

Semi Blind Channel Estimation for IRS-assisted MIMO Communication Systems

Venkata Naga Sai Praneeth Atmuri

Department of Electrical & Computer Engineering
McGill University
Montréal, Québec, Canada

August 11, 2023

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of
Master of Science. ©2023 Venkata Naga Sai Praneeth

Abstract

Intelligent Reflective Surface (IRS) has emerged as a promising approach in our pursuit of continuous enhancements in terms of speed, connectivity, security, latency, and other aspects of wireless communications. IRS is a controllable planar meta-surface that is made of multiple passive reflective elements, each possessing the ability to modify the properties of incident waves by manipulating their phase. This remarkable technology allows for controlling the wireless propagation environment and has a wide range of applications, including passive beamforming, sensing, localization, and security among others. To fully exploit the potential of IRS, accurate knowledge of the channel state information is required. Therefore, in this thesis, we focus on the channel estimation component of an IRS-aided Multiple Input-Mulitple Output (MIMO) communication system. Specifically, we adopt the semi-blind approach and develop a channel estimation algorithm based on the Expectation Maximization (EM) framework. In contrast to other research works, our algorithm is applicable not only to Single Input Multiple Output (SIMO) systems but to general MIMO systems, and assumes that the transmitted data is sourced from a discrete constellation, e.g., 4-QAM modulation scheme. Further, we consider two different channel estimation protocols: non-superimposed and superimposed. The non-superimposed protocol utilizes a traditional pilot-data structure, where the pilots are transmitted first followed by data symbols. On the other hand, in the superimposed protocol both the data and pilot symbols are transmitted simultaneously. To deal with the complexity involved in

computing Expectation step (E-step) of the EM algorithm for discrete data sets, a detection-based E-step is introduced where with the help of the Zero Forcing (ZF), Minimum Mean Square Error (MMSE), and Soft Decision Fixed Complexity (SDFC) detectors, we estimate the conditional probability of the unknown data symbols given the received signal. Simulation results show that the proposed algorithm performs better than a EM-based method from the recent literature, while the use of detectors for estimating the data vectors, allows us to trade off accuracy and time complexity. Finally, it is shown that the non-superimposed protocol performs comparatively better than the superimposed protocol as we increase the number of pilot symbols.

Abrégé

La surface réfléchissante intelligente (IRS) est apparue comme une approche prometteuse dans notre quête d'améliorations continues en termes de vitesse, de connectivité, de sécurité, de latence et d'autres aspects des communications sans fil. L'IRS est une méta-surface plane contrôlable composée de plusieurs éléments réfléchissants passifs, chacun possédant la capacité de modifier les propriétés des ondes incidentes en manipulant leur phase. Cette technologie remarquable permet de contrôler l'environnement de propagation sans fil et offre un large éventail d'applications, notamment la formation passive de faisceaux, la détection, la localisation et la sécurité. Pour exploiter pleinement le potentiel de l'IRS, une connaissance précise des informations sur l'état du canal est nécessaire. C'est pourquoi, dans cette thèse, nous nous concentrons sur la composante d'estimation du canal d'un système à entrées multiples et à modulation de fréquence assistée par l'IRS. Le système de communication à entrées et sorties multiples (MIMO) sont assistées par l'IRS. Plus précisément, nous adoptons l'approche semi-aveugle et nous nous concentrons sur l'estimation du canal, nous adoptons l'approche semi-aveugle et nous développons un algorithme d'estimation de canal basé sur la maximisation des attentes (EM). Contrairement à d'autres travaux de recherche, notre algorithme est applicable non seulement aux systèmes à entrée unique et sortie multiple (SIMO), mais aussi aux systèmes MIMO généraux. Systèmes MIMO généraux, et suppose que les données transmises proviennent d'une constellation discrète, par ex. constellation discrète, par exemple un schéma de modulation 4-QAM. En outre, nous considérons deux

protocoles d'estimation de canal différents : non superposé et superposé. Le protocole non superposé utilise une structure pilote-données traditionnelle, dans laquelle les pilotes sont transmis en premier, suivis des symboles de données. En revanche, dans le protocole superposé, les données et les symboles pilotes sont transmis simultanément. Pour faire face à la complexité du calcul de l'étape d'attente (E-step) de l'algorithme EM pour les ensembles de données discrètes, une étape E basée sur la détection est introduite où, avec l'aide des détecteurs Zero Forcing (ZF), Minimum Mean Square Error (MMSE), et Soft Decision Fixed Complexity (SDFC), nous estimons la probabilité conditionnelle des symboles de données inconnus étant donné le signal reçu. Les résultats des simulations montrent que l'algorithme proposé est plus performant qu'une méthode basée sur l'EM tirée de la littérature récente, tandis que l'utilisation de détecteurs pour l'estimation des vecteurs de données nous permet de faire un compromis entre la précision et la complexité temporelle. Enfin, il est démontré que le protocole non superposé est comparativement plus performant que le protocole superposé à mesure que l'on augmente le nombre de symboles pilotes.

Acknowledgements

I express an enormous amount of gratitude to my thesis supervisor, Dr. Ioannis Psaromiligkos, for his continuous support and significant assistance. Working under his guidance has been an honor, and my research has benefited greatly from his help in acquiring suitable equations. I also want to thank Prof. Jan Bajcsy from the bottom of my heart for choosing my application and allowing me to attend McGill University. I also thank the ECE department and the Financial aid department at McGill University for their generous financial assistance. I would also like to thank Ph.D. student Alexander Fernandes for his valuable assistance and insightful ideas whenever needed. I am incredibly thankful to thank my roommates and friends Vishal, Manoj, Nithila, Barfi, Rhythm, Akriti, Shubham, Nancy, Nehal, Aishwarya, Sandeep, Aashika and all others in Canada for unflagging support throughout the difficult research environment brought on by the COVID-19 epidemic. I also want to deeply appreciate my childhood friend Adharsh for helping me with my academics and never-ending support during challenging times. A special thanks to Prakriti for sharing her research, and life experiences with me and genuinely inspiring me to be the best version of myself.

Last but not least, I extend my gratitude to Vibha for her unwavering support throughout my journey and for helping me to stay focused to achieve my goal. I owe overwhelming gratitude to my parents - Bala Murali Krishna and Parvathi, and my brother Sree Sai Ganesh along with my sister-in-law Monisha for being there for me through thick and thin. I cherish

this close relationship because their support and affection have been my constant source of strength. I consider myself incredibly lucky to have such a wonderful family. Lastly, I would like to thank all my friends and family back in India for showing their constant support.

Contents

Abstract	i
Abrégé	iii
Acknowledgements	v
1 Introduction	1
1.1 Motivation	1
1.2 Thesis Contributions	2
1.3 Thesis Outline	4
1.4 Author Contributions	5
2 Literature Review	6
2.1 Intelligent Reflective Surface	6
2.2 Basic System Model	9
2.3 IRS and Relays	11
2.4 Channel estimation: Literature Review	12
2.4.1 Fundamental methods	13
2.4.2 Compressed Sensing and Matrix Factorization /Decomposition	15
2.4.3 Machine learning and Deep Learning	16
2.4.4 Semi-blind channel estimation	19
2.5 Summary	20

3 Background	22
3.1 Maximum likelihood estimation	22
3.2 Expectation Maximization	23
3.3 Expectation Maximization with Gaussian Prior	24
3.4 Kronecker Product, Khatri Rao Product, and Vectorization	27
3.5 Summary	29
4 EM-based Channel Estimation and Detection: Non-Superimposed Method	31
4.1 System model	32
4.2 EM algorithm	35
4.2.1 E-step	36
4.2.2 M-step	37
4.3 Computationally efficient variants of the EM algorithm	40
4.3.1 EM with detector	41
4.3.2 Log-max approximation	47
4.4 Complexity	48
4.5 Summary	49
5 EM-based Channel Estimation: Superimposed Method	50
5.1 System Model	51
5.2 EM algorithm	52
5.3 Summary	55
6 Simulation Results and Discussions	56
6.1 Non-Superimposed method	56
6.2 Comparison	60
6.3 Superimposed method	63

6.4	Summary	65
7	Conclusions and Future Work	66
A	Gaussian distribution: MIMO-IRS	68

List of Figures

2.1	IRS architecture. Adapted from [1]	7
2.2	Anomalous Reflection on a continuously changing surface impedance vs normal mirror that follows Snell's law (adapted from [2]).	8
2.3	Illustration of SISO-IRS aided communication system	10
4.1	System model for MIMO-IRS aided communication system	32
4.2	Non-Superimposed protocol for time period T	33
5.1	Superimposed protocol for time interval T	51
6.1	NMSE versus T_p with $T_d = 50$, number of IRS elements $N = 15$, and for a 2×4 MIMO system	58
6.2	NMSE versus T_d with $T_p = 20$, $N = 15$ and for a 2×4 MIMO system	59
6.3	NMSE vs Number of users with $T_p = 20$, $T_d = 50$, number of IRS elements $N = 15$ and $n_{tx} = 2$	60
6.4	NMSE versus T_p with $T_d = 50$, number of IRS elements $N = 32$ and for a 1×8 SIMO system	62
6.5	NMSE versus T_d with $T_p = 16$, number of IRS elements $N = 32$, and for a 1×8 SIMO system	62
6.6	NMSE vs T for $N = 15$ and for a 2×4 MIMO system	64

List of Tables

6.1	Parameters used for T_p vs NMSE	57
6.2	Time complexity for non-superimposed method with $T_p = 20$, $T_d = 50$, $N = 16$ for 2×4 MIMO system	61
6.3	Parameters used for comparison	63
6.4	Parameters used for Superimposed method	64

List of Acronyms

AF	Amplify-and-Forward.
BS	Base Station.
CRB	Cramer-Rao Bound.
CSI	Channel State Information.
DF	Decode-and-Forward.
EM	Expectation Maximization.
E-step	Expectation step.
IRS	Intelligent Reflective Surface.
LLF	Log-Likelihood Function.
LLR	Log-Likelihood Ratio.
LS	Least Squares.
MIMO	Multiple Input-Mulitple Output.
MISO	Multiple Input Single Output.
ML	Maximum likelihood.
MMSE	Minimum Mean Square Error.
NMSE	Normalized minimum mean squared error.
OFDM	Orthogonal Frequency-Division Multiplexing.
PARAFAC	PARAllel FACTor.
PM	Partial Marginalization.

QAM	Quadrature amplitude modulation.
SDFC	Soft Decision Fixed Complexity.
SIMO	Single Input Multiple Output.
SISO	Single Input Single Output.
SNR	Signal-to-noise ratio.
UE	User Equipment.
ZF	Zero Forcing.

List of Notation

x	Scalars are denoted by lower-case and not bold
\mathbf{x}	Vectors are denoted by lower-case bold
\mathbf{X}	Matrices are denoted by upper-case bold
\otimes	Kronecker product
\diamond	Khatri-Rao product
\odot	Hadamard product
$\mathbb{C}^{x \times y}$	Space of $x \times y$ complex matrices.
$(\cdot)^H$	Hermitian operator
$(\cdot)^+$	Pseudo inverse of a matrix
$\ \cdot\ ^2$	Norm
$\mathcal{CN}(\mu, \sigma^2)$	Circularly symmetric complex Gaussian distribution with mean μ and variance σ^2
\mathbf{I}_x	Identity matrix with size $x \times x$

Chapter 1

Introduction

1.1 Motivation

Recent technological advances have led to a need for better performance in terms of speed and connectivity due to the increase in cellular networks and the widespread use of wireless devices. It is projected that between 2020 to 2030 the global data traffic will progressively increase yearly by 55% [3]. Every successive generation has been primarily driven by the need to meet the increasing market requirements such as data rate, speed, latency, throughput, etc. With ongoing improvements and deployment of 5G cellular networks in most parts of the world, researchers now focus on 6G [4]. The next generation is anticipated to include numerous upgrades that will not only strive to meet market demands but also address the current communication bottlenecks, such as the high power consumption and interruptions in connectivity stemming from the unpredictability and uncontrollability of the wireless environment [3, 5]. Until lately, improvements have typically been made at the transmitter or the receiver, but an smart radio environment is necessary to address the issues outlined earlier. The term "smart radio environment" refers to the capability of controlling and programming the environment in addition to the

transmitters and receivers. A promising new technology to achieve this is Intelligent Reflective Surface (IRS) which has gained more interest in the past few years [1, 6]. An IRS is a planar meta-surface consisting of numerous passive elements, each of which can independently induce a controllable amplitude or/and phase change to the incident electromagnetic wave [7]. Due to their controllable reflection capability, IRSs allow the network designer to control the wireless environment. For example, they can be used for passive beamforming or help create a smart environment by installing them on walls [8]. However, to fully leverage the IRS capabilities, it is required that the channel state information is accurately known. Therefore, developing accurate channel estimation algorithms while employing this technology is crucial and is the primary goal of this thesis.

1.2 Thesis Contributions

In wireless communication, three channel estimation techniques are majorly dealt with training-based, blind-based, and semi-blind-based channel estimation. Among these techniques, a widely used channel estimation technique in the literature is training-based channel estimation. In this approach, the transmitted pilot symbols are known to the receiver to facilitate channel estimation. Multiple pieces of research deal with training-based channel estimation for Multiple Input-Mulitple Output (MIMO)-Intelligent Reflective Surface (IRS) systems [9–16]. Secondly, blind channel estimation is where the pilot symbols are avoided for transmission, and the channel estimation depends majorly on the received data symbols. Lastly, semi-blind channel estimation combines training and blind-based channel estimation. This method utilizes both pilot and received data symbols for channel estimation [17–22].

In this thesis, we consider the semi-blind channel estimation for the MIMO-IRS system as it has attracted relatively little research attention, thereby being the focus of this thesis. Due to the ease of system modeling and derivations, previous researches have mainly concentrated

on Gaussian distributed symbols. The discrete constellation data is overlooked due to the computational complexity involved. Regardless of the challenges, discrete constellation data is widely implemented in practical applications because discrete constellation points offer defined values and distances between them, making it simpler to encode, transmit, and decode [23]. Therefore, in this work, we deal mainly with discrete constellation data, e.g., 4-Quadrature amplitude modulation (QAM) scheme, although the same experiments can be performed with other constellations. We propose a system model for the MIMO-IRS-aided communication system that incorporates a non-superimposed protocol, where the pilot symbols are transmitted first, followed by the data symbols. A very commonly used algorithm to perform semi-blind channel estimation is Expectation Maximization (EM) algorithm, which is an iterative algorithm and is particularly useful to deal with hidden data (received data symbols) or incomplete data (channel parameters) [24, 25]. Therefore, this thesis uses the EM algorithm to perform the channel estimation. The proposed algorithm has high computational complexity, and to cope with it, we propose a detection-based Expectation step. Using detectors such as the Zero Forcing (ZF), Minimum Mean Square Error (MMSE), and Soft Decision Fixed Complexity (SDFC), we compute an estimate of the conditional probability of unknown data symbols given the received signal. Moreover, we propose another protocol, namely superimposed protocol, where both pilot and data symbols are transmitted simultaneously. Similar to the channel estimation process used for the non-superimposed protocol, we use the EM algorithm to estimate the channels for the superimposed protocol. Finally, we briefly discuss the challenges associated with assuming Gaussian distribution for data in the proposed MIMO-IRS-aided communication system.

1.3 Thesis Outline

The structure for the thesis is as follows

- In Chapter 2, we provide a detailed explanation of IRS technology, its definition, architecture, and applications. We then introduce the basic system model for SISO-IRS-aided communication system. We discuss the comparison between IRS-based systems and relay-based systems. Finally, we end the chapter by providing a detailed review of different signal processing techniques (Fundamental methods, Compressed sensing, Matrix Factorization/Decomposition, Machine learning, and Deep Learning) used in literature to perform channel estimation for IRS-aided communication systems.
- In Chapter 3, we focus our discussion on the background information required to comprehend this thesis. Therefore we first introduce Maximum likelihood (ML) algorithm and discuss the challenges that occur while dealing with hidden or incomplete data. We then discuss the EM algorithm, which can serve as an alternative to ML algorithm to deal with the random nuisance elements. In addition, we introduce the Gaussian prior based EM algorithm for Multiple Input Single Output (MISO)-IRS aided communication system as mentioned by authors of [17]. Finally, we provide the mathematical background details, such as a discussion on the Kronecker product, Khatri-Rao product, and Vectorization with the related properties that will be used in this thesis.
- In Chapter 4, we propose the system model for the non-superimposed protocol, and using EM framework, we estimate the channels. As mentioned earlier in this chapter, the proposed algorithm is computationally complex and incredibly challenging to work with discrete constellation data. Therefore, we propose two methods to deal with this complexity: (i) log-max approximation and (ii) detection based E-step. In the log-max

approximation, we detect the data vector that is associated with the highest value of the conditional probability of unknown data symbols given the received signal. In the case of detection based E-step, we estimate the conditional probability of unknown data symbols given the received signal by using detectors such as ZF, MMSE, and SDFC. We end the chapter by comparing the computational complexity of the proposed method with detection based E-step.

- In Chapter 5, we introduce the system model for the superimposed protocol. Similar to Chapter 4, we use the EM framework to estimate the channels.
- In Chapter 6, we provide the simulation results obtained from implementing the proposed algorithms in Chapter 4 and Chapter 5 and analyze these simulation results.
- Finally, in Chapter 7, we discuss the overall conclusion before laying down the groundwork for future studies.

1.4 Author Contributions

The contribution of the author of this thesis, Venkata Naga Sai Praneeth, includes reviewing related literature in Chapters 2,3, developing the proposed algorithms in Chapters 4,5, obtaining simulation results in Chapter 6, and finally providing a comprehensive conclusion in Chapter 7. Note these contributions were made under the guidance of the author's supervisor, Dr. Ioannis Psaromiligkos.

Chapter 2

Literature Review

In this chapter, firstly, we discuss the basic principles behind the operation of IRS. Moreover, we provide an intuitive discussion of the basic IRS system model used in the literature. Following this, we present the differences between IRSs and relays, a question that commonly arises when IRSs are discussed. Finally, we provide a comprehensive study of innovative signal processing techniques (Fundamental methods, Compressed sensing, Matrix Factorization/Decomposition, Machine learning, and Deep Learning) used in literature to estimate the channels of IRS-based communication systems.

2.1 Intelligent Reflective Surface

An IRS is a programmable flat surface consisting of several passive components, each of which can individually induce a change in the incident electromagnetic's amplitude or phase [1]. As mentioned by authors of [1], more elaborately, IRS is made up of planar arrays of 2D smart metasurfaces composed of several meta-atoms. These meta-atoms maintain an electrical thickness that is comparable to the subwavelength of the interest's operating frequency. By carefully considering its size, geometry, and orientation, an IRS can be tuned. Additionally, tunability may be attained by using a variety of components, including positive-intrinsic-

negative (PIN) diodes, field-effect transistors (FETs), or micro-electromechanical systems (MEMS) switches. The IRS's hardware architecture consists of three layers and a smart controller as shown in Figure 2.1. In essence, the outer layer is manufactured with the objective that it interacts with incident waves. The purpose of the middle layer is essentially to prevent signal energy leakage. Finally, the innermost layer features a control board that aids in configuring the IRS's tunability feature [1].

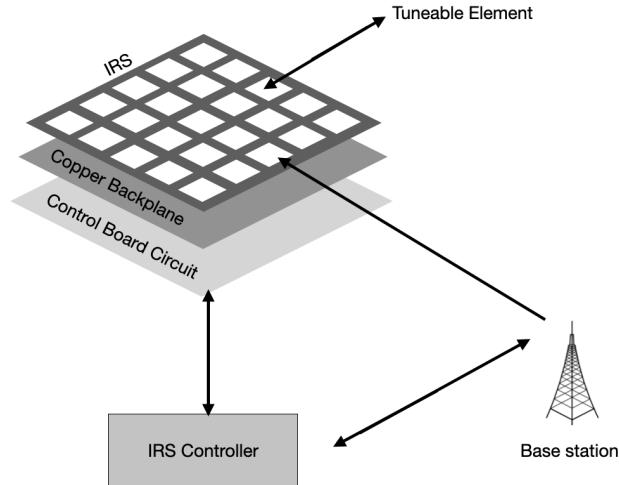


Figure 2.1: IRS architecture. Adapted from [1]

An anomalous reflection means that the incident angle θ_i from the transmitter to the surface is not equal to the reflected angle θ_r from the surface to the receiver [26]. An IRS's primary objective is to control the scattered field to direct the main beam towards the intended receiver's location. This implies that the incident wave's angle from the transmitter and the angle of observation directed to the receiver are different. Hence it is viable to state that IRS exploits anomalous reflection [27].

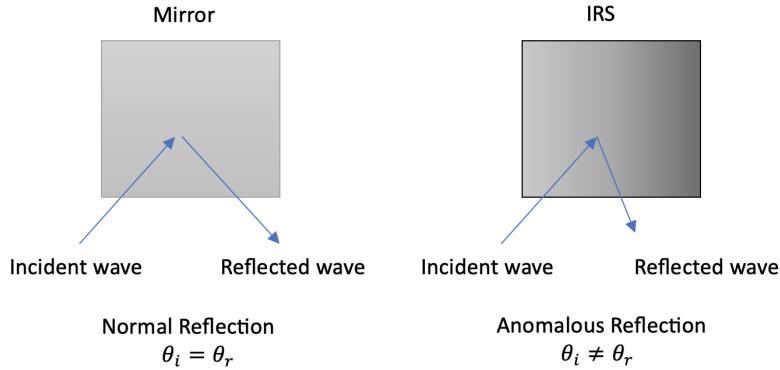


Figure 2.2: Anomalous Reflection on a continuously changing surface impedance vs normal mirror that follows Snell’s law (adapted from [2]).

The left image in Figure 2.2 depicts a mirror reflecting light in accordance with Snell’s law, with the incident angle θ_i equal to the reflected angle θ_r . The anomalous reflection scenario is shown in the right image, where the surface impedance continuously changes, and the incident angle θ_i is not equal to the reflected angle θ_r [2]. When the IRS’s reflecting components are carefully managed and optimized, the reflected electromagnetic waves are constructively added to enhance the signal that is being received [28]. By employing numerous IRSs that are carefully coordinated within the network, it may be possible to deal with channel fading and improve wireless communication capacity and reliability [29].

2.2 Basic System Model

This section presents a basic system model used in the literature to describe the operation of IRS-equipped wireless systems [29]. First, the single-user Single Input Single Output (SISO) case is considered for a narrow-band system. Figure 2.3 illustrates a point-to-point communication system involving an IRS with N passive reflective elements, a base station (BS), and a user. Due to high path loss, it can be assumed that the signal reflected by IRS more than once has negligible power. Let x be a complex-valued baseband transmit signal, $\beta_n \in [0, 1]$ and $\theta_n \in [0, 2\pi)$ be the amplitude attenuation and phase shift induced by IRS element n where $n \in \{0, 1, \dots, N\}$, respectively. The channels from the BS to the n th element of the IRS, from the n th element to the user, and from the BS to the user are given by $h_{BS,n}$, $h_{SU,n}$ and h_{BU} , respectively. Assuming all IRS elements reflect the incident signals without signal coupling. Then the received signal at the user from all the N IRS elements at a time $t \in \{1, \dots, T\}$ is given by

$$y_t = (\mathbf{h}_{BS}^H \Theta \mathbf{h}_{SU} + h_{BU})x + n, \quad (2.1)$$

where $\mathbf{h}_r^H = [h_{BS,1}, \dots, h_{BS,N}]$, $\Theta = \text{diag}(\beta_1 e^{j\theta_1}, \dots, \beta_N e^{j\theta_N})$, $\mathbf{H}_{SU} = [h_{SU,1}, \dots, h_{SU,N}]^T$, and $n \in \mathcal{CN}(0, \sigma^2)$ is additive white Gaussian noise.

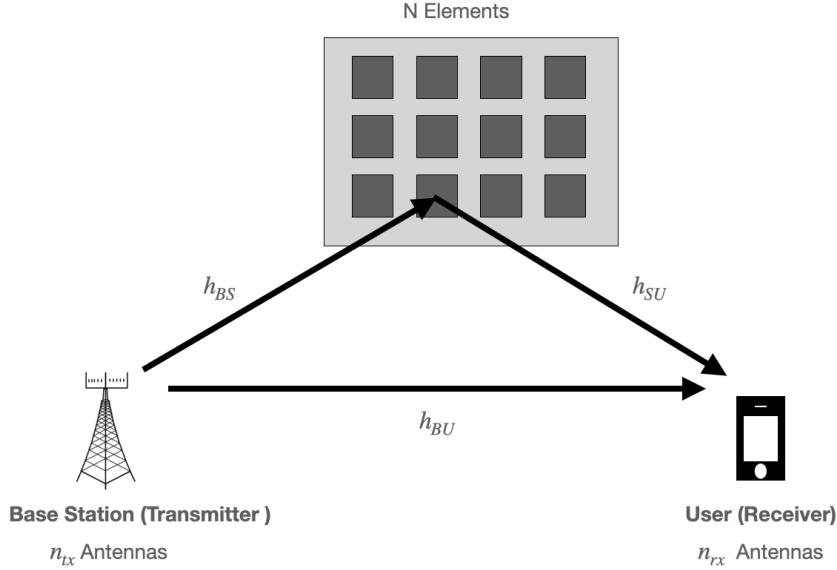


Figure 2.3: Illustration of SISO-IRS aided communication system

Let us extend the system model to the MIMO case. Consider the same setup as in Figure 2.3, but now the BS is equipped with M antennas, and the user is equipped with L antennas. Let the transmitted signal vector be $\mathbf{x} \in \mathbb{C}^{1 \times M}$. The channels from the BS to the IRS, from the IRS to the user and from the BS to the user are given by $\mathbf{H}_{BS}^H \in \mathbb{C}^{M \times N}$, $\mathbf{H}_{SU} \in \mathbb{C}^{N \times L}$, and $\mathbf{H}_{BU} \in \mathbb{C}^{M \times L}$, respectively. Then, the received signal from all the IRS elements at the user is given by

$$\mathbf{y} = (\mathbf{H}_{BS}^H \Theta \mathbf{H}_{SU} + \mathbf{H}_{BU})\mathbf{x} + \mathbf{n} \quad (2.2)$$

The single-user MIMO case can be further extended to the multi-user MIMO case. Assume there are K users. In this case, the baseband equivalent channels from BS to IRS, from IRS to user $k \in \{1, \dots, K\}$ and from the BS to user k are given by $\mathbf{H}_{BS,k}^H \in \mathbb{C}^{M \times N}$, $\mathbf{H}_{SU,k} \in \mathbb{C}^{N \times L}$, and $\mathbf{H}_{BU,k}^H \in \mathbb{C}^{M \times L}$, respectively. Let the transmitted data vector for user k be \mathbf{x}_k . Similar

to (2.2), the received signal from all the IRS elements at the user k is expressed as

$$\mathbf{y} = \sum_{k=1}^K (\mathbf{H}_{BS,k}^H \Theta \mathbf{H}_{SU,k} + \mathbf{H}_{BU,k}) \mathbf{x}_k + \mathbf{n}, \quad (2.3)$$

where \mathbf{n} in $\mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise.

2.3 IRS and Relays

Relays are a well-known technology that can potentially be used to turn a single non-line-of-sight (NLOS) link into multiple line-of-sight (LOS) ones. The two relay protocols that are most frequently used are Amplify-and-Forward (AF) and Decode-and-Forward (DF). The former increases the signal's strength and transmits it without decoding, while the latter decodes, processes, and then retransmits the signal [30]. Although relay-assisted communication and IRS-assisted communication share many similarities, they also differ significantly [31]. The advantage of the IRS system is that it forwards incident electromagnetic waves in a particular direction with minimal power consumption. Due to the minimal usage of RF chains, this system consumes very little energy, whereas AF relaying consumes more energy for a self-evident reason: the presence of an amplifier unit in AF relaying communication. In addition, IRS operates in full-duplex mode and thereby avoids self-interference, whereas AF relaying works in a half-duplex manner in practice, which is less spectrally efficient than full-duplex IRS communication. However, AF relaying can be implemented in full-duplex mode, but it is cost-inefficient as it requires self-interference cancellation techniques. Another difference the authors of [31] pointed out is that they noted that the IRS configured to follow anomalous reflection are not affected by the additive noise. In contrast, relays are affected by additive noise due to the presence of active electronic components. As a result, this noise component affects the performance of relays negatively. For example, AF relaying amplifies this noise component,

thus negatively affecting the system's performance.

Choosing an IRS or a relay completely depends on the system's specific requirements, as these two technologies have pros and cons. For instance, it was pointed out by the authors of [31, 32] that a larger IRS surface outperforms traditional relay-aided communication in terms of data rates with reduced complexity. However, the authors of [33] proved that a simple RF multi-antenna full-duplex relay could outperform IRS in terms of throughput under specific conditions. The authors of [34] proposed a new IRS-relay-aided communication system to show that the proposed system improves the data rates than relay-aided and IRS-aided communications individually. Moreover, the authors of [35] focused the channel estimation and pilot optimization for a two-way relaying IRS-assisted system. A particular pilot pattern was introduced to improve the performance and reduce the complexity of the system. The results showed that a certain phase shifter (one to four-bit) would be needed to achieve optimal MSE performance, depending on the phase matrix of IRS (Hadamard matrix or DFT matrix).

2.4 Channel estimation: Literature Review

The CSI is often presumed to be completely known to the transceiver; however, in reality, this assumption is not necessarily valid. Furthermore, because IRS is a passive technology, achieving accurate CSI for an IRS-aided communication system may be challenging. As a result, a substantial amount of the research concentrates on channel estimation. Multiple signal processing techniques, including the fundamental methods (MMSE-LS), compressed sensing, matrix factorization, and deep learning, have been used by researchers to solve the channel estimation problem [13].

2.4.1 Fundamental methods

One of the most fundamental and significant experiments was carried out by authors of [12], where they used the ON/OFF protocol for a MISO-IRS-assisted system to perform channel estimation. In essence, the ON/OFF protocol involves turning all IRS elements OFF while estimating the direct channel (BS to UE) and while estimating the cascaded channels (BS to IRS, IRS to UE) each of the IRS elements is turned ON one after the other [13]. Another protocol that was taken into consideration while estimating the channel was the complete reflection of IRS elements (all of the passive elements switched ON with maximum reflection amplitude) [14]. Moreover, the authors of [14] proposed a transmission protocol to perform channel estimation and reflection optimization for the SISO IRS-Orthogonal Frequency-Division Multiplexing (OFDM) communication system. To reduce the complexity of channel estimation when all the elements are ON, they grouped adjacent passive elements with high correlation into sub-surfaces. They designed a new IRS reflection pattern to perform channel estimation at BS based on the uplink pilot signals from the user. Based on the channel estimate, optimization of IRS reflection coefficients was performed to maximize the signal path in the time domain. The authors of [15] used the always-ON protocol (without turning OFF the selected IRS elements) to avoid the power reflection loss that results from using the "ON/OFF" protocol. Additionally, the pilot overhead for an uplink multiuser MISO-IRS-aided system was reduced by implementing the common-link structure, which essentially separates the cascaded channel coefficients.

In addition to the training protocols, handling the pilot power is quite important. As a result, the authors of [16] obtained the power allocation necessary for a single-user SISO IRS-assisted system to achieve an accurate channel estimation. Moreover, the authors analyzed the effect of channel estimation on the passive beamforming performance of the system. As a result, they observed that low-gain channels required higher power allocation with a specified given power, which increased the IRS's passive beamforming performance even

with fewer IRS elements. Another interesting research on imperfect CSI was performed using the conventional MMSE and ZF detectors [36]. The authors considered a multiuser massive MIMO-IRS-assisted system and proposed an MMSE-based method to estimate the cascaded and direct channels. Quite interestingly, it was assumed that the number of users is the minimum limit to choosing the length of pilots. Furthermore, the authors explored the uplink achievable rates for the proposed method and derived both upper and lower bounds for the rate. The analysis revealed that these bounds effectively mitigate challenges raised from IRS phase shift quantization errors. Finally, the authors of [36] proposed a low-complexity majorization-minimization-based algorithm to optimize the sum user rate and the minimum user rate. The study showed that maximizing the sum rate is achieved by placing the IRS closer to a random user than the BS.

Manufacturing an IRS with a large number of reflective elements with high-resolution phase shifts is generally costly, making continuous phase shifts challenging to use. Therefore the authors of [37] used discretized phase shifts. They designed the IRS elements (grouping and partitions) such that channel estimation for SISO single-user IRS-assisted system occurs progressively. Based on this estimate, they proposed the progressive passive beamforming of the IRS elements to improve the achievable rate. Another piece of research that demonstrates the significance and difficulties that arise when CSI knowledge is not known is discussed by the authors in [38]. Considering the pilot overhead that results from least-squares channel estimation, the authors of [38] proposed their research on the achievable rate of a SISO uplink single-user IRS-assisted system. In the case of assuming perfect CSI, the attainable rate rises monotonically as the number of IRS elements K increases [6]. However, the authors of [38] noted that monotonicity does not always hold while imperfect CSI is assumed. Therefore, they demonstrated that an optimal number of IRS elements K^* exists such that the achievable rate can be maximized by balancing IRS power gains and accounting for the channel estimation overhead. Furthermore, the authors provide analytical approximations for the optimal number of IRS components K^* and

variations of K^* with statistical characteristics of the channel based on the upper bound on the average achievable rate obtained. In [39], the authors approached the channel estimation problem differently. The channel estimation problem for the downlink SISO-IRS-assisted system was converted to an optimization problem. To solve this problem, they have used the Lagrange optimization technique and a dual ascent-based algorithm. The performance of the system was validated by deriving the Cramer-Rao Bound (CRB) and comparing it with the conventional LS method.

2.4.2 Compressed Sensing and Matrix Factorization / Decomposition

Using high frequencies in wireless communication, such as THz or mmWave, would guarantee high sparsity and low rank in the angular domain [13]. As a result, many researchers have employed compressed sensing techniques to deal with sparsity. The authors of [40] converted the channel estimation problem to sparse signal recovery for a single user downlink mmWave MISO IRS-assisted system, ignoring the direct link from BS to the user. Consequently, the pilot overhead is reduced by taking advantage of the channel matrix's sparse feature. They identified a sparse representation of cascaded channels using Katri-Rao and Kronecker properties. The cascaded channel was then estimated using compressed-sensing techniques like the orthogonal matching pursuit, a low-complexity algorithm. Instead of taking THz or mmWave into account, the authors of [41] chose large channel matrices for their uplink multi-user MISO IRS-assisted system. As a result, they have used matrix decompositions like the PARAllel FACtor (PARAFAC) decomposition, which effectively divides a high-dimensional matrix into several rank-one matrices, to estimate the channels. Applying the PARAFAC, they were able to create different forms for the unknown channels. They used two iterative channel estimating methodologies to estimate the newly obtained forms, (i) alternating least squares (ALS) and (ii) vector

approximate message passing (VAMP). To validate the performance of the ALS-based channel estimates, the CRB was computed. Further along these lines, another intriguing research models the single-user MIMO IRS-assisted system to a keyhole MIMO system model [42]. With clutters surrounding the transmitter and receiver, the channel rank can reduce to one, resulting in a keyhole MIMO channel where most signal energy passes through small holes. They separated cascaded channels and estimated them independently using eigenvalue decomposition (EVD) techniques to estimate these newly modeled channels. A numerical analysis of the proposed algorithm revealed that it had a low estimation time overhead and a lower estimation error.

2.4.3 Machine learning and Deep Learning

The advancement of technology has led to the use of machine learning or deep learning (DL) algorithms for channel estimations [14,39]. The DL network is provided with various channel characteristics for training, thus enabling the system to adapt to changes in the environment (eg. user movement). The authors of [43] proposed a twin convolution neural network (CNN) trained on the pilot symbols to estimate the direct and cascaded channels. In their studies, the authors assumed that each user could access a CNN network for estimating their channel in a multi-user mmWave massive MISO-IRS aided system. Based on the Saleh-Valenzuela models, the authors compared the performance gains of the proposed scheme to state-of-the-art deep learning-based techniques using numerical simulations. As the proposed method in [43] uses sequential twin CNNs to estimate the channels, it is possible that the proposed method could suffer from an error propagation issue. Therefore, the authors of [44] have proposed a single CNN structure to estimate the direct and cascaded channels independently by avoiding error propagation. In this study, the authors focused on downlink MISO-OFDM-aided systems and relied on the fact, described in [14], that cascaded and direct channels can be unfolded by selecting the proper reflection coefficients. Accordingly, they estimated

the channels first using the CNN model and then calculated the individual channels using those estimated channels.

As discussed earlier, THz waves bring sparsity into the picture. The authors of [45] tackle the sparsity issue with deep learning concepts. A downlink THz IRS -MIMO communication system was considered, and the channel estimation problem was formulated as the sparse recovery problem. To estimate and recover the channels, a two-stage neural network was developed. In particular, for the first stage, the neural network replicates the noisy linear measurement process and presents the relationship between the received pilot signal, the sensing matrix, and the sparse signal. In the second stage, they reconstructed the desired channels using the sparse recovery function obtained from the proposed network. In another intriguing study focused on the reduction of pilot overhead for an uplink single-user MIMO-IRS-assisted system [46], the authors proposed a channel estimation framework involving three stages. The three-stage model was a synthetic deep neural network in which the direct channel was estimated when IRS elements were partly turned ON. The cascaded channels were estimated for the IRS elements that are turned ON. Finally, the cascaded channels for the IRS elements that are not turned ON were predicted. In addition, the designed model addresses the issue of error propagation without adversely affecting the system. Ultimately, the proposed model improved performance without a high pilot overhead and exact knowledge of channel statistics.

Along the lines of reduction of channel estimation overhead, the authors of [47] proposed a channel prediction method. An interesting aspect of their research is that they compared the channel estimation to image super-resolution in computer vision as IRS elements are normally arranged in two dimensions. The authors chose to perform channel estimation for only a subset of the IRS elements and predicted the channels for the remaining elements. It was demonstrated that accurate control of reflection coefficients on IRS can be maintained with reduced overhead. The semi-definite relaxation technique was applied to optimize the reflection coefficients, and for channel prediction, a new end-to-end deep-learning model was

designed. By comparing their method with grouping-based methods [14], [6], they showed that the proposed method provides improved channel estimation for instantaneous rate. Additionally, the trade-off study of channel estimation overhead versus channel prediction found that increasing elemental size was not always efficient. A different approach is used in another research [48] aimed at reducing channel estimation overhead. They proposed a long-term CSI based transmission design for a downlink multi-user MISO IRS-assisted system. The long-term CSI means that the variation of CSI is very minimal over time and in a relative reference to instantaneous CSI, the long-term CSI changes slowly. It is difficult to obtain the exact expression for multi-user ergodic data rate; therefore, they proposed a novel deep deterministic policy gradient (DDPG) algorithm to solve the optimization problem. So, they estimate the long-term CSI and obtain a set of offline instantaneous CSI samples, based on which they train the DDPG network. Along with reducing the computational complexity, their algorithm outperformed the conventional instantaneous-CSI-based scheme regarding net throughput.

Moreover, the authors of [49] proposed a joint channel estimation scheme that was based on multi-task learning (MTL) for mmWave multi-user MISO-IRS aided communication systems to reduce the pilot overhead. With the help of shared pilot signals, the direct and cascaded channels were jointly estimated simultaneously. The proposed MTL network was divided into top and bottom layers, with the top layers focusing on direct and cascaded channel estimation tasks and the bottom layers attempting to suppress the noise effect using deep residual shrinkage network architectures. Compared to the traditional channel estimation systems, simulation results demonstrated that the suggested method provides improved estimation accuracy with minimal pilot overhead.

2.4.4 Semi-blind channel estimation

Until now, the research that has been discussed has been related to training or blind channel estimation. However, let us now dive into semi-blind channel estimation. To our knowledge, there have only been a handful of studies that have used semi-blind channel estimation methods for MIMO-IRS-aided systems at this time [17–22]. The most recent work focuses on semi-blind channel estimation for a massive MIMO-IRS system proposed by authors of [19]. They applied the approximate message passing framework by converting the channel estimation problem to a trilinear inference problem and computed the MMSE iteratively. A joint estimation of the channel and the partially unknown transmit signal (as part of the transmit signal is considered the pilot signal) was performed. In the end, their model provided accurate estimates with reduced pilot overhead. Another interesting research [18] focuses on semi-blind channel estimation using conventional methods for a massive multi-user MISO-IRS-aided communication system. The authors proposed a multi-stage channel estimation method, where the first stage is to estimate the direct channel with pilot symbols, followed by the data-aided estimation for the cascaded channel. More precisely, the data-aided estimation is as follows, for stage-1 only the first element of IRS (where the IRS is assumed to have a total of N elements) is turned ON, and the cascaded channel is estimated. In the case of the stage- N , only the first and N th element of IRS is turned ON while the rest are all turned OFF. Finally, stage- $(N + 1)$ is where all the elements of IRS are turned ON, and the channel estimated is performed along with data detection for all users. By using this method, it was shown that the achievable sum rate performed better than the state-of-the-art algorithms. Furthermore, the outage probability for the proposed system was discussed.

So far, the semi-blind channel estimation was dealt only with massive MIMO. However, authors of [21] proposed a joint symbol detection and channel estimation for a multiuser mmWave MIMO-IRS aided system. To obtain the full knowledge of CSIs and information

symbols at UEs without a training phase, they proposed the PARATUCK2 model along with the use of hybrid Kronecker or Khatri-Rao factorization and low-rank property of mmWave channels. The results showed that the channel estimation and symbol detection error is minimal compared to existing methods. The authors of [20] provided a very comprehensive study of the semi-blind joint channel and symbol estimation for a MIMO-IRS-assisted system. They considered a direct link between BS and the user, unlike the case presented by authors of [17], where they ignored the direct link. Using the trilinear alternating least squares estimation method that utilizes the PARATUCK tensor model, they estimated the direct and cascaded channels iteratively. Interestingly, they discussed the joint detection of channels and symbols and designed a coding matrix for IRS elements along with a phase shift matrix to improve the performance of the receiver. A unique discussion regarding the impact of residual reflection on the performance of the proposed system was discussed. Finally, they obtained the analytical expressions for CRB. Another research that the first author of [22] conducted focused on joint channel and symbol detection for multiuser MIMO-IRS systems. This research ignored the direct link and dealt with the cascaded channel alone. Although the authors stuck to the PARATUCK tensor model of the reflected signals, they proposed a two-stage closed-form receiver using Kronecker and Khatri-Rao factorization. The proposed method was shown to perform better than the pilot-based Khatri-Rao and Kronecker factorization methods. The final research that exists that uses semi-blind channel estimation for the MISO system is presented by authors of [17] and a detailed discussion of this research paper will be provided in Chapter 3.

2.5 Summary

A detailed review of the IRS in terms of its working, architecture, and difference over relays was discussed in this chapter. Firstly, a basic system model for IRS-aided communication

often used in the literature was explored. In addition, a review of the literature on channel estimation from a signal-processing perspective was given. The literature review that was discussed had three major parts,

1. Fundamental methods - LS, MMSE, and other traditional estimation methods were used for channel estimation with different protocols
2. Compressed Sensing and Matrix Factorization/Decomposition - PARAFAC, VAMP, EVD, etc were used for channel estimation
3. Machine learning and Deep learning - Majorly CNNs and variations of CNNs were used for channel estimation

Lastly, a discussion of the semi-blind approach to channel estimation using signal processing methods was presented. The fundamental ideas and background necessary to comprehend this thesis will be covered in more detail in the next chapter.

Chapter 3

Background

In this chapter, we discuss the essential topics required to comprehend this thesis. We initially provide a quick overview of maximum likelihood estimation before concentrating on EM algorithm, which forms the core of this thesis. The description of Sections 3.1 and 3.2 are referenced from the Ph.D. thesis [25]. In addition, we conduct an in-depth study of [17] and their methodology to perform semi-blind channel estimation for MISO-IRS aided system. In the final section of this chapter, we outline the mathematical operations such as the Kronecker product, Khatri-Rao product, and Vectorization operator and their related properties used in this thesis.

3.1 Maximum likelihood estimation

Maximum likelihood (ML) estimation is a channel estimation technique most commonly used in the literature. The statistical distribution is often used to describe the observations in estimation problems. Let us consider an example with y_1, \dots, y_N observations and each of the observations depends on the symbols s_1, \dots, s_N , that is chosen from a set $S = \{\eta_1, \dots, \eta_M\}$. The observations are parameterized with respect to parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l]^T$. The ML

estimate of $\boldsymbol{\theta}$ given the observations y_1, \dots, y_N is

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} f(y_1, \dots, y_N; \boldsymbol{\theta}) \quad (3.1)$$

where $f(y_1, \dots, y_N; \boldsymbol{\theta})$ is the likelihood function. The likelihood function for a particular observation y_i is given as

$$f(y_i; \boldsymbol{\theta}) = \sum_{j=1}^M f(y_i | s_i = \eta_j, \boldsymbol{\theta}) P(s_j = \eta_j). \quad (3.2)$$

For N independent observations, the joint likelihood function will be

$$f(y_1, \dots, y_N; \boldsymbol{\theta}) = \prod_{i=1}^N \left(\sum_{j=1}^M f(y_i | s_i = \eta_j, \boldsymbol{\theta}) P(s_j = \eta_j) \right). \quad (3.3)$$

It is common for (3.2) to have a simple form, but when more observations are taken into account, the likelihood function becomes quite difficult to calculate, as shown in (3.3). Consequently, many applications employ EM algorithm instead of calculating the closed-form solution in (3.3) due to the computational complexity involved.

3.2 Expectation Maximization

An Expectation Maximization (EM) algorithm is an iterative approach that emphasizes the determination of local maximum likelihood estimates of parameters. Two iterative steps comprise the EM algorithm, the E-step, and the M-step. Let us consider the same example as in section 3.1. Let the unknown data symbols or the latent data be $\mathbf{s} = [s_1, \dots, s_N]^T$, the observation vector $\mathbf{y} = [y_1, \dots, y_N]^T$ be the incomplete data, and the $\{\mathbf{y}, \mathbf{s}\}$ represents complete data. Consider the t^{th} step of EM iteration, the E-step calculates the expectation of the Log-Likelihood Function (LLF) of the complete data set with respect to the conditional

PMF $f(\mathbf{s}|\mathbf{y}; \boldsymbol{\theta}^{(t)})$ of the hidden data, \mathbf{s} , given the observations and the current estimate $\boldsymbol{\theta}^{(t)}$ of $\boldsymbol{\theta}$. The E-step of t^{th} iteration is given as

$$\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = \mathbb{E}\{\mathcal{L}(\mathbf{y}, \mathbf{s}; \boldsymbol{\theta}) | \mathbf{y}; \boldsymbol{\theta}^{(t)}\} \quad (3.4)$$

The next step called the M-step, is to maximize the expected value obtained in the E-step with respect to $\boldsymbol{\theta}$ to get an estimate of $\boldsymbol{\theta}^{(t)}$ as shown in (3.5)

$$\boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}). \quad (3.5)$$

The algorithm iterates until the parameters converge to the local maximum. As the algorithm is sensitive to initialization, extra care should be taken when selecting the initial value (value at the first iteration) for the parameter.

3.3 Expectation Maximization with Gaussian Prior

This section discusses [17] in detail as the authors have provided a semi-blind channel estimation for the IRS-assisted MISO system. The authors assumed a Gaussian prior for the data symbols and formulated a maximum likelihood problem for cascaded channels, ignoring direct channels. Nevertheless, due to challenges associated with maximum likelihood estimation, cascaded channels were estimated iteratively using the EM algorithm. The channels were invariant, which means they do not alter between transmission blocks. The following was the uplink system model that was considered for the MISO system: N passive IRS elements, the BS is equipped with M antennas, single antenna user, a total of $L = L_d + L_p$ symbols where L_d data symbols and L_p pilot symbols were used. The received signal at BS was given as,

$$\mathbf{y}_l = \mathbf{G}^H \boldsymbol{\phi}_l s_l + \mathbf{z}_l$$

where $l \in 1, \dots, L$, $\mathbf{G} = \text{diag}(\mathbf{h}^H)\mathbf{F}$, \mathbf{h} represents the channel between user to IRS, \mathbf{F}^H represents the channel between IRS and user, with AWGN noise at each BS antenna and $\boldsymbol{\phi}_l$ represents the phase shift vector of IRS at time l . Therefore, the pilot received signal is given by $\mathbf{Y}_p = [\mathbf{y}_1, \dots, \mathbf{y}_p]$ and the subsequent data received signal is represented as $\mathbf{Y}_d = [\mathbf{y}_{p+1}, \dots, \mathbf{y}_L]$. The transmitted pilot symbols are defined by $\mathbf{s}_p \in \mathbb{C}^{L_p \times 1}$, and the phase shift matrix during the pilot transmission is given by $\Phi_p \in \mathbb{C}^{N \times L_p}$. The ML estimate has been formulated using this system model. However, as discussed in Section 3.2, the ML estimate involves high dimensional integration with respect to unknown data symbols at the base station. This results in difficulties in finding a closed-form solution. Therefore, EM algorithm has been proposed as a solution in response. Following the traditional method of matching the columns of IRS elements with the columns of *pilot* \times *pilot* point DFT matrix, they presented the initial estimate $i = 0$ of the cascaded channel as

$$\hat{\mathbf{G}}^{(0)} = (\mathbf{Y}_p(\Phi_p \text{diag}(\mathbf{s}_p))^\dagger)^H$$

It was assumed that the number of pilots should be smaller than the product of the number of IRS elements and the number of transmit antennas. However, the number of pilots should not be less than the number of IRS elements since this makes it very difficult to implement the channel estimation methods proposed so far for IRS-assisted systems. After carefully applying the EM algorithm with the Gaussian priori assumption, the data symbols and the final estimate of the cascaded channel are given in the following equation.

$$\hat{\mathbf{G}}^{(i+1)} = [\mathbf{g}_{[1]}^{(i+1)*}, \dots, \mathbf{g}_{[M]}^{(i+1)*}].$$

The term $(\mathbf{g}_{[m]})^T$ denotes the m th row of $\mathbf{G}^H \in \mathbb{C}^{M \times N}$ and is given by

$$(\mathbf{g}_{[m]}^{(i+1)}) = \left(\sum_{l=1}^{L_p} ([\mathbf{y}_l]_m s_l^* \boldsymbol{\phi}_l^H) + \sum_{l=L_p+1}^L [\mathbf{y}_l]_m \mathbb{E}\{s_l | \mathbf{Y}, \hat{\mathbf{G}}^{(i)}\}^* \boldsymbol{\phi}_l^H \right)$$

$$\times \left(\sum_{l=1}^{L_p} (\phi_l |s_l|^2 \phi_l^H) + \sum_{l=L_p+1}^L \phi_l \mathbb{E}\{|s_l|^2 |Y, \hat{G}^{(i)}\}^* \phi_l^H \right)^{-1}. \quad (3.6)$$

The mean and the conditional probability $\mathbb{E}\{s_d^2 | Y, \hat{G}^{(i)}\}$ in (3.6) are described in the (3.7) and (3.8), respectively.

$$\begin{aligned} \mathbb{E}\{s_d | Y, \hat{G}^{(i)}\} &= \mathbb{E}\{s_d \mathbf{y}_d^H | \hat{G}^{(i)}\} \mathbb{E}\{\mathbf{y}_d \mathbf{y}_d^H | \hat{G}^{(i)}\}^{-1} \mathbf{y}_d \\ &= \mathbb{E}\{s_d (\mathbf{G}^H \phi_d s_d + \mathbf{z}_d)^H | \hat{G}^{(i)}\} \mathbb{E}\{(\mathbf{G}^H \phi_d s_d + \mathbf{z}_d) (\mathbf{G}^H \phi_d s_d + \mathbf{z}_d)^H | \hat{G}^{(i)}\}^{-1} \mathbf{y}_d \\ &= \mathbb{E}\{s_d (\phi_d^H \mathbf{G} s_d + \mathbf{z}_d^H) | \hat{G}^{(i)}\} \mathbb{E}\{(\mathbf{G}^H \phi_d s_d + \mathbf{z}_d) (\mathbf{G}^H \phi_d s_d + \mathbf{z}_d)^H | \hat{G}^{(i)}\}^{-1} \mathbf{y}_d \\ &= (\mathbb{E}\{\phi_d^H \mathbf{G} s_d^2 | \hat{G}^{(i)}\} + \mathbb{E}\{s_d \mathbf{z}_d^H | \hat{G}^{(i)}\}) \\ &\quad \times \mathbb{E}\{(\mathbf{G}^H \phi_d s_d + \mathbf{z}_d) (\mathbf{G}^H \phi_d s_d + \mathbf{z}_d)^H | \hat{G}^{(i)}\}^{-1} \mathbf{y}_d \\ &= (\phi_d^H \mathbb{E}\{\mathbf{G} s_d^2 | \hat{G}^{(i)}\} + \mathbb{E}\{\mathbf{z}_d^H\} \mathbb{E}\{s_d | \hat{G}^{(i)}\}) \\ &\quad \times \mathbb{E}\{(\mathbf{G}^H \phi_d s_d + \mathbf{z}_d) (\mathbf{G}^H \phi_d s_d + \mathbf{z}_d)^H | \hat{G}^{(i)}\}^{-1} \mathbf{y}_d \\ &= P \phi_d^H \hat{G}^{(i)} \mathbb{E}\{(\mathbf{G}^H \phi_d \phi_d^H \mathbf{G} s_d^2 + \mathbf{z}_d \phi_d^H \mathbf{G} s_d + \mathbf{G}^H \phi_d \mathbf{z}_d^H s_d + \mathbf{z}_d \mathbf{z}_d^H) | \hat{G}^{(i)}\}^{-1} \mathbf{y}_d \\ &= P \phi_d^H \hat{G}^{(i)} (\mathbb{E}\{\mathbf{G}^H \phi_d \phi_d^H \mathbf{G} s_d^2 | \hat{G}^{(i)}\} + \mathbb{E}\{\mathbf{z}_d \mathbf{z}_d^H | \hat{G}^{(i)}\})^{-1} \mathbf{y}_d \\ &= P \phi_d^H \hat{G}^{(i)} (\mathbb{E}\{\mathbf{G}^H \phi_d \phi_d^H \mathbf{G} s_d^2 | \hat{G}^{(i)}\} + \mathbb{E}\{\mathbf{z}_d \mathbf{z}_d^H\})^{-1} \mathbf{y}_d \\ &= P \phi_d^H (\hat{G}^{(i)}) (P(\hat{G}^{(i)})^H \phi_d \phi_d^H (\hat{G}^{(i)}) + \sigma^2 \mathbb{I}_M)^{-1} \mathbf{y}_d, \end{aligned} \quad (3.7)$$

where $\mathbb{E}\{\mathbf{z}_d^H\} = 0$ (as it is the mean of the noise).

$$\mathbb{E}\{|s_d|^2 | Y, \hat{G}^{(i)}\} = \sigma^2 (\phi_l^H \hat{G}^{(i)} (\hat{G}^{(i)})^H \phi_l + \sigma^2)^{-1} + (\mathbb{E}\{s_d | Y, \hat{G}^{(i)}\})^2 \quad (3.8)$$

Simulations demonstrate that the proposed algorithm enhanced channel estimation accuracy at the expense of increased calculation complexity while requiring fewer training pilots. As the benchmark methods, the ON/OFF method [12] and DFT based [50] were chosen. They primarily investigated the effect of pilots and data symbols on the accuracy of the channel

estimate individually. The proposed method gave superior accuracy in the CSI estimation than the benchmark models, even when using fewer pilot symbols than IRS elements. Prior to achieving saturation, it was shown that the accuracy was still superior to other approaches even when the number of pilot symbols exceeded the number of IRS elements. An increase in data symbols indicated improved NMSE performance. Additionally, it was demonstrated that the proposed method performs better and improves spectral efficiency. However, the main limitation of this algorithm is that the data symbols are assumed to be Gaussian distributed, but this is not the case in reality. Therefore, to overcome this limitation, this thesis provides an extensive study about choosing discrete data symbols and methods to reduce the complexity involved. The simulation results (Chapter 6) in this thesis provides a detailed comparison study between the proposed method and the method proposed by authors of [17].

3.4 Kronecker Product, Khatri Rao Product, and Vectorization

We frequently use Kronecker Products, Khatri Rao Products, and Vectorization in this thesis. Therefore, we briefly discuss these operations and the properties we will use later in the thesis. The Kronecker product of two matrices \mathbf{A} , \mathbf{B} with dimensions $m \times n$, $p \times q$, respectively, is denoted by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \cdots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix},$$

that returns a $mp \times nq$ matrix. The Khatri Rao product is often referred to as the column-wise Kroncker product. The definition of this product is as the following.

$$\mathbf{A} \diamond \mathbf{B} = \left[\mathbf{a}_1 \otimes \mathbf{b}_1, \dots, \mathbf{a}_n \otimes \mathbf{b}_n \right],$$

where dimensions of \mathbf{A} and \mathbf{B} is $m \times n$ and $p \times n$ respectively and the dimensions of the final product would be $mp \times nq$ matrix. Note, \mathbf{a}_i and \mathbf{b}_i are the column vectors of matrices \mathbf{A} and \mathbf{B} with $i = 1, \dots, n$. Finally, vectorization is an operation that rearranges a matrix to a vector form. Basically, every column of a $m \times n$ matrix \mathbf{A} is stacked vertically as defined below,

$$\text{vec } \mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix},$$

where \mathbf{a}_i represents the column vector of \mathbf{A} with $i = 1, \dots, n$. The above-defined operations have multiple applications in the field of wireless communication [51], statistics [52], and solving linear matrix equations [52]. Let us look at the related properties we will use in the thesis.

1. Assume \mathbf{A} , \mathbf{B} and, \mathbf{C} are matrices with dimensions $p \times n$, $n \times n$ and, $n \times m$ respectively.

Then,

$$\text{vec } \mathbf{ABC} = (\mathbf{C}^T \diamond \mathbf{A}) \text{diag}(\mathbf{B}),$$

where diag returns the main diagonal of \mathbf{B} .

For the following few properties the matrices are defined as $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times p}$, and $\mathbf{C} \in \mathbb{C}^{p \times s}$.

2. $\text{vec } \mathbf{ABC} = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec } \mathbf{B}$

3. $(\mathbf{A} \otimes \mathbf{B}) \text{vec } \mathbf{C} = \text{vec } \mathbf{BCA}^T$
4. $\text{vec } \mathbf{AB} = (\mathbf{B}^T \otimes \mathbf{I}_k) \text{vec } \mathbf{A}$
5. $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$
6. $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$
7. If $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times p}$, $\mathbf{C} \in \mathbb{C}^{p \times s}$, and $\mathbf{D} \in \mathbb{C}^{s \times r}$ then, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$
8. If $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{p \times q}$, and $\mathbf{C} \in \mathbb{C}^{r \times s}$.then $\text{vec}(\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}) = (\mathbf{I}_n \otimes \mathbf{K}^{(m,q,s)} \otimes \mathbf{I}_{pr})(\mathbf{I}_{mnp} \otimes \mathbf{K}^{(p,s)} \otimes \mathbf{I}_r)(\text{vec } \mathbf{A} \otimes \text{vec } \mathbf{B} \otimes \text{vec } \mathbf{C})$ [53]. Note, The commutation matrix is given as $\mathbf{K}^{(p,r)} = \begin{bmatrix} \mathbf{K}^{(1,1)} & \dots & \mathbf{K}^{(1,r)} \\ \vdots & \ddots & \vdots \\ \mathbf{K}^{(p,1)} & \dots & \mathbf{K}^{(p,r)} \end{bmatrix}$, where the m,n entry of $p \times r$ matrix \mathbf{K}_{ij} is given by $\mathbf{K}_{ij}(m, n) = 1$ when $i = m$ and $j = n$, otherwise $\mathbf{K}_{ij}(m, n) = 0$
9. If \mathbf{a} is a column vector then $\text{vec}(\mathbf{a}^T) = \mathbf{a}$
10. $(\mathbf{A}_1 \otimes \dots \otimes \mathbf{A}_n)(\mathbf{B}_1 \otimes \dots \otimes \mathbf{B}_n) = \mathbf{A}_1 \mathbf{B}_1 \otimes \dots \otimes \mathbf{A}_n \mathbf{B}_n$ [54] (with appropriate dimensions of the matrices)

3.5 Summary

In this chapter we reviewed the Expectation Maximization algorithm. Following this, we discussed [17] research which focuses on continuous data, whereas the proposed method utilizes discrete data such as QAM data. Finally, we defined the mathematical operations used predominately in this thesis, such as the Kronecker product, Khatri-Rao product, and Vectorization. We provided the properties that are utilized in the subsequent chapters. In the following chapter, we define the problem statement, use the EM algorithm to optimize,

and talk about a non-superimposed protocol. Moreover, we propose approximation methods in motivation to lessen the computational complexity involved.

Chapter 4

EM-based Channel Estimation and Detection: Non-Superimposed Method

Following the introduction to the various background topics in Chapter 3, we introduce the system model for the MIMO-IRS-aided communication system in section 4.1. In this chapter, We adopt the non-superimposed symbol protocol, which is the traditional data-pilot structure protocol that is being used in the literature [55–58]. This protocol transmits the pilots first, followed by data symbols. We perform semi-blind channel estimation based on the EM algorithm for non-superimposed symbol protocol. To further focus on computational complexity reduction, we propose various methods such as log-max approximation and using detectors for detecting the symbol vectors.

4.1 System model

Consider a single-user IRS-assisted half-duplex MIMO communication system that operates over flat-fading channels. Figure 4.1 illustrates a downlink system involving an IRS with N passive reflective elements, a Base Station (BS) with n_{tx} antennas, and a single user with n_{rx} antennas. A block of $n_{tx} \times T_p$ pilot symbols followed by a block of $n_{tx} \times T_d$ data symbols are transmitted, for a total of $n_{tx} \times T$ symbols, with $T = T_p + T_d$ as in Figure 4.2.

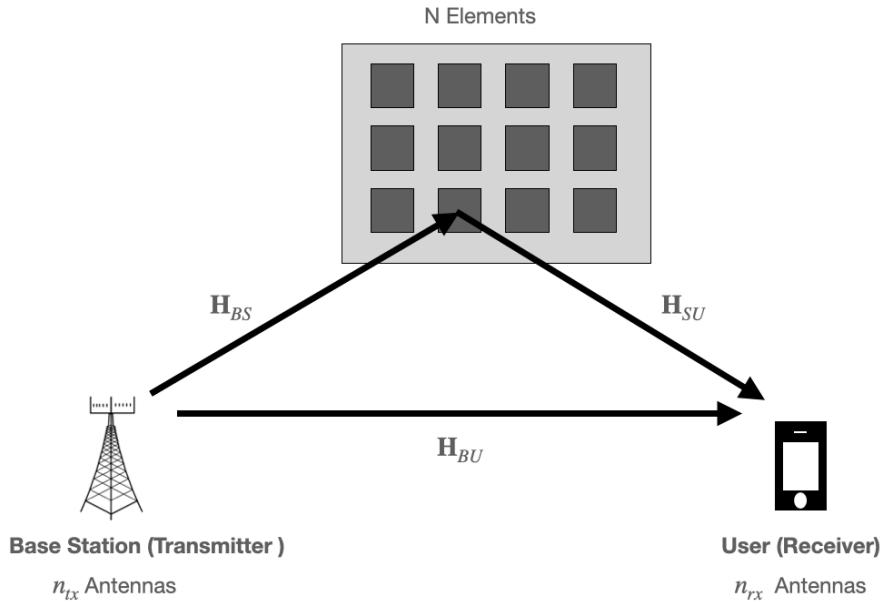


Figure 4.1: System model for MIMO-IRS aided communication system

The data symbols are randomly selected from a constellation $\mathcal{X} = \{\xi_1, \xi_2, \dots, \xi_M\}$ of size M . The transmitted pilot signal vector at time $t \in \{1, 2, \dots, T_p\}$ is represented as $\mathbf{x}_{p,t} \in \mathbb{C}^{n_{tx}}$. Similarly, the transmitted data signal vector at time $t \in \{T_p + 1, \dots, T\}$ is represented as $\mathbf{x}_{d,t} \in \mathbb{C}^{n_{tx}}$. Let $\mathbf{s}_p = [\mathbf{x}_{p,1}^T, \dots, \mathbf{x}_{p,T_p}^T]^T \in \mathbb{C}^{n_{tx}T_p \times 1}$ and $\mathbf{s}_d = [\mathbf{x}_{d,T_p+1}^T, \dots, \mathbf{x}_{d,T}^T]^T \in \mathbb{C}^{n_{tx}T_d \times 1}$. Moreover, the average transmitted power of BS during the pilot transmission is P , i.e., $\mathbf{s}_p^H \mathbf{s}_p = T_p P$. Similarly, the power used during the data transmission is $\mathbb{E}\{\mathbf{s}_d^H \mathbf{s}_d\} = T_d P$. The channels from BS to IRS, from IRS to the user, and the direct channel are denoted as $\mathbf{H}_{BS} \in$

$\mathbb{C}^{N \times n_{tx}}$, $\mathbf{H}_{SU} \in \mathbb{C}^{n_{rx} \times N}$, and $\mathbf{H}_{BU} \in \mathbb{C}^{n_{rx} \times n_{tx}}$, respectively. Collectively, the IRS elements are represented as $\Phi_t = \text{diag}(\boldsymbol{\psi}_t) \in \mathbb{C}^{N \times N}$, where $\boldsymbol{\psi}_t = [\beta_{1,t} e^{j\alpha_{1,t}}, \dots, \beta_{N,t} e^{j\alpha_{N,t}}]^T \in \mathbb{C}^N$, β_n and α_n denote the amplitude attenuation and phase shift induced by the n th IRS element, $n = \{0, 1, 2, \dots, N\}$.

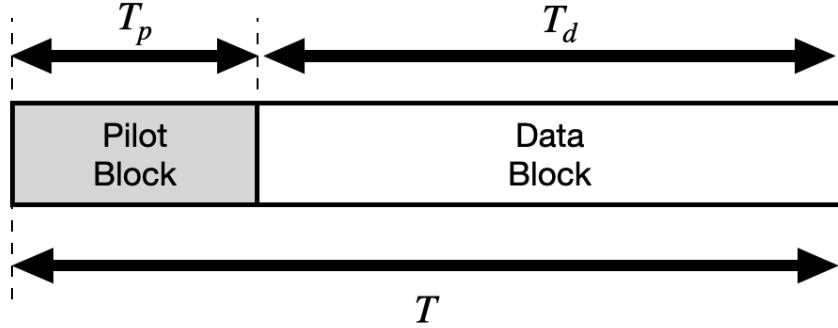


Figure 4.2: Non-Superimposed protocol for time period T

The received signals by the user at time t are given by

$$\mathbf{y}_{p,t} = (\mathbf{H}_{BU} + \mathbf{H}_{SU}\Phi_t\mathbf{H}_{BS})\mathbf{x}_{p,t} + \mathbf{n}_{p,t}, \quad \text{and} \quad (4.1)$$

$$\mathbf{y}_{d,t} = (\mathbf{H}_{BU} + \mathbf{H}_{SU}\Phi_t\mathbf{H}_{BS})\mathbf{x}_{d,t} + \mathbf{n}_{d,t}, \quad (4.2)$$

where the noise vectors $\mathbf{n}_{p,t}$ and $\mathbf{n}_{d,t}$ are assumed to be Gaussian with mean zero and covariance $\sigma^2 \mathbf{I}_{n_{rx}}$. Using the properties of the Khatri-Rao product and Kronecker product, we obtain

$$\begin{aligned} \mathbf{y}_{p,t} &= \text{vec}(\mathbf{I}_{n_{rx}}(\mathbf{H}_{BU} + \mathbf{H}_{SU}\Phi_t\mathbf{H}_{BS})\mathbf{x}_{p,t}) + \text{vec}(\mathbf{n}_{p,t}) \\ &= (\mathbf{x}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}) \text{vec}(\mathbf{H}_{BU} + \mathbf{H}_{SU}\Phi_t\mathbf{H}_{BS}) + \mathbf{n}_{p,t} \end{aligned}$$

4. EM-based Channel Estimation and Detection: Non-Superimposed Method34

$$\begin{aligned}
&\stackrel{(a)}{=} (\mathbf{x}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}) \begin{bmatrix} \text{vec}(\mathbf{H}_{BU}) & \text{vec}(\mathbf{H}_{BS}^T \diamond \mathbf{H}_{SU}) \end{bmatrix} \begin{bmatrix} 1 \\ \psi_t \end{bmatrix} + \mathbf{n}_{p,t} \\
&\stackrel{(b)}{=} (\tilde{\psi}_t^T \otimes \mathbf{x}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}) \begin{bmatrix} \text{vec}(\mathbf{H}_{BU}) \\ \text{vec}(\mathbf{H}_{BS}^T \diamond \mathbf{H}_{SU}) \end{bmatrix} + \mathbf{n}_{p,t} \\
&= \mathbf{Z}_{p,t} \mathbf{h} + \mathbf{n}_{p,t}, \tag{4.3}
\end{aligned}$$

where (a) and (b) follow Property [1] and Property [2] respectively. Moreover, $\tilde{\psi}_t = \begin{bmatrix} 1 \\ \psi_t \end{bmatrix} \in \mathbb{C}^{(N+1)}$, $\mathbf{h} = \begin{bmatrix} \text{vec}(\mathbf{H}_{BU}) \\ \text{vec}(\mathbf{H}_{BS}^T \diamond \mathbf{H}_{SU}) \end{bmatrix} \in \mathbb{C}^{n_{rx}n_{tx}(N+1) \times 1}$, and

$$\mathbf{Z}_{p,t} = \mathbf{z}_{p,t}^T \otimes \mathbf{I}_{n_{rx}} \in \mathbb{C}^{n_{rx} \times n_{rx}n_{tx}(N+1)} \tag{4.4}$$

with

$$\mathbf{z}_{p,t} = \tilde{\psi}_t \otimes \mathbf{x}_{p,t} \tag{4.5}$$

Similarly, the received data vector at time t is given by

$$\mathbf{y}_{d,t} = \mathbf{Z}_{d,t} \mathbf{h} + \mathbf{n}_{d,t} \tag{4.6}$$

where $\mathbf{Z}_{d,t} = \mathbf{z}_{d,t}^T \otimes \mathbf{I}_{n_{rx}} \in \mathbb{C}^{n_{rx} \times n_{rx}n_{tx}(N+1)}$ with $\mathbf{z}_{d,t} = \tilde{\psi}_t \otimes \mathbf{x}_{d,t}$. Combining the received vectors from T_d and T_p symbols together, we have

$$\mathbf{y}_p = \begin{bmatrix} \mathbf{y}_{p,1} \\ \vdots \\ \mathbf{y}_{p,T_p} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{p,1} \\ \vdots \\ \mathbf{Z}_{p,T_p} \end{bmatrix} \mathbf{h} + \begin{bmatrix} \mathbf{n}_{p,1} \\ \vdots \\ \mathbf{n}_{p,T_p} \end{bmatrix} = \mathbf{Z}_p \mathbf{h} + \mathbf{n}_p, \tag{4.7}$$

$$\mathbf{y}_d = \begin{bmatrix} \mathbf{y}_{d,T_p+1} \\ \vdots \\ \mathbf{y}_{d,T_d} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{d,T_p+1} \\ \vdots \\ \mathbf{Z}_{d,T_d} \end{bmatrix} \mathbf{h} + \begin{bmatrix} \mathbf{n}_{d,T_p+1} \\ \vdots \\ \mathbf{n}_{d,T_d} \end{bmatrix} = \mathbf{z}_d \mathbf{h} + \mathbf{n}_d, \quad (4.8)$$

Therefore, the received pilot signal \mathbf{y}_p is complex Gaussian with mean $\mathbb{E}\{\mathbf{y}_p\} = \mathbf{z}_p \mathbf{h}$ and covariance $\mathbb{E}\{\mathbf{y}_p \mathbf{y}_p^H\} = \sigma^2 \mathbf{I}_{n_{rx} T_p}$. Similarly, the received data signal \mathbf{y}_d conditioned on \mathbf{z}_d is complex Gaussian with mean $\mathbb{E}\{\mathbf{y}_d\} = \mathbf{z}_d \mathbf{h}$ and covariance $\mathbb{E}\{\mathbf{y}_d \mathbf{y}_d^H\} = \sigma^2 \mathbf{I}_{n_{rx} T_d}$. We are interested in estimating the channel vector \mathbf{h} from (4.7), and (4.8).

4.2 EM algorithm

The unknown parameter that must be estimated is $\boldsymbol{\theta} \triangleq \mathbf{h}$. The complete data set is given by $\{\mathbf{y}_p, \mathbf{y}_d, \mathbf{z}_d\}$. The incomplete data set comprises the observed vectors $\{\mathbf{y}_p, \mathbf{y}_d\}$ and the hidden data (latent variable) is \mathbf{z}_d . Hence the likelihood function of the complete data is

$$\begin{aligned} f(\mathbf{y}_p, \mathbf{y}_d, \mathbf{z}_d; \boldsymbol{\theta}) &= \prod_{t=1}^{T_p} f(\mathbf{y}_{p,t}; \boldsymbol{\theta}) \prod_{t=T_p+1}^T f(\mathbf{y}_{d,t} | \mathbf{z}_{d,t}; \boldsymbol{\theta}) p(\mathbf{z}_{d,t}) \\ &= \frac{1}{M^{T_d n_{tx}} (\pi \sigma^2)^{T_p + T_d}} e^{-\frac{1}{\sigma^2} \|\mathbf{y}_p - \mathbf{z}_p \boldsymbol{\theta}\|^2} e^{-\frac{1}{\sigma^2} \|\mathbf{y}_d - \mathbf{z}_d \boldsymbol{\theta}\|^2} \end{aligned} \quad (4.9)$$

where $p(\mathbf{z}_{d,t}) = \frac{1}{M^{n_{tx}}}$. Using (4.9), the LLF of the complete data set is

$$\mathcal{L}(\mathbf{y}_p, \mathbf{y}_d, \mathbf{z}_d; \boldsymbol{\theta}) = -T_d n_{tx} \log M - (T_p + T_d) \log(\pi \sigma^2) - \frac{1}{\sigma^2} (\|\mathbf{y}_p - \mathbf{z}_p \boldsymbol{\theta}\|^2 - \|\mathbf{y}_d - \mathbf{z}_d \boldsymbol{\theta}\|^2) \quad (4.10)$$

In the next two subsections, we derive the two steps (E-step and M-step) of the EM algorithm.

4. EM-based Channel Estimation and Detection: Non-Superimposed Method 36

4.2.1 E-step

To calculate the E step, we use (3.4) and (4.10). We have

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)}) &= \mathbb{E}_{\mathbf{y}_p, \mathbf{y}_d, \mathbf{z}_d; \boldsymbol{\theta} | \mathbf{y}_p, \mathbf{y}_d; \boldsymbol{\theta}^{(l)}} [-T_d n_{tx} \log M - (T_p + T_d) \log(\pi \sigma^2) \\
&\quad - \frac{1}{\sigma^2} \|\mathbf{y}_p - \mathbf{z}_p \boldsymbol{\theta}\|^2 - \frac{1}{\sigma^2} \|\mathbf{y}_d - \mathbf{z}_d \boldsymbol{\theta}\|^2] \\
&= -T_d n_{tx} \log M - (T_p + T_d) \log(\pi \sigma^2) - \frac{1}{\sigma^2} \sum_{t=1}^{T_p} \|\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta}\|^2 \\
&\quad - \frac{1}{\sigma^2} \mathbb{E}_{\mathbf{y}_p, \mathbf{y}_d, \mathbf{z}_d; \boldsymbol{\theta} | \mathbf{y}_p, \mathbf{y}_d; \boldsymbol{\theta}^{(l)}} [\|\mathbf{y}_d - \mathbf{z}_d \boldsymbol{\theta}\|^2] \\
&= -T_d n_{tx} \log M - (T_p + T_d) \log(\pi \sigma^2) - \frac{1}{\sigma^2} \sum_{t=1}^{T_p} \|\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta}\|^2 \\
&\quad - \frac{1}{\sigma^2} \sum_{t=T_p+1}^T \sum_{j_1=1}^M \sum_{j_2=1}^M \cdots \sum_{j_{n_{tx}}=1}^M \beta_{t,j_1,j_2,\dots,j_{n_{tx}}}^{(l)} \|\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}} \boldsymbol{\theta}\|^2 \quad (4.11)
\end{aligned}$$

where $\tilde{\mathbf{Z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}} = \tilde{\mathbf{z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}}^T \otimes \mathbf{I}_{n_{rx}}$ with $\tilde{\mathbf{z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}} = \tilde{\boldsymbol{\psi}}_t \otimes [\xi_{j_1}, \xi_{j_2}, \dots, \xi_{j_{n_{tx}}}]$, and $\beta_{t,j_1,j_2,\dots,j_{n_{tx}}}^{(l)}$ is the conditional probability $P(\tilde{\mathbf{Z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}} | \mathbf{y}_p, \mathbf{y}_d; \boldsymbol{\theta}^{(l)})$ of the t th data vector at the l th iteration, given by

$$\beta_{t,j_1,j_2,\dots,j_{n_{tx}}}^{(l)} = \frac{e^{-\frac{\|\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}} \boldsymbol{\theta}^{(l)}\|^2}{\sigma^2}}}{\sum_{k_1=1}^M \sum_{k_2=1}^M \cdots \sum_{k_{n_{tx}}=1}^M e^{-\frac{\|\mathbf{y}_{d,t} - \tilde{\mathbf{z}}_{d,t,k_1,\dots,k_{n_{tx}}} \boldsymbol{\theta}^{(l)}\|^2}{\sigma^2}}}. \quad (4.12)$$

To simplify notation, we will denote any sum of the form $\sum_{j_1=1}^M \cdots \sum_{j_{n_{tx}}=1}^M$ as a sum over all n_{tx} -tuples $\mathbf{j} = (j_1, j_2, \dots, j_{n_{tx}}) \in \mathcal{J} = \{1, 2, \dots, M\}^{n_{tx}}$. Similarly, $\tilde{\mathbf{Z}}_{d,t,j_1,\dots,j_{n_{tx}}}$ and $\tilde{\mathbf{z}}_{d,t,j_1,j_2,\dots,j_{n_{tx}}}$ will be denoted by $\tilde{\mathbf{Z}}_{d,t,\mathbf{j}}$ and $\tilde{\mathbf{z}}_{d,t,\mathbf{j}}$, respectively. Using this notation, eq. (4.11) becomes

$$\mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)}) = -T_d n_{tx} \log M - (T_p + T_d) \log(\pi \sigma^2) - \frac{1}{\sigma^2} \sum_{t=1}^{T_p} \|\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta}\|^2$$

$$-\frac{1}{\sigma^2} \sum_{t=T_p+1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \left\| \mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \boldsymbol{\theta} \right\|^2 \quad (4.13)$$

4.2.2 M-step

Maximizing (4.13) with respect to $\boldsymbol{\theta}$ is equivalent to minimizing the following term with respect to $\boldsymbol{\theta}$:

$$\begin{aligned} \delta(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)}) &\triangleq \sum_{t=1}^{T_p} (\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta})^H (\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \boldsymbol{\theta})^H (\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \boldsymbol{\theta}) \\ &= \sum_{t=1}^{T_p} (\mathbf{y}_{p,t}^H - \boldsymbol{\theta}^H \mathbf{Z}_{p,t}^H) (\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\mathbf{y}_{d,t}^H - \boldsymbol{\theta}^H \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H) (\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \boldsymbol{\theta}). \end{aligned}$$

Differentiating the above equation with respect to $\boldsymbol{\theta}^H$ and equating the derivative to 0, we get $\boldsymbol{\theta}^{(l+1)}$. Specifically,

$$\begin{aligned} \frac{d\delta(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)})}{d\boldsymbol{\theta}^H} &= \sum_{t=1}^{T_p} (-\mathbf{Z}_{p,t}^H) (\mathbf{y}_{p,t} - \mathbf{Z}_{p,t} \boldsymbol{\theta}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (-\tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H (\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \boldsymbol{\theta})) \\ &= \sum_{t=1}^{T_p} (-\mathbf{Z}_{p,t}^H \mathbf{y}_{p,t} + \mathbf{Z}_{p,t}^H \mathbf{Z}_{p,t} \boldsymbol{\theta}) + \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (-\tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t} + \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \boldsymbol{\theta}) \end{aligned}$$

Note that $\mathbf{y}_p - \mathbf{Z}_p \boldsymbol{\theta}$ and $\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,\mathbf{j}} \boldsymbol{\theta}$ are independent of $\boldsymbol{\theta}^H$. Equating the above to 0 we have

$$\left(\sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{Z}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{d,\mathbf{j}} \right) \boldsymbol{\theta} = \sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{y}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t}$$

which gives

$$\begin{aligned} \boldsymbol{\theta}^{(l+1)} &= \left(\sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{Z}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{d,\mathbf{j}} \right)^{-1} \\ &\quad \times \left(\sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{y}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t} \right) \end{aligned} \quad (4.14)$$

4. EM-based Channel Estimation and Detection: Non-Superimposed Method38

We can further simplify the expression in (4.14) as follows. First, using property 5 and property 7 we note that

$$\mathbf{Z}_{p,t}^H \mathbf{Z}_{p,t} = (\mathbf{z}_{p,t}^T \otimes \mathbf{I}_{n_{rx}})^H (\mathbf{z}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}) \quad (4.15)$$

$$= (\mathbf{z}_{p,t}^* \otimes \mathbf{I}_{n_{rx}}) (\mathbf{z}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}) \quad (4.16)$$

$$= (\mathbf{z}_{p,t}^* \mathbf{z}_{p,t}^T) \otimes \mathbf{I}_{n_{rx}} \quad (4.17)$$

Similarly,

$$\tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} = (\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^T) \otimes \mathbf{I}_{n_{rx}} \quad (4.18)$$

Combining the equation (4.17) and (4.18) , we get

$$\begin{aligned} & \sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{Z}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \\ &= \sum_{t=1}^{T_p} (\mathbf{z}_{p,t}^* \mathbf{z}_{p,t}^T) \otimes \mathbf{I}_{n_{rx}} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^T) \otimes \mathbf{I}_{n_{rx}} \\ &= \left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t}^* \mathbf{z}_{p,t}^T + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^T \right) \otimes \mathbf{I}_{n_{rx}} \end{aligned} \quad (4.19)$$

which leads to

$$\begin{aligned} & \left(\sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{Z}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{d,t,\mathbf{j}} \right)^{-1} \\ &\stackrel{(a)}{=} \left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t}^* \mathbf{z}_{p,t}^T + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^T \right)^{-1} \otimes \mathbf{I}_{n_{rx}} \end{aligned} \quad (4.20)$$

where (a) uses property 6. Then, using property 3 we see that

$$\mathbf{Z}_{p,t}^H \mathbf{y}_{p,t} = (\mathbf{z}_{p,t}^T \otimes \mathbf{I}_{n_{rx}})^H \mathbf{y}_{p,t} \quad (4.21)$$

4. EM-based Channel Estimation and Detection: Non-Superimposed Method 39

$$= (\mathbf{z}_{p,t}^* \otimes \mathbf{I}_{n_{rx}}) \text{vec}(\mathbf{y}_{p,t}) \quad (4.22)$$

$$= \text{vec}(\mathbf{I}_{n_{rx}} \mathbf{y}_{p,t} \mathbf{z}_{p,t}^H) \quad (4.23)$$

$$= \text{vec}(\mathbf{y}_{p,t} \mathbf{z}_{p,t}^H) \quad (4.24)$$

$$= \text{vec}(\mathbf{y}_{p,t} \cdot \mathbf{1} \cdot \mathbf{z}_{p,t}^H) \quad (4.25)$$

$$= (\mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t}) \text{vec}(\mathbf{1}) \quad (4.26)$$

$$= \mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t} \quad (4.27)$$

Similarly,

$$\tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t} = \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \otimes \mathbf{y}_{d,t} \quad (4.28)$$

which, together with eq. (4.27) gives

$$\sum_{t=1}^{T_p} \mathbf{Z}_{p,t}^H \mathbf{y}_{p,t} + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t} = \sum_{t=1}^{T_p} (\mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \otimes \mathbf{y}_{d,t}) \quad (4.29)$$

Finally, let us define the following operator $\text{mat}_{q,r}(\mathbf{v})$ that restructures the vector \mathbf{v} into a $q \times r$ matrix such that

$$\text{vec}(\text{mat}_{q,r}(\mathbf{v})) = \mathbf{v} \quad (4.30)$$

Therefore, using eq. (4.29) and eq. (4.20), the channel estimate in eq. (4.14) can be written as

$$\begin{aligned} \boldsymbol{\theta}^{(l+1)} &= \left(\left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t}^* \mathbf{z}_{p,t}^T + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^T \right)^{-1} \otimes \mathbf{I}_{n_{rx}} \right) \\ &\quad \times \left(\sum_{t=1}^{T_p} (\mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \otimes \mathbf{y}_{d,t}) \right) \end{aligned}$$

4. EM-based Channel Estimation and Detection: Non-Superimposed Method 40

$$\begin{aligned}
&= \left(\left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t}^* \mathbf{z}_{p,t}^T + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^T \right)^{-1} \otimes \mathbf{I}_{n_{rx}} \right) \\
&\quad \times \text{vec} \left(\text{mat}_{n_{rx}, n_{tx}(N+1)} \left(\sum_{t=1}^{T_p} (\mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \otimes \mathbf{y}_{d,t}) \right) \right) \\
&= \text{vec} \left(\text{mat}_{n_{rx}, n_{tx}(N+1)} \left(\sum_{t=1}^{T_p} (\mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t}) + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^* \otimes \mathbf{y}_{d,t}) \right) \right. \\
&\quad \left. \times \left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t} \mathbf{z}_{p,t}^H + \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{d,t,\mathbf{j}} \tilde{\mathbf{z}}_{d,t,\mathbf{j}}^H \right)^{-1} \right) \tag{4.31}
\end{aligned}$$

We run this EM algorithm till the estimate is converged. Note in this thesis we set the convergence threshold value and once the threshold is reached the algorithm is stopped. The proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Semi Blind channel estimation using EM Algorithm

Input: $\mathbf{y}_p, \mathbf{y}_d, \mathbf{Z}_p$

Output: Estimate $\boldsymbol{\theta}$

- 1: Initialize $\boldsymbol{\theta}^{(0)} = (\mathbf{Z}_p)^\dagger \mathbf{y}_p$ and set $l = 0$
 - 2: **repeat**
 - 3: Compute $\beta_{t,\mathbf{j}}^{(l)}$ using (4.12) for each possible value of constellation
 - 4: Compute $\boldsymbol{\theta}^{(l+1)}$ using (4.31)
 - 5: **Until** $\boldsymbol{\theta}^{(l+1)} - \boldsymbol{\theta}^{(l)} < \epsilon$, where ϵ is the convergence threshold.
-

4.3 Computationally efficient variants of the EM algorithm

Working with discrete constellation data for an iterative algorithm such as the EM algorithm is computationally complex. A popular solution to avoid the computational complexity of such a system is to assume Gaussian distributed data [59]. Based on the results presented by the authors of [59], the authors of [17] presented a MISO-IRS system and assumed Gaussian

distributed data, as explained in Section 3.3. However, it is more challenging to assume the Gaussian distribution data for a MIMO-IRS-aided communication system, nevertheless; Appendix A briefly examines this. Therefore, to reduce the computational complexity of the proposed algorithm, we discuss several approaches in this section.

4.3.1 EM with detector

It can be noticed from (4.13) and (4.31) that the factor $\beta_{t,j}^{(l)}$ is calculated for every possible data vector. It is this step that contributes the most to the computation cost of the EM algorithm. In this section, we propose detection-based E-step to reduce the computational complexity. This approach involves estimating the conditional probability of unknown data symbols given the received signal using traditional linear detectors such as the Zero Forcing (ZF), Minimum Mean Square Error (MMSE), and a Soft Decision Fixed Complexity (SDFC) [60] for the detection. The two-step algorithm is as follows:

- **Step 1:** Use a detector to estimate the data vector $\hat{\mathbf{x}}_{d,t}$ during the E-step
- **Step 2:** Replace the sums $\sum_{j \in \mathcal{J}} \beta_{t,j}^{(l)} (\tilde{\mathbf{z}}_{d,t,j}^* \otimes \mathbf{y}_{d,t})$ and $\sum_{j \in \mathcal{J}} \beta_{t,j}^{(l)} \tilde{\mathbf{z}}_{d,t,j} \tilde{\mathbf{z}}_{d,t,j}^H$ in (4.31) by the single term corresponding to $\hat{\mathbf{x}}_{d,t}$.

Therefore, we longer compute $\beta_{t,j}^{(l)}$ and we only need to calculate $\hat{\mathbf{z}}_{d,t}^H \mathbf{y}_{d,t}$ and $\hat{\mathbf{z}}_{d,t} \hat{\mathbf{z}}_{d,t}^H$, where $\hat{\mathbf{z}}_{d,t}^H = \tilde{\boldsymbol{\psi}}_t \otimes \hat{\mathbf{x}}_{d,t}$. Using ZF, MMSE detectors, (4.31) changes to

$$\begin{aligned} \boldsymbol{\theta}^{(l+1)} = & \text{vec} \left(\text{mat}_{n_{rx}, n_{tx}(N+1)} \left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t} + \sum_{t=T_p+1}^{T_d} \hat{\mathbf{z}}_{d,t}^* \otimes \mathbf{y}_{d,t} \right) \right. \\ & \times \left. \left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t} \mathbf{z}_{p,t}^H + \sum_{t=T_p+1}^{T_d} \hat{\mathbf{z}}_{d,t} \hat{\mathbf{z}}_{d,t}^H \right)^{-1} \right). \end{aligned} \quad (4.32)$$

In what follows, we consider a simple MIMO system and present the generic equation for the detector before developing the equation for the proposed system. In an up-link MIMO

4. EM-based Channel Estimation and Detection: Non-Superimposed Method42

system with an N -antenna base station and K users, each equipped with a single antenna, the received signal is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (4.33)$$

where $\mathbf{H} \in \mathbb{C}^{N \times K}$, $\mathbf{x} \in \mathbb{C}^K$, $\mathbf{n} \in \mathbb{C}^N$ are the channel matrix, data vector, and noise vector, respectively.

ZF Detector

ZF is a linear detector that ignores the noise effect and works well in interference-limited scenarios. The estimate of the data signal for the generic MIMO system (4.33) is given as [61], [62]

$$\hat{\mathbf{x}} = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{y} = \mathbf{H}^\dagger \mathbf{y}. \quad (4.34)$$

It is evident from the equation above that \mathbf{H} is inverted, which eliminates the channel's effect on the estimation of the data symbols. Now we apply the ZF detector to estimate the data vector for the proposed system model. Therefore, using (4.2) and (4.34), we can calculate the data symbol estimate $\hat{\mathbf{x}}_{d,t}$

$$\hat{\mathbf{x}}_{d,t} = (\hat{\mathbf{H}}_{BU} + \hat{\mathbf{H}}_{SU} \Phi_t \hat{\mathbf{H}}_{BS})^\dagger \mathbf{y}_{d,t}. \quad (4.35)$$

It is crucial to keep in mind that the E-step employs the cascaded channel rather than the individual channels. Therefore, the channel matrix $\hat{\mathbf{h}}$ (the parameter $\boldsymbol{\theta}^{(l)}$ that is used in EM algorithm) is decomposed to obtain the individual channels. The channel in the E-step of

4. EM-based Channel Estimation and Detection: Non-Superimposed Method43

the l^{th} iteration would be

$$\hat{\mathbf{h}} = \begin{bmatrix} \text{vec}(\hat{\mathbf{H}}_{BU}) \\ \text{vec}(\hat{\mathbf{H}}_{BS}^T \diamond \hat{\mathbf{H}}_{SU}) \end{bmatrix} = \begin{bmatrix} \hat{h}_1 \\ \vdots \\ \hat{h}_{n_{tx}n_{rx}} \\ \vdots \\ \hat{h}_{n_{tx}n_{rx}(N+1)} \end{bmatrix}$$

The first $n_{tx}n_{rx}$ elements of $\hat{\mathbf{h}}$ will be the elements of vectorized direct channel $\text{vec}(\hat{\mathbf{H}}_{BU}) = \hat{\mathbf{h}}_{BU}$. Now, we employ the Kronecker product's Property 2, to obtain the cascaded term:

$$\text{vec}(\hat{\mathbf{H}}_{SU}\Phi_t\hat{\mathbf{H}}_{BS}) = \text{vec}(\hat{\mathbf{H}}_{BS}^T \diamond \hat{\mathbf{H}}_{SU})\text{diag}(\Phi_t).$$

The terms after the first $n_{tx}n_{rx}$ in $\hat{\mathbf{h}}$ will be the elements of $\text{vec}(\hat{\mathbf{H}}_{BS}^T \diamond \hat{\mathbf{H}}_{SU})$. Product of $\text{vec}(\hat{\mathbf{H}}_{BS}^T \diamond \hat{\mathbf{H}}_{SU})$ with the $\text{diag}(\Phi_t)$ at a particular iteration, would give us $\text{vec}(\hat{\mathbf{H}}_{SU}\Phi_t\hat{\mathbf{H}}_{BS})$, which can further be de-vectorized to get the actual cascaded channel $\hat{\mathbf{H}}_{SU}\Phi_t\hat{\mathbf{H}}_{BS}$. Adding the direct and cascaded channel would give us the estimate of the total channel for the iteration l of E-step. On obtaining the channel value, we calculate the data estimate using (4.35). Once we get the data estimate, we pass it through a QAM detector to calculate the closest data vector $\hat{\mathbf{x}}_{d,t}$. Finally, after calculating the data vector estimate, we then calculate the corresponding $\hat{\mathbf{z}}_{d,t}^H \mathbf{y}_{d,t}$ and $\hat{\mathbf{z}}_{d,t}^H \hat{\mathbf{z}}_{d,t}$.

Minimum Mean Square Error

Another linear detector aims at estimating the transmitted symbols by minimizing the MSE between the transmitted and estimated signal. The MMSE estimate for the data signal in (4.33) is given as [61]

$$\hat{\mathbf{x}} = \left(\mathbf{H}^H \mathbf{H} + \frac{K}{SNR} \mathbf{I} \right)^{-1} \mathbf{H}^H \mathbf{y}. \quad (4.36)$$

4. EM-based Channel Estimation and Detection: Non-Superimposed Method44

The fact that the MMSE detector considers the noise effect and requires knowledge of SNR clearly distinguishes it from the ZF detector. Now, with this detector, we use the same approach to the ZF detector to estimate the data vector for the proposed method. The data vector $\hat{\mathbf{x}}_{d,t}$ for the proposed system is calculated using

$$\hat{\mathbf{x}}_{d,t} = \left(\mathbf{G} \mathbf{G}^H + \frac{n_{tx}}{\sigma^2} \mathbf{I}_{n_{tx}} \right)^{-1} \mathbf{G}^H \mathbf{y}_{d,t} \quad (4.37)$$

where $\mathbf{G} = (\hat{\mathbf{H}}_{BU} + \hat{\mathbf{H}}_{SU} \Phi_t \hat{\mathbf{H}}_{BS})$. The obtained data vector is passed through the QAM detector to calculate the final data vector. Finally, using the estimated data vector we calculate corresponding $\hat{\mathbf{z}}_{d,t}^H \mathbf{y}_{d,t}$ and $\hat{\mathbf{z}}_{d,t}^H \hat{\mathbf{z}}_{d,t}$. The proposed algorithm with the detectors ZF and MMSE to detect the data vector is summarized in Algorithm 2.

Algorithm 2 Estimating the channel using the detectors - ZF, MMSE

Input: $\mathbf{y}_p, \mathbf{y}_d, \mathbf{Z}_p$

Output: Estimate θ

- 1: Initialize $\theta^{(0)} = (\mathbf{Z}_p)^\dagger \mathbf{y}_p$ and set $l = 0$
 - 2: **repeat**
 - 3: Estimate data symbols using the detectors
 - 4: Pass through the constellation detector
 - 5: Use the new data vector while computing $\hat{\mathbf{z}}_{d,t}^H \mathbf{y}_{d,t}$ and $\hat{\mathbf{z}}_{d,t}^H \hat{\mathbf{z}}_{d,t}$
 - 6: Compute $\theta^{(l+1)}$, according to (4.31)
 - 7: **Until** $\theta^{(l+1)} - \theta^{(l)} < \epsilon$, where ϵ is the convergence threshold.
-

SDFC Detector - Partial Marginalization

The authors of [60] proposed a novel two-step fixed complexity soft detection method based on Partial Marginalization (PM). We will briefly review this method before we apply it to our situation. Consider the system as described in (4.33) with the data signal \mathbf{x} carrying bits $b_i \in \{1, -1\}$, $i = 1, \dots, K \log_2(M)$, where M represents the constellation size. The

4. EM-based Channel Estimation and Detection: Non-Superimposed Method45

posteriori LLR for detection of b_i given the received signal \mathbf{y} is as follows

$$L(b_i|\mathbf{y}) = \log \left(\frac{P(b_i = 1|\mathbf{y})}{P(b_i = -1|\mathbf{y})} \right) \quad (4.38)$$

$$= \log \left(\frac{\sum_{\mathbf{x}: b_i=1} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2)}{\sum_{\mathbf{x}: b_i=-1} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2)} \right) \quad (4.39)$$

where $\mathbf{x} : b_i = 1$ represents all the vectors \mathbf{x} such that $b_i = 1$, and assuming all the bits are equiprobable (i.e.), $P(b_i = 1) = P(b_i = -1) = \frac{1}{2}$. Now calculating the above expression is quite complex as there are altogether 2^{MK} sum terms. Therefore, to overcome this complexity, the authors [60] devised the two-step partial marginalization method. To perform the two-step SDFC-PM method, they have used a permutation heuristic (as in Algorithm 3 which gives the symbol index permutation \mathcal{I} on $[1, \dots, K]$ as well) to rearrange the columns of \mathbf{H} such that the symbol error rate is minimized. A user-chosen parameter p is used to partition the data and channel matrix. Therefore we rewrite $\mathbf{H} = [\mathbf{h}_{\mathcal{I}_1}, \dots, \mathbf{h}_{\mathcal{I}_p}, \dots, \mathbf{h}_{\mathcal{I}_K}]$ and $\mathbf{x} = [x^{\mathcal{I}_1}, \dots, x^{\mathcal{I}_p}, \dots, x^{\mathcal{I}_K}]^T$ as

$$\mathbf{H}_a \triangleq [\mathbf{h}_{\mathcal{I}_1}, \dots, \mathbf{h}_{\mathcal{I}_p}], \quad \mathbf{H}_b \triangleq [\mathbf{h}_{\mathcal{I}_{p+1}}, \dots, \mathbf{h}_{\mathcal{I}_K}], \quad (4.40)$$

$$\mathbf{x}^a \triangleq [x^{\mathcal{I}_1}, \dots, x^{\mathcal{I}_p}]^T, \quad \mathbf{x}^b \triangleq [x^{j(\mathcal{I}_{p+1})}, \dots, x^{\mathcal{I}_K}]^T \quad (4.41)$$

The two-step PM approximation of (4.39) is as follows [63],

$$L(b_i|\mathbf{y}_i) \approx \log \left(\frac{\sum_{\mathbf{x}^a: b_i=1} \max_{\mathbf{x}^b} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}_a \mathbf{x}^a - \mathbf{H}_b \mathbf{x}^b\|^2)}{\sum_{\mathbf{x}^a: b_i=-1} \max_{\mathbf{x}^b} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}_a \mathbf{x}^a - \mathbf{H}_b \mathbf{x}^b\|^2)} \right) \quad (4.42)$$

$$\approx \log \left(\frac{\sum_{\mathbf{x}^a: b_i=1} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}_a \mathbf{x}^a - \mathbf{H}_b \hat{\mathbf{x}}^b\|^2)}{\sum_{\mathbf{x}^a: b_i=-1} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{H}_a \mathbf{x}^a - \mathbf{H}_b \hat{\mathbf{x}}^b\|^2)} \right) \quad (4.43)$$

where a total of $K - p$ of the K sums are approximated by maximizing over \mathbf{x}^b in (4.42), which is the first step and referred to as the exact marginalization. The second step uses a

4. EM-based Channel Estimation and Detection: Non-Superimposed Method46

low-complexity method such as a ZF or ZF-Decision Feedback solution $\hat{\mathbf{x}}^b$ to approximate the maximization in (4.43).

Algorithm 3 Permutation of columns of \mathbf{H}

Input: \mathbf{H}

Output: \mathcal{I}

- 1: Let $\mathcal{I} = [\emptyset]$ and let $\mathcal{I}^c = [1, \dots, n_{tx}]$
 - 2: Calculate $\delta = \text{diag}\{(\mathbf{H}^H \mathbf{H})^{-1}\}$. Let k be the index of the largest term of δ
 - 3: Set $\mathcal{I} = [\mathcal{I}, \mathcal{I}_k^c]$.
 - 4: Remove the k th column from \mathbf{H} .
 - 5: Remove the k th element of \mathcal{I}^c .
 - 6: **repeat** till \mathbf{H} is empty.
-

Now, using a similar method to that of [60], we propose a two-step method where we first perform exact marginalization followed by approximate marginalization. Let us now consider the system model in (4.2) and partition \mathbf{H} and $\mathbf{x}_{d,t}$ in a similar manner as in (4.40),(4.41) using the permutation heuristic given in Algorithm 3. We have

$$\mathbf{H}_a \triangleq [\mathbf{h}_{\mathcal{I}_1}, \dots, \mathbf{h}_{\mathcal{I}_p}], \quad \mathbf{H}_b \triangleq [\mathbf{h}_{\mathcal{I}_{p+1}}, \dots, \mathbf{h}_{\mathcal{I}_{n_{tx}}}],$$

$$\mathbf{x}_{d,t}^a \triangleq [x_{d,t}^{\mathcal{I}_1}, \dots, x_{d,t}^{\mathcal{I}_p}]^T, \quad \mathbf{x}_{d,t}^b \triangleq [x_{d,t}^{\mathcal{I}_{(p+1)}}, \dots, x_{d,t}^{\mathcal{I}_{n_{tx}}}]^T$$

where $\mathbf{H} = (\mathbf{H}_{BU} + \mathbf{H}_{SU} \Phi_t \mathbf{H}_{BS})$. Let p be the user parameter that refers to the partition of the symbols (p symbols follow exact marginalization and remaining $n_{tx}M - p$ symbols follow approximate marginalization). Therefore, we can rewrite (4.2) as $\mathbf{y}_{d,t} = \mathbf{H}_a \mathbf{x}_{d,t}^a + \mathbf{H}_b \mathbf{x}_{d,t}^b + \mathbf{n}_{d,t}$ and the data signal estimate is as follows,

$$\hat{\mathbf{x}}_{d,t}^b \triangleq \arg \min_{\mathbf{x}_{d,t}^a \in \mathcal{X}} \|(\mathbf{H}_b^H \mathbf{H})^{-1} \mathbf{H}_b^H (\mathbf{y}_{d,t} - \mathbf{H}_a \mathbf{x}_{d,t}^a)\|^2.$$

Once we find the above value, it is then passed on to a QAM detector. After obtaining the estimate of $\mathbf{x}_{d,t}^b$ from QAM detector, for each possible values of $\mathbf{x}_{d,t}^a$, we proceed to compute $\beta_{t,j}^{(l)}$ in the E-step with the reduced number of symbols in $\mathbf{x}_{d,t}^a$. Therefore the E-step in (4.13)

changes as in (4.44),

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)}) = & -(T_d n_{tx}) \log M - (T_p + T_d) \log(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{t=1}^{T_p} \|\mathbf{y}_{p,t} - \mathbf{Z}_{p,t}\boldsymbol{\theta}\|^2 \\ & - \frac{1}{\sigma^2} \sum_{t=T_p+1}^{T_d} \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,j_1, \dots, j_p}^{(l)} \|\mathbf{y}_{d,t} - \tilde{\mathbf{Z}}_{d,t,\mathbf{j}}\boldsymbol{\theta}\|^2 \end{aligned} \quad (4.44)$$

where $\mathbf{j} = (j_1, j_2, \dots, j_p) \in \mathcal{J} = \{1, 2, \dots, M\}^p$ and $\tilde{\mathbf{Z}}_{d,t,\mathbf{j}} = \tilde{\boldsymbol{\psi}}_t^T \otimes [\xi_{j_1}, \dots, \xi_{j_p}, \hat{\mathbf{x}}_{d,t}^b] \otimes \mathbf{I}_{n_{rx}}$. After computing the E-step with the detector, the M-step is executed as previously conducted, without any modification.

4.3.2 Log-max approximation

This algorithm is inspired by the max-log approximation discussed by the authors in [60], where they approximate each of the sums in (4.39) with their largest term as follows

$$L(b_i|\mathbf{y}) = \log \left(\frac{\max_{\mathbf{x}: b_i=1} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{Hx}\|^2)}{\max_{\mathbf{x}: b_i=-1} \exp(-\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{Hx}\|^2)} \right). \quad (4.45)$$

Therefore we use this idea of approximation to circumvent the computational complexity. Initially, we calculated terms $\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t}$ and $\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{z}}_{d,t,\mathbf{j}}$ in (4.31) for every possible combination of data symbols. However choosing the maximum value of $\beta_{t,\mathbf{j}}^{(l)}$ and obtaining the corresponding data vector, eliminates the need to calculate $\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^H \mathbf{y}_{d,t}$ and $\tilde{\mathbf{z}}_{d,t,\mathbf{j}}^H \tilde{\mathbf{z}}_{d,t,\mathbf{j}}$ for every possible combination data symbols. Hence, the proposed approximation reduces the computational complexity of the proposed system. The following is the two-step algorithm, which we refer to in this thesis as the log-max algorithm:

- **Step 1:** Calculate all $\beta_{t,\mathbf{j}}^{(l)}$ values for $\mathbf{j} \in \mathcal{J}$ using (4.12) and store them $\boldsymbol{\gamma} \in \mathbb{C}^{M^{n_{tx}}}$.
- **Step 2:** Choose the data vector that is associated with $\max\{\boldsymbol{\gamma}\}$ value and compute $\boldsymbol{\theta}^{(l+1)}$.

4. EM-based Channel Estimation and Detection: Non-Superimposed Method48

Basically eq. (4.31) changes as follows,

$$\boldsymbol{\theta}^{(l+1)} = \text{vec} \left(\text{mat}_{n_{rx}, n_{tx}(N+1)} \left(\sum_{t=1}^{T_p} (\mathbf{z}_{p,t}^* \otimes \mathbf{y}_{p,t}) + \sum_{t=T_p+1}^{T_d} \max(\boldsymbol{\gamma}) (\tilde{\mathbf{z}}_{d,t}^* \otimes \mathbf{y}_{d,t}) \right) \times \left(\sum_{t=1}^{T_p} \mathbf{z}_{p,t} \mathbf{z}_{p,t}^H + \sum_{t=T_p+1}^{T_d} \max(\boldsymbol{\gamma}) \tilde{\mathbf{z}}_{d,t} \tilde{\mathbf{z}}_{d,t}^H \right)^{-1} \right) \quad (4.46)$$

4.4 Complexity

In this section, we compare the complexity of non-superimposed and non-superimposed protocol with detectors. The majority of the computational complexity in the proposed approaches is due to matrix multiplications and inversions. However, the complexity varies in these models while calculating the conditional probability of hidden data given observed data. Now, let us look at the complexity of the non-superimposed method. In (4.31), the term $\beta_{t,j}^{(l)}$ is calculated for each possible point on the constellation for every transmit antenna n_{tx} . Accordingly, the complexity of calculating $\beta_{t,j}^{(l)}$ would be $\mathcal{O}(T_d M^{n_{tx}})$ (focusing mainly on the data symbols). However, by estimating the data symbols (using detectors to estimate the data symbols), we reduce the complexity but at the cost of accuracy. Let us examine the /acrshort[zf] detector in more detail, this detector is used to estimate the data vector, thereby preventing the necessity of calculating $\beta_{t,j}^{(l)}$ value. Therefore the complexity changes from $\mathcal{O}(T_d M^{n_{tx}})$ to $\mathcal{O}(T_d)$. This argument remains the same for the MMSE detector case. However, since the SDFC detector estimates only a portion p of data symbols rather than estimating all data symbols, the complexity changes from $\mathcal{O}(T_d M^{n_{tx}})$ to $\mathcal{O}(T_d M^p)$ where $p < n_{tx}$. This indicates that while SDFC is more computationally complex than ZF or MMSE detectors, but it would provide us with more accurate estimations.

4.5 Summary

In this chapter, we introduced the system model for the MIMO-IRS-aided communication system. For the proposed system model, we estimated the channels using the EM algorithm in a semi-blind manner for the non-superimposed protocol. Additionally, we propose detection based E-step focusing on minimizing computational complexity. We have utilized both SDFC detector and conventional detectors like the ZF and MMSE for detection. Further, we discuss the log-max approximation that reduces the computational complexity of the actual EM algorithm but at the cost of accuracy. Finally, we contrast the computational complexity of exact computations (i.e., those performed without the use of detectors) with those performed with detectors. In the following chapter, we propose a superimposed protocol and carry out channel estimation in a similar procedure.

Chapter 5

EM-based Channel Estimation: Superimposed Method

In this chapter, we make use of the superimposed protocol. As the name suggests, in this protocol, the data and pilot symbols are transmitted simultaneously, as in Figure 5.1. An interesting advantage of this method is that it helps suppress pilot contamination, which refers to the reuse of the same pilot symbols by different users. It is important to note that pilot contamination could potentially lead to inaccurate channel estimation [64]. However, there is a trade-off as superimposed protocol could introduce pilot and data symbol interference, impacting the quality of channel estimation [65]. Despite this performance degradation, several research studies [65–67] have shown that the superimposed method outperforms the non-superimposed method in terms of achievable sum rate. Therefore in this chapter, we investigate the channel estimation using the superimposed protocol for MIMO-IRS-aided communication system by following the methods similar to that in Chapter 4. We propose a system model for this protocol and utilize the EM algorithm to estimate the channels.

5.1 System Model

The system model remains similar to the one in Section 4.1, however, the transmission protocol is different. Over a data frame of length T , the BS transmits pilot and data symbols simultaneously. Specifically, at time $t = 1, 2, \dots, T$ the BS transmits

$$\mathbf{x}_t = \mathbf{x}_{p,t} + \mathbf{x}_{d,t},$$

where $\mathbf{x}_{p,t} \in \mathbb{C}^{n_{tx}}$ are the known pilot symbols, and $\mathbf{x}_{d,t} \in \mathbb{C}^{n_{tx}}$ are the data symbols. The total power P is divided among the data and pilot symbols as follows: $\mathbb{E}\{\mathbf{x}_{p,t}^H \mathbf{x}_{p,t}\} = \alpha P$, and $\mathbb{E}\{\mathbf{x}_{d,t}^H \mathbf{x}_{d,t}\} = (1 - \alpha)P$. Each element of $\mathbf{x}_{d,t}$, $x_{d,t}[n]$, $n = 1, \dots, n_{tx}$, is randomly selected from a constellation $\mathcal{X} = \{\xi_1, \xi_2, \dots, \xi_M\}$ of size M .

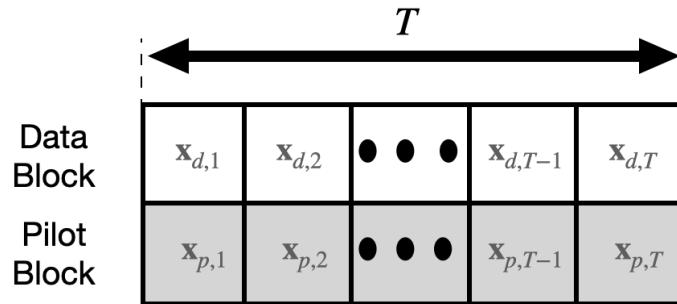


Figure 5.1: Superimposed protocol for time interval T

The received signals by the user due to the transmission of \mathbf{x}_t is given by

$$\mathbf{y}_t = (\mathbf{H}_{BU} + \mathbf{H}_{SU} \Phi_t \mathbf{H}_{BS}) \mathbf{x}_t + \mathbf{n}_t \quad (5.1)$$

where the noise vector \mathbf{n}_t is assumed to be Gaussian with mean zero and covariance $\sigma^2 \mathbf{I}_{n_{rx}}$.

Following the same procedure as in Section 4.1 we can rewrite (5.1) as

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{h} + \mathbf{n}_t \quad (5.2)$$

where $\tilde{\psi}_t = \begin{bmatrix} 1 \\ \psi_t \end{bmatrix} \in \mathbb{C}^{(N+1) \times 1}$, $\mathbf{h} = \begin{bmatrix} \text{vec}(\mathbf{H}_{BU}) \\ \text{vec}(\mathbf{H}_{BS}^T \diamond \mathbf{H}_{SU}) \end{bmatrix} \in \mathbb{C}^{n_{rx} n_{tx}(N+1) \times 1}$, and

$$\mathbf{Z}_t = \tilde{\psi}_t^T \otimes \mathbf{x}_t^T \otimes \mathbf{I}_{n_{rx}} = \underbrace{\tilde{\psi}_t^T \otimes \mathbf{x}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}}_{\mathbf{Z}_{p,t}} + \underbrace{\tilde{\psi}_t^T \otimes \mathbf{x}_{d,t}^T \otimes \mathbf{I}_{n_{rx}}}_{\mathbf{Z}_{d,t}} \in \mathbb{C}^{n_{rx} \times n_{rx} n_{tx}(N+1)} \quad (5.3)$$

Stacking the received signals over the whole data frame, we get

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_T \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_{p,1} \\ \vdots \\ \mathbf{Z}_{p,T} \end{bmatrix} \mathbf{h} + \begin{bmatrix} \mathbf{Z}_{d,1} \\ \vdots \\ \mathbf{Z}_{d,T} \end{bmatrix} \mathbf{h} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_T \end{bmatrix} = \mathbf{Z}_p \mathbf{h} + \mathbf{Z}_d \mathbf{h} + \mathbf{n}, \quad (5.4)$$

We are interested in estimating the cascaded channel \mathbf{h} from (5.4).

5.2 EM algorithm

The unknown parameter that must be estimated is $\boldsymbol{\theta} \triangleq \mathbf{h}$. The incomplete data set is $\{\mathbf{y}\}$, the hidden data (latent variable) is $\{\mathbf{Z}_d\}$ and the complete data set is $\{\mathbf{y}, \mathbf{Z}_d\}$. Hence the likelihood function of the complete data is

$$\begin{aligned} f(\mathbf{y}, \mathbf{Z}_d; \boldsymbol{\theta}) &= \prod_{t=1}^T f(\mathbf{y}_t | \mathbf{Z}_{d,t}; \boldsymbol{\theta}) p(\mathbf{Z}_{d,t}) \\ &= \prod_{t=1}^T \frac{1}{\pi \sigma^2} e^{-\frac{1}{\sigma^2} \|\mathbf{y}_t - \mathbf{Z}_{p,t} \boldsymbol{\theta} - \mathbf{Z}_{d,t} \boldsymbol{\theta}\|^2} \frac{1}{M^{n_{tx}}} \\ &= \frac{1}{M^{Tn_{tx}} (\pi \sigma^2)^T} e^{-\frac{1}{\sigma^2} \|\mathbf{y} - (\mathbf{Z}_p + \mathbf{Z}_d) \boldsymbol{\theta}\|^2}, \mathbf{Z}_d \in \mathcal{X}^{Tn_{tx}} \end{aligned} \quad (5.5)$$

because $p(\mathbf{Z}_t) = \frac{1}{M^{n_{tx}}}$. From (5.5), the LLF of the complete data set is

$$L(\mathbf{\eta}, \mathbf{z}_d; \boldsymbol{\theta}) = -Tn_{tx} \log M - T \log(\pi\sigma^2) - \frac{1}{\sigma^2} \|\mathbf{\eta} - (\mathbf{z}_p + \mathbf{z}_d)\boldsymbol{\theta}\|^2 \quad (5.6)$$

- **E-step**

For the E-step, we use (3.4). At the l^{th} iteration, we have

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)}) &= \mathbb{E}_{\mathbf{\eta}, \mathbf{z}_d; \boldsymbol{\theta} | \mathbf{\eta}; \boldsymbol{\theta}^{(l)}} \left(-Tn_{tx} \log M - T \log(\pi\sigma^2) - \frac{1}{\sigma^2} \|\mathbf{\eta} - (\mathbf{z}_p + \mathbf{z}_d)\boldsymbol{\theta}\|^2 \right) \\ &= -Tn_{tx} \log M - T \log(\pi\sigma^2) - \frac{1}{\sigma^2} \mathbb{E}_{\mathbf{\eta}, \mathbf{z}_d; \boldsymbol{\theta} | \mathbf{\eta}; \boldsymbol{\theta}^{(l)}} (\|\mathbf{\eta} - (\mathbf{z}_p + \mathbf{z}_d)\boldsymbol{\theta}\|^2) \\ &= -Tn_{tx} \log M - T \log(\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \|\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta}\|^2 \end{aligned} \quad (5.7)$$

where $\tilde{\mathbf{Z}}_{t,\mathbf{j}} = \tilde{\mathbf{z}}_{t,\mathbf{j}}^T \otimes \mathbf{I}_{n_{rx}}$, $\tilde{\mathbf{z}}_{t,\mathbf{j}} = \tilde{\psi}_t \otimes (\xi_{j_1} + x_{p,t}[1], \xi_{j_2} + x_{p,t}[2], \dots, \xi_{j_{n_{tx}}} + x_{p,t}[n_{tx}])$ and $\beta_{t,\mathbf{j}}^{(l)} = P(\tilde{\mathbf{Z}}_{t,\mathbf{j}} | \mathbf{\eta}; \boldsymbol{\theta}^{(l)})$ is the conditional probability of the t th data-pilot vector at the l th iteration given by

$$\beta_{t,\mathbf{j}}^{(l)} = \frac{e^{-\frac{1}{\sigma^2} \|\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,j_1, \dots, j_{n_{tx}}} \boldsymbol{\theta}^{(l)}\|^2}}{\sum_{k_1=1}^M \cdots \sum_{k_{n_{tx}}=1}^M e^{-\frac{1}{\sigma^2} \|\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,k_1, \dots, k_{n_{tx}}} \boldsymbol{\theta}^{(l)}\|^2}}. \quad (5.8)$$

- **M-step**

Maximizing (5.7) with respect to $\boldsymbol{\theta}$ is equivalent to minimizing the following term:

$$\begin{aligned} \delta(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)}) &\triangleq \sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta})^H (\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta}) \\ &= \sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\mathbf{y}_t^H - \boldsymbol{\theta}^H \tilde{\mathbf{Z}}_{t,\mathbf{j}}^H) (\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta}). \end{aligned}$$

Differentiating the above equation with respect to $\boldsymbol{\theta}^H$, equating the derivative to 0,

and solving for $\boldsymbol{\theta}$ gives us $\boldsymbol{\theta}^{(l+1)}$. Specifically,

$$\begin{aligned}\frac{d\delta(\boldsymbol{\theta}; \boldsymbol{\theta}^{(l)})}{d\boldsymbol{\theta}^H} &= \sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (-\tilde{\mathbf{Z}}_{t,\mathbf{j}}^H (\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta})) \\ &= \sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (-\tilde{\mathbf{Z}}_{t,\mathbf{j}}^H \mathbf{y}_t + \tilde{\mathbf{Z}}_{t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta})\end{aligned}$$

Note that $\mathbf{y}_t - \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta}$ is independent of $\boldsymbol{\theta}^H$. Equating the above to 0 we have

$$\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{t,\mathbf{j}} \boldsymbol{\theta} = \sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{t,\mathbf{j}}^H \mathbf{y}_t \quad (5.9)$$

which gives

$$\boldsymbol{\theta}^{(l+1)} = \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{t,\mathbf{j}}^H \tilde{\mathbf{Z}}_{t,\mathbf{j}} \right)^{-1} \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{Z}}_{t,\mathbf{j}}^H \mathbf{y}_t \right) \quad (5.10)$$

Using the results obtained from (4.18), (4.28), and (4.30), we can further simplify (5.10) as follows:

$$\begin{aligned}\boldsymbol{\theta}^{(l+1)} &= \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{t,\mathbf{j}}^* \tilde{\mathbf{z}}_{t,\mathbf{j}}^T) \otimes \mathbf{I}_{n_{rx}} \right)^{-1} \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{t,\mathbf{j}}^* \otimes \mathbf{y}_t) \right) \\ &= \left(\left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{t,\mathbf{j}}^* \tilde{\mathbf{z}}_{t,\mathbf{j}}^T) \right)^{-1} \otimes \mathbf{I}_{n_{rx}} \right) \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{t,\mathbf{j}}^* \otimes \mathbf{y}_t) \right) \\ &= \left(\left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{t,\mathbf{j}}^* \tilde{\mathbf{z}}_{t,\mathbf{j}}^T \right)^{-1} \otimes \mathbf{I}_{n_{rx}} \right) \\ &\quad \text{vec} \left(\text{mat}_{n_{rx}, n_{tx}(N+1)} \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{t,\mathbf{j}}^* \otimes \mathbf{y}_t) \right) \right) \\ &= \text{vec} \left(\text{mat}_{n_{rx}, n_{tx}(N+1)} \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} (\tilde{\mathbf{z}}_{t,\mathbf{j}}^* \otimes \mathbf{y}_t) \right) \left(\sum_{t=1}^T \sum_{\mathbf{j} \in \mathcal{J}} \beta_{t,\mathbf{j}}^{(l)} \tilde{\mathbf{z}}_{t,\mathbf{j}} \tilde{\mathbf{z}}_{t,\mathbf{j}}^H \right)^{-1} \right) \quad (5.11)\end{aligned}$$

5.3 Summary

The semi-blind channel estimation method for MIMO-IRS-aided communication system applying the EM algorithm using superimposed protocol is introduced in this chapter. In the superimposed protocol, pilot and data symbols are transmitted simultaneously. Furthermore, we introduce a term α that allows splitting the power accordingly to the data or pilot symbols. Finally, we perform channel estimation for the proposed method based on the EM algorithm. The performance of the non-superimposed and superimposed methods and related computational complexity will be examined in the following chapter.

Chapter 6

Simulation Results and Discussions

In this chapter, we discuss the performance of the methods proposed in Chapter 4 (non-superimposed method) and Chapter 5 (super-imposed method). Furthermore, we compare the simulation results obtained for the non-superimposed method and the method presented by the authors of [17]. Finally, we provide a succinct discussion on the complexity of the non-superimposed method compared to the computationally efficient variants of the EM algorithm, such as the log-max algorithm and detection-based E-step.

6.1 Non-Superimposed method

In this section, we examine the impact of various parameters such as the number of pilot symbols (T_p), number of receiver antennas (n_{rx}), and number of data symbols (T_d) on the performance of the non-superimposed method. The estimation of the channels includes operations such as matrix multiplications and inversions which could result in floating point errors. Therefore, it is necessary to run the experiment for multiple number of Monte Carlo iterations to ensure accurate results. Hence throughout this chapter, the simulation results are obtained by averaging over 100 Monte Carlo iterations. We use the NMSE to evaluate the performance of the proposed methods. The Normalized minimum mean squared error

(NMSE) is defined as

$$NMSE = \frac{\|\boldsymbol{\theta}^{(l+1)} - \boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^2}, \quad (6.1)$$

where $\boldsymbol{\theta}^{(l+1)}$ is the estimated channel and $\boldsymbol{\theta}$ is the actual channel.

In Figure 6.1, we present the simulation result for NMSE vs T_p . The parameters used to perform the simulation are given in Table 6.1. We compared the proposed non-superimposed method with the approximating methods such as the log-max method and with detection-based E-step (as discussed in Section 4.3.1). One can interpret from this plot that the proposed algorithm without the approximating techniques performed the best with the increase in the number of pilot symbols. Although, in every case, the NMSE decreases as the number of pilot symbols is increased. Moreover, from the plot, one can see that the SDFC detector performance lies in between the MMSE, ZF detectors and the proposed non-superimposed method, which is an expected behavior as discussed in Chapter 4.

Table 6.1: Parameters used for T_p vs NMSE

Symbols	Description
T_d	50
N	15
n_{rx}	4
n_{tx}	2
Modulation scheme (M)	QAM (4)
SNR	20 dB
r	1
Convergence threshold ϵ	0.5

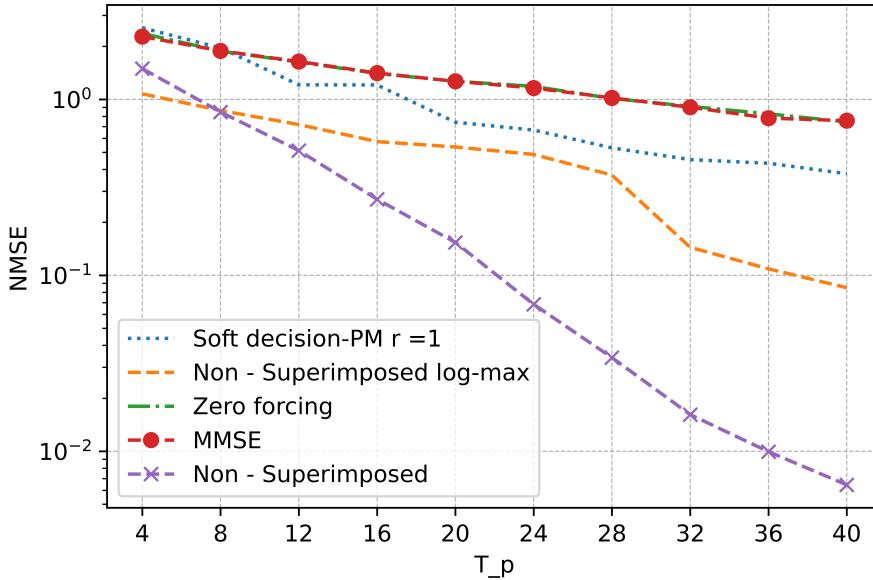


Figure 6.1: NMSE versus T_p with $T_d = 50$, number of IRS elements $N = 15$, and for a 2×4 MIMO system

In Figure 6.2, the parameters used are similar to the previous experiment as in Table 6.1 with the change in $T_p = 20$. The simulation result for variation in NMSE versus T_d (for fixed T_p) is shown in this plot. The proposed non-superimposed method was compared to approximate methods, such as detection-based E-step and the log-max method. The performance looks very similar to the previous experiment. The plot illustrates that NMSE improves as the number of data symbols increases. Therefore, it can be said that lesser pilot overhead results in improved NMSE.

Moreover, both Figure 6.1 and Figure 6.2 show that the performance of ZF and MMSE detectors are very similar. The unexpected increase in Log-max performance, contrary to the anticipated decrease, can be seen in Figure 6.2. It is possible that the unexpected result was caused by the use of different software packages to estimate Log-max performance. The Python package mpmath was used to achieve high-precision floating-point arithmetic with

Log-max precision. Alternatively, the Python package gmpy2 was used to perform other computations. The lack of array data structures in gmpy2 as well as viable alternatives is one of the limitations.

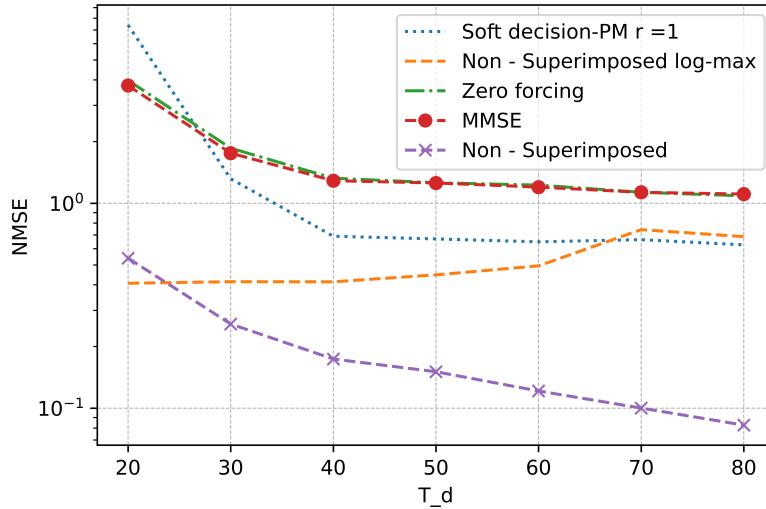


Figure 6.2: NMSE versus T_d with $T_p = 20$, $N = 15$ and for a 2×4 MIMO system

In Figure 6.3, the parameters used are similar to Table 6.1 except the only changes are $T_p = 20$ and $T_d = 50$. The simulation result for variation in NMSE versus n_{rx} is shown in this plot. This graph demonstrates that the proposed model works better when the number of receiver antennas is increased. This is a very significant finding because the graph can also be interpreted as a single antenna in a multi-user MISO system. In this instance, we may state that the x -axis would be the number of users rather than the number of receiver antennas, and as we increase the number of single-antenna users, the NMSE drops.

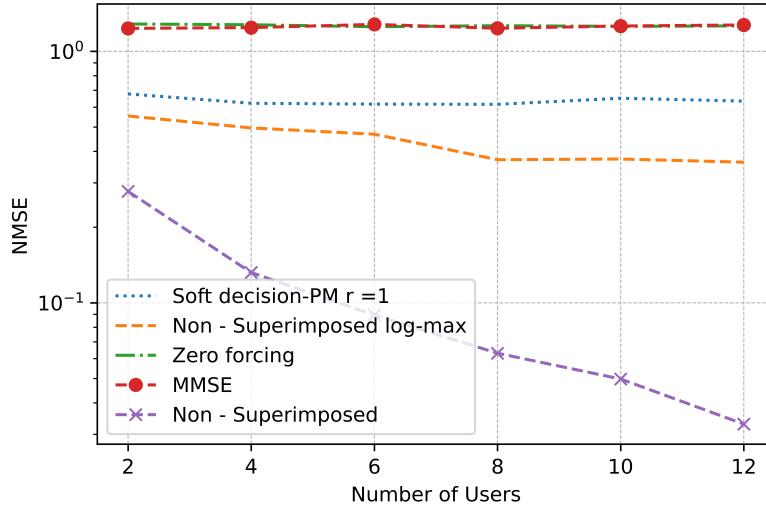


Figure 6.3: NMSE vs Number of users with $T_p = 20$, $T_d = 50$, number of IRS elements $N = 15$ and $n_{tx} = 2$

As mentioned earlier in this thesis, we proposed the approximation methods for the non-superimposed method to deal with the computational complexity. Therefore from Table 6.2, it can be confirmed that the proposed algorithm with detectors takes lesser time to compute but at the cost of accuracy as noticed in Figures 6.1, 6.2 and 6.3. It is evident from this table that the use of SDFC is comparatively higher in complexity but lesser than the proposed algorithm without detectors, which agrees with the discussion in Section 4.4. Finally, the values in Table 6.2 are obtained for 100 monte carlo iterations and a set number of EM iterations 10.

6.2 Comparison

In this section, we compare the proposed method with the method proposed by authors of [17]. The parameters used for the following two simulations are as in Table 6.3, which are similar to that presented in [17]. As mentioned earlier, the following experiments are

Table 6.2: Time complexity for non-superimposed method with $T_p = 20$, $T_d = 50$, $N = 16$ for 2×4 MIMO system

Method	Time Complexity
No detectors	$361.66s$
Log Max	$0.75s$
ZF	$0.089s$
MMSE	$0.087s$
SDFC	$91.05s$

performed for about 100 Monte Carlo iterations, and the results are averaged over all the iterations.

From Figure 6.4 for a fixed number of data symbols, it can be observed that the proposed method performs better than the approach put forth by the authors of [17] as the number of pilot symbols is increased. Moreover, in Figure 6.5 for a fixed number of pilot symbols, the proposed method performs better than the method presented by the authors of [17] as the number of data symbols increases. Therefore, the CSI estimate for the proposed method provides a more accurate value even with fewer pilots.

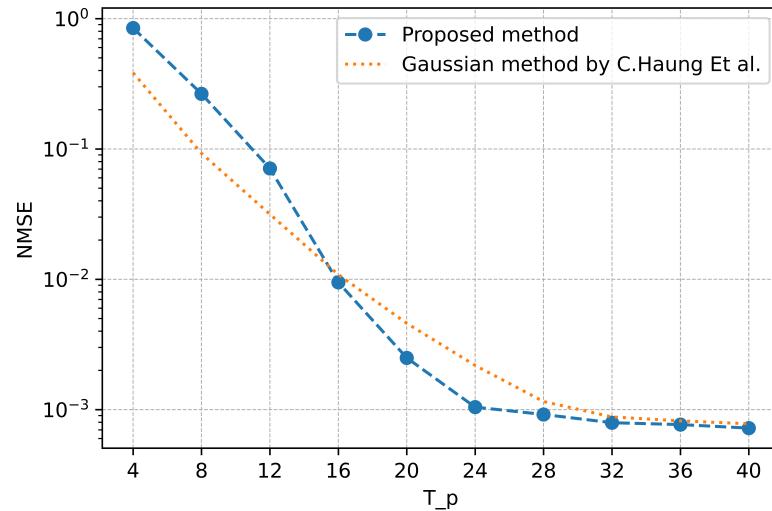


Figure 6.4: NMSE versus T_p with $T_d = 50$, number of IRS elements $N = 32$ and for a 1×8 SIMO system

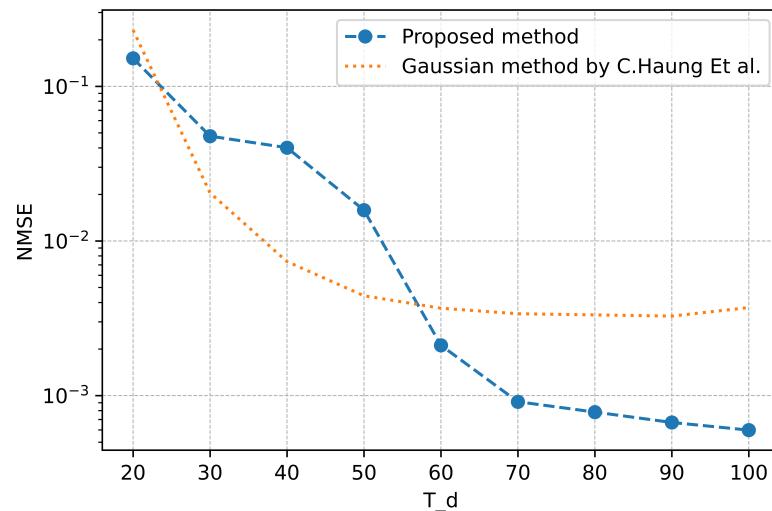


Figure 6.5: NMSE versus T_d with $T_p = 16$, number of IRS elements $N = 32$, and for a 1×8 SIMO system

Table 6.3: Parameters used for comparison

Symbols	Description
T_d	50
N	32
n_{rx}	1
n_{tx}	8
Modulation scheme(M)	QAM (4)
SNR	20 dB
Convergence threshold ϵ	10^{-1}

6.3 Superimposed method

This section discusses the simulation results obtained while applying the superimposed method. Figure 6.6 shows the graph for NMSE versus T for the MIMO-IRS system using the superimposed protocol. This graph illustrates an intriguing and expected result: the NMSE gets more accurate as the value of α rises. This results from a larger α value suggesting a bigger power distribution among the pilot symbols than the data symbols. The parameters used for this experiment are given in Table 6.4

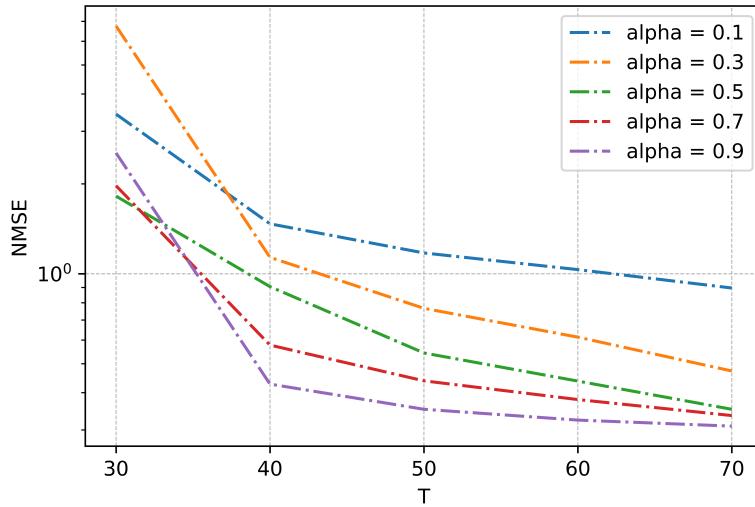


Figure 6.6: NMSE vs T for $N = 15$ and for a 2×4 MIMO system

Table 6.4: Parameters used for Superimposed method

Symbols	Description
N	15
n_{rx}	4
n_{tx}	2
Modulation scheme(M)	QAM (4)
SNR	20 dB
Convergence threshold ϵ	10^{-1}

6.4 Summary

One of the key takeaways from this chapter is that the proposed algorithm performs better than the method presented by authors of [17]. Additionally, we showcased various results for the proposed protocols and analyzed them. A brief discussion regarding the time complexity of the proposed algorithms is presented. In the following chapter, we discuss the conclusions and possible future research.

Chapter 7

Conclusions and Future Work

In this thesis, we proposed an EM-based semi-blind channel estimation for a MIMO-IRS-aided communication system. An intriguing feature of this was that we worked mainly with discrete constellations, particularly 4-QAM, despite its inherent computational complexity. We used two data-transmission protocols: (i) non-superimposed protocol, a traditional pilot-data structure, and (ii) superimposed protocol, where pilot and data symbols are transmitted simultaneously. For the non-superimposed method, we derived the equations necessary to compute the estimate of the channels. Nevertheless, this method involves higher computational complexity. To mitigate this complexity, we introduced two computationally efficient methods: (i) log-max approximation and (ii) detection based E-step. In the latter, we used detectors such as a ZF, MMSE, and SDFC. The results obtained in Chapter 6 affirm that the approximation methods are indeed computationally efficient compared to the proposed non-superimposed method without approximations. Furthermore, it is evident from the results that these methods (log-max approximation and detection based E-step) reduce the complexity of the proposed method at the cost of channel estimate accuracy. Moreover, in the case of the superimposed method, we derived equations similar to the non-superimposed method, and the results show that the accuracy

obtained is not optimal when compared to the superimposed method. This performance could be because the superimposed method introduces pilot-data interference, thereby affecting the channel estimate accuracy. Finally, the results obtained from the simulations show that the proposed method provides more accurate channel estimates than the method provided by authors of [17].

There are various potential research directions for this thesis. Firstly, a very significant extension would be working on achieving semi-blind channel estimation for the MIMO-IRS system with the assumption of Gaussian distributed data (a brief discussion regarding the challenges that occur with this assumption is provided in Appendix A). Secondly, this thesis can be extended for full-duplex semi-blind channel estimation for the MIMO-IRS system, allowing transmission and reception simultaneously and further improving system performance. Moreover, implementing a deep learning model to estimate the conditional probability of data given received signal which could potentially reduce the complexity of the system. Finally, an exciting area to explore could be investigating the EM-based semi-blind channel estimation of MIMO intelligent omni-surface-aided communication systems.

Appendix A

Gaussian distribution: MIMO-IRS

In this section, we discuss the limitations that occur while assuming Gaussian distributed data for a MIMO-IRS system. The closest research that was performed was for the MISO-IRS system with the assumption of a Gaussian distributed data [17]. So let us first look at the system model to perform this study and draw the conclusions towards the end of this appendix. Let us further vectorize (4.3), (4.6) and we would get the following,

$$\mathbf{y}_{p,t} \stackrel{(a)}{=} \mathbf{H}\mathbf{z}_{p,t} + \mathbf{n}_{p,t} \quad (\text{A.1})$$

$$\mathbf{y}_{d,t} \stackrel{(a)}{=} \mathbf{H}\mathbf{z}_{d,t} + \mathbf{n}_{d,t}, \quad (\text{A.2})$$

where $\mathbf{H} = \mathbf{h}^T \otimes \mathbf{I}_{n_{rx}} \in \mathbb{C}^{n_{rx} \times n_{rx}^2 n_{tx} N}$, $\mathbf{z}_{p,t} = \text{vec}(\mathbf{Z}_{p,t}) = \text{vec}(\boldsymbol{\psi}_t^T \otimes \mathbf{x}_{p,t}^T \otimes \mathbf{I}_{n_{rx}}) \in \mathbb{C}^{n_{rx}^2 n_{tx} N}$, $\mathbf{n}_{p,t} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{n_{rx}})$ with (a) following property 4. Combining the data for T_d and T_p time period, we have

$$\mathbf{y}_p = \begin{bmatrix} \mathbf{y}_{p,1} \\ \vdots \\ \mathbf{y}_{p,T_p} \end{bmatrix} = \mathbf{H} \begin{bmatrix} \mathbf{z}_{p,1} \\ \vdots \\ \mathbf{z}_{p,T_p} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{p,1} \\ \vdots \\ \mathbf{n}_{p,T_p} \end{bmatrix} = \mathbf{H}\tilde{\mathbf{z}}_p + \mathbf{w}_p, \quad (\text{A.3})$$

$$\mathbf{y}_d = \begin{bmatrix} \mathbf{y}_{d,T_p+1} \\ \vdots \\ \mathbf{y}_{d,T_d} \end{bmatrix} = \mathbf{H} \begin{bmatrix} \mathbf{z}_{d,T_p+1} \\ \vdots \\ \mathbf{z}_{d,T_d} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{d,T_p+1} \\ \vdots \\ \mathbf{n}_{d,T_d} \end{bmatrix} = \mathbf{H}\tilde{\mathbf{z}}_d + \mathbf{w}_d \quad (\text{A.4})$$

The data symbols $\mathbf{x}_{d,t}$ are assumed to follow complex normal distribution with 0 mean and $\sigma_x^2 \mathbf{I}_{n_{tx}}$. The mean of $\mathbf{z}_{d,t}$ is 0 and the covariance of $\mathbf{z}_{d,t}$ is calculated as follows

$$\begin{aligned} \Sigma_t &= \mathbb{E}\{\mathbf{z}_{d,t}\mathbf{z}_{d,t}^H\} = \mathbb{E}\{\text{vec}(\boldsymbol{\psi}_t^T \otimes \mathbf{x}_{d,t}^T \otimes \mathbf{I}_{n_{rx}}) \text{vec}(\boldsymbol{\psi}_t^T \otimes \mathbf{x}_{d,t}^T \otimes \mathbf{I}_{n_{rx}})^H\} \\ &\stackrel{(a)}{=} \mathbb{E}\{\left(\mathbf{PQ}(\text{vec} \boldsymbol{\psi}_t^T \otimes \text{vec} \mathbf{x}_{d,t}^T \otimes \text{vec} \mathbf{I}_{n_{rx}})\right) \left(\mathbf{PQ}(\text{vec} \boldsymbol{\psi}_t^T \otimes \text{vec} \mathbf{x}_{d,t}^T \otimes \text{vec} \mathbf{I}_{n_{rx}})\right)^H\} \\ &\stackrel{(b)}{=} \mathbb{E}\{\left(\mathbf{PQ}(\boldsymbol{\psi}_t \otimes \mathbf{x}_{d,t} \otimes \text{vec} \mathbf{I}_{n_{rx}})(\boldsymbol{\psi}_t \otimes \mathbf{x}_{d,t} \otimes \text{vec} \mathbf{I}_{n_{rx}})^H \mathbf{Q}^H \mathbf{P}^H\right)\} \\ &\stackrel{(c)}{=} \mathbb{E}\{\left(\mathbf{PQ}(\boldsymbol{\psi}_t \otimes \mathbf{x}_{d,t} \otimes \text{vec} \mathbf{I}_{n_{rx}})(\boldsymbol{\psi}_t^H \otimes \mathbf{x}_{d,t}^H \otimes (\text{vec} \mathbf{I}_{n_{rx}})^H) \mathbf{Q}^H \mathbf{P}^H\right)\} \\ &\stackrel{(d)}{=} \mathbb{E}\{\mathbf{PQ}(\boldsymbol{\psi}_t \boldsymbol{\psi}_t^H \otimes \mathbf{x}_{d,t} \mathbf{x}_{d,t}^H \otimes \text{vec} \mathbf{I}_{n_{rx}} (\text{vec} \mathbf{I}_{n_{rx}})^H) \mathbf{Q}^H \mathbf{P}^H\} \\ &= \mathbf{PQ}(\boldsymbol{\psi}_t \boldsymbol{\psi}_t^H \otimes \mathbb{E}\{\mathbf{x}_{d,t} \mathbf{x}_{d,t}^H\} \otimes \text{vec} \mathbf{I}_{n_{rx}} (\text{vec} \mathbf{I}_{n_{rx}})^H) \mathbf{Q}^H \mathbf{P}^H \\ &\stackrel{(e)}{=} \sigma_x^2 \mathbf{PQ}(\boldsymbol{\psi}_t \boldsymbol{\psi}_t^H \otimes \mathbf{I}_{n_{tx}} \otimes \text{vec} \mathbf{I}_{n_{rx}} (\text{vec} \mathbf{I}_{n_{rx}})^H) \mathbf{Q}^H \mathbf{P}^H \end{aligned} \quad (\text{A.5})$$

where $\mathbf{P} = (\mathbf{I}_N \otimes \mathbf{K}^{(1,n_{tx}n_{rx})} \otimes \mathbf{I}_{n_{rx}})$, $\mathbf{Q} = (\mathbf{I}_{Nn_{tx}} \otimes \mathbf{K}^{(1,n_{rx})} \otimes \mathbf{I}_{n_{rx}})$, $\mathbf{K}^{(p,r)}$ is the commutation matrix with $mn \times mn$ dimensions, (a), (b), (c), (d), and (e) use the following property 8, 9, 5, 10 and the data is assumed to be Gaussian distributed, (i.e) $\mathbb{E}\{\mathbf{x}_{d,t} \mathbf{x}_{d,t}^H\} = \sigma_x^2 \mathbf{I}_{n_{tx}}$ respectively. Now, by carefully looking at the covariance of $\mathbf{z}_{d,t}$ in (A.5), we can say it is a very sparse matrix and is not invertible, as we have Kronecker products with the identity matrix. Even though the hidden variable is $\mathbf{x}_{d,t}$, we can't directly access it since they are related by $\mathbf{z}_{d,t} = \text{vec}(\boldsymbol{\psi}_t^T \otimes \mathbf{x}_{d,t}^T \otimes \mathbf{I}_{n_{rx}})$. Hence, we find the covariance of $\mathbf{z}_{d,t}$ rather than $\mathbf{x}_{d,t}$. However, if we move forward with this covariance matrix and apply the EM algorithm, we ideally require the covariance of $\mathbf{z}_{d,t}$ to be invertible. From (A.2), it can be noted that the channel matrix \mathbf{H} is tall and sparse, whereas $\mathbf{z}_{d,t}$ is short and sparse matrix thereby leading to degrees of freedom mismatch. Therefore, assuming Gaussian distribution data for the MIMO-IRS

system and applying a semi-blind approach based on the EM algorithm has limitations.

Bibliography

- [1] Q. Wu and R. Zhang, “Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network,” *IEEE Comm. Magazine*, vol. 58, no. 1, pp. 106–112, 2020.
- [2] H. A. M. Alwazani, *Channel estimation and user association for distributed intelligent reflecting surfaces assisted MISO communications*. PhD thesis, University of British Columbia, 2021.
- [3] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretyakov, “Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead,” *IEEE journal on selected areas in communications*, vol. 38, no. 11, pp. 2450–2525, 2020.
- [4] Z. Qadir, K. N. Le, N. Saeed, and H. S. Munawar, “Towards 6G Internet of Things: Recent advances, use cases, and open challenges,” *ICT Express*, vol. 9, no. 3, pp. 296–312, 2023.
- [5] K. Anoh, C. H. See, Y. Dama, R. A. Abd-Alhameed, and S. Keates, “6G Wireless Communication Systems: Applications, Opportunities and Challenges,” *Future Internet*, vol. 14, no. 12, p. 379, 2022.

- [6] Q. Wu and R. Zhang, “Intelligent Reflecting Surface enhanced wireless network: Joint active and passive beamforming design,” in *Proc. of 2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2018.
- [7] W. Long, R. Chen, M. Moretti, W. Zhang, and J. Li, “A promising technology for 6G wireless networks: Intelligent reflecting surface,” *Journal of Communications and Information Networks*, vol. 6, no. 1, pp. 1–16, 2021.
- [8] A. I. Salameh and M. El Tarhuni, “From 5G to 6G—challenges, technologies, and applications,” *Future Internet*, vol. 14, no. 4, p. 117, 2022.
- [9] Z. Wang, L. Liu, and S. Cui, “Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6607–6620, 2020.
- [10] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, “Deep residual network empowered channel estimation for IRS-assisted multi-user communication systems,” in *ICC 2021-IEEE International Conference on Communications*, pp. 1–7, IEEE, 2021.
- [11] S. Jung, J.-W. Lee, and C. Lee, “RSS-based channel estimation for IRS-aided wireless energy transfer system,” *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14860–14873, 2021.
- [12] D. Mishra and H. Johansson, “Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4659–4663, IEEE, 2019.
- [13] B. Zheng, C. You, W. Mei, and R. Zhang, “A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless

- communications," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1035–1071, 2022.
- [14] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518–522, 2019.
- [15] H. Guo and V. K. N. Lau, "Uplink Cascaded Channel Estimation for Intelligent Reflecting Surface Assisted Multiuser MISO Systems," *IEEE Transactions on Signal Processing*, vol. 70, pp. 3964–3977, 2022.
- [16] An, Jiancheng and Wang, Li and Xu, Chao and Gan, Lu and Hanzo, Lajos, "Optimal pilot power based channel estimation improves the throughput of intelligent reflective surface assisted systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16202–16206, 2020.
- [17] C. Huang, J. Xu, W. Zhang, W. Xu, and D. W. K. Ng, "Semi-Blind Channel Estimation for RIS-Assisted MISO Systems Using Expectation Maximization," *IEEE Transactions on Vehicular Technology*, 2022.
- [18] A. S. Alwakeel and A. Elzanaty, "Semi-Blind Channel Estimation for Intelligent Reflecting Surfaces in Massive MIMO Systems," *IEEE Access*, vol. 10, pp. 127783–127797, 2022.
- [19] Z.-Q. He, H. Liu, X. Yuan, Y.-J. A. Zhang, and Y.-C. Liang, "Semi-Blind Channel Estimation for RIS-Aided Massive MIMO: A Trilinear AMP Approach," in *2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 2822–2827, IEEE, 2021.
- [20] G. T. de Araújo, A. L. de Almeida, R. Boyer, and G. Fodor, "Semi-Blind Joint Channel and Symbol Estimation for IRS-Assisted MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 71, pp. 1184–1199, 2023.

- [21] J. Du, X. Luo, X. Li, M. Zhu, K. M. Rabie, and F. Kara, "Semi-Blind Joint Channel Estimation and Symbol Detection for RIS-Empowered Multiuser mmWave Systems," *IEEE Communications Letters*, vol. 27, no. 1, pp. 362–366, 2022.
- [22] G. T. de Araújo, P. R. Gomes, A. L. de Almeida, G. Fodor, and B. Makki, "Semi-blind joint channel and symbol estimation in IRS-assisted multiuser MIMO networks," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1553–1557, 2022.
- [23] I. Santamaria, P. M. Crespo, C. Lameiro, and P. J. Schreier, "Information-Theoretic Analysis of a Family of Improper Discrete Constellations," *Entropy*, vol. 20, no. 1, 2018.
- [24] C. Aldana, E. de Carvalho, and J. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3280–3292, 2003.
- [25] S. Abdallah, *Spectrally-efficient approaches to channel estimation for amplify-and-forward two-way relay networks*. PhD thesis, McGill University, 2013.
- [26] L. Liang, M. Qi, J. Yang, X. Shen, J. Zhai, W. Xu, B. Jin, W. Liu, Y. Feng, C. Zhang, *et al.*, "Anomalous terahertz reflection and scattering by flexible and conformal coding metamaterials," *Advanced optical materials*, vol. 3, no. 10, pp. 1374–1380, 2015.
- [27] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, 2019.
- [28] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three myths and two critical questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020.
- [29] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Transactions on Communications*, 2021.

- [30] G. Levin and S. Loyka, “Amplify-and-forward versus decode-and-forward relaying: Which is better?,” in *22th International Zurich seminar on communications (IZS)*, Eidgenössische Technische Hochschule Zürich, 2012.
- [31] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, *et al.*, “Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.
- [32] E. Björnson, Ö. Özdogan, and E. G. Larsson, “Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?,” *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2019.
- [33] A. Bazrafkan, M. Poposka, Z. Hadzi-Velkov, and N. Zlatanov, “A Simple Single RF-Chain Multi-Antenna Full-Duplex Relay Can Outperform An Intelligent Reflecting Surface,” *arXiv preprint arXiv:2104.07980*, 2021.
- [34] X. Ying, U. Demirhan, and A. Alkhateeb, “Relay aided intelligent reconfigurable surfaces: Achieving the potential without so many antennas,” *arXiv preprint arXiv:2006.06644*, 2020.
- [35] Z. Sun, X. Wang, S. Feng, X. Guan, F. Shu, and J. Wang, “Pilot optimization and channel estimation for two-way relaying network aided by IRS with finite discrete phase shifters,” *IEEE Transactions on Vehicular Technology*, 2023.
- [36] K. Zhi, C. Pan, G. Zhou, H. Ren, M. Elkashlan, and R. Schober, “Is RIS-aided massive MIMO promising with ZF detectors and imperfect CSI?,” *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 10, pp. 3010–3026, 2022.

- [37] C. You, B. Zheng, and R. Zhang, “Intelligent reflecting surface with discrete phase shifts: Channel estimation and passive beamforming,” in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2020.
- [38] N. K. Kundu and M. R. Mckay, “Large intelligent surfaces with channel estimation overhead: Achievable rate and optimal configuration,” *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 986–990, 2021.
- [39] J. Lin, G. Wang, R. Fan, T. A. Tsiftsis, and C. Tellambura, “Channel estimation for wireless communication systems assisted by large intelligent surfaces,” *arXiv preprint arXiv:1911.02158*, 2019.
- [40] P. Wang, J. Fang, H. Duan, and H. Li, “Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems,” *IEEE Signal Processing Letters*, vol. 27, pp. 905–909, 2020.
- [41] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, “Channel estimation for RIS-empowered multi-user MISO wireless communications,” *IEEE Transactions on Communications*, 2021.
- [42] S. Eddine Zegrar, L. Afeef, and H. Arslan, “Reconfigurable intelligent surface (RIS): Eigenvalue Decomposition-Based Separate Channel Estimation,” in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2021.
- [43] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, “Deep channel learning for large intelligent surfaces aided mm-wave massive MIMO systems,” *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1447–1451, 2020.

- [44] S. Liu, M. Lei, and M.-J. Zhao, “Deep Learning Based Channel Estimation for Intelligent Reflecting Surface Aided MISO-OFDM Systems,” in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pp. 1–5, IEEE, 2020.
- [45] Z. Li, Z. Chen, X. Ma, and W. Chen, “Channel Estimation for Intelligent Reflecting Surface Enabled Terahertz MIMO Systems: A Deep Learning Perspective,” in *2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, pp. 75–79, IEEE, 2020.
- [46] S. Gao, P. Dong, Z. Pan, and G. Y. Li, “Deep Multi-Stage CSI Acquisition for Reconfigurable Intelligent Surface Aided MIMO Systems,” *IEEE Communications Letters*, vol. 25, no. 6, pp. 2024–2028, 2021.
- [47] S. Yin, Y. Li, Y. Tian, and R. Yu, “Intelligent Reflecting Surface Enhanced Wireless Communications with Deep-learning-based Channel Prediction,” *IEEE Transactions on Vehicular Technology*, 2021.
- [48] H. Ren, C. Pan, L. Wang, W. Liu, Z. Kou, and K. Wang, “Long-term CSI-based design for RIS-aided multiuser MISO systems exploiting deep reinforcement learning,” *IEEE Communications Letters*, vol. 26, no. 3, pp. 567–571, 2022.
- [49] W. Xie, J. Xiao, P. Zhu, and C. Yu, “Multi-Task Learning-Based Channel Estimation for RIS Assisted Multi-User Communication Systems,” *IEEE Communications Letters*, 2021.
- [50] T. L. Jensen and E. De Carvalho, “An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5000–5004, IEEE, 2020.

- [51] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.
- [52] S. Liu, G. Trenkler, *et al.*, "Hadamard, Khatri-Rao, Kronecker and other matrix products," *International Journal of Information and Systems Sciences*, vol. 4, no. 1, pp. 160–177, 2008.
- [53] Y. S. Yun and C. Kang, "Some Results on Kronecker Products and Commutation Matrices," *East Asian mathematical journal*, vol. 29, no. 3, pp. 259–268, 2013.
- [54] R. A. Horn, R. A. Horn, and C. R. Johnson, *Topics in matrix analysis*. Cambridge university press, 1994.
- [55] X. Hou, C. Yang, and B. K. Lau, "Impact of nonorthogonal training on performance of downlink base station cooperative transmission," *IEEE transactions on vehicular technology*, vol. 60, no. 9, pp. 4633–4639, 2011.
- [56] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 787–801, 2014.
- [57] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *2010 IEEE international conference on communications*, pp. 1–5, IEEE, 2010.
- [58] Y. Jin, J. Zhang, S. Jin, and B. Ai, "Channel estimation for cell-free mmWave massive MIMO through deep learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10325–10329, 2019.
- [59] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, 2017.

- [60] E. G. Larsson and J. Jalden, “Fixed-Complexity Soft MIMO Detection via Partial Marginalization,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3397–3407, 2008.
- [61] M. A. Albreem, M. Juntti, and S. Shahabuddin, “Massive MIMO detection techniques: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.
- [62] D. J. Costello, “Fundamentals of wireless communication (tse, d. and viswanath, P.)[book review],” *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 919–920, 2009.
- [63] D. Persson and E. G. Larsson, “Partial marginalization soft MIMO detection with higher order constellations,” *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 453–458, 2010.
- [64] N. Garg, H. Ge, and T. Ratnarajah, “Generalized superimposed training scheme in IRS-assisted cell-free massive MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 1157–1171, 2022.
- [65] Y. Zhang, X. Qiao, L. Yang, and H. Zhu, “Superimposed pilots are beneficial for mitigating pilot contamination in cell-free massive MIMO,” *IEEE Communications Letters*, vol. 25, no. 1, pp. 279–283, 2020.
- [66] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, “On superimposed pilot for channel estimation in multicell multiuser MIMO uplink: Large system analysis,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1492–1505, 2015.
- [67] D. Verenzuela, E. Björnson, and L. Sanguinetti, “Spectral and energy efficiency of superimposed pilots in uplink massive MIMO,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7099–7115, 2018.