,3$, since the LHS of (4.32) and (4.33) are increasing functions of β_m and γ_m, respectively, and the LHS of (4.34) is an increasing function of β_m, γ_m, and δ_n. Moreover, in (4.36), β_m is a monotonically decreasing function of μ_1 and μ_3, γ_m is monotonically decreasing with respect to μ_2 and μ_3 in (4.37), while in (4.38), δ_n is a monotonically decreasing function of μ_3.

When α is not fixed, i.e., α is an optimization variable, the problem (4.31)-(4.35) as a whole is not a convex optimization problem. However, the following theorem states that (4.31) is a unimodal function of α.

THEOREM 4.2: The objective function (4.31) subjecting to (4.32)-(4.35) is a unimodal (quasi-convex) function with respect to α.

PROOF: See Appendix 4.B. □

To verify Theorem 4.2, a plot of the MSE value over a range of feasible values of α is shown in Fig. 4.2. We consider the case where all nodes have the same

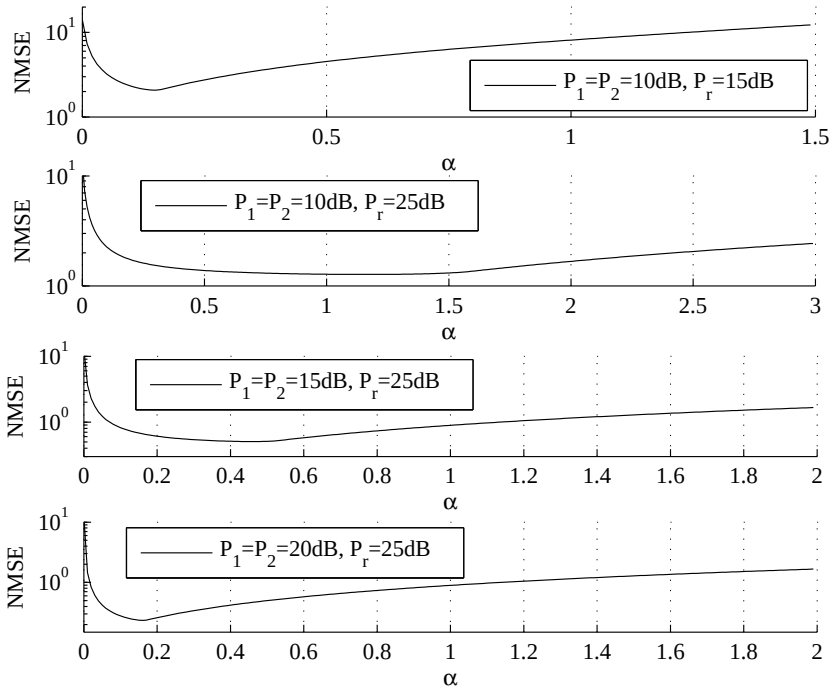

Figure 4.2: NMSE versus α for different $p_1 = p_2$ and p_r with $N = 2$ and $Q = 4$.

number of antennas, i.e., $N_s = N_r = N = 2$, and $Q = 4$. Fig. 4.2 shows the NMSE versus α for different $p_1 = p_2$ and p_r. Note that the NMSE is obtained by dividing (4.31) with $L = (5Q - 2)N$. It can be seen from Fig. 4.2 that (4.31) is a unimodal function of α. For a unimodal function, the minimum value can be efficiently found by the GSS [100] technique. Hence, the optimal α for the problem (4.31)-(4.35) can be found by applying the GSS technique as described in Table 4.1, where ε is a positive constant close to 0, and $\phi > 0$ is the reduction factor. It is shown in [100] that the optimal $\phi = 1.618$, also known as the golden ratio. It can be seen from Fig. 4.2 that the optimal value of α varies with p_1, p_2, and p_r. For fixed p_1 and p_2, the optimal α has a larger value when p_r increases. For a given p_r, the optimal value of α decreases when p_1 and p_2 increases.

Since at each iteration, the GSS method reduces the interval containing the optimal α to 0.618 times of the interval at the preceding iteration, the length of the interval of uncertainty after the nth iteration is $\Gamma_n = (0.618)^n \Gamma_0$, where Γ_0 is the length of the initial feasible interval [100]. Therefore, the complexity of the GSS method depends on the number of iterations, which is determined by the desired accuracy.

Table 4.1: Procedure of applying the golden section search (GSS) to find the optimal α in the problem (4.31)-(4.35).

---

1. Set a feasible bound $[a, b]$ on α.
2. Define $c_1 = (\phi - 1)a + (2 - \phi)b$ and $c_2 = (2 - \phi)a + (\phi - 1)b$.
3. Solve the problem (4.31)-(4.35) for $\alpha = c_1$;
   Compute the MSE value defined in (4.31), $f_{\text{MSE}}(c_1)$ for $\alpha = c_1$.
4. Repeat Step 3 for $\alpha = c_2$.
5. If $f_{\text{MSE}}(c_1) < f_{\text{MSE}}(c_2)$, then assign $b = c_2$.
   Otherwise, assign $a = c_1$.
6. If $|b - a| \leq \varepsilon$, then end.
   Otherwise, go to step 2.

4.3.3 Retrieving the Multipath Channel Vectors

Based on (4.12) and (4.14), the second-hop channels $\mathbf{h}_{k,n}^{ir}$, $i = 1, 2$ can be directly obtained from $\hat{\boldsymbol{\theta}}_{i,k}$. The first-hop channels $\mathbf{h}_{n,m}^{r1}$ and $\mathbf{h}_{n,m}^{r2}$ can be estimated based on $\hat{\boldsymbol{\theta}}_{i,k}$ as follows. Since $\mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r1} = \mathbf{T}(\mathbf{h}_{k,n}^{ir})\mathbf{h}_{n,m}^{r1}$, where $\mathbf{T}(\mathbf{h})$ stands for a $(2Q-1) \times Q$ circulant matrix taking $[\mathbf{h}^T, \mathbf{0}_{1\times(Q-1)}]^T$ as its first column, we have

$$\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r1} = \sum_{n=1}^{N_r} \mathbf{T}(\mathbf{h}_{k,n}^{ir})\mathbf{h}_{n,m}^{r1} = \mathbf{d}_{k,m}^{i1}, \quad k, m = 1, \cdots, N_s \tag{4.42}$$

$$\sum_{n=1}^{N_r} \mathbf{h}_{k,n}^{ir} * \mathbf{h}_{n,m}^{r2} = \sum_{n=1}^{N_r} \mathbf{T}(\mathbf{h}_{k,n}^{ir})\mathbf{h}_{n,m}^{r2} = \mathbf{d}_{k,m}^{i2}, \quad k, m = 1, \cdots, N_s. \tag{4.43}$$

Equations (4.42) and (4.43) can be represented in matrix form as

$$\boldsymbol{\Psi}_{ir}\mathbf{h}_m^{r1} = \mathbf{e}_m^{i1}, \quad m = 1, \cdots, N_s \tag{4.44}$$

$$\boldsymbol{\Psi}_{ir}\mathbf{h}_m^{r2} = \mathbf{e}_m^{i2}, \quad m = 1, \cdots, N_s \tag{4.45}$$

where

$$\boldsymbol{\Psi}_{ir} \triangleq \begin{bmatrix} \mathbf{T}(\mathbf{h}_{1,1}^{ir}), & \cdots, & \mathbf{T}(\mathbf{h}_{1,N_r}^{ir}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}(\mathbf{h}_{N_s,1}^{ir}), & \cdots, & \mathbf{T}(\mathbf{h}_{N_s,N_r}^{ir}) \end{bmatrix}, \ \mathbf{h}_m^{ri} \triangleq \begin{bmatrix} \mathbf{h}_{1,m}^{ri} \\ \vdots \\ \mathbf{h}_{N_r,m}^{ri} \end{bmatrix}, \ i = 1, 2,$$

$$\mathbf{e}_m^{ij} \triangleq \begin{bmatrix} \mathbf{d}_{1,m}^{ij} \\ \vdots \\ \mathbf{d}_{N_s,m}^{ij} \end{bmatrix}, \ j = 1, 2. \tag{4.46}$$

In the following, we develop an LMMSE estimator to retrieve the first-hop multipath channel vectors $\{\mathbf{h}_{n,m}^{ri}\}$. Taking into account the estimation errors in $\mathbf{h}_{k,m}^{ir}$ and $\mathbf{d}_{n,m}^{ij}$, we have

$$\mathbf{h}_{k,m}^{ir} = \hat{\mathbf{h}}_{k,m}^{ir} + \boldsymbol{\kappa}_{k,m}^{ir}, \quad \mathbf{d}_{n,m}^{ij} = \hat{\mathbf{d}}_{n,m}^{ij} + \boldsymbol{\iota}_{n,m}^{ij} \tag{4.47}$$

where $\hat{\mathbf{h}}_{k,m}^{ir}$ and $\hat{\mathbf{d}}_{n,m}^{ij}$ are the estimates of $\mathbf{h}_{k,m}^{ir}$ and $\mathbf{d}_{n,m}^{ij}$, respectively, obtained from $\hat{\boldsymbol{\theta}}_{i,k}$, and $\boldsymbol{\kappa}_{k,m}^{ir}$ and $\boldsymbol{\iota}_{n,m}^{ij}$ are the estimation error vectors. Substituting (4.47) back into (4.44) and (4.45), we have

$$\left(\hat{\boldsymbol{\Psi}}_{ir} + \boldsymbol{\Delta}_i\right) \mathbf{h}_m^{rj} = \hat{\mathbf{e}}_m^{ij} + \mathbf{g}_m^{ij}, \quad m = 1, \cdots, N_s, \ j = 1, 2 \tag{4.48}$$

where

$$\hat{\mathbf{\Psi}}_{ir} \triangleq \begin{bmatrix} \mathbf{T}(\hat{\mathbf{h}}^{ir}_{1,1}), & \cdots, & \mathbf{T}(\hat{\mathbf{h}}^{ir}_{1,N_r}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}(\hat{\mathbf{h}}^{ir}_{N_s,1}), & \cdots, & \mathbf{T}(\hat{\mathbf{h}}^{ir}_{N_s,N_r}) \end{bmatrix}, \hat{\mathbf{e}}^{ij}_m \triangleq \begin{bmatrix} \hat{\mathbf{d}}^{ij}_{1,m} \\ \vdots \\ \hat{\mathbf{d}}^{ij}_{N_s,m} \end{bmatrix}$$

$$\mathbf{\Delta}_i \triangleq \begin{bmatrix} \mathbf{T}(\boldsymbol{\kappa}^{ir}_{1,1}), & \cdots, & \mathbf{T}(\boldsymbol{\kappa}^{ir}_{1,N_r}) \\ \vdots & \ddots & \vdots \\ \mathbf{T}(\boldsymbol{\kappa}^{ir}_{N_s,1}), & \cdots, & \mathbf{T}(\boldsymbol{\kappa}^{ir}_{N_s,N_r}) \end{bmatrix}, \mathbf{g}^{ij}_m \triangleq \begin{bmatrix} \boldsymbol{\iota}^{ij}_{1,m} \\ \vdots \\ \boldsymbol{\iota}^{ij}_{N_s,m} \end{bmatrix}.$$

We can rewrite (4.48) as

$$\hat{\mathbf{e}}^{ij}_m = \hat{\mathbf{\Psi}}_{ir}\mathbf{h}^{rj}_m + \boldsymbol{\varepsilon}^{ij}_m, \quad m = 1, \cdots, N_s,\ j = 1, 2 \tag{4.49}$$

where $\boldsymbol{\varepsilon}^{ij}_m$ is the equivalent estimation error vector given by

$$\boldsymbol{\varepsilon}^{ij}_m \triangleq \mathbf{\Delta}_i\mathbf{h}^{rj}_m - \mathbf{g}^{ij}_m. \tag{4.50}$$

Using a linear estimator to estimate $\mathbf{h}^{rj}_m$ at node i, we have

$$\hat{\mathbf{h}}^{rj}_m = \mathbf{V}^H_{ijm}\hat{\mathbf{e}}^{ij}_m, \quad m = 1, \cdots, N_s,\ j = 1, 2 \tag{4.51}$$

where $\mathbf{V}_{ijm}$ is the weight matrix of the LMMSE estimator at node i. From (4.49) and (4.51), the sum MSE of the first-hop channel estimation at node i is given by

$$\begin{aligned} \mathrm{MSE}_i &= \sum_{j=1}^{2}\sum_{m=1}^{N_s} \mathrm{tr}\big(\mathrm{E}\big[(\hat{\mathbf{h}}^{rj}_m - \mathbf{h}^{rj}_m)(\hat{\mathbf{h}}^{rj}_m - \mathbf{h}^{rj}_m)^H\big]\big) \\ &= \sum_{j=1}^{2}\sum_{m=1}^{N_s} \mathrm{tr}\big(\mathrm{E}\big[(\mathbf{V}^H_{ijm}\mathbf{\Psi}_{ir} - \mathbf{I}_{N_rQ})\mathbf{R}_{\mathbf{h}^{rj}_m}(\mathbf{V}^H_{ijm}\mathbf{\Psi}_{ir} - \mathbf{I}_{N_rQ})^H \\ &\qquad + \mathbf{V}^H_{ijm}\mathbf{R}_{\boldsymbol{\varepsilon}^{ij}_m}\mathbf{V}_{ijm}\big]\big) \end{aligned} \tag{4.52}$$

where $\mathbf{R}_{\boldsymbol{\varepsilon}^{ij}_m} \triangleq \mathrm{E}\left[\boldsymbol{\varepsilon}^{ij}_m\big(\boldsymbol{\varepsilon}^{ij}_m\big)^H\right]$ is the estimation error covariance matrix and $\mathbf{R}_{\mathbf{h}^{rj}_m} \triangleq \mathrm{E}\left[\mathbf{h}^{rj}_m\big(\mathbf{h}^{rj}_m\big)^H\right]$ is the covariance matrix of $\mathbf{h}^{rj}_m$. From (4.46), we have

$$\mathbf{R}_{\mathbf{h}^{rj}_m} = \mathrm{Bdiag}[\mathbf{D}^{rj}_{1,m}, \cdots, \mathbf{D}^{rj}_{N_r,m}], \quad m = 1, \cdots, N_s,\ j = 1, 2$$

where $\mathbf{D}_{n,m}^{rj} \triangleq \text{diag}[\sigma_{n,m,1}^{rj}, \cdots, \sigma_{n,m,Q}^{rj}]$.

The estimation error covariance matrix $\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}$ is obtained from (4.50) as

$$\begin{aligned}\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}} &= \text{E}\left[\left(\boldsymbol{\Delta}_i\mathbf{h}_m^{rj} - \mathbf{g}_m^{ij}\right)\left(\boldsymbol{\Delta}_i\mathbf{h}_m^{rj} - \mathbf{g}_m^{ij}\right)^H\right] \\ &= \text{E}\left[\boldsymbol{\Delta}_i\mathbf{R}_{\mathbf{h}_m^{rj}}\boldsymbol{\Delta}_i^H\right] + \text{E}\left[\mathbf{g}_m^{ij}\left(\mathbf{g}_m^{ij}\right)^H\right], \quad m = 1, \cdots, N_s,\ j = 1, 2.\end{aligned}$$

Due to the circulant structure of $\boldsymbol{\Delta}_i$ and the fact that $\mathbf{R}_{\mathbf{h}_m^{rj}}$ is a diagonal matrix, we have

$$\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}} = \text{Bdiag}\left[\mathbf{R}_{i,j}^{1,m}, \cdots, \mathbf{R}_{i,j}^{N_s,m}\right] + \mathbf{R}_{\mathbf{g}_m^{ij}} \quad m = 1, \cdots, N_s,\ j = 1, 2$$

where $\mathbf{R}_{\mathbf{g}_m^{ij}} = \text{E}\left[\mathbf{g}_m^{ij}\left(\mathbf{g}_m^{ij}\right)^H\right]$ can be obtained from (4.30) and $\mathbf{R}_{i,j}^{k,m} = \sum_{n=1}^{Nr}\text{diag}(\boldsymbol{\sigma}_{n,m}^{rj} * \mathbf{d}_{\boldsymbol{\kappa}_{k,n}^i})$. Here $\boldsymbol{\sigma}_{n,m}^{rj}$ is defined in the line after (4.19) and $\mathbf{d}_{\boldsymbol{\kappa}_{k,n}^i}$ contains the diagonal elements of $\mathbf{R}_{\boldsymbol{\kappa}_{k,n}^i} \triangleq \text{E}\left[\boldsymbol{\kappa}_{k,n}^{ir}\left(\boldsymbol{\kappa}_{k,n}^{ir}\right)^H\right]$, which can be obtained from the MSE expression (4.30).

The weight matrices $\mathbf{V}_{ijm}, j = 1, 2, m = 1, \cdots, N_s$, that minimize MSE_i in (4.52) are given by

$$\mathbf{V}_{ijm} = \left(\boldsymbol{\Psi}_{ir}\mathbf{R}_{\mathbf{h}_m^{rj}}\boldsymbol{\Psi}_{ir}^H + \mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}\right)^{-1}\boldsymbol{\Psi}_{ir}\mathbf{R}_{\mathbf{h}_m^{rj}}. \tag{4.53}$$

Substituting (4.53) back into (4.52), we obtain the MSE of channel estimation at node i as

$$\text{MSE}_i = \sum_{j=1}^{2}\sum_{m=1}^{N_s}\text{tr}\left(\left[\left(\mathbf{R}_{\mathbf{h}_m^{rj}}\right)^{-1} + \boldsymbol{\Psi}_{ir}^H\left(\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}\right)^{-1}\boldsymbol{\Psi}_{ir}\right]^{-1}\right). \tag{4.54}$$

It can be seen from (4.54) that the MSE of the first-hop channel estimation depends on the covariance matrix of the second-hop channel estimation error $\mathbf{R}_{\boldsymbol{\varepsilon}_m^{ij}}$. When the MSE of the second-hop channel estimation increases, the MSE of the first-hop channel estimation also increases.

4.4 Numerical Examples

In this section, we study the performance of the proposed superimposed channel training algorithm for two-way MIMO relay systems operating in frequency-selective fading environments through numerical simulations. We consider a three-node two-way MIMO relay system where all nodes are equipped with the

same number of antennas, i.e., $N_s = N_r = N$. For simplicity, we assume that all channel taps have unit variances. We use the shortest length of training sequence possible with $L = (5Q-2)N$. For all scenarios, the NMSE of channel estimation at nodes 1 and 2 are computed.

For the first three simulation examples, we assume that all nodes have the same transmission power $p_i = p$, $i = 1, 2, r$. In the first example, we investigate the performance of the superimposed channel training algorithm for different α. Fig. 4.3 shows the NMSE of the proposed algorithm versus p with different α when $N = 2$ and $Q = 4$. The optimal α curve is obtained by applying the GSS technique to the proposed superimposed channel training algorithm to obtain the optimal α for different p. It can be observed from Fig. 4.3 that the optimal α curve consistently has the lowest MSE level for all p. This proves that the GSS technique is able to obtain the optimal α at different p efficiently.

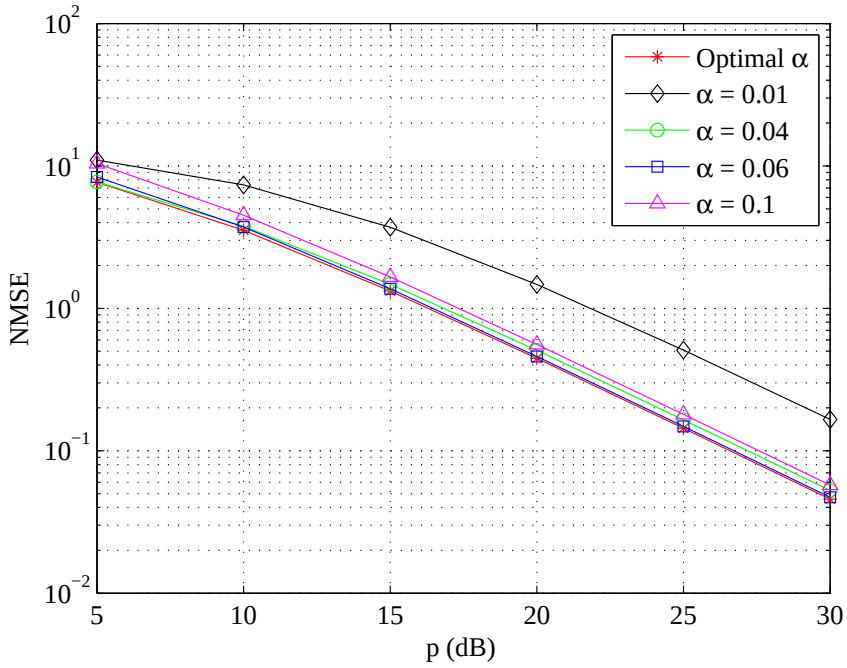

Figure 4.3: Example 4.1: NMSE versus p for different α with $N = 2$ and $Q = 4$.

Interestingly, we notice from Fig. 4.3 that the optimal α varies with respect to p, indicating that using a constant α is strictly suboptimal. Although the NMSE with $\alpha = 0.06$ is close to the NMSE using the optimal α for p between 10dB and 30dB, $\alpha = 0.06$ yields a higher NMSE than $\alpha = 0.04$ at $p = 5$dB. Moreover, for other simulation examples (e.g. different N and Q), the NMSE with $\alpha = 0.06$

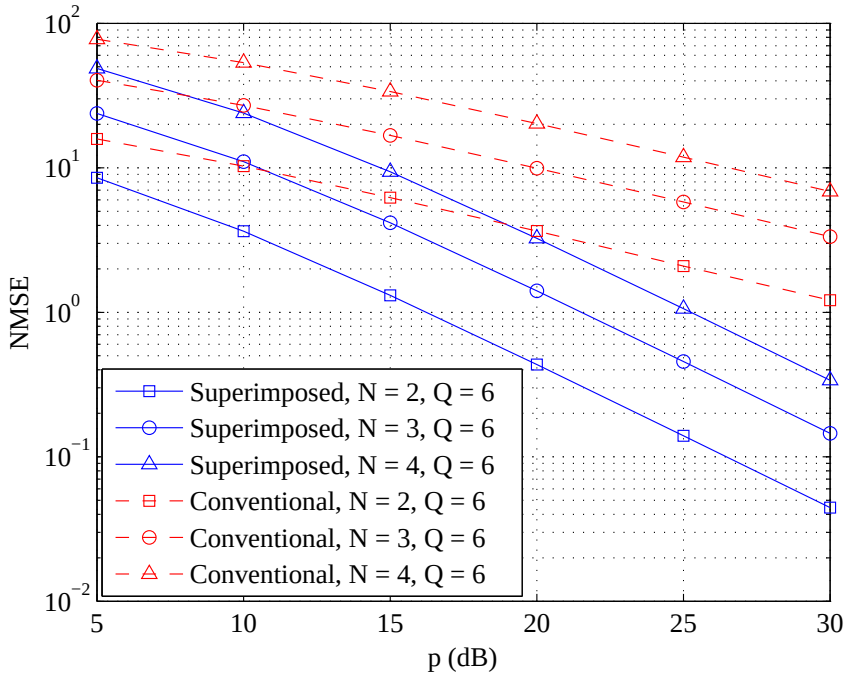

Figure 4.4: Example 4.2: NMSE versus p for different N with $Q = 6$.

might not be close to the NMSE using the optimal α. In practical systems, a table containing the value of the optimal α at different p, N, and Q can be constructed for reference.

In the second example, we study the performance of the proposed superimposed channel training algorithm when the optimal α is used under different simulation parameters. We compare the proposed algorithm with the conventional two-stage MMSE channel estimation algorithm, where the second-hop channel matrices are estimated at the first stage by using the training sequence sent from the relay node, and the first-hop channel matrices are estimated at the second stage by exploiting the training signals sent from the source nodes [86]. Fig. 4.4 demonstrates the NMSE performance of both methods versus p for different N and $Q = 6$. As expected, when the number of antennas increases, the NMSE of channel estimation at both sides also increases as there are more unknowns to be estimated. It can also be seen from Fig. 4.4 that the performance of the proposed algorithm is always better than the conventional two-stage channel estimation method, especially at high power levels.

Fig. 4.5 demonstrates the NMSE performance of the algorithm proposed in Section 4.3.3 which retrieves the individual CSI $\{\mathbf{h}_{k,n}^{ir}\}$ and $\{\mathbf{h}_{n,m}^{ri}\}$. It can be

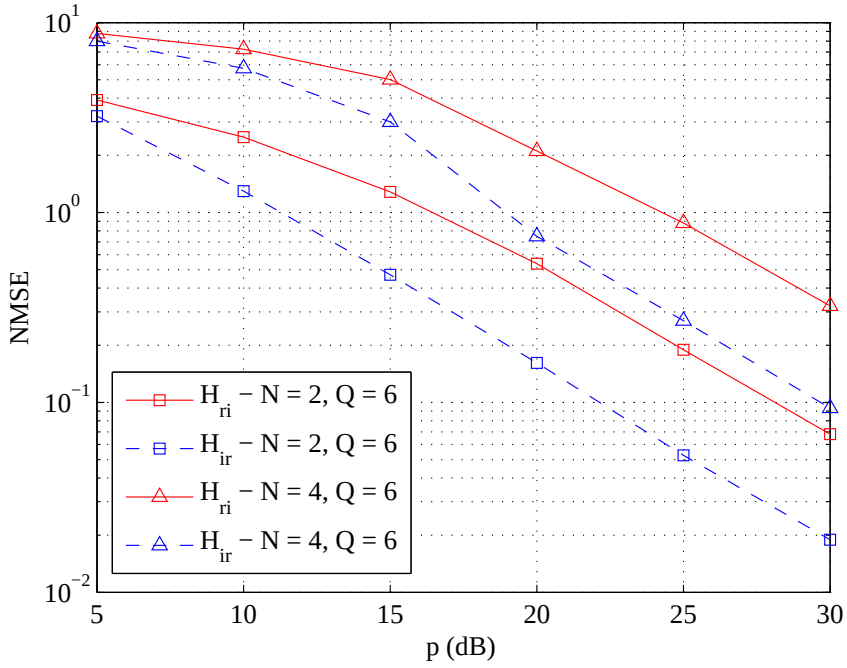

Figure 4.5: Example 4.2: Individual channel NMSE versus p for different N with $Q = 6$.

observed that the NMSE performance for the estimation of $\{\mathbf{h}_{k,n}^{ir}\}$ is always better than that for the estimation of $\{\mathbf{h}_{n,m}^{ri}\}$, as the estimation of $\{\mathbf{h}_{n,m}^{ri}\}$ depends on the estimation of $\{\mathbf{h}_{k,n}^{ir}\}$.

In the third example, the effect of the number of multipath Q on the performance of the proposed superimposed channel training algorithm is investigated. The results are shown in Fig. 4.6 for the case of $N = 2$. It can be seen that the NMSE performance of channel estimation improves when Q increases, as all channel taps are set to have unit variance. It can also be seen from Fig. 4.6 that such improvement diminishes when Q becomes larger.

The fourth simulation example studies the scenario where the power constraints at the source nodes and relay node are different. Fig. 4.7 shows the NMSE of the proposed algorithm versus $p_1 = p_2 = p_s$ for different fixed p_r when $N = 2$ and $Q = 6$. It can be seen that as expected, the proposed algorithm has a better NMSE performance when the power at the source/relay node is increased.

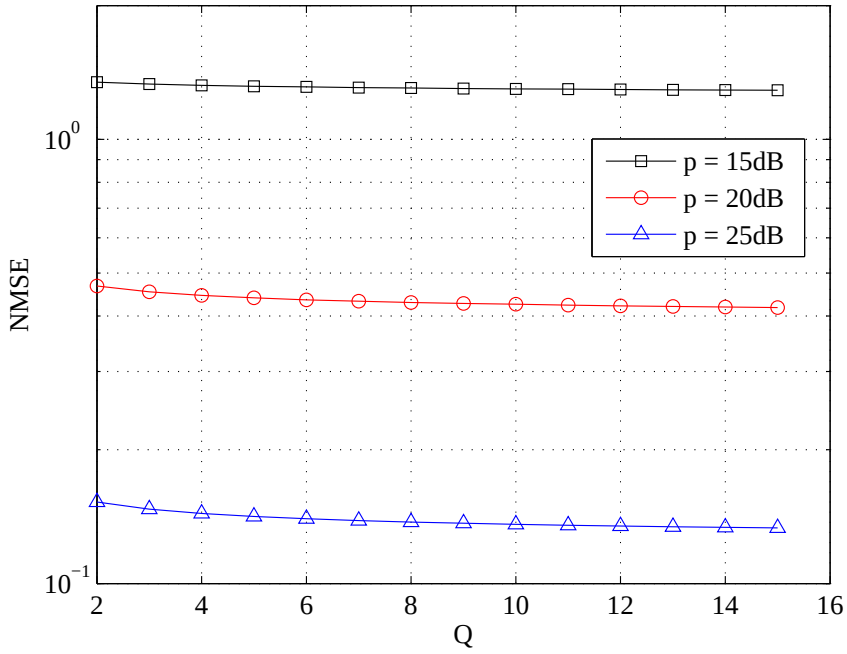

Figure 4.6: Example 4.3: NMSE versus Q for different p and $N = 2$.

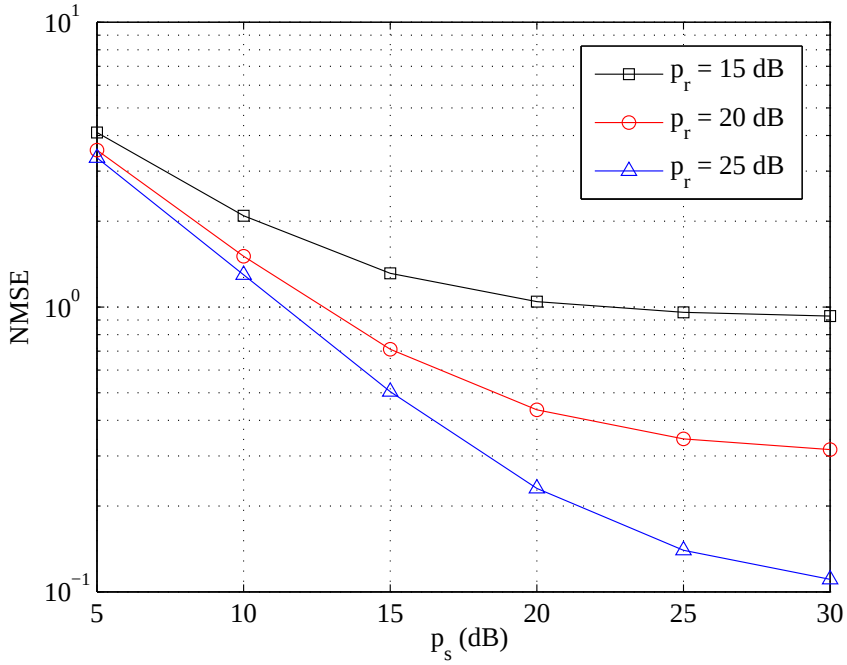

Figure 4.7: Example 4.4: NMSE versus p_s for different p_r with $N = 2$ and $Q = 6$.

4.5 Conclusions

We have applied the method of superimposed channel training to two-way MIMO relay communication systems in frequency-selective fading environments. The proposed algorithm can efficiently estimate the individual CSI for two-way MIMO relay systems with frequency-selective fading channels. We also derived the optimal structure of the training sequences that minimize the MSE of the channel estimation and optimize the power allocation between the source and relay training sequences.

4.A Proof of Theorem 4.1

The MSE in (4.24) can be rewritten as

$$\begin{aligned}\text{MSE} \;=\; \sum_{i=1}^{2}\sum_{k=1}^{N_s} \text{tr}\Bigg(\Bigg[\Bigg(\begin{array}{ccc}\mathbf{C}_{i1}^{k} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{i2}^{k} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{ir}^{k}\end{array}\Bigg)^{-1} \\ +\eta_{i,k}\begin{pmatrix}\sqrt{\alpha}\boldsymbol{\Phi}^{H}(\mathbf{s})\\ \sqrt{\alpha}\boldsymbol{\Phi}^{H}(\mathbf{t})\\ \boldsymbol{\Phi}^{H}(\mathbf{r})\end{pmatrix}\Big(\sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{s}),\ \sqrt{\alpha}\boldsymbol{\Phi}(\mathbf{t}),\ \boldsymbol{\Phi}(\mathbf{r})\Big)\Bigg]^{-1}\Bigg)\end{aligned} \tag{4.55}$$

where

$$\eta_{i,k} \triangleq \left(\alpha\sum_{n=1}^{N_r}\sum_{j=1}^{Q}\sigma_{k,n,j}^{ir}+1\right)^{-1}, \quad i=1,2, \quad k=1,\cdots,N_s. \tag{4.56}$$

It can be seen that (4.55) is minimized only if all off-diagonal matrices of the second term are zero, i.e.,

$$\boldsymbol{\Phi}^{H}(\mathbf{s})\boldsymbol{\Phi}(\mathbf{t})=\mathbf{0} \quad \boldsymbol{\Phi}^{H}(\mathbf{s})\boldsymbol{\Phi}(\mathbf{r})=\mathbf{0} \quad \boldsymbol{\Phi}^{H}(\mathbf{r})\boldsymbol{\Phi}(\mathbf{t})=\mathbf{0}. \tag{4.57}$$

Based on (4.7)-(4.9) and (4.57), we have that for $m, n = 1, \cdots, N_s$, $p = 1, \cdots, N_r$

$$\begin{aligned} &\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \mathbf{0}, \quad \mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0}, \\ &\mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{0}. \end{aligned} \tag{4.58}$$

Using (4.57), MSE in (4.55) can be written as

$$\begin{aligned} \text{MSE} \;=\; & \sum_{i=1}^{2}\sum_{k=1}^{N_s} \text{tr}\Big(\big[\big(\mathbf{C}_{i1}^k\big)^{-1} + \alpha\eta_{i,k}\mathbf{\Phi}^H(\mathbf{s})\mathbf{\Phi}(\mathbf{s})\big]^{-1} \\ & + \big[\big(\mathbf{C}_{i2}^k\big)^{-1} + \alpha\eta_{i,k}\mathbf{\Phi}^H(\mathbf{t})\mathbf{\Phi}(\mathbf{t})\big]^{-1} + \big[\big(\mathbf{C}_{ir}^k\big)^{-1} + \eta_{i,k}\mathbf{\Phi}^H(\mathbf{r})\mathbf{\Phi}(\mathbf{r})\big]^{-1}\Big). \end{aligned} \tag{4.59}$$

Since from (4.18) and (4.19), $\mathbf{C}_{i1}^k$, $\mathbf{C}_{i2}^k$, and $\mathbf{C}_{ir}^k$ are all diagonal, to minimize (4.59), $\mathbf{\Phi}^H(\mathbf{s})\mathbf{\Phi}(\mathbf{s})$, $\mathbf{\Phi}^H(\mathbf{t})\mathbf{\Phi}(\mathbf{t})$, and $\mathbf{\Phi}^H(\mathbf{r})\mathbf{\Phi}(\mathbf{r})$ must be diagonal, and together with (4.7)-(4.9), we have

$$\begin{aligned} &\mathbf{C}_{2Q-1}^H(\mathbf{s}_m)\mathbf{C}_{2Q-1}(\mathbf{s}_m) = \mathbf{D}_{s,m}, \quad \mathbf{C}_{2Q-1}^H(\mathbf{t}_n)\mathbf{C}_{2Q-1}(\mathbf{t}_n) = \mathbf{D}_{t,n}, \\ &\mathbf{C}_Q^H(\mathbf{r}_p)\mathbf{C}_Q(\mathbf{r}_p) = \mathbf{D}_{r,p} \end{aligned} \tag{4.60}$$

where $\mathbf{D}_{s,m}$ and $\mathbf{D}_{t,n}$ are $(2Q-1)\times(2Q-1)$ diagonal matrices, while $\mathbf{D}_{r,p}$ is a $Q \times Q$ diagonal matrix.

It is worth noting that (4.58) and (4.60) do not change the value of $\mathbf{s}_m^H\mathbf{s}_m$, $\mathbf{t}_m^H\mathbf{t}_m$, and $\mathbf{r}_n^H\mathbf{r}_n$ in the constraints (4.25)-(4.27). Moreover, it can be deduced that $\text{tr}\big(\mathbf{C}_Q(\mathbf{s}_m)\mathbf{D}_{n,m}^{r1}\mathbf{C}_Q^H(\mathbf{s}_m)\big)$ in the constraint (4.27) is minimized if $\mathbf{C}_Q^H(\mathbf{s}_m)\mathbf{C}_Q(\mathbf{s}_m)$ is diagonal and its diagonal elements are in the inverse order to that of $\mathbf{D}_{n,m}^{r1}$ [101]. Similarly, the term of $\text{tr}\big(\mathbf{C}_Q(\mathbf{t}_m)\mathbf{D}_{n,m}^{r2}\mathbf{C}_Q^H(\mathbf{t}_m)\big)$ in (4.27) is minimized if $\mathbf{C}_Q^H(\mathbf{t}_m)\mathbf{C}_Q(\mathbf{t}_m)$ is diagonal and its diagonal elements are in the inverse order to that of $\mathbf{D}_{n,m}^{r2}$. Obviously, these two requirements are satisfied by (4.60).

Considering (4.58), (4.60), and the circulant structure of $\mathbf{C}_{2Q-1}(\mathbf{s}_m)$, $\mathbf{C}_{2Q-1}(\mathbf{t}_n)$, and $\mathbf{C}_Q(\mathbf{r}_p)$, we have

$$\mathbf{D}_{s,m} = \beta_m\mathbf{I}_{2Q-1}, \quad \mathbf{D}_{t,n} = \gamma_n\mathbf{I}_{2Q-1}, \quad \mathbf{D}_{r,p} = \delta_p\mathbf{I}_Q$$

where $\mathbf{s}_m^H\mathbf{s}_m = \beta_m$, $\mathbf{t}_n^H\mathbf{t}_n = \gamma_n$, and $\mathbf{r}_p^H\mathbf{r}_p = \delta_p$. □

4.B Proof of Theorem 4.2

By introducing $\xi_{1,m} \triangleq \alpha\beta_m$, $\xi_{2,m} \triangleq \alpha\gamma_m$, $m = 1, \cdots, N_s$, the problem (4.31)-(4.35) can be equivalently rewritten as

$$\min_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\boldsymbol{\delta},\alpha} \quad \sum_{m=1}^{N_s}\sum_{k=1}^{N_s}\sum_{q=1}^{2Q-1}\sum_{i=1}^{2}\left(\frac{1}{c_{k,m,q}^{i1}+\xi_{1,m}\eta_{i,k}}+\frac{1}{c_{k,m,q}^{i2}+\xi_{2,m}\eta_{i,k}}\right) + \sum_{n=1}^{N_r}\sum_{k=1}^{N_s}\sum_{p=1}^{Q}\sum_{i=1}^{2}\frac{1}{c_{k,n,p}^{ir}+\delta_n\eta_{i,k}} \tag{4.61}$$

$$\text{s.t.} \quad \mathbf{1}^T\boldsymbol{\xi}_i \le \alpha p_i, \qquad i = 1, 2 \tag{4.62}$$

$$\boldsymbol{\kappa}_1^T\boldsymbol{\xi}_1 + \boldsymbol{\kappa}_2^T\boldsymbol{\xi}_2 + \mathbf{1}^T\boldsymbol{\delta} \le p_r - \alpha L N_r \tag{4.63}$$

$$\alpha > 0, \;\; \xi_{i,m} \ge 0, i = 1, 2, \; m = 1, \cdots, N_s, \;\; \delta_n \ge 0, n = 1, \cdots, N_r \tag{4.64}$$

where $\boldsymbol{\kappa}_i \triangleq [\kappa_{i,1}, \cdots, \kappa_{i,N_s}]^T$, $\boldsymbol{\xi}_i \triangleq [\xi_{i,1}, \cdots, \xi_{i,N_s}]^T$, $i = 1, 2$, and $\mathbf{1}$ is a column vector of all ones with a commensurate dimension.

Let us first ignore the effect of α on all $\eta_{i,k}$ by treating them as known variables. Then the problem (4.61)-(4.64) becomes a convex optimization problem, as (4.61) is a convex function of $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, $\boldsymbol{\delta}$, and (4.62)-(4.64) are linear inequality constraints. When α has a sufficiently small value, the value of (4.61) is strongly governed by the constraints in (4.62), since the constraint (4.63) is inactive compared with the constraints in (4.62) for small value of α. Once α increases from a small value, the feasible region specified by (4.62) expands, and thus, the value of (4.61) decreases.

On the other hand, when α is large (close to $p_r/(LN_r)$), the value of (4.61) is strongly governed by the constraint (4.63), as the constraints in (4.62) are inactive compared with that of (4.63) when α is large. Once α decreases from a large value, the feasible region specified by (4.63) expands, leading to the decreasing of (4.61).

Now we consider the effect of α on $\eta_{i,k}$. It can be seen from (4.56) that $\eta_{i,k}$ monotonically decreases with increasing α, and (4.61) increases when $\eta_{i,k}$ decreases. From the analysis above, it can be deduced that when α increases from a significantly small positive number, the objective function (4.61) starts to decrease since the potential decrease of (4.61) due to the expanded feasible region of (4.62) dominates the potential increase of (4.61) caused by the decreasing $\eta_{i,k}$. The value of (4.61) keeps decreasing till a 'turning point' where the decreasing of $\eta_{i,k}$ starts to dominate the effect of relaxed feasible region in (4.62). After such

turning point, the value of (4.61) is monotonically increasing with an increasing α. Therefore, the objective function (4.31) subjecting to (4.32)-(4.35) is a unimodal function with respect to α. □

Chapter 5

Blind Channel Estimation for MIMO Relay Systems

In this chapter, we propose a blind channel estimation and signal retrieving algorithm for two-hop MIMO relay systems. We first introduce the background knowledge on blind channel estimation for one-way MIMO relay systems in Section 5.1. The system model of a one-way MIMO relay system is presented in Section 5.2. We propose a new blind channel estimation algorithm which integrates two BSS methods to estimate the individual CSI of the source-relay and relay-destination links. In particular, a first-order Z-domain precoding technique is developed in Section 5.3 for the blind estimation of the relay-destination channel matrix, where the signals received at the relay node are pre-processed by a set of precoders before being transmitted to the destination node. With the estimated signals at the relay node, we propose an algorithm based on the constant modulus and signal mutual information properties in Section 5.4 to estimate the source-relay channel matrix. Compared with training-based MIMO relay channel estimation approaches, the proposed algorithm has a better bandwidth efficiency as no bandwidth is wasted for sending the training sequences. Numerical examples are shown in Section 5.5 to demonstrate the performance of the proposed algorithm. Conclusions are drawn in Section 5.6. We show the proof of Theorem 5.1 and Corollary 5.1 in Section 5.A and 5.B, respectively.

5.1 Introduction

In general, there are two types of channel estimation techniques, which are training-based channel estimation and blind channel estimation. Several training-based channel estimation algorithms have been proposed in Chapter 2, 3, and 4. A tensor-based channel estimation algorithm is developed in [102] for a two-way MIMO relay system. Since the algorithm in [102] exploits the channel reciprocity in a two-way relay system, its application in one-way MIMO relay systems is not straightforward.

The main drawback of the training-based channel estimation algorithms is the high cost involved in sending the training sequences, considering the limited bandwidth available for wireless communication. Moreover, in some applications such as asynchronous wireless network and message interception, training-based algorithms are unrealistic and not suitable for implementation [71], [72]. In these applications, blind channel estimation techniques, which do not require training sequences, become important. Recursive least squares (RLS) and least mean squares (LMS) subspace-based adaptive algorithms were proposed in [73] for blind channel estimation in code-division multiple access (CDMA) systems. A subspace-based blind channel estimation algorithm with reduced time averaging was proposed in [74] for MIMO-OFDM systems. However, the algorithms in [73] and [74] were developed for point-to-point (single-hop) communication systems, and the extension to MIMO relay systems is not straightforward. A blind channel estimation based on the deterministic maximum likelihood (DML) approach was developed in [75] for two-way relay networks with constant-modulus signaling. However, this algorithm only estimates the cascaded source-relay-destination channel in a single-input single-output (SISO) relay system, and does not provide the estimation of the individual second-hop channel in MIMO relay systems which is important for the optimal receiver design at the destination node.

In this chapter, we develop a blind channel estimation algorithm for two-hop MIMO relay communication systems by exploiting the link between BSS and channel estimation. BSS techniques are able to separate a mixture of signals into individual source signals, without the knowledge (or little knowledge) of the source signals or the channel between the source and receiver. The proposed

algorithm integrates two BSS methods to estimate the instantaneous CSI for the individual source-relay and relay-destination links. We would like to note that channel matrices of the first-hop and second-hop are estimated at the destination node. The advantage of directly estimating both channel matrices at the destination node is to avoid sending the CSI from the relay node to the destination node [68], [82]. As the blind channel estimation algorithm we proposed uses the communication data for channel estimation, unlike [86], there is no need for sending training signal from the relay node to the destination node. Therefore, the proposed algorithm does not require more signalling bits.

In particular, we first develop a first-order Z-domain precoding technique for the blind estimation of the relay-destination channel matrix using signals received at the destination node. In this algorithm, the signals received at the relay node are filtered by properly designed precoders before being transmitted to the destination node. By utilizing the Z-domain properties of the precoded signals, an estimation criterion is derived to recover the relay-destination channel matrix and signals received at the relay node. Note that in this algorithm, the order of the precoders is fixed to one, while a second-order Z-domain precoding algorithm was developed in [76] for blind separation of spatially correlated signals. Obviously, the computational complexity of the first-order precoder is smaller than that of the second-order precoder.

With the estimated received signals at the relay node, we then develop a blind channel estimation algorithm based on the constant modulus and signal mutual information (MI) properties to estimate the source-relay channel matrix. The constant modulus property of many modulated communication signals such as phase-shift keying (PSK) and quadrature amplitude modulation (QAM) is exploited in this blind estimation algorithm. However, using the constant modulus property of signals alone does not guarantee the complete separation of the source signals and the channel matrix, as the constant modulus algorithm might capture the same signal even though there are multiple signal streams. To overcome this problem, we minimize a cost function which includes the MI of the estimated signals in addition to the constant modulus property, to ensure that all estimated signals are distinct. This algorithm does not have the problem of estimation error

propagation as in [103] and [104]. A similar method was adopted in [72] for the extraction of unknown source signals, essentially in single-hop (point-to-point) MIMO wireless networks.

Comparing the proposed blind channel estimation algorithm with the training-based channel techniques, the former one has a better bandwidth efficiency as all the bandwidth is used for the transmission of the communication signals. We would like to note that the proposed algorithm can be applied in two-hop MIMO relay systems with multiple distributed source nodes and multiple distributed relay nodes.

5.2 System Model

Let us consider a three-node two-hop MIMO communication system where the source node transmits information to the destination node through a relay node as shown in Fig. 5.1. The source, relay, and destination nodes are equipped with N_s, N_r, and N_d antennas, respectively. We would like to mention that the algorithm developed in this chapter can be easily extended to MIMO relay systems with multiple sources and relay nodes. In this chapter, we assume that the direct link between the source node and the destination node is sufficiently weak and thus can be ignored. This scenario occurs when the direct link is blocked by obstacles, such as tall buildings or mountains.

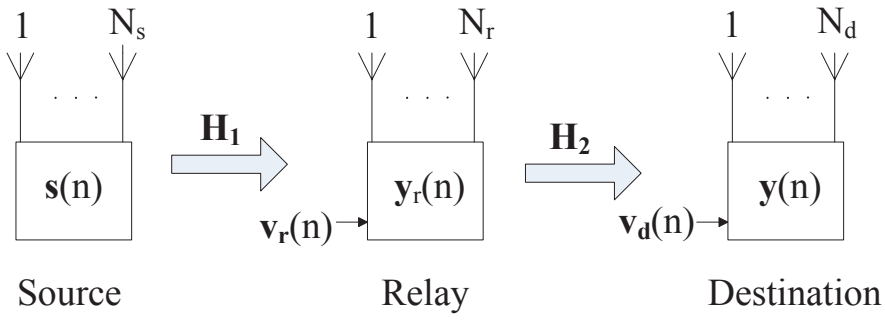

Figure 5.1: Block diagram of a general two-hop MIMO relay communication system.

The communication process is completed in two time slots. In the first time slot, the source signal vector $\mathbf{s}(n) = [s_1(n), s_2(n), \cdots, s_{N_s}(n)]^T$ is transmitted

from the source node. The signal vector received at the relay node can be expressed as

$$\mathbf{y}_r(n) = \mathbf{H}_1\mathbf{s}(n) + \mathbf{v}_r(n) \tag{5.1}$$

where $\mathbf{y}_r(n)$ is the $N_r \times 1$ received signal vector, $\mathbf{H}_1$ is the $N_r \times N_s$ MIMO channel matrix between the source node and the relay node, and $\mathbf{v}_r(n)$ is the $N_r \times 1$ noise vector at the relay node.

In the second time slot, each received signal stream in $\mathbf{y}_r(n)$ is preprocessed separately by a first-order precoder $p_i(z)$ as

$$p_i(z) = 1 - r_i z^{-1}, \quad i = 1, \cdots, N_r \tag{5.2}$$

where r_i is the zero of the precoder $p_i(z)$. Note that all zeros are distinct and satisfy $0 < |r_i| < 1$, for $i = 1, \cdots, N_r$, and are known at the destination node. From (5.2), the ith precoded signal at the relay node can be written as

$$\begin{aligned} x_i(n) &= p_i(z) y_{r,i}(n) \\ &= y_{r,i}(n) - r_i y_{r,i}(n-1), \quad i = 1, \cdots, N_r \end{aligned} \tag{5.3}$$

where $y_{r,i}(n)$ is the ith element of $\mathbf{y}_r(n)$. It is worth noting that the precoding operation (5.3) can be readily implemented at physically distributed relay nodes, as there is no need for cooperation among different signal streams. The first-order precoding operation in (5.3) serves for the blind estimation of the relay-destination channel matrix, where the estimation criterion will be derived by exploiting the Z-domain properties of the precoders as shown in Section 5.3.

The precoded signal vector $\mathbf{x}(n) = [x_1(n), x_2(n), \cdots, x_{N_r}(n)]^T$ is transmitted to the destination node, and the received signal vector at the destination node can be expressed as

$$\mathbf{y}(n) = \mathbf{H}_2\mathbf{x}(n) + \mathbf{v}_d(n) \tag{5.4}$$

where $\mathbf{H}_2$ is the $N_d \times N_r$ channel matrix between the relay node and the destination node and $\mathbf{v}_d(n) = [v_{d,1}(n), v_{d,2}(n), \cdots, v_{d,N_d}(n)]^T$ is the noise vector at the destination node. We assume that:

1. All noises are i.i.d. AWGN.

2. The source signals in $\mathbf{s}(n)$ are temporally white and have constant modulus.

3. The noises are independent of the source signals.

4. The number of antennas at the receiving sides is equal or greater than that of the transmitting sides, i.e., $N_d \geq N_r \geq N_s$.

The model in (5.4) has a similar structure to the classical BSS problem. In BSS techniques, signal separation is usually achieved by exploiting the statistical properties of the source signals, either based on the higher-order statistics (HOS) or second-order statistics (SOS). Independent component analysis (ICA) is one example of the HOS-based BSS methods, and is generally applied for non-Gaussian source signals. One of the drawbacks of the HOS-based methods is the large number of data samples required for a satisfactory result. On the contrary, the number of data samples required by the SOS-based BSS methods is generally much smaller than the HOS-based BSS techniques. However, the SOS-based BSS methods usually require the source signals to be mutually uncorrelated. This limits the application of the SOS-based BSS methods in MIMO relay communication systems as the signals received at the relay node ($\mathbf{y}_r$ in (5.1)) are mutually correlated.

A second-order precoding-based BSS algorithm has been developed in [76] to separate mutually correlated sources. However, this algorithm might not be applicable to MIMO relay systems. This is because the algorithm in [76] does not allow any source signal to be linear combination of the other source signals, while in a MIMO relay system, the received signal at the relay node (5.1) is a linear combination of the source signals. Thus, when the noise at the relay node is sufficiently small, the signal at the relay node does not satisfy the requirement of the second-order precoding technique. This motivates us to develop the first-order precoding technique for blind channel estimation in MIMO relay systems as presented in the next section.

5.3 First-Order Z-Domain Precoding Based Channel Estimation

In this section, we develop a first-order Z-domain precoding algorithm for the blind estimation of the relay-destination channel matrix $\mathbf{H}_2$. The main idea of this approach is to preprocess the received signals at the relay node with the first-order Z-domain precoders before retransmitting them to the destination node. Then, by utilizing the Z-domain properties of the precoders, this blind channel estimation aims to find a separation matrix $\mathbf{B}_1$ to separate $\mathbf{x}(n)$ and $\mathbf{H}_2$ in (5.4) with only the observable output at the destination node $\mathbf{y}(n)$. Compared with [76], the first-order precoding technique requires less transmission time at the relay node and simplifies the implementation of the precoders at the relay node in practical MIMO relay systems.

Let $\mathbf{B}_1 = [\mathbf{b}_{1,1}, \mathbf{b}_{1,2}, \cdots, \mathbf{b}_{1,N_r}]$ be an $N_d \times N_r$ matrix, the desired outcome of the blind channel estimation algorithm is given by

$$\hat{\mathbf{x}}(n) = \mathbf{B}_1^H \mathbf{y}(n) = \mathbf{\Lambda}\mathbf{x}(n) + \mathbf{B}_1^H \mathbf{v}_d(n) \tag{5.5}$$

where $\hat{\mathbf{x}}(n)$ is an estimation of the precoded signal vector and $\mathbf{\Lambda} \triangleq \mathbf{B}_1^H \mathbf{H}_2$ is a diagonal matrix of scaling ambiguity inherited in the blind estimation algorithm. Note that the permutation ambiguity usually associated with BSS methods does not exist in (5.5) due to the filtering operation (5.3) at the relay node before retransmitting the signals, as each signal stream in $\mathbf{y}_r(n)$ is preprocessed by a distinct precoder. The scaling ambiguity can be resolved, for example, through normalization as in [83] and [105]. Once the separation matrix $\mathbf{B}_1$ is obtained, $\mathbf{H}_2$ and $\mathbf{y}_r(n)$ can be efficiently estimated as shown later on. In the following subsection, we will first propose an estimation criterion by exploiting the Z-domain properties of the precoders and find the separation matrix $\mathbf{B}_1$ based on this criterion.

5.3.1 Estimation Criterion

Let us define the autocorrelation matrix of $\mathbf{y}_r(n)$ at time lag k as

$$\begin{aligned}\mathbf{C}_{\mathbf{y}_r\mathbf{y}_r}(k) &= \mathrm{E}\big[\mathbf{y}_r(n)\mathbf{y}_r(n-k)^H\big] \\ &= \mathbf{H}_1\mathbf{C}_{\mathbf{ss}}(k)\mathbf{H}_1^H + \mathbf{C}_{\mathbf{v_r v_r}}(k) \end{aligned} \tag{5.6}$$

where $\mathbf{C}_{\mathbf{ss}}(k) = \mathrm{E}\big[\mathbf{s}(n)\mathbf{s}(n-k)^H\big]$ and $\mathbf{C}_{\mathbf{v_r v_r}}(k) = \mathrm{E}\big[\mathbf{v}_r(n)\mathbf{v}_r(n-k)^H\big]$ are the autocorrelation matrices of $\mathbf{s}(n)$ and $\mathbf{v}_r(n)$, respectively. Note that $\mathbf{C}_{\mathbf{v_r v_r}}(k) = \mathbf{0}$ for $k \neq 0$ as the noises are temporally independent. Based on (5.6), the power spectral matrix of $\mathbf{y}_r(n)$ is defined as

$$\mathbf{Q}_{\mathbf{y}_r\mathbf{y}_r}(z) = \sum_{k=-\infty}^{\infty} \mathbf{C}_{\mathbf{y}_r\mathbf{y}_r}(k) z^{-k}. \tag{5.7}$$

As the noise covariance matrix at the relay node $\mathbf{C}_{\mathbf{v_r v_r}}(0)$ is of full rank, the following proposition is established.

PROPOSITION 5.1: The power spectral matrix $\mathbf{Q}_{\mathbf{y}_r\mathbf{y}_r}(z)$ is of full rank at $z = r_i$ for $i = 1, \cdots, N_r$.

Let us denote the autocorrelation matrices of $\mathbf{y}(n)$ and $\mathbf{v}_d(n)$ as $\mathbf{C}_{\mathbf{yy}}(k)$ and $\mathbf{C}_{\mathbf{v_d v_d}}(k)$, respectively. It follows from (5.4) that

$$\begin{aligned}\mathbf{C}_{\mathbf{yy}}(k) &= \mathrm{E}\big[\mathbf{y}(n)\mathbf{y}(n-k)^H\big] \\ &= \mathbf{H}_2\mathbf{C}_{\mathbf{xx}}(k)\mathbf{H}_2^H + \mathbf{C}_{\mathbf{v_d v_d}}(k) \end{aligned} \tag{5.8}$$

where $\mathbf{C}_{\mathbf{xx}}(k) = \mathrm{E}\big[\mathbf{x}(n)\mathbf{x}(n-k)^H\big]$ is the autocorrelation matrix of $\mathbf{x}(n)$ and $\mathbf{C}_{\mathbf{v_d v_d}}(k) = \mathrm{E}\big[\mathbf{v}_d(n)\mathbf{v}_d(n-k)^H\big] = \mathbf{0}$ for $k \neq 0$ as the noises are temporally independent. Similarly, the power spectral matrix of $\mathbf{y}(n)$ can be derived based on (5.3), (5.7), and (5.8) as

$$\begin{aligned}\mathbf{Q}_{\mathbf{yy}}(z) &= \sum_{k=-\infty}^{\infty} \mathbf{C}_{\mathbf{yy}}(k) z^{-k} \\ &= \mathbf{H}_2\mathbf{Q}_{\mathbf{xx}}(z)\mathbf{H}_2^H + \mathbf{Q}_{\mathbf{v_d v_d}}(z) \\ &= \mathbf{H}_2\mathbf{P}(z)\mathbf{Q}_{\mathbf{y}_r\mathbf{y}_r}(z)\mathbf{P}(z^{-1})^H\mathbf{H}_2^H + \mathbf{Q}_{\mathbf{v_d v_d}}(z) \end{aligned} \tag{5.9}$$

where $\mathbf{Q}_{\mathbf{xx}}(z) = \sum_{k=-\infty}^{\infty} \mathbf{C}_{\mathbf{xx}}(k) z^{-k}$ and $\mathbf{Q}_{\mathbf{v_d v_d}}(z) = \sum_{k=-\infty}^{\infty} \mathbf{C}_{\mathbf{v_d v_d}}(k) z^{-k}$ are the power spectral matrices of $\mathbf{x}(n)$ and $\mathbf{v}_d(n)$, respectively, and $\mathbf{P}(z) = \mathrm{diag}(p_1(z), p_2(z), \cdots, p_{N_r}(z))$ is a diagonal matrix.

Let us introduce

$$\mathbf{T}_i(z) = \mathbf{P}_i(z)\mathbf{Q}_{\mathbf{y}_r\mathbf{y}_r}(z)\mathbf{P}(z^{-1})^H, \quad i = 1, \cdots, N_r$$

where $\mathbf{P}_i(z)$ is the matrix $\mathbf{P}(z)$ with the ith diagonal entry replaced by zero, i.e.,

$$\mathbf{P}_i(z) = \text{diag}(p_1(z), \cdots, p_{i-1}(z), 0, p_{i+1}(z), \cdots, p_{N_r}(z)). \tag{5.10}$$

It can be shown that for any r_i,

$$\text{rank}(\mathbf{P}_i(r_i)) = N_r - 1$$

while the matrix $\mathbf{P}(r_i^{-1})^H$ is of full rank, since r_i^{-1} is not a zero of any precoder. It can be shown using (5.10) that all elements in the ith row of $\mathbf{T}_i(r_i)$ are zero. Using these results and Proposition 5.1, the following lemma can be established.

LEMMA 5.1: The rank of $\mathbf{T}_i(r_i)$ is $N_r - 1$, for $i = 1, \cdots, N_r$, and all rows of $\mathbf{T}_i(r_i)$ except for the ith row are linearly independent.

Let $\mathbf{H}_{2,i}$ be equal to $\mathbf{H}_2$ with the ith column replaced by a zero vector, i.e.,

$$\mathbf{H}_{2,i} = [\mathbf{h}_{2,1}, \cdots, \mathbf{h}_{2,i-1}, \mathbf{0}, \mathbf{h}_{2,i+1}, \cdots, \mathbf{h}_{2,N_r}]. \tag{5.11}$$

We can rewrite (5.9) as

$$\begin{aligned}
\mathbf{Q}_{\mathbf{yy}}(r_i) &= \mathbf{H}_2\mathbf{P}(r_i)\mathbf{Q}_{\mathbf{y}_r\mathbf{y}_r}(r_i)\mathbf{P}(r_i^{-1})^H\mathbf{H}_2^H + \mathbf{Q}_{\mathbf{v_d v_d}}(r_i) \\
&= \mathbf{H}_{2,i}\mathbf{P}_i(r_i)\mathbf{Q}_{\mathbf{y}_r\mathbf{y}_r}(r_i)\mathbf{P}(r_i^{-1})^H\mathbf{H}_2^H + \mathbf{C}_{\mathbf{v_d v_d}}(0) \\
&= \mathbf{H}_{2,i}\mathbf{T}_i(r_i)\mathbf{H}_2^H + \mathbf{C}_{\mathbf{v_d v_d}}(0).
\end{aligned} \tag{5.12}$$

Assuming that $\mathbf{C}_{\mathbf{v_d v_d}}(0)$ can be estimated, which will be shown later, and removed from (5.12), we have

$$\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{H}_{2,i}\mathbf{T}_i(r_i)\mathbf{H}_2^H. \tag{5.13}$$

The following theorem establishes the estimation criterion for our blind channel estimation algorithm.

THEOREM 5.1: For $i = 1, \cdots, N_r$, $\mathbf{b}_{1,i}$ is an $N_d \times 1$ separation vector ensuring

$$\mathbf{b}_{1,i}^H\mathbf{H}_2 = [0, \cdots, 0, c_i, 0, \cdots, 0], \quad c_i \neq 0 \tag{5.14}$$

if and only if

$$\begin{cases} \mathbf{b}_{1,i}^H\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{0} & (5.15) \\ \mathbf{b}_{1,i}^H\mathbf{C}_{\mathbf{yy}}(1)\mathbf{b}_{1,i} \neq 0. & (5.16) \end{cases}$$

PROOF: See Appendix 5.A. □

Theorem 5.1 holds when the autocorrelation matrix of $\mathbf{y}(n)$ has a time lag of $\tau = 1$, i.e., $\mathbf{C_{yy}}(1)$. Interestingly, it is shown in the following corollary that Theorem 5.1 is not valid for $\mathbf{C_{yy}}(\tau)$ with other time lag values.

COROLLARY 5.1: Theorem 5.1 does not hold for $\mathbf{C_{yy}}(\tau)$, $\tau \neq 1$.

PROOF: See Appendix 5.B. □

It can be seen that the proposed first-order precoding algorithm has different requirements on the selection of parameters compared with the second-order precoding algorithm in [76]. The implementation of the first-order Z-domain precoding based blind channel estimation algorithm is shown in the following subsection.

5.3.2 Algorithm Implementation

The following blind channel estimation procedures are applied to obtain the relay-destination channel matrix $\mathbf{H}_2$.

1. Compute the autocorrelation matrix of $\mathbf{y}(n)$ as

$$\mathbf{C_{yy}}(k) \approx \frac{1}{L}\sum_{n=0}^{L-1}\mathbf{y}(n)\mathbf{y}(n-k)^H \tag{5.17}$$

   where $L \geq N_d$ is the number of samples of the received signal.

2. Compute the power spectral matrix of $\mathbf{y}(n)$ as

$$\mathbf{Q_{yy}}(r_i) \approx \sum_k \mathbf{C_{yy}}(k) r_i^{-k}, \quad i = 1, \cdots, N_r. \tag{5.18}$$

3. Estimate the noise covariance matrix $\mathbf{C_{v_d v_d}}(0)$. It follows from (5.8) that

$$\mathbf{C_{yy}}(0) = \mathbf{H}_2\mathbf{C_{xx}}(0)\mathbf{H}_2^H + \mathbf{C_{v_d v_d}}(0). \tag{5.19}$$

   Since the noises are assumed to be i.i.d. white Gaussian, we have

$$\mathbf{C_{v_d v_d}}(0) = \sigma_{v_d}^2\mathbf{I}_{N_d} \tag{5.20}$$

   where $\sigma_{v_d}^2$ is the noise variance. Let us introduce the EVD of

$$\mathbf{C_{yy}}(0) = \mathbf{U}_Y\mathbf{\Lambda}_Y\mathbf{U}_Y^H \tag{5.21}$$

where $\mathbf{U}_Y$ is the unitary eigenvector matrix and $\mathbf{\Lambda}_Y$ is the diagonal eigenvalue matrix with descending diagonal elements. Obviously, from (5.19) there is

$$\mathbf{\Lambda}_Y = \mathbf{\Lambda}_X + \sigma_{v_d}^2 \mathbf{I}_{N_d} \tag{5.22}$$

where $\mathbf{\Lambda}_X$ is the eigenvalue matrix of $\mathbf{H}_2\mathbf{C}_{\mathbf{xx}}(0)\mathbf{H}_2^H$ with descending diagonal elements.

If $N_d > N_r$, i.e., $\mathbf{H}_2$ is a tall matrix, from (5.22), we have

$$\begin{aligned} \lambda_{y,i} &= \lambda_{x,i} + \sigma_{v_d}^2, \quad i = 1, \cdots, N_r \\ \lambda_{y,i} &= \sigma_{v_d}^2, \quad i = N_r + 1, \cdots, N_d \end{aligned} \tag{5.23}$$

where $\lambda_{y,i}$, $i = 1, \cdots, N_d$, and $\lambda_{x,j}$, $j = 1, \cdots, N_r$, are the diagonal elements of $\mathbf{\Lambda}_Y$ and $\mathbf{\Lambda}_X$, respectively. From (5.23), we can estimate $\sigma_{v_d}^2$ as

$$\sigma_{v_d}^2 = \frac{1}{N_d - N_r} \sum_{i=N_r+1}^{N_d} \lambda_{y,i}. \tag{5.24}$$

If $N_d = N_r$, i.e., $\mathbf{H}_2$ is a square matrix, the noise covariance matrix can be estimated prior to the transmission of data, i.e., when $\mathbf{y}(n) = \mathbf{v}_d(n)$, $n = 1, \cdots, J$, we have

$$\mathbf{C}_{\mathbf{v_d v_d}}(0) \approx \frac{1}{J} \sum_{n=0}^{J-1} \mathbf{y}(n)\mathbf{y}(n)^H.$$

4. Estimate $\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i)$ as

$$\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) \triangleq \mathbf{Q}_{\mathbf{yy}}(r_i) - \mathbf{C}_{\mathbf{v_d v_d}}(0), \quad i = 1, \cdots, N_r. \tag{5.25}$$

5. Obtain separation matrix $\mathbf{B}_1$ as follows. From Lemma 5.1, it can be seen that $\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i)$ has a rank of $N_r - 1$. Since $\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i)$ is an $N_d \times N_d$ matrix, there are $N_d - N_r + 1$ zero singular values. As we assume $N_d \geq N_r$, there exists at least one zero singular value. Let $\mathbf{V}_i$ be an $N_d \times (N_d - N_r + 1)$ matrix whose columns consist of the $N_d - N_r + 1$ left singular vectors corresponding to the zero singular values of $\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i)$, and column vector $\mathbf{u}_i$ be the eigenvector corresponding to any nonzero eigenvalue λ of $\mathbf{V}_i^H\mathbf{C}_{\mathbf{yy}}(1)\mathbf{V}_i$. It can be proven that

$$\mathbf{u}_i^H\mathbf{V}_i^H\bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{0}$$

and

$$\mathbf{u}_i^H \mathbf{V}_i^H \mathbf{C}_{\mathbf{yy}}(1) \mathbf{V}_i \mathbf{u}_i = \lambda \mathbf{u}_i^H \mathbf{u}_i \neq 0.$$

Then, the separation vector $\mathbf{b}_{1,i}$ can be selected as $\mathbf{b}_{1,i}^H = \mathbf{u}_i^H \mathbf{V}_i^H$. The operations in this step are carried out for $i = 1, \cdots, N_r$.

6. The precoded signals can be estimated by

$$\hat{\mathbf{x}}(n) = \mathbf{B}_1^H \mathbf{y}(n), \quad n = 1, \cdots, L. \tag{5.26}$$

7. The relay-destination channel matrix is estimated as

$$\hat{\mathbf{H}}_2 = \mathbf{Y} \hat{\mathbf{X}}^\dagger \tag{5.27}$$

where $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \cdots, \mathbf{y}(L)]$ and $\hat{\mathbf{X}} = [\hat{\mathbf{x}}(1), \hat{\mathbf{x}}(2), \cdots, \hat{\mathbf{x}}(L)]$. Note that since $L \geq N_r$, we have the right inverse of $\hat{\mathbf{X}}$ as

$$\hat{\mathbf{X}}^\dagger = \hat{\mathbf{X}}^T (\hat{\mathbf{X}} \hat{\mathbf{X}}^T)^{-1}. \tag{5.28}$$

5.4 Channel Estimation Based on Signal MI Modified Constant Modulus Algorithm

In this section, we develop a signal MI modified constant modulus algorithm to estimate the first-hop channel matrix $\mathbf{H}_1$. Based on the estimated precoded signals $\hat{x}_i(n)$, $i = 1, \cdots, N_r$, the signals received at the relay node can be estimated by

$$\hat{y}_{r,i}(n) = \hat{x}_i(n) + r_i \hat{y}_{r,i}(n-1), \quad i = 1, \cdots, N_r. \tag{5.29}$$

Let us introduce an $N_r \times N_s$ separation matrix $\mathbf{B}_2$ and let

$$\hat{\mathbf{s}}(n) = \mathbf{B}_2^H \hat{\mathbf{y}}_r(n) = \mathbf{C}\mathbf{s}(n) + \mathbf{B}_2^H \mathbf{v}_r(n) \tag{5.30}$$

where $\hat{\mathbf{s}}(n)$ is the estimated source signal vector and $\mathbf{C} \triangleq \mathbf{B}_2^H \mathbf{H}_1$. This blind channel estimation algorithm aims to obtain the separation matrix $\mathbf{B}_2$ in order to recover the first-hop channel $\mathbf{H}_1$, only from the estimated relay channel output signals $\hat{\mathbf{y}}_r(n)$. Obviously, the estimation of $\mathbf{H}_1$ is affected by the accuracy of the

estimation of $\mathbf{y}_r(n)$. Due to the unaccessible source signals, there are inherent scaling and permutation ambiguities in this algorithm, i.e.,

$$\mathbf{C} = \mathbf{B}_2^H \mathbf{H}_1 = \mathbf{P}\boldsymbol{\Delta}$$

where $\mathbf{P}$ is a permutation matrix and $\boldsymbol{\Delta}$ is a diagonal matrix.

5.4.1 Development of the Algorithm

The general cost function for the constant modulus algorithm is given by

$$\sum_{i=1}^{N_s} \mathrm{E}\big[(|\hat{s}_i(n)|^2 - \gamma)^2\big]$$

where $\hat{s}_i(n)$ is the ith element of $\hat{\mathbf{s}}(n)$ and γ is a priori constant dispersion. As mentioned earlier, the constant modulus algorithm is capable of retrieving one source signal at a time. However, it does not guarantee the extraction of all source signals as the constant modulus algorithm might extract the same signal.

Similar to [72], we propose to exploit the MI property of the estimated signals, along with the constant modulus algorithm, to ensure that the channel matrix and source signals are completely separated. In particular, the following cost function with the addition of the MI term is minimized

$$J(\mathbf{B}_2) = \sum_{i=1}^{N_s} \mathrm{E}\big[(|\hat{s}_i(n)|^2 - \gamma)^2\big] + \beta \left[\sum_{i=1}^{N_s} \log(r_{ii}) - \log|\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}| \right] \tag{5.31}$$

where β is a positive real number that balances the constant modulus term and the MI term, r_{ii} is the diagonal element of $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$, and $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} \triangleq \mathrm{E}\big[\hat{\mathbf{s}}(n)\hat{\mathbf{s}}(n)^H\big]$ is the covariance matrix of $\hat{\mathbf{s}}(n)$. From [72], we have the following proposition.

PROPOSITION 5.2: The MI term is zero when $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$ is a diagonal matrix, i.e., when the elements of $\hat{\mathbf{s}}(n)$ are uncorrelated.

Proposition 5.2 is important to ensure that all source signals are separated from the channel matrix $\mathbf{H}_1$ at the destination node. The cost function (5.31) can be rewritten as

$$\begin{aligned} J(\mathbf{B}_2) \;=\; & \mathrm{E}\left[\sum_{i=1}^{N_s} \left(\mathbf{e}_i^T \mathbf{B}_2^H \hat{\mathbf{y}}_r(n) \hat{\mathbf{y}}_r(n)^H \mathbf{B}_2 \mathbf{e}_i - \gamma \right)^2 \right] \\ & + \beta \left[\sum_{i=1}^{N_s} \log\left(\mathbf{e}_i^T \mathbf{B}_2^H \mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r} \mathbf{B}_2 \mathbf{e}_i \right) - \log\left| \mathbf{B}_2^H \mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r} \mathbf{B}_2 \right| \right] \end{aligned}$$

where $\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r} \triangleq \mathrm{E}\big[\hat{\mathbf{y}}_r(n)\hat{\mathbf{y}}_r(n)^H\big]$ is the covariance matrix of $\hat{\mathbf{y}}_r(n)$ and $\mathbf{e}_i$ is an $N_s \times 1$ column vector whose elements are zero except for the ith element which is one. The gradient of $J(\mathbf{B}_2)$ is given by

$$\begin{aligned}\nabla J(\mathbf{B}_2) &= \frac{\partial J(\mathbf{B}_2)}{\partial \mathbf{B}_2^*} \\ &= 2\sum_{i=1}^{N_s} \mathrm{E}\Big[\big(|\hat{s}_i(n)|^2 - \gamma\big)\hat{\mathbf{y}}_r(n)\mathbf{e}_i^T\big(\hat{\mathbf{y}}_r(n)^H\mathbf{B}_2\mathbf{e}_i\big)\Big] \\ &\quad + \beta\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}\mathbf{B}_2\Big[(\mathrm{diag}(\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}))^{-1} - \mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}^{-1}\Big]. \end{aligned} \tag{5.32}$$

5.4.2 Algorithm Implementation

The procedure of applying the signal MI modified constant modulus algorithm to estimate the source-relay channel matrix $\mathbf{H}_1$ is listed below.

1. Initialize $\mathbf{B}_2^{(0)}$ and $\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(0)}$; Set $i = 1$.

2. Update $\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(i)}$ through

$$\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(i)} = (1-\kappa)\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(i-1)} + \kappa\hat{\mathbf{y}}_r(i)\hat{\mathbf{y}}_r(i)^H \tag{5.33}$$

   where $0 < \kappa < 1$ is a small positive real number.

3. Estimate $\hat{\mathbf{s}}(i) = \Big(\mathbf{B}_2^{(i-1)}\Big)^H\hat{\mathbf{y}}_r(i)$.

4. Calculate $\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}^{(i)} = \Big(\mathbf{B}_2^{(i-1)}\Big)^H\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(i)}\mathbf{B}_2^{(i-1)}$.

5. From step 1)-4), an estimation of (5.32) is obtained by removing the expectation operator E in the equation. Let us denote this estimation as $\hat{\nabla}J(\mathbf{B}_2)$.

6. Update the separation matrix $\mathbf{B}_2$ as

$$\mathbf{B}_2^{(i)} = \mathbf{B}_2^{(i-1)} - \mu\hat{\nabla}J(\mathbf{B}_2)|_{\mathbf{B}_2=\mathbf{B}_2^{(i-1)}}. \tag{5.34}$$

7. Repeat steps 2) – 6) for $i = 2, 3, \cdots, L$ to obtain $\mathbf{B}_2 = \mathbf{B}_2^{(L)}$.

8. The source signals are estimated as

$$\hat{\mathbf{s}}(n) = \mathbf{B}_2^H\hat{\mathbf{y}}_r(n), \quad n = 1, \cdots, L \tag{5.35}$$

9. Estimate the source-relay channel matrix as

$$\hat{\mathbf{H}}_1 = \hat{\mathbf{Y}}_r \hat{\mathbf{S}}^{\dagger} \tag{5.36}$$

where $\hat{\mathbf{Y}}_r = [\hat{\mathbf{y}}_r(1), \hat{\mathbf{y}}_r(2), \cdots, \hat{\mathbf{y}}_r(L)]$ and $\hat{\mathbf{S}} = [\hat{\mathbf{s}}(1), \hat{\mathbf{s}}(2), \cdots, \hat{\mathbf{s}}(L)]$. Note that since $L \geq N_s$, we have the right inverse of $\hat{\mathbf{S}}$ as

$$\hat{\mathbf{S}}^{\dagger} = \hat{\mathbf{S}}^T (\hat{\mathbf{S}} \hat{\mathbf{S}}^T)^{-1}. \tag{5.37}$$

We would like to note that the algorithm proposed in [72] was developed for blind signal separation in one-hop systems, whereas in this chapter we apply this algorithm for channel estimation in two-hop MIMO relay communication systems.

5.5 Numerical Examples

In this section, we study the performance of the proposed blind MIMO relay channel estimation algorithm through numerical simulations. We consider a three-node two-hop MIMO relay system with N_s, N_r, and N_d antennas equipped at the source, relay, and destination node, respectively. For the proposed first-order Z-domain precoding based channel estimation algorithm, the zeros of the precoders in (5.2) are chosen as

$$r_i = \eta_i e^{\frac{j\pi(2i-1)}{2N_r}}, \quad i = 1, \cdots, N_r \tag{5.38}$$

where $j = \sqrt{-1}$ and $0 < \eta_i < 1$, $i = 1, \cdots, N_r$. This model ensures that all zeros are distinct and satisfy $0 < |r_i| < 1$, $i = 1, \cdots, N_r$, and the angles of zeros are equally spaced on the Z-plane. For the signal MI modified constant modulus based channel estimation algorithm, unless explicitly mentioned, the matrices $\mathbf{B}_2^{(n)}$ and $\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(n)}$ are initialized as $\mathbf{B}_2^{(0)} = \left[\mathbf{I}_{N_s}, \mathbf{0}_{N_s \times (N_r - N_s)}\right]^H$ and $\mathbf{R}_{\hat{\mathbf{y}}_r\hat{\mathbf{y}}_r}^{(0)} = \mathbf{I}_{N_r}$, respectively. We choose $\mu = 0.0005$, $\kappa = 0.05$, $\beta = 1$, and $\gamma = 1$. The simulation parameters are assigned with reasonable values. The step size of the gradient descent algorithm μ is chosen to be small enough to ensure the convergence of the algorithm, while γ is chosen to be 1 as the absolute value of the source signals has a constant unit value. We assume that the channel matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are complex Gaussian distributed with zero mean and unit variance, and channels do

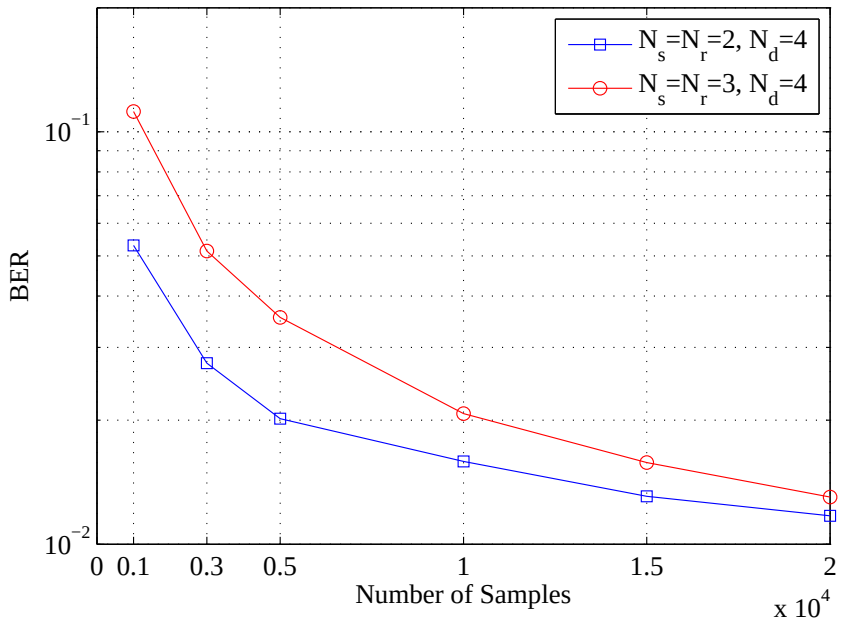

Figure 5.2: Example 5.1: BER versus number of samples for different N_s and N_r with $\mathrm{SNR_{r-d}} = \mathrm{SNR_{s-r}} = 20\mathrm{dB}$.

not change within one cycle of transmission. All simulation results are averaged over 1000 random channel realizations. The SNR of the source-relay and relay-destination link is denoted as $\mathrm{SNR_{s-r}}$ and $\mathrm{SNR_{r-d}}$, respectively.

In the first example, we evaluate the performance of the proposed blind channel estimation algorithm at various number of samples L of the received signal. Fig. 5.2 shows the bit-error-rate (BER) of the proposed algorithm versus L for various N_s and N_r with $\mathrm{SNR_{s-r}}$ and $\mathrm{SNR_{r-d}}$ fixed at 20dB. It can be seen from Fig. 5.2 that the BER performance of the proposed algorithm improves when L increases. This is because in the proposed first-order Z-domain precoding based channel estimation algorithm, the accuracy of estimating the autocorrelation matrix $\mathbf{C_{yy}}(k)$ is affected by L, i.e., the estimated $\mathbf{C_{yy}}(k)$ approaches its theoretical at a large L. Moreover, the performance of the signal MI modified constant modulus algorithm improves when a larger L is used as more iterations are involved in finding the separation matrix. In the following simulation examples, the number of samples is chosen as $L = 5000$ to achieve a good tradeoff between the performance and the computational complexity.

In the second example, we study the performance of the proposed blind estimation algorithms in finding the separation matrix. For each channel realization,

the mean interference rejection level (MIRL) for the first-order Z-domain precoding algorithm is calculated as

$$\mathrm{MIRL}_{\mathbf{H}_2} = 10\log_{10}\left(\frac{1}{N_r(N_r-1)}\sum_{i=1}^{N_r}\sum_{j=1,j\neq i}^{N_r}\frac{|(\mathbf{B}_1^H\mathbf{H}_2)_{ij}|^2}{|(\mathbf{B}_1^H\mathbf{H}_2)_{ii}|^2}\right) \tag{5.39}$$

while the MIRL of the signal MI modified constant modulus algorithm is given by

$$\mathrm{MIRL}_{\mathbf{H}_1} = 10\log_{10}\left(\frac{1}{N_s(N_s-1)}\sum_{i=1}^{N_s}\frac{\sum_{j=1}^{N_s}|(\mathbf{C})_{ij}|^2 - \max_j(|(\mathbf{C})_{ij}|^2)}{\max_j(|(\mathbf{C})_{ij}|^2)}\right). \tag{5.40}$$

Note that a smaller value of MIRL indicates a better performance of the blind channel estimation algorithm.

Fig. 5.3 shows the MIRL for the proposed blind channel estimation algorithms versus $\mathrm{SNR}_{\mathrm{r-d}}$ with $N_s = N_r = 3$, $N_d = 4$, and $\mathrm{SNR}_{\mathrm{s-r}} = 20$dB. It can be seen from Fig. 5.3 that the MIRL performance of the proposed blind channel estimation algorithm improves with the increase of $\mathrm{SNR}_{\mathrm{r-d}}$. Interestingly, the first-order Z-domain precoding technique performs better than the signal MI modified constant modulus algorithm, as the latter algorithm is affected by the accuracy of the estimation of $\mathbf{y}_r(n)$. Note that for the first-order Z-domain precoding technique, theoretically the derivation of the separation matrix is not affected by the noise at the destination node, thus only a small improvement is observed when $\mathrm{SNR}_{\mathrm{r-d}}$ increases. A plot of the MIRL of the proposed blind channel estimation algorithms versus $\mathrm{SNR}_{\mathrm{r-d}}$ for $\mathrm{SNR}_{\mathrm{s-r}} = 20$dB and different N_s and N_r is shown in Fig. 5.4. It can be seen from Fig. 5.4 that when the number of antennas at the source node and relay node increases, the MIRL also increases.

In the third example, we demonstrate the performance of the proposed blind channel estimation algorithms in terms of the NMSE. For the relay-destination channel, the NMSE is calculated as

$$\mathrm{NMSE}_{\mathbf{H}_2} = \frac{\|\mathbf{H}_2 - \hat{\mathbf{H}}_2\|_F^2}{N_r N_d}. \tag{5.41}$$

Similarly, the NMSE for the estimation of the source-relay channel matrix is given by

$$\mathrm{NMSE}_{\mathbf{H}_1} = \frac{\|\mathbf{H}_1 - \hat{\mathbf{H}}_1\|_F^2}{N_s N_r}. \tag{5.42}$$

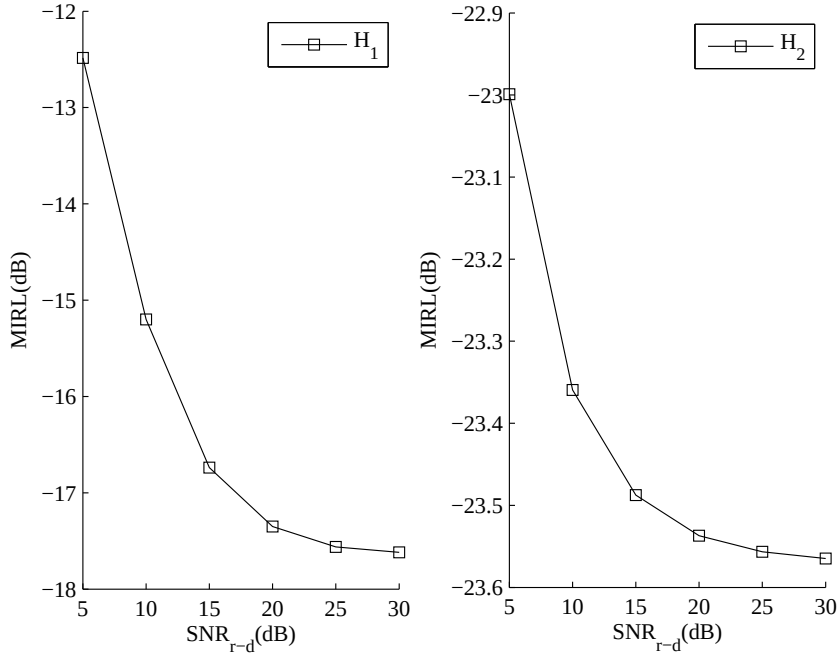

Figure 5.3: Example 5.2: MIRL versus $\mathrm{SNR_{r-d}}$ for $N_s = N_r = 3$, $N_d = 4$, and $\mathrm{SNR_{s-r}} = 20\mathrm{dB}$.

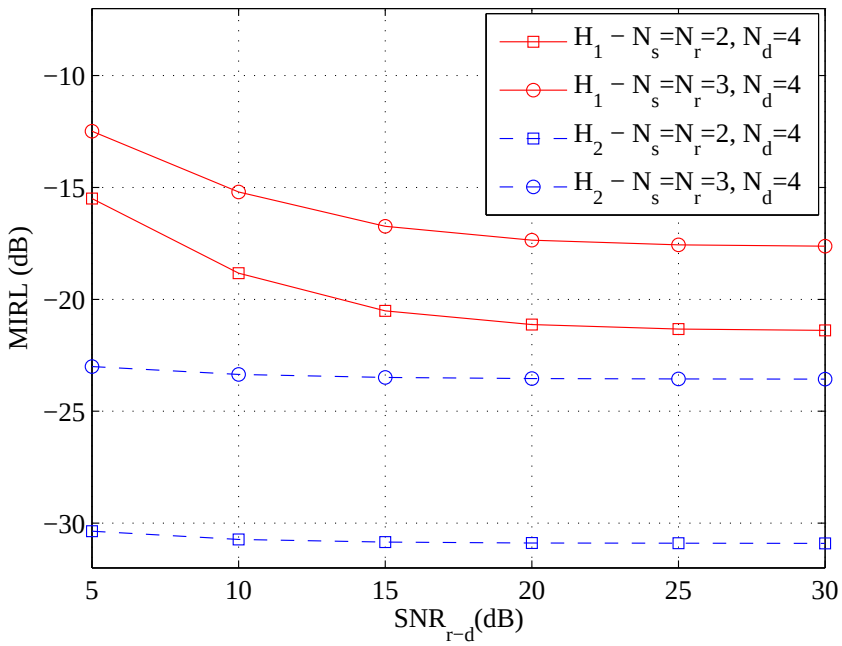

Figure 5.4: Example 5.2: MIRL versus $\mathrm{SNR_{r-d}}$ for different N_s and N_r with $\mathrm{SNR_{s-r}} = 20\mathrm{dB}$.

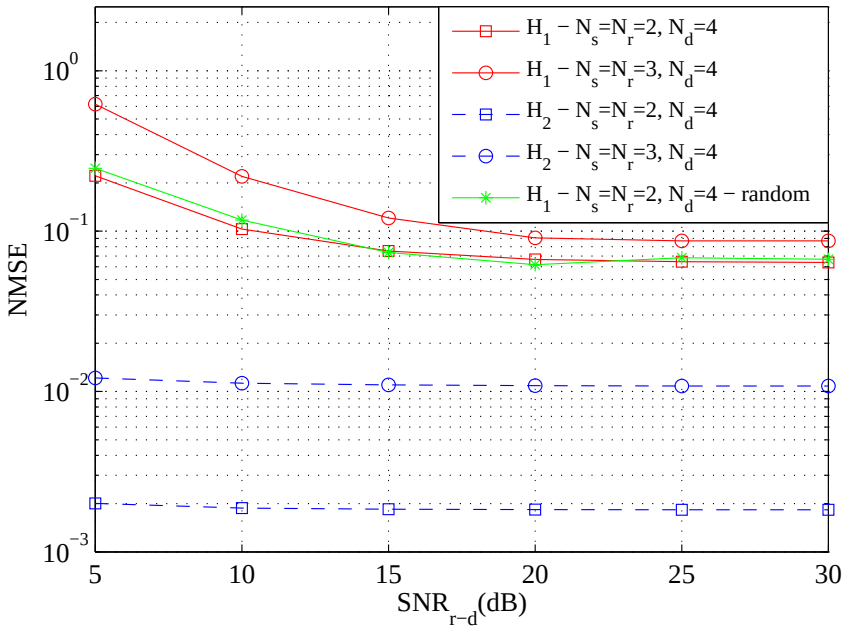

Figure 5.5: Example 5.3: Normalized MSE versus $\mathrm{SNR}_{\mathrm{r-d}}$ for different N_s and N_r with $\mathrm{SNR}_{\mathrm{s-r}} = 20\mathrm{dB}$.

Similar to [76], the scaling ambiguity in estimating $\mathbf{H}_2$ is removed by minimizing the mean square error between $\mathbf{H}_2$ and $\hat{\mathbf{H}}_2$. The scaling and permutation ambiguity in the estimation of $\mathbf{H}_1$ is removed by minimizing the mean square error between $\mathbf{s}(n)$ and $\hat{\mathbf{s}}(n)$.

Fig. 5.5 shows the NMSE of the proposed blind estimation algorithm versus $\mathrm{SNR}_{\mathrm{r-d}}$ for different N_s and N_r with $\mathrm{SNR}_{\mathrm{s-r}}$ fixed at 20dB. It can be seen from Fig. 5.5 that the NMSE of estimating $\mathbf{H}_1$ and $\mathbf{H}_2$ decreases when the number of antennas at the source and relay nodes decreases. Note that only small improvement is observed in the estimation of $\mathbf{H}_2$ when $\mathrm{SNR}_{\mathrm{r-d}}$ increases as theoretically, the estimation of $\mathbf{H}_2$ is not affected by $\mathrm{SNR}_{\mathrm{r-d}}$. We also investigate the performance of signal MI modified constant modulus channel estimation scheme when the algorithm is initialized with random matrices. It can be seen from Fig. 5.5 that the NMSE of the first-hop channel estimation with random matrices initialization is very similar to the NMSE when the scheme is initialized with identity matrix.

In the fourth example, we compare the proposed blind MIMO relay channel estimation algorithm with the training-based MIMO relay channel estimation algorithm developed in [86], where the training sequences are optimized with proper adjustment of simulation parameters for a fair comparison. The channel

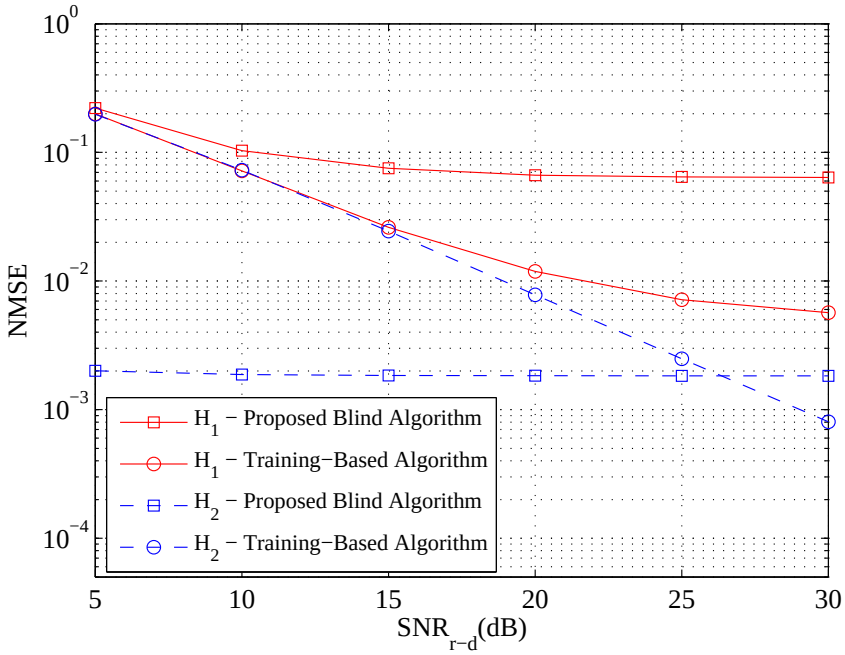

Figure 5.6: Example 5.4: Normalized MSE versus $\mathrm{SNR_{r-d}}$ for $N_s = N_r = 2$ and $N_d = 4$ with $\mathrm{SNR_{s-r}} = 20\mathrm{dB}$.

correlation matrices used in the training-based algorithm [86] are set to identity matrices to have the same statistical distribution as the channel model used in our proposed blind channel estimation algorithm. Fig. 5.6 shows the NMSE performance of estimating $\mathbf{H}_2$ and $\mathbf{H}_1$ versus $\mathrm{SNR_{r-d}}$ with $\mathrm{SNR_{s-r}} = 20\mathrm{dB}$, $N_s = N_r = 2$, and $N_d = 4$. The MSE performance of two algorithms versus $\mathrm{SNR_{r-d}}$ with $\mathrm{SNR_{s-r}} = 20\mathrm{dB}$, $N_s = N_r = 3$, and $N_d = 4$ is demonstrated in Fig. 5.7. It can be seen from Figs. 5.6 and 5.7 that at low SNR, the performance of the proposed algorithm is comparable to that of the training-based algorithm. However, at high SNR, the training-based algorithm outperforms the proposed algorithm at the expense of bandwidth efficiency.

Fig. 5.8 illustrates the BER performance of two algorithms versus $\mathrm{SNR_{r-d}}$ when $N_s = N_r = 2$, $N_d = 4$, and $\mathrm{SNR_{s-r}}$ is fixed at 20dB. As a benchmark, we also show the BER performance of the MIMO relay system when the channel matrices are perfectly known. It can be seen from Fig. 5.8 that the BER performance of our proposed blind channel estimation algorithm is close to the performance of the training-based algorithm.

Finally, we compare the computational complexity of the proposed blind channel estimation algorithm and the training-based channel estimation technique.

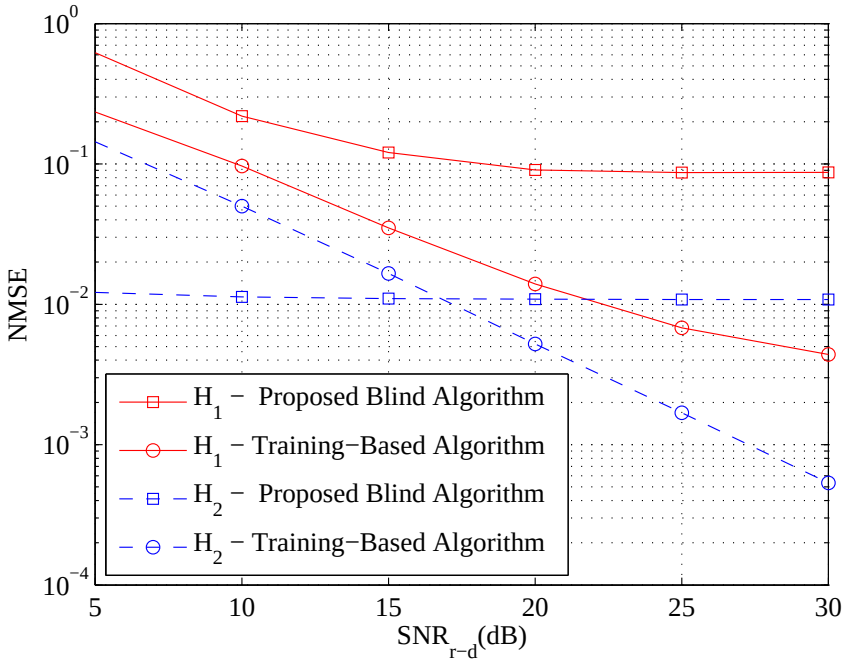

Figure 5.7: Example 5.4: Normalized MSE versus $\mathrm{SNR}_{\mathrm{r-d}}$ for $N_s = N_r = 3$ and $N_d = 4$ with $\mathrm{SNR}_{\mathrm{s-r}} = 20\mathrm{dB}$.

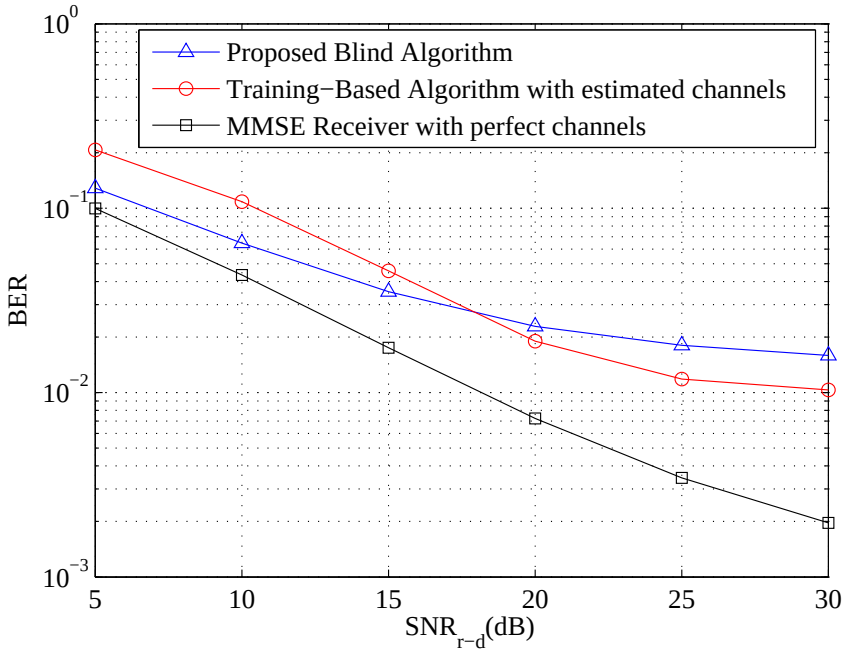

Figure 5.8: Example 5.4: BER versus $\mathrm{SNR}_{\mathrm{r-d}}$ for $N_s = N_r = 2$ and $N_d = 4$ with $\mathrm{SNR}_{\mathrm{s-r}} = 20\mathrm{dB}$.

The complexity of the first-order Z-domain precoding based channel estimation algorithm is governed by the EVD and the SVD operations required in deriving the separation matrix, while the complexity of the signal MI modified constant modulus algorithm is governed by the matrix inversion operation in the gradient descent method. Thus, the computational complexity of the proposed blind channel estimation algorithm can be estimated as $\mathcal{O}(N_d^3 + LN_s^3)$, where the first term represents the complexity of the first-order Z-domain precoding based channel estimation algorithm, and the second term is the complexity of the signal MI modified constant modulus algorithm.

The complexity of the training-based channel estimation technique can be estimated as $\mathcal{O}(d_\nu d_c N_r^2 + d_a d_{\mu_1} d_{\lambda_F} N_s + d_a d_{\mu_2} d_{\mu_3} d_{\lambda_S} N_d)$, where $d_\nu, d_{\mu_1}, d_{\mu_2}$, and d_{μ_3} stand for the number of iterations required to obtain the optimal Lagrangian multipliers associated with the optimization problem in [86], d_c and d_{λ_S} represent the number of bisection operations required to obtain the optimal training sequences, d_{λ_F} is the number of bisection operations required to derive the optimal relay amplification matrix, and d_a stands for the number of iterations required to find the local optimal solution to the problem.

5.6 Conclusions

We have developed a new blind channel estimation algorithm for two-hop MIMO relay systems. The proposed algorithm is able to estimate the individual source-relay and relay-destination CSI at the destination node, which is necessary for retrieving the source signals at the destination node. In particular, a novel first-order Z-domain precoding technique has been developed for the blind estimation of the relay-destination channel matrix. The proposed algorithm has a similar BER performance to the training-based channel estimation algorithm, and better bandwidth efficiency as all the bandwidth is used for sending communication signals. The proposed algorithm can be extended to other MIMO relay communication systems such as multiuser MIMO relay systems with multiple relay nodes.

5.A Proof of Theorem 5.1

We prove Theorem 5.1 through verifying the necessity and sufficiency conditions. Assuming that (5.14) is satisfied, we prove the necessity of (5.15) as follow

$$\mathbf{b}_{1,i}^H \bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{b}_{1,i}^H \mathbf{H}_{2,i} \mathbf{T}_i(r_i) \mathbf{H}_2^H = \mathbf{0}. \tag{5.43}$$

Since we assumed the source signals to be temporally white, from (5.3), (5.8), and (5.14), we prove the necessity condition for (5.16) as

$$\begin{aligned}
\mathbf{b}_{1,i}^H \mathbf{C}_{\mathbf{yy}}(1) \mathbf{b}_{1,i} &= \mathbf{b}_{1,i}^H \mathbf{H}_2 \mathbf{C}_{\mathbf{xx}}(1) \mathbf{H}_2^H \mathbf{b}_{1,i} \\
&= c_i c_i^* \mathrm{E}\big[x_i(n) x_i(n-1)^*\big] \\
&= |c_i|^2 \mathrm{E}\Big[\big(y_{r,i}(n) - r_i y_{r,i}(n-1)\big)\big(y_{r,i}(n-1) - r_i y_{r,i}(n-2)\big)^*\Big] \\
&= |c_i|^2 \mathrm{E}\big[-r_i y_{r,i}(n-1) y_{r,i}(n-1)^*\big] \\
&= -|c_i|^2 r_i \sigma_i^2 \\
&\neq 0
\end{aligned}$$

where $\sigma_i^2 \triangleq \mathrm{E}[y_{r,i}(n-1) y_{r,i}(n-1)^*]$.

Now we prove the sufficiency of (5.15) and (5.16). Since $\mathbf{b}_{1,i}^H \bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{0}$, from (5.13) we have

$$\mathbf{b}_{1,i}^H \mathbf{H}_{2,i} \mathbf{T}_i(r_i) \mathbf{H}_2^H = \mathbf{0}. \tag{5.44}$$

The matrix $\mathbf{H}_2^H$ is of full row rank, and thus implying that

$$\mathbf{b}_{1,i}^H \mathbf{H}_{2,i} \mathbf{T}_i(r_i) = \mathbf{0}. \tag{5.45}$$

From Lemma 5.1, all the rows of the matrix $\mathbf{T}_i(r_i)$ excluding the ith row are linearly independent, and therefore we obtain that

$$\mathbf{b}_{1,i}^H \mathbf{h}_{2,j} = 0, \quad j = 1, \cdots, N_r, \quad j \neq i. \tag{5.46}$$

Subsequently, from (5.11) and (5.46), we have

$$\mathbf{b}_{1,i}^H \mathbf{H}_2 = \left[0, \cdots, 0, \mathbf{b}_{1,i}^H \mathbf{h}_{2,i}, 0, \cdots, 0\right]. \tag{5.47}$$

Next, we consider $\mathbf{b}_{1,i}^H \mathbf{C}_{\mathbf{yy}}(1) \mathbf{b}_{1,i} \neq 0$. From (5.8), we have

$$\mathbf{b}_{1,i}^H \mathbf{H}_2 \mathbf{C}_{\mathbf{xx}}(1) \mathbf{H}_2^H \mathbf{b}_{1,i} \neq 0 \tag{5.48}$$

which implies that $\mathbf{b}_{1,i}^H \mathbf{H}_2 \neq 0$, and from (5.47), we can infer that

$$\mathbf{b}_{1,i}^H \mathbf{H}_2 = [0, \cdots, 0, c_i, 0, \cdots, 0] \tag{5.49}$$

where $c_i = \mathbf{b}_{1,i}^H \mathbf{h}_{2,i} \neq 0$. □

5.B Proof of Corollary 5.1

5.B.1 For $\tau = 0$

For the case of $N_d > N_r$, the channel matrix $\mathbf{H}_2$ has a row-rank deficiency, i.e., the rows of $\mathbf{H}_2$ are linearly dependent. Subsequently, an $N_d \times 1$ non-zero vector $\mathbf{b}_i$ exists such that

$$\mathbf{b}_i^H \mathbf{H}_2 = \mathbf{0}. \tag{5.50}$$

From (5.13) and (5.50), we have

$$\mathbf{b}_i^H \bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{b}_i^H \mathbf{H}_{2,i} \mathbf{T}_i(r_i) \mathbf{H}_2^H = \mathbf{0}. \tag{5.51}$$

Based on (5.8) and (5.50), there is

$$\begin{aligned} \mathbf{b}_i^H \mathbf{C}_{\mathbf{yy}}(0) \mathbf{b}_i &= \mathbf{b}_i^H \mathbf{H}_2 \mathbf{C}_{\mathbf{xx}}(0) \mathbf{H}_2^H \mathbf{b}_i + \mathbf{b}_i^H \mathbf{C}_{\mathbf{v_d v_d}}(0) \mathbf{b}_i \\ &= \mathbf{b}_i^H \mathbf{C}_{\mathbf{v_d v_d}}(0) \mathbf{b}_i \\ &= \sigma_w^2 \mathbf{b}_i^H \mathbf{b}_i \\ &\neq 0. \end{aligned} \tag{5.52}$$

It can be observed from (5.50)-(5.52) that $\mathbf{b}_i^H \bar{\mathbf{Q}}_{\mathbf{yy}}(r_i) = \mathbf{0}$ and $\mathbf{b}_i^H \mathbf{C}_{\mathbf{yy}}(0) \mathbf{b}_i \neq 0$ do not guarantee (5.14).

5.B.2 For $\tau \geq 2$

Assuming (5.14) is satisfied, we have

$$\begin{aligned} \mathbf{b}_i^H \mathbf{C}_{\mathbf{yy}}(\tau) \mathbf{b}_i &= \mathbf{b}_i^H \mathbf{H}_2 \mathbf{C}_{\mathbf{xx}}(\tau) \mathbf{H}_2^H \mathbf{b}_i \\ &= c_i c_i^* \mathrm{E}\big[x_i(n) x_i(n-\tau)^*\big] \\ &= |c_i|^2 \mathrm{E}\Big[\big(y_{r,i}(n) - r_i y_{r,i}(n-1)\big)\big(y_{r,i}(n-\tau) - r_i y_{r,i}(n-\tau-1)\big)^*\Big] \\ &= 0. \end{aligned}$$

This indicates that no separation vector $\mathbf{b}_i$ can satisfy the condition $\mathbf{b}_i^H \mathbf{C}_{\mathbf{yy}}(\tau) \mathbf{b}_i \neq 0$ for time lag $\tau \geq 2$. □

Chapter 6

Conclusions and Future Work

In practical MIMO wireless relay communication networks, the knowledge of the instantaneous CSI is unavailable at both transmitter and receiver. As CSI is important for the retrieval of the source information and the optimization of MIMO relay networks, it is necessary to estimate the instantaneous CSI. In this thesis, several efficient channel estimation techniques have been developed for cooperative MIMO wireless communication systems. Some final remarks and possible future works are given in Section 6.1 and Section 6.2, respectively.

6.1 Concluding Remarks

This thesis investigates the channel estimation problem for MIMO relay networks using two methods, the training-based scheme and the blind technique. Chapter 2, 3, and 4 focus on the development of training-based channel estimation algorithm while Chapter 5 addresses the channel estimation issue using the blind technique. In Chapter 2, a robust channel estimation algorithm for one-way MIMO relay networks is studied. This robust algorithm incorporates the effect of the mismatch between the estimated and true relay-destination channel into the estimation of source-relay channel. It has been proven that the overall channel estimation algorithm performs better when the mismatch is taken into consideration.

In Chapter 3, two channel estimation algorithms, which are the superimposed channel training and two-stage channel estimation schemes, are proposed and

compared for frequency-flat two-way MIMO relay communication systems. The proposed algorithms are able to estimate the individual channel matrices involved in the communication efficiently. In comparison, the two-stage channel estimation scheme has a better performance than the superimposed channel training algorithm at the expense of higher computational complexity.

This work is extended in Chapter 4 by applying the method of superimposed channel training to a more general situation where two-way MIMO relay communication networks are operating in the frequency-selective fading environments. The proposed algorithm can efficiently estimate the individual channel matrices for two-way MIMO relay networks with frequency-selective fading channels. The optimal structure of the training sequences and the optimal power allocation between the source and relay training sequences are derived.

Last but not least, a blind channel estimation algorithm for two-hop MIMO relay systems is developed in Chapter 5. This algorithm integrates two BSS methods, namely first-order Z-domain precoding technique and signal MI modified constant modulus algorithm, to estimate the individual CSI for the source-relay and relay-destination links. Compared with training-based methods, the proposed blind channel estimation algorithm has a better spectrum efficiency as no transmission of training sequence is required.

6.2 Future Works

Several efficient channel estimation algorithms for cooperative MIMO wireless communication networks have been developed in this thesis. Nonetheless, the works in this thesis can be further extended in many possible ways. In Chapter 2, a local optimal solution is found for the channel estimation problem in one-way MIMO relay network. It will be interesting to develop an algorithm that gives a global solution to the channel estimation problem in Chapter 2.

Channel estimation problem for MIMO relay networks with frequency-flat fading channels and frequency-selective fading channels have been investigated in Chapter 3 and 4, respectively. These works can be extended to consider the channel estimation issue for MIMO relay networks operating in time-selective

fading environments, where the users are moving at a high speed.

In Chapter 5, a blind channel estimation algorithm is proposed for one-way MIMO relay networks. Further research can be done on developing blind channel estimation algorithms for two-way MIMO relay networks. Moreover, any possible method to improve the convergence and accuracy of the blind channel estimation algorithm can be an interesting future work.

This thesis focuses on the MIMO relay networks with the amplify-and-forward relaying scheme. It will also be interesting to compare the works in this thesis with the channel estimation algorithms developed for MIMO relay networks with decode-and-forward relaying scheme in terms of the complexity and performance. Such tradeoff between the complexity and performance can be investigated to maximize the benefits of the MIMO relay networks.

In addition, the algorithms derived in this thesis do not consider OFDM modulated relay networks as an option. The channel estimation problem for MIMO-OFDM relay networks can be included in future works.

Finally, any novel channel estimation algorithm, either using the training-based or blind technique, is of great interest.

References

[1] E. C. Van Der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120-154, 1971.

[2] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572-584, Sep. 1979.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939-1948, Nov. 2003.

[4] 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. 1927-1938, Nov. 2003.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[6] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037-3063, Sep. 2005.

[7] D. Gunduz and E. Erkip, "Source and channel coding for cooperative relaying," *IEEE Trans. Inform. Theory*, vol. 53, no. 10, pp. 3454-3475, Oct. 2007.

[8] W. Su, A. Sadek, and K. J. R. Liu, "Cooperative communication protocols in wireless networks: Performance analysis and optimum power allocation," *Wireless Pers. Commun.*, vol. 44, no. 2, pp. 181-217, Jan. 2008.

[9] B. Sirkeci-Mergen and M. Gastpar, "On the broadcast capacity of wireless networks with cooperative relays," *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3847-3861, Aug. 2010.

[10] G. Huang, Y. Wang, and J. Coon, "Performance of multihop decode-and-forward and amplify-and-forward relay networks with channel estimation," *Proc. 2011 IEEE Pacific Rim Conf. Commun. Computers and Signal Process. (PacRim)*, Victoria, BC, Aug. 23-26, 2011, pp. 352-357.

[11] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear non-regenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837-4851, Dec. 2009.

[12] Y. Rong and Y. Hua, "Optimality of diagonalization of multi-hop MIMO relays," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6068-6077, Dec. 2009.

[13] Y. Rong, "Linear non-regenerative multicarrier MIMO relay communications based on MMSE criterion," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1918-1923, Jul. 2010.

[14] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1099-1109, Aug. 2004.

[15] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16J relay-based wireless access networks: an overview," *IEEE Wireless Commun.*, vol. 15, no. 5, pp. 56-63, Oct. 2008.

[16] Q. Li, R. Q. Hu, Y. Qian, and G. Wu, "Cooperative communications for wireless networks: techniques and applications in LTE-advanaced systems," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 22-29, Apr. 2012.

[17] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading Environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311-335, 1998.

[18] K. Du and M. N. S. Swamy, *Wireless Communication Systems: From RF subsystems to 4G enabling technologies*. Cambridge University Press, 2010.

[19] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[20] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications - A comprehensive literature survey," *IEEE Commun. Surveys & Tutorials*, vol. 11, no. 2, pp. 87-105, 2009.

[21] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073-1096, May. 2003.

[22] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585-595, Nov/Dec. 1999.

[23] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41-59, 1996.

[24] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281-302, Apr. 2003.

[25] J. P. Kermoal, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1211-1226, Nov. 2002.

[26] H. Boche, A. Bourdoux, J. R. Fonollosa, T. Kaiser, A. Molisch, and W. Utschick, "Antennas: state of the art," *IEEE Veh. Technol. Magazine*, vol. 1, no. 1, pp. 8-17, Mar. 2006.

[27] T. Kaiser, A. Bourdoux, M. Rupp, and U. Heute, "Implementation aspects and testbeds for MIMO systems," *EURASIP Journal Applied Signal Process.*, vol. 2006, article 69217, pp. 1-3, 2006.

[28] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications - a key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198-218, Feb. 2004.

[29] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[30] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998.

[31] J. H. Winters, "The diversity gain of transmit diversity in wireless systems with Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 47, no. 1, pp. 119-123, Feb. 1998.

[32] T. Paul and T. Ogunfunmi, "Wireless LAN comes of age: Understanding the IEEE 802.11n amendment," *IEEE Circuits and Syst. Magazine*, vol. 8, no. 1, pp. 28-54, 2008.

[33] A. Greenspan, M. Klerer, J. Tomcik, R. Canchi, and J. Wilson, "IEEE 802.20: Mobile broadband wireless access for the twenty-first century," *IEEE Commun. Magazine*, vol. 46, no. 7, pp. 56-63, July 2008.

[34] A. Schwager, D. Schneider, W. Baschlin, A. Dilly, and J. Speidel, "MIMO PLC: Theory, measurements and system setup," *Proc. 2011 IEEE Int. Symposium Power Line Commun. Its Applications (ISPLC)*, Udine, Apr. 3-6, 2011, pp. 48-53.

[35] L. T. Berger, A. Schwager, P. Pagani, and D. M. Schneider, "MIMO power line communications," *IEEE Commun. Surveys & Tutorials*, to appear, 2014.

[36] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398-1407, Apr. 2007.

[37] I. Hammerström and A. Wittneben, "Power allocation schemes for amplify-and-forward MIMO-OFDM relay links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2798-2802, Aug. 2007.

[38] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 29-43, Jan. 2005.

[39] M. R. A. Kandaker and Y. Rong, "Precoding design for MIMO relay multicasting," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3544-3555, Jul. 2013.

[40] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on transceiver design for amplify-and-forward MIMO relay systems," *IEEE J. Select. Areas Commun.*, vol. 30, no. 8, pp. 1331-1346, Sep. 2012.

[41] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774-1786, May 2007.

[42] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimization of a MIMO relay network," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5062-5073, Oct 2008.

[43] Y. Rong, "Optimal joint source and relay beamforming for MIMO relays with direct link," *IEEE Commun. Letters*, vol. 14, no. 5, pp. 390-392, May 2010.

[44] Y. Rong, "Robust design for linear non-regenerative MIMO relays with imperfect channel state information," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2455-2460, May 2011.

[45] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "On the design of amplify-and-forward MIMO-OFDM relay systems with QoS requirements specified as schur-convex functions of the MSEs," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1871-1877, May 2013.

[46] Y. Rong and M. R. A. Kandaker, "On uplink-downlink duality of multi-hop MIMO relay channel," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1923-1931, Jun. 2011.

[47] X. Tang and Y. Hua, "Optimal waveform design for MIMO relaying," *Proc. 2005 IEEE 6th Workshop Signal Processing Advances in Wireless Commun.*, New York, USA, Jun. 5-8, 2005, pp. 289-293.

[48] O. Munoz, J. Vidal, and A. Agustin, "Non-regenerative MIMO relaying with channel state information," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2005 (ICASSP'05)*, Mar. 18-23, 2005, vol. 3, pp. III/361-III/364.

[49] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Commun. Surveys & Tutorials*, vol. 9, no. 2, pp. 18-48, 2007.

[50] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 12-25, Nov. 2004.

[51] M. L. Moher and J. H. Lodge, "TCMP-a modulation and coding strategy for Rician fading channels," *IEEE J. Select. Areas Commun.*, vol. 7, no. 9, pp. 1347-1355, Dec. 1989.

[52] S. Sampei and T. Sunaga, "Rayleigh fading compensation method for 16QAM in digital land mobile radio channels," in *Proc. IEEE 39th Vehicular Technol. Conf. 1989*, San Francisco, California, May. 1-3, 1989, vol. 2, pp. 640-646.

[53] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. on Vehicular Technology*, vol. 40, no. 4, pp. 686-693, Nov. 1991.

[54] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, Sep. 22-24, 1999.

[55] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 549-553, Oct. 2002.

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

[57] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," *Proc. Global Telecommun. Conf., 1999 (GLOBECOM'99)*, Rio de Janeireo, 1999, vol. 4, pp. 2162-2166.

[58] A. Scaglione and A. Vosoughi, "Turbo estimation of channel and symbols in precoded MIMO systems," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2004 (ICASSP'04)*, May 17-21, 2004, vol. 4, pp. IV/413-IV/416.

[59] M. Biguesh and A. B. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP Journal Applied Signal Process. 2004:9*, pp. 1330-1339, 2004.

[60] M. Biguesh and A. B. Gershman, "MIMO channel estimation: Optimal training and tradeoffs between estimation techniques," in *Proc. IEEE Int. Conf. Commun. 2004*, Jun. 20-24, 2004.

[61] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 884-893, Mar. 2006.

[62] M. Biguesh, S. Gazor, and M. H. Shariat, "Optimal training sequence for MIMO wireless systems in colored environment," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3144-3153, Aug. 2009.

[63] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807-1820, Mar. 2010.

[64] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 453-466, Mar. 2005.

[65] H. Zhang, Y. Li, A. Reid, and J. Terry, "Optimum training symbol design for MIMO OFDM in correlated fading channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2343-2347, Sep. 2006.

[66] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Robust Channel Estimation Algorithm for Dual-Hop MIMO Relay Channels," *Proc. 23rd IEEE Int. Symposium Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sep. 9-12, 2012, pp. 2376-2381.

[67] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Superimposed Channel Training for Two-Way MIMO Relay Systems," *Proc. 13th IEEE Int. Conf. Commun. Syst. (ICCS)*, Singapore, Nov. 21-23, 2012, pp. 21-25.

[68] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel training algorithms for two-way MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3988-3998, Aug. 2013.

[69] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Estimation for Frequency-Selective Two-Way MIMO Relay Systems," *Proc. Int. Symposium Inf. Theory Its Applications (ISITA'2014)*, Melbourne, Australia, Oct. 26-29, 2014.

[70] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Channel Estimation for Two-Way MIMO Relay Systems in Frequency-Selective Fading Environments," *IEEE Trans. Wireless Commun.*, to appear, 2014.

[71] A. Ikhlef and D. L. Guennec, "A simplified constant modulus algorithm for blind recovery of MIMO QAM and PSK signals: A criterion with convergence analysis," *EURASIP Journal Wireless Commun. Network.*, article ID 90401, Sep. 2007.

[72] Y. Xiang, N. Gu, and K. L. Wong, "Adaptive blind source separation using constant modulus criterion and signal mutual information," in *Proc. IEEE Int. Conf. Industrial Technol.*, pp. 1371-1375, Dec. 2005.

[73] X. G. Doukopoulos and G. V. Moustakides, "Adaptive power techniques for

blind channel estimation in CDMA systems," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 1110-1120, Mar. 2005.

[74] C.-C. Tu and B. Champagne, "Subspace-based blind channel estimation for MIMO-OFDM systems with reduced time averaging," *IEEE Trans. on Vehicular Technology*, vol. 59, pp. 1539-1544, Mar. 2010.

[75] S. Abdallah and I. Psaromiligkos,, "Blind channel estimation for amplify-and-forward two-way relay networks employing M-PSK modulation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3604-3615, 2012.

[76] Y. Xiang, D. Peng, Y. Xiang, and S. Guo, "Novel Z-domain precoding method for blind separation of spatially correlated signals," *IEEE Trans. Neural Networks and Learning Systems*, vol. 24, pp. 94-105, Jan. 2013.

[77] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Blind Estimation of MIMO Relay Channels," *Proc. IEEE Workshop on Statistical Signal Processing (SSP'2014)*, Gold Coast, Australia, Jun. 29-Jul. 2, 2014, pp. 400-403.

[78] C. W. R. Chiong, Y. Rong, and Y. Xiang, "Blind Channel Estimation and Signal Retrieving for MIMO Relay Systems," *IEEE Trans. Signal Process.*, revised, Sep. 2014.

[79] F. Raphel and S. M. Sameer, "A novel interim channel estimation technique for MIMO mimicking AF cooperative relay systems," *Proc. National Conf. Commun.*, Feb. 2012.

[80] M. Yuksel and E. Erkip, "Diversity-multiplexing tradeoff in cooperative wireless systems," *Proc. 40th Annual Conf. Inform. Sciences Syst.*, pp. 1062-1067, Mar. 2006.

[81] P. Lioliou, M. Viberg, and M. Matthaiou, "Bayesian channel estimation techniques for AF MIMO relaying systems," *Proc. IEEE Vehicular Techn. Conf.*, Sep. 2011.

[82] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay systems via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224-2233, June 2012.

[83] P. Lioliou and M. Viberg, "Least-squares based channel estimation for MIMO relays," *Proc. International ITG Workshop on Smart Antennas*, pp. 90-95, Feb. 2008.

[84] P. Lioliou, M. Viberg, and M. Coldrey, "Efficient channel estimation techniques for amplify and forward relaying systems," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3150-3155, Nov. 2012.

[85] F. Gao, B. Jiang, X. Gao, and X. Zhang, "Superimposed training based channel estimation for OFDM modulated amplify-and-forward relay networks," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 2029-2039, Jul. 2011.

[86] T. Kong and Y. Hua, "Optimal design of source and relay pilots for MIMO relay channel estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4438-4446, Sep. 2011.

[87] Q. Ma and C. Tepedelenlioglu, "Antenna selection for space-time coded systems with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 710-719, Feb. 2007.

[88] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory.* Englewood Cilffs, NJ: Prentice Hall, 1993.

[89] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772-781, Sep. 1978.

[90] D. S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 503-513, Mar. 2000.

[91] A. Gupta and D. Nagar, *Matrix Variate Distributions.* London, U.K.: Chapman & Hall/CRC, 2000.

[92] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Probability Statistics*, Berkeley, CA, vol. 1, pp. 611-644, 1961.

[93] S. Xu and Y. Hua, "Optimal design of spatial source-and-relay matrices for a non-regenerative two-way MIMO relay system," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1645-1655, May 2011.

[94] R. Wang, M. Tao, and Y. Huang, "Linear precoding designs for amplify-and-forward multiuser two-way relay systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4457-4469, Dec. 2012.

[95] Y. Rong, "Joint source and relay optimization for two-way linear non-regenerative MIMO relay communications," *IEEE Trans. Signal Process.*, vol. 60, no. 12, Dec. 2012.

[96] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024-3033, Oct. 2009.

[97] C. K. Ho, R. Zhang, and Y.-C. Liang, "Two-way relaying over OFDM: Optimized tone permutation and power allocation," in *Proc. IEEE ICC*, pp. 3908-3912, May 2008.

[98] Z. Fang, J. Shi, and H. Shan, "Comparison of channel estimation schemes for MIMO two-way relaying systems," in *Proc. Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, vol. 1, pp. 719-722, Jul. 2011.

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

[100] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Spring Street, NY: Springer Science+Business Media, LCC, 2007.

[101] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. 2nd Ed., Springer, 2009.

[102] F. Roemer and M. Haardt, "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Trans. Signal Process.*, vol. 58, pp. 5720-5735, Nov. 2010.

[103] H. Furukawa, Y. Kamio, and H. Sasaoka, "Cochannel interference reduction and path-diversity reception technique using CMA adaptive array antenna

in digital land mobile communications," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 605-616, Mar. 2001.

[104] J. J. Shynk, A. V. Keerthi, and A. Mathur, "Steady-state analysis of the multistage constant modulus array," *IEEE Trans. Signal Process.*, vol. 44, pp. 948-962, Apr. 1996.

[105] P. Lioliou, M. Viberg, and M. Coldrey "Performance analysis of relay channel estimation," in *Proc. IEEE Asilomar*, Pacific Grove, CA, USA, Nov. 2009, pp. 1533-1537.