Spectrally-efficient Approaches to Channel Estimation for Amplify-and-forward Two-way Relay Networks

Saeed Abdallah



Department of Electrical & Computer Engineering McGill University Montreal, Canada

August 2013

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

© 2013 Saeed Abdallah

Abstract

Relay networks constitute one of the key technologies that are being developed for use in next generation wireless systems. In relay networks, the communication between the source and the destination is aided by dedicated nodes (relays) that convey the source's message to the destination. The use of multiple relays makes it possible to implement a distributed form of spatial diversity. The use of relays improves the coverage, capacity and reliability in the network.

Relay networks were initially designed to support only unidirectional communication. However, two-way relay networks (TWRNs) have recently been proposed to support bidirectional communication and have attracted the attention of many researchers because of their high spectral efficiency. In particular, TWRNs employing the amplify-and-forward (AF) protocol are appealing because of the minimal processing requirements at the relay. Effective operation of AF TWRNs requires accurate channel state information for selfinterference cancellation and coherent decoding. This information is also needed in several important applications such as beamforming, relay selection and power allocation.

The majority of works on channel estimation for AF TWRNs follow the training-based approach, which requires the transmission of pilots known to both terminals. Despite its robustness and simplicity, the training-based approach consumes much needed bandwidth resources, which undermines the spectral efficiency of TWRNs. Blind channel estimation avoids the costly training burden by relying only on the received data samples to estimate the channel parameters. Another alternative approach is semi-blind estimation, a hybrid of blind and training-based approaches. Semi-blind estimation requires fewer pilots than training-based estimation by utilizing data samples in addition to the pilots.

The main objective of this thesis is to investigate blind and semi-blind channel estimation for AF TWRNs as a means for achieving substantially better tradeoffs between accuracy and spectral efficiency than possible using the training-based approach.

In the first part of the thesis, we consider blind channel estimation for flat-fading channel conditions. Using the deterministic maximum likelihood (DML) approach, we propose new algorithms for blind channel estimation in AF TWRNs that employ constant-modulus signalling. Assuming M-PSK modulation, we prove that the proposed estimators are consistent and approach the true channel with high probability at high SNR. Using simulations, we show that the DML estimator offers a superior tradeoff between accuracy and spectral efficiency than the pilot-based LS estimator. The corresponding Cramer-Rao bound (CRB) is also derived.

Still within the context of flat-fading channels, the second part of the thesis focuses on semi-blind channel estimation. Assuming nonreciprocal channels, we derive the exact CRB for semi-blind channel estimation in AF TWRNs that employ square QAM. The derived bound, which has not been reported before in the context of TWRNs, is based on the true likelihood function that incorporates the exact statistics of the transmitted data symbols. Using the new bound, we show that the training overhead can be significantly reduced by employing semi-blind estimation, as even a limited number of data samples can lead to substantial improvements in estimation accuracy. To demonstrate the achievability of these gains, we derive an expectation maximization (EM)-based semi-blind algorithm that performs very closely to the derived CRB and requires only a small number of iterations to converge.

In the last part of the thesis, we consider semi-blind channel estimation for OFDM-based TWRNs operating in frequency selective channel conditions. In contrast to previous works, we focus on nonreciprocal channels as this is a more realistic assumption in the frequency selective scenario. To assist in the estimation of the individual channels, superimposed training is adopted at the relay. Our proposed semi-blind estimation algorithm is based on the Gaussian ML approach. We also derive the CRB, and we design the pilot vectors of the terminals and relay to optimize estimation performance. Our simulations show that the proposed method provides significant improvements in estimation accuracy even with a limited number of OFDM data symbols.

Overall, the work presented in this thesis demonstrates that blind and semi-blind approaches to channel estimation are viable and practical alternatives to the training-based approach as they can provide substantially better tradeoffs between accuracy and spectral efficiency at an affordable computational cost.

Sommaire

Les réseaux à relais constituent l'une des technologies clé de la prochaine génération des systèmes de communication sans fil. Dans les réseaux à relais, la communication entre une source et une destination est assistée par des nœuds dédiés (ou relais) qui relayent le message de la source jusqu'à la destination. L'usage de plusieurs relais permet de mettre en œuvre une forme distribuée de diversité spatiale. L'utilisation de relais améliore la couverture, la capacité et la fiabilité des réseaux.

Bien que les réseaux à relais aient d'abord été conçus pour une communication unidirectionnelle, les réseaux à relais bidirectionnels (*two-way relay networks* ou TWRNs) ont été récemment proposés et ont suscité l'intérêt de nombreux chercheurs à cause de leur grande efficacité spectrale. En particulier, les TWRNs utilisant le protocole de relayage amplifier-et-transférer (*amplify-and-forward* ou AF) sont attrayants à cause de leurs faibles exigences de traitement aux relais. Le bon fonctionnent des AF TWRNs nécessite une étape d'estimation précise du canal pour la suppression d'auto-interférence et pour le décodage cohérent. Cette étape est aussi essentielle pour la formation de faisceaux, la sélection de relais et l'allocation de puissance.

La plupart des travaux sur l'estimation du canal pour les AF TWRNs suivent l'approche basée sur l'entrainement, ce qui nécessite la transmission de pilotes connus des deux terminaux. Malgré sa robustesse et sa simplicité, cette approche consomme de précieuses ressources de communication afin de transmettre les pilotes, ce qui compromet l'efficacité spectrale des TWRNs. L'estimation aveugle évite ce fardeau en se fiant seulement sur les données reçues pour estimer les paramètres du canal. Une autre option est l'estimation semi-aveugle, une approche hybride de l'approche aveugle et de celle basée sur l'entrainement. L'estimation semi-aveugle nécessite moins de pilotes que l'approche basée sur l'entrainement puisqu'elle utilise aussi des échantillons de données.

L'objectif de cette thèse est d'examiner l'estimation aveugle et semi-aveugle du canal pour les AF TWRNs afin d'obtenir un meilleur compromi entre l'efficacité spectrale et la précision que que celui offert par l'approche basée sur l'entrainement.

Dans la première partie de cette thèse, nous considérons l'estimation aveugle du canal pour des conditions d'évanouissement plat. Employant le principe du maximum de vraisemblance déterminée (*deterministic maximum likelihood* ou DML), nous proposons de nouveaux algorithmes pour l'estimation et la détection aveugle et conjointe du canal pour les AF TWRNs qui utilisent une signalisation à module constant. Supposant une modulation M-PSK, nous prouvons que les estimateurs proposés sont consistants et se rapprochent du vrai canal avec une grande probabilité lorsque le rapport signal sur bruit est élevé. En utilisant des simulations, nous montrons que le DML offre un meilleur compromis entre l'efficacité spectrale et la précision que l'estimateur LS employant des pilotes. La borne de Cramér-Rao (*Cramer-Rao bound* ou CRB) correspondante est aussi établie.

Toujours en rapport avec l'évanouissement plat, la deuxième partie de cette thèse porte sur l'estimation semi-aveugle du canal.

En supposant des canaux non-réciproques, nous établissons la CRB exacte pour l'estimation semi-aveugle du canal pour les AF TWRNs qui utilisent une constellation QAM carrée. Cette borne n'a jamais été déterminée auparavant dans le contexte des TWRNs. Elle est basée sur la vraie fonction de vraisemblance qui tient compte des statistiques exactes des symboles de données transmis. En utilisant cette nouvelle borne, nous montrons que la complexité attribuable à l'entrainement peut être considérablement réduite en employant un estimation semi-aveugle car même une quantité limitée de données peut engendrer une amélioration substantielle de la précision de l'estimation. Afin de montrer que ces gains sont réalisables, nous concevons un algorithme semi-aveugle basé sur la méthode d'espérancemaximisation (EM), et nous montrons que cet algorithme se rapproche tout près de la CRB et qu'il converge en un petit nombre d'itérations.

Dans la dernière partie de cette thèse, nous considérons l'estimation semi-aveugle pour les TWRNs qui sont basés sur l'OFDM et qui opèrent dans un environnement sélectif en fréquence. Contrairement aux travaux antérieurs, nous nous penchons sur des canaux nonréciproques, ceci étant une hypothèse plus réaliste dans un scénario sélectif en fréquence. Pour faciliter l'estimation des canaux individuels, un entrainement superposé est adopté aux relais. L'algorithme semi-aveugle que nous proposons est basé sur le principe ML gaussien. Nous établissons la CRB, et nous concevons des séquences pilotes pour les terminaux et pour les relais de façon à optimiser la performance de l'estimation. Nos simulations révèlent qu'en utilisant seulement un nombre limité de symboles de données OFDM, la méthode proposée fournit une nette amélioration de la précision de l'estimation.

Somme toute, le travail présenté dans cette thèse établit que les approches aveugles et semi-aveugles de l'estimation du canal sont des solutions de rechange pratiques et viables pour l'approche basée sur l'entrainement car elles offrent un meilleur compromis entre l'efficacité spectrale et la précision à un cout de calcul abordable.

Acknowledgments

In the name of God, the most merciful and compassionate,

First and foremost, I thank God the Almighty for all His blessings and for making it possible for me to complete this journey through His pure bounty and generosity.

I would like to sincerely thank Prof. Ioannis Psaromiligkos for being my supervisor and mentor throughout my graduate studies. I will ever be grateful to him for his excellent guidance and his unrelenting encouragement and support throughout this journey. Working with Prof. Psaromiligkos has been a thoroughly enjoyable and invaluable learning experience which has greatly contributed to my professional development. I am deeply inspired by his patience, professionalism and devotion to research.

I would also like to thank my PhD Committee members, Prof. Benoit Champagne and Prof. Warren Gross for their insights and advice which contributed positively to my research. I am also grateful to the members of my PhD defense committee for all their insightful feedback on my work.

I am deeply grateful to my colleagues and lab mates, François Côte, Brian Lawrence, Ali Shahrad, Ardvan Salehi, Stefanos Koskinas, Mahmoud Watfa, Jad Kabbara, and Ahmad Rashid for all the beautiful times we spent together. I am especially indebted to François Côte for all the valuable discussions that we had and the technical insights that he shared with me and for his French translation of the abstract of this thesis. I also want to thank Ali Shahrad for his help in proving Lemma 1 of this thesis.

During my time at McGill, I have come to know many sincere friends whose friendship I will always cherish. I want to thank my dear friends Omar Fawzi, Ahmad Shawqi, Kosai Alchaghoury, Mostafa Momen, Alaa al Kaadry, Mohammad Auais, Ahmad Abu El Haija, Faris Abu Risheh and Kusai Al Abdallah.

Last but not least, I want to thank my family to whom I am deeply indebted for their immense love, patience and support which has inspired me throughout my graduate studies. vi

Contents

1	Introduction		1
	1.1	Relay Networks	2
		1.1.1 Basic Concepts	2
		1.1.2 Two-way Relay Networks	3
	1.2	Two-way Relay Channel Estimation	4
		1.2.1 Individual and Cascaded Channels	5
		1.2.2 The Channel Reciprocity Assumption	6
	1.3	Previous Works	6
	1.4	Motivation	8
	1.5	Thesis Organization and Contributions	10
2	Bac	kground	15
	2.1	Maximum Likelihood Estimation	16
		2.1.1 ML Estimation in the Presence of Random Nuisance Parameters	16
		2.1.2 The Expectation Maximization Algorithm	18
	2.2	Cramer-Rao Bound	19
		2.2.1 The CRB in the Presence of Random Nuisance Parameters	20
	2.3	Multipath Fading	21
		2.3.1 Slow and Fast Fading	21
		2.3.2 Flat-fading and Frequency-selective fading Channels	22
	2.4	OFDM Systems	23
3	Blir	nd Channel Estimation for Nonreciprocal Flat-fading Channels under	
	\mathbf{MP}	SK modulation	27
	3.1	Introduction	27

	3.2	System Model	8
	3.3	Proposed Channel Estimation Algorithms)
		3.3.1 Deterministic ML Approach	0
		3.3.2 Constrained ML Approach for BPSK	2
		3.3.3 Gaussian ML Approach	4
	3.4	Asymptotic Behavior Analysis	õ
		3.4.1 Consistency of the DML Estimator	ő
		3.4.2 High SNR Behavior of the DML Estimator	8
		3.4.3 High SNR Behavior of the MCML Estimator $(M=2)$	9
		3.4.4 MSE Performance of the GML Estimator	0
		3.4.5 Cramer-Rao Bounds	0
	3.5	Simulation Results	2
	3.6	Conclusions $\ldots \ldots 44$	4
4	Blir	nd Channel Estimation for Reciprocal Flat-fading Channels under	
	\mathbf{MP}	PSK-modulation 49	9
	4.1	Introduction	9
	4.2	System Model	0
	4.3	Channel Estimation Algorithms	0
	4.4	Asymptotic Behavior Analysis	2
		4.4.1 High SNR Behavior of the two Estimators	2
		4.4.2 Consistency of Estimators	3
	4.5	Cramer-Rao Bounds	4
	4.6	Simulation Results	6
	4.7	Conclusions	7
5	Exa	act CRBs for Semi-blind Channel Estimation under Square QAM 59	9
-	5.1	Introduction	9
	5.2	System Model	1
	5.3	Cramer-Bao Bounds 6	2
	0.0	5.3.1 Exact Cramer-Rao Bound	2
		5.3.2 Modified Cramer-Bao Bound	2
	5.4	Simulation Besults	3
	5.5	Conclusions	5

Contents

6	$\mathbf{E}\mathbf{M}$	-based	Semi-blind Channel Estimation for Flat-fading Channels		81
	6.1	Introduction			81
	6.2	System	n Model		82
		6.2.1	Received Signal for Nonreciprocal Channels		83
		6.2.2	Received Signal for Reciprocal Channels		83
	6.3	Propo	sed Channel Estimation Algorithms		83
		6.3.1	The EM Algorithm for Nonreciprocal Channels		84
		6.3.2	The EM Algorithm for Reciprocal Channels		86
	6.4	CRB f	for the Reciprocal Scenario		90
	6.5	Simula	ation Results		93
	6.6	Conclu	usions		95
7	Sem	ni-blino	d Channel Estimation for OFDM-modulated AF TWRNs		101
	7.1	Introd	uction		101
	7.2	System	n Model		102
		7.2.1	Transmission at the Terminals		103
		7.2.2	Processing at the Relay		104
		7.2.3	Received Vectors at the Terminal		106
	7.3	Semi-l	blind Channel Estimation		107
		7.3.1	Proposed Algorithm		107
		7.3.2	Pilot-based Least Squares Estimation		109
		7.3.3	Design of the pilot vectors		110
	7.4	Crame	er-Rao Bounds		111
		7.4.1	The CRB for Pilot-based Estimation		112
		7.4.2	The CRB for Semi-blind Estimation		113
		7.4.3	Power Allocation at the Relay		115
	7.5	Simula	ation Results		115
	7.6	Conclu	usions		117
8	Con	clusio	ns and Future Work		121
	8.1	Summ	ary and Conclusions		121
	8.2	Direct	ions for Future Work		123

 $\mathbf{i}\mathbf{x}$

\mathbf{A}	Der	ivations of Select Proofs and Results from Chapter 3	127
	A.1	Derivation of Eq. (3.22)	127
	A.2	Proof of Lemma 1	128
	A.3	Proof of Lemma 2	130
	A.4	Proof of Lemma 3	132
В	Der	ivations of Select Proofs and Results from Chapter 4	135
	B.1	Proof of Lemma 4	135
	B.2	Steepest Descent Implementation	136
С	Der	ivations of Select Proofs and Results from Chapter 5	139
	C.1	Proof of Eqs. (5.30) and (5.31)	139
	C.2	Proof of Eq. (5.37) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots	140
	C.3	Proof of Eq. (5.43)	141
	C.4	Proof of Eq. (5.54)	144
	C.5	Derivation of Eqs. (5.65) and (5.66)	145
D	Der	ivations of Select Proofs and Results from Chapter 7	149
	D.1	BFGS Method	149
	D.2	Proof of Theorem 3	150
Re	efere	nces	153

List of Figures

1.1	Example of a relay network	2
1.2	AF TWRN operating in flat-fading channel conditions	4
2.1	A typical OFDM transceiver.	24
2.2	OFDM block with CP	25
3.1	MSE (a) vs SNR for the DML, GML and MCML algorithms (BPSK)	45
3.2	MSE (a) vs number of samples for the DML, GML and MCML algorithms	
	(BPSK)	45
3.3	MSE (a) vs SNR for DML and GML algorithms (QPSK)	46
3.4	MSE (a) vs number of samples for the DML and GML algorithms (QPSK).	46
3.5	MSE ($ b $) vs SNR for the DML and GML algorithms (QPSK)	47
3.6	SER vs SNR of DML and LS estimators (Case I)	47
3.7	SER vs SNR of DML and LS estimators (Case II)	48
4.1	MSE (a) vs SNR for the DML and MSEV algorithms (reciprocal channels,	
	QPSK)	57
4.2	Average number of SD and linesearch iterations (reciprocal channels, QPSK).	58
4.3	MSE (a) vs number of samples for the DML and MSEV (reciprocal channels,	
	QPSK)	58
5.1	Exact semi-blind and pilot-based CRBs (a) vs SNR	76
5.2	Exact semi-blind and pilot-based CRBs (b) vs SNR	77
5.3	Exact semi-blind CRB (a) vs number of samples	77
5.4	Exact semi-blind CRB (b) vs number of samples	78
5.5	Contour plot of ratio of semi-blind to pilot-based CRBs (a)	78

5.6	Contour plot of ratio of semi-blind to pilot-based CRBs (b)	79
6.1	MSE of EM algorithm vs SNR (nonreciprocal channels, 4 EM iterations)	96
6.2	MSE of EM algorithm vs SNR (reciprocal channels, 4 EM iterations)	97
6.3	MSE of EM algorithm vs number of samples (nonreciprocal channels, $10 EM$	
	iterations)	97
6.4	MSE of EM algorithm vs number of samples (reciprocal channels, 10 EM	
	iterations)	98
6.5	MSE of EM algorithm vs number of EM iterations (nonreciprocal channels,	
	15dB)	98
6.6	MSE of EM algorithm vs number of EM iterations (reciprocal channels, 15dB).	99
6.7	SER performance of the EM algorithm and the LS estimator plotted versus	
	SNR ($N = 36$, $L = 4$, 6 EM iterations)	99
7.1	OFDM frame structure.	103
7.2	Block diagram of OFDM-based AF TWRN	104
7.3	Illustration of superimposed training at the relay.	105
7.4	OFDM TWRNs: semi-blind and pilot-based CRBs vs power allocation factor.	118
7.5	OFDM TWRNs: MSE of the semi-blind and pilot-based estimators vs SNR.	119
7.6	Average number of BFGS iterations for convergence.	119
7.7	OFDM TWRNs: MSE of the semi-blind and pilot-based estimators vs num-	
	ber of data blocks.	120
7.8	OFDM TWRNs: MSE for the semi-blind and pilot-based estimators for	
	cascaded channels.	120

List of Acronyms

AF	Amplify and Foward
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase Shift Keying
CFO	Carrier Frequency Offset
CM	Constant Modulus
CML	Constrained Maximum Likelihood
CP	Cyclic Prefix
CPM	Continuous Phase Modulation
CRB	Cramer Rao Bound
DF	Decode and Forward
DFT	Discrete Fourier Transform
DML	Deterministic Maximum Likelihood
ECM	Expectation Conditional Maximization
EDGE	Enhanced Data Rates for GSM Evolution
EM	Expectation Maximization
FIM	Fisher Information Matrix
FIR	Finite Impulse Response
GML	Gaussian Maximum Likelihood
IBI	Inter-block Interference
IDFT	Inverse Discrete Fourier Transform
ISI	Inter-symbol Interference
LS	Least Squares
MCRB	Modified Cramer Rao Bound
MFIM	Modified Fisher Information Matrix

MIMO	Multiple-input Multiple-output
ML	Maximum Likelihood
MPSK	M-ary Phase Shift Keying
MSE	Mean Squared Error
MSEV	Minimum Sample Envelope Variance
OFDM	Orthogonal Frequency Division Multiplexing
OWRN	One-way Relay Network
PMF	Probability Mass Function
PDF	Probability Density Function
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RV	Random Variable
SCCP	Single Carrier Cyclic Prefix
SD	Steepest Descent
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
TDD	Time Division Duplexing
TWRN	Two-way Relay Network
3GPP-LTE	Third Generation Partnership Project - Long Term Evolution

Mathematical Notation

J	$\sqrt{-1}$
x^*	Conjugate of x
$\Re\{x\}$	Real part of x
$\Im\{x\}$	Imaginary part of x
$\operatorname{sgn}(\alpha)$	Sign of α (for real α)
x	The magnitude of x
$\angle x$	The phase of x
mod	The modulo operator
$x \perp y$	$\angle x - \angle y = \frac{\pi}{2} \mod \pi$
\boldsymbol{x}	The column vector \boldsymbol{x}
$\ m{x}\ $	The 2-norm of \boldsymbol{x}
x * y	The linear convolution of \boldsymbol{x} and \boldsymbol{y}
\boldsymbol{A}	The matrix \boldsymbol{A}
$[oldsymbol{A}]_{ij}$	The (i, j) -th element of \boldsymbol{A}
$oldsymbol{A}^T$	The transpose of \boldsymbol{A}
$oldsymbol{A}^H$	The Hermitian of \boldsymbol{A}
$\operatorname{tr}(\boldsymbol{A})$	The trace of \boldsymbol{A}
A	The determinant of \boldsymbol{A}
$oldsymbol{A}^{-1}$	The inverse of \boldsymbol{A}
$(oldsymbol{A})^\dagger$	Moore-Penrose Pseudo-inverse of \boldsymbol{A}
\boldsymbol{I}_N	The $N \times N$ identity matrix
$\mathcal{CCN}(\mu,\sigma^2)$	Circularly complex Normal random variable with mean μ and variance σ^2
$\mathcal{CCN}(oldsymbol{\mu}, oldsymbol{\Sigma})$	Circularly complex Normal random vector with mean μ and covariance Σ
$\mathbb{E}\left\{\cdot ight\}$	The statistical expectation operation

xvi

Chapter 1

Introduction

Wireless communications is currently experiencing an unprecedented growth that is driven by the ever increasing market demand for fast, uninterrupted access to information. The last few years have witnessed a phenomenal increase in wireless data traffic [1, 2]. This trend is expected to continue as an 18-fold increase in global data traffic is projected for the period 2011-2016 [2]. This has placed an immense strain on the currently deployed wireless infrastructure. In response, intense efforts are currently being made to design the next generation of wireless systems that will provide high data rates, extended coverage, enhanced system capacity as well as improved reliability.

Relaying [3,4] is one of the key technologies currently being developed for use in next generation wireless systems. In relay networks, the communication between the source and the destination is aided by specialized nodes (i.e., relays) that convey the source's message to the destination. Relay networks are capable of implementing a distributed form of spatial diversity by recruiting multiple relay nodes that collectively form a virtual antenna array. It is now widely accepted that the use of relays can improve coverage, throughput and reliability [4].

Relaying has also been adopted as a key feature in the 3GPP Long Term Evolution (LTE) standard for next generation wireless systems. Relay nodes (RNs) are used to support two-hop communication between the base station (eNB) and user equipment (UE). RNs have a lower transmission power than the eNBs and cover a smaller area, thus providing a layer of low-power nodes on top of the conventional base stations [5]. The deployment of RNs can extend the cell range and improve both coverage and capacity, especially at the cell edge. They can also be used for dead spot mitigation by filling coverage holes in the

macro network, which may be caused by large obstacles [6]. Relays are well suited for these purposes because of their deployment flexibility and hardware feasibility. In particular, they can be deployed without incurring high site acquisition and backhaul costs [7].

1.1 Relay Networks



Fig. 1.1 An example of a relay network with one source, one destination, and multiple relays.

1.1.1 Basic Concepts

A typical relay network consists of a source terminal, a destination terminal, and one or more relay nodes that aid the transmission between the source and the destination (see Fig. 1.1). Relay nodes are commonly assumed to operate in a half-duplex mode, i.e., they can either transmit or receive at any given time¹. Until recently, research has focused on relay networks that support only unidirectional communication, commonly referred to as one-way relay networks (OWRNs). In OWRNs, the communication between the source and the destination occurs in two phases. In the first phase, the source transmits to the relay, and in the second phase the relay forwards a processed version of the received signal to the destination. The relay processes the signal it receives according to one of several protocols. The main two relaying protocols are amplify-and-forward (AF) and decode-andforward (DF) [8]. According to the AF protocol, the relay simply amplifies the received signal and then forwards it to the destination. By amplifying the received signal, both the information bearing signal component and the noise are amplified. Nonetheless, this

¹Although it is possible to use full-duplex relays that can transmit and receive at the same time, their hardware implementation is much more costly as they require sophisticated signal isolation techniques to prevent interference between the transmitted and received signals.

protocol is appealing due to its simplicity and the minimal processing required at the relay which can be transparent to the modulation and coding employed by the source. In the DF protocol, the relay decodes the message in the received signal, re-encodes it into a new codeword and then transmits it to the destination. While the DF protocal avoids noise amplification, it obviously requires more complex processing at the relay which has to be aware of the modulation and coding employed by the source.

1.1.2 Two-way Relay Networks

The above-described OWRNs have been specifically developed for the scenario when only one of the two terminals has a message to transmit and acts as the source while the other terminal acts solely as a destination. It becomes spectrally inefficient to employ OWRNs, however, when bidirectional communication is needed, i.e., when both terminals have messages to transmit to each other. In particular, a single round of bidirectional exchange between the two terminals requires 4 time slots to be completed using OWRNs. Two-way relay networks (TWRNs) [9–12], which are the focus of this thesis, have recently been developed to achieve spectrally efficient bidirectional communication. Using TWRNs, bidirectional communication can be achieved over only two time slots, i.e., at twice the communication rate of OWRNs [8]. In the first time slot, both terminals simultaneously transmit to the relay. Employing either the DF or the AF protocol, the relay then processes the received signal made up of the superposition of the signals transmitted from the two terminals. It then broadcasts the processed signal to both terminals in the second time slot. Since each terminal knows its own transmitted message, it can use this knowledge to extract its intended message.

Applying the DF protocol in TWRNs requires complicated processing at the relay. The relay has to decode the received signals from both users and then combine the two decoded messages using a combining scheme such as superposition coding [8], or XORing [13]. To achieve this, the relay needs a high processing capacity, knowledge of the coding schemes employed at the two terminals as well as accurate knowledge of the channels from each terminal to the relay.

AF TWRNs, on the other hand, do not have such requirements and can be implemented using minimal processing at the relay. The relay does not have to decode the messages of the two sources; rather, it just amplifies the overall received signal and then broadcasts it to the terminals. In this case, the received signal at each terminal at the end of the broadcast



Fig. 1.2 AF TWRN operating in flat-fading channel conditions.

phase consists of a self-interference term containing the message originally transmitted by the terminal as well as the term containing the intended message, in addition to noise. Assuming perfect knowledge of the channel parameters at the terminals, each terminal can cancel out the self-interference term before decoding the signal of interest. Because of the simplicity of this approach, it has received the attention of researchers and has been the subject of many recent studies. For example, power allocation strategies for AF TWRNs have been developed in [14,15], while relay selection strategies have been proposed in [16,17]. For AF TWRNs with multiple relays, optimal beamforming strategies have been derived in [18, 19]. Furthermore, the achievable rates for AF TWRNs have been studied in [9]. In this thesis, we consider TWRNs that are based on the AF protocol. A TWRN employing the AF protocol is illustrated in Fig. 1.2, where \mathcal{T}_1 , \mathcal{T}_2 and \mathcal{R} denote the two terminals and the relay, respectively.

1.2 Two-way Relay Channel Estimation

The self-interference experienced by each terminal in AF TWRNs can be completely cancelled out when perfect knowledge of the channel parameters is available. In practice, however, the channel parameters have to be estimated. Hence, their accurate estimation is essential in order to minimize the impact of the residual self-interference and avoid the potential performance degradation. Accurate channel estimates are also essential to perform coherent decoding. Furthermore, they are needed in several applications such as relay selection, power allocation and distributed beamforming. In fact, the majority of works on these applications assume perfect knowledge of the channel [14–19]. While traditional estimation methods developed for point-to-point systems can readily be applied for channel estimation in DF TWRNs, the same does not hold for AF TWRNs because of the differences in the received signal structure.

The focus of this thesis is the development of highly accurate, spectrally efficient channel estimation algorithms for AF TWRNs at an affordable computational cost. Before describing our approach to this problem, we will discuss some assumptions regarding the AF TWRN channel that affect the formulation of the estimation problem. First there is the type of fading under consideration. In this thesis we consider two types of fading channels: 1) flat fading channels which occur when the delay spread of the channel is much smaller than the symbol period, and 2) frequency selective channels which occur when the delay spread is greater than the symbol period [20]. In the former case the channel can be modelled as a complex multiplicative coefficient and in the latter as a linear finite impulse response (FIR) filter. The two types of fading are discussed in more detail in Chapter 2. Other considerations that affect the formulation of the estimation problem include whether we are interested in estimating the end-to-end channel or the individual channels across each link in the network, as well as whether or not channel reciprocity is assumed to hold. These two issues are discussed below.

1.2.1 Individual and Cascaded Channels

For the flat-fading scenario, we denote by h_1 , g_1 , h_2 and g_2 the complex coefficients corresponding to the flat-fading channels across the links $\mathcal{T}_1 \to \mathcal{R}$, $\mathcal{T}_2 \to \mathcal{R}$, $\mathcal{R} \to \mathcal{T}_1$ and $\mathcal{R} \to \mathcal{T}_2$, respectively, as shown in Fig. 1.2. Similarly, for the frequency-selective scenario we denote by h_1 , g_1 , h_2 and g_2 the complex vectors of the FIR filter coefficients corresponding to the channels across the same links. As we can see from Fig. 1.2, the self-interference signal component in the received signal at \mathcal{T}_1 at the end of the second transmission phase experiences an overall fading due to the links $\mathcal{T}_1 \to \mathcal{R}$ and $\mathcal{R} \to \mathcal{T}_1$, while the information-bearing component experiences an overall fading due to the links $\mathcal{T}_2 \to \mathcal{R}$ and $\mathcal{R} \to \mathcal{T}_1$. Hence, in the flat-fading scenario, the overall self-interference channel is given by $b \triangleq h_2g_1$. In the frequency-selective scenario, the overall self-interference channel is obtained as² $a \triangleq h_2 * h_1$ and the overall information-bearing channel is obtained as² $a \triangleq h_2 * h_1$ and the overall links are often referred to as individual channels, while the parameters a, b, (or a and b) are referred to as the cascaded or composite channels. Knowledge of the cascaded channel

²We use x * y to denote the linear convolution between the vectors x and y.

parameters is sufficient for detection purposes [21]. On the other hand, knowledge of the individual channels can be useful in other applications such as beamforming.

1.2.2 The Channel Reciprocity Assumption

Most works on TWRNs consider time division duplexing (TDD). In this setting, it is very common to assume that the channels are reciprocal [15, 16, 18, 21, 22], i.e., that the channel from each terminal to the relay is equal to the channel from the relay back to the same terminal. The basis for this assumption is a basic principle in antenna theory that the two channel coefficients across the forward and reverse links between two antennas are the equal [23]. In practice, however, the mismatch between the RF front ends, including RF gains and baseband circuitry, ruins the reciprocity of the overall forward and reverse links despite the reciprocity of the radio propagation channel [23–27]. Factors such as carrier frequency offset, timing offset, sampling clock deviations, etc... all contribute to making the channels nonreciprocal [24], and special calibration would be required to restore reciprocity. Hence, channel reciprocity does not hold in a strict sense but is, in fact, an approximation. For the flat-fading scenario, our work will cover both reciprocal and nonreciprocal channels. For the frequency selective scenario, however, we will only consider nonreciprocal channels since OFDM systems are more vulnerable to the RF front end imperfections that affect the reciprocity assumption [24].

In the next section we will present an overview of previous works on channel estimation for AF TWRNs.

1.3 Previous Works

A number of recent works have addressed the problem of channel estimation for AF TWRNs, covering both the flat-fading and frequency-selective environments [21,22,28–34]. These works have adopted a training-based approach to channel estimation. This approach requires each terminal to transmit a pilot sequence that is known to the other terminal. The channel parameters are then estimated using either a maximum likelihood (ML) or a least squares (LS) based method [35].

Channel estimation for AF TWRNs in flat-fading channel conditions was considered in [21, 28, 29]. In [21], the training-based ML channel estimator was derived for the estimation of the cascaded channels at the terminals. The corresponding Cramer-Rao Bound (CRB) on the variance of unbiased estimators was also derived. The CRB was then employed as the criterion to design the optimal training sequences. It was shown that the CRB is minimized when the training sequences transmitted by the two terminals are orthogonal to each other. In [28], the ML approach was used to acquire initial estimates of the individual channels at the relay. The relay used these estimates to denoise the training signal and to allocate the training power such that channel estimation performance is optimized at the terminals. The case when the relay and the two terminals are equipped with multiple antennas was investigated in [29], where an algebraic tensor-based method was proposed to estimate the individual channels at the terminals by exploiting the structure of the cascaded MIMO channels and appropriate design of an amplification matrix at the relay.

Channel estimation for AF TWRNs operating in frequency selective fading conditions was studied in [22,30,31]. In [22], OFDM transision was adopted to combat the multipath effects, and LS estimation was employed to estimate the cascaded channels at the terminals. The pilot sequences of the two terminals were designed to minimize the MSE of the LS estimator. It was also demonstrated that, thanks to the channel reciprocity assumption, the estimates of the individual channels can be acquired up to a sign ambiguity from the estimates of the cascaded channels using a search over all vectors with length equal to the number of subcarriers and whose entries take values in the set $\{\pm 1\}$. The works in [30] and [31] considered channel estimation for single carrier cyclic prefix³ (SCCP)-based AF TWRNs for the single and multiple antenna cases, respectively. In both works, the cascaded channels were estimated at the terminals using the LS approach and the pilot sequences were designed to minimize the resulting mean squared error (MSE).

Another problem of interest for OFDM systems is carrier frequency offset (CFO) estimation [37], which is sometimes performed jointly with channel estimation. The CFO, which results from the mismatch between the local oscillators at the different nodes in the network, can cause performance degradation by compromising subcarrier orthogonality. A number of studies have considered joint CFO and channel estimation for AF TWRNs [32–34]. A LS approach was employed to jointly estimate the cascaded channels as well as the end-to-end CFO for conventional and zero-padded (ZP) OFDM systems in [32] and [33], respectively.

³Single carrier cyclic prefix (SCCP) transmission with frequency-domain equalization (FDE) was first proposed in [36]. It delivers very similar performance to OFDM systems in terms of multipath mitigation, but with a much smaller peak-to-average power ratio (PAPR).

To facilitate the estimation of the individual channels and the individual CFOs, superimposed training [38] at the relay was adopted in [34]. Specifically, the relay superimposes its own pilots over the received signal carrying the pilots from the two terminals before broadcasting it back to the terminals. A LS approach was employed to obtain initial estimates of the channel parameters and the CFO, and an iterative ML-based procedure was then used to improve the accuracy of the acquired estimates.

1.4 Motivation

The above-mentioned works on channel estimation for AF TWRNs all follow the trainingbased approach, i.e., they require each terminal to transmit pilot symbols that are known to the other terminal in order to estimate the desired channel parameters. Despite the robustness and simplicity of this approach, the transmission of known pilots consumes much needed bandwidth resources and thus undermines the spectral efficiency of TWRNs. Given the high demand for spectrum utilization, it is important to find more efficient estimation algorithms that can provide the desired accuracy without imposing a heavy training burden. It is also desired to reduce the impact of channel uncertainties on the performance of AF TWRNs by developing methods that can achieve superior accuracy to that provided by the conventional training-based methods developed thus far.

Blind channel estimation [39–42] avoids the use of training pilots and relies only on the received data samples for the estimation of the desired parameters. This approach has been widely studied in point-to-point communication systems because of its high spectral efficiency [39]. To estimate the channel blindly, this approach exploits the structure of the received signal and the properties of the data symbols, such as their distribution or their finite alphabet property. Blind methods are usually based either on the ML criterion, or on exploiting the moments of the received signal, i.e., its second-order or higher order characteristics [39]. Among the most popular types of blind algorithms are subspace-based algorithms [43–45], which estimate the channel by exploiting the orthogonality between the signal and noise subspaces. It should be noted, however, that blind techniques often suffer from an inherent rotational ambiguity [39], which means that it is not possible to acquire all the desired information about the channel blindly. Thus, the use of a number of pilots, albeit small, remains necessary to resolve this ambiguity.

Another alternative to training-based estimation is semi-blind estimation [41, 46–48],

1.4 Motivation

which is a hybrid of the blind and training-based approaches. Similar to the latter, it uses pilot symbols, but it also incorporates into the estimation the received data samples as well as the received training samples. This makes the semi-blind approach more flexible than the purely blind approach and eliminates the need for separate ambiguity resolution. By utilizing the data samples in conjunction with the pilots, semi-blind estimation requires a smaller number of pilots, which makes it more spectrally efficient than training-based estimation [48]. It is also capable of achieving higher accuracy than that possible using purely blind or purely training-based estimation [48]. The expectation maximization (EM) framework provides one of the most popular algorithms that can be employed for semi-blind estimation [49–51]. The EM algorithm is an iterative method commonly used to avoid the high complexity of direct ML estimation when the likelihood function incorporates unknown random parameters (such as the data symbols) in addition to the unknown deterministic parameters (such as the channel parameters).

Despite being extensively studied in point-to-point communications, little effort has so far been made to investigate the application of blind and semi-blind approaches for channel estimation in relay networks. Due to the differences in the received signal structure, the blind and semi-blind estimation methods developed for point-to-point systems cannot be directly applied in AF TWRNs. Hence, there is a need to develop new blind and semiblind channel estimation algorithms that are specifically tailored for the AF TWRN signal model. In fact, AF TWRNs constitute a promising candidate for the application of such techniques because of the presence of the known self-interference symbols embedded in the received data samples. These self-interference symbols may be perceived as pseudo-pilots whose knowledge can be used to extract valuable information about the channel.

The development of blind and semi-blind estimation algorithms for AF TWRNs is still in its early stages. In [52], an algorithm was developed for the estimation of the cascaded channels blindly in OFDM-based AF TWRNs based on the second-order statistics of the received signal. However, the proposed algorithm requires a very large number of OFDM blocks to achieve accurate estimation accuracy. In [53], a semi-blind algorithm was developed for joint data detection and the estimation of the cascaded channels in MIMO-OFDM AF TWRNs based on the expectation conditional maximization (ECM) method with soft interference cancellation.

1.5 Thesis Organization and Contributions

This thesis proposes new methods for channel estimation in AF TWRNs based on the blind and semi-blind approaches. The primary goal is to achieve substantial improvements in spectral efficiency and/or estimation accuracy over the conventional training-based methods. We will achieve this goal by designing algorithms that not only have a manageable computational complexity but also require only a limited number of data samples. We will also provide the appropriate theoretical tools for analyzing the performance of these algorithms. We will consider both the flat-fading and the frequency-selective environments as well as both the reciprocal and nonreciprocal channel assumptions.

Throughout this thesis, we will focus on the classical AF TWRN with two single antenna terminals and one single antenna relay operating in the half-duplex mode. This setting is sufficient for the purpose of demonstrating the feasibility and high potential of the blind and semi-blind estimation. Although similar gains may be possible in the multiple antenna scenario, the generalization from the single antenna case to the multiple antenna case is not necessarily straightforward and is left for future work.

As our focus is on channel estimation performance, we will assume perfect timing synchronization between the two terminals and the relay. This assumption is very common in works on channel estimation [21, 22, 28–34]. In practice, synchronization of the timing offset would be handled by a separate block that precedes the channel estimation block. Similarly, we also assume perfect frequency synchronization. In practice, frequency synchronization requires estimation and compensation of the CFO. CFO estimation can be handled by a separate block, although sometimes it may also be handled in conjunction with channel estimation, as done in [34].

In the rest of this section, we describe the different chapters of this thesis and the research contributions presented in each one.

Chapter 2 provides a background on some of the important concepts that will be used in the thesis. In particular, we provide a brief background on ML estimation and discuss its asymptotic properties. We also discuss different ways of applying the ML criterion when the received signal involves random nuisance parameters. We then introduce the EM algorithm which can serve as a low-complexity alternative to ML estimation in the presence of random nuisance parameters. In addition, we introduce the CRB which will be adopted as a benchmark on estimation performance. We also discuss the modified CRB (MCRB) which is a more tractable variant of the CRB used in the presence of random nuisance parameters. After a brief discussion on multipath fading, we finally provide an overview of OFDM transmission, which is commonly employed in frequency selective environments to prevent the detrimental effects of ISI.

In Chapter 3, we focus on AF TWRNs that employ constant modulus (CM) signalling. Assuming nonreciprocal flat-fading channels, we propose an algorithm for blind channel estimation based on the deterministic maximum likelihood (DML) approach, which treats the data symbols as deterministic unknowns. Assuming *M*-PSK modulation, we show that the resulting estimator is consistent and approaches the true channel with high probability at high signal-to-noise ratio (SNR) for modulation orders higher than 2. For BPSK (2-PSK), however, the DML algorithm performs poorly. Motivated by this, we propose an alternative algorithm that yields much better performance by taking into account the BPSK structure of the data symbols. For comparative purposes, we also investigate the Gaussian maximumlikelihood (GML) approach which treats the data symbols as Gaussian-distributed nuisance parameters. We also derive the corresponding CRB and use Monte-Carlo simulations to investigate the mean squared error (MSE) performance of the proposed algorithms. By comparing the symbol-error rate (SER) performance of the DML algorithm with that of the training-based LS estimator, we demonstrate that the DML offers a superior tradeoff between accuracy and spectral efficiency than the LS estimator.

In Chapter 4, we consider a very similar problem to Chapter 3, but we focus on reciprocal channels instead. We derive the corresponding DML channel estimator and prove that it approaches the true channel with high probability at high SNR but, unlike the nonreciprocal case, is not consistent. We then propose an alternative estimator which acquires the channel estimate by minimizing the sample variance of the envelope of the received signal after self-interference cancellation. This estimator is consistent and has favorable high-SNR performance, and can be implemented at a low complexity using the steepest descent algorithm. We also derive the CRB for the reciprocal case.

Still focusing on flat-fading channels, in Chapter 5 we shift our attention to semi-blind estimation which utilizes both pilots and data samples. Assuming nonreciprocal channels, we derive the CRB for semi-blind channel estimation in AF TWRNs employing square QAM. In contrast to the CRBs derived in the previous two chapters, which treated the data symbols as deterministic unknowns, the CRB derived in this chapter is exact as it is based on the true likelihood function that takes into account the statistics of the transmitted data symbols. The derivation of this bound is a challenging task due to the complicated nature of the likelihood function when the statistics of the transmitted data symbols are taken into account, and to the best of our knowledge it has not been reported before for AF TWRNs. Using this new bound, we show that substantial improvements in estimation accuracy are possible by exploiting even a limited number of the data samples in addition to the pilot symbols. These improvements depend on the modulation order: the lower the order the higher the gain in accuracy. Because of the superior accuracy of the semi-blind approach, it requires fewer pilots than the training-based approach, thus yielding a better tradeoff between accuracy and spectral efficiency. Finally, we also derive the more tractable modified CRB which, for low modulation orders, can be used as an approximation of the exact CRB at high SNR.

In Chapter 6, we focus on the design of semi-blind algorithms to achieve the accuracy gains predicted by the CRB analysis in Chapter 5. Direct ML estimation is intractable in this case due to the high complexity of the true likelihood function. Instead, we implement semi-blind estimation using the iterative EM approach. We derive semi-blind EM-based estimators for both nonreciprocal and reciprocal channels. In both cases, the complexity of the EM steps is linear in the number of data samples for any given modulation order. We use simulations to show that, for both nonreciprocal and reciprocal channels, the derived EM algorithms perform very closely to the CRBs and require only a small number of iterations to converge. We also show that the EM algorithm can provide a significant improvement in throughput (as high as 27% for QPSK modulation) since a smaller number of pilots would be needed to achieve the same SER performance as the LS estimator. This confirms the practicality of the semi-blind approach and the achievability of its predicted gains.

Until this point, we have focused on flat-fading channels. In Chapter 7, we shift our attention to OFDM-based TWRNs operating in frequency selective channel conditions. To the best of our knowledge, semi-blind channel estimation has not been considered before for OFDM-based TWRNs. Moreover, all previous works on channel estimation for OFDM TWRNs have considered reciprocal channels, which reduces the number of channel vectors to be estimated from three to two. In contrast to previous works, we consider channel estimation for the more realistic case of nonreciprocal channels. Semi-blind estimation is performed using a single OFDM pilot block and a limited number of OFDM data blocks. To assist in the estimation of the individual channels, we adopt a superimposed training strategy at the relay [34]. More specifically, the relay superimposes its own pilot symbols

over the received pilot OFDM block before broadcasting it. Our proposed semi-blind estimator is based on the Gaussian ML criterion in which the transmitted data are treated as Gaussian-distributed. The resulting semi-blind ML estimator reduces to a nonlinear minimization problem, which we solve numerically. We also design the pilot vectors of the two terminals and the relay to optimize estimation performance. Furthermore, we derive the semi-blind and pilot-based CRBs as estimation performance benchmarks. Finally, we use simulation studies to show that the proposed semi-blind approach provides significant improvements in estimation accuracy over the conventional pilot-based approach and that it closely approaches the semi-blind CRB.

The contributions in this thesis have lead to a number of publications in peer-reviewed journals and refereed conferences, as listed below:

Journal Articles (published)

1. Saeed Abdallah and Ioannis N. Psaromiligkos, "Blind channel estimation for amplifyand-forward two-way relay networks employing *M*-PSK modulation", *IEEE Transactions* on Signal Processing, vol. 60, no. 7, pp. 3604 - 3615, Jul. 2012.

2. Saeed Abdallah and Ioannis N. Psaromiligkos, "Partially-blind estimation of reciprocal channels for AF two-way relay networks employing *M*-PSK modulation", *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1649 - 1654, May 2012.

3. Saeed Abdallah and Ioannis N. Psaromiligkos, "EM-based Semi-blind Channel Estimation in Amplify-and-Forward Two-Way Relay Networks", *accepted for publication in IEEE Wireless Communications Letters*, Jun. 2013.

Journal Articles (under review)

1. Saeed Abdallah and Ioannis N. Psaromiligkos, "Semi-blind channel estimation with superimposed training for OFDM-modulated AF two-way relaying", *submitted to IEEE Transactions on Wireless Communications*, Feb. 2013.

Peer-Reviewed Conference Papers

1. Saeed Abdallah and Ioannis N. Psaromiligkos, "Semi-blind channel estimation for OFDMbased amplify-and-forward two-way relay networks", *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, CA, May 2013.

2. Saeed Abdallah and Ioannis N. Psaromiligkos, "Blind channel estimation for MPSKbased amplify-and-forward two-way relaying", *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.

 Saeed Abdallah and Ioannis N. Psaromiligkos, "Semi-blind channel estimation for amplifyand-forward two-way relay networks employing constant-modulus constellations", in Proc. 44th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, Mar. 2010.

Other

1. Saeed Abdallah and Ioannis N. Psaromiligkos, "Exact Cramer-Rao bounds for semiblind channel estimation in amplify-and-forward two-way relay networks employing square QAM modulation", *ArXiv pre-print cs.IT/1207.5483*, Jul. 2012.

Chapter 2

Background

This chapter provides a brief background on several important concepts that will be used in the thesis. The first concept we cover is the widely employed maximum-likelihood (ML) criterion for estimation which will be the basis for several of the channel estimation algorithms proposed in our work. We discuss the intuition behind ML estimation and some of its appealing asymptotic properties. For the scenario when the received signal involves random nuisance parameters whose estimation is not strictly required, we discuss several ways of applying ML estimation, depending on how the nuisance parameters are treated. We then introduce the Expectation Maximization (EM) algorithm, which can provide efficient, low complexity estimation in the presence of random nuisance parameters and which will be employed in Chapter 6.

Another important concept that we cover is the Cramer-Rao bound (CRB) which is the most commonly employed benchmark on the performance of practical estimators. In addition to the standard CRB, we discuss several variants of the CRB that can be employed when the received signal involves random nuisance parameters, such as the modified CRB (MCRB) [54], a less tight but more tractable variant of the CRB.

Furthermore, this chapter provides a brief overview on the phenomenon of multipath fading, and introduces the different types of fading channels that will be considered this thesis, namely flat fading and frequency selective fading channels. Finally, since Chapter 7 considers frequency-selective channels where OFDM transmission is commonly employed to combat interference (ISI), we also provide and overview of OFDM transmission.

2.1 Maximum Likelihood Estimation

ML estimation [35] is one of the most popular and widely studied approaches to parameter estimation. The philosophy of ML estimation is simple and intuitive. Given a set of statistically independent observations, say r_1, \ldots, r_N , generated from some known distribution which is parametrized with respect to an unknown deterministic parameter or set of parameters, say¹ $\boldsymbol{\theta} \triangleq [\theta_1, \ldots, \theta_K]^T$, the ML approach is to estimate $\boldsymbol{\theta}$ by the value $\hat{\boldsymbol{\theta}}$ which most likely generated the observations, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} f(r_1, \dots, r_N; \boldsymbol{\theta}), \qquad (2.1)$$

where $f(r_1, \ldots, r_N; \boldsymbol{\theta})$ is the joint probability density function of the observations r_1, \ldots, r_N , parametrized w.r.t. $\boldsymbol{\theta}$, also known as the likelihood function. Since the observations are independent, we may rewrite (2.1) as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} f(r_1; \boldsymbol{\theta}) f(r_2; \boldsymbol{\theta}) \dots f(r_N; \boldsymbol{\theta}).$$
(2.2)

ML estimation has many appealing theoretical properties. Assuming that the observations are independent and identically distributed (i.i.d.), the ML estimator is consistent under mild conditions [55], which means that it converges in probability to the true value of the parameter as the number of observed samples becomes large, i.e.²,

$$\lim_{N \to \infty} P(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| > \epsilon) = 0, \quad \forall \epsilon > 0.$$
(2.3)

where $\|\boldsymbol{x}\|$ denotes the 2-norm of \boldsymbol{x} . Furthermore, for i.i.d. observations, the ML estimator is also asymptotically unbiased under mild conditions [56], which means that³ $\lim_{N\to\infty} \mathbb{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$.

2.1.1 ML Estimation in the Presence of Random Nuisance Parameters

A commonly encountered situation in practical estimation problems is when the observations depend on random nuisance parameters whose estimation is not strictly required, and

¹For generality we consider the case of a vector of parameters. The case of a single parameter follows in a straightforward manner.

 $^{^{2}}P(E)$ denotes the probability of the event E.

 $^{{}^{3}\}mathbb{E}[\cdot]$ denotes the statistical expectation operator.

which are only known through their statistical distribution. This is often the case in the context of channel estimation in wireless communications, where the transmitted signal is modulated by data symbols that are chosen with a certain probability distribution from some finite-alphabet constellation.

Suppose that each observation r_i depends on an unknown data symbol d_i , which is chosen from the set $S = \{\xi_1, \ldots, \xi_M\}$ according to some predefined probability distribution. Then the likelihood function for the observation r_i may be expressed as

$$f(r_i; \theta) = \sum_{k=1}^{M} f(r_i, d_i = \xi_k; \theta) = \sum_{k=1}^{M} f(r_i | d_i = \xi_k; \theta) P(d_i = \xi_k).$$
(2.4)

The conditional likelihood terms $f(r_i|d_i = \xi_k; \boldsymbol{\theta})$ often take a simple form. Specifically, each would have the standard Gaussian form when the signal is embedded in additive white Gaussian noise. However, the overall likelihood function in (2.4) is considerably more complicated since it is a weighted sum of the conditional likelihood terms, which corresponds to a Gaussian mixture in the case of additive white Gaussian noise. For Nindependent observations, the joint likelihood function is given by

$$f(r_1,\ldots,r_N;\boldsymbol{\theta}) = \prod_{i=1}^N \left(\sum_{k=1}^M f(r_i | d_i = \xi_k; \boldsymbol{\theta}) P(d_i = \xi_k) \right).$$
(2.5)

The likelihood function in (2.5) may be called the true likelihood function since it incorporates the exact statistics of the data symbols. Unfortunately, the complicated form of (2.5) makes exact ML estimation (i.e., direct application of (2.1) using the true likelihood function) very challenging in the presence of random nuisance parameters.

One way to avoid the highly complicated likelihood function associated with exact ML estimation is to ignore the statistics of the unknown data symbols and treat them as deterministic instead. This approach is called deterministic maximum-likelihood (DML) estimation. In this case, the ML criterion can be used to jointly estimate both the original desired parameters as well as the data symbols:

$$\{\hat{\boldsymbol{\theta}}, \hat{d}_1, \dots, \hat{d}_N\} = \arg \max_{\boldsymbol{\theta}, d_1, \dots, d_N} f(r_1, \dots, r_N | d_1, \dots, d_N; \boldsymbol{\theta})$$

= $\arg \max_{\boldsymbol{\theta}, d_1, \dots, d_N} f(r_1 | d_1; \boldsymbol{\theta}) f(r_2 | d_2; \boldsymbol{\theta}) \dots f(r_N | d_N; \boldsymbol{\theta}).$ (2.6)

In contrast to exact ML estimation, with DML estimation the number of unknown parameters grows linearly with the number of samples. The original parameters (θ) which affect all the samples are called structural parameters, while the data symbols, each of which affects a single observation, are called incidental parameters [57]. Since the value of the data symbols can vary between observations, the observations are no longer i.i.d. The estimation of structural parameters in the presence of incidental parameters has been investigated in the literature [57–60] and is referred to as the Incidental Parameter Problem. One drawback of DML estimation is that the asymptotic properties of conventional ML estimators (e.g., consistency), which hold under mild conditions when the dimension of the parameter space is fixed, do not necessarily hold in the presence of incidental parameters [58]. In fact, the ML estimator may not be consistent even when a consistent estimator exists [57].

Another alternative to exact ML estimation is to approximate the data symbols as Gaussian distributed [61]. This approach, called Gaussian ML (GML), considerably simplifies the likelihood function without introducing extra parameters. Furthermore, it often makes it feasible to obtain closed-form estimates of the desired parameters. However, this simplicity may come at the price of a lower estimation accuracy since the Gaussian approximation is used instead of the true statistics of the data symbols. Since the data symbols are not Gaussian distributed in reality, some of the asymptotic properties of conventional ML estimators may not hold.

2.1.2 The Expectation Maximization Algorithm

The EM framework [49–51] provides a convenient low-complexity method for approximating the true ML solution in the presence of random nuisance parameters, which can be also perceived as missing information. Starting with arbitrary values of the unknown parameters, the EM algorithm iterates between calculating the conditional expectation of the complete-data log-likelihood and maximizing this expectation with respect to the unknown parameters. The latter maximization is typically easier than maximizing the true likelihood function.

The basic idea of the EM algorithm is that there is the set $\boldsymbol{d} \triangleq [d_1, \ldots, d_N]^T$ of hidden or missing data that would make the estimation of $\boldsymbol{\theta}$ easier if they were known. The observation vector $\boldsymbol{r} \triangleq [r_1, \ldots, r_N]^T$ represents incomplete data, while the complete data is $\{\boldsymbol{r}, \boldsymbol{d}\}$. An iteration of the EM algorithm, say the *t*-th one, consists of two steps. The
first step, called the expectation step (E-step) consists of evaluating the term

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = \mathbb{E}\left\{ \log f(\boldsymbol{r}, \boldsymbol{d}; \boldsymbol{\theta}) | \boldsymbol{r}; \boldsymbol{\theta}^{(t)} \right\}, \qquad (2.7)$$

which is the expectation of the log-likelihood of the complete data, $\log f(\mathbf{r}, \mathbf{d}; \boldsymbol{\theta})$, with respect to the conditional PDF (or PMF) $f(\mathbf{d}|\mathbf{r}; \boldsymbol{\theta}^{(t)})$ of the hidden data given the observations and the current estimate $\boldsymbol{\theta}^{(t)}$ of $\boldsymbol{\theta}$. The second step of the EM algorithm is the maximization step (M-step) which consists of maximizing the expectation $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$ in order to obtain an updated estimate $\boldsymbol{\theta}^{(t+1)}$, i.e.,

$$\boldsymbol{\theta}^{(t+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}).$$
(2.8)

The two steps are repeated until the algorithm converges. The algorithm can be initialized using an arbitrarily chosen $\theta^{(0)}$. Although the EM algorithm does not always converge to the true ML solution, it produces estimates which monotonically increase in likelihood and is guaranteed to converge to a stationary point of the likelihood function under fairly general conditions [62]. However, the choice of the initial point may affect the rate of convergence of the algorithm as well as the final point of convergence.

2.2 Cramer-Rao Bound

The CRB [35] is a fundamental lower limit on the variance of any unbiased estimator and is the most widely used benchmark on the performance of practical estimators. The CRB for the estimation of the parameter vector $\boldsymbol{\theta}$ can be obtained by evaluating the Fisher information matrix (FIM), defined as

$$\boldsymbol{I}(\boldsymbol{\theta}) = \mathbb{E}\left\{\frac{\partial \mathcal{L}(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \mathcal{L}(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{T}}\right\} = -\mathbb{E}\left\{\frac{\partial^{2} \mathcal{L}(\boldsymbol{r};\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{T}}\right\}$$
(2.9)

where $\mathcal{L}(\boldsymbol{r};\boldsymbol{\theta}) \triangleq \log f(\boldsymbol{r};\boldsymbol{\theta})$ is the log-likelihood function. Let $\tilde{\boldsymbol{C}} \triangleq \mathbb{E}\left\{ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right\}$ be the error covariance matrix for the unbiased estimation of $\boldsymbol{\theta}$, then for any unbiased estimator of $\boldsymbol{\theta}$, we have that⁴ [35]

$$\tilde{\boldsymbol{C}} \ge \boldsymbol{I}(\boldsymbol{\theta})^{-1}.$$
(2.10)

⁴For square matrices A and B, the notation $A \ge B$ means that the difference A - B is positive semi-definite.

Hence, the mean-squared error (MSE) for unbiased estimators of θ is bounded as follows

$$\mathbb{E}\left\{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^{2}\right\} \geq \operatorname{tr}\left(\boldsymbol{I}(\boldsymbol{\theta})^{-1}\right).$$
(2.11)

The CRB is thus given by

$$CRB_{\boldsymbol{\theta}} = \operatorname{tr}\left(\boldsymbol{I}(\boldsymbol{\theta})^{-1}\right).$$
 (2.12)

The CRB is known to be tight for a wide range of estimators, provided that the number of observations is sufficiently large and/or SNR is high [54]. An estimator is said to be efficient if it attains the CRB. One of the attractive properties of ML estimation is that if an efficient estimator exists then the ML estimator is efficient. Furthermore, even when an efficient estimator does not exist, the ML estimator is asymptotically efficient for i.i.d. observations under mild conditions [56], which means that it approaches the CRB as the number of observations becomes large.

2.2.1 The CRB in the Presence of Random Nuisance Parameters

When the observations involve random nuisance parameters in the form of data symbols that are known only through their distribution, the standard CRB, which is based on the true likelihood function (see (2.5)), can be very difficult to evaluate. We will derive such a bound in Chapter 5 of this thesis. However, due to the considerable analytical derivations required to obtain such bounds, it is also worth it to consider other, simpler variants of the CRB which may be used in the presence of random nuisance parameters.

One way to simplify the derivation of the CRB is to ignore the statistics of the nuisance parameters and treat them as deterministic, as done in DML estimation. In this case, the CRB can be obtained for the joint estimation of both $\boldsymbol{\theta}$ and the data symbols. Letting $\tilde{\boldsymbol{\theta}} \triangleq [\boldsymbol{\theta}^T, \boldsymbol{d}^T]^T$ be the augmented parameter vector, the corresponding FIM is given by

$$\boldsymbol{I}(\tilde{\boldsymbol{\theta}}) = \mathbb{E}\left\{\frac{\partial \mathcal{L}(\boldsymbol{r}|\boldsymbol{d};\boldsymbol{\theta})}{\partial \tilde{\boldsymbol{\theta}}} \frac{\partial \mathcal{L}(\boldsymbol{r}|\boldsymbol{d};\boldsymbol{\theta})}{\partial \tilde{\boldsymbol{\theta}}^{T}}\right\}.$$
(2.13)

As we can see from (2.13), the size of the FIM increases with the number of observations. The resulting CRB for the estimation of $\boldsymbol{\theta}$ is given by

$$CRB_D = \sum_{i=1}^{K} [\boldsymbol{I}(\tilde{\boldsymbol{\theta}})^{-1}]_{ii}.$$
(2.14)

This bound is tight for the class of estimators which ignore the statistics of the data symbols, such as the ones that will be considered in Chapters 3 and 4. However, estimators that take into account the statistics of the data may outperform the bound.

Another variant of the CRB that is commonly used in the presence of random nuisance parameters is the modified CRB (MCRB) [54,63]. In contrast to the previous bound, this bound takes into account the statistics of the nuisance parameters and it applies to the general class of estimators for which the standard CRB applies. The MCRB is obtained using a modified version of the FIM, denoted as the modified FIM (MFIM), given by

$$\boldsymbol{I}_{M}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{r},\boldsymbol{d}} \left\{ \frac{\partial \log f(\boldsymbol{r}|\boldsymbol{d};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f(\boldsymbol{r}|\boldsymbol{d};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{T}} \right\}$$

$$= \mathbb{E}_{\boldsymbol{d}} \mathbb{E}_{\boldsymbol{r}|\boldsymbol{d}} \left\{ \frac{\partial \log f(\boldsymbol{r}|\boldsymbol{d};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f(\boldsymbol{r}|\boldsymbol{d};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{T}} \right\}.$$

$$(2.15)$$

Thus, the FIM for the estimation of $\boldsymbol{\theta}$ is first obtained while treating the data symbols as deterministic unknowns and then averaged using the statistics of the data symbols to yield the MFIM. It is proved in [64] that

$$\boldsymbol{I}(\boldsymbol{\theta})^{-1} - \boldsymbol{I}_M(\boldsymbol{\theta})^{-1} \ge \boldsymbol{0}, \qquad (2.16)$$

which shows that the MCRB is a valid bound, though looser than the standard CRB.

2.3 Multipath Fading

Multipath fading [20] refers to the attenuation due to the interference between different copies of the signal which arrive at the receiver through different paths, with slightly different time delays, amplitudes and phase offsets. The multipath phenomenon can severely degrade the received signal power, and thus it requires proper compensation at the receiver.

2.3.1 Slow and Fast Fading

The channel coherence time T_c is the period during which the channel remains approximately unchanged, or in other words the period during which the fading process is highly correlated [20]. Hence, it characterizes the time varying nature of the wireless channel. The fading is designated as slow when the symbol period T is small relative to the channel coherence time, i.e., $T \ll T_c$. Slow fading channels that are static over the duration of several symbol periods are described as quasi-static within the period T_c . The fading is considered to be fast when $T > T_c$. In our work we will focus on quasi-static channels.

2.3.2 Flat-fading and Frequency-selective fading Channels

The maximum delay spread of the channel, denoted as τ_{max} , is the time duration for the arrival of the multipath components. The maximum delay spread is closely related to another quantity, the channel coherence bandwidth B_c which measures the frequency range over which the channel frequency response is nearly flat, i.e., highly correlated. In particular, we have [20]

$$B_c \approx \frac{1}{\tau_{max}}.\tag{2.17}$$

An important characteristic of fading channels is their frequency selectivity. A wireless channel is classified as frequency-flat if the signal bandwidth is much smaller than the coherence bandwidth of the channel, which means that all the spectral components of the received signal are affected in the same way. In the time domain, this means that the delay spread τ_{max} is much smaller than the symbol period T, and the different multipath components cannot be resolved at the receiver. A flat-fading channel can be modelled as a random complex multiplicative coefficient, whose distribution is chosen depending on the propagation environment. For instance, when the multipath components of the signal do not contain a line-of-sight (LOS) component, such as commonly the case in urban areas, the channel coefficient can be modelled as a noncircular complex Gaussian random variable [65].

On the other hand, when the signal bandwidth is greater than the coherence bandwidth of the channel, the different spectral components of the signal experience different amplitude gains and phase shifts, and the channel is called frequency selective. In the time domain, the multipath delay spread τ_{max} is greater than the symbol period, and the different multipath components are resolvable at the receiver. The presence of multiple resolvable versions of the transmitted symbol waveform results in ISI at the receiver. The channel is referred to as wideband in this case and is commonly modelled in baseband as a linear FIR filter. The filter coefficients can be modelled as zero-mean circularly complex Gaussian RVs, and their variance is commonly assumed to follow an exponentially decaying power delay profile [20]. Frequency selectivity requires a more sophisticated form of channel equalization to avoid the detrimental effects of ISI. The most popular solution for addressing this problem is OFDM, which we describe next.

2.4 OFDM Systems

OFDM has emerged as the dominant technology for broadband multicarrier communication [65]. The basic idea of OFDM is to transform the frequency-selective channel into many parallel flat-fading subchannels, by converting the serial input data stream into parallel lowrate streams that are modulated on separate subcarriers and transmitted simultaneously. This process substantially simplifies the task of channel equalization, which can be performed in the frequency-domain at each subcarrier using a single-tap equalizer. To avoid interference between consecutive OFDM symbols, a Cyclic Prefix (CP) can be inserted into each OFDM symbol, at the cost of a small reduction in spectral efficiency. OFDM is also characterized by its high bandwidth efficiency because the subcarriers are allowed to overlap in the frequency domain while maintaining orthogonality between their respective time domain waveforms. OFDM can be implemented efficiently in hardware by using the Inverse Fast Fourier Transform (IFFT) block for modulation, and the Fast Fourier Transform (FFT) block for demodulation, which significantly reduces hardware complexity. Because of its numerous advantages, OFDM has been adopted in various wireless standards, such as European digital audio broadcast (DAB) [66], digital video broadcast (DVB) [67], as well as the 3GPP LTE standards for next generation broadband wireless systems [68].

In what follows, we first describe block-based transmission under frequency selective channel conditions to illustrate the effects of ISI and then provide a detailed description of OFDM transmission. Let $\mathbf{h} \triangleq [h_1, \ldots, h_L]^T$ be the vector of FIR filter coefficients representing the frequency selective channel. Let $\tilde{x}(n)$ be the input sequence, the *n*th received sample is given by

$$y(n) = \sum_{k=1}^{L} h(k)\tilde{x}(n-k) + v(n), \qquad (2.18)$$

where v(n) is additive White Gaussian noise. Now suppose that transmission occurs in consecutive blocks of length N > L and let $\tilde{\boldsymbol{x}}_0$ and $\tilde{\boldsymbol{x}}_1$ be two consecutive transmitted blocks. We focus on the received signal corresponding to the transmission of $\tilde{\boldsymbol{x}}_1 = [\tilde{x}_{11}, \ldots, \tilde{x}_{1N}]^T$.



Fig. 2.1 A typical OFDM transceiver.

The first N received samples corresponding to the transmission of \tilde{x}_1 are given by⁵

$$\boldsymbol{y} = \boldsymbol{H}\tilde{\boldsymbol{x}}_1 + \boldsymbol{H}_0\tilde{\boldsymbol{x}}_0 + \boldsymbol{v}, \qquad (2.19)$$

where \boldsymbol{H} and $\boldsymbol{\check{H}}$ are the $N \times N$ matrices given by

$$\boldsymbol{H} = \begin{bmatrix} h_1 & 0 & \dots & 0 \\ h_2 & h_1 & 0 & \dots & 0 \\ \vdots & \ddots & & \ddots & \\ 0 & \dots & 0 & h_L & \dots & h_1 \end{bmatrix}$$
(2.20)

and

$$\boldsymbol{H}_{0} = \begin{bmatrix} h_{L} & \dots & h_{2} \\ \mathbf{0}_{N \times (N-L+1)} & \vdots & \ddots & & \\ & 0 & \dots & 0 & h_{L} \\ & & & \mathbf{0}_{(N-L+1) \times (L-1)} \end{bmatrix}, \quad (2.21)$$

respectively and \boldsymbol{v} is the additive noise vector. We can see from (2.19) that, in addition to the ISI, the first L-1 received samples also experience interblock interference (IBI) due to the preceding transmitted block $\tilde{\boldsymbol{x}}_{0}$.

We now describe how OFDM operates. An illustration of the OFDM transceiver is shown in Fig. 2.1. Starting with the $N \times 1$ vector of constellation symbols $\tilde{\boldsymbol{x}}$, we obtain its inverse discrete Fourier transform (IDFT), given by $\boldsymbol{x} = \boldsymbol{F}^H \tilde{\boldsymbol{x}}$, where \boldsymbol{F} is the $N \times N$ normalized discrete Fourier transform (DFT) matrix whose (m, n)th entry is $1/\sqrt{N}e^{-j2\pi(m-1)(n-1)/N}$. Before transmitting \boldsymbol{x} , a cyclic prefix (CP) of length $L_{CP} \geq L-1$

⁵Although the total number of received samples which bear information about \tilde{x}_1 is N + L - 1, we focus only on the first N samples since that the last L - 1 samples would contain information from the block transmitted after \tilde{x}_1 .



Fig. 2.2 OFDM block with CP.

is inserted, as shown in Fig. 2.2. In particular, a copy of the last L_{CP} symbols of the vector \boldsymbol{x} is inserted at the beginning of \boldsymbol{x} . The resulting vector, \boldsymbol{x}' , is then serially transmitted.

At the receiver side, the received vector is the result of the linear convolution between the channel impulse response h and the transmitted vector x', and thus has a length of $N + L_{CP} + L - 1$. The receiver discards the first L_{CP} received samples (which are affected by IBI) and collects N samples of the received signal, resulting in the vector

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v} \tag{2.22}$$

where $\check{\boldsymbol{H}}$ is the $N \times N$ circulant matrix with first column $\bar{\boldsymbol{h}} \triangleq [\boldsymbol{h}^T, \boldsymbol{0}_{1 \times N - L}]^T$. The circulant structure of $\check{\boldsymbol{H}}$ plays an important role in the functionality of OFDM systems. Specifically, $\check{\boldsymbol{H}}$ may be expressed as $\check{\boldsymbol{H}} = \boldsymbol{F}^H \tilde{\boldsymbol{H}} \boldsymbol{F}$, where $\tilde{\boldsymbol{H}}$ is the diagonal matrix whose diagonal elements are given by $\tilde{\boldsymbol{h}} = \sqrt{N} \boldsymbol{F} \bar{\boldsymbol{h}} = [\tilde{h}_1, \dots, \tilde{h}_N]^T$, which is the N-point (non-normalized) DFT of \boldsymbol{h} . Hence, (2.22) becomes

$$\boldsymbol{y} = \boldsymbol{F}^H \tilde{\boldsymbol{H}} \tilde{\boldsymbol{x}} + \boldsymbol{v}. \tag{2.23}$$

Let $\tilde{\boldsymbol{y}} = \boldsymbol{F} \boldsymbol{y}$ be the DFT \boldsymbol{y} , then

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{v}},$$
(2.24)

where $\tilde{\boldsymbol{v}} \triangleq \boldsymbol{F}\boldsymbol{v}$ is the DFT of \boldsymbol{v} . From (2.24), we see that the *i*th element of $\tilde{\boldsymbol{y}}$ is given by $\tilde{y}_i = \tilde{h}_i \tilde{x}_i + \tilde{v}_i$, which means that the *i*th transmitted symbol \tilde{x}_i is affected by the flat-fading channel coefficient \tilde{h}_i which is the *i*th element in the *N*-pt DFT of \boldsymbol{h} . Hence, by using a CP and applying the IFFT and FFT operations at the transmitter and receiver, respectively, the frequency selective channel has been successfully converted into *N* parallel flat-fading subchannels. For each of these subchannels, a single tap equalizer is sufficient to perform equalization, which can be followed by detection. It is also worth noting that the noise

statistics are not affected since \tilde{v} is still white Gaussian with the same statistics as v. OFDM can be easily extended to the context of AF TWRNs, as we shall see in Chapter 7.

Chapter 3

Blind Channel Estimation for Nonreciprocal Flat-fading Channels under MPSK modulation

3.1 Introduction

In this chapter, we consider blind channel estimation for AF TWRNs that employ constantmodulus (CM) signalling under nonreciprocal flat-fading channel conditions. Because of its constant envelope, CM signaling permits the use of inexpensive and energy-efficient nonlinear amplifiers. In fact, CM signalling in the form of continuous-phase modulation (CPM) is used in the well-known GSM cellular standard, while 8-PSK modulation is employed in the Enhanced Data Rates for GSM Evolution (EDGE) [69]. Moreover, QPSK modulation is supported in the 3rd Generation Partnership Program (3GPP) Long Term Evolution (LTE) and LTE-Advanced wireless standards [70].

We propose a deterministic ML (DML)-based algorithm that estimates the cascaded channel parameters blindly by treating the data symbols as deterministic unknowns. While the proposed algorithm may be applied to any type of CM signalling, we analyze its asymptotic performance assuming that the terminals employ M-PSK modulation. Noting that consistency is not guaranteed for ML estimators when the data symbols are treated as deterministic unknowns [58,59], we prove that our DML estimator is consistent when the channel parameters belong to compact sets. We also study the asymptotic behavior of the DML estimator at high SNR and prove that it approaches the true channel with high probability for modulation orders higher than 2. For M = 2, however, the DML estimator performs poorly, and we propose an alternative estimator based on the constrained ML (CML) approach which provides much better performance by explicitly taking into account the BPSK structure of the data symbols. As a simple alternative to the DML approach, we also consider the Gaussian ML (GML) estimator which is obtained by treating the data symbols as Gaussian-distributed nuisance parameters. When CM signalling is employed, the GML estimator takes the form of a sample average which is consistent and can be updated online but suffers from an error floor at high SNR. We also derive two CRBs for our estimation problem. The first bound is obtained by treating the data symbols as deterministic unknown parameters and the second is the MCRB discussed in Chapter 2.

Monte Carlo simulations are used to investigate the performance of the proposed algorithms. For M > 2, we show that the DML estimator outperforms the GML estimator at medium-to-high SNR and approaches the CRB at high SNR. For M = 2, we show that the CML-inspired estimator outperforms the GML estimator except at very low SNR. We also investigate the tradeoffs of following the blind approach by comparing the symbol-error rate (SER) performance of the DML estimator with that of the training-based LS estimator. We show that the DML approach provides a better tradeoff between accuracy and spectral efficiency.

The work presented in this chapter originally appeared in [71]. A more elaborated version, including mathematical proofs and new theoretical contributions appeared in [72].

The remainder of this chapter is organized as follows. In Section 3.2, we present our system model. In Section 3.3, we present the proposed algorithms. In Section 3.4, we analyze the asymptotic behavior of our estimators and derive the CRB. We show our simulation results and comparisons in Section 3.5. Finally, our conclusions are discussed in Section 3.6.

3.2 System Model

We consider the typical half-duplex TWRN with two source nodes, \mathcal{T}_1 and \mathcal{T}_2 , and a single relaying node \mathcal{R} , shown in Fig. 1.2. The network operates in quasi-static flat-fading channel conditions. Each data transmission period is divided into two phases. In the first phase, \mathcal{T}_1 and \mathcal{T}_2 simultaneously transmit to \mathcal{R} the *M*-PSK data symbols s_1 and s_2 , respectively. The symbols s_1 and s_2 are of the form $s_1 = \sqrt{P_1}e^{j\phi_1}$ and $s_2 = \sqrt{P_2}e^{j\phi_2}$, where P_1 and P_2 are the transmission powers of \mathcal{T}_1 , \mathcal{T}_2 , respectively, and ϕ_1 , ϕ_2 are the information-bearing phases randomly and independently chosen from the set $S_M = \{(2\ell - 1)\pi/M, \ell = 1, \ldots, M\}$. The received signal at the relay during the first transmission phase is given by

$$r = h_1 s_1 + g_1 s_2 + n, (3.1)$$

where h_1 and g_1 are the complex coefficients of the flat-fading channels $\mathcal{T}_1 \to \mathcal{R}$ and $\mathcal{T}_2 \to \mathcal{R}$, respectively, modelled as¹ $\mathcal{CCN}(0, \gamma^2)$, and n is the additive white noise modelled as $\mathcal{CCN}(0, \sigma^2)$. In the second phase, \mathcal{R} broadcasts the amplified signal Ar, where A is the amplification factor, assumed to be known at both terminals as is common practice (cf. [21, 28]). The amplified signal passes through the channels h_2 and g_2 to reach \mathcal{T}_1 and \mathcal{T}_2 , respectively. The coefficients h_2 and g_2 are also modelled as $\mathcal{CCN}(0, \gamma^2)$. Furthermore, we assume that h_1 and h_2 are correlated with a correlation coefficient of $\varrho \triangleq \mathbb{E}\{h_1h_2\}$. Similarly, g_1 and g_2 are correlated with the same correlation coefficient. The complex channel coefficients h_1, h_2, g_1 and g_2 remain fixed during the estimation period. To maintain an average power of P_r at the relay over the long term, the amplification factor is chosen as [21]

$$A = \sqrt{\frac{P_r}{\gamma^2 P_1 + \gamma^2 P_2 + \sigma^2}}.$$
 (3.2)

Without loss of generality, we consider channel estimation at terminal \mathcal{T}_1 and assume that a similar process is taking pace at terminal \mathcal{T}_2 . The received signal at \mathcal{T}_1 in the second transmission phase is

$$z = Ah_2h_1s_1 + Ah_2g_1s_2 + Ah_2n + \eta$$

= $Aas_1 + Abs_2 + Ah_2n + \eta$, (3.3)

where η is $\mathcal{CCN}(0, \sigma^2)$ and $a \triangleq h_2h_1$, $b \triangleq h_2g_1$ are the cascaded channels parameters whose knowledge is sufficient for detection purposes. The noise variance σ^2 is assumed to be known at \mathcal{T}_1 . The term Aas_1 represents the self-interference at \mathcal{T}_1 . We will focus in this chapter on the estimation of the cascaded channels a and b. Under the CM assumption, it is sufficient for detection to know a and $\phi_b \triangleq \angle b$. Let $\tau \triangleq |h_2|^2$, we can see from (3.3) that τ is also identifiable when the noise variance σ^2 is known. However, τ is not needed

¹The notation $\mathcal{CCN}(\mu, \sigma^2)$ is used to denote a circularly complex Normal random variable with mean μ and variance σ^2 .

in detection.

3.3 Proposed Channel Estimation Algorithms

Estimation is performed at \mathcal{T}_1 using N received samples, z_i , $i = 1, \ldots, N$, of the form given by (3.3). The time index *i* is used to indicate the realizations of s_1 , s_2 , ϕ_1 , ϕ_2 , n, η , that gave rise to each sample z_i . Let $\boldsymbol{z} \triangleq [z_1, \ldots, z_N]^T$ be the vector of received samples. This vector can be expanded as

$$\boldsymbol{z} = Aa\boldsymbol{s}_1 + Ab\boldsymbol{s}_2 + Ah_2\boldsymbol{n} + \boldsymbol{\eta}, \tag{3.4}$$

where $\boldsymbol{s}_1 \triangleq [s_{11}, \ldots, s_{1N}]^T$, $\boldsymbol{s}_2 \triangleq [s_{21}, \ldots, s_{2N}]^T$, $\boldsymbol{n} \triangleq [n_1, \ldots, n_N]^T$ and $\boldsymbol{\eta} \triangleq [\eta_1, \ldots, \eta_N]^T$.

3.3.1 Deterministic ML Approach

To avoid dealing with a complicated likelihood function, we treat the transmitted symbols s_{2i} , $i = 1, \ldots, N$ as deterministic unknowns. We also ignore the finite alphabet constraint that restricts the phases ϕ_{2i} to the set S_M . Nonetheless, the actual statistics of these phases will be used in Section 3.4 to analyze the behavior of the resulting estimator. Due to the above assumptions, the DML approach can be used to blindly estimate the sums $\phi_{bi} \triangleq \phi_{2i} + \phi_b$, but it cannot be used to obtain separate estimates of ϕ_b and ϕ_{2i} , $i = 1, \ldots, N$. However, since *M*-PSK modulation is assumed, the Viterbi-Viterbi algorithm [73] may be used to blindly estimate ϕ_b . A few pilots, however, are still needed to resolve the resulting *M*-fold ambiguity in ϕ_b .

We let $\boldsymbol{\theta} \triangleq [a, |b|, \tau, \phi_{b1}, \dots, \phi_{bN},]^T$ be the vector of unknown parameters for the DML algorithm. With \boldsymbol{s}_2 assumed deterministic and \boldsymbol{s}_1 known, the received vector \boldsymbol{z} is complex Gaussian with $\mathbb{E}\{\boldsymbol{z}\} = Aa\boldsymbol{s}_1 + Ab\boldsymbol{s}_2$ and covariance $\mathbb{E}\{\boldsymbol{z}\boldsymbol{z}^H\} = \sigma^2(A^2\tau + 1)\boldsymbol{I}_N$ where \boldsymbol{I}_N is the $N \times N$ identity matrix. Hence, the log-likelihood function is given by

$$\mathcal{L}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{1}{\sigma^2 (A^2 \tau + 1)} \|\boldsymbol{z} - Aa\boldsymbol{s}_1 - Ab\boldsymbol{s}_2\|^2 - N \log \left(\pi \sigma^2 (A^2 \tau + 1)\right) = -\frac{1}{\sigma^2 (A^2 \tau + 1)} \sum_{i=1}^N \left| z_i - Aas_{1i} - A|b| \sqrt{P_2} e^{j\phi_{bi}} \right|^2 - N \log \left(\pi \sigma^2 (A^2 \tau + 1)\right).$$
(3.5)

Let \hat{a} , $|\hat{b}|$, $\hat{\tau}$, $\hat{\phi}_{bi}$ be the ML estimates of a, |b|, τ and ϕ_{bi} , $i = 1, \ldots, N$, respectively. It

is straightforward to show that $\hat{\phi}_{bi} = \angle (z_i - A\hat{a}s_{1i})$, which when substituted into $\mathcal{L}(\boldsymbol{z};\boldsymbol{\theta})$ yields the updated log-likelihood

$$\mathcal{L}'(\boldsymbol{z}; a, |b|, \tau) = -\frac{1}{\sigma^2 (A^2 \tau + 1)} \sum_{i=1}^N \left(|z_i - Aas_{1i}| - A|b| \sqrt{P_2} \right)^2 - N \log \left(\pi \sigma^2 (A^2 \tau + 1) \right).$$
(3.6)

Maximizing (3.6) with respect to |b|, we get

$$|\hat{b}| = \frac{\sum_{i=1}^{N} |z_i - A\hat{a}s_{1i}|}{NA\sqrt{P_2}}.$$
(3.7)

Substituting $|\hat{b}|$ in place of |b| in (3.6), we obtain

$$\mathcal{L}''(\boldsymbol{z}; a, \tau) = -\frac{1}{\sigma^2 (A^2 \tau + 1)} \sum_{i=1}^N \left(|z_i - Aas_{1i}| - \frac{1}{N} \sum_{k=1}^N |z_k - Aas_{1k}| \right)^2 - N \log \left(\pi \sigma^2 (A^2 \tau + 1) \right).$$
(3.8)

Hence, the DML estimate of a is given by²

$$\hat{a} = \arg\min_{u \in \mathbb{C}} \sum_{i=1}^{N} \left(|z_i - Aus_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aus_{1k}| \right)^2.$$
(3.9)

The objective function in (3.9) does not have a closed-form solution. However, it can be solved using numerical optimization methods such as the steepest-descent algorithm or quasi-Newton type methods [74,75]. The computational complexity of the steepest descent implementation is O(N), i.e., it is linear in the number of data samples (see Appendix B.2). Two-dimensional grid search can also be employed to obtain the solution. Finally, the estimate of the parameter τ is neither needed for detection nor for estimating a, |b|, but is

²Since $a = h_2 h_1$ and $\tau = |h_2|^2$, the parameters a and τ are not completely decoupled. In particular, a can take any complex value when $\tau \neq 0$, but a = 0 when $\tau = 0$. However, we have ignored this coupling between a and τ to simplify the derivation of the DML algorithm, since the case $h_2 = 0$ is not of practical interest.

provided for completeness and is given by

$$\tau = \max\left\{0, \frac{1}{A^2} \left[\frac{1}{N\sigma^2} \sum_{i=1}^{N} \left(|z_i - Aas_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aas_{1k}|\right)^2 - 1\right]\right\}.$$
 (3.10)

The DML estimator in (3.9) has the following interpretation. Let $\tilde{z}_i(u) \triangleq z_i - Aus_{1i}$, $i = 1, \ldots, N$ be the "cleaned" versions of the received signal samples after self-interference has been removed using the complex value u as an estimate of a. The signals $\tilde{z}_i(u)$ are independently generated realizations of the RV $\tilde{z}(u) \triangleq A(a-u)s_1 + Abs_2 + Ah_2n + \eta$. The quantity

$$W_N(u) = \frac{1}{N-1} \sum_{i=1}^N \left(|\tilde{z}_i(u)| - \frac{1}{N} \sum_{k=1}^N |\tilde{z}_k(u)| \right)^2$$
(3.11)

which is the objective function in (3.9), scaled by $\frac{1}{N-1}$, also represents the sample-variance of the envelope of $\tilde{z}(u)$. We demonstrate in Section 3.4 that the value u = a which completely cancels the interference also minimizes the variance of the envelope of $\tilde{z}(u)$. Hence, the variance of the envelope of $\tilde{z}(u)$ may be seen as a measure of the level of self-interference.

By treating the transmitted symbols s_{2i} as deterministic unknowns, the estimator in (3.9) ignores the underlying structure of the phases ϕ_{2i} , $i = 1, \ldots, N$. As we shall see in Section 3.4, the DML estimator in (3.9) performs poorly for BPSK modulation as its objective function experiences infinitely many global minima at high SNR. As an alternative, we will consider a constrained ML (CML) approach by solving the likelihood function in (3.5) subject to the constraint that $\phi_{2i} \in S_M, i = 1, \ldots, N$. For higher order modulations, it is difficult to apply this approach and still obtain a closed-form objective function. However, as we shall see next, this approach becomes more feasible for BPSK modulation.

3.3.2 Constrained ML Approach for BPSK

As we will see shortly, a straightforward application of the CML approach for M = 2 is not feasible as it results in a 3-dimensional search. However, it is possible to obtain a CML-inspired estimator which possesses a closed-form objective function by utilizing just a few pilot symbols to estimate the phase ϕ_b .

Under BPSK modulation, the CML approach maximizes the same objective function

as in (3.5), but subject to the constraint that $s_{2i} = \pm \sqrt{P_2}$, $i = 1, \ldots, N$. It is sufficient to focus on the first term in (3.5).Let

$$\mathcal{J}(\boldsymbol{z}; a, |b|, \phi_b, \boldsymbol{s}_2) \triangleq \sum_{i=1}^{N} \left| z_i - Aas_{1i} - A|b| \sqrt{P_2} e^{j(\phi_{bi})} \right|^2$$
(3.12)

be our objective function. The resulting estimate of s_{2i} is

$$\hat{s}_{2i} = \sqrt{P_2} \operatorname{sgn} \left(\Re \{ e^{-j\phi_b} (z_i - Aas_{1i}) \} \right).$$
 (3.13)

Substituting (3.13) back into (3.12), we obtain

$$\mathcal{J}'(\boldsymbol{z}; a, |b|, \phi_b) = \sum_{i=1}^{N} |z_i - Aas_{1i}|^2 + NA^2 |b|^2 P_2 - 2A|b|\sqrt{P_2} \sum_{i=1}^{N} |\Re\{e^{-j\phi_b}(z_i - Aas_{1i})\}|.$$
(3.14)

From (3.14), we see that the CML estimate of |b| is given by

$$|\hat{b}|_{c} = \frac{1}{NA\sqrt{P_{2}}} \sum_{i=1}^{N} |\Re\{e^{-j\phi_{b}}(z_{i} - Aas_{1i})\}|.$$
(3.15)

Substituting (3.15) into (3.14), we obtain the updated objective function

$$\mathcal{J}''(\boldsymbol{z}; a, \phi_b) = \sum_{i=1}^{N} |z_i - Aas_{1i}|^2 - \frac{1}{N} \left(\sum_{i=1}^{N} |\Re\{e^{-j\phi_b}(z_i - Aas_{1i})\}| \right)^2$$
(3.16)

which we have to solve for a and ϕ_b . The CML estimate of ϕ_b in terms of a is the solution of the following maximization problem

$$\hat{\phi}_{bc} = \arg \max_{\psi \in [0,2\pi)} \sum_{i=1}^{N} |z_i - Aas_{1i}| |\cos(\angle (z_i - Aas_{1i}) - \psi)|.$$
(3.17)

Unfortunately, the maximization in (3.17) does not have a closed-form solution, which means that we cannot proceed to obtain an objective function that only depends on a, like the one in (3.9). Thus, a strict application of the CML approach requires the use of a 3-dimensional search to estimate a and ϕ_b .

34 Blind Estimation for Nonreciprocal Channels under MPSK Modulation

To get around this problem, we propose to replace ϕ_b in (3.16) with a pilot-based estimate of ϕ_b and then proceed to estimate a. We let $t_{1\ell}$, $t_{2\ell}$ and \ddot{z}_{ℓ} , $\ell = 1, \ldots, L$, be the L pilot symbols transmitted by \mathcal{T}_1 and \mathcal{T}_2 and the corresponding samples received at \mathcal{T}_1 , respectively. We may thus estimate ϕ_b by $\hat{\phi}_b = \angle \sum_{\ell=1}^L (\bar{z}_\ell - Aat_{1\ell})t_{2\ell}$. Substituting $\hat{\phi}_b$ into (3.16), we obtain the following estimate of a:

$$\hat{a}_{c} = \arg\min_{u \in \mathbb{C}} \sum_{i=1}^{N} |z_{i} - Aus_{1i}|^{2} - \frac{1}{N} \bigg(\sum_{k=1}^{N} \bigg| \Re \bigg\{ (z_{k} - Aus_{1k}) e^{-j \angle \sum_{\ell=1}^{L} (\bar{z}_{\ell} - Aut_{1\ell}) t_{2\ell}} \bigg\} \bigg| \bigg)^{2}.$$
(3.18)

We refer to the estimator in (3.18) as the modified CML (MCML) estimator.

3.3.3 Gaussian ML Approach

In deriving the DML and the MCML estimators, we treated the data symbols as deterministic unknowns. Another approach commonly used to deal with nuisance parameters is Gaussian ML estimation [76]. In this case, the data symbols s_{2i} , i = 1, ..., N are treated as i.i.d. complex Gaussian random variables with mean zero and variance P_2 . Under this assumption, the total noise variance becomes $\sigma_g^2 \triangleq P_2|b|^2 + \sigma^2(A^2\tau + 1)$. The resulting log-likelihood function for a is

$$\mathcal{L}(\boldsymbol{z}; a) = -\frac{1}{\sigma_g^2} \|\boldsymbol{z} - Aa\boldsymbol{s}_1\|^2 - N\log(\pi\sigma_g^2).$$
(3.19)

Hence, the GML estimate of a is

$$\hat{a}_g = \frac{\sum_{i=1}^N s_{1i}^* z_i}{\sum_{i=1}^N |s_{1i}|^2} = \frac{1}{NAP_1} \sum_{i=1}^N s_{1i}^* z_i.$$
(3.20)

Thus, the GML estimate of *a* has the form of a computationally inexpensive sample-average. It can be easily updated at run-time by updating the average term as new samples arrive. The GML estimator will serve as a low-complexity benchmark with which to compare the performance of the proposed DML and MCML estimators.

3.4 Asymptotic Behavior Analysis

In this section, we analyze theoretically the asymptotic behavior of the proposed estimators. Since the parameters |b| and τ are not required for detection, our analysis focuses on the estimation of a. To simplify our notation, we let $C \triangleq \sigma^2(A^2\tau + 1)$ be the overall noise variance at \mathcal{T}_1 . Regarding the DML estimator, the fact that we treat the phases ϕ_{bi} , i = $1, \ldots, N$ as deterministic unknowns means that the number of real unknown parameters for N (complex) samples is N+4. Because the dimension of the parameter space grows linearly with the number of samples, the DML estimator falls within a special class of ML estimators that are based on "partially-consistent observations" [58]. As discussed in Chapter 2, the deterministic parameters can be classified into two groups. The first group are the incidental parameters ϕ_{bi} , $i = 1, \ldots, N$. Each incidental parameter affects only one sample. The other group of parameters, a, |b|, and τ , are the structural parameters which affect all received samples. As mentioned in Chapter 2, the estimation of structural parameters in the presence of incidental parameters is referred to as the Incidental Parameter Problem [57–60]. It is well-known that the asymptotic properties of ML estimators, such as consistency, which hold when the dimension of the parameter space is fixed do not necessarily hold in the presence of incidental parameters [58]. It thus becomes important to investigate the asymptotic behavior of the DML estimator. We do this by explicitly taking into account the fact that ϕ_{1i} and ϕ_{2i} are equiprobably and independently chosen from the set S_M . Regarding the powers P_1 , P_2 and P_r , the most common convention in TWRNs is to set $P_1 = P_2 = P_r$ (cf. [22, 28, 77]), even though some works use slightly different setups. For instance, the authors in [21] choose $P_2 = 2P_1$ and $P_r = \frac{1}{2}(P_1 + P_2)$. In our analysis we consider the general case of $P_1 = \alpha P_2$ and $P_r = \beta (P_1 + P_2)$ for some $\alpha, \beta > 0$.

The first asymptotic property we address is the consistency of the DML estimator. We are chiefly concerned with the consistency of the estimator of a in (3.9), since |b| is not required for detection. We will prove that the estimator in (3.9) is consistent when the parameter spaces of a and b are restricted to compact sets. This is true even for BPSK modulation. The second aspect we address is how the DML estimator behaves at high SNR (we will define SNR shortly). We show that the estimator in (3.9) approaches the true channel with high probability at high SNR for M > 2. However, the objective function exhibits infinitely many global minima at high SNR for M = 2.

We also analyze the high SNR behavior of the MCML estimator in (3.18), showing that

it approaches the true channel with high probability for L = 1 (single pilot). For L > 1, the pilots can be chosen such that the MCML estimator always approaches the true channel at high SNR.

3.4.1 Consistency of the DML Estimator

In this section, we study the behavior of the DML estimator as $N \to \infty$. We demonstrate that the DML estimator is consistent when the channel parameters a, b belong to compact sets.

Before proceeding, we note that the estimator of a in (3.9) belongs to the class of extremum estimators. An estimator $\hat{\omega}$ is called an extremum estimator (cf. [78], [79]) if there is an objective function $\Sigma_N(\omega)$ such that $\hat{\omega} = \arg \min_{\omega \in \Omega} \Sigma_N(\omega)$, where Ω is the set of parameter values. The fundamental theorem for the consistency of extremum estimators can be summarized as follows:

Theorem 1 (Newey & McFadden [79, Ch. 36, Thm. 2.1]). If ω belongs to a compact set Ω and $\Sigma_N(\omega)$ converges uniformly in probability to $\Sigma_o(\omega)$ as $N \to \infty$, where $\Sigma_o(\omega)$ is continuous and uniquely minimized at $\omega = \omega_o$, then $\hat{\omega}$ converges in probability to ω_o .

Thus, we need to establish that, as $N \to \infty$, the objective function $W_N(u)$ in (3.11) converges uniformly in probability to a deterministic function of u which has a unique global minimum at u = a. Letting $v \triangleq a - u$ and $V_N(v) \triangleq W_N(a - v)$, we obtain

$$V_N(v) = \frac{1}{N-1} \sum_{i=1}^N \left(|y_i(v)| - \frac{1}{N} \sum_{k=1}^N |y_k(v)| \right)^2, \tag{3.21}$$

where $y_i(v) \triangleq \tilde{z}_i(a-v)$. The terms $y_i(v)$, i = 1, ..., N are independently generated realizations of the RV $y(v) \triangleq \tilde{z}(a-v)$, and $V_N(v)$ is the sample variance of the envelope of y(v). Let $\mathcal{V}(v)$ be the variance of |y(v)|, $\varphi_k \triangleq \frac{2\pi k}{M}$, and $\theta_k(v) \triangleq \phi_v - \phi_b + \varphi_k$ for k = 0, ..., M - 1. We show in Appendix A.1 that

$$\mathcal{V}(v) = A^2 |v|^2 P_1 + A^2 |b|^2 P_2 + C - \left(\sqrt{\frac{\pi C}{4M^2}} \sum_{k=0}^{M-1} L_{1/2} \left(-\lambda_k(v) \right) \right)^2, \tag{3.22}$$

where $L_{1/2}(\cdot)$ is the Laguerre polynomial [80] with parameter 1/2, and

$$\lambda_k(v) = \frac{1}{C} \left(A^2 |v|^2 P_1 + A^2 |b|^2 P_2 + 2A^2 |v| |b| \sqrt{P_1 P_2} \cos \theta_k(v) \right).$$
(3.23)

The behavior of $\mathcal{V}(v)$ for v = 0, which corresponds to u = a, is described in the following lemma:

Lemma 1. The variance $\mathcal{V}(v)$ of the random variable |y(v)| has a unique global minimum occurring at v = 0.

Proof. See Appendix A.2.

To apply Theorem 1, it remains to establish that $V_N(v)$ converges uniformly in probability to $\mathcal{V}(v)$. Since $V_N(v)$ is the sample variance of |y(v)| and |y(v)| has a finite fourth central moment, $V_N(v)$ converges in probability to $\mathcal{V}(v)$ [81]. The following lemma holds regarding uniform convergence in probability which is a stricter requirement than convergence in probability:

Lemma 2. Assuming a, b and v all belong to compact sets, then $V_N(v)$ converges uniformly in probability to $\mathcal{V}(v)$.

Proof. See Appendix A.3.

From Lemma 1 and Lemma 2, we see that all of the conditions of Theorem 1 are met if the channel parameters a and b belong to compact sets. Hence, the following theorem holds.

Theorem 2. If the channel parameters a, b belong to the compact sets C_1 and C_2 , then the following estimator of a:

$$\hat{a} = \arg \min_{u \in B_1} \sum_{i=1}^{N} \left(|z_i - Aus_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aus_{1k}| \right)^2$$

is consistent.

The compactness assumption for a and b can be satisfied by assuming that the magnitudes of g_1 , h_1 and h_2 are bounded. If we treat g_1 , h_1 and h_2 as complex Gaussian random variables, there is no upper bound on |a| and |b|, strictly speaking, but we can always choose a sufficiently large ξ such that $Prob(|a|, |b| \leq \xi) = 1 - \epsilon$, where ϵ can be made arbitrarily small.

3.4.2 High SNR Behavior of the DML Estimator

We now investigate the behavior of the DML estimator at high transmit SNR. The SNR is defined as $\gamma \triangleq \frac{P_2}{\sigma^2}$. Let

$$X(v) = \sum_{i=1}^{N} \left(|Avs_{1i} + Abs_{2i}| - \frac{1}{N} \sum_{k=1}^{N} |Avs_{1k} + Abs_{2k}| \right)^2$$
(3.24)

be the objective function in (3.9) in the limit as $\sigma \to 0$. The following lemma describes the behavior of the DML estimator as $\sigma \to 0$.

Lemma 3. For fixed (finite) N, the DML estimator approaches the true channel a as $\sigma \to 0$, except when the data symbols are such that the phase differences $\phi_{1i} - \phi_{2i}$, i = 1, ..., N take at most two distinct values, in which case the objective function encounters infinitely many global minima as $\sigma \to 0$. Therefore, assuming M-PSK transmission, $\hat{a} \to a$ as $\sigma \to 0$ with probability $P_{M,N} = 1 - \left(\frac{2}{M}\right)^{N-1} (M-1)$.

Proof. See Appendix A.4.

As we can see from Lemma 3, the DML estimator approaches the true channel except in the event that the phase differences $\phi_{1i} - \phi_{2i}$, $i = 1, \ldots, N$ happen to take at most two distinct values. When this event occurs, the objective function of the estimator will have infinitely many global minima at high SNR. In fact, the occurrence of this event also results in a singular Fisher information matrix as we shall see in Section 3.4.5. For M > 2, this event is unlikely for sufficiently large N. For instance, at M = 4, the probability of this event is less than 5.6×10^{-9} for $N \ge 30$. Thus, the DML estimator approaches the true channel with high probability for M > 2 as long as the sample size is not very small. In this case, the average MSE performance of the estimator keeps improving with SNR, and it can effectively achieve arbitrary accuracy for sufficiently high SNR. For M = 2, however, $P_{2,N} = 0$ and the estimator always encounters infinitely many global minima at high SNR because the difference $\phi_{1i} - \phi_{2i}$ cannot take more than two distinct values for $i = 1, \ldots, N$. In fact, for M = 2 and for any non-zero v, b, there are only two possible values that the terms $|vs_{1i} + bs_{2i}|$, i = 1, ..., N can take, and they are $|\sqrt{P_1}v + \sqrt{P_2}b|$ and $|\sqrt{P_1}v - \sqrt{P_2}b|$. Whenever $v \perp b$, these two values are equal, which means that the terms $|vs_{1i} + bs_{2i}|, i = 1, \ldots, N$ are all equal regardless of the values of $s_{1i}, s_{2i}, i = 1, \ldots, N$, i.e., X(v) = 0. Therefore, all values of v such that $v \perp b$ are global minimizers of the

objective function X(v), and the estimator is not able to identify the true channel. Hence, the DML estimator performs poorly for M = 2 and its MSE performance deteriorates at high SNR. Despite this behavior, the DML estimator is consistent for M = 2 for any fixed $\sigma > 0$, because for $\sigma > 0$ only v = 0 minimizes the variance $\mathcal{V}(v)$. This explains why, as we shall see, the DML estimator performs better at low SNR than high SNR for M = 2. Since the performance of the estimator will also degrade for very low SNR, the MSE for M = 2 exhibits a U-shaped behavior when plotted versus SNR. This is confirmed by our simulation results in Section 3.5.

3.4.3 High SNR Behavior of the MCML Estimator (M=2)

We now investigate the high SNR behavior of the MCML estimator and show that it effectively avoids the problem of infinitely many global minima at high SNR. Let

$$Y(v) \triangleq \sum_{i=1}^{N} |Avs_{1i} + Abs_{2i}|^2 - \frac{1}{N} \left(\sum_{k=1}^{N} |Avs_{1k} + Abs_{2k}| \cdot |\cos(\Delta_k)| \right)^2$$
(3.25)

be the objective function in (3.18) as $\sigma \to 0$, expressed in terms of v = a - u, where

$$\Delta_k = \angle (Avs_{1k} + Abs_{2k}) - \angle (LAbP_2 + Av\sum_{\ell=1}^L t_{1\ell}t_{2\ell}).$$
(3.26)

Clearly, $Y(v) \ge 0$ and Y(0) = 0. It remains to investigate whether it is possible to have Y(v) = 0 for $v \ne 0$, i.e., whether there could be other global minima besides v = 0. For this to happen, all terms $|Avs_{1k} + Abs_{2k}|$, k = 1, ..., N have to be equal and all terms $|\cos(\Delta_k)|$, k = 1, ..., N have to be equal to 1. The first requirement can be met if $v \perp b$ or if the products $s_{1i}s_{2i}$ are all equal for i = 1, ..., N. As for the second requirement, it is only satisfied when

$$\angle (Avs_{1k} + Abs_{2k}) = \angle (LAbP_2 + Av\sum_{\ell=1}^{L} t_{1\ell}t_{2\ell}) \mod \pi$$
(3.27)

for k = 1, ..., N. Let \mathcal{H} be the event that the requirement in (3.27) is satisfied for k = 1, ..., N. For L = 1, we have $P(\mathcal{H}) = 1/2^N$. For L > 1, however, the occurrence of \mathcal{H} can be completely avoided if the pilots are chosen such that the products $t_{1\ell}t_{2\ell}$, $\ell = 1, ..., L$ are

not all equal. If the pilots are randomly chosen, then $P(\mathcal{H}) = 1/2^{N+L-1}$. Hence, depending on the choice of the pilots, the MCML estimator either always approaches the true channel or approaches it with a probability of $1 - 1/2^{N+L-1}$. Thus, the MCML estimator effectively resolves the problem of infinitely many global minima at high SNR even when a single pilot is used.

3.4.4 MSE Performance of the GML Estimator

The estimator of a in (3.20) can be expanded as

$$\hat{a}_g = a + \frac{b}{N\sqrt{\alpha}} \sum_{i=1}^N e^{j(\phi_{2i} - \phi_{1i})} + \frac{1}{N\sqrt{\alpha}P_2} \sum_{i=1}^N e^{-j\phi_{1i}} h_2 n_i + \frac{1}{NA\sqrt{\alpha}P_2} \sum_{i=1}^N e^{-j\phi_{1i}} \eta_i.$$
(3.28)

It is straightforward to check that the estimator is unbiased. The resulting MSE is

$$\mathbb{E}\{|\hat{a}_g - a|^2\} = \frac{|b|^2}{N\alpha} + \frac{|h_2|^2\sigma^2}{N\alpha P_2} + \frac{\sigma^2}{NA^2\alpha P_2}.$$
(3.29)

Since $\mathbb{E}\{|\hat{a}_g - a|^2\} \to 0$ as $N \to \infty$, the estimator is consistent. Clearly, as $\sigma \to 0$, the MSE performance of this estimator is limited by an error floor of $|b|^2/N\alpha$.

3.4.5 Cramer-Rao Bounds

In this section, we obtain CRBs for the estimation of a and |b|. The first bound is derived by treating the data symbols s_{21}, \ldots, s_{2N} as deterministic unknowns. We exclude the parameter τ from our CRB derivation since its Fisher information is decoupled from the Fisher information of the other parameters. Let $\boldsymbol{\theta}_R \triangleq [\Re\{a\}, \Im\{a\}, |b|, \phi_{b1}, \ldots, \phi_{bN}]^T$ be the vector of real unknown parameters (excluding τ), and let $\boldsymbol{I}(\boldsymbol{\theta}_R)$ be the corresponding Fisher information matrix (FIM). The matrix $\boldsymbol{I}(\boldsymbol{\theta}_R)$ is given by

$$\boldsymbol{I}(\boldsymbol{\theta}_R) = \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{J}_2 \\ \boldsymbol{J}_2^T & \boldsymbol{J}_3 \end{bmatrix}, \qquad (3.30)$$

where

$$\boldsymbol{J}_{1} = \frac{2A^{2}}{C} \begin{bmatrix} NP_{1} & 0 & \Re\{e^{j\phi_{b}}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{2}\}\\ 0 & NP_{1} & \Im\{e^{j\phi_{b}}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{2}\}\\ \Re\{e^{j\phi_{b}}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{2}\} & \Im\{e^{j\phi_{b}}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{2}\} & NP_{2} \end{bmatrix},$$
(3.31)

$$\boldsymbol{J}_{2} = \frac{2A^{2}}{C} \begin{bmatrix} \Im\{b^{*}s_{11}s_{21}^{*}\} & \dots & \Im\{b^{*}s_{1N}s_{2N}^{*}\}\\ \Re\{b^{*}s_{11}s_{21}^{*}\} & \dots & \Re\{b^{*}s_{1N}s_{2N}^{*}\}\\ 0 & \dots & 0 \end{bmatrix},$$
(3.32)

and

$$\boldsymbol{J}_{3} = \frac{2A^{2}}{C} |b|^{2} P_{2} \boldsymbol{I}_{N}.$$
(3.33)

The resulting CRBs on the estimation of a and |b| are

$$CRB_a = [\mathbf{I}^{-1}(\boldsymbol{\theta}_R)]_{11} + [\mathbf{I}^{-1}(\boldsymbol{\theta}_R)]_{22},$$
 (3.34)

and

$$CRB_{|b|} = [\boldsymbol{I}^{-1}(\boldsymbol{\theta}_R)]_{33}.$$
(3.35)

Moreover, using the Schur-complement property and letting $\tilde{\boldsymbol{J}} \triangleq \boldsymbol{J}_1 - \boldsymbol{J}_2 \boldsymbol{J}_3^{-1} \boldsymbol{J}_2^T$, we obtain³

$$CRB_a = [\tilde{\boldsymbol{J}}^{-1}]_{11} + [\tilde{\boldsymbol{J}}^{-1}]_{22},$$
 (3.36)

and

$$CRB_{|b|} = [\tilde{\boldsymbol{J}}^{-1}]_{33}.$$
 (3.37)

The bounds CRB_a and $CRB_{|b|}$ are obtained by treating the data as deterministic parameters. Hence, they apply for the class of estimation algorithms that treat the data as deterministic, such as the proposed DML estimator. The bounds exist whenever \tilde{J} is invertible. It can be shown that $\det(\tilde{J}) = 0$ whenever the data symbols are such that the differences $\phi_{1i} - \phi_{2i}$, $i = 1, \ldots, N$ take at most two distinct values. This is the same condition that results in an objective function with infinitely many global minima at high SNR. Hence, the bounds do not exist for BPSK modulation since this condition is always met for M = 2. Since the bounds in (3.36) and (3.37) are both functions of s_1 and s_2 , they hold for the particular realizations of s_1 and s_2 under consideration. In Section 3.5, we use Monte-Carlo simulations to average CRB_a and $CRB_{|b|}$ over many realizations of s_1 and s_2 .

Another variant of the CRB commonly used in the presence of random nuisance parameters is the MCRB, introduced in Chapter 2. Unlike the previous bound, the MCRB

³Although it is possible to evaluate the expressions in (3.36) and (3.37) and obtain closed-form expressions for CRB_a and $CRB_{|b|}$, the resulting expressions are quite lengthy and are omitted for the sake of brevity.

takes into account the statistics of the data symbols, and thus it applies for a wider class of estimators. The MCRB is obtained using the MFIM matrix (see Chapter 2, Eq. 2.15). Letting $\boldsymbol{\theta}' \triangleq [\Re\{a\}, \Im\{a\}, |b|]^T$, the MFIM is given by

$$\boldsymbol{I}_{m}(\boldsymbol{\theta}') \triangleq \mathbb{E}\{\boldsymbol{J}_{1}\} = \frac{1}{C} \begin{bmatrix} A^{2}NP_{1} & 0 & 0\\ 0 & 2A^{2}NP_{1} & 0\\ 0 & 0 & 2A^{2}NP_{2} \end{bmatrix}.$$
 (3.38)

Hence,

$$MCRB_a = \frac{C}{A^2 N P_1}$$
 and $MCRB_{|b|} = \frac{C}{2A^2 N P_2}$. (3.39)

The bounds in (3.39) have the advantage of possessing a simple closed form. However, they are not as tight as the bounds in (3.36) and (3.37) when considering estimators that treat the data as deterministic. Hence, we will only consider the bounds in (3.36) and (3.37) in our simulation results.

3.5 Simulation Results

In this section, we use Monte-Carlo simulations to numerically investigate the performance of the proposed algorithms. Our results are obtained assuming $P_r = P_1 = P_2$, and they are averaged using a set of 300 independent realizations of the channel parameters h_1 , h_2 , g_1 and g_2 . These realizations are generated by modelling h_1 and h_2 as correlated complex Gaussian random variables with mean zero, variance 1, and correlation coefficient $\rho = 0.3$. Similarly we model g_1 and g_2 as correlated complex Gaussian random variables with the same mean, variance and correlation coefficient, but independent of h_1 and h_2 . To generate correlated complex Gaussian random variables we follow the approach proposed in [82]. The DML and MCML estimates are obtained using a two-dimensional grid-search with a step-size of 10^{-3} . Unless otherwise mentioned, MSE results are for the estimation of a.

We begin by comparing the MSE performance of the DML, GML and MCML estimators for M = 2. We do not show the CRB in this case, since the FIM is singular. The MSE performance of the three estimators is plotted versus SNR⁴ for N = 45 in Fig. 3.1 and versus N for an SNR of 20 dB in Fig. 3.2. For the MCML estimator, 2 pilots are employed to obtain an estimate of ϕ_b . Both plots show that the DML estimator performs poorly for

⁴The SNR is defined as $10 \log \frac{P_2}{\sigma^2}$

BPSK modulation and is outperformed by the GML and MCML estimators. Fig. 3.1 shows the U-shaped behavior of the DML estimator for M = 2 described in Section 3.4.2. The MCML estimator is superior to the GML estimator except at very low SNR. Moreover, as the SNR increases, the MCML estimator improves steadily while the GML estimator encounters an error floor. Fig. 3.2 also demonstrates the superiority of the MCML estimator to the GML estimator.

Next, we study the MSE performance of the DML and GML estimators for M = 4 (QPSK modulation). The MSE performance of the two estimators is plotted versus SNR for N = 45 in Fig. 3.3 and versus N for an SNR of 20 dB in Fig. 3.4. The bound CRB_a is included as a reference in both plots. As we can see in Fig. 3.3, the DML estimator outperforms the GML estimator, except at low SNR. Moreover, the MSE performance of the DML estimator improves steadily with SNR and approaches CRB_a , while that of the GML estimator goes below CRB_a at low SNR. This should not be a surprise since CRB_a is derived by treating s_2 as deterministic, and the GML estimator is biased in this case. As we can see in Fig. 3.4, both estimators improve as N increases, but the DML estimator is much closer to CRB_a .

In Fig. 3.5, the MSE performance of the DML and the GML estimators for the estimation of |b| and the associated CRB are plotted versus SNR for M = 4. For the GML estimator, we obtain an estimate of |b| by substituting \hat{a}_g in (3.7). Fig. 3.5 shows that the GML estimator is slightly better except at high SNR where it appears to encounter an error floor and the DML estimator becomes better and approaches $CRB_{|b|}$.

Our next goal is to compare the SER performance of the DML estimator with that of the training-based LS estimator in order to investigate the tradeoffs between accuracy and spectral efficiency that result from following the blind approach. We focus on small sample sizes since this is more suitable for modern-day cellular systems. We note that, when channels are nonreciprocal, the training-based LS estimator is an efficient estimator that coincides with the training-based ML estimator. As a reference, we also plot the SER performance assuming perfect channel knowledge. The phase ϕ_b is estimated blindly using the Viterbi-Viterbi algorithm⁵ and a small number of pilots is employed to resolve the resulting *M*-fold ambiguity using the unique word method [83]. In Fig. 3.6, we show the

⁵Using the Viterbi-Viterbi algorithm, we estimate ϕ_b by $\hat{\phi}_b \triangleq \frac{1}{M} \angle \sum_{i=1}^N |\tilde{z}_i(\hat{a})|^2 e^{jM \angle \tilde{z}_i(\hat{a}) + \pi}$, where $\tilde{z}_i(\hat{a}), i = 1, \ldots, N$ are the resulting N samples after the estimate \hat{a} is used to remove self-interference.

44 Blind Estimation for Nonreciprocal Channels under MPSK Modulation

SER performance of the two algorithms assuming that the channel is fixed for the duration of 20 samples. For the DML estimator, 2 pilots and 18 data symbols are transmitted. The 2 pilots are used to resolve the M-fold ambiguity, and all 20 samples are used to blindly estimate a. For the LS estimator, we estimate a and b using 4 pilot symbols⁶ and we transmit 16 data symbols. As we can see from Fig. 3.6, the SER performance of the DML estimator is very close to that of the LS estimator (approximately 0.6 dB away). In Fig. 3.7, we assume that the channel is fixed for the duration of 40 samples. In the DML case 4 pilots and 36 data symbols are transmitted, and in the LS case 8 pilots and 32 data symbols are transmitted. The performance of DML estimator is again very close to that of the LS estimator (approximately 0.4 dB away), and it is only 1.5 dB away from the performance under perfect CSI. In both examples, for the DML algorithm we use 90% of the channel coherence time to transmit data and 10% to transmit pilots, while for the LS algorithm we are use 80% of the coherence time to transmit data and 20% to transmit pilots, which demonstrates that the DML estimator offers a better tradeoff between accuracy and spectral efficiency.

3.6 Conclusions

In this chapter, we proposed the DML algorithm for blind channel estimation in AF TWRNs employing *M*-PSK modulation. The DML estimator was derived by treating the data symbols as deterministic unknowns. For comparison, we also derived the GML estimator by treating the data symbols as Gaussian-distributed nuisance parameters. We showed that the DML estimator is consistent. For M > 2, we showed that its MSE performance is superior to that of the GML estimator for medium-to-high SNR and that it approaches the true channel with high probability as the SNR increases. In contrast, the GML estimator suffers from an error floor at high SNR. We also compared the SER performance of the DML estimator with that of the training-based LS estimator and demonstrated that the DML approach provides a better tradeoff between accuracy and spectral efficiency. For the case M = 2 where the DML estimator performs poorly, we proposed the MCML estimator which explicitly takes into account the structure of the BPSK signal. This estimator outperforms the GML estimator except at very low SNR and approaches the true channel at high SNR.

⁶For the training-based LS algorithm, the pilots are chosen such that the pilot vectors from the two terminals are orthogonal to each other.



Fig. 3.1 Average MSE of the DML, GML and MCML algorithms for the estimation of a, plotted versus SNR for M = 2 and N = 45.



Fig. 3.2 Average MSE of the DML, GML and MCML algorithms for estimation of a, plotted versus N for M = 2 and an SNR of 20 dB.



Fig. 3.3 Average MSE of the DML and GML algorithms for the estimation of a and the bound CRBa plotted versus SNR for M = 4 and N = 45.



Fig. 3.4 Average MSE of the DML and GML algorithms for the estimation of a and the bound CRBa plotted versus N for M = 4 and an SNR of 20 dB.



Fig. 3.5 Average MSE of the DML and GML algorithms for the estimation of |b| and the bound $CRB_{|b|}$ plotted versus SNR for M = 4 and N = 45.



Fig. 3.6 Average SER versus SNR for the DML and LS estimators for M = 4, assuming the channel is fixed for 20 samples. We use 2 pilots to resolve the M-fold ambiguity in ϕ_b for the DML estimator, and 4 pilots for LS estimation.



Fig. 3.7 Average SER versus SNR for the DML and LS estimators for M = 4, assuming the channel is fixed for 40 samples. We use 4 pilots to resolve the M-fold ambiguity in ϕ_b for the DML estimator, and 8 pilots for LS estimation.

Chapter 4

Blind Channel Estimation for Reciprocal Flat-fading Channels under MPSK-modulation

4.1 Introduction

In the previous chapter, we considered blind channel estimation under nonreciprocal flatfading channel conditions for AF TWRNs that employ CM signalling in the form of M-PSK modulation. Treating the data symbols as deterministic unknowns, we derived the DML channel estimator. We also proved that the DML estimator is consistent and that it approaches the true channel with high probability as SNR increases for M > 2.

In this chapter, we shift our focus from nonreciprocal to reciprocal flat-fading channels. Assuming CM signalling, we derive the DML estimator for the reciprocal case and investigate its asymptotic behavior. We show that for M > 2, the DML estimator approaches the true channel with high probability at high SNR. However, in contrast to the nonreciprocal case, we prove that the DML estimator is not consistent. As an alternative to the DML estimator, we propose to estimate the channel by minimizing the sample variance of the envelope of the received signal after self-interference cancellation. This criterion is inspired by the DML estimator for nonreciprocal channels in Chapter 3. We refer to this estimator as the minimum sample envelope variance (MSEV) estimator. The asymptotic behavior of the MSEV estimator is similar to that of the DML estimator in Chapter 3, i.e., it is consistent and approaches the true channel with high probability at high SNR for M > 2. We also derive two CRBs as estimation performance benchmarks for the case of reciprocal channels. The first bound is obtained by treating the data symbols as deterministic unknown parameters while the second is the MCRB. Monte-Carlo simulations are then used to obtain the mean-squared error (MSE) of the two estimators, demonstrating that the MSEV estimator outperforms the DML estimator. In summary, the main contributions of this chapter are: (i) analysis of the high SNR performance of the DML estimator for reciprocal channels; (ii) investigation of the consistency of the DML estimator for reciprocal channels; (iii) application of the MSEV criterion to reciprocal channels; (iv) derivation of two CRBs on the variance of unbiased estimators for reciprocal channels.

The work presented in this chapter first appeared in [84]. A more elaborate version with new theoretical contributions appeared in [85].

The remainder of this chapter is organized as follows. In Section 4.2, we present our system model. In Section 4.3, we present the DML and MSEV estimators for the case of reciprocal channels. In Section 4.4, we analyze the high SNR behavior and the consistency of the two estimators. The CRBs are derived in Section 4.5. Our simulation results are shown in Section 4.6. Finally, our conclusions are in Section 4.7.

4.2 System Model

The system model considered in this chapter is the same as the one in Chapter 3, with the exception that the channels are reciprocal, i.e., that $h_1 = h_2 = h$ and $g_1 = g_2 = g$. Hence, the received signal at \mathcal{T}_1 during the second phase of transmission is given by

$$z = Aas_1 + Abs_2 + Ahn + \eta, \tag{4.1}$$

where $a \triangleq h^2$, $b \triangleq gh$, and s_1 , s_2 , n and η have the same definitions as in Chapter 3. The channel coefficients h and g are modelled as $\mathcal{CCN}(0, \gamma^2)$. As in Chapter 3, it is sufficient for detection purposes to know a and $\phi_b \triangleq \angle b$.

4.3 Channel Estimation Algorithms

As in Chapter 3, the vectors $\boldsymbol{z} \triangleq [z_1 \dots z_N]^T$, $\boldsymbol{s}_1 \triangleq [s_{11}, \dots, s_{1N}]^T$, $\boldsymbol{s}_2 \triangleq [s_{21}, \dots, s_{2N}]^T$, $\boldsymbol{n} = [n_1, \dots, n_N]$, and $\boldsymbol{\eta} = [\eta_1, \dots, \eta_N]$ denote the received vector at \mathcal{T}_1 , the transmitted symbol vectors of \mathcal{T}_1 and \mathcal{T}_2 , the noise vector at \mathcal{R} and the noise vector at \mathcal{T}_1 during N successive transmissions, respectively.

We begin by deriving the DML estimator for the reciprocal case. Similar to Chapter 3, the DML is derived by treating the data symbols as deterministic unknowns. The unknown parameters are collected in the vector $\boldsymbol{\theta} \triangleq [a, |b|, \psi_1, \dots, \psi_N]^T$ where $\psi_i \triangleq \phi_{2i} + \phi_b$, $i = 1, \dots, N$. As before, the DML estimate of ψ_i is $\hat{\psi}_i = \angle(z_i - A\hat{a}s_{1i})$ and that of |b| is $|\hat{b}| = \frac{1}{NA\sqrt{P_2}} \sum_{i=1}^N |z_i - A\hat{a}s_{1i}|$. The DML estimator of a is given by

$$\hat{a} = \arg\min_{u \in \mathbb{C}} \left\{ \frac{\sum_{i=1}^{N} \left(|z_i - Aus_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aus_{1k}| \right)^2}{\sigma^2 (A^2 |u| + 1)} + N \log \left(A^2 |u| + 1 \right) \right\}.$$
(4.2)

The structure of the DML estimator of a in (4.2) is clearly different from that of the DML estimator in the nonreciprocal case (Chapter 3, (3.9)). In particular, the objective function in (3.9) appears as the numerator of the first term in (4.2), but (4.2) additionally has a denominator term and a logarithmic term that are not present in (3.9). As pointed out in Chapter 3, the DML estimator of a in the nonreciprocal case minimizes the sample variance of the envelope of the received signal after self-interference cancellation. Given the differences between (3.9) and (4.2), it would be insightful to apply the criterion used in (3.9) to the case of reciprocal channels and compare the resulting performance with that of the DML estimator in (4.2). Hence, we will consider the estimator

$$\hat{a}_{v} = \arg\min_{u \in \mathbb{C}} \sum_{i=1}^{N} \left(|\tilde{z}_{i}(u)| - \frac{1}{N} \sum_{k=1}^{N} |\tilde{z}_{k}(u)| \right)^{2}$$
(4.3)

where $\tilde{z}_i(u) \triangleq (z_i - Aus_{1i})$, i = 1, ..., N. The signals $\tilde{z}_1(u), ..., \tilde{z}_N(u)$ can be viewed as realizations of the RV $\tilde{z}(u) = A(a-u)t_1 + Abt_2 + Ahn + \eta$, and the objective function in (4.3) can be seen, after scaling it by $\frac{1}{N-1}$, as the sample variance of $|\tilde{z}(u)|$. Since the above estimator does not retain its DML interpretation in the reciprocal case, we will refer to it as the minimum sample envelope variance (MSEV) estimator.

The solutions for (4.2) and (4.3) may be obtained using numerical methods such as the steepest-descent algorithm or quasi-Newton type algorithms [75]. Since the objective functions in (4.2) and (4.3) are nonconvex, the performance of such methods will depend on the availability of good initial estimates. A simple way to obtain an initial estimate is the sample average estimator employed in Chapter 3 (see Eq. (3.20)). The steepest descent implementation for (4.2) and (4.3) is detailed in Appendix B.2, and its computational complexity is O(N), i.e., it is linear in the number of data samples. In Section 4.6, we show that steepest descent algorithm yields almost identical performance to two-dimensional grid search, and that it requires only a small number of iterations to converge.

4.4 Asymptotic Behavior Analysis

As we did in Chapter 3, we focus in our analysis on the estimation of a since |b| is not required for detection under *M*-PSK modulation. We consider the consistency of the two estimators and their behavior at high transmit SNR. The transmit SNR is defined as $\gamma \triangleq \frac{P_2}{\sigma^2}$.

4.4.1 High SNR Behavior of the two Estimators

As the MSEV estimator has the same structure as the estimator in (3.9), the high SNR behavior in Lemma 3 of Chapter 3 applies to the MSEV estimator as well. Hence, for fixed N, the MSEV estimator approaches the true channel as $\sigma \to 0$ with probability $1 - \left(\frac{2}{M}\right)^{N-1} (M-1)$, i.e., it approaches the true channel with high probability for M > 2.

We now consider the high SNR behavior of the DML estimator in (4.2). We first note that multiplying the objective function in (4.2) with σ^2 does not affect the solution. After multiplying the objective function in (4.2) with σ^2 and taking the limit as $\sigma \to 0$, we obtain the objective function

$$\tilde{X}(v) = \frac{X(v)}{A^2 |a - v| + 1}$$
(4.4)

where, as in Chapter 3, $v \triangleq a - u$ and

$$X(v) = \sum_{i=1}^{N} \left(|Avs_{1i} + Abs_{2i}| - \frac{1}{N} \sum_{k=1}^{N} |Avs_{1k} + Abs_{2k}| \right)^2.$$
(4.5)

From Lemma 3 in Chapter 3, we know that, with probability $P_{M,N} = 1 - \left(\frac{2}{M}\right)^{N-1} (M-1)$, X(v) has a unique global minimum at v = 0, which corresponds to u = a. Since X(0) = 0, it is clear from (4.4) that $\tilde{X}(v)$ also has a unique global minimum at v = 0 with probability $P_{M,N}$. Hence, the DML estimator also approaches the true channel with high probability

as SNR increases for M > 2.

4.4.2 Consistency of Estimators

In Chapter 3, it was proved that the DML estimator is consistent under the nonreciprocal channel assumption. It can be easily verified that the proof in Chapter 3 is not affected if channel reciprocity is assumed. Hence, the MSEV estimator is consistent.

We now investigate the consistency of the DML estimator. Let $Y_N(u)$ be the objective function of the DML estimator in (4.2) scaled by $\frac{1}{N-1}$. Thus,

$$Y_N(u) \triangleq \frac{W_N(u)}{\sigma^2(A^2|u|+1)} + \frac{N}{N-1}\log\left(A^2|u|+1\right),$$
(4.6)

where

$$W_N(u) \triangleq \frac{1}{N-1} \sum_{i=1}^N \left(|\tilde{z}_i(u)| - \frac{1}{N} \sum_{k=1}^N |\tilde{z}_k(u)| \right)^2$$
(4.7)

is the the scaled version of the objective function of the MSEV estimator, which is also the sample variance of $|\tilde{z}(u)|$. As $N \to \infty$, $Y_N(u)$ converges in probability to

$$\mathcal{Y}(u) \triangleq \frac{\mathcal{W}(u)}{\sigma^2(A^2|u|+1)} + \log\left(A^2|u|+1\right),\tag{4.8}$$

where

$$\mathcal{W}(u) = A^2 |a - u|^2 P_1 + A^2 |b|^2 P_2 + \sigma^2 (A^2 |a| + 1) - \frac{\pi \sigma^2 (A^2 |a| + 1)}{4M^2} \left(\sum_{k=0}^{M-1} L_{1/2} \left(-\lambda_k (a - u) \right) \right)^2$$
(4.9)

is the variance of $|\tilde{z}(u)|$ (see Chapter 3, (3.22)) and

$$\lambda_k(v) \triangleq \frac{1}{\sigma^2(A^2|a|^2+1)} \left(A^2|v|^2 P_1 + A^2|b|^2 P_2 + 2A^2|v||b|\sqrt{P_1 P_2} \cos\left(\angle v - \phi_b + 2\pi k/M\right) \right).$$
(4.10)

To find out whether $\mathcal{Y}(u)$ has an extremum at u = a, we analyze the behavior of the partial derivatives $\dot{\mathcal{Y}}_{\Re}(u) \triangleq \frac{\partial \mathcal{Y}(u)}{\partial \Re\{u\}}$ and $\dot{\mathcal{Y}}_{\Im}(u) \triangleq \frac{\partial \mathcal{Y}(u)}{\partial \Im\{u\}}$. Because of the symmetry of $\mathcal{Y}(u)$

with respect to $\Re\{u\}$ and $\Im\{u\}$, it is sufficient to consider $\mathcal{Y}_{\Re}(u)$. We have

$$\dot{\mathcal{Y}}_{\Re}(u) = \frac{A^2 \Re\{u\}}{|u|(A^2|u|+1)} \left(1 - \frac{\mathcal{W}(u)}{\sigma^2(A^2|u|+1)}\right) + \frac{1}{\sigma^2(A^2|u|+1)} \frac{\partial \mathcal{W}(u)}{\partial \Re\{u\}}.$$
(4.11)

Moreover, it can be verified that $\frac{\partial \mathcal{W}(u)}{\partial \Re\{u\}}\Big|_{u=a} = 0$. Therefore,

$$\dot{\mathcal{Y}}_{\Re}(a) = \frac{A^2 \Re\{a\}}{|a|(A^2|a|+1)} \left(1 - \frac{\mathcal{W}(a)}{\sigma^2(A^2|a|+1)}\right).$$
(4.12)

For the factor $\left(1 - \frac{\mathcal{W}(a)}{\sigma^2(A^2|a|+1)}\right)$ we have the following lemma

Lemma 4. For any $a \in \mathbb{C}$, $\left(1 - \frac{\mathcal{W}(a)}{\sigma^2(A^2|a|+1)}\right) > 0$.

Proof. See Appendix B.1.

Hence, $\dot{\mathcal{Y}}_{\Re}(a)$ is zero only when $\Re\{a\} = 0$. Similarly, $\dot{\mathcal{Y}}_{\Im}(a)$ is zero only when $\Im\{a\}$ is zero. Since $\mathcal{Y}(u)$ is differentiable at u = a (for $a \neq 0$) and u = a is not a boundary point, this implies that $\mathcal{Y}(u)$ does not have an extremum at u = a for $a \neq 0$. Hence, the DML estimator is not consistent [79]. The inconsistency of the DML estimator should not come as a surprise since the data symbols are treated as deterministic unknowns. Due to this assumption, the number of parameters is not fixed but grows linearly with the number of samples. Hence, as discussed in Chapter 2, the usual asymptotic properties of ML estimators, such as consistency, do not necessarily hold in this case.

4.5 Cramer-Rao Bounds

We derive two Cramer-Rao bounds for the estimation problem under consideration. The first bound is derived by treating the phases ψ_1, \ldots, ψ_N as deterministic unknowns. The vector of unknown real parameters is $\boldsymbol{\theta}_R \triangleq [\Re\{a\}, \Im\{a\}, |b|, \psi_1, \ldots, \psi_N]^T$, and the corresponding FIM is given by

$$\boldsymbol{I}(\boldsymbol{\theta}_R) = \begin{bmatrix} \boldsymbol{J}_1 & \boldsymbol{J}_2 \\ \boldsymbol{J}_2^T & \boldsymbol{J}_3 \end{bmatrix}, \qquad (4.13)$$
where

$$\boldsymbol{J}_{1} = \begin{bmatrix} \frac{2A^{2}NP_{1}}{\sigma^{2}(A^{2}|a|+1)} + \frac{A^{4}\Re\{a\}^{2}N}{|a|^{2}(A^{2}|a|+1)^{2}} & \frac{A^{4}\Re\{a\}\Im\{a\}N}{|a|^{2}(A^{2}|a|+1)^{2}} \\ \frac{A^{4}\Re\{a\}\Im\{a\}N}{|a|^{2}(A^{2}|a|+1)^{2}} & \frac{2A^{2}NP_{1}}{\sigma^{2}(A^{2}|a|+1)} + \frac{A^{4}\Im\{a\}^{2}N}{|a|^{2}(A^{2}|a|+1)^{2}} \end{bmatrix}$$

$$\left[\Re\{e^{j\phi_{b}}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{2}\} - \Im\{e^{j\phi_{b}}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{2}\} \right]$$

$$(4.14)$$

$$\boldsymbol{J}_{2}^{T} = \frac{2A^{2}}{\sigma^{2} (A^{2}|a|+1)} \begin{bmatrix} \Re\{b^{*}s_{11}s_{21}^{*}\} & \Re\{b^{*}s_{11}s_{21}^{*}\} \\ \Im\{b^{*}s_{1N}s_{2N}^{*}\} & \Re\{b^{*}s_{1N}s_{2N}^{*}\} \\ \Im\{b^{*}s_{1N}s_{2N}^{*}\} & \Re\{b^{*}s_{1N}s_{2N}^{*}\} \end{bmatrix},$$
(4.15)

and

$$\boldsymbol{J}_{3} = \frac{1}{\sigma^{2} \left(A^{2} |a|+1\right)} \left[2A^{2} N P_{2}, 2A^{2} |b|^{2} P_{2}, \dots, 2A^{2} |b|^{2} P_{2}\right]^{T}$$
(4.16)

Assuming $I(\boldsymbol{\theta}_R)$ is invertible, the CRB for the estimation of a is given by the sum of the first two diagonal entries in the inverse of $I(\boldsymbol{\theta}_R)$, i.e., $CRB_a = [I^{-1}(\boldsymbol{\theta}_R)]_{11} + [I^{-1}(\boldsymbol{\theta}_R)]_{22}$. Let \tilde{J} be the 2 × 2 top left submatrix of $I^{-1}(\boldsymbol{\theta}_R)$. Using the Schur-complement property, we have that $\tilde{J} = (J_1 - J_2 J_3^{-1} J_2^T)^{-1}$, i.e.,

$$CRB_a = \operatorname{tr}\left(\left(\boldsymbol{J}_1 - \boldsymbol{J}_2\boldsymbol{J}_3^{-1}\boldsymbol{J}_2^T\right)^{-1}\right).$$
(4.17)

Because the symbols s_1 are known and the symbols s_2 are treated as deterministic unknowns, CRB_a is a function of s_1 and s_2 , and it thus applies for the particular realizations of s_1 and s_2 under consideration.

As we did in Chapter 3, we will also consider the MCRB, commonly used in the presence of random nuisance parameters. Let $\boldsymbol{\theta}' \triangleq [\Re\{a\}, \Im\{a\}, |b|]^T$, the MFIM is given by

$$\boldsymbol{I}_{m}(\boldsymbol{\theta}') = \begin{bmatrix} \frac{2A^{2}NP_{1}}{\sigma^{2}(A^{2}|a|+1)} + \frac{A^{4}\Re\{a\}^{2}N}{|a|^{2}(A^{2}|a|+1)^{2}} & \frac{A^{4}\Re\{a\}\Im\{a\}N}{|a|^{2}(A^{2}|a|+1)^{2}} & 0\\ \frac{A^{4}\Re\{a\}\Im\{a\}N}{|a|^{2}(A^{2}|a|+1)^{2}} & \frac{2A^{2}NP_{1}}{\sigma^{2}(A^{2}|a|+1)} + \frac{A^{4}\Im\{a\}^{2}N}{|a|^{2}(A^{2}|a|+1)^{2}} & 0\\ 0 & 0 & \frac{2A^{2}NP_{2}}{\sigma^{2}(A^{2}|a|+1)} \end{bmatrix}.$$
(4.18)

The resulting MCRBs on a and |b| are given by

$$MCRB_{a} = [\mathbf{I}_{m}^{-1}(\boldsymbol{\theta}')]_{11} + [\mathbf{I}_{m}^{-1}(\boldsymbol{\theta}')]_{22}$$

= $\frac{4\sigma^{2}P_{1}(A^{2}|a|+1)^{2} + \sigma^{4}A^{2}(A^{2}|a|+1)}{4NA^{2}P_{1}^{2}(A^{2}|a|+1) + 2N\sigma^{2}A^{4}P_{1}},$ (4.19)

and

$$MCRB_{|b|} = [\mathbf{I}_m^{-1}(\boldsymbol{\theta}')]_{33} = \frac{\sigma^2(A^2|a|+1)}{2NA^2P_2}.$$
(4.20)

4.6 Simulation Results

In this section, we compare the MSE performance of the DML estimator and the MSEV estimator using Monte-Carlo simulations. Our results are generated assuming M = 4(QPSK) and $P_r = P_1 = P_2$, and are averaged over the same set of 100 realizations of the channel parameters g, h which are independently generated from the complex Gaussian distribution with mean zero and variance 1. The minimizers of the objective functions in (4.2) and (4.3) are obtained using the steepest descent (SD) algorithm. The initial points are obtained using the sample average estimator ((3.20)), while the step size is chosen using backtracking line search [74]. As a reference, we also show the MSE performance for the two estimators when the solutions for (4.2) and (4.3) are obtained using grid search (GS) with a step size of 10^{-3} . We also show the bounds CRB_a and $MCRB_a$, where CRB_a is averaged over many realizations of s_1 and s_2 .

Fig. 4.1 shows the average MSE of the two estimators versus SNR for N = 100. The MSEV estimator outperforms the DML estimator and the performance gap is most significant at low to medium SNR. At high SNR, both estimators approach the bound CRB_a . We can also see from Fig. 4.1 that, for both algorithms, the MSE performance is almost identical for the steepest descent implementation and the grid search implementation.

The bar plots in Fig. 4.2 show for both estimators the average number of SD iterations required for convergence and the average number of line search iterations to find the step size for a single steepest descent iteration, respectively. The average number of iterations ranges between 6.7 and 13.3 for the DML estimator and between 5.6 and 7.4 for the MSEV estimator. Hence, in both cases, only a small number of iterations is needed to achieve convergence, which shows that the SD algorithm is a reliable and computationally efficient method for solving (4.2) and (4.3). Moreover, the MSEV estimator on average requires less SD iterations and less linesearch iterations than the DML and is thus more efficient.

Fig. 4.3 shows the average MSE of the two estimators versus N for an SNR of 15dB. The MSEV estimator has a superior MSE performance which improves steadily as N increases. The gap between the MSE performances of the two estimators becomes more significant as the sample size increases, in accordance with the fact that the MSEV estimator is consistent

while the DML estimator is not.

4.7 Conclusions

In this chapter, we compared two blind channel estimation algorithms for AF TWRNs assuming channel reciprocity and *M*-PSK modulation. The first estimator was the DML estimator for reciprocal channels obtained by treating the data symbols as deterministic unknowns. The second was the MSEV estimator which minimized the sample variance of the envelope of the received signal after self-interference cancellation and was inspired by the DML estimator for nonreciprocal channels derived in Chapter 3. We showed that both estimators approach the true channel with high probability as SNR increases. However, the MSEV estimator is consistent while the DML estimator is not. We also derived two CRBs on the variance of unbiased estimators. Monte-Carlo simulations were used to compare the MSE performance of the two estimators, showing that the MSEV estimator performs better than the DML estimator and that the steepest descent algorithm can be used to provide accurate low-complexity implementations for both estimators.



Fig. 4.1 Average MSE of the DML estimator and the MSEV estimator and the bounds CRB_a and $MCRB_a$ versus SNR for N = 100, and M = 4.



Fig. 4.2 Average number of steepest descent iterations (a), and line search iterations for a single steepest descent iteration (b) for different SNR values (N = 100, M = 4).



Fig. 4.3 Average MSE of the DML estimator and the MSEV estimator, and the bounds CRB_a and $MCRB_a$ versus N for an SNR of 15 dB, and M = 4.

Chapter 5

Exact CRBs for Semi-blind Channel Estimation under Square QAM

5.1 Introduction

In Chapters 3 and 4, we focused on blind channel estimation for AF TWRNs that employ M-PSK modulation under flat-fading channel conditions. We showed in Chapter 3 that the proposed blind DML estimator provides a better tradeoff between accuracy and spectral efficiency than the training-based LS estimator. However, as we pointed out in Chapter 3, it is not possible to obtain all the required channel information blindly as pilots are still needed to resolve the phase ambiguity. Still within the context of flat-fading channel conditions, in this chapter we shift our focus from blind channel estimation to semi-blind channel estimation. Semi-blind estimation [41,46–48] is a hybrid of the blind and training-based approaches. Although it employs pilot symbols, it incorporates into the estimation the received data samples as well as the received training samples. This makes the semi-blind approach more flexible than the purely blind approach and eliminates the need for separate ambiguity resolution. By utilizing the data samples in conjunction with the pilots, semi-blind estimation [48]. It is also capable of achieving higher accuracy than that possible using purely blind or purely training-based estimation [48].

In order to evaluate the potential of semi-blind channel estimation for AF TWRNs, it is very useful to know the corresponding semi-blind CRB on achievable estimation accuracy. For AF TWRNs employing pilot-based channel estimation in flat-fading conditions, the CRB has been derived in [21]. In contrast to pilot-based estimation, however, the received signal in the cases of blind and semi-blind estimation involves random nuisance parameters in the form of unknown data symbols. As discussed in Chapter 2, the derivation of the exact CRB is a challenging task in this case because of the complicated nature of the true likelihood function which takes into account the statistics of the transmitted data symbols. In Chapters 3 and 4, we avoided dealing with a complicated likelihood function by treating the data symbols as deterministic unknowns. However, the resulting CRBs apply to the smaller class of estimators that treat the data symbols as deterministic unknowns, such as the those derived in Chapters 3 and 4. The MCRB is another alternative bound which takes into account the statistics of the atta symbols, but it is known to be less tight than the exact CRB.

To the best of our knowledge, the exact CRB based on the true likelihood function that incorporates the statistics of the transmitted data symbols has not been reported before for AF TWRNs. In this chapter, we fill this gap and derive the exact CRB for semi-blind channel estimation¹ in AF TWRNs. The derivation of the exact CRB necessarily depends on the underlying modulation scheme. We focus on square QAM, a commonly used modulation scheme in high data-rate applications due to its high bandwidth efficiency [86]. Using the derived bounds, we quantify the spectral efficiency advantages of the semi-blind approach over the pilot-based approach. We show that, by employing even a limited number of data samples, the semi-blind approach can provide substantially higher accuracy. Equivalently, the semi-blind approach makes it possible to employ a much smaller number of pilots to achieve a given level of estimation accuracy, thus providing a better tradeoff between accuracy and spectral efficiency. The derived CRB also quantifies the effect of the modulation order on the achievable accuracy and shows that the lower modulation order, the higher the achievable accuracy. Finally, we also derive the more tractable MCRB. Our simulations show that the MCRB is significantly looser than the exact CRB for high modulation orders, but is a reasonable approximation for the exact CRB at high SNR for low modulation orders.

The rest of the chapter is organized as follows. In Section 5.2 we present the system model. Our derivations of the exact CRB and the modified CRB are presented in Section 5.3. Simulation results are presented in 5.4. Finally, our conclusions are discussed in Section 5.5.

¹The derived bounds also cover the blind and the pilot-based approaches as special cases.

5.2 System Model

We consider the half-duplex AF TWRN shown in Fig. 1.2, operating in flat-fading nonreciprocal channel conditions. As in the previous chapters, each transmission period is divided into two phases. In the first phase, \mathcal{T}_1 and \mathcal{T}_2 simultaneously transmit to \mathcal{R} , and in the second phase \mathcal{R} broadcasts an amplified version of the received signal to both terminals.

The semi-blind approach employs both pilot and data samples to estimate the channel parameters. More specifically, prior to transmitting data symbols, each terminal transmits a block of L pilot symbols. We denote by $\mathbf{t}_1 \triangleq [t_{11}, \ldots, t_{1L}]^T$ and $\mathbf{t}_2 \triangleq [t_{21}, \ldots, t_{2L}]^T$ the vectors containing the pilot symbols transmitted by \mathcal{T}_1 and \mathcal{T}_2 , respectively. The corresponding received signal vector at \mathcal{R} is $\bar{\mathbf{r}} = h_1 \mathbf{t}_1 + g_1 \mathbf{t}_2 + \boldsymbol{\omega}$, where $\boldsymbol{\omega}$ is $\mathcal{CCN}(0, \sigma^2 \mathbf{I}_L)$). Then, the relay broadcasts $A\bar{\mathbf{r}}$, where A > 0 is the amplification factor. The corresponding received signal vector at terminal \mathcal{T}_1 is

$$\bar{\boldsymbol{z}} = Ah_2h_1\boldsymbol{t}_1 + Ah_2g_1\boldsymbol{t}_2 + Ah_2\boldsymbol{\omega} + \boldsymbol{\omega}_1$$
(5.1)

where $\boldsymbol{\omega}_1$ is also $\mathcal{CCN}(0, \sigma^2 \boldsymbol{I}_L)$.

After transmitting the L pilots, \mathcal{T}_1 and \mathcal{T}_2 transmit N data symbols each. We denote by $\mathbf{s}_1 \triangleq [s_{11}, \ldots, s_{1N}]^T$ and $\mathbf{s}_2 \triangleq [s_{21}, \ldots, s_{2N}]^T$ the transmitted data symbol vectors of \mathcal{T}_1 and \mathcal{T}_2 , respectively. The received signal vector at \mathcal{R} is $\mathbf{r} = h_1 \mathbf{s}_1 + g_1 \mathbf{s}_2 + \mathbf{n}$, and the corresponding received signal vector at \mathcal{T}_1 is

$$\boldsymbol{z} = Ah_1h_2\boldsymbol{s}_1 + Ag_1h_2\boldsymbol{s}_2 + Ah_2\boldsymbol{n} + \boldsymbol{n}_1$$
(5.2)

where \boldsymbol{n} and \boldsymbol{n}_1 are $\mathcal{CCN}(0, \sigma^2 \boldsymbol{I}_N)$. The channel coefficients h_1, h_2, g_1 and g_2 remain fixed during the transmission of the L pilot symbols and the N data symbols.

We assume that both terminals employ square QAM. Without loss of generality, we focus on the derivation of the CRB for channel estimation at terminal \mathcal{T}_1 . The total number of points in the square QAM constellation employed by \mathcal{T}_2 is $M = 2^{2p}$, where $p = 1, 2, 3, \ldots$ Denoting by d the intersymbol distance and letting $d_p \triangleq \frac{d}{2}$, the set of constellation points used by \mathcal{T}_2 is given by $S = \{\pm d_p(2i-1) \pm jd_p(2\ell-1)\}, i, \ell = 1, \ldots, 2^{p-1}$ [87]. The average transmitted power at \mathcal{T}_2 is $P_2 = \mathbb{E}\{|s_{2k}|^2\} = \frac{M-1}{6}d^2$. The noise variance σ^2 is assumed to be known at \mathcal{T}_1 . We are interested in deriving the CRBs for the estimation of the cascaded channel parameters $a \triangleq h_2h_1$ and $b \triangleq h_2g_1$ which are sufficient for detection.

5.3 Cramer-Rao Bounds

In this section, we derive the exact CRBs for the estimation of a and b. In deriving these bounds, we also have to take into account the parameter $\tau \triangleq |h_2|^2$ which, though not required for detection, appears in the likelihood function and thus affects the estimation performance. The unknown parameters are thus a, b and τ and are collected into the real vector $\boldsymbol{\theta} \triangleq [\Re\{a\}, \Im\{a\}, \Re\{b\}, \Im\{b\}, \tau]^T$. To simplify our notation, we let $a_R \triangleq \Re\{a\}$, $a_I \triangleq \Im\{a\}, b_R \triangleq \Re\{b\}, b_I \triangleq \Im\{b\}$.

5.3.1 Exact Cramer-Rao Bound

To derive the exact CRB for semi-blind channel estimation, we need to consider the joint likelihood of \bar{z} and z. Let $\tilde{z} \triangleq [\bar{z}^T, z^T]^T$, the likelihood of \tilde{z} is given by

$$f(\tilde{\boldsymbol{z}};\boldsymbol{\theta}) = \frac{1}{(\pi\sigma^2(A^2\tau+1))^{N+L}} e^{-\frac{\|\tilde{\boldsymbol{z}}-Aat_1-Abt_2\|^2}{\sigma^2(A^2\tau+1)}} \prod_{k=1}^N \frac{1}{M} \left(\sum_{s_2 \in S} e^{-\frac{|z_k-Aas_{1k}-Abs_2|^2}{\sigma^2(A^2\tau+1)}}\right).$$
(5.3)

Let $C \triangleq \sigma^2 (A^2 \tau + 1)$. The resulting log-likelihood function is

$$\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta}) = -(N+L)\log(\pi C) - \frac{1}{C} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_1 - Ab\boldsymbol{t}_2\|^2 - N\log M + \sum_{k=1}^N \log\left(\sum_{s_2 \in S} e^{-\frac{1}{C}|\boldsymbol{z}_k - Aas_{1k} - Abs_2|^2}\right)$$
(5.4)

Let $I(\theta)$ be the corresponding FIM, and let $I_{x,y}$ be the joint Fisher information between the parameters x and y. The matrix $I(\theta)$ is given by

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\mathbb{E}\left\{\frac{\partial^{2}\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{T}}\right\} = \begin{bmatrix}\boldsymbol{I}_{aa} & \boldsymbol{I}_{ab} & \boldsymbol{I}_{a\tau}\\ \boldsymbol{I}_{ab}^{T} & \boldsymbol{I}_{bb} & \boldsymbol{I}_{b\tau}\\ \boldsymbol{I}_{a\tau}^{T} & \boldsymbol{I}_{b\tau}^{T} & \boldsymbol{I}_{\tau,\tau}\end{bmatrix},$$
(5.5)

where

$$\boldsymbol{I}_{aa} = \begin{bmatrix} I_{a_R,a_R} & I_{a_R,a_I} \\ I_{a_R,a_I} & I_{a_I,a_I} \end{bmatrix}, \quad \boldsymbol{I}_{ab} = \begin{bmatrix} I_{a_R,b_R} & I_{a_R,b_I} \\ I_{a_I,b_R} & I_{a_I,b_I} \end{bmatrix}, \quad (5.6)$$

$$\boldsymbol{I}_{bb} = \begin{bmatrix} I_{b_R, b_R} & I_{b_R, b_I} \\ I_{b_R, b_I} & I_{b_I, b_I} \end{bmatrix}, \quad \boldsymbol{I}_{a\tau} = \begin{bmatrix} I_{a_R, \tau} \\ I_{a_I, \tau} \end{bmatrix}, \quad \boldsymbol{I}_{b\tau} = \begin{bmatrix} I_{b_R, \tau} \\ I_{b_I, \tau} \end{bmatrix}.$$
(5.7)

The exact CRB is then obtained by taking the inverse of $I(\theta)$. In particular, the exact CRBs for parameters a and b are

$$CRB_a = [\boldsymbol{I}(\boldsymbol{\theta})^{-1}]_{11} + [\boldsymbol{I}(\boldsymbol{\theta})^{-1}]_{22}, \qquad (5.8)$$

and

$$CRB_b = [\boldsymbol{I}(\boldsymbol{\theta})^{-1}]_{33} + [\boldsymbol{I}(\boldsymbol{\theta})^{-1}]_{44}.$$
(5.9)

In what follows, we will obtain analytical expressions for all the elements of $I(\theta)$. In summary, the analytical expressions for I_{aa} , I_{bb} , I_{ab} , $I_{a\tau}$, $I_{b\tau}$ and $I_{\tau\tau}$ are given in Eqns. (5.38), (5.47), (5.58), (5.59), (5.60) and (5.57), respectively.

Before proceeding to obtain analytical expressions for the elements of $I(\theta)$, we will first factorize the likelihood function by taking into account the symmetric structure of square QAM, following the approach proposed in [87]. We begin by rewriting the likelihood function in (5.3) as

$$f(\tilde{\boldsymbol{z}};\boldsymbol{\theta}) = \frac{1}{(\pi C)^{N+L} M^N} e^{-\frac{1}{C} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_1 - Ab\boldsymbol{t}_2\|^2} \\ \times \prod_{k=1}^N \left(e^{-\frac{1}{C} |\boldsymbol{z}_k - Aas_{1k}|^2} \sum_{s_2 \in S} e^{-\frac{A^2}{C} |\boldsymbol{b}|^2 |s_2|^2 + \frac{2A}{C} \Re\{(\boldsymbol{z}_k - Aas_{1k})^* bs_2\}} \right).$$
(5.10)

Now, we let

$$D_{k}(\boldsymbol{\theta}) \triangleq \sum_{s_{2} \in S} e^{-\frac{1}{C}A^{2}|b|^{2}|s_{2}|^{2} + \frac{2A}{C}\Re\{(z_{k} - Aas_{1k})^{*}bs_{2}\}}.$$
(5.11)

The inherent symmetry of the constellation set allows us to write (5.11) as a sum over the symbols in the first quadrant. If Q_1 is the set of constellation symbols that lie in the first quadrant, S can be partitioned as² $S = Q_1 \cup (-Q_1) \cup Q_1^* \cup (-Q_1^*)$. Hence, we may rewrite $D_k(\boldsymbol{\theta})$ as:

$$D_{k}(\boldsymbol{\theta}) = \sum_{s_{2} \in Q_{1}} e^{-\frac{1}{C}A^{2}|b|^{2}|s_{2}|^{2}} \left(e^{\frac{2A}{C}\Re\{(z_{k} - Aas_{1k})^{*}bs_{2}\}} + e^{\frac{2A}{C}\Re\{(z_{k} - Aas_{1k})^{*}bs_{2}^{*}\}} + e^{-\frac{2A}{C}\Re\{(z_{k} - Aas_{1k})^{*}bs_{2}^{*}\}} + e^{-\frac{2A}{C}\Re\{(z_{k} - Aas_{1k})^{*}bs_{2}^{*}\}} \right)$$
(5.12)

²For a set $B = \{b_1, b_2, \ldots\}$, the notations B^* and -B denote the sets $\{b_1^*, b_2^*, \ldots\}$ and $\{-b_1, -b_2, \ldots\}$, respectively.

Noting that $\Re\{(z_k - Aas_{1k})^*bs_2\} = \Re\{(z_k - Aas_{1k})^*b\}\Re\{s_2\} - \Im\{(z_k - Aas_{1k})^*b\}\Im\{s_2\}$ and that for $s_2 \in Q_1$, we have $\Re\{s_2\}, \Im\{s_2\} \in \{(2i-1)d_p, i = 1, \dots, 2^{p-1}\}, (5.12)$ becomes

$$D_{k}(\boldsymbol{\theta}) = 4 \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} e^{-\frac{1}{C}A^{2}|b|^{2}((2i-1)^{2}+(2\ell-1)^{2})d_{p}^{2}} \times \cosh\left[\frac{2A}{C}(2i-1)d_{p}\Re\{(z_{k}-Aas_{1k})^{*}b\}\right] \cosh\left[\frac{2A}{C}(2\ell-1)d_{p}\Im\{(z_{k}-Aas_{1k})^{*}b\}\right].$$
(5.13)

From (5.13), we can see that, similar to [87], $D_k(\boldsymbol{\theta})$ can be expressed as

$$D_k(\boldsymbol{\theta}) = 4F_{\theta}(u_k)F_{\theta}(v_k), \qquad (5.14)$$

where

$$F_{\theta}(t) \triangleq \sum_{i=1}^{2^{p-1}} e^{-\frac{1}{C}A^2 d_p^2 |b|^2 (2i-1)^2} \cosh\left(\frac{2Ad_p}{C}(2i-1)t\right),$$
(5.15)

$$u_k \triangleq \Re\{(z_k - Aas_{1k})^*b\} = \Re\{z_k - Aas_{1k}\}b_R + \Im\{z_k - Aas_{1k}\}b_I$$
(5.16)

and

$$v_k \triangleq \Im\{(z_k - Aas_{1k})^*b\} = \Re\{z_k - Aas_{1k}\}b_I - \Im\{z_k - Aas_{1k}\}b_R.$$
 (5.17)

For convenience, we further simplify our notation by letting $\beta_i \triangleq \frac{Ad_p}{C}(2i-1)$ and $\gamma_i \triangleq \frac{A^2d_p^2}{C}(2i-1)^2$ for $i = 1, \ldots, 2^{p-1}$. Using the newly defined β_i, γ_i , we can write $F_{\theta}(t)$ as

$$F_{\theta}(t) \triangleq \sum_{i=1}^{2^{p-1}} e^{-\gamma_i |b|^2} \cosh(2\beta_i t) \,.$$
 (5.18)

Using (5.14), the likelihood function becomes

$$f(\tilde{\boldsymbol{z}};\boldsymbol{\theta}) = \frac{1}{(\pi C)^{N+L}} e^{-\frac{1}{C} \|\tilde{\boldsymbol{z}} - Aa\boldsymbol{t}_1 - Ab\boldsymbol{t}_2\|^2} \prod_{k=1}^N \frac{1}{M} \left(e^{-\frac{1}{C} |\boldsymbol{z}_k - Aas_{1k}|^2} 4F_{\boldsymbol{\theta}}(\boldsymbol{u}_k) F_{\boldsymbol{\theta}}(\boldsymbol{v}_k) \right).$$
(5.19)

As we shall see shortly, the RVs u_k and v_k are independent, a fact that will simplify our derivation of the FIM. It is also useful to define two new RVs, $x_k \triangleq \Re\{z_k - Aas_{1k}\}$ and

 $y_k \triangleq \Im\{z_k - Aas_{1k}\}$. The pair $\{u_k, v_k\}$ and the pair $\{x_k, y_k\}$ are related through the following linear transformation

$$\begin{bmatrix} u_k \\ v_k \end{bmatrix} = \begin{bmatrix} b_R & b_I \\ b_I & -b_R \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix}.$$
 (5.20)

Both pairs of RVs will be used in deriving the elements of $I(\theta)$. It is easy to see that the joint PDF of x_k and y_k is

$$f_{X,Y}(x,y) = \frac{4}{\pi MC} e^{-\frac{x^2 + y^2}{C}} F_{\theta}(b_R x + b_I y) F_{\theta}(b_I x - b_R y).$$
(5.21)

Using (5.20) and (5.21), we obtain the joint PDF of u_k and v_k :

$$f_{U,V}(u,v) = \frac{4}{\pi MC|b|^2} e^{-\frac{u^2+v^2}{C|b|^2}} F_{\theta}(u) F_{\theta}(v).$$
(5.22)

It is clear from (5.22) that the RVs u_k and v_k are i.i.d. with respective PDFs

$$f_U(x) = f_V(x) = \frac{2}{\sqrt{M\pi C|b|^2}} e^{-\frac{x^2}{C|b|^2}} F_\theta(x).$$
(5.23)

Going back to the log-likelihood function in (5.4), we may now rewrite it as

$$\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta}) = -(N+L)\log(\pi C) + N\log\frac{4}{M} - \frac{1}{C}\|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_1 - Ab\boldsymbol{t}_2\|^2 - \frac{1}{C}\|\boldsymbol{z} - Aa\boldsymbol{s}_1\|^2 + \sum_{k=1}^N \log F_{\boldsymbol{\theta}}(u_k) + \sum_{k=1}^N \log F_{\boldsymbol{\theta}}(v_k).$$
(5.24)

From (5.24) we see that the main task in obtaining analytical expressions for the elements of $I(\theta)$ is the evaluation of the expectations $\mathbb{E}\left\{\frac{\partial^2 \log F_{\theta}(u_k)}{\partial \theta_i \partial \theta_j}\right\}$ and $\mathbb{E}\left\{\frac{\partial^2 \log F_{\theta}(v_k)}{\partial \theta_i \partial \theta_j}\right\}$ for $i, j = 1, \ldots, 5$. Letting

$$B_{k}^{(ij)} \triangleq \frac{\frac{\partial^{2} F_{\theta}(u_{k})}{\partial \theta_{i} \partial \theta_{j}}}{F_{\theta}(u_{k})}, \qquad G_{k}^{(ij)} \triangleq \frac{\frac{\partial F_{\theta}(u_{k})}{\partial \theta_{i}} \frac{\partial F_{\theta}(u_{k})}{\partial \theta_{j}}}{F_{\theta}(u_{k})^{2}}, \tag{5.25}$$

and

$$H_{k}^{(ij)} \triangleq \frac{\frac{\partial^{2} F_{\theta}(v_{k})}{\partial \theta_{i} \partial \theta_{j}}}{F_{\theta}(v_{k})}, \qquad W_{k}^{(ij)} \triangleq \frac{\frac{\partial F_{\theta}(v_{k})}{\partial \theta_{i}} \frac{\partial F_{\theta}(v_{k})}{\partial \theta_{j}}}{F_{\theta}(v_{k})^{2}}, \tag{5.26}$$

we have that

$$\frac{\partial^2 \log F_{\theta}(u_k)}{\partial \theta_i \partial \theta_j} = B_k^{(ij)} - G_k^{(ij)}, \qquad (5.27)$$

and

$$\frac{\partial^2 \log F_\theta(v_k)}{\partial \theta_i \partial \theta_j} = H_k^{(ij)} - W_k^{(ij)}.$$
(5.28)

Despite the factorization of the log-likelihood function, the derivation of analytical expressions for the elements of $I(\theta)$ requires tedious calculations. Due to space limitations, we will provide detailed derivations for only some of these elements. For the remaining elements we will provide only the resulting analytical expressions.

1) Derivation of $I_{a,a}$:

We begin with the submatrix $I_{a,a}$ of $I(\theta)$ and start by finding I_{a_R,a_R} . We have

$$\mathbb{E}\left\{\frac{\partial^{2}\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial a_{R}^{2}}\right\} = -\frac{2A^{2}}{C}\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1} - \frac{2A^{2}}{C}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1} + \sum_{k=1}^{N}\mathbb{E}\left\{B_{k}^{(11)} - G_{k}^{(11)}\right\} + \sum_{k=1}^{N}\mathbb{E}\left\{H_{k}^{(11)} - W_{k}^{(11)}\right\}.$$
(5.29)

We show in Appendix C.1 that

$$\mathbb{E}\left\{B_{k}^{(11)}\right\} = \frac{8A^{2}}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} \beta_{i}^{2} \Re\{s_{1k}^{*}b\}^{2},$$
(5.30)

and

$$\mathbb{E}\left\{H_k^{(11)}\right\} = \frac{8A^2}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} \beta_i^2 \Im\{s_{1k}^*b\}^2.$$
(5.31)

To obtain $\mathbb{E}\left\{G_k^{(11)}\right\}$, we need the first derivative of $F_\theta(u_k)$ with respect to a_R , which is given by

$$\frac{\partial F_{\theta}(u_k)}{\partial a_R} = -2A\Re\{s_{1k}^*b\}\sum_{i=1}^{2^{p-1}}\beta_i e^{-\gamma_i|b|^2}\sinh[2\beta_i u_k].$$
(5.32)

From (5.32), we see that we can evaluate $\mathbb{E}\left\{G_k^{(11)}\right\}$ using the PDF in (5.23). We thus

obtain

$$\mathbb{E}\left\{G_k^{(11)}\right\} = \Re\{s_{1k}^*b\}^2\Upsilon,\tag{5.33}$$

where

$$\Upsilon = \frac{8A^2}{\sqrt{\pi MC|b|^2}} \int_{-\infty}^{\infty} \frac{\mathcal{P}^2(t)}{F_{\theta}(t)} e^{-\frac{t^2}{C|b|^2}} dt,$$
(5.34)

and

$$\mathcal{P}(t) = \sum_{i=1}^{2^{p-1}} \beta_i e^{-\gamma_i |b|^2} \sinh[2\beta_i t].$$
(5.35)

Moreover, it can be easily verified that $\mathbb{E}\left\{W_k^{(11)}\right\} = \Im\{s_{1k}^*b\}^2\Upsilon$. Hence,

$$I_{a_R,a_R} = \frac{2A^2}{C} \boldsymbol{t}_1^H \boldsymbol{t}_1 + \frac{2A^2}{C} \boldsymbol{s}_1^H \boldsymbol{s}_1 - |b|^2 \boldsymbol{s}_1^H \boldsymbol{s}_1 \left(\frac{8A^2}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} \beta_i^2 - \Upsilon\right).$$
(5.36)

Furthermore, it can be easily shown that $I_{a_I,a_I} = I_{a_R,a_R}$. Finally, we prove in Appendix C.2 that

$$I_{a_R,a_I} = 0. (5.37)$$

The resulting analytical expression for $\boldsymbol{I}_{a,a}$ is

$$\boldsymbol{I}_{aa} = \left(\frac{2A^2}{C} \left(\boldsymbol{t}_1^H \boldsymbol{t}_1 + \boldsymbol{s}_1^H \boldsymbol{s}_1\right) - |b|^2 \boldsymbol{s}_1^H \boldsymbol{s}_1 \left(\frac{8A^2}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} \beta_i^2 - \Upsilon\right)\right) \boldsymbol{I}_2, \quad (5.38)$$

where I_2 is the 2 × 2 identity matrix.

2) Derivation of $I_{b,b}$:

We next consider the submatrix $I_{b,b}$. Before deriving the elements of $I_{b,b}$, we let

$$q_{1}(x_{k}, y_{k}) \triangleq \frac{\partial F_{\theta}(b_{R}x_{k} + b_{I}y_{k})}{\partial b_{R}}$$

= $-\sum_{i=1}^{2^{p-1}} 2\gamma_{i}b_{R}e^{-\gamma_{i}|b|^{2}}\cosh[2\beta_{i}(b_{R}x_{k} + b_{I}y_{k})] + \sum_{i=1}^{2^{p-1}} 2\beta_{i}x_{k}e^{-\gamma_{i}|b|^{2}}\sinh[2\beta_{i}(b_{R}x_{k} + b_{I}y_{k})].$
(5.39)

and

$$q_{2}(x_{k}, y_{k}) \triangleq \frac{\partial F_{\theta}(u_{k})}{\partial b_{I}}$$

= $-\sum_{i=1}^{2^{p-1}} 2\gamma_{i}b_{I}e^{-\gamma_{i}|b|^{2}} \cosh[2\beta_{i}(b_{R}x_{k}+b_{I}y_{k})] + \sum_{i=1}^{2^{p-1}} 2\beta_{i}y_{k}e^{-\gamma_{i}|b|^{2}} \sinh[2\beta_{i}(b_{R}x_{k}+b_{I}y_{k})].$
(5.40)

and we also define for i,j=1,2

$$\Gamma_{ij} \triangleq \mathbb{E}\left\{\frac{q_i(x_k, y_k)q_j(x_k, y_k)}{F_{\theta}(u_k)}\right\} = \frac{4}{\pi MC} \iint_{-\infty}^{\infty} \frac{q_i(x, y)q_j(x, y)F_{\theta}(b_I x - b_R y)}{F_{\theta}(b_R x + b_I y)} e^{-\frac{x^2 + y^2}{C}} dx dy.$$
(5.41)

We now consider I_{b_R,b_R} . We have

$$\mathbb{E}\left\{\frac{\partial^{2}\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial b_{R}^{2}}\right\} = -\frac{2A^{2}}{C}\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{2} + \sum_{k=1}^{N}\mathbb{E}\left\{B_{k}^{(33)} - G_{k}^{(33)}\right\} + \sum_{k=1}^{N}\mathbb{E}\left\{H_{k}^{(33)} - W_{k}^{(33)}\right\}.$$
 (5.42)

We show in Appendix C.3 that

$$\mathbb{E}\left\{B_{k}^{(33)}\right\} = \mathbb{E}\left\{H_{k}^{(33)}\right\} = \frac{16}{M}\sum_{i=1}^{2^{p-1}}\sum_{\ell=1}^{2^{p-1}}\gamma_{i}\gamma_{\ell}b_{I}^{2}.$$
(5.43)

Moreover, we have that $\mathbb{E}\left\{G_k^{(33)}\right\} = \mathbb{E}\left\{W_k^{(33)}\right\} = \Gamma_{11}$. Hence,

$$I_{b_R,b_R} = \frac{2A^2}{C} \boldsymbol{t}_2^H \boldsymbol{t}_2 - \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_I^2 + 2N\Gamma_{11}.$$
(5.44)

A very similar approach can be followed to evaluate I_{b_I,b_I} , thus obtaining

$$I_{b_I,b_I} = \frac{2A^2}{C} \boldsymbol{t}_2^H \boldsymbol{t}_2 - \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_R^2 + 2N\Gamma_{22}.$$
 (5.45)

As for the element I_{b_R,b_I} , it can be shown that (we skip the derivation for brevity)

$$I_{b_R,b_I} = \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_R b_I + 2N\Gamma_{12}.$$
 (5.46)

The resulting analytical expression for $\boldsymbol{I}_{b,b}$ is given by

$$\boldsymbol{I}_{bb} = \begin{bmatrix} \frac{2A^2}{C} \boldsymbol{t}_2^H \boldsymbol{t}_2 - \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_I^2 + 2N\Gamma_{11} & \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_R b_I + 2N\Gamma_{12} \\ \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_R b_I + 2N\Gamma_{12} & \frac{2A^2}{C} \boldsymbol{t}_2^H \boldsymbol{t}_2 - \frac{32N}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell b_R^2 + 2N\Gamma_{22}. \end{bmatrix}$$
(5.47)

3) Derivation of $I_{\tau,\tau}$:

For the fifth diagonal element of $\boldsymbol{I}(\boldsymbol{\theta}), I_{\tau,\tau}$, we have

$$\mathbb{E}\left\{\frac{\partial^{2}\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial\tau^{2}}\right\} = \frac{(N+L)A^{4}\sigma^{4}}{C^{2}} - \frac{2A^{4}\sigma^{4}}{C^{3}}\left(\mathbb{E}\left\{\|\boldsymbol{z}_{p} - Aa\boldsymbol{t}_{1} - Ab\boldsymbol{t}_{2}\|^{2}\right\} + \mathbb{E}\left\{\|\boldsymbol{z} - Aa\boldsymbol{s}_{1}\|^{2}\right\}\right) + \sum_{k=1}^{N}\mathbb{E}\left\{B_{k}^{(55)} - G_{k}^{(55)}\right\} + \sum_{k=1}^{N}\mathbb{E}\left\{H_{k}^{(55)} - W_{k}^{(55)}\right\}.$$
(5.53)

It can be easily shown that $\mathbb{E}\{\|\boldsymbol{z}_p - Aa\boldsymbol{t}_1 - Ab\boldsymbol{t}_2\|^2\} = LC$ and $\mathbb{E}\{\|\boldsymbol{z} - Aa\boldsymbol{s}_1\|^2\} = NA^2|b|^2P_2 + NC$. Moreover, we show in Appendix C.4 that

$$\mathbb{E}\left\{B_{k}^{(55)}\right\} = \mathbb{E}\left\{H_{k}^{(55)}\right\} = \sum_{i=1}^{2^{p-1}} \frac{2A^{4}\sigma^{4}}{\sqrt{M}} \left(\beta_{i}^{4}|b|^{4} + \frac{4}{C}\beta_{i}^{2}|b|^{2}\right).$$
(5.54)

To obtain $\mathbb{E}\left\{G_k^{(55)}\right\}$, we first take the derivative $F_\theta(u_k)$ with respect to τ :

$$g(u_k) \triangleq \frac{\partial F_{\theta}(u_k)}{\partial \tau} = \sum_{i=1}^{2^{p-1}} A^2 \sigma^2 \beta_i^2 |b|^2 e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k] - \sum_{i=1}^{2^{p-1}} \frac{2A^2 \sigma^2}{C} \beta_i u_k e^{-\gamma_i |b|^2} \sinh[2\beta_i u_k].$$
(5.55)

Letting $\Lambda \triangleq \mathbb{E}\left\{G_k^{(55)}\right\}$, we thus get

$$\Lambda = \frac{2}{\sqrt{\pi MC|b|^2}} \int_{-\infty}^{\infty} \frac{g(t)^2}{F_{\theta}(t)} e^{-\frac{t^2}{C|b|^2}} dt.$$
(5.56)

Clearly, we also have that $\mathbb{E}\left\{W_k^{55}\right\} = \Lambda$. Using the above results, we get

$$I_{\tau,\tau} = (N+L)\frac{A^4\sigma^4}{C^2} + \frac{2NA^6\sigma^4}{C^3}|b|^2P_2 - 4N\sum_{i=1}^{2^{p-1}}\frac{A^4\sigma^4}{\sqrt{M}}\left(\beta_i^4|b|^4 + \frac{4}{C}\beta_i^2|b|^2\right) + 2N\Lambda.$$
(5.57)

4) Remaining Submatrices of $I(\theta)$:

The remaining submatrices of $I(\theta)$ can be obtained by following similar approaches to those used so far. More specifically, it can be shown that

$$\boldsymbol{I}_{ab} = \begin{bmatrix} \frac{2A^2}{C} \Re\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & -\frac{2A^2}{C} \Im\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} \\ \frac{2A^2}{C} \Im\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & \frac{2A^2}{C} \Re\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} \end{bmatrix},$$
(5.58)

$$\boldsymbol{I}_{a\tau} = \begin{bmatrix} 0\\0 \end{bmatrix} \tag{5.59}$$

and

$$\boldsymbol{I}_{b\tau} = \begin{bmatrix} 8N \sum_{i=1}^{2^{p-1}} \frac{A^2 \sigma^2}{\sqrt{M}} \beta_i^2 b_R + 2N\Delta_1 \\ 8N \sum_{i=1}^{2^{p-1}} \frac{A^2 \sigma^2}{\sqrt{M}} \beta_i^2 b_I + 2N\Delta_2 \end{bmatrix}$$
(5.60)

where Δ_i , i = 1, 2 is given by

$$\Delta_{i} = \frac{4}{\pi MC} \iint_{-\infty}^{\infty} \frac{g(b_{R}x + b_{I}y)q_{i}(x,y)}{F_{\theta}(b_{R}x + b_{I}y)} F_{\theta}(b_{I}x - b_{R}y)e^{-\frac{x^{2}+y^{2}}{C}}dxdy.$$
(5.61)

Having derived the FIM matrix $I(\theta)$, the exact CRBs on a and b are given as

$$CRB_a = [\mathbf{I}(\boldsymbol{\theta})^{-1}]_{11} + [\mathbf{I}(\boldsymbol{\theta})^{-1}]_{22},$$
 (5.62)

and

$$CRB_b = [I(\theta)^{-1}]_{33} + [I(\theta)^{-1}]_{44}.$$
 (5.63)

When the pilot vectors t_1 and t_2 are orthogonal, CRB_a has the following closed-form expression

$$CRB_{a} = 2\left(\frac{2A^{2}}{C}\left(\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1} + \boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}\right) - |b|^{2}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}\left(\frac{8A^{2}}{\sqrt{M}}\sum_{i=1}^{2^{p-1}}\beta_{i}^{2} - \Upsilon\right)\right)^{-1}.$$
(5.64)

The term Υ in (5.64) is a single integral over the interval $(-\infty, \infty)$ (see (5.34)). This integral can be accurately approximated in a numerically efficient way using Gauss-Hermite polynomials [80]. Moreover, the integrand decays rapidly as |t| increases. Hence, the integral can be accurately approximated by a finite integral over an interval [-T, T] for finite T, where Riemann integration methods can be employed. For modulation orders M = 4 and M = 16, it is also possible to obtain a closed-form approximation for Υ at high SNR. In particular, we show in Appendix C.5 that, at high SNR, we may approximate Υ for M = 4 and M = 16 as

$$\Upsilon \approx \frac{2A^4 d_p^2}{C^2} \left(1 + \operatorname{erf}\left(\sqrt{\gamma_1 |b|^2}\right) \right)$$
(5.65)

and

$$\Upsilon \approx \frac{A^4 d_p^2}{C^2} \left(10 + \operatorname{erf}\left(\sqrt{\gamma_1 |b|^2}\right) + 9 \operatorname{erf}\left(\sqrt{\gamma_2 |b|^2}\right) \right),$$
(5.66)

respectively.

Substituting (5.65) into (5.64) and noting that $\sum_{i=1}^{2^{p-1}} \beta_i^2 = \frac{A^2 d_p^2}{6C^2} (M-1)\sqrt{M}$, we obtain for M = 4

$$CRB_{a} \approx 2\left(\frac{2A^{2}}{C}\left(\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1} + \boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}\right) + \frac{2A^{4}d_{p}^{2}}{C^{2}}|b|^{2}\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}\left(\mathrm{erf}\left(\sqrt{\gamma_{1}|b|^{2}}\right) - 1\right)\right)^{-1}.$$
(5.67)

Since $\boldsymbol{s}_1^H \boldsymbol{s}_1 = NP_1$ for M = 4, we may rewrite (5.67) as

$$CRB_a \approx \left(\frac{A^2}{C} \left[\boldsymbol{t}_1^H \boldsymbol{t}_1 + NP_1 \left(1 + \gamma_1 |b|^2 \operatorname{erf}\left(\sqrt{\gamma_1 |b|^2}\right) - \gamma_1 |b|^2 \right) \right] \right)^{-1}.$$
 (5.68)

Eq. (5.68) relates CRB_a to the overall training power $(\boldsymbol{t}_1^H \boldsymbol{t}_1)$, the overall self-interference

power NP_1 and the effective energy of the received data symbols which is a scaled version of $\gamma_1|b|^2$. Since $1 + x \operatorname{erf}(x) - x > 0$ for x > 0, we can see from (5.68) that CRB_a decreases as the number of data samples increases, or in other words, that the estimation accuracy improves as the number of data samples increases. Assuming all other parameters are fixed, CRB_a decreases with respect to the number of data samples as $\frac{1}{\alpha N+\delta}$ for $\alpha, \delta > 0$. This is verified in our simulation results in Section 5.4.

For M = 16, we substitute (5.66) into (5.64), obtaining

$$CRB_{a} \approx \left(\frac{A^{2}}{C} \left[\boldsymbol{t}_{1}^{H} \boldsymbol{t}_{1} + \boldsymbol{s}_{1}^{H} \boldsymbol{s}_{1} \left(1 + \gamma_{1} |b|^{2} \operatorname{erf} \left(\sqrt{\gamma_{1} |b|^{2}} \right) + \gamma_{2} |b|^{2} \operatorname{erf} \left(\sqrt{\gamma_{2} |b|^{2}} \right) - \gamma_{1} |b|^{2} - \gamma_{2} |b|^{2} \right) \right] \right)^{-1}$$
(5.69)

Inspecting (5.69) and (5.68), we see that, compared to the denominator for M = 4, the denominator for M = 16 has the extra term $\gamma_2 |b|^2 \operatorname{erf}(\sqrt{\gamma_2 |b|^2}) - \gamma_2 |b|^2$. As this term is always negative, we see that CRB_a increases as the modulation order M increases from 4 to 16. This trend is supported by our simulation results in Section 5.4 which show that both CRB_a and CRB_b increase with M. Noting that for large $N \ \mathbf{s}_1^H \mathbf{s}_1 \approx NP_1$, we see that, similar to situation for M = 4, CRB_a decreases with respect to N as $\frac{1}{\alpha N + \delta}$.

As for CRB_b , the closed-form expression is complicated because $I_{b\tau}$ is nonzero and because the elements of I_{bb} and $I_{b\tau}$ involve the double integrals Γ_{11} , Γ_{12} , Γ_{22} , Δ_1 and Δ_2 . These integrals can be evaluated using the two-dimensional Gauss Hermite quadrature [88]. Moreover, the integrands of Γ_{11} , Γ_{12} , Γ_{22} , Δ_1 and Δ_2 decay rapidly as |x| and |y| increase. Hence, numerical methods for integrations over finite intervals such as the Gaussian quadrature method [80] can be used to efficiently compute these integrals.

The bounds in (5.62) and (5.63) provide convenient benchmarks for the performance of estimators of a and b. They yield the bounds for the blind case³ for L = 0 and the bounds for the fully pilot-based case for N = 0. In Section 5.4, we will use these bounds to compare the semi-blind and the pilot-based approaches. We next consider the more tractable MCRB.

5.3.2 Modified Cramer-Rao Bound

Due to its tractability, the MCRB is commonly used in the presence of random nuisance parameters [54]. As we discussed in Chapter 2, the MCRB is obtained using the MFIM

³Although the parameter b suffers from an inherent ambiguity in the absence of pilots and is only locally identifiable in this case, the CRB would still be defined, as pointed out in [89].

matrix (see Chapter 2, Eq. (2.15)). For our case, the MFIM matrix is given by

$$\boldsymbol{J} = \frac{2A^2}{C} \begin{bmatrix} \left(\boldsymbol{t}_1^H \boldsymbol{t}_1 + \boldsymbol{s}_1^H \boldsymbol{s}_1\right) & 0 & \Re\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & -\Im\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & 0 \\ 0 & \left(\boldsymbol{t}_1^H \boldsymbol{t}_1 + \boldsymbol{s}_1^H \boldsymbol{s}_1\right) & \Im\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & \Re\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & 0 \\ \Re\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & \Im\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & \left(\boldsymbol{t}_2^H \boldsymbol{t}_2 + NP_2\right) & 0 & 0 \\ -\Im\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & \Re\{\boldsymbol{t}_1^H \boldsymbol{t}_2\} & 0 & \left(\boldsymbol{t}_2^H \boldsymbol{t}_2 + NP_2\right) & 0 \\ 0 & 0 & 0 & 0 & \left(N + L\right)\frac{A^2\sigma^4}{2C} \end{bmatrix}$$
(5.70)

Denoting by $MCRB_a$ and $MCRB_b$ the resulting bounds for a and b, respectively, we have

$$MCRB_{a} = \frac{C\left(\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{2} + NP_{2}\right)}{A^{2}\left(\left(\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1} + \boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}\right)\left(\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{2} + NP_{2}\right) - \boldsymbol{t}_{1}^{H}\boldsymbol{t}_{2}\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{1}\right)\right)}$$
(5.71)

and

$$MCRB_{b} = \frac{C}{A^{2}(\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{2} + NP_{2})} \left(1 + \frac{\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{2}\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{1}}{\left((\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1} + \boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1})(\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{2} + NP_{2}) - \boldsymbol{t}_{1}^{H}\boldsymbol{t}_{2}\boldsymbol{t}_{2}^{H}\boldsymbol{t}_{1}\right)} \right). \quad (5.72)$$

The above expressions are much simpler and more tractable than those in (5.8) and (5.9). However, as we shall see in Section 5.4, they are significantly less tight than the exact CRBs and do not convey the impact of the modulation order on the achievable estimation accuracy.

5.4 Simulation Results

In this section, we use MATLAB simulations to investigate the behavior of the derived CRBs for a and b and to compare the semi-blind and the pilot-based approaches. The CRBs are evaluated using the Monte-Carlo approach⁴. All plots are averages over a set of 100 independent realizations of the channel parameters h_1 , h_2 , g_1 and g_2 . As we did in Chapter 3, we generate these realizations by modelling h_1 and h_2 as correlated complex Gaussian random variables with mean zero, variance 1, and a correlation coefficient $\rho = 0.3$. Similarly, we model g_1 and g_2 as correlated complex Gaussian random variables with the same mean, variance and correlation coefficient, but independent of h_1 and h_2 . In order

⁴Although more numerically-efficient approaches are possible, find the best numerical implementation for CRB_a and CRB_b is beyond the scope of this thesis.

to investigate the effect of the modulation order on the CRB, we consider four modulation orders for the data symbols, M = 4, M = 16, M = 64 and M = 256. The pilot symbols are always generated using the modulation order⁵ $M_p = 4$ and the pilot vectors are chosen to be orthogonal to each other.

We begin by comparing the CRBs of the semi-blind and pilot-based approaches, assuming that both use the same number of pilots. Our goal is to investigate how substantial is the improvement in accuracy that results from using the information available in the data samples. We denote by L_p the number of pilots for the pilot-based CRB. We plot the semi-blind and the pilot-based CRBs versus SNR^6 for parameter a in Fig. 5.1 and for parameter b in Fig. 5.2. The number of pilots for the two bounds is $L = L_p = 8$, and the number of transmitted data symbols is N = 32. We also plot the corresponding MCRB for M = 4 and M = 256 in both figures. As we can see from the two plots, the semi-blind CRB is significantly lower than the pilot-based CRB for all modulation orders. Moreover, the gain in accuracy depends on the modulation order: the lower the modulation order the higher the gain. At high SNR, the accuracy gain is close to 4-fold for M = 4, M = 16 and M = 64, and close to 2-fold for M = 256, for both a and b. These gains are available at a relatively small number of data samples (N = 32), which demonstrates the practical worth of the semi-blind approach. We can also see from Figs. 5.1 and 5.2 that CRB_a and CRB_b behave differently at low SNR: CRB_a is significantly lower than the pilot-based CRB while CRB_b approaches the pilot-based CRB. This difference in behavior is due to presence of known self-interference symbols which make the estimation of a an easier task. In addition, Figs. 5.1 and 5.2 show that the MCRB is generally loose compared to the true CRB and is not sensitive to the modulation order. However, the MCRB provides a good approximation of the exact CRB at high SNR for M = 4 and M = 16.

We next consider the effect of the number of data samples on the semi-blind CRB. In Figs. 5.3 and 5.4, we plot versus N the ratio of the semi-blind CRB to the pilot-based CRB for parameters a and b, respectively. The number of pilots for the two bounds is $L = L_p = 8$, and the SNR is set at 20 dB. We see from the two plots that, as expected, the larger the number of samples, the higher the accuracy of the semi-blind approach. At N = 100, the semi-blind CRB for a is approximately 10 times lower than the pilot-based

⁵We note that when the pilot symbol vectors are orthogonal, the modulation order of the pilot symbols has no impact on CRB, and only the overall training power of each terminal matters, as we can see from (5.38) and (5.47).

⁶The SNR is defined as $10 \log \frac{P_2}{\sigma^2}$.

CRB for M = 4, 5 times lower for M = 16, and 3 times lower for M = 64, M = 256; the semi-blind CRB for b is approximately 5 times lower than the pilot-based CRB for M = 4, 3 times lower for M = 16 and 2 times lower for M = 64, M = 256. We note that the sample size N is constrained by the coherence time of the channel during which the channel parameters a, b remain fixed. Hence, the longer the channel coherence time the more attractive the semi-blind approach becomes.

So far we have compared the semi-blind and pilot-based CRBs when both use the same number pilots to show the superior accuracy of the semi-blind approach. We will next show that the semi-blind approach can also be employed to improve the spectral efficiency of the system by providing the same or higher accuracy than the pilot-based approach while using a much smaller number of pilots. For this purpose, we generate contour plots of the ratio of the semi-blind CRB to the pilot-based CRB, varying the number of pilots and data samples for the semi-blind CRB, while fixing number of pilots for the pilot-based CRB at $L_p = 10$. The SNR is set at 20 dB. The resulting contour plots for a and b are shown in Figs. 5.5 and 5.6, respectively. The regions in each plot for which the ratio is lower than 1 represent the combinations of L and N for which the semi-blind CRB is lower than the pilot-based CRB. We can see from the two plots that using just 4 pilots and 15 data samples, the semi-blind CRB becomes lower than the pilot-based CRB, which means that the semi-blind approach potentially achieves better accuracy while using 60% less pilots, thus providing a much better tradeoff between accuracy and spectral efficiency. This also means that the overall power that should be allocated for transmitting pilots is much less in the semi-blind scenario than in the pilot-based scenario.

5.5 Conclusions

In this chapter, we derived the exact CRB for semi-blind channel estimation in AF TWRNs that employ square QAM. Using the derived bound, we showed that the semi-blind approach provides substantial accuracy gains over the pilot-based approach. Only a limited number of data samples is needed to achieve these gains. Moreover, the accuracy gain depends on the modulation order, the lower the modulation order the higher the gain. Hence, semi-blind estimation makes it possible to use fewer pilots, resulting in better tradeoffs between accuracy and spectral efficiency. As a more tractable alternative to the exact CRB, we derived the MCRB. Although the MCRB has much simpler expressions, it is signifi-

cantly looser than the exact CRB, and it does not convey the impact of the modulation order on the achievable estimation accuracy. In the next chapter, we will focus on the design of of efficient low-complexity semi-blind algorithms whose performance approaches the exact CRB.



Fig. 5.1 Semi-blind and pilot-based CRBs for the estimation of a plotted versus SNR for N = 32 and L = 8. We also plot $MCRB_a$ for M = 4 and M = 256.



Fig. 5.2 Semi-blind and pilot-based CRBs for the estimation of b plotted versus SNR for N = 32 and L = 8.



Fig. 5.3 Ratio of semi-blind CRB for a to the pilot-based CRB for a plotted versus N. We use 8 pilots for both bounds, and we set the SNR to 20 dB.



Fig. 5.4 Ratio of semi-blind CRB for b to the pilot-based CRB for b plotted versus N. We use 8 pilots for both bounds, and we set the SNR to 20 dB.



Fig. 5.5 Contour plot of the ratio of the semi-blind CRB for a at M = 16 to the pilot-based CRB. We set the SNR to 20 dB, and we fix the number of pilots for the pilot-based CRB at 10.



Fig. 5.6 Contour plot of the ratio of the semi-blind CRB for *b* at M = 16 to the pilot-based CRB. We set the SNR to 20 dB, and we fix the number of pilots for the pilot-based CRB at 10.

Chapter 6

EM-based Semi-blind Channel Estimation for Flat-fading Channels

6.1 Introduction

In Chapter 5, we derived the CRB for semi-blind channel estimation in AF TWRNs employing square QAM. This was done assuming nonreciprocal channels and flat-fading channel conditions. The derived CRB is exact in the sense that it is based on the true-likelihood function that takes into account the actual statistics of the transmitted data symbols. By comparing the derived semi-blind CRB with the pilot-based CRB, we showed that the semiblind approach can provide substantial improvements in estimation accuracy and can be employed to significantly reduce the training overhead. These improvements were possible by incorporating only a limited number of data samples.

While the CRB analysis in Chapter 5 provides a very strong indication of the superior potential of the semi-blind approach, an equally important task is to design semi-blind estimation algorithms that can realize those gains and perform closely to the derived bounds, preferably at an affordable computational complexity. The focus of this chapter will be the accomplishment of this task. Unfortunately, the true likelihood function is highly complicated (see Chapter 5, Eq. (5.3)). Unlike the transmitted pilots, the exact values of the transmitted data symbols are not known beforehand, i.e., they constitute missing information and only their distribution (PMF) is known. The part of the likelihood function corresponding to the unknown data symbols has a mixed-Gaussian structure, instead of the much simpler Gaussian structure that would result if the transmitted data were known beforehand. Because of this, it is very difficult to perform exact ML estimation by maximizing the true likelihood function. In such scenarios where there is missing/incomplete data, the expectation maximization (EM) framework introduced in Chapter 2 provides a convenient low-complexity method for approximating the true ML solution. Starting with arbitrary values of the unknown parameters, the EM algorithm iterates between calculating the conditional expectation of the complete-data log-likelihood and maximizing this expectation with respect to the unknown parameters. Although the EM does not always converge to the true ML solution, it produces estimates with monotonically increasing likelihood.

In this chapter we derive the semi-blind EM-based channel estimator under flat-fading channel conditions for both nonreciprocal and reciprocal channels. The derived EM iterations have a low computational cost, and only a small number of iterations is needed to achieve convergence. Furthermore, we describe how the semi-blind CRB can be numerically evaluated for the case of reciprocal channels. Using simulations, we show that the EM algorithm performs very close to the semi-blind CRB and, even with a limited number of data samples, provides substantially better accuracy than the training-based least-squares (LS) estimator. We also show that the EM algorithm provides a significant improvement in throughput since a smaller number of pilots would be needed to achieve the same symbol-error rate (SER) as the LS estimator. The proposed EM methods thus combine high accuracy with computational efficiency, illustrating the practicality of semi-blind channel estimation for AF TWRNs.

The rest of the chapter is organized as follows. In Section 6.2 we present the system model. In Section 6.3, we derive the EM algorithm for both nonreciprocal and reciprocal channels. The numerical evaluation of the CRB for the reciprocal scenario is discussed in Section 6.4. Simulation results are presented in 6.5. Finally, our conclusions are discussed in Section 6.6. Part of the work presented in this chapter has appeared in [90].

6.2 System Model

The system model for the nonreciprocal scenario is identical to the one considered in Chapter 5, and its description is not repeated here to avoid redundancy. However, since we will also be considering the reciprocal scenario as well, we will highlight the differences in the received signal structure under the two assumptions. These differences will have a significant impact on the derivation of the EM algorithm.

6.2.1 Received Signal for Nonreciprocal Channels

In the nonreciprocal scenario, the received signal vector at terminal \mathcal{T}_1 corresponding to the *L* transmitted pilots is given by

$$\bar{\boldsymbol{z}} = Ah_1h_2\boldsymbol{t}_1 + Ag_1h_2\boldsymbol{t}_2 + Ah_2\boldsymbol{\omega} + \boldsymbol{\omega}_1 \tag{6.1}$$

and the received vector corresponding to the N transmitted data symbols is

$$\boldsymbol{z} = Ah_1h_2\boldsymbol{s}_1 + Ag_1h_2\boldsymbol{s}_2 + Ah_2\boldsymbol{n} + \boldsymbol{n}_1 \tag{6.2}$$

where all the terms in (6.1) and (6.2) have the same definitions as in Chapter 5. As before, the unknown channel parameters are $a = h_1h_2$, $b = g_1h_2$ and $\tau = |h_2|^2$.

6.2.2 Received Signal for Reciprocal Channels

In the reciprocal scenario, we have that $h_1 = h_2 = h$ and $g_1 = g_2 = g$. Hence, the received vector corresponding to the transmitted pilots is

$$\bar{\boldsymbol{z}} = Ah^2 \boldsymbol{t}_1 + Agh \boldsymbol{t}_2 + Ah\boldsymbol{\omega} + \boldsymbol{\omega}_1 \tag{6.3}$$

and the received vector corresponding to the N transmitted data symbols is

$$\boldsymbol{z} = Ah^2\boldsymbol{s}_1 + Agh\boldsymbol{s}_2 + Ah\boldsymbol{n} + \boldsymbol{n}_1.$$
(6.4)

In this case, the unknown channel parameters are $a = h^2$ and b = gh. There is no need to define a third parameter since the likelihood function can be expressed in terms of a and b.

Unlike Chapter 5, we do not restrict our attention in this chapter to square QAM. Rather, we just assume that the data symbols s_{21}, \ldots, s_{2N} are equiprobably drawn from the set of discrete constellation points $S = \{\xi_1, \ldots, \xi_M\}$, of size M.

6.3 Proposed Channel Estimation Algorithms

In this section, we derive the EM-based channel estimation algorithms for the nonreciprocal and reciprocal scenarios.

6.3.1 The EM Algorithm for Nonreciprocal Channels

We first derive the EM algorithm for the case of nonreciprocal channels. We will incorporate both the received pilot samples as well as the received data samples into the EM formulation¹. The unknown deterministic parameters are collected into the vector $\boldsymbol{\theta} = [a, b, \tau]^T$. The observed vectors $\{\bar{\boldsymbol{z}}, \boldsymbol{z}\}$ are the incomplete data, and the data symbols \boldsymbol{s}_2 are the hidden data. Hence, the complete data is $\{\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_2\}$, and the corresponding likelihood function is given by

$$f(\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_{2}; \boldsymbol{\theta}) = \frac{1}{M^{N}} f(\bar{\boldsymbol{z}}, \boldsymbol{z} | \boldsymbol{s}_{2}; \boldsymbol{\theta})$$

$$= \frac{1}{M^{N} (\pi \sigma^{2} (A^{2} \tau + 1))^{N+L}} e^{-\frac{\|\bar{\boldsymbol{z}} - Aat_{1} - Abt_{2}\|^{2}}{\sigma^{2} (A^{2} \tau + 1)}} e^{-\frac{\|\boldsymbol{z} - Aas_{1} - Abs_{2}\|^{2}}{\sigma^{2} (A^{2} \tau + 1)}}.$$
(6.5)

The resulting log-likelihood function (LLF) is

$$\mathcal{L}(\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_{2}; \boldsymbol{\theta}) = -N \log M - (N+L) \log(\pi \sigma^{2} (A^{2}\tau + 1)) - \frac{1}{\sigma^{2} (A^{2}\tau + 1)} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_{1} - Ab\boldsymbol{t}_{2}\|^{2} - \frac{1}{\sigma^{2} (A^{2}\tau + 1)} \|\boldsymbol{z} - Aa\boldsymbol{s}_{1} - Ab\boldsymbol{s}_{2}\|^{2}.$$
(6.6)

We let $\boldsymbol{\theta}^{(t)} = [a^{(t)}, b^{(t)}, \tau^{(t)}]^T$ be the estimate of $\boldsymbol{\theta}$ at iteration t. We also denote by $\beta_{i,j}^{(t)} \triangleq P(s_{2i} = \xi_j | z_i; \boldsymbol{\theta}^{(t)})$ the posterior PMF of the *i*th data symbol conditioned on $\boldsymbol{\theta}^{(t)}$, given by

$$\beta_{i,j}^{(t)} = \frac{f(z_i|s_{2i} = \xi_j; \theta^{(t)})}{\sum\limits_{k=1}^{M} f(z_i|s_{2i} = \xi_k; \theta^{(t)})} = \frac{e^{-\frac{|z_i - Aa^{(t)}s_{1i} - Ab^{(t)}\xi_j|^2}{\sigma^2(A^2\tau^{(t)} + 1)}}}{\sum\limits_{k=1}^{M} e^{-\frac{|z_i - Aa^{(t)}s_{1i} - Ab^{(t)}\xi_k|^2}{\sigma^2(A^2\tau^{(t)} + 1)}}}.$$
(6.7)

The two steps of the EM algorithm for the nonreciprocal scenario are as follows: E-step

 $^{^1\}mathrm{Another}$ possibility is to formulate the EM using just the data samples and use the pilot samples only for initialization.

We have²

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = \mathbb{E}\left\{ \mathcal{L}(\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_{2}; \boldsymbol{\theta}) | P(\boldsymbol{s}_{2} | \boldsymbol{z}; \boldsymbol{\theta}^{(t)}) \right\}$$

$$= -(N+L) \log(\pi \sigma^{2} (A^{2} \tau + 1)) - \frac{1}{\sigma^{2} (A^{2} \tau + 1)} \| \bar{\boldsymbol{z}} - Aa \boldsymbol{t}_{1} - Ab \boldsymbol{t}_{2} \|^{2}$$

$$- \frac{1}{\sigma^{2} (A^{2} \tau + 1)} \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} | \boldsymbol{z}_{i} - Aa \boldsymbol{s}_{1i} - Ab \boldsymbol{\xi}_{j} |^{2}.$$
 (6.8)

M-step

We now solve for $a^{(t+1)}$, $b^{(t+1)}$ and $\tau^{(t+1)}$. We have

$$\left\{a^{(t+1)}, b^{(t+1)}, \tau^{(t+1)}\right\} = \arg \max_{\boldsymbol{\theta} = [a, b, \tau]^T} Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}\right).$$
(6.9)

It can be verified that the solutions for (6.9) are given by

$$a^{(t+1)} = \frac{(\boldsymbol{s}_{1}^{H}\boldsymbol{z} + \boldsymbol{t}_{1}^{H}\bar{\boldsymbol{z}})\left(\sum_{i=1}^{N}\sum_{j=1}^{M}\beta_{i,j}^{(t)}|\xi_{j}|^{2} + \boldsymbol{t}_{2}^{H}\boldsymbol{t}_{2}\right) - \left(\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{2} + \sum_{i=1}^{N}\sum_{j=1}^{M}\beta_{i,j}^{(t)}\boldsymbol{s}_{1i}^{*}\xi_{j}\right)\left(\boldsymbol{t}_{2}^{H}\bar{\boldsymbol{z}} + \sum_{i=1}^{N}\sum_{j=1}^{M}\beta_{i,j}^{(t)}\xi_{j}^{*}\boldsymbol{z}_{i}\right)}{A(\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1} + \boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}) - A\left|\sum_{i=1}^{N}\sum_{j=1}^{M}\beta_{i,j}^{(t)}\xi_{j}^{*}\boldsymbol{s}_{1i} + \boldsymbol{t}_{2}^{H}\boldsymbol{t}_{1}\right|^{2}},$$

$$(6.10)$$

$$b^{(t+1)} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \xi_{j}^{*} z_{i} - A \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} a^{(t+1)} \xi_{j}^{*} s_{1i} + \mathbf{t}_{2}^{H} \bar{\mathbf{z}} - A a^{(t+1)} \mathbf{t}_{2}^{H} \mathbf{t}_{2}}{A \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} |\xi_{j}|^{2} + A \mathbf{t}_{2}^{H} \mathbf{t}_{2}}, \quad (6.11)$$

and

$$\tau^{(t+1)} = \max\left(0, \frac{1}{A^2} \left(\frac{1}{(N+L)\sigma^2} \left(\sum_{i=1}^N \sum_{j=1}^M \beta_{i,j}^{(t)} |z_i - Aa^{(t+1)}s_{1i} - Ab^{(t+1)}\xi_j|^2 + \|\bar{\boldsymbol{z}} - Aa^{(t+1)}\boldsymbol{t}_1 - Ab^{(t+1)}\boldsymbol{t}_2\|^2\right) - 1\right)\right).$$
(6.12)

²We ignore the term $N \log M$ since it has no impact on the solution.

The expression for $\tau^{(t+1)}$ is obtained by taking into account the constraint $\tau \ge 0$.

The EM algorithm can be initialized using arbitrary values. However, since L pilot samples are available, we can speed up the convergence by employing the pilot-based ML estimates to initialize the EM procedure. The pilot-based ML estimates of a, b, τ are given by

$$a^{(0)} = \frac{\boldsymbol{t}_2^H \boldsymbol{t}_2 \boldsymbol{t}_1^H \bar{\boldsymbol{z}} - \boldsymbol{t}_2^H \bar{\boldsymbol{z}}}{A(\boldsymbol{t}_1^H \boldsymbol{t}_1 \boldsymbol{t}_2^H \boldsymbol{t}_2 - \boldsymbol{t}_2^H \boldsymbol{t}_1)},\tag{6.13}$$

$$b^{(0)} = \frac{\boldsymbol{t}_2^H \boldsymbol{t}_2 \boldsymbol{t}_1^H \boldsymbol{t}_1 \boldsymbol{t}_2^H \bar{\boldsymbol{z}} - \boldsymbol{t}_2^H \boldsymbol{t}_1 \boldsymbol{t}_2^H \boldsymbol{t}_1 \boldsymbol{t}_1^H \bar{\boldsymbol{z}}}{\boldsymbol{t}_2^H \boldsymbol{t}_2 (\boldsymbol{t}_1^H \boldsymbol{t}_1 \boldsymbol{t}_2^H \boldsymbol{t}_2 - \boldsymbol{t}_2^H \boldsymbol{t}_1)},$$
(6.14)

and

$$\tau^{(0)} = \max\left(0, \frac{1}{A^2} \left(\frac{1}{L\sigma^2} \left(\left\|\bar{\boldsymbol{z}} - Aa^{(0)}\boldsymbol{t}_1 - Ab^{(0)}\boldsymbol{t}_2\right\|^2\right) - 1\right)\right).$$
(6.15)

It is clear from (6.10), (6.11) and (6.12) that the complexity of each EM iteration is O(MN), i.e., it is linear in the number of data samples for a given modulation order. Moreover, as we shall see in Section 6.5, the number of EM iterations needed is achieve convergence is very small, which shows that the overall computational cost of the EM algorithm is low.

6.3.2 The EM Algorithm for Reciprocal Channels

We now derive the EM algorithm for the case of reciprocal channels. Under channel reciprocity, we have only two parameters to estimate, $a = h^2$ and b = gh, i.e., $\boldsymbol{\theta} = [a, b]^T$. As before, \boldsymbol{s}_2 represents the hidden data, and the complete data set is $\{\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_2\}$. As we shall see, the derivation of the EM algorithm is more involved under the reciprocity assumption. The likelihood function for the complete data is given by

$$f(\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_2; \boldsymbol{\theta}) = \frac{1}{M^N(\pi\sigma^2(A^2|a|+1))^{N+L}} e^{-\frac{\|\bar{\boldsymbol{z}}-Aa\boldsymbol{t}_1-Ab\boldsymbol{t}_2\|^2}{\sigma^2(A^2|a|+1)}} e^{-\frac{\|\boldsymbol{z}-Aa\boldsymbol{s}_1-Ab\boldsymbol{s}_2\|^2}{\sigma^2(A^2|a|+1)}}.$$
 (6.16)

The resulting LLF is

$$\mathcal{L}(\bar{\boldsymbol{z}}, \boldsymbol{z}, \boldsymbol{s}_{2}; \boldsymbol{\theta}) = -(N+L) \log(\pi \sigma^{2} (A^{2}|a|+1)) - \frac{1}{\sigma^{2} (A^{2}|a|+1)} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_{1} - Ab\boldsymbol{t}_{2}\|^{2} - \frac{1}{\sigma^{2} (A^{2}|a|+1)} \|\boldsymbol{z} - Aa\boldsymbol{s}_{1} - Ab\boldsymbol{s}_{2}\|^{2}.$$
(6.17)

Furthermore, the posterior PMF of the ith data symbol is given by

$$\beta_{i,j}^{(t)} = \frac{f(z_i|s_{2i} = \xi_j; \theta^{(t)})}{\sum_{k=1}^{M} f(z_i|s_{2i} = \xi_k; \theta^{(t)})} = \frac{e^{-\frac{1}{\sigma^2(A^2|a^{(t)}|+1)}|z_i - Aa^{(t)}s_{1i} - Ab^{(t)}\xi_j|}}{\sum_{k=1}^{M} e^{-\frac{1}{\sigma^2(A^2|a^{(t)}|+1)}|z_i - Aa^{(t)}s_{1i} - Ab^{(t)}\xi_k|}}.$$
(6.18)

E-step

We have

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}) = -N \log M - (N+L) \log(\pi \sigma^2 (A^2 |a|+1)) - \frac{1}{\sigma^2 (A^2 |a|+1)} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_1 - Ab\boldsymbol{t}_2\|^2 - \frac{1}{\sigma^2 (A^2 |a|+1)} \sum_{i=1}^N \sum_{j=1}^M \beta_{i,j}^{(t)} |z_i - Aas_{1i} - Ab\xi_j|^2.$$
(6.19)

M-step

The derivation of the M-step is more complicated for the reciprocal case because of the term |a| that appears in logarithmic term and the denominators of the last two terms in the RHS of (6.19). We need to obtain the values $a^{(t+1)}$, $b^{(t+1)}$ such that

$$\left\{a^{(t+1)}, b^{(t+1)}\right\} = \arg \max_{\boldsymbol{\theta} = [a,b]^T} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}).$$
(6.20)

Regarding $b^{(t+1)}$, it can be easily verified that the value of b that maximizes $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)})$ for a given value of a is

$$b_{o}(a) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \xi_{j}^{*} z_{i} - Aa \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \xi_{j}^{*} s_{1i} + \boldsymbol{t}_{2}^{H} \bar{\boldsymbol{z}} - Aa \boldsymbol{t}_{2}^{H} \boldsymbol{t}_{2}}{A \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} |\xi_{j}|^{2} + A \boldsymbol{t}_{2}^{H} \boldsymbol{t}_{2}}.$$
(6.21)

Substituting $b_o(a)$ in place of b in (6.19), we obtain the following updated objective function

that depends only on a:

$$Q(a; \boldsymbol{\theta}^{(t)}) = -N \log(\pi \sigma^2 (A^2 |a| + 1)) - \frac{1}{G^2 \sigma^2 (A^2 |a| + 1)} \|G\bar{\boldsymbol{z}} - AaG\boldsymbol{t}_1 - A\mathcal{I}\boldsymbol{t}_2 + A^2 a\mathcal{X}\boldsymbol{t}_2\|^2 - \frac{1}{\sigma^2 (A^2 |a| + 1)} \frac{1}{G^2} \sum_{i=1}^N \sum_{j=1}^M \beta_{i,j}^{(t)} |G\boldsymbol{z}_i - AaG\boldsymbol{s}_{1i} - A\mathcal{I}\boldsymbol{\xi}_j + A^2 a\mathcal{X}\boldsymbol{\xi}_j|^2,$$
(6.22)

where

$$G = A(\sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} |\xi_j|^2 + \mathbf{t}_2^H \mathbf{t}_2), \quad \mathcal{I} = \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \xi_j^* \mathbf{z}_i + \mathbf{t}_2^H \bar{\mathbf{z}}, \quad \mathcal{X} = \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \xi_j^* s_{1i} + \mathbf{t}_2^H \mathbf{t}_1.$$
(6.23)

In order to maximize (6.22) with respect to a, we will maximize it first with respect to the phase $\phi_a \triangleq \angle a$ and then the amplitude |a|. Maximizing (6.22) with respect to ϕ_a is equivalent to minimizing the following term with respect to ϕ_a

$$\Lambda(a;\boldsymbol{\theta}^{(t)}) \triangleq \left\| G\bar{\boldsymbol{z}} - AaG\boldsymbol{t}_1 - A\mathcal{I}\boldsymbol{t}_2 + A^2a\mathcal{X}\boldsymbol{t}_2 \right\|^2 + \sum_{i=1}^N \sum_{j=1}^M \beta_{i,j}^{(t)} \left| Gz_i - AaGs_{1i} - A\mathcal{I}\xi_j + A^2a\mathcal{X}\xi_j \right|^2.$$

$$\tag{6.24}$$

Moreover, we can expand $\Lambda(a; \boldsymbol{\theta}^{(t)})$ as

$$\Lambda(a;\boldsymbol{\theta}^{(t)}) = \left\| G\bar{\boldsymbol{z}} - A\mathcal{I}\boldsymbol{t}_2 \right\|^2 + |a|^2 \left\| A^2 \mathcal{X}\boldsymbol{t}_2 - AG\boldsymbol{t}_1 \right\|^2 + 2\Re \left\{ a(G\bar{\boldsymbol{z}} - A\mathcal{I}\boldsymbol{t}_2)^H (A^2 \mathcal{X}\boldsymbol{t}_2 - AG\boldsymbol{t}_1) \right\} \\ + \sum_{i=1}^N \sum_{j=1}^M \beta_{i,j}^{(t)} \left(\left| Gz_i - A\mathcal{I}\xi_j \right|^2 + |a|^2 \left| A^2 \mathcal{X}\xi_j - AGs_{1i} \right|^2 + 2\Re \left\{ a(Gz_i - A\mathcal{I}\xi_j)^* (A^2 \mathcal{X}\xi_j - AGs_{1i}) \right\} \right).$$

$$(6.25)$$

Minimizing (6.25) with respect to ϕ_a , we obtain

$$\phi_{a}^{(t+1)} = \pi - \angle \left(\sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \left(Gz_{i} - A\mathcal{I}\xi_{j} \right)^{*} \left(A^{2}\mathcal{X}\xi_{j} - AGs_{1i} \right) + \left(G\bar{\boldsymbol{z}} - A\mathcal{I}\boldsymbol{t}_{2} \right)^{H} \left(A^{2}\mathcal{X}\boldsymbol{t}_{2} - AG\boldsymbol{t}_{1} \right) \right).$$

$$(6.26)$$

88

Substituting $\phi_a^{(t+1)}$ into (6.22), we obtain the following function that depends only on |a|:

$$Q(|a|;\boldsymbol{\theta}^{(t)}) = -N\log(\pi\sigma^2(A^2|a|+1)) - \frac{\breve{U}}{\sigma^2(A^2|a|+1)} - \frac{\breve{V}|a|^2}{\sigma^2(A^2|a|+1)} + \frac{2\breve{W}|a|}{\sigma^2(A^2|a|+1)},$$
(6.27)

where

$$\check{U} = \frac{1}{G^2} \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} |Gz_i - A\mathcal{I}\xi_j|^2 + \frac{1}{G^2} ||G\bar{z} - A\mathcal{I}t_2||^2,$$
(6.28)

$$\breve{V} = \frac{1}{G^2} \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} |A^2 \mathcal{X} \xi_j - AGs_{1i}|^2 + \frac{1}{G^2} ||A^2 \mathcal{X} t_2 - AGt_1||^2,$$
(6.29)

and

$$\breve{W} = \frac{1}{G^2} \bigg| \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{i,j}^{(t)} \big(Gz_i - A\mathcal{I}\xi_j \big)^* \big(A^2 \mathcal{X}\xi_j - AGs_{1i} \big) + \big(G\bar{\boldsymbol{z}} - A\mathcal{I}\boldsymbol{t}_2 \big)^H \big(A^2 \mathcal{X}\boldsymbol{t}_2 - AG\boldsymbol{t}_1 \big) \bigg|.$$

$$(6.30)$$

The derivative of (6.27) w.r.t. |a| is

$$\frac{d}{d|a|}Q(|a|;\boldsymbol{\theta}^{(t)}) = -\frac{NA^2}{A^2|a|+1} + \frac{A^2\breve{U}}{\sigma^2(A^2|a|+1)^2} - \frac{A^2\breve{V}|a|^2 + 2\breve{V}|a|}{\sigma^2(A^2|a|+1)^2} + \frac{2\breve{W}}{\sigma^2(A^2|a|+1)^2}.$$
 (6.31)

Setting $\frac{d}{d|a|}Q(|a|; \boldsymbol{\theta}^{(t)}) = 0$, we obtain the quadratic equation

$$A^{2}\breve{V}|a|^{2} + (2\breve{V} + NA^{4}\sigma^{2})|a| + NA^{2}\sigma^{2} - A^{2}\breve{U} - 2\breve{W} = 0.$$
 (6.32)

Solving (6.32), we finally get

$$|a|^{(t+1)} = \frac{-(2\breve{V} + NA^4\sigma^2) + \sqrt{(2\breve{V} + NA^4\sigma^2)^2 - 4A^2\breve{V}(NA^2\sigma^2 - A^2\breve{U} - 2\breve{W})}}{2A^2\breve{V}}.$$
 (6.33)

Similar to the nonreciprocal scenario, we can see from (6.21), (6.26) and (6.33) that the computational complexity of each EM iterations is O(MN), which confirms the computational efficiency of the EM approach.

6.4 CRB for the Reciprocal Scenario

The exact CRB for semi-blind channel estimation in the nonreciprocal scenario was derived analytically in Chapter 5 assuming square QAM. A similar approach may be followed to derive the semi-blind CRB for the reciprocal scenario. To avoid redundancy, however, we will only show in this section how the semi-blind CRB for the reciprocal scenario can be numerically evaluated using the Monte-Carlo approach. As in Chapter 5, we focus on square QAM.

We let $\tilde{\boldsymbol{z}} \triangleq [\bar{\boldsymbol{z}}^T, \boldsymbol{z}^T]^T$ and denote by $\boldsymbol{\theta}_R \triangleq [a_R, a_I, b_R, b_I]^T$ the vector of real channel parameters, where $a_R \triangleq \Re\{a\}, a_I \triangleq \Im\{a\}, b_R \triangleq \Re\{b\}, b_I \triangleq \Im\{b\}$. Following the approach in Chapter 5, the LLF of $\tilde{\boldsymbol{z}}$ may be expressed as

$$\mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta}_{R}) = -(N+L)\log(\pi\sigma^{2}(A^{2}|a|+1)) + N\log\frac{4}{M} - \frac{1}{\sigma^{2}(A^{2}|a|+1)}\|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_{1} - Ab\boldsymbol{t}_{2}\|^{2} - \frac{1}{\sigma^{2}(A^{2}|a|+1)}\|\boldsymbol{z} - Aa\boldsymbol{s}_{1}\|^{2} + \sum_{k=1}^{N}\log F_{\theta}(u_{k}) + \sum_{k=1}^{N}\log F_{\theta}(v_{k}).$$
(6.34)

where

$$F_{\theta}(x) \triangleq \sum_{i=1}^{2^{p-1}} e^{-\gamma_i |b|^2} \cosh\left(2\beta_i x\right), \quad \beta_i \triangleq \frac{Ad_p}{\sigma^2 (A^2 |a|+1)} (2i-1), \quad \gamma_i \triangleq \frac{A^2 d_p^2}{\sigma^2 (A^2 |a|+1)} (2i-1)^2,$$
(6.35)

and

$$u_k = \Re\{(z_k - Aas_{1k})^*b\}, \ v_k = \Im\{(z_k - Aas_{1k})^*b\}.$$
(6.36)

We denote by $I(\theta_R)$ the corresponding FIM, and by $I_{x,y}$ the joint Fisher information between the parameters x and y, where $x, y \in \{a_R, a_I, b_R, b_I\}$. Hence,

$$I_{x,y} = \mathbb{E}\left\{\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial x}\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial y}\right\}$$
(6.37)

and

$$\boldsymbol{I}(\boldsymbol{\theta}_{R}) = \mathbb{E}\left\{\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta}_{R})}{\partial \boldsymbol{\theta}_{R}}\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta}_{R})}{\partial \boldsymbol{\theta}_{R}^{T}}\right\} = \begin{bmatrix}\boldsymbol{I}_{aa} & \boldsymbol{I}_{ab}\\ \boldsymbol{I}_{ab}^{T} & \boldsymbol{I}_{bb}\end{bmatrix},$$
(6.38)
where

$$\boldsymbol{I}_{aa} = \begin{bmatrix} I_{a_{R},a_{R}} & I_{a_{R},a_{I}} \\ I_{a_{R},a_{I}} & I_{a_{I},a_{I}} \end{bmatrix}, \quad \boldsymbol{I}_{ab} = \begin{bmatrix} I_{a_{R},b_{R}} & I_{a_{R},b_{I}} \\ I_{a_{I},b_{R}} & I_{a_{I},b_{I}} \end{bmatrix}, \quad \boldsymbol{I}_{bb} = \begin{bmatrix} I_{b_{R},b_{R}} & I_{b_{R},b_{I}} \\ I_{b_{R},b_{I}} & I_{b_{I},b_{I}} \end{bmatrix}. \quad (6.39)$$

To approximate the expectation in (6.38) using Monte-Carlo simulations, we generate a large number of realizations of the vectors \boldsymbol{z} and $\bar{\boldsymbol{z}}$ and then average the product $\frac{\partial \mathcal{L}(\tilde{z};\theta)}{\partial \theta_R} \frac{\partial \mathcal{L}(\tilde{z};\theta)}{\partial \theta_R^T}$ over all realizations. The CRB can then be obtained by taking the inverse of I_{θ_R} . To do this, however, we need expressions for $\frac{\partial L(\tilde{z};\theta)}{\partial a_R}$, $\frac{\partial L(\tilde{z};\theta)}{\partial a_I}$, $\frac{\partial L(\tilde{z};\theta)}{\partial b_R}$ and $\frac{\partial L(\tilde{z};\theta)}{\partial b_I}$. We have

$$\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial a_{R}} = -\frac{(N+L)A^{2}a_{R}}{(A^{2}|a|+1)|a|} - \frac{2A}{\sigma^{2}(A^{2}|a|+1)} \left(A\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1}a_{R} - \Re\{\boldsymbol{t}_{1}^{H}(\bar{\boldsymbol{z}}-Ab\boldsymbol{t}_{2})\}\right) \\
+ \frac{A^{2}a_{R}}{\sigma^{2}(A^{2}|a|+1)^{2}|a|} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_{1} - Ab\boldsymbol{t}_{2}\|^{2} - \frac{2A}{\sigma^{2}(A^{2}|a|+1)} \left(A\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}a_{R} - \Re\{\boldsymbol{s}_{1}^{H}\boldsymbol{z}\}\right) \\
+ \frac{A^{2}a_{R}}{\sigma^{2}(A^{2}|a|+1)^{2}|a|} \|\boldsymbol{z} - Aa\boldsymbol{s}_{1}\|^{2} + \sum_{k=1}^{N} \frac{\frac{\partial F_{\theta}(u_{k})}{\partial a_{R}}}{F_{\theta}(u_{k})} + \sum_{k=1}^{N} \frac{\frac{\partial F_{\theta}(v_{k})}{\partial a_{R}}}{F_{\theta}(v_{k})}, \tag{6.40}$$

where

$$\frac{\partial F_{\theta}(u_{k})}{\partial a_{R}} = \sum_{i=1}^{2^{p-1}} \frac{A^{2}a_{R}}{|a|} \sigma^{2} \beta_{i}^{2} |b|^{2} e^{-\gamma_{i}|b|^{2}} \cosh[2\beta_{i}u_{k}] + \sum_{i=1}^{2^{p-1}} e^{-\gamma_{i}|b|^{2}} \sinh[2\beta_{i}u_{k}] \left[-2\frac{\beta_{i}A^{2}a_{R}u_{k}}{\sigma^{2}(A^{2}|a|+1)|a|} + 2\beta_{i}\frac{\partial u_{k}}{\partial a_{R}} \right],$$

$$\frac{\partial F_{\theta}(v_{k})}{\partial a_{R}} = \sum_{i=1}^{2^{p-1}} \frac{A^{2}a_{R}}{|a|} \sigma^{2} \beta_{i}^{2} |b|^{2} e^{-\gamma_{i}|b|^{2}} \cosh[2\beta_{i}v_{k}] + \sum_{i=1}^{2^{p-1}} e^{-\gamma_{i}|b|^{2}} \sinh[2\beta_{i}v_{k}] \left[-2\frac{\beta_{i}A^{2}a_{R}v_{k}}{\sigma^{2}(A^{2}|a|+1)|a|} + 2\beta_{i}\frac{\partial v_{k}}{\partial a_{R}} \right],$$

$$(6.41)$$

$$+ \sum_{i=1}^{2^{p-1}} e^{-\gamma_{i}|b|^{2}} \sinh[2\beta_{i}v_{k}] \left[-2\frac{\beta_{i}A^{2}a_{R}v_{k}}{\sigma^{2}(A^{2}|a|+1)|a|} + 2\beta_{i}\frac{\partial v_{k}}{\partial a_{R}} \right],$$

and

 $i{=}1$

$$\frac{\partial u_k}{\partial a_R} = -A\Re\{s_{1k}^*b\}, \quad \frac{\partial v_k}{\partial a_R} = -A\Im\{s_{1k}^*b\}.$$
(6.43)

Similarly, we have

$$\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial a_{I}} = -\frac{(N+L)A^{2}a_{I}}{(A^{2}|a|+1)|a|} - \frac{2A}{\sigma^{2}(A^{2}|a|+1)} \left(A\boldsymbol{t}_{1}^{H}\boldsymbol{t}_{1}a_{I} - \Im\{\boldsymbol{t}_{1}^{H}(\bar{\boldsymbol{z}}-Ab\boldsymbol{t}_{2})\} \right)
+ \frac{A^{2}a_{I}}{\sigma^{2}(A^{2}|a|+1)^{2}|a|} \|\bar{\boldsymbol{z}} - Aa\boldsymbol{t}_{1} - Ab\boldsymbol{t}_{2}\|^{2} - \frac{2A}{\sigma^{2}(A^{2}|a|+1)} \left(A\boldsymbol{s}_{1}^{H}\boldsymbol{s}_{1}a_{I} - \Im\{\boldsymbol{s}_{1}^{H}\boldsymbol{z}\} \right)
+ \frac{A^{2}a_{I}}{\sigma^{2}(A^{2}|a|+1)^{2}|a|} \|\boldsymbol{z} - Aa\boldsymbol{s}_{1}\|^{2} + \sum_{k=1}^{N} \frac{\frac{\partial F_{\theta}(u_{k})}{\partial a_{I}}}{F_{\theta}(u_{k})} + \sum_{k=1}^{N} \frac{\frac{\partial F_{\theta}(v_{k})}{\partial a_{I}}}{F_{\theta}(v_{k})}.$$
(6.44)

The expressions for $\frac{\partial F_{\theta}(u_k)}{\partial a_I}$ and $\frac{\partial F_{\theta}(v_k)}{\partial a_I}$ can be obtained by replacing a_R with a_I in (6.41) and (6.42), respectively and noting that

$$\frac{\partial u_k}{\partial a_I} = -A\Im\{s_{1k}^*b\}, \quad \frac{\partial v_k}{\partial a_I} = A\Re\{s_{1k}^*b\}. \tag{6.45}$$

We next consider $\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial b_R}$ and $\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial b_I}$. We have

$$\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial b_R} = \frac{1}{\sigma^2 (A^2|a|+1)} \left(2A\Re\{\boldsymbol{t}_2^H(\bar{\boldsymbol{z}}-Aa\boldsymbol{t}_1)\} \right) + \sum_{k=1}^N \frac{\frac{\partial F_{\theta}(u_k)}{\partial b_R}}{F_{\theta}(u_k)} + \sum_{k=1}^N \frac{\frac{\partial F_{\theta}(v_k)}{\partial b_R}}{F_{\theta}(v_k)}, \quad (6.46)$$

where

$$\frac{\partial F_{\theta}(u_k)}{\partial b_R} = -\sum_{i=1}^{2^{p-1}} 2\gamma_i b_R e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k] + \sum_{i=1}^{2^{p-1}} 2\beta_i \Re\{z_k - Aas_{1k}\} e^{-\gamma_i |b|^2} \sinh[2\beta_i u_k)],\tag{6.47}$$

and

$$\frac{\partial F_{\theta}(v_k)}{\partial b_R} = -\sum_{i=1}^{2^{p-1}} 2\gamma_i b_R e^{-\gamma_i |b|^2} \cosh[2\beta_i v_k] - \sum_{i=1}^{2^{p-1}} 2\beta_i \Im\{z_k - Aas_{1k}\} e^{-\gamma_i |b|^2} \sinh[2\beta_i v_k)].$$
(6.48)

Similarly,

$$\frac{\partial \mathcal{L}(\tilde{\boldsymbol{z}};\boldsymbol{\theta})}{\partial b_I} = \frac{1}{\sigma^2 (A^2|a|+1)} \left(2A\Im\{\boldsymbol{t}_2^H(\bar{\boldsymbol{z}}-Aa\boldsymbol{t}_1)\} \right) + \sum_{k=1}^N \frac{\frac{\partial F_{\theta}(u_k)}{\partial b_I}}{F_{\theta}(u_k)} + \sum_{k=1}^N \frac{\frac{\partial F_{\theta}(v_k)}{\partial b_I}}{F_{\theta}(v_k)}, \quad (6.49)$$

where

$$\frac{\partial F_{\theta}(u_k)}{\partial b_I} = -\sum_{i=1}^{2^{p-1}} 2\gamma_i b_I e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k] + \sum_{i=1}^{2^{p-1}} 2\beta_i \Im\{z_k - Aas_{1k}\} e^{-\gamma_i |b|^2} \sinh[2\beta_i u_k], \quad (6.50)$$

and

$$\frac{\partial F_{\theta}(v_k)}{\partial b_I} = -\sum_{i=1}^{2^{p-1}} 2\gamma_i b_I e^{-\gamma_i |b|^2} \cosh[2\beta_i v_k] + \sum_{i=1}^{2^{p-1}} 2\beta_i \Re\{z_k - Aas_{1k}\} e^{-\gamma_i |b|^2} \sinh[2\beta_i v_k].$$
(6.51)

Finally, utilizing the expressions in (6.40), (6.44), (6.46) and (6.49), it is possible to numerically evaluate all the elements of $I(\theta_R)$ via the Monte Carlo approach by generating a large number of realizations of the vectors \boldsymbol{z} and $\bar{\boldsymbol{z}}$. The CRB is then obtained by taking the inverse of $I(\theta_R)$.

6.5 Simulation Results

In this section, we investigate through simulations the MSE performance of the derived EM algorithm and compare it to the pilot-based LS estimator for both nonreciprocal and reciprocal channels. Our results are obtained assuming that $P_1 = P_2 = P_r$, where P_1, P_2 and P_r are the transmission powers at the terminal \mathcal{T}_1 and \mathcal{T}_2 and the relay \mathcal{R} , respectively. For the nonreciprocal scenario, the channel parameters are generated exactly as in Chapter 5. That is, we model h_1 and h_2 as correlated complex Gaussian RVs with mean zero, variance 1, and a correlation coefficient $\rho = 0.3$; we also model g_1 and g_2 as correlated complex Gaussian RVs with the same mean, variance and correlation coefficient, but independent of h_1 and h_2 . For the reciprocal scenario, we model h and g as independent complex Gaussian RVs with mean zero and variance 1. For both scenarios, we average our results over 100 independent realizations of the channel parameters. As in Chapter 5, the data symbols are generated using square QAM modulation. To explore the impact of the modulation order on the performance of the EM algorithm, we consider in our simulations the modulation orders M = 4, M = 16 and M = 64. Unless mentioned otherwise, the number of pilots and the number of data symbols are set at L = 8 and N = 32, respectively. In our plots we consider the total MSE, which is the sum of the MSE for the estimation of a and the estimation of b.

In Fig. 6.1, we plot the MSE performance of the derived semi-blind EM algorithm and the pilot-based LS estimator versus SNR³ for the nonreciprocal scenario. The channel estimates are obtained after 4 iterations of the EM algorithm. For comparison, we also plot the MSE of the LS estimator that only uses the pilot samples, as well as the semi-blind CRB (obtained in Chapter 5). As we can see from Fig. 6.1, the MSE performance of the EM algorithm is very close to the CRB for the whole SNR range and for all modulation orders. Moreover, the EM algorithm provides substantially higher accuracy than the LS estimator. This demonstrates that the improvements in estimation accuracy predicted in Chapter 5 can indeed be realized at a low computational cost since only 4 EM iterations were used.

In Fig. 6.2, we plot the MSE performance of the semi-blind EM algorithm and the pilotbased LS estimator versus SNR for the reciprocal scenario. The EM estimates are again generated using 4 iterations. We also plot the semi-blind CRB for the reciprocal scenario whose numerical evaluation was discussed in Section 6.4. Similarly to the nonreciprocal scenario, the EM algorithm performs very closely to the semi-blind CRB and provides a significant improvement in accuracy over the LS estimator.

We next consider the effect of the number of data samples on the performance of the EM algorithm. In Figs. 6.3 and 6.4, we plot the MSE performance of the EM algorithm versus the number of data samples, normalized w.r.t. the MSE of the LS estimator (which does not depend on the number of data samples) for the nonreciprocal and reciprocal cases, respectively. We use 8 pilots and 10 EM iterations, and set the SNR at 15dB. The (normalized) semi-blind CRB is also plotted as a reference. As we can see in both plots, the EM algorithm performs close to the CRB even for large data sizes. Moreover, the larger the number of data samples, the higher the accuracy gain, which shows that, as noted in Chapter 5, semi-blind estimation becomes more attractive as the channel coherence time increases.

We next consider the convergence behavior of the EM algorithm. In Fig. 6.5, we plot the MSE of the EM algorithm versus the number of iterations for N = 32 and N = 100, assuming 8 pilots and an SNR of 15dB. Fig. 6.5 shows that the number of iterations needed to achieve convergence is small for all modulation orders. In all cases, convergence is achieved within at most 12 iterations (as few as 4 iterations are sufficient in some cases). Convergence becomes slightly slower as the modulation order increases and as the number of

³As before, the SNR is defined as $10 \log \frac{P_2}{\sigma^2}$

data samples increases. Very similar trends are observed for the case of reciprocal channels in Fig. 6.6.

In Fig. 6.7, we compare the SER performance of the EM algorithm to that of the LS estimator, as well as to the SER performance for perfect channel state information (CSI) assuming reciprocal channels. We assume that the channel is fixed for the duration of 40 samples and employ 4 pilots for the both estimators, and N = 36 data samples for the EM algorithm. We use QPSK modulation (M = 4). As we can see in Fig. 6.7, the EM algorithm provides better SER performance than the LS estimator and is much closer to the performance under perfect CSI. Hence, the higher estimation accuracy of the EM algorithm also results in an SNR gain, since a lower SNR is required to achieve the same SER as the LS estimator. Similar results can be obtained for the nonreciprocal case.

Finally, in Table I, we compare the achievable throughput of the proposed EM algorithm and the LS estimator for the reciprocal case at different SNR and SER values, assuming that the channel is constant for the duration of 40 data samples and that QPSK modulation is used. To obtain the achievable throughput, we first determine through simulations the average number of pilots required by each algorithm to achieve a certain SER performance at a given SNR. The corresponding number of data symbols that can be transmitted is then divided by 40 to obtain the throughput. As we can see from Table I, substantial improvements in throughput (up to 27%) are possible by using the EM algorithm. Similar results can be obtained for the nonreciprocal case.

6.6 Conclusions

In this chapter, we derived the EM algorithm for semi-blind channel estimation assuming flat-fading channel conditions. This was done for both nonreciprocal and reciprocal channels. In both cases, our simulations showed that the derived EM algorithm outperforms the pilot-based LS estimator even with a limited number of data symbols and performs very close to the corresponding semi-blind CRB. The proposed algorithms require only a small number of low-complexity iterations to converge. Hence, the accuracy gains of semi-blind estimation predicted in Chapter 5 through CRB analysis can indeed be realized at an affordable computational price. Finally, by virtue of its higher accuracy the EM algorithm requires a smaller number of pilots compared to the LS estimator, which allows for transmitting more data symbols, resulting in a significant improvement in throughput and spectral efficiency. Equivalently, for the same number of pilots the EM algorithm has a superior SER performance compared to the LS estimator.

SNR (dB)	SER	Throughput (EM/LS)
10	3×10^{-1}	0.88 / 0.78
15	1.5×10^{-1}	0.8 / 0.68
20	6×10^{-2}	0.8 / 0.63
25	1.5×10^{-2}	$0.85 \ / \ 0.58$
30	2×10^{-3}	$0.95 \ / \ 0.73$

Table 6.1Data throughput for the EM and LS estimators.



Fig. 6.1 MSE performance of the EM algorithm for nonreciprocal channels along with the corresponding semi-blind CRBs versus SNR (N = 32, L = 8, 4 EM iterations).



Fig. 6.2 MSE performance of the EM algorithm for reciprocal channels along with the corresponding semi-blind CRBs plotted versus SNR (N = 32 and L = 8, 4 EM iterations).



Fig. 6.3 MSE performance of the EM algorithm for nonreciprocal channels along with the corresponding semi-blind CRBs versus N, normalized w.r.t. the MSE of the pilot-based LS estimator (L = 8, SNR 15dB, 10 EM iterations).



Fig. 6.4 MSE performance of the EM algorithm for reciprocal channels along with the corresponding semi-blind CRBs plotted versus N, normalized w.r.t. the MSE of the pilot-based LS estimator (L = 8, SNR 15dB, 10 EM iterations).



Fig. 6.5 MSE performance of the EM algorithm for nonreciprocal channels plotted versus the number of EM iterations (N = 32, 100, L = 8, SNR 15dB).



Fig. 6.6 MSE performance of the EM algorithm for reciprocal channels plotted versus the number of EM iterations (N = 32, 100, L = 8, SNR 15dB).



Fig. 6.7 SER performance of the EM algorithm and the LS estimator plotted versus SNR (N = 36, L = 4, 6 EM iterations).

Chapter 7

Semi-blind Channel Estimation for OFDM-modulated AF TWRNs

7.1 Introduction

In Chapters 3–6, we developed blind and semi-blind channel estimation algorithms for AF TWRNs under flat-fading conditions and showed that the proposed methods can provide substantially better tradeoffs between accuracy and spectral efficiency than the conventional training-based methods. Motivated by our results for flat-fading channels, in this chapter we consider semi-blind channel estimation for AF TWRNs in frequency-selective channels. To avoid the detrimental effects of ISI, we assume that OFDM transmission is employed (see Chapter 2). Previous works on channel estimation for OFDM-based AF TWRNs have assumed channel reciprocity [22,34,91,92], which reduces the number of channel parameters that need to be estimated. In contrast, we will consider nonreciprocal channels as this is a more realistic assumption due to the sensitivity of OFDM systems to RF front end imperfections that ruin channel reciprocity [24,26].

We assume that each terminal transmits a single OFDM pilot block followed by a number of OFDM data blocks. As done in [22, 32, 34], we focus on the estimation of the individual channels, rather than the cascaded channels. Although the cascaded channels are sufficient for detection, the individual channels are needed in other applications such as beamforming [93]. Estimating the individual channels also makes it possible to use the information in the covariance matrix of the received signal as this matrix cannot be expressed entirely in terms of the cascaded channels. To assist in the estimation of the

individual channels, we adopt a superimposed training strategy at the relay during the broadcast of the OFDM pilot blocks, as done in [34,91]. More specifically, the relay superimposes its own pilot symbols over the received OFDM pilot block before broadcasting it. Our proposed semi-blind estimator is based on the Gaussian ML (GML) approach, i.e., we treat the transmitted data symbols as Gaussian distributed. The resulting semi-blind GML channel estimator reduces to a nonlinear minimization problem, which we solve numerically using an iterative quasi-Newton method. In addition, we derive sufficient conditions for the optimality of the three pilot vectors (the two at the terminals and the one at the relay) and provide an example of pilot design that satisfies these conditions. Furthermore, we derive the semi-blind and pilot-based CRBs which serve as benchmarks on estimation performance. Using simulation results, we show that the proposed semi-blind estimator closely approaches the semi-blind CRB and provides substantial improvements in accuracy over the pilot-based approach while using only a limited number of OFDM data blocks. This improvement in performance also holds when the data symbols are drawn from discrete constellations. The proposed semi-blind method may also be used to estimate the cascaded channels which can be obtained by linear convolution after estimating the individual channels. We show in our simulations that estimating the individual channels semi-blindly first and then acquiring from them estimates of the cascaded channels provides superior accuracy over pilot-based estimation of the cascaded channels. Finally, we also show that semi-blind estimation results in improved SER performance compared to pilot-based estimation.

The rest of this chapter is organized as follows. In Section 7.2 we present the system model. The proposed semi-blind channel estimation algorithm and the pilot design are presented in Section 7.3. The semi-blind and pilot-based CRBs are derived in Section 7.4. Simulation results are presented in Section 7.5. Finally, our conclusions are discussed in Section 7.6.

7.2 System Model

We consider a half-duplex AF TWRN with two source nodes, \mathcal{T}_1 and \mathcal{T}_2 , and a single relaying node \mathcal{R} . The network operates in frequency selective channel conditions. To compensate for the effects of the frequency selective fading, OFDM transmission with N



Fig. 7.1 Structure of the OFDM frame transmitted by the terminals.

subcarriers¹ is employed. Each round of data exchange between \mathcal{T}_1 and \mathcal{T}_2 consists of two phases. In the first phase, the two terminals simultaneously transmit an OFDM frame to \mathcal{R} . In the second phase, \mathcal{R} broadcasts an amplified version of the received frame to both terminals.

7.2.1 Transmission at the Terminals

Each OFDM frame transmitted by the terminals is composed of one pilot block and K data blocks. Furthermore, each OFDM block (pilot or data) consists of N time-domain symbols and a cyclic prefix (CP) of appropriate length which is inserted to avoid interblock interference (see Fig. 7.1). We denote by $\tilde{\mathbf{t}}_1 = [\tilde{t}_{11}, \ldots, \tilde{t}_{1N}]^T$ and $\tilde{\mathbf{t}}_2 = [\tilde{t}_{21}, \ldots, \tilde{t}_{2N}]^T$ the $N \times 1$ frequency-domain pilot symbol vectors of \mathcal{T}_1 and \mathcal{T}_2 and by $\tilde{\mathbf{s}}_{1k}$ and $\tilde{\mathbf{s}}_{2k}$, $k = 1, \ldots, K$, the $N \times 1$ frequency-domain data symbol vectors of \mathcal{T}_1 and \mathcal{T}_2 , respectively. The corresponding time-domain pilot and data symbol vectors are $\mathbf{t}_1 = \mathbf{F}^H \tilde{\mathbf{t}}_1$, $\mathbf{t}_2 = \mathbf{F}^H \tilde{\mathbf{t}}_2$, $\mathbf{s}_{1k} = \mathbf{F}^H \tilde{\mathbf{s}}_{1k}$ and $\mathbf{s}_{2k} = \mathbf{F}^H \tilde{\mathbf{s}}_{2k}$, $k = 1, \ldots, K$, where \mathbf{F} is the $N \times N$ normalized discrete Fourier transform (DFT) matrix whose (p,q)th entry is $1/\sqrt{N}e^{-j2\pi(p-1)(q-1)/N}$. Moreover, we assume that the average transmission powers of \mathcal{T}_1 and \mathcal{T}_2 during pilot transmission are P_1 and P_2 , respectively, i.e., $\tilde{\mathbf{t}}_1^H \tilde{\mathbf{t}}_1 = NP_1$ and $\tilde{\mathbf{t}}_2^H \tilde{\mathbf{t}}_2 = NP_2$. For simplicity, we also assume that the same average transmission powers are employed by the terminals for data transmission, i.e., $\mathbb{E} \left\{ \mathbf{s}_{1k}^H \mathbf{s}_{1k} \right\} = NP_1$ and $\mathbb{E} \left\{ \mathbf{s}_{2k}^H \mathbf{s}_{2k} \right\} = NP_2$.

We denote by $\boldsymbol{h}_1 \triangleq [h_{11}, \ldots, h_{1L_1}]^T$ the L_1 -tap baseband channel from \mathcal{T}_1 to \mathcal{R} and by $\boldsymbol{g}_1 \triangleq [g_{11}, \ldots, g_{1J_1}]^T$ the J_1 -tap baseband channels from \mathcal{T}_2 to \mathcal{R} . The elements of each

¹We note that N is used in this chapter to denote the number of subcarriers, whereas in the previous chapters it was used to denote the number of data samples.



Fig. 7.2 OFDM-based two-way relay network with superimposed training at the relay.

channel vector are modelled as independent circular complex Gaussian random variables with mean zero and whose variance follows the exponential power decay profile of [94]. More specifically, the variance of the ℓ th channel tap is $\sigma_{\ell}^2 = e^{-(\ell-1)/10}$, where $\ell = 1, \ldots, L_1$ for \mathbf{h}_1 and $\ell = 1, \ldots, J_1$ for \mathbf{g}_1 . We let $\sigma_{h1}^2 = \sum_{\ell=1}^{L_1} \sigma_{\ell}^2$ and $\sigma_{g1}^2 = \sum_{\ell=1}^{J_1} \sigma_{\ell}^2$. Moreover, we assume quasi-static channel conditions, such that the channels \mathbf{h}_1 and \mathbf{g}_1 are fixed for the frame duration, which includes 1 pilot block and K data blocks, but may vary between consecutive frames.

7.2.2 Processing at the Relay

The received vector at the relay corresponding to the transmitted pilot blocks after CP removal is given by

$$r = H_1 t_1 + G_1 t_2 + n,$$
 (7.1)

where \boldsymbol{H}_1 and \boldsymbol{G}_1 are $N \times N$ circulant matrices with first columns $[\boldsymbol{h}_1^T, \boldsymbol{0}_{1 \times (N-L_1)}]^T$ and $[\boldsymbol{g}_1^T, \boldsymbol{0}_{1 \times (N-J_1)}]^T$, respectively, and \boldsymbol{n} is the circular complex white Gaussian noise vector with mean zero and covariance² $\sigma^2 \boldsymbol{I}_N$, denoted as $\mathcal{CCN}(0, \sigma^2 \boldsymbol{I}_N)$. Similarly, the received vectors corresponding to the K data blocks are

$$\boldsymbol{r}_{k} = \boldsymbol{H}_{1}\boldsymbol{s}_{1k} + \boldsymbol{G}_{1}\boldsymbol{s}_{2k} + \boldsymbol{n}_{k},$$
 (7.2)

104

 $^{{}^{2}\}boldsymbol{I}_{N}$ denotes the $N \times N$ identity matrix.



Fig. 7.3 Illustration of superimposed training at the relay.

where $\boldsymbol{n}_1, \ldots, \boldsymbol{n}_K$ are i.i.d. $\mathcal{CCN}(0, \sigma^2 \boldsymbol{I}_N)$.

The relay amplifies the received pilot block using an amplification factor $A_p > 0$, and then, to assist channel estimation at the terminals, superimposes over the amplified vector $A_p \mathbf{r}$ a time-domain pilot vector $\mathbf{t}_3 = \mathbf{F}^H \tilde{\mathbf{t}}_3$, where $\tilde{\mathbf{t}}_3 = [\tilde{p}_{31}, \ldots, \tilde{p}_{3N}]^T$ is the corresponding frequency-domain vector [34]. The resulting signal vector is given by

$$\breve{\boldsymbol{r}} = A_p \boldsymbol{r} + \boldsymbol{t}_3 = A_p \boldsymbol{H}_1 \boldsymbol{t}_1 + A_p \boldsymbol{G}_1 \boldsymbol{t}_2 + A_p \boldsymbol{n} + \boldsymbol{t}_3.$$
(7.3)

The average transmission power of the relay over the long term (i.e., over many OFDM frames) is set at P_r . In other words, $\mathbb{E}\left\{\breve{r}^H\breve{r}\right\} = NP_r$, where the expectation takes into account both the channel statistics and the noise statistics. This power is divided between the amplified signal vector $A_p \mathbf{r}$ and the superimposed vector \mathbf{t}_3 as follows: $P_3 = \alpha P_r$ is allocated to the superimposed pilot and $(1-\alpha)P_r$ is allocated to the amplified term, where $0 < \alpha < 1$. Using the statistics of the time-domain channels \mathbf{h}_1 and \mathbf{g}_1 , it can be shown that the average power of the received signal at the relay is

$$\frac{1}{N}\mathbb{E}\left\{\boldsymbol{r}^{H}\boldsymbol{r}\right\} = \sigma_{h1}^{2}P_{1} + \sigma_{g1}^{2}P_{2} + \sigma^{2}.$$
(7.4)

Hence, to allocate $(1 - \alpha)P_r$ to the amplified signal, the amplification factor should be set as

$$A_p = \sqrt{\frac{(1-\alpha)P_r}{\sigma_{h1}^2 P_1 + \sigma_{g1}^2 P_2 + \sigma^2}}.$$
(7.5)

The relay also amplifies the received data vectors \mathbf{r}_k , using the amplification factor $A_d > 0$, but without superimposing a pilot. In this case, the relay can maintain an average

transmission power of P_r when broadcasting the amplified data vectors by setting A_d as³

$$A_d = \sqrt{\frac{P_r}{\sigma_{h1}^2 P_1 + \sigma_{g1}^2 P_2 + \sigma^2}}.$$
(7.6)

The processing at the relay, including superimposed training, is illustrated in Fig. 7.3. Before broadcasting the frame that contains the amplified pilot and information-bearing vectors, the relay inserts a CP into each block in the frame.

7.2.3 Received Vectors at the Terminal

Assuming nonreciprocal channels, the channel from each terminal to the relay is different from the channel from the relay back to the terminal. Without loss of generality, we focus on channel estimation at \mathcal{T}_1 . The baseband channel from \mathcal{R} back to \mathcal{T}_1 is denoted by $\boldsymbol{h}_2 \triangleq [h_{21}, \ldots, h_{2L_2}]^T$. The elements of \boldsymbol{h}_2 are modelled in the same way as those of \boldsymbol{h}_1 and \boldsymbol{g}_1 and are assumed fixed for the frame duration. We let $\sigma_{h_2}^2 = \sum_{\ell=1}^{L_2} \sigma_{\ell}^2$. The pilot-bearing received signal block at \mathcal{T}_1 after CP removal is given by

$$y = A_p H_2 H_1 t_1 + A_p H_2 G_1 t_2 + H_2 t_3 + A_p H_2 n + w,$$
(7.7)

where \boldsymbol{H}_2 is the $N \times N$ circulant matrix with first column $[\boldsymbol{h}_2^T, \boldsymbol{0}_{1 \times (N-L_2)}]^T$, and the noise vector \boldsymbol{w} is $\mathcal{CCN}(0, \sigma^2 \boldsymbol{I}_N)$. Similarly, the K data-bearing received signal blocks are given by

$$\boldsymbol{z}_{k} = A_{d}\boldsymbol{H}_{2}\boldsymbol{H}_{1}\boldsymbol{s}_{1k} + A_{d}\boldsymbol{H}_{2}\boldsymbol{G}_{1}\boldsymbol{s}_{2k} + A_{d}\boldsymbol{H}_{2}\boldsymbol{n}_{k} + \boldsymbol{w}_{k}, \qquad (7.8)$$

where $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$ are also $\mathcal{CCN}(0, \sigma^2 \boldsymbol{I}_N)$, with the difference that they do not contain the signal component corresponding to superimposed training.

The goal of our work is to accurately estimate the individual channel vectors h_1 , h_2 and g_1 . We propose to estimate these vectors using a semi-blind approach which provides enhanced estimation performance by incorporating both the pilot-bearing vector y and the information-bearing vectors z_1, \ldots, z_K into the estimation process.

³We note that the use of different amplification factors during pilot and data transmission is necessary to satisfy the power constraint at the relay since superimposed training is only used during pilot transmission.

7.3 Semi-blind Channel Estimation

In this section, we present the proposed semi-blind channel estimation algorithm. For comparison purposes, we also consider fully pilot-based estimation and derive the corresponding pilot-based least-squares (LS) channel estimator.

7.3.1 Proposed Algorithm

The unknown channel parameters to be estimated are collected into the vector $\boldsymbol{\theta} \triangleq [\boldsymbol{h}_1^T, \boldsymbol{h}_2^T, \boldsymbol{g}_1^T]^T$. The first step towards the development of the semi-blind estimator is the derivation of the joint likelihood function of the vectors $\boldsymbol{y}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$. This likelihood function depends on the specific constellation from which the frequency-domain data symbol vectors $\tilde{\boldsymbol{s}}_{21}, \ldots, \tilde{\boldsymbol{s}}_{2K}$ are drawn. Unfortunately, taking into account the discrete statistics of these vectors would result in an intractable likelihood function. Instead, we obtain a tractable likelihood function by resorting to the Gaussian approximation, i.e., by modelling the data vectors $\tilde{\boldsymbol{s}}_{21}, \ldots, \tilde{\boldsymbol{s}}_{2K}$ as i.i.d. $\mathcal{CCN}(0, P_2 \boldsymbol{I}_N)$. Nonetheless, we will show in Section 7.5 that the impact of this approximation on the performance of the proposed estimator is minimal and that the proposed estimator performs well when the data symbol vectors are drawn from discrete constellations.

Under the Gaussian assumption, the joint likelihood can be expressed in terms of the first and second order statistics of the vectors $\boldsymbol{y}, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$. Let us denote by $\boldsymbol{\mu}$ and \boldsymbol{C} the mean and covariance matrix of \boldsymbol{y} , respectively. From (7.7), we see that

$$\boldsymbol{\mu} = \mathbb{E}\left\{\boldsymbol{z}\right\} = A_p \boldsymbol{H}_2 \boldsymbol{H}_1 \boldsymbol{t}_1 + A_p \boldsymbol{H}_2 \boldsymbol{G}_1 \boldsymbol{t}_2 + \boldsymbol{H}_2 \boldsymbol{t}_3, \tag{7.9}$$

and

$$\boldsymbol{C} = A_p^2 \sigma^2 \boldsymbol{H}_2 \boldsymbol{H}_2^H + \sigma^2 \boldsymbol{I}_N.$$
(7.10)

Furthermore, let us denote by μ_k the mean of \boldsymbol{z}_k and by \boldsymbol{Q} the corresponding covariance matrix. From (7.8) we obtain

$$\boldsymbol{\mu}_{k} = \mathbb{E}\left\{\boldsymbol{z}_{k}\right\} = A_{d}\boldsymbol{H}_{2}\boldsymbol{H}_{1}\boldsymbol{s}_{1k}, \qquad (7.11)$$

and

$$\boldsymbol{Q} = A_d^2 P_2 \boldsymbol{H}_2 \boldsymbol{G}_1 \boldsymbol{G}_1^H \boldsymbol{H}_2^H + A_d^2 \sigma^2 \boldsymbol{H}_2 \boldsymbol{H}_2^H + \sigma^2 \boldsymbol{I}_N.$$
(7.12)

Collecting the data-bearing vectors into a single vector $\breve{z} \triangleq [z_1^T, \ldots, z_K^T]^T$, the joint likelihood function of \boldsymbol{y} and \breve{z} is given by⁴

$$f(\boldsymbol{y}, \boldsymbol{\breve{z}}; \boldsymbol{\theta}) = \frac{1}{\pi^N |\boldsymbol{C}|} e^{-(\boldsymbol{y}-\boldsymbol{\mu})^H \boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})} \prod_{k=1}^K \frac{1}{\pi^N |\boldsymbol{Q}|} e^{-(\boldsymbol{z}_k - \boldsymbol{\mu}_k)^H \boldsymbol{Q}^{-1}(\boldsymbol{z}_k - \boldsymbol{\mu}_k)}.$$
 (7.13)

Hence, the corresponding joint log-likelihood function is

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{\breve{z}}; \boldsymbol{\theta}) = -(K+1)N\log\pi - \log|\boldsymbol{C}| - K\log|\boldsymbol{Q}| - (\boldsymbol{y} - \boldsymbol{\mu})^{H}\boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) - \sum_{k=1}^{K} (\boldsymbol{z}_{k} - \boldsymbol{\mu}_{k})^{H}\boldsymbol{Q}^{-1}(\boldsymbol{z}_{k} - \boldsymbol{\mu}_{k}).$$
(7.14)

Therefore, the semi-blind GML estimates of $\boldsymbol{h}_1,\,\boldsymbol{h}_2$ and \boldsymbol{g}_1 are

$$\{\hat{\boldsymbol{h}}_{1}^{(s)}, \hat{\boldsymbol{h}}_{2}^{(s)}, \hat{\boldsymbol{g}}_{1}^{(s)}\} = \underset{\boldsymbol{h}_{1}, \boldsymbol{h}_{2}, \boldsymbol{g}_{1}}{\arg\min} \log |\boldsymbol{C}| + K \log |\boldsymbol{Q}| + (\boldsymbol{y} - \boldsymbol{\mu})^{H} \boldsymbol{C}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) + \sum_{k=1}^{K} (\boldsymbol{z}_{k} - \boldsymbol{\mu}_{k})^{H} \boldsymbol{Q}^{-1} (\boldsymbol{z}_{k} - \boldsymbol{\mu}_{k}).$$
(7.15)

The objective function in (7.15) is nonconvex, and the solution for the minimization problem may be obtained using standard numerical techniques [75]. In this work, we use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [75], which is the most popular quasi-Newton method and is known for its robustness and efficiency. It avoids the computation of the Hessian matrix and requires no matrix inversion. To initialize the BFGS algorithm, we will use the channel estimates provided by the pilot-based LS estimator, which we will derive next. Backtracking linesearch is used to find the step size at each iteration [74]. While the BFGS method is known to work well in practice, convergence to the global minimum is not guaranteed since the objective function is nonconvex. In addition, convergence to a local minimum requires a sophisticated linesearch⁵ to be theoretically guaranteed. Nonetheless, as we shall see in Section 7.5, BFGS with backtracking linesearch coupled with using the pilot-based LS estimator for initialization results in performance that is very close to the semi-blind CRB, which indicates that convergence to the global minimum occurs most of

 $^{{}^{4}|\}boldsymbol{A}|$ denotes the determinant of \boldsymbol{A} .

⁵To guarantee local convergence for the BFGS method, the linesearch has to satisfy the Wolfe conditions [75]. Nonetheless, conventional backtracking linesearch is often used.

the time.

7.3.2 Pilot-based Least Squares Estimation

In [34], the LS estimator was derived for pilot-based channel estimation with superimposed training in the context of reciprocal channels. In that case, $\mathbf{h}_1 = \mathbf{h}_2$ and $\mathbf{g}_1 = \mathbf{g}_2$, i.e., only two channel vectors need to be estimated. In our work, we consider the case of nonreciprocal channels where all three channel vectors \mathbf{h}_1 , \mathbf{h}_2 and \mathbf{g}_1 have to be estimated. Due to the structure of the received pilot signal block (see (7.7)), the conventional LS approach cannot be applied directly to estimate the individual channel vectors. Instead, similar to what is done in [34], we adopt a two-step approach whereby we first obtain LS estimates of the cascaded time-domain channels⁶ $\mathbf{a} \triangleq \mathbf{h}_1 * \mathbf{h}_2$ and $\mathbf{b} \triangleq \mathbf{g}_1 * \mathbf{h}_2$ and the individual channel \mathbf{h}_2 , and then extract from them estimates of the individual channels \mathbf{h}_1 and \mathbf{g}_1 . We let $M_1 = L_1 + L_2 - 1$ and $M_2 = J_1 + L_2 - 1$ be the lengths of \mathbf{a} and \mathbf{b} , respectively. Eq. (7.7) can be rewritten in terms of the cascaded channels \mathbf{a} and \mathbf{b} as

$$\boldsymbol{y} = A_p \boldsymbol{S}_1 \boldsymbol{a} + A_p \boldsymbol{S}_2 \boldsymbol{b} + \boldsymbol{S}_3 \boldsymbol{h}_2 + A_p \boldsymbol{H}_2 \boldsymbol{n} + \boldsymbol{w}$$
$$= \underbrace{\begin{bmatrix} A_p \boldsymbol{S}_1 & A_p \boldsymbol{S}_2 & \boldsymbol{S}_3 \end{bmatrix}}_{\Omega} \underbrace{\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{h}_2 \end{bmatrix}}_{\boldsymbol{q}} + A_p \boldsymbol{H}_2 \boldsymbol{n} + \boldsymbol{w}. \tag{7.16}$$

where S_1 is the $N \times M_1$ circulant matrix with first column t_1 , S_2 is the $N \times M_2$ circulant matrix with first column t_2 and S_3 is the $N \times L_2$ circulant matrix with first column t_3 . Assuming that $N \ge M_1 + M_2 + L_2$, \boldsymbol{q} may be estimated using the LS approach by⁷

$$\hat{\boldsymbol{q}} = \boldsymbol{\Omega}^{\dagger} \boldsymbol{y}. \tag{7.17}$$

The estimates \hat{a} , \hat{b} and \hat{h}_2 of a, b and h_2 are obtained from the corresponding entries of \hat{q} . To obtain the estimates for the individual channels h_1 and g_1 , we note that the cascaded channels a and b may be expressed as

$$\boldsymbol{a} = \boldsymbol{B}_1 \boldsymbol{h}_1, \qquad \boldsymbol{b} = \boldsymbol{B}_2 \boldsymbol{g}_1 \tag{7.18}$$

⁶We use x * y to denote the linear convolution between the vectors x and y.

 $^{{}^{7}}A^{\dagger}$ denotes the Moore Penrose pseudo-inverse of A.

where \boldsymbol{B}_1 is the $M_1 \times L_1$ circulant matrix with first column $[\boldsymbol{h}_2^T, \boldsymbol{0}_{1 \times (L_1-1)}]^T$ and \boldsymbol{B}_2 is the $M_2 \times J_1$ circulant matrix with first column $[\boldsymbol{h}_2^T, \boldsymbol{0}_{1 \times (J_1-1)}]^T$. We may thus estimate \boldsymbol{h}_1 and \boldsymbol{g}_1 as follows:

$$\hat{\boldsymbol{h}}_1 = \left(\hat{\boldsymbol{B}}_1\right)^{\dagger} \hat{\boldsymbol{a}}, \quad \hat{\boldsymbol{g}}_1 = \left(\hat{\boldsymbol{B}}_2\right)^{\dagger} \hat{\boldsymbol{b}}, \tag{7.19}$$

where \hat{B}_1 and \hat{B}_2 are formed using the estimate \hat{h}_2 . The LS estimator will be also be used as a reference to demonstrate the performance enhancement provided by the semi-blind approach.

7.3.3 Design of the pilot vectors

A convenient criterion for designing the pilot vectors \tilde{t}_1 , \tilde{t}_2 and \tilde{t}_3 is to choose them such that they minimize the mean-squared error (MSE) of the LS estimator of q. This criterion is attractive because the MSE for the estimation of q has a simple closed-form expression, unlike the MSE for the estimation of the individual channels. The MSE for the LS estimation of q is given by

$$\mathbb{E}\left\{\|\hat{\boldsymbol{q}}-\boldsymbol{q}\|^{2}\right\} = \sigma^{2}(1+A^{2}\sigma_{h2}^{2})\mathrm{tr}\left((\boldsymbol{\Omega}^{H}\boldsymbol{\Omega})^{-1}\right),\tag{7.20}$$

where the expectation takes into account both the noise and the channel statistics. The problem looks similar to pilot design for channel estimation in 3×1 MISO systems. However, the approach used in [95] cannot be directly applied since the two terminals and the relay have their individual powers. A similar type of minimization was considered in the context of OFDM-based TWRNs in [22], but it involved the design of only two pilot vectors since superimposed training was not employed. The following theorem describes sufficient conditions for the pilot vectors \tilde{t}_1 , \tilde{t}_2 and \tilde{t}_3 to be optimal.

Theorem 3. For fixed P_1 , P_2 and P_3 , the MSE in (7.20) is minimized by any three training vectors \tilde{t}_1 , \tilde{t}_2 and \tilde{t}_3 that satisfy the following conditions:

1.
$$|\tilde{p}_{ji}|^2 = P_j \text{ for } j = 1, 2, 3, i = 1, \dots, N.$$

2. $\sum_{i=1}^N \tilde{p}_{1i}^* \tilde{p}_{2i} e^{j2\pi(i-1)m_1/N} = 0, \forall m_1 \in \{1 - M_2, \dots, M_1 - 1\},$
3. $\sum_{i=1}^N \tilde{p}_{1i}^* \tilde{p}_{3i} e^{j2\pi(i-1)m_2/N} = 0, \forall m_2 \in \{1 - L_2, \dots, M_1 - 1\},$
4. $\sum_{i=1}^N \tilde{p}_{2i}^* \tilde{p}_{3i} e^{j2\pi(i-1)m_3/N} = 0, \forall m_3 \in \{1 - L_2, \dots, M_2 - 1\}.$

Proof. See Appendix D.2.

The above theorem can be seen as a generalization of Theorem 3.5 in [22] to the case of 3 pilot vectors. The following is an example of optimal training vectors, inspired by the pilot design in [22]:

$$\tilde{p}_{1i} = \sqrt{P_1}, \ \tilde{p}_{2i} = \sqrt{P_2} e^{j2\pi(i-1)\kappa/N}, \ \tilde{p}_{3i} = \sqrt{P_3} e^{j2\pi(i-1)\delta/N}$$
(7.21)

for any values κ , δ that satisfy

$$\kappa \in \{M_2, \dots, N - M_1\}, \delta \in \{L_2, \dots, N - M_1\}$$

$$\kappa - \delta \in \{L_2, \dots, N - M_2\}.$$
(7.22)

For the existence of integer values κ and δ that simultaneously satisfy the above constraints, the number of subcarriers N should satisfy

$$N \ge \max(M_1 + M_2, M_1 + L_2), \tag{7.23}$$

and

$$\max(L_2, M_1 + M_2 - N) \le \min(N - M_2, N - M_1 - L_2).$$
(7.24)

We note that pilot-design for OFDM-based AF TWRNS with superimposed training was also considered in another recent work [96]. However in [96] the pilots are designed to minimize the Bayesian CRB, which results in different optimality conditions from those in Theorem 3. In particular, only a single orthogonality condition is required in [96], $S_1^H S_2 = 0$, whereas optimality in our case also requires $S_1^H S_3 = 0$ and $S_2^H S_3 = 0$. Hence, pilots that satisfy the conditions of Theorem 3 (as the ones used in the simulations section) not only minimize the MSE of the LS estimator but also the Bayesian CRB.

7.4 Cramer-Rao Bounds

In this section, we will derive the CRBs for the estimation of the channel parameters h_1 , h_2 and g_1 for both the semi-blind and the pilot-based scenarios. The joint log-likelihood in (7.14) may be decomposed into the sum of two terms

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{\breve{z}}; \boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{y}; \boldsymbol{\theta}) + \mathcal{L}(\boldsymbol{\breve{z}}; \boldsymbol{\theta}), \qquad (7.25)$$

where

$$\mathcal{L}(\boldsymbol{y};\boldsymbol{\theta}) = -N\log\pi - \log|\boldsymbol{C}| - (\boldsymbol{y} - \boldsymbol{\mu})^{H}\boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$$
(7.26)

is the log-likelihood function for the pilot-bearing received vector and

$$\mathcal{L}(\breve{\boldsymbol{z}};\boldsymbol{\theta}) = -KN\log\pi - K\log|\boldsymbol{Q}| - \sum_{k=1}^{K} (\boldsymbol{z}_k - \boldsymbol{\mu}_k)^H \boldsymbol{Q}^{-1} (\boldsymbol{z}_k - \boldsymbol{\mu}_k)$$
(7.27)

is the log-likelihood function for the K data-bearing received vectors.

7.4.1 The CRB for Pilot-based Estimation

The Fisher information matrix (FIM) for fully pilot-based estimation is defined as

$$\Gamma_p = \mathbb{E}\left\{\frac{\partial \mathcal{L}(\boldsymbol{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \frac{\partial \mathcal{L}(\boldsymbol{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right\}.$$
(7.28)

From [35], we know that the (i, j)th element of Γ_p is given by⁸

$$[\boldsymbol{\Gamma}_{P}]_{ij} = \frac{\partial \boldsymbol{\mu}^{H}}{\partial \theta_{i}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_{j}} + \operatorname{tr} \left(\boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial \theta_{i}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial \theta_{j}} \right).$$
(7.29)

Hence, the FIM may be expressed as

$$\Gamma_p = \Lambda_p + \Sigma_p, \tag{7.30}$$

where

$$\boldsymbol{\Lambda}_{p} = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{h}_{1}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_{1}^{T}} & \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{h}_{1}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_{2}^{T}} & \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{h}_{1}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{g}_{1}^{T}} \\ \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{h}_{2}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_{1}^{T}} & \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{h}_{2}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_{2}^{T}} & \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{h}_{2}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{g}_{1}^{T}} \\ \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{g}_{1}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_{1}^{T}} & \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{g}_{1}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_{2}^{T}} & \frac{\partial \boldsymbol{\mu}^{H}}{\partial \boldsymbol{g}_{1}^{*}} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{g}_{1}^{T}} \end{bmatrix},$$
(7.31)

and

$$[\mathbf{\Sigma}_{p}]_{ij} = \operatorname{tr}\left(\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_{i}^{*}}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_{j}}\right).$$
(7.32)

Moreover, we have that $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_1^T} = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}}{\partial h_{11}}, \dots, \frac{\partial \boldsymbol{\mu}}{\partial h_{1L_1}} \end{bmatrix}$, $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{h}_2^T} = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}}{\partial h_{21}}, \dots, \frac{\partial \boldsymbol{\mu}}{\partial h_{2L_2}} \end{bmatrix}$ and $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{g}_1^T} = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}}{\partial g_{11}}, \dots, \frac{\partial \boldsymbol{\mu}}{\partial g_{1J_1}} \end{bmatrix}$. Denoting by $\boldsymbol{\Upsilon}_{(m)}(\boldsymbol{x})$ the $N \times m$ circulant matrix with first column \boldsymbol{x} , we

⁸ $[\mathbf{A}]_{ij}$ refers to the (i, j)th element of the matrix \mathbf{A} .

may evaluate the terms $\frac{\partial \mu}{\partial h_1^T}$, $\frac{\partial \mu}{\partial h_2^T}$ and $\frac{\partial \mu}{\partial g_1^T}$ using the following properties [92]:

Using the above, we obtain

$$\frac{\partial \boldsymbol{\mu}}{\boldsymbol{h}_1^T} = A_p \boldsymbol{H}_2 \boldsymbol{\Upsilon}_{L_1}(\boldsymbol{t}_1), \qquad (7.33)$$

$$\frac{\partial \boldsymbol{\mu}}{\boldsymbol{h}_2^T} = A_p \boldsymbol{H}_1 \boldsymbol{\Upsilon}_{L_2}(\boldsymbol{t}_1) + A_p \boldsymbol{G}_1 \boldsymbol{\Upsilon}_{L_2}(\boldsymbol{t}_2) + \boldsymbol{\Upsilon}_{L_2}(\boldsymbol{t}_3), \qquad (7.34)$$

and

$$\frac{\partial \boldsymbol{\mu}}{\boldsymbol{g}_1^T} = A_p \boldsymbol{H}_2 \boldsymbol{\Upsilon}_{J_1}(\boldsymbol{t}_2). \tag{7.35}$$

In order to evaluate the terms in (7.32), we can use the expression for C in (7.10) to show that

$$\frac{\partial \boldsymbol{C}}{\partial h_{1i}} = \boldsymbol{0}_{N \times N}, \quad \frac{\partial \boldsymbol{C}}{\partial g_{1i}} = \boldsymbol{0}_{N \times N}, \quad \frac{\partial \boldsymbol{C}}{\partial h_{2i}} = A^2 \sigma^2 \boldsymbol{E}_i \boldsymbol{H}_2^H, \quad (7.36)$$

where \boldsymbol{E}_i is the $N \times N$ circulant matrix with first column⁹ \boldsymbol{e}_i .

Finally, the pilot-based CRB is given by $CRB_{\theta}^{(p)} = \operatorname{tr}(\boldsymbol{\Gamma}_p^{-1})$. In particular,

$$CRB_{\boldsymbol{h}_{1}}^{(p)} = \sum_{i=1}^{L_{1}} [\boldsymbol{\Gamma}_{p}^{-1}]_{ii}, \quad CRB_{\boldsymbol{h}_{2}}^{(p)} = \sum_{i=L_{1}+1}^{L_{1}+L_{2}} [\boldsymbol{\Gamma}_{p}^{-1}]_{ii}, \quad (7.37)$$

and

$$CRB_{g_1} = \sum_{i=L_1+L_2+1}^{L_1+L_2+J_1} [\Gamma_p^{-1}]_{ii}.$$
(7.38)

7.4.2 The CRB for Semi-blind Estimation

The FIM for semi-blind estimation is given by

$$\Gamma_s = \Gamma_p + \Gamma_d \tag{7.39}$$

⁹The vector e_i is the $N \times 1$ basis vector with the *i*th element 1 and the remaining elements 0.

where

$$\Gamma_d = \mathbb{E}\left\{\frac{\partial \mathcal{L}(\check{\boldsymbol{z}};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \frac{\partial \mathcal{L}(\check{\boldsymbol{z}};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right\}.$$
(7.40)

We can see from (7.27) that Γ_d may be expressed as

$$\Gamma_d = \sum_{k=1}^{K} \Lambda_d^{(k)} + K \Sigma_d, \qquad (7.41)$$

where

$$\boldsymbol{\Lambda}_{d}^{(k)} = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{h}_{1}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{h}_{1}^{T}} & \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{h}_{1}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{h}_{2}^{T}} & \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{h}_{1}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{g}_{1}^{T}} \\ \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{h}_{2}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{h}_{1}^{T}} & \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{h}_{2}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{h}_{2}^{*}} & \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{h}_{2}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{g}_{1}^{*}} \\ \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{g}_{1}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{h}_{1}^{T}} & \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{g}_{1}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{h}_{2}^{*}} & \frac{\partial \boldsymbol{\mu}_{k}^{H}}{\partial \boldsymbol{g}_{1}^{*}} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{\mu}_{k}}{\partial \boldsymbol{g}_{1}^{*}} \end{bmatrix},$$
(7.42)

and

$$[\boldsymbol{\Sigma}_d]_{ij} = \operatorname{tr} \left(\boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{Q}}{\partial \theta_i^*} \boldsymbol{Q}^{-1} \frac{\partial \boldsymbol{Q}}{\partial \theta_j} \right).$$
(7.43)

Moreover, since $\boldsymbol{\mu}_k = A_d \boldsymbol{H}_2 \boldsymbol{H}_1 \boldsymbol{s}_{1k}$, we have that

$$\frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{h}_1^T} = A_d \boldsymbol{H}_2 \boldsymbol{\Upsilon}_{L_1}(\boldsymbol{s}_{1k}), \ \frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{h}_2^T} = A_d \boldsymbol{H}_1 \boldsymbol{\Upsilon}_{L_2}(\boldsymbol{s}_{1k}),$$
(7.44)

and

$$\frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{g}_1^T} = \boldsymbol{0}_{N \times N}. \tag{7.45}$$

To evaluate the terms in (7.43), we can use the expression for \boldsymbol{Q} in (7.12) to obtain

$$\frac{\partial \boldsymbol{Q}}{\partial h_{1i}} = \boldsymbol{0}_{N \times N}, \ \frac{\partial \boldsymbol{Q}}{\partial h_{2i}} = A_d^2 P_2 \boldsymbol{E}_i \boldsymbol{G}_1 \boldsymbol{G}_1^H \boldsymbol{H}_2^H + A_d^2 \sigma^2 \boldsymbol{E}_i \boldsymbol{H}_2^H, \tag{7.46}$$

and

$$\frac{\partial \boldsymbol{Q}}{\partial g_{1i}} = A_d^2 P_2 \boldsymbol{H}_2 \boldsymbol{E}_i \boldsymbol{G}_1^H \boldsymbol{H}_2^H.$$
(7.47)

Finally, the semi-blind CRB is given by $CRB_{\theta}^{(s)} = \operatorname{tr}(\Gamma_s^{-1})$. In particular,

$$CRB_{\boldsymbol{h}_{1}}^{(s)} = \sum_{i=1}^{L_{1}} [\boldsymbol{\Gamma}_{s}^{-1}]_{ii}, \ CRB_{\boldsymbol{h}_{2}}^{(s)} = \sum_{i=L_{1}+1}^{L_{1}+L_{2}} [\boldsymbol{\Gamma}_{s}^{-1}]_{ii},$$
(7.48)

and

$$CRB_{g_1}^{(s)} = \sum_{i=L_1+L_2+1}^{L_1+L_2+J_1} [\Gamma_s^{-1}]_{ii}.$$
(7.49)

7.4.3 Power Allocation at the Relay

The factor α determines the fraction of the average power, P_r , that the relay allocates to the superimposed pilot. Given that P_r is fixed, the more power the relay allocates to the superimposed pilot the less it allocates to the pilots transmitted by the terminals. Both the superimposed pilot and the pilots transmitted by the terminals are essential for the estimation process, and it is desired to choose a value of α that optimizes estimation performance. Since the pilot-based estimator is used to initialize the semi-blind estimator, α should be chosen so that both estimators perform well. One way to choose α is to find the value that minimizes the average CRB over the channel distribution. Unfortunately, determining this value analytically is a very challenging task due to the complicated expressions of the CRBs. In Section 7.5, we plot the CRBs for both estimators versus α at several SNR values (averaged over many channel realizations). It turns out that it is not critical to choose the optimal value of α , as the optimal value lies within a flat region of the CRBs for both estimators. In other words, no fine-tuning with respect to α is required and it is sufficient to choose α by inspection. Moreover, there is no need to change the value α according to SNR.

7.5 Simulation Results

In this section, we investigate through simulations the performance of the proposed semiblind algorithm and compare it to that of the pilot-based LS estimator. Our results are obtained using $P_1 = P_2 = \frac{1}{2}P_r$. We assume that h_1 and h_2 have 5 taps each, while g_1 has 4 taps. We average our results over a set of 300 independent realizations of h_1 , h_2 and g_1 . The taps of each channel vector are modelled as independent and zero mean complex Gaussian random variables whose variances follow the exponential power decay profile. The variance of the ℓ -th channel tap is $\sigma_{\ell}^2 = e^{-(\ell-1)/10}$ and the vectors h_1 , h_2 and g_1 are assumed independent of each other. The number of subcarriers is set at N = 64. Unless mentioned otherwise, the number of data symbol blocks is K = 10, and the vectors s_{1k} , s_{2k} , $k = 1, \ldots, K$ are modelled as i.i.d. $CCN(0, P_2 I_N)$. Furthermore, the pilot vectors t_1 , t_2 and t_3 are chosen according to the example provided in Section 7.3.3, with $\kappa = 20$ and $\delta = 40$. As we mentioned earlier, the BFGS algorithm is employed to solve the minimization problem in (7.15), in conjunction with backtracking linesearch [74] that is used to find the appropriate step size at each iteration. We initialize the BFGS algorithm using the channel estimates provided by the pilot-based LS estimator and we assume that the BFGS algorithm has converged when $\|\nabla \mathcal{L}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}^{\{n\}})\|^2 < 10^{-4}$, where $\boldsymbol{\theta}^{\{n\}}$ is the value of estimate of $\boldsymbol{\theta}$ at the *n*th iteration. The inverse-Hessian approximation is initialized using matrix $\beta \boldsymbol{I}$ where β is set at 0.005.

In Fig. 7.4, we plot the semi-blind and pilot-based CRBs versus the parameter α for several SNR values (6 dB, 15 dB, 24 dB). As we can see from Fig. 7.4, for both estimators the CRB curve is relatively flat for a wide range of α (approximately $0.25 \leq \alpha \leq 0.6$ for the pilot-based estimator and $0.4 \leq \alpha \leq 0.7$ for the semi-blind estimator) which includes the optimal value. Since the pilot-based estimator is used to initialize the semi-blind estimator, α should be chosen to lie within the intersection of the flat regions of the pilot-based and semi-blind estimators. However, it is not necessary to choose exactly the optimal value of α as it is sufficient to choose a value within this region by inspection. In our subsequent simulations, we use the value $\alpha = 0.55$, which lies within the flat-region for both estimators.

In Fig. 7.5, we plot the MSE performance of the semi-blind algorithm and the pilotbased LS estimator versus SNR along with the corresponding semi-blind and pilot-based CRBs. The MSE performance of the semi-blind estimator is shown for the case where the symbol vectors $\tilde{\boldsymbol{s}}_{1k}$, $\tilde{\boldsymbol{s}}_{2k}$, $k = 1, \ldots, K$ are $\mathcal{CCN}(0, P_2 \boldsymbol{I}_N)$ as well as the case where they are generated using QPSK modulation. As we can see in Fig. 7.5, the performance of the semi-blind estimator is almost the same for the two cases, which illustrates the minimal impact of the proposed Gaussian approximation. In both cases, the semi-blind estimator provides a substantial improvement in accuracy over the pilot-based estimator. In particular, the semi-blind estimator is approximately 3 times more accurate than the pilot-based estimator at low SNR and twice as accurate at medium SNR. Furthermore, the MSE of the semi-blind algorithm closely approaches the semi-blind CRB as SNR increases. In fact, it almost overlaps with the CRB in the medium-to-high SNR range. At very high SNR, the performance of the semi-blind estimator overlaps with that of the pilot-based estimator, which shows that the pilot-based estimator provides highly accurate estimates in this case and incorporating the data samples no longer provides any substantial gain in accuracy.

In Fig. 7.6, we show the average number of iterations needed for the BFGS algorithm to converge at different SNR values for Gaussian-distributed data symbols. The average number of iterations ranges from 32 to 44.5, which shows that the performance enhancement of the semi-blind approach comes at a moderate computational load.

We next consider the effect of the number of data blocks on the performance of the semi-blind estimator. In Fig. 7.7, we plot the MSE performance of the semi-blind algorithm versus the number of OFDM data blocks K, along with the semi-blind CRB for Gaussian data. The SNR is set at 20dB. As expected the accuracy of the semi-blind estimator improves as K increases, which shows that the longer the coherence time of the channel the more attractive the semi-blind approach becomes.

In the above simulation results, we investigated the performance of semi-blind estimation for the individual channels h_1 , h_2 and g_1 . We now investigate whether the proposed algorithm is also superior to pilot-based estimation when it is only desirable to estimate the cascaded channels a, b. To answer this question, we apply the semi-blind estimator using the same settings as in Fig. 7.5 (QPSK data), and then obtain the estimates for a and bfrom the estimates of h_1 , h_2 and g_1 using linear convolution. For comparison purposes, we use the LS algorithm to estimate the cascaded channels directly, without employing superimposed training at the relay, while allocating the same training power at the terminals and the relay as we did in the semi-blind scenario. The resulting MSE performances for the two algorithms are plotted versus SNR in Fig. 7.8. As we can see, at low-to-medium SNR the semi-blind approach provides substantial accuracy gains over the LS estimator for the estimation of cascaded channels.

7.6 Conclusions

In this chapter, we proposed a semi-blind channel estimator for OFDM-based AF TWRNs based on the Gaussian ML approach. To assist in the estimation of the individual channels, we employed superimposed training at the relay. The resulting GML estimates were obtained numerically using the BFGS algorithm. We also derived conditions for the optimality of the training pilots and provided examples of optimal pilot vectors. As performance benchmarks, we derived the the CRBs for semi-blind and pilot-based estimation. We then used simulation studies to compare the proposed estimator to the conventional pilot-based estimator and showed that the proposed estimator provides a substantial improvement in accuracy. The MSE performance of the semi-blind algorithm also closely approaches the derived semi-blind CRB. These performance gains are achieved at a reasonable computational cost, which clearly establishes the merit and practicality of semi-blind channel estimation for OFDM-based AF TWRNs.



Fig. 7.4 Semi-blind and pilot-based CRBs versus α for SNR = 6, 15, 24dB.



Fig. 7.5 MSE performance of the semi-blind and pilot-based (LS) estimators along with the corresponding semi-blind and pilot-based CRBs plotted versus SNR for the cases of Gaussian-distibuted and QPSK-distributed data symbols.



Fig. 7.6 The average number of iterations until convergence of the BFGS algorithm at different SNR values assuming Gaussian-distributed data symbols.



Fig. 7.7 MSE performance of the semi-blind and pilot-based estimators along with the corresponding semi-blind and pilot-based CRBs plotted versus K for SNR= 20dB.



Fig. 7.8 MSE performance of the semi-blind algorithm and the pilot-based (LS) algorithm for the estimation of the cascaded channels plotted versus SNR for QPSK data symbols.

Chapter 8

Conclusions and Future Work

8.1 Summary and Conclusions

In this thesis, we considered the problem of channel estimation in AF TWRNs where accurate channel state information is essential to avoid performance degradation. The conventional approach to this problem has been a training-based one that completely relies on the transmission of known pilots by the terminals, and thus diminishes the spectral efficiency of TWRNs. In contrast, our work focused on blind and semi-blind approaches to estimation. Blind estimation relies only on the data samples, while semi-blind estimation is a hybrid approach that exploits both pilots and data samples. Based on these two approaches, we developed new channel estimation algorithms for AF TWRNs in both flat-fading and frequency-selective channel conditions, with the goal of achieving superior tradeoffs between accuracy and spectral efficiency than the conventional training-based approach.

In Chapter 3, we focused on the problem of blind channel estimation for AF TWRNs employing constant modulus signalling. Assuming nonreciprocal flat-fading channels, we proposed an algorithm for blind channel estimation based on the DML approach, which treats the data symbols as deterministic unknowns. We showed that, for M-PSK modulation, the proposed estimator is consistent and approaches the true channel with high probability at high SNR for modulation orders higher than 2. An alternative algorithm was proposed for the case of BPSK modulation where the DML algorithm performed poorly. For comparative purposes, we investigated the GML approach which treats the data symbols as Gaussian-distributed nuisance parameters. We also derived the corresponding CRB. Using simulations, we showed that the DML estimator performs better than the GML estimator at medium-to-high SNR and approaches the derived CRB at high SNR, unlike the GML estimator which encounters an error-floor. We also used SER simulations to show that the DML estimator provides superior tradeoffs between accuracy and spectral efficiency than the pilot-based LS estimator.

In Chapter 4, we turned our attention to reciprocal channels. In this case, the DML approach was shown to result in an inconsistent estimator. As an alternative, we proposed the MSEV estimator which minimized the sample variance of the envelope of the received signal after self-interference cancellation. This estimator is consistent and approaches the true channel with high probability at high SNR. The corresponding CRB for the case of reciprocal channels was also derived. Simulations were used to show that the MSEV estimator outperforms the DML estimator and can be accurately implemented using steepest descent at a computational complexity that is linear in the number of data samples.

Still within the context of flat-fading channels, we then considered semi-blind channel estimation which uses both pilots and data samples to estimate the channel parameters. In Chapter 5, we derived the exact CRB for semi-blind estimation of nonreciprocal channels in AF TWRNs employing square QAM. Unlike the bounds in the previous chapters which treated the data symbols as deterministic unknowns, this bound was based on the true likelihood function that takes into account the statistics of the data symbols. Using the derived bound, we showed that incorporation into the estimation process of even a limited number of data samples leads to substantial accuracy gains over the pilot-based approach and significantly reduces the number of required pilot symbols. These improvements hold for all modulation orders and are highest at low modulation orders. The MCRB was also derived as a more tractable alternative for the exact CRB.

In Chapter 6, we showed that the performance gains of semi-blind estimation promised by the CRB analysis in Chapter 5 were indeed achievable and with an affordable computational cost. To avoid the high complexity of direct ML estimation based on the true likelihood function, semi-blind estimation was implemented using the iterative EM approach. EM-based algorithms were derived for both nonreciprocal and reciprocal channels. For both cases, the complexity of the EM steps was linear in the number of data samples for a fixed modulation order. Moreover, the proposed algorithms performed very closely to the exact CRBs and required only a small number of iterations to converge. Using SER simulations, we showed the EM algorithm provides a significant improvement in data

123

throughput compared to the LS estimator due to its reduced pilot requirements.

Finally, in Chapter 7 we considered semi-blind channel estimation for OFDM-based TWRNs operating in frequency selective channel conditions. Unlike previous works, we assumed that the channels were nonreciprocal as this is more realistic for OFDM systems in frequency selective environments. To assist in the estimation of the individual channels, we adopted a superimposed training strategy at the relay whereby the relay superimposed its own pilot symbols over the received pilot-bearing OFDM block before broadcasting it. Our proposed semi-blind estimator was based on the Gaussian ML approach which assumes that the data symbols were Gaussian-distributed, and the resulting estimates were obtained numerically using the BFGS algorithm. The pilot vectors of the two terminals and the relay were designed to optimize estimation performance, and the corresponding semi-blind and pilot-based CRBs were derived. Using simulations, we showed that the proposed semi-blind estimator closely approaches the semi-blind CRB and provides substantial improvements in accuracy over the pilot-based approach while using only a limited number of OFDM data blocks. These improvements also hold when the data is drawn from discrete constellations such as QPSK.

In summary, in this thesis we investigated the application of blind and semi-blind approaches to channel estimation in AF TWRNs. We developed blind and semi-blind methods that achieve significantly better tradeoffs between accuracy and spectral efficiency than the conventional training based approach for both flat-fading and frequency selective channel conditions. These gains were achieved at an affordable computational cost and with a limited number of data symbols. We thus demonstrated that the (semi)-blind channel estimation approaches are viable and practical alternatives to the training-based approach. In particular, they are convenient solutions for applications that require high estimation accuracy and/or highly efficient spectrum utilization.

8.2 Directions for Future Work

The results presented in our work showed that the application of blind and semi-blind estimation techniques in AF TWRNs is very promising and that significant improvements in performance are feasible. This motivates us to extend the application of these techniques in several directions within the context of AF TWRNs. Semi-blind estimation, in particular, lends itself easily to wide application because of its flexibility and because no separate ambiguity resolution is needed.

Like most works on channel estimation for TWRNs, our work has focused on the single relay scenario. Channel estimation for TWRNs with multiple relays remains an open research problem. The use of multiple relays makes it possible to apply beamforming strategies to enhance the quality of the received signal [97]. These strategies require highly accurate channel information, and semi-blind estimation techniques can be employed in order to achieve the desired estimation accuracy with minimal training costs. In AF TWRNs with multiple relays, both the effective self-interference channel and the effective informationbearing channel are sums of channel components corresponding to the different relaying paths. One way to make the estimation of the channels corresponding to each relaying path feasible is to allocate a separate training phase for each relay during which the other relays are silent. Once initial estimates have been acquired for all the channels, all the relays can again be used simultaneously. Semi-blind estimation can be employed to improve the accuracy of the initial estimates and to reduce the training costs of the first phase. In fact, the EM approach that we adopted in Chapter 6 may be a convenient way of applying semi-blind estimation in this scenario at a low computational cost while also taking into account the statistics of the data symbols.

Another important problem where efficient estimation algorithms are needed is channel estimation for MIMO AF TWRNs. In MIMO AF TWRNs, both the terminals and the relay are equipped with multiple antennas and channel estimation is more challenging due to the large number of channel parameters involved. In this case, a significant training overhead may be needed if the conventional pilot-based methods are employed. Semi-blind channel estimation based on the Gaussian ML approach which we applied in OFDM-based AF TWRNs in Chapter 7 may also be a convenient way of improving the estimation accuracy and reducing the training overhead for MIMO AF TWRNs. As we did in Chapter 7, we can use superimposed training to assist in the estimation of the individual channels. The number of pilot vectors should be chosen to guarantee the identifiability of the channel parameters. Moreover, an appropriate low-complexity iterative method will be needed to obtain the ML channel estimates.

Finally, another problem of interest is joint CFO and channel estimation for AF TWRNs in frequency selective environments. The CFO results from the frequency mismatch between the oscillators at different nodes in the network. Accurate estimation of the CFO is necessary to perform frequency synchronization and to avoid a severe degradation in the system performance. Several methods for joint CFO and channel estimation in AF TWRNs have been proposed [32–34], but they all follow the training-based approach. In fact, the system model in Chapter 7 can be easily extended to account for the presence of CFO. Given the superior channel estimation performance of the semi-blind Gaussian ML approach, it is a promising candidate for the joint estimation of both the CFO and the channel parameters.
Appendix A

Derivations of Select Proofs and Results from Chapter 3

A.1 Derivation of Eq. (3.22)

In this appendix, we derive the closed-form expression for the variance $\mathcal{V}(v)$ of |y(v)|. We recall that

$$y(v) = Av\sqrt{P_1}e^{j\phi_1} + Ab\sqrt{P_2}e^{j\phi_2} + Ah_2n + \eta.$$
 (A.1)

The variance of |y(v)| is $\mathcal{V}(v) = \mathbb{E}\{|y(v)|^2\} - \mathbb{E}\{|y(v)|\}^2$. It can be easily shown that

$$\mathbb{E}\{|y(v)|^2\} = A^2|v|^2P_1 + A^2|b|^2P_2 + C.$$
(A.2)

To evaluate $\mathbb{E}\{|y(v)|\}$, we first obtain the conditional expectation $\mathbb{E}\{|y(v)| \mid \phi_1, \phi_2\}$. We can see from (A.1) that $\Re\{y(v)\}$ and $\Im\{y(v)\}$ conditioned on ϕ_1 and ϕ_2 are Gaussian-distributed with conditional means

$$\mathbb{E}\{\Re(y(v)) \mid \phi_1, \phi_2\} = A|v|\sqrt{P_1}\cos(\phi_v + \phi_1) + A|b|\sqrt{P_2}\cos(\phi_b + \phi_2),$$
$$\mathbb{E}\{\Im(y(v)) \mid \phi_1, \phi_2\} = A|v|\sqrt{P_1}\sin(\phi_v + \phi_1) + A|b|\sqrt{P_2}\sin(\phi_b + \phi_2)$$

and a conditional variance of C/2. Hence, when conditioned on ϕ_1 and ϕ_2 , |y(v)| is a noncentral Chi random variable whose mean is [98]

$$\mathbb{E}\{|y(v)| \mid \phi_1, \phi_2\} = \sqrt{\frac{\pi C}{4}} L_{1/2}(-\lambda(v; \phi_1, \phi_2))$$

where

$$\lambda(v;\phi_1,\phi_2) \triangleq \frac{1}{C} \left(\mathbb{E}\{\Re(y(v))|\phi_1,\phi_2\}^2 + \mathbb{E}\{\Im(y(v))|\phi_1,\phi_2\}^2 \right) \\ = \frac{A^2}{C} \left(|v|^2 P_1 + |b|^2 P_2 + 2|v||b|\sqrt{P_1 P_2}\cos(\phi_v - \phi_b + \phi_1 - \phi_2) \right),$$
(A.3)

 $L_{1/2}(x) = e^{x/2} \left[(1-x)I_0(x/2) + xI_1(x/2) \right]$ is the Laguerre polynomial with parameter 1/2, and $I_{\varsigma}(\cdot)$ is the Modified Bessel Function of the First Kind of order ς [80]. The last equation shows that $\lambda(v; \phi_1, \phi_2)$ depends on the difference $(\phi_1 - \phi_2) \mod 2\pi$, i.e., $\lambda(v; \phi_1, \phi_2) = \lambda(v; \phi_1 - \phi_2)$. The difference, $(\phi_1 - \phi_2)$ takes the values $\varphi_k = \frac{2\pi k}{M}$, $k = 0, \ldots, M - 1$ with equal probability. Letting $\lambda_k(v) \triangleq \lambda(v; \phi_1 - \phi_2 = \varphi_k)$, the unconditional mean $\mathbb{E}\{|y(v)|\}$ is given by

$$\mathbb{E}\{|y(v)|\} = \sum_{k=0}^{M-1} \sqrt{\frac{\pi C}{4M^2}} L_{1/2}\left(-\lambda_k(v)\right), \qquad (A.4)$$

which completes the proof of (3.22).

A.2 Proof of Lemma 1

In this appendix, we show that $\mathcal{V}(v)$ has a unique global minimum at v = 0. Let $\nu \triangleq \sqrt{\frac{1}{C} (A^2 |v|^2 P_1)}, \zeta \triangleq \sqrt{\frac{1}{C} (A^2 |b|^2 P_2)}$, and let

$$D(\nu) \triangleq \nu^2 + \zeta^2 - \frac{\pi}{4M^2} \left(\sum_{k=0}^{M-1} L_{1/2} \left(- \left[\nu^2 + \zeta^2 + 2\nu\zeta\cos\theta_k(v) \right] \right) \right)^2.$$
(A.5)

It is sufficient to prove that $D(\nu)$ has a unique global minimum at $\nu = 0$. It is straightforward to verify that $L_{1/2}(-x)$ is a strictly concave function. Therefore,

$$\frac{1}{M} \sum_{k=0}^{M-1} L_{1/2} \left(-\left[\nu^2 + \zeta^2 + 2\nu\zeta\cos\theta_k(v)\right] \right) \le L_{1/2} \left(-\frac{1}{M} \sum_{k=0}^{M-1} \left[\nu^2 + \zeta^2 + 2\nu\zeta\cos(\theta_k(v))\right] \right) = L_{1/2} \left(-\left[\nu^2 + \zeta^2\right] \right),$$
(A.6)

where we have used the fact that $\sum_{k=0}^{M-1} \cos(\phi_v - \phi_b + \frac{2k\pi}{M}) = 0$. Hence,

$$D(\nu) \ge \nu^2 + \zeta^2 - \frac{\pi}{4} \left(L_{1/2} \left(- \left[\nu^2 + \zeta^2 \right] \right) \right)^2.$$
 (A.7)

For $\zeta \neq 0$, the equality holds if and only if $\nu = 0$. Let $F(\nu) \triangleq \nu^2 + \zeta^2 - \frac{\pi}{4} \left(L_{1/2} \left(- \left[\nu^2 + \zeta^2 \right] \right) \right)^2$. It is sufficient to show that $F(\nu)$ has a unique global minimum at $\nu = 0$. We have

$$\frac{d}{d\nu}F(\nu) = 2\nu - \nu\pi L_{1/2}\left(-\left[\nu^2 + \zeta^2\right]\right)\mathcal{Z}(\nu^2 + \zeta^2),$$
(A.8)

where $\mathcal{Z}(x) \triangleq \frac{1}{2}e^{-x/2} \left(I_0(x/2) + I_1(x/2) \right)$. Since $\frac{d}{d\nu}F(\nu) = 0$ for $\nu = 0$, it is sufficient to show that $\frac{dF(\nu)}{d\nu} > 0$ for $\nu > 0$, i.e., that $\frac{\pi}{2}L_{1/2} \left(-\left[\nu^2 + \zeta^2\right] \right) \mathcal{Z}(\nu^2 + \zeta^2) < 1$ for $\nu > 0$. Let $\rho \triangleq \nu^2 + \zeta^2$ and let

$$\mathcal{S}(\rho) \triangleq \frac{\pi}{2} L_{1/2}(-\rho) \,\mathcal{Z}(\rho) \\ = \frac{\pi}{4} e^{-\rho} \Big[(1+\rho) I_0 \Big(\frac{\rho}{2}\Big) + \rho I_1 \Big(\frac{\rho}{2}\Big) \Big] \Big[I_0 \Big(\frac{\rho}{2}\Big) + I_1 \Big(\frac{\rho}{2}\Big) \Big] , \tag{A.9}$$

we have to show that $S(\rho) < 1$ for $\rho > 0$. We will do this by showing that $\frac{d}{d\rho}S(\rho) > 0$ for $\rho > 0$ and that $\lim_{\rho \to \infty} S(\rho) = 1$. We have

$$\frac{d}{d\rho}\mathcal{S}(\rho) = \frac{\pi}{2}e^{-\rho} \bigg[\frac{1}{4}I_0 \bigg(\frac{\rho}{2}\bigg)^2 - \frac{1}{4}I_1 \bigg(\frac{\rho}{2}\bigg)^2 - \frac{1}{2\rho}I_0 \bigg(\frac{\rho}{2}\bigg) I_1 \bigg(\frac{\rho}{2}\bigg) \bigg].$$
(A.10)

Let $U(\rho) \triangleq \rho I_0 \left(\frac{\rho}{2}\right)^2 - \rho I_1 \left(\frac{\rho}{2}\right)^2 - 2I_0 \left(\frac{\rho}{2}\right) I_1 \left(\frac{\rho}{2}\right)$. Hence $\frac{d}{d\rho} \mathcal{S}(\rho) = \frac{\pi}{8\rho} e^{-\rho} U(\rho)$. Moreover, $\frac{d}{d\rho} U(\rho) = \frac{2}{\rho} I_0 \left(\frac{\rho}{2}\right) I_1 \left(\frac{\rho}{2}\right) > 0$. Thus, $U(\rho)$ is strictly increasing for $\rho > 0$. Since U(0) = 0, we have that $U(\rho) > 0$ for $\rho > 0$, which implies that $\frac{d}{d\rho} \mathcal{S}(\rho) > 0$ for $\rho > 0$. It remains to show that $\lim_{\rho \to \infty} S(\rho) = 1$. For large arguments, $I_{\varsigma}(\cdot)$ has the following asymptotic expansion [80]

$$I_{\varsigma}(x) \approx \frac{e^x}{\sqrt{2\pi x}} \left\{ 1 - \frac{\varepsilon - 1}{8x} + \frac{(\varepsilon - 1)(\varepsilon - 9)}{2!(8x)^2} - \dots \right\},\tag{A.11}$$

where $\varepsilon = 4\varsigma^2$. We rewrite $\mathcal{S}(\rho)$ as

$$\mathcal{S}(\rho) = \frac{\pi}{4} e^{-\rho} \left[(1+\rho) I_0 \left(\frac{\rho}{2}\right)^2 + (1+2\rho) I_0 \left(\frac{\rho}{2}\right) I_1 \left(\frac{\rho}{2}\right) + \rho I_1 \left(\frac{\rho}{2}\right)^2 \right].$$
(A.12)

Using the expansion in (A.11), we obtain

$$\lim_{\rho \to \infty} \frac{\pi}{4} e^{-\rho} (1+\rho) I_0 \left(\frac{\rho}{2}\right)^2 = \lim_{\rho \to \infty} \frac{\pi}{4} e^{-\rho} \rho I_1 \left(\frac{\rho}{2}\right)^2 = \frac{1}{4},$$
(A.13)

and

$$\lim_{\rho \to \infty} \frac{\pi}{4} e^{-\rho} (1+2\rho) I_0\left(\frac{\rho}{2}\right) I_1\left(\frac{\rho}{2}\right) = \frac{1}{2}.$$
 (A.14)

Therefore, $\lim_{\rho \to \infty} S(\rho) = 1$, which completes the proof.

A.3 Proof of Lemma 2

In this appendix, we prove that $V_N(v)$ converges uniformly in probability to $\mathcal{V}(v)$ when a, b and v belong to compact sets. Suppose that $v \in \mathcal{C}$. By definition, $V_N(v)$ converges uniformly in probability to $\mathcal{V}(v)$ when $\sup_{v \in \mathcal{C}} |V_N(v) - \mathcal{V}(v)|$ converges in probability to zero as $N \to \infty$. Since $y_i(v) = \tilde{z}_i(a - v) = z_i - A(a - v)s_{1i}$, $V_N(v)$ is a function of the parameter v, the observations \boldsymbol{z} and known data symbols \boldsymbol{t}_1 . According to Lemma 2.9 in [28, Ch. 36], a sufficient condition for uniform convergence in probability when \mathcal{C} is compact is the existence of a function $\mathcal{F}_N(\boldsymbol{z}, \boldsymbol{s}_1)$ with bounded expectation $\mathbb{E}\{\mathcal{F}_N(\boldsymbol{z}, \boldsymbol{s}_1)\}$ such that for all $v_1, v_2 \in \mathcal{C}$, $|V_N(v_1) - V_N(v_2)| \leq \mathcal{F}_N(\boldsymbol{z}, \boldsymbol{s}_1)|v_1 - v_2|$.

Using the triangular inequality, we obtain

$$\begin{aligned} \left| V_{N}(v_{1}) - V_{N}(v_{2}) \right| &\leq \frac{1}{N-1} \sum_{i=1}^{N} \left| \left(|y_{i}(v_{1})| - \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{1})| \right)^{2} - \left(|y_{i}(v_{2})| - \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{2})| \right)^{2} \right| \\ &= \frac{1}{N-1} \sum_{i=1}^{N} \left(\left| |y_{i}(v_{1})| - |y_{i}(v_{2})| - \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{1})| + \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{2})| \right| \right. \\ & \left. \times \left| |y_{i}(v_{1})| + |y_{i}(v_{2})| - \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{1})| - \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{2})| \right| \right) \right. \end{aligned}$$
(A.15)

For $i = 1, \ldots, N$, let

$$\Upsilon_{i}(v_{1}, v_{2}) \triangleq \left| |y_{i}(v_{1})| - |y_{i}(v_{2})| - \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{1})| + \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{2})| \right|$$
(A.16)

and

$$\Lambda_i(v_1, v_2) \triangleq \left| |y_i(v_1)| + |y_i(v_2)| - \frac{1}{N} \sum_{k=1}^N |y_k(v_1)| - \frac{1}{N} \sum_{k=1}^N |y_k(v_2)| \right|.$$
(A.17)

Using the triangular inequality again, we get

$$\Upsilon_{i}(v_{1}, v_{2}) \leq |y_{i}(v_{1}) - y_{i}(v_{2})| + \frac{1}{N} \sum_{k=1}^{N} |y_{k}(v_{1}) - y_{k}(v_{2})|$$

$$= 2AP_{1}|v_{1} - v_{2}|,$$
(A.18)

and

$$\Lambda_i(v_1, v_2) \le |y_i(v_1)| + |y_i(v_2)| + \frac{1}{N} \sum_{k=1}^N |y_k(v_1)| + \frac{1}{N} \sum_{k=1}^N |y_k(v_2)|.$$
(A.19)

Noting that $|y_i(v)| \leq |z_i| + A|a|P_1 + AP_1|v|$, we further obtain

$$\Lambda_i(v_1, v_2) \le 2|z_i| + \frac{2}{N} \sum_{k=1}^N |z_k| + 4AP_1|a| + 2AP_1(|v_1| + |v_2|).$$
(A.20)

Since the set C is compact, there exists $\mathcal{U} > 0$ such that $|v| \leq \mathcal{U}, \forall v \in C$. Hence, we obtain

the following upper bound on $\Lambda_i(v_1, v_2)$:

$$\Lambda_i(v_1, v_2) \le 2|z_i| + \frac{2}{N} \sum_{k=1}^N |z_k| + 2AP_1|a| + 4A\mathcal{U}.$$
 (A.21)

Combining (A.18) and (A.21), we obtain

$$\left|V_{N}(v_{1}) - V_{N}(v_{2})\right| \leq \frac{8AP_{1}}{N-1} \left[ANP_{1}\mathcal{U} + ANP_{1}|a| + \sum_{i=1}^{N} |z_{i}|\right] |v_{1} - v_{2}|.$$
(A.22)

Noting that $z_i = y_i(a)$, we have from (A.4) that

$$\mathbb{E}\{|z_i|\} \le \sqrt{\frac{\pi C}{4}} L_{1/2} \left(-\frac{1}{C} \left(A|a|P_1 + A|b|P_2 \right)^2 \right).$$
(A.23)

Since both a and b belong to compact sets, there exists $\mathcal{M}_1 > 0$ and $\mathcal{M}_2 > 0$ such that $|a| \leq \mathcal{M}_1$ and $|b| \leq \mathcal{M}_2$. Hence,

$$\mathbb{E}\{|z_i|\} \le \sqrt{\frac{\pi C}{4}} L_{1/2} \left(-\frac{1}{C} \left(A \mathcal{M}_1 P_1 + A \mathcal{M}_2 P_2 \right)^2 \right).$$
(A.24)

Letting, $\mathcal{G} \triangleq \sqrt{\frac{\pi C}{4}} L_{1/2} \left(-\frac{1}{C} \left(A \mathcal{M}_1 P_1 + A \mathcal{M}_2 P_2 \right)^2 \right)$, we obtain

$$\mathbb{E}\left\{\frac{8AP_1}{N-1}\left[ANP_1\mathcal{U} + ANP_1|a| + \sum_{i=1}^N |z_i|\right]\right\} \le 16A^2(\mathcal{U} + \mathcal{M}_1) + 16A\mathcal{G},\tag{A.25}$$

which completes the proof.

A.4 Proof of Lemma 3

In this appendix, we prove Lemma 3. It is obvious from (3.24) that $X(v) \ge 0$ with equality if and only if the terms $|Avs_{1i} + Abs_{2i}|$, i = 1, ..., N are all equal. Because of the constant modulus nature of the data symbols, this occurs at v = 0 for any M, which means that there is always a global minimum at v = 0 (i.e., at u = a). To see whether we also have X(v) = 0 for some $v \neq 0$, we rewrite $|Avs_{1i} + Abs_{2i}|$ as

$$|Avs_{1i} + Abs_{2i}| = |A|v|\sqrt{P_1}e^{j\phi_v + \phi_{1i}} + A|b|\sqrt{P_2}e^{j\phi_b + \phi_{2i}}|$$

= |A|v|\sqrt{P_1}e^{j\phi_v - \phi_b + \phi_{1i} - \phi_{2i}} + A|b|\sqrt{P_2}|. (A.26)

Let $\chi_i \triangleq \cos(\phi_v - \phi_b + \phi_{1i} - \phi_{2i})$, i = 1, ..., N. It is clear from (A.26) that X(v) is zero whenever the terms χ_i , i = 1, ..., N are all equal. We denote by \mathcal{E} the event that the terms χ_i , i = 1, ..., N are equal. Let $\psi_i \triangleq \phi_{1i} - \phi_{2i}$, then the values ψ_i , i = 1, ..., N are i.i.d. realizations of the discrete uniform random variable Ψ which takes values from the set $S_{\Psi} = \left\{\frac{2\ell\pi}{M}, \ell = 0, ..., M - 1\right\}$. We will now show that the event \mathcal{E} occurs if and only if the values ψ_i , i = 1, ..., N are chosen from the same size 2 subset of S_{Ψ} . Suppose that ψ_1 is fixed, and that we are choosing the remaining phases such that $\cos(\psi_i + \vartheta) = \cos(\psi_1 + \vartheta)$ for i = 2, ..., N. If ψ_{κ} is different from ψ_1 for some index κ , then $\cos(\psi_{\kappa} + \phi_v - \phi_b) = \cos(\psi_1 + \phi_v - \phi_b)$ can only be satisfied if $2(\phi_v - \phi_b) = (-\psi_1 - \psi_{\kappa})$. For the remaining phases with indices i = 2, ..., N, $i \neq \kappa$, the equality $\cos(\psi_i + \phi_v - \phi_b) = \cos(\psi_1 + \phi_v - \phi_b)$ holds only if $\psi_i = \psi_1$ or $\psi_i = \psi_{\kappa}$. Therefore, the terms χ_i , i = 1, ..., N are equal if and only if the phases ψ_i , i = 1, ..., N take at most two distinct values, i.e, the probability that \mathcal{E} occurs is

$$P\left(\mathcal{E}\right) = \binom{M}{2} \frac{2^{N}}{M^{N}} = \left(\frac{2}{M}\right)^{N-1} (M-1), \qquad (A.27)$$

where $\binom{M}{2}$ is the number of distinct subsets of size 2 that can be chosen from a set of size M. Suppose that \mathcal{E} occurs and that ψ_{Σ} is the sum of the two distinct phase values, then X(v) = 0 for all the values of v that satisfy $2(\phi_v - \phi_b) = -\psi_{\Sigma}$, which means that there are infinitely many global minimizers of X(v). Hence, the probability that X(v) has a unique minimum at v = 0 is $P_{M,N} = 1 - \left(\frac{2}{M}\right)^{N-1} (M-1)$.

Appendix B

Derivations of Select Proofs and Results from Chapter 4

B.1 Proof of Lemma 4

We have

$$\left(1 - \frac{\mathcal{W}(a)}{\sigma^2(A^2|a|+1)}\right) = \frac{\pi}{4} \left(L_{1/2} \left(-\frac{A^2|b|^2 P_2}{\sigma^2(A^2|a|+1)}\right)\right)^2 - \frac{A^2|b|^2 P_2}{\sigma^2(A^2|a|+1)}.$$
(B.1)

Let $\mathcal{Q}(x) \triangleq \frac{\pi}{4} (L_{1/2}(-x))^2 - x$, it is sufficient to show that $\mathcal{Q}(x) > 0$ for x > 0. We know from Appendix A.2 that $\mathcal{Q}(x)$ is strictly decreasing for x > 0. Using this fact, we will establish that $\mathcal{Q}(x) > 0$ for x > 0 by showing that $\lim_{x \to \infty} \mathcal{Q}(x) = \frac{1}{2}$. We can expand $\mathcal{Q}(x)$ as

$$\mathcal{Q}(x) = \frac{\pi}{4} e^{-x} \left[(1+x)^2 I_0 \left(\frac{x}{2}\right)^2 + 2x(1+x) I_0 \left(\frac{x}{2}\right) I_1 \left(\frac{x}{2}\right) + x^2 I_1 \left(\frac{x}{2}\right)^2 \right] - x.$$
(B.2)

Using the expansion in (A.11), we obtain the following approximations for large x

$$\frac{\pi}{4}e^{-x}(1+x)^2 I_0\left(\frac{x}{2}\right)^2 \approx \frac{x}{4} + \frac{5}{8},\tag{B.3}$$

$$\frac{\pi}{4}e^{-x}2x(1+x)I_0\left(\frac{x}{2}\right)I_1\left(\frac{x}{2}\right) \approx \frac{x}{2} + \frac{1}{4},$$
(B.4)

$$\frac{\pi}{4}e^{-x}x^2I_1\left(\frac{x}{2}\right)^2 \approx \frac{x}{4} - \frac{3}{8}.$$
(B.5)

Substituting these approximations into (B.2), we have $\lim_{x\to\infty} Q(x) = \frac{1}{2}$, which completes the proof.

B.2 Steepest Descent Implementation

In this appendix, we provide the details of the steepest descent implementation for the objective functions of the DML estimator in (4.2) and MSEV estimator in (4.3). Let

$$\mathcal{B}(u) \triangleq \frac{\sum_{i=1}^{N} \left(|z_i - Aut_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aut_{1k}| \right)^2}{\sigma^2 (A^2 |u| + 1)} + N \log \left(A^2 |u| + 1 \right)$$
(B.6)

and

$$\mathcal{D}(u) \triangleq \sum_{i=1}^{N} \left(|z_i - Aut_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aut_{1k}| \right)^2,$$
(B.7)

be the DML and MSEV objective functions, respectively. The steepest descent algorithm follows the direction of the negative gradient. Thus, for the DML estimator, the update equation is

$$\hat{a}^{(k+1)} = \hat{a}^{(k)} - \mu^{(k)} \nabla \mathcal{B}(\hat{a}^{(k)}), \tag{B.8}$$

where $\mu^{(k)}$ is the step size and $\nabla \mathcal{B}(u)$ is the gradient of $\mathcal{B}(u)$, given by

$$\nabla \mathcal{B}(u) = \frac{\partial \mathcal{B}(u)}{\partial \Re\{u\}} + j \frac{\partial \mathcal{B}(u)}{\partial \Im\{u\}}.$$
(B.9)

The partial derivatives $\frac{\partial \mathcal{B}(u)}{\partial \Re\{u\}}$ and $\frac{\partial \mathcal{B}(u)}{\partial \Im\{u\}}$ are given by

$$\frac{\partial \mathcal{B}(u)}{\partial \Re\{u\}} = \frac{1}{\sigma^2} \left[\frac{|u| (A^2 |u| + 1) \frac{\partial \mathcal{D}(u)}{\partial \Re\{u\}} - A^2 \mathcal{D}(u) \Re\{u\}}{|u| (A^2 |u| + 1)^2} \right] + \frac{A^2 \Re\{u\}}{|u| (A^2 |u| + 1)}$$
(B.10)

and

$$\frac{\partial \mathcal{B}(u)}{\partial \Im\{u\}} = \frac{1}{\sigma^2} \left[\frac{|u|(A^2|u|+1)\frac{\partial \mathcal{D}(u)}{\partial \Im\{u\}} - A^2 \mathcal{D}(u)\Im\{u\}}{|u|(A^2|u|+1)^2} \right] + \frac{A^2\Im\{u\}}{|u|(A^2|u|+1)}.$$
 (B.11)

For the MSEV estimator, the update equation is

$$\hat{a}_{v}^{(k+1)} = \hat{a}_{v}^{(k)} - \delta^{(k)} \nabla \mathcal{D}(\hat{a}_{v}^{(k)}), \tag{B.12}$$

where $\delta^{(k)}$ is the step size and $\nabla \mathcal{D}(u)$ is the gradient of $\mathcal{D}(u)$, given by

$$\nabla \mathcal{D}(u) = \frac{\partial \mathcal{D}(u)}{\partial \Re\{u\}} + j \frac{\partial \mathcal{D}(u)}{\partial \Im\{u\}}.$$
 (B.13)

The partial derivatives $\frac{\partial \mathcal{D}(u)}{\partial \Re\{u\}}$ and $\frac{\partial \mathcal{D}(u)}{\partial \Im\{u\}}$ are given by

$$\frac{\partial \mathcal{D}(u)}{\partial \Re\{u\}} = 2\sum_{i=1}^{N} \left(|z_i - Aut_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aut_{1k}| \right) \times \left[-\frac{\Re\{At_{1i}^*(z_i - Aut_{1i})\}}{|z_i - Aut_{1i}|} + \frac{1}{N} \sum_{k=1}^{N} \frac{\Re\{At_{1k}^*(z_k - Aut_{1k})\}}{|z_k - Aut_{1k}|} \right]$$
(B.14)

and

$$\frac{\partial \mathcal{D}(u)}{\partial \Im\{u\}} = 2 \sum_{i=1}^{N} \left(|z_i - Aut_{1i}| - \frac{1}{N} \sum_{k=1}^{N} |z_k - Aut_{1k}| \right) \times \left[-\frac{\Im\{At_{1i}^*(z_i - Aut_{1i})\}}{|z_i - Aut_{1i}|} + \frac{1}{N} \sum_{k=1}^{N} \frac{\Im\{At_{1k}^*(z_k - Aut_{1k})\}}{|z_k - Aut_{1k}|} \right].$$
(B.15)

To determine the convergence of the steepest descent algorithm, we test the difference in the value of the objective function between consecutive iterations. We assume that convergence is achieved for when the difference is below 10^{-5} . It is clear from (B.6)–(B.7), (B.10)–(B.11) and (B.14)–(B.15) that the number of required operations to implement the steepest descent algorithm for both estimators is O(N), i.e., it is linear in the number of received samples. The step sizes $\mu^{(k)}$ and $\delta^{(k)}$ are found using backtracking line search [74], describedd in the following pseudo-code.

Given the descent direction $\nabla \mathcal{B}(\hat{a}^{(k)})$ and some $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$ $\mu^{(k)} := 1$ while $\mathcal{B}\left(\hat{a}^{(k)} - \mu^{(k)} \nabla \mathcal{B}(\hat{a}^{(k)})\right) > \mathcal{B}(a^{(k)}) - \alpha \mu^{(k)} |\nabla \mathcal{B}(\hat{a}^{(k)}))|^2$ $\mu^{(k)} := \beta \mu^{(k)}$ end

Appendix C

Derivations of Select Proofs and Results from Chapter 5

C.1 Proof of Eqs. (5.30) and (5.31)

In this appendix, we prove (5.30) and (5.31). Recalling that $u_k = \Re\{z_k - Aas_{1k}\}b_R + \Im\{z_k - Aas_{1k}\}b_I$, the second derivative of $F_{\theta}(u_k)$ with respect to a_R is given by

$$\frac{\partial^2 F_{\theta}(u_k)}{\partial a_R^2} = 4A^2 \Re\{s_{1k}^*b\}^2 \sum_{i=1}^{2^{p-1}} \beta_i^2 e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k].$$
(C.1)

Thus,

$$\mathbb{E}\left\{B_{k}^{(11)}\right\} = 4A^{2}\Re\{s_{1k}^{*}b\}^{2}\sum_{i=1}^{2^{p-1}}\beta_{i}^{2}e^{-\gamma_{i}|b|^{2}}\mathbb{E}\left\{\frac{\cosh[2\beta_{i}u_{k}]}{F_{\theta}(u_{k})}\right\}.$$
(C.2)

Using the PDF for u_k in (5.23), we obtain:

$$\mathbb{E}\left\{\frac{\cosh[2\beta_i u_k]}{F_{\theta}(u_k)}\right\} = \frac{2}{\sqrt{\pi MC|b|^2}} \int_{-\infty}^{\infty} \cosh[2\beta_i t] e^{-\frac{t^2}{C|b|^2}} dt.$$
(C.3)

Moreover, for $\alpha > 0$, it can be verified that

$$\int_{-\infty}^{\infty} e^{-\alpha t^2 - 2\delta t} dt = \sqrt{\frac{\pi}{\alpha}} e^{\frac{\delta^2}{\alpha}}.$$
 (C.4)

Hence,

$$\mathbb{E}\left\{\frac{\cosh[2\beta_i u_k]}{F_{\theta}(u_k)}\right\} = \frac{2}{\sqrt{M}}e^{\gamma_i|b|^2} \tag{C.5}$$

and

$$\mathbb{E}\left\{B_{k}^{(11)}\right\} = \frac{8A^{2}}{\sqrt{M}}\Re\{s_{1k}^{*}b\}^{2}\sum_{i=1}^{2^{p-1}}\beta_{i}^{2}.$$
(C.6)

Using very similar steps, we can also show that

$$\mathbb{E}\left\{H_k^{(11)}\right\} = \frac{8A^2}{\sqrt{M}}\Im\{s_{1k}^*b\}^2 \sum_{i=1}^{2^{p-1}}\beta_i^2.$$
(C.7)

C.2 Proof of Eq. (5.37)

In this appendix, we show that $\mathbb{E}\left\{\frac{\partial^2 \mathcal{L}(\boldsymbol{q};\boldsymbol{\theta})}{\partial a_R \partial a_I}\right\} = 0$. We have

$$\mathbb{E}\left\{\frac{\partial^2 \mathcal{L}(\boldsymbol{q};\boldsymbol{\theta})}{\partial a_R \partial a_I}\right\} = \sum_{k=1}^N \mathbb{E}\left\{B_k^{(12)} - G_k^{(12)}\right\} + \sum_{k=1}^N \mathbb{E}\left\{H_k^{12} - W_k^{(12)}\right\}.$$
 (C.8)

We can prove that $\mathbb{E}\left\{\frac{\partial^2 \mathcal{L}(\boldsymbol{q};\boldsymbol{\theta})}{\partial a_R \partial a_I}\right\} = 0$ by showing that $\mathbb{E}\left\{B_k^{(12)}\right\} = -\mathbb{E}\left\{H_k^{(12)}\right\}$ and $\mathbb{E}\left\{G_k^{(12)}\right\} = -\mathbb{E}\left\{W_k^{(12)}\right\}$. We have

$$\frac{\partial^2 F_{\theta}(u_k)}{\partial a_R \partial a_I} = 4A^2 \Re\{s_{1k}^* b\} \Im\{s_{1k}^* b\} \sum_{i=1}^{2^{p-1}} \beta_i^2 e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k]$$
(C.9)

and

$$\frac{\partial^2 F_{\theta}(v_k)}{\partial a_R \partial a_I} = -4A^2 \Re\{s_{1k}^*b\} \Im\{s_{1k}^*b\} \sum_{i=1}^{2^{p-1}} \beta_i^2 e^{-\gamma_i |b|^2} \cosh[2\beta_i v_k].$$
(C.10)

Hence,

$$\mathbb{E}\left\{B_{k}^{(12)}\right\} = 4A^{2}\Re\{s_{1k}^{*}b\}\Im\{s_{1k}^{*}b\}\sum_{i=1}^{2^{p-1}}\beta_{i}^{2}e^{-\gamma_{i}|b|^{2}}\mathbb{E}\left\{\frac{\cosh[2\beta_{i}u_{k}]}{F_{\theta}(u_{k})}\right\},\qquad(C.11)$$

and

$$\mathbb{E}\left\{H_{k}^{(12)}\right\} = -4A^{2}\Re\{s_{1k}^{*}b\}\Im\{s_{1k}^{*}b\}\sum_{i=1}^{2^{p-1}}\beta_{i}^{2}e^{-\gamma_{i}|b|^{2}}\mathbb{E}\left\{\frac{\cosh[2\beta_{i}v_{k}]}{F_{\theta}(v_{k})}\right\}.$$
(C.12)

Since u_k and v_k are i.i.d., it is clear that $\mathbb{E}\left\{B_k^{(12)}\right\} = -\mathbb{E}\left\{H_k^{(12)}\right\}$. It can be shown in a similar manner that $\mathbb{E}\left\{G_k^{(12)}\right\} = -\mathbb{E}\left\{W_k^{(12)}\right\}$.

C.3 Proof of Eq. (5.43)

In this appendix, we prove (5.43). The second derivative of $F_{\theta}(u_k)$ with respect to b_R is given by

$$\frac{\partial^2 F_\theta(u_k)}{\partial b_R^2} = T_1 + T_2 + T_3 \tag{C.13}$$

where

$$T_{1} \triangleq \sum_{i=1}^{2^{p-1}} (4\gamma_{i}^{2}b_{R}^{2} - 2\gamma_{i})e^{-\gamma_{i}|b|^{2}}\cosh[2\beta_{i}(b_{R}x_{k} + b_{I}y_{k})]$$

$$T_{2} \triangleq -\sum_{i=1}^{2^{p-1}} 8\beta_{i}\gamma_{i}b_{R}x_{k}e^{-\gamma_{i}|b|^{2}}\sinh[2\beta_{i}(b_{R}x_{k} + b_{I}y_{k})]$$

$$T_{3} \triangleq \sum_{i=1}^{2^{p-1}} 4\beta_{i}^{2}x_{k}^{2}e^{-\gamma_{i}|b|^{2}}\cosh[2\beta_{i}(b_{R}x_{k} + b_{I}y_{k})].$$
(C.14)

We next find $\mathbb{E}\left\{\frac{T_1}{F_{\theta}(u_k)}\right\}$, $\mathbb{E}\left\{\frac{T_2}{F_{\theta}(u_k)}\right\}$, and $\mathbb{E}\left\{\frac{T_3}{F_{\theta}(u_k)}\right\}$. Using (C.3), we obtain

$$\mathbb{E}\left\{\frac{T_1}{F_{\theta}(u_k)}\right\} = \frac{1}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} (8\gamma_i^2 b_R^2 - 4\gamma_i).$$
(C.15)

We next consider $\mathbb{E}\left\{\frac{T_2}{F_{\theta}(u_k)}\right\}$. We have

$$\mathbb{E}\left\{\frac{T_2}{F_{\theta}(u_k)}\right\} = -\sum_{i=1}^{2^{p-1}} 8\beta_i \gamma_i b_R e^{-\gamma_i |b|^2} \mathbb{E}\left\{\frac{x_k \sinh[2\beta_i (b_R x_k + b_I y_k)]}{F_{\theta}(b_R x_k + b_I y_k)}\right\}$$
(C.16)

Using the PDF $f_{X,Y}(x,y)$ in (5.21), we obtain

$$\mathbb{E}\left\{\frac{x_{k}\sinh\left[2\beta_{i}(b_{R}x_{k}+b_{I}y_{k})\right]}{F_{\theta}(b_{R}x_{k}+b_{I}y_{k})}\right\} = \frac{4}{\pi MC} \iint_{-\infty}^{\infty} x\sinh\left[2\beta_{i}(b_{R}x+b_{I}y)\right]F_{\theta}(b_{I}x-b_{R}y)e^{-\frac{x^{2}+y^{2}}{C}}dxdy$$

$$= \frac{1}{\pi MC}\sum_{\ell=1}^{2^{p-1}} e^{-\gamma_{\ell}|b|^{2}} \iint_{-\infty}^{\infty} x\left(e^{2\beta_{i}(b_{R}x+b_{I}y)}-e^{-2\beta_{i}(b_{R}x+b_{I}y)}\right)\left(e^{2\beta_{\ell}(b_{I}x-b_{R}y)}+e^{-2\beta_{\ell}(b_{I}x-b_{R}y)}\right)e^{-\frac{x^{2}+y^{2}}{C}}dxdy$$

$$= \frac{1}{\pi MC}\sum_{\ell=1}^{2^{p-1}} e^{-\gamma_{\ell}|b|^{2}} \left[\int_{-\infty}^{\infty} xe^{-\frac{x^{2}}{C}+(2\beta_{i}b_{R}+2\beta_{\ell}b_{I})x}dx\int_{-\infty}^{\infty} e^{-\frac{y^{2}}{C}+(2\beta_{i}b_{I}-2\beta_{\ell}b_{R})y}dy$$

$$+ \int_{-\infty}^{\infty} xe^{-\frac{x^{2}}{C}+(2\beta_{i}b_{R}-2\beta_{\ell}b_{I})x}dx\int_{-\infty}^{\infty} e^{-\frac{y^{2}}{C}+(2\beta_{i}b_{I}+2\beta_{\ell}b_{R})y}dy$$

$$- \int_{-\infty}^{\infty} xe^{-\frac{x^{2}}{C}-(2\beta_{i}b_{R}-2\beta_{\ell}b_{I})x}dx\int_{-\infty}^{\infty} e^{-\frac{y^{2}}{C}-(2\beta_{i}b_{I}+2\beta_{\ell}b_{R})y}dy$$

$$- \int_{-\infty}^{\infty} xe^{-\frac{x^{2}}{C}-(2\beta_{i}b_{R}+2\beta_{\ell}b_{I})x}dx\int_{-\infty}^{\infty} e^{-\frac{y^{2}}{C}-(2\beta_{i}b_{I}-2\beta_{\ell}b_{R})y}dy$$

$$(C.17)$$

In the above expression, the single integrals in x can be evaluated using the following result which holds for $\alpha > 0$ [80]:

$$\int_{-\infty}^{\infty} t e^{-\alpha t^2 - 2\delta t} dt = -\sqrt{\frac{\pi}{\alpha^3}} \delta e^{\frac{\delta^2}{\alpha}}$$
(C.18)

while the single integrals in y can be evaluated using (C.4). After evaluating all the single integrals in (C.17), we finally obtain

$$\mathbb{E}\left\{\frac{x\sinh[2\beta_i(b_Rx+b_Iy)]}{F_{\theta}(b_Rx+b_Iy)}\right\} = \frac{4C}{M}2^{p-1}e^{\gamma_i|b|^2}\beta_i b_R.$$
(C.19)

Therefore

$$\mathbb{E}\left\{\frac{T_2}{F_{\theta}(u_k)}\right\} = -\frac{16}{\sqrt{M}} \sum_{i=1}^{2^{P-1}} \gamma_i^2 b_R^2.$$
(C.20)

We next consider $\mathbb{E}\left\{\frac{T_3}{F_{\theta}(u_k)}\right\}$. We have

$$\mathbb{E}\left\{\frac{T_3}{F_{\theta}(u_k)}\right\} = \sum_{i=1}^{2^{p-1}} 4\beta_i^2 e^{-\gamma_i |b|^2} \mathbb{E}\left\{\frac{x_k^2 \cosh[2\beta_i (b_R x_k + b_I y_k)]}{F_{\theta}(b_R x_k + b_I y_k)}\right\}.$$
 (C.21)

To evaluate (C.21), we can follow the same approach that we used in (C.17). However, instead of using (C.18), we would use the following result, which holds for $\alpha > 0$ [80]:

$$\int_{-\infty}^{\infty} t^2 e^{-\alpha t^2 - 2\delta t} dt = \sqrt{\frac{\pi}{\alpha^5}} \delta^2 e^{\frac{\delta^2}{\alpha}} + \frac{1}{2} \sqrt{\frac{\pi}{\alpha^3}} e^{\frac{\delta^2}{\alpha}}.$$
 (C.22)

After some calculations (the details are skipped for brevity) we obtain

$$\mathbb{E}\left\{\frac{x_k^2 \cosh[2\beta_i (b_R x_k + b_I y_k)]}{F_\theta (b_R x_k + b_I y_k)}\right\} = e^{\gamma_i |b|^2} \left(\frac{C}{\sqrt{M}} + \frac{2C}{\sqrt{M}}\gamma_i b_R^2 + \frac{4C}{M} \sum_{\ell=1}^{2^{p-1}} \gamma_\ell b_I^2\right).$$
(C.23)

Hence,

$$\mathbb{E}\left\{\frac{T_3}{F_{\theta}(u_k)}\right\} = \frac{4}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} \gamma_i + \frac{8}{\sqrt{M}} \sum_{i=1}^{2^{p-1}} \gamma_i^2 b_R^2 + \frac{16}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_i \gamma_\ell \Im\{b\}^2.$$
(C.24)

Adding (C.15), (C.20) and (C.24), we finally obtain

$$\mathbb{E}\left\{B_{k}^{(33)}\right\} = \frac{16}{M} \sum_{i=1}^{2^{p-1}} \sum_{\ell=1}^{2^{p-1}} \gamma_{i} \gamma_{\ell} \Im\{b\}^{2}.$$
(C.25)

Following very similar steps, it can be shown that $\mathbb{E}\left\{H_k^{(33)}\right\} = \mathbb{E}\left\{B_k^{(33)}\right\}$.

C.4 Proof of Eq. (5.54)

In this appendix, we prove (5.54). Taking the derivative of $F_{\theta}(u_k)$ twice with respect to τ , we obtain:

$$\frac{\partial^2 F_{\theta}(u_k)}{\partial \tau^2} = \sum_{i=1}^{2^{p-1}} \left(A^4 \sigma^4 \beta_i^4 |b|^4 - 2 \frac{A^4 \sigma^4}{C} \beta_i^2 |b|^2 \right) \times e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k] \\ + \sum_{i=1}^{2^{p-1}} \left(4 \frac{A^4 \sigma^4}{C^2} \beta_i - 4 \frac{A^4 \sigma^4}{C} \beta_i^3 |b|^2 \right) u_k e^{-\gamma_i |b|^2} \sinh[2\beta_i u_k] + 4 \sum_{i=1}^{2^{p-1}} \frac{A^4 \sigma^4}{C^2} \beta_i^2 u_k^2 e^{-\gamma_i |b|^2} \cosh[2\beta_i u_k].$$
(C.26)

Since $B_k^{(55)} = \frac{\frac{\partial^2 F_{\theta}(u_k)}{\partial \tau^2}}{F_{\theta}(u_k)}$, it is clear from (C.26) that we need to evaluate $\mathbb{E}\left\{\frac{\cosh[2\beta_i u_k]}{F_{\theta}(u_k)}\right\}$, $\mathbb{E}\left\{\frac{u_k \sinh[2\beta_i u_k]}{F_{\theta}(u_k)}\right\}$ and $\mathbb{E}\left\{\frac{u_k^2 \cosh[2\beta_i u_k]}{F_{\theta}(u_k)}\right\}$. The first term is found in (C.5). For the second term, we have

$$\mathbb{E}\left\{\frac{u_k \sinh[2\beta_i u_k]}{F_{\theta(u_k)}}\right\} = \frac{2}{\sqrt{\pi MC|b|^2}} \int_{-\infty}^{\infty} t \sinh[2\beta_i t] e^{-\frac{t^2}{C|b|^2}} dt.$$
(C.27)

The above integral can be evaluated using the result in (C.18), thus obtaining

$$\mathbb{E}\left\{\frac{u_k \sinh[2\beta_i u_k]}{F_{\theta(u_k)}}\right\} = \frac{2}{\sqrt{M}} C|b|^2 e^{\gamma_i |b|^2} \beta_i.$$
(C.28)

For the third term, we have

$$\mathbb{E}\left\{\frac{u_k^2\cosh[2\beta_i u_k]}{F_\theta(u_k)}\right\} = \frac{2}{\sqrt{\pi MC|b|^2}} \int_{-\infty}^{\infty} t^2 \sinh[2\beta_i t] e^{-\frac{t^2}{C|b|^2}} dt.$$
(C.29)

We evaluate the above integral using (C.22), obtaining

$$\mathbb{E}\left\{\frac{u_k^2 \cosh[2\beta_i u_k]}{F_{\theta}(u_k)}\right\} = \frac{1}{\sqrt{M}} e^{\gamma_i |b|^2} (2C^2 |b|^4 \beta_i^2 + C|b|^2).$$
(C.30)

Using (C.5), (C.28) and (C.30), we finally see that

$$\mathbb{E}\left\{B_{k}^{(55)}\right\} = \sum_{i=1}^{2^{p-1}} \frac{2A^{4}\sigma^{4}}{\sqrt{M}} \left(\beta_{i}^{4}|b|^{4} + \frac{4}{C}\beta_{i}^{2}|b|^{2}\right).$$
(C.31)

The above derivation can be replicated by replacing u_k with v_k to show that $\mathbb{E}\left\{H_k^{(55)}\right\} = \mathbb{E}\left\{B_k^{(55)}\right\}$.

C.5 Derivation of Eqs. (5.65) and (5.66)

In this appendix, we derive closed-form approximations for the integral Υ in (5.34) in the high SNR regime for modulation orders M = 4 and M = 16. We have

$$\Upsilon = \frac{8A^2}{\sqrt{\pi MC|b|^2}} \Xi \tag{C.32}$$

where

$$\Xi = \int_{-\infty}^{\infty} \frac{\mathcal{P}^{2}(t)}{F_{\theta}(t)} e^{-\frac{t^{2}}{C|b|^{2}}} dt = \int_{-\infty}^{\infty} \frac{\sum_{i=1}^{2^{p-1}} \left(\beta_{i}e^{-\gamma_{i}|b|^{2}}\sinh(2\beta_{i}t)\right)^{2}}{\sum_{i=1}^{2^{p-1}} e^{-\gamma_{i}|b|^{2}}\cosh(2\beta_{i}t)} e^{\frac{-t^{2}}{C|b|^{2}}} dt$$

$$= 2\int_{0}^{\infty} \frac{\sum_{i=1}^{2^{p-1}} \left(\beta_{i}e^{-\gamma_{i}|b|^{2}}\sinh(2\beta_{i}t)\right)^{2}}{\sum_{i=1}^{2^{p-1}} e^{-\gamma_{i}|b|^{2}}\cosh(2\beta_{i}t)} e^{\frac{-t^{2}}{C|b|^{2}}} dt.$$
(C.33)

Since $\beta_i > 0$, the following approximation holds at high SNR for t > 0

$$\cosh(2\beta_i t) \approx \sinh(2\beta_i t) \approx \frac{1}{2}e^{2\beta_i t}.$$
 (C.34)

Using (C.34), we can approximate Ξ at high SNR as

$$\Xi \approx 2 \int_0^\infty \frac{\sum_{i=1}^{2^{p-1}} \left(\frac{1}{2}\beta_i e^{-\gamma_i |b|^2} e^{2\beta_i t}\right)^2}{\sum_{i=1}^{2^{p-1}} \frac{1}{2} e^{-\gamma_i |b|^2} e^{2\beta_i t}} e^{\frac{-t^2}{C|b|^2}} dt.$$
(C.35)

The integral in (C.35) depends on the modulation order M, which determines the number of terms in the numerator and denominator of the integrand. As we shall see next, the integral can be evaluated in closed-form for M = 4, and it can be accurately approximated for M = 16. For M = 4, (C.35) becomes

$$\Xi \approx 2 \int_0^\infty \frac{1}{2} \beta_1^2 e^{-\gamma_1 |b|^2} e^{2\beta_1 t} e^{\frac{-t^2}{C|b|^2}} dt.$$
(C.36)

Moreover, for $\alpha > 0$ we have

$$\int_{0}^{\infty} e^{-\alpha t^{2} + 2\delta t} dt = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}} e^{\frac{\delta^{2}}{\alpha}} \left(1 + \operatorname{erf}\left(\sqrt{\frac{\delta^{2}}{\alpha}}\right) \right).$$
(C.37)

We thus obtain

$$\Xi \approx \frac{1}{2} \beta_1^2 \sqrt{\pi C |b|^2} \left(1 + \operatorname{erf}(\sqrt{\gamma_1 |b|^2}) \right), \qquad (C.38)$$

hence

$$\Upsilon \approx \frac{2A^4 d_p^2}{C^2} \left(1 + \operatorname{erf}\left(\sqrt{\gamma_1 |b|^2}\right) \right).$$
(C.39)

We now consider the case M = 16. In this case, (C.35) becomes

$$\begin{split} \Xi \approx & 2 \int_{0}^{\infty} \frac{\left(\frac{1}{2}\beta_{1}e^{-\gamma_{1}|b|^{2}}e^{2\beta_{1}t} + \frac{1}{2}\beta_{2}e^{-\gamma_{2}|b|^{2}}e^{2\beta_{2}t}\right)^{2}}{\frac{1}{2}e^{-\gamma_{1}|b|^{2}}e^{2\beta_{1}t} + \frac{1}{2}e^{-\gamma_{2}|b|^{2}}e^{2\beta_{2}t}}e^{\frac{-t^{2}}{C|b|^{2}}}dt \\ &= \frac{A^{2}d_{p}^{2}}{C^{2}}\int_{0}^{\infty} \frac{\left(e^{-\gamma_{1}|b|^{2}}e^{2\beta_{1}t} + 3e^{-\gamma_{2}|b|^{2}}e^{2\beta_{2}t}\right)^{2}}{e^{-\gamma_{1}|b|^{2}}e^{2\beta_{1}t} + e^{-\gamma_{2}|b|^{2}}e^{2\beta_{2}t}}e^{\frac{-t^{2}}{C|b|^{2}}}dt \\ &= \frac{A^{2}d_{p}^{2}}{C^{2}}\int_{0}^{\infty} \left[e^{-\gamma_{1}|b|^{2}}e^{2\beta_{1}t} + 5e^{-\gamma_{2}|b|^{2}}e^{2\beta_{2}t} + \frac{4e^{-2\gamma_{2}|b|^{2}}e^{4\beta_{2}t}}{e^{-\gamma_{1}|b|^{2}}e^{2\beta_{2}t}}\right]e^{\frac{-t^{2}}{C|b|^{2}}}dt. \end{split}$$
(C.40)

As t increases, the term $e^{-\gamma_2|b|^2}e^{2\beta_2 t}$ dominates the term $e^{-\gamma_1|b|^2}e^{2\beta_1 t}$, and we may use the approximation

$$\frac{4e^{-2\gamma_2|b|^2}e^{4\beta_2t}}{e^{-\gamma_1|b|^2}e^{2\beta_1t} + e^{-\gamma_2|b|^2}e^{2\beta_2t}} \approx 4e^{-\gamma_2|b|^2}e^{2\beta_2t}.$$
 (C.41)

to obtain

$$\Xi \approx \frac{A^2 d_p^2}{C^2} \int_0^\infty \left(e^{-\gamma_1 |b|^2} e^{2\beta_1 t} + 9e^{-\gamma_2 |b|^2} e^{2\beta_2 t} \right) e^{\frac{-t^2}{C|b|^2}} dt$$

$$= \frac{A^2 d_p^2}{2C^2} \sqrt{\pi C |b|^2} \left(10 + \operatorname{erf}\left(\sqrt{\gamma_1 |b|^2}\right) + 9\operatorname{erf}\left(\sqrt{\gamma_2 |b|^2}\right) \right)$$
(C.42)

Hence,

$$\Upsilon \approx \frac{A^4 d_p^2}{C^2} \left(10 + \operatorname{erf}\left(\sqrt{\gamma_1 |b|^2}\right) + 9\operatorname{erf}\left(\sqrt{\gamma_2 |b|^2}\right) \right).$$
(C.43)

Appendix D

Derivations of Select Proofs and Results from Chapter 7

D.1 BFGS Method

The BFGS method for optimization belongs to the class of quasi-Newton optimization methods [75]. The classical Newton method [74] requires the evaluation of the inverse of the Hessian matrix at each iteration. The BFGS method avoids the high complexity of evaluating the inverse Hessian by approximating it using rank-two updates that only require the the evaluation of the gradient. The following pseudo-code describes the steps of the BFGS method.

Starting with an initial estimate \boldsymbol{x}_0 and an approximation \boldsymbol{R}_0 of the Hessian matrix the following steps are repeated until convergence: 1. Obtain the search direction: $\boldsymbol{p}_k = -\boldsymbol{R}_k^{-1}\nabla f(\boldsymbol{x}_k)$ 2. Use linesearch to find the step size α_k . 3. Update the estimate: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$. 3. Set $\boldsymbol{u}_k = \alpha_k \boldsymbol{p}_k$, $\boldsymbol{v}_k = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$ 4. Update the inverse Hessian approximation: $\boldsymbol{R}_{k+1}^{-1} = \boldsymbol{R}_k^{-1} + \frac{(\boldsymbol{u}_k^T \boldsymbol{v}_k + \boldsymbol{v}_k^T \boldsymbol{R}_k^{-1} \boldsymbol{v}_k) \boldsymbol{u}_k \boldsymbol{u}_k^T}{(\boldsymbol{u}_k^T \boldsymbol{v}_k)^2} - \frac{\boldsymbol{R}_k^{-1} \boldsymbol{v}_k \boldsymbol{u}_k^T + \boldsymbol{u}_k \boldsymbol{v}_k^T \boldsymbol{R}_k^{-1}}{\boldsymbol{u}_k^T \boldsymbol{v}_k}$

D.2 Proof of Theorem 3

We let

$$\boldsymbol{D} \triangleq \boldsymbol{\Omega}^{H} \boldsymbol{\Omega} = \begin{bmatrix} A^{2} \boldsymbol{S}_{1}^{H} \boldsymbol{S}_{1} & A^{2} \boldsymbol{S}_{1}^{H} \boldsymbol{S}_{2} & A \boldsymbol{S}_{1}^{H} \boldsymbol{S}_{3} \\ A^{2} \boldsymbol{S}_{2}^{H} \boldsymbol{S}_{1} & A^{2} \boldsymbol{S}_{2}^{H} \boldsymbol{S}_{2} & A \boldsymbol{S}_{2}^{H} \boldsymbol{S}_{3} \\ A \boldsymbol{S}_{3}^{H} \boldsymbol{S}_{1} & A \boldsymbol{S}_{3}^{H} \boldsymbol{S}_{2} & \boldsymbol{S}_{3}^{H} \boldsymbol{S}_{3} \end{bmatrix} .$$
(D.1)

Assuming that Ω is full-column rank, D is positive definite, and the MSE of the LS estimator is given by tr(D^{-1}). Applying the Cauchy-Schwartz inequality, we can lower-bound this MSE as follows [99]

$$\operatorname{tr}\left(\boldsymbol{D}^{-1}\right) \geq \sum_{i=1}^{M_1+M_2+L_2} \frac{1}{[\boldsymbol{D}]_{ii}} \tag{D.2}$$

where the equality holds if and only if D is diagonal. Furthermore, using the circulant property of the matrices S_1 , S_2 and S_3 , we may express them as

$$egin{aligned} oldsymbol{S}_1 &= \sqrt{N}oldsymbol{F}^H ilde{oldsymbol{S}}_1 oldsymbol{F}_{M_1}, \ oldsymbol{S}_2 &= \sqrt{N}oldsymbol{F}^H ilde{oldsymbol{S}}_2 oldsymbol{F}_{M_2}, \ oldsymbol{S}_3 &= \sqrt{N}oldsymbol{F}^H ilde{oldsymbol{S}}_3 oldsymbol{F}_{L_2}, \end{aligned}$$

where \mathbf{F}_m is the $N \times m$ matrix that has the first m columns of \mathbf{F} and $\tilde{\mathbf{S}}_1$, $\tilde{\mathbf{S}}_2$, $\tilde{\mathbf{S}}_3$ are the $N \times N$ diagonal matrices with diagonal elements $\tilde{\mathbf{t}}_1$, $\tilde{\mathbf{t}}_2$ and $\tilde{\mathbf{t}}_3$, respectively. Hence, $\mathbf{S}_1^H \mathbf{S}_1 = N \mathbf{F}_{M_1}^H \tilde{\mathbf{S}}_1^H \tilde{\mathbf{S}}_1 \mathbf{F}_{M_1}$, $\mathbf{S}_2^H \mathbf{S}_2 = N \mathbf{F}_{M_2}^H \tilde{\mathbf{S}}_2^H \tilde{\mathbf{S}}_2 \mathbf{F}_{M_2}$, and $\mathbf{S}_3^H \mathbf{S}_3 = N \mathbf{F}_{L_2}^H \tilde{\mathbf{S}}_3^H \tilde{\mathbf{S}}_3 \mathbf{F}_{L_2}$. Moreover, it can be easily verified that the diagonal elements of the matrix $\mathbf{S}_1^H \mathbf{S}_1$ are all equal to $\operatorname{tr}(\tilde{\mathbf{S}}_1^H \tilde{\mathbf{S}}_1) = N P_1$. Similarly, those of $\mathbf{S}_2^H \mathbf{S}_2$ are equal to $\operatorname{tr}(\tilde{\mathbf{S}}_2^H \tilde{\mathbf{S}}_2) = N P_2$ and those of $\mathbf{S}_3^H \mathbf{S}_3$ are equal to $\operatorname{tr}(\tilde{\mathbf{S}}_3^H \tilde{\mathbf{S}}_1) = N P_3$. We can thus rewrite the RHS of the inequality in (D.2) as

$$\sum_{i=1}^{M_1+M_2+L_2} \frac{1}{[\boldsymbol{D}]_{ii}} = \frac{M_1}{NA_p^2 P_1} + \frac{M_2}{NA_p^2 P_2} + \frac{L_2}{NP_3}.$$
 (D.3)

Clearly, the MSE of the LS estimator is smallest when the inequality in (D.2) holds with equality, i.e., when D is diagonal. This happens if and only if the following conditions

hold:

$$S_1^H S_2 = 0, \ S_1^H S_3 = 0, \ S_2^H S_3 = 0,$$
 (D.4)

$$\boldsymbol{F}_{M_1}^H \tilde{\boldsymbol{S}}_1^H \tilde{\boldsymbol{S}}_1 \boldsymbol{F}_{M_1} = N P_1 \boldsymbol{I}_{M_1}, \qquad (D.5)$$

$$\boldsymbol{F}_{M_2}^H \tilde{\boldsymbol{S}}_2^H \tilde{\boldsymbol{S}}_2 \boldsymbol{F}_{M_2} = N P_2 \boldsymbol{I}_{M_2}, \tag{D.6}$$

$$\boldsymbol{F}_{L_2}^H \tilde{\boldsymbol{S}}_3^H \tilde{\boldsymbol{S}}_3 \boldsymbol{F}_{L_2} = N P_3 \boldsymbol{I}_{L_2}. \tag{D.7}$$

The conditions (D.5), (D.6) and (D.7) are satisfied when all the individual frequency domain symbols within each pilot vector have the same power, i.e., when

$$|\tilde{t}_{ji}|^2 = P_j, \text{ for } i = 1, \dots, N, \ j = 1, 2, 3.$$
 (D.8)

Moreover, the conditions in (D.4) are equivalent to the following three conditions:

$$\sum_{i=1}^{N} \tilde{t}_{1i}^* \tilde{t}_{2i} e^{j2\pi(i-1)m_1/N} = 0, \ \forall m_1 \in \{1 - M_2, \dots, M_1 - 1\},$$
(D.9)

$$\sum_{i=1}^{N} \tilde{t}_{1i}^* \tilde{t}_{3i} e^{j2\pi(i-1)m_2/N} = 0, \ \forall m_2 \in \{1 - L_2, \dots, M_1 - 1\},$$
(D.10)

$$\sum_{i=1}^{N} \tilde{t}_{2i}^{*} \tilde{t}_{3i} e^{j2\pi(i-1)m_3/N} = 0, \ \forall m_3 \in \{1 - L_2, \dots, M_2 - 1\}.$$
 (D.11)

The following is an example of training vectors \tilde{t}_1 , \tilde{t}_2 and \tilde{t}_3 that satisfy the conditions in (D.8)–(D.11), which is inspired by the proposed pilot design for the case of two pilot vectors in [22]. We let $\tilde{t}_{1i} = \sqrt{P_1}$, $\tilde{t}_{2i} = \sqrt{P_2}e^{j2\pi(i-1)\kappa/N}$, $\tilde{t}_{3i} = \sqrt{P_3}e^{j2\pi(i-1)\delta/N}$, where κ and δ satisfy the following constraints

$$\kappa \in \{M_2, \dots, N - M_1\}\tag{D.12}$$

$$\delta \in \{L_2, \dots, N - M_1\} \tag{D.13}$$

$$\kappa - \delta \in \{L_2 \dots, N - M_2\}. \tag{D.14}$$

For the existence of integer values κ and δ that simultaneously satisfy the above constraints,

the number of carriers ${\cal N}$ should satisfy

$$N \ge \max(M_1 + M_2, M_1 + L_2),$$
 (D.15)

and

$$\max(L_2, M_1 + M_2 - N) \le \min(N - M_2, N - M_1 - L_2)$$
(D.16)

We can always find a sufficiently large N to satisfy these two constraints.

References

- [1] E. Halepovic, C. Williamson, and M. Ghaderi, "Wireless data traffic: a decade of change," *IEEE Network*, vol. 23, no. 2, pp. 20–26, Mar. 2009.
- [2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016," *Cisco White Paper*, Feb. 2012.
- [3] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.
- [4] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [5] Q. Li, R. Q. Hu, Y. Qian, and G. Wu, "Cooperative communications for wireless networks: Techniques and applications in LTE-Advanced systems," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 22–29, Apr. 2012.
- [6] F. Boccardi, V. Braun, O. Stanze, M. Baker, A. Tulino, and A. Weber, "An industrial perspective of relaying for cellular systems," in *Proc. 44th Asilomar Conf. on Signals.*, *Syst., Comput.*, Pacific Grove, CA, Nov. 2010.
- [7] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, C. Koutsimanis, and X. Shen, "Relaying operation in 3GPP LTE: Challenges and solutions," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 156–162, Feb. 2012.
- [8] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.
- [9] —, "Spectral efficient signaling for half-duplex relay channels," in *Proc. 39th Asilo*mar Conf. on Signals, Syst. and Comput., Pacific Grove, CA, Oct. 2005, pp. 1066–1071.
- [10] S. Berger, M. Kuhn, A. Wittneben, T. Unger, and A. Klein, "Recent advances in amplify-and-forward two-hop relaying," *IEEE Commun. Mag.*, vol. 47, no. 7, pp. 50– 56, 2009.

- [11] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Glasgow, Scotland, Jun. 2007, pp. 707–712.
- [12] A. Argyriou and A. Pandharipande, "Cooperative protocol for analog network coding in distributed wireless networks," vol. 9, no. 10, pp. 3112–3119, Oct. 2010.
- [13] P. Larsson, N. Johansson, and K.-E. Sunell, "Coded bi-directional relaying," in Proc. IEEE 63rd Vehic. Techol. Conf (VTC), Melbourne, Australia, May 2006.
- [14] Y. U. Jang, E. R. Jeong, and Y. Lee, "A two-step approach to power allocation for OFDM signals over two-way amplify-and-forward relay," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2426–2430, Apr. 2010.
- [15] Y. Zhang, Y. Ma, and R. Tafazolli, "Power allocation for bidirectional AF relaying over Rayleigh fading channels," *IEEE Commun. Lett.*, vol. 14, no. 2, pp. 145–147, Feb. 2010.
- [16] L. Song, "Relay selection for two-way relaying with amplify-and-forward protocols," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1954–1959, May 2011.
- [17] J. Zheng, B. Bai, and Y. Li, "Outage-optimal opportunistic relaying for two-way amplify and forward relay channel," *Electron. Lett.*, vol. 46, pp. 595–596, Apr. 2010.
- [18] V. Havary-Nassab, S. Shahbazpanahi, and A. Grami, "Optimal distributed beamforming for two-way relay networks," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1238–1250, Mar. 2010.
- [19] M. Zeng, R. Zhang, and S. Cui, "On design of collaborative beamforming for two-way relay networks," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2284–2295, May 2011.
- [20] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. New York: Wiley, 2005.
- [21] F. Gao, R. Zhang, and Y.-C. Liang, "Optimal channel estimation and training design for two-way relay networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3024–3033, Oct. 2009.
- [22] —, "Channel estimation for OFDM modulated two-way relay networks," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4443–4455, Nov. 2009.
- [23] K. Iniewski, *Wireless Technologies: Circuits, Systems and Devices.* Vancouver, Canada: CRC Press, 2007.
- [24] S. Haile, "Investigation of channel reciprocity for OFDM TDD systems," Master's thesis, Dept. of Elec. Eng., University of Waterloo, Waterloo, Ont., Canada.

- [25] Z.-J. Liu, D.-C. Sun, J.-L. Wang, and K.-C. Yi, "Impact and compensation of I/Q imbalance on channel reciprocity of time-division-duplexing multiple-input multipleoutput systems," *IET Commun.*, vol. 7, no. 7, pp. 663–672, Feb. 2013.
- [26] J. Haartsen, "Impact of non-reciprocal channel conditions in broadband TDD systems," in *Proc. IEEE Internation. Symp. Indoor Mobile Radio Commun. (PIMRC)*, Cannes, France, Sep. 2008.
- [27] M. Sanchez-Fernandez and A. Lozano, "Doppler sensitivity of link reciprocity in TDD MIMO systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, St. Louis, Missouri, Nov. 2005.
- [28] B. Jiang, F. Gao, X. Gao, and A. Nallanathan, "Channel estimation and training design for two-way relay networks with power allocation," *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2022–2032, Jun. 2010.
- [29] F. Roemer and M. Haardt, "Tensor-based channel estimation (TENCE) and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5720–5735, Nov. 2010.
- [30] T.-H. Pham, Y.-C. Liang, A. Nallanathan, and H. K. Garg, "On the design of optimal training sequence for bi-directional relay networks," *IEEE Signal Process. Lett.*, vol. 16, no. 3, pp. 200–203, Mar. 2009.
- [31] —, "Optimal training sequences for channel estimation in bi-directional relay networks with multiple antennas," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 474–479, Feb. 2010.
- [32] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, "Joint CFO and channel estimation for OFDM-based two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 456–465, Feb. 2011.
- [33] —, "Joint CFO and channel estimation for ZP-OFDM modulated two-way relay networks," in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Sydney, Australia, Apr. 2010.
- [34] G. Wang, F. Gao, X. Zhang, and C. Tellambura, "Superimposed training-based joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," *EURASIP J. Wireless Commun. and Network.*, vol. 2010, Jul. 2010, ID=403936.
- [35] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Englewood Cliffs, NJ: Prentice-Hall, 1993.

- [36] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.
- [37] J. J. van de Beek, M. Sandell, and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1800–1805, Jul. 1997.
- [38] J. K. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, no. 9, pp. 413–415, Sep. 2003.
- [39] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.
- [40] S. Zhou and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1402– 1414, Aug. 2001.
- [41] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699 –1712, Jul. 2002.
- [42] C. Shin, R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.
- [43] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on secondorder statistics: a time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.
- [44] J. Cardoso, E. Moulines, P. Duhamel, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
- [45] K. Berberidis, B. Champagne, G. Moustakides, H. Poor, and P. Stoica, "Advances in subspace-based techniques for signal processing and communications," *EURASIP J. Adv. Signal Process.*, vol. 2007, Jan. 2007, ID=48612.
- [46] V. Buchoux, O. Cappe, E. Moulines, and A. Gorokhov, "On the performance of semiblind subspace-based channel estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1750 –1759, Jun. 2000.
- [47] H. A. Cirpan and M. K. Tsatsanis, "Stochastic maximum likelihood methods for semiblind channel estimation," *IEEE Signal Process. Lett.*, vol. 5, no. 1, pp. 21–24, Jan. 1998.

- [48] E. de Carvalho and D. T. Slock, "Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1053–1064, Apr. 2004.
- [49] A. Kocian and B. H. Fleury, "EM-based joint data detection and channel estimation of DS-CDMA signals," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709–1720, Oct. 2003.
- [50] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.
- [51] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, "Comparison of EM-based algorithms for MIMO channel estimation," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 216–226, Jan. 2007.
- [52] X. Liao, L. Fan, and F. Gao, "Blind channel estimation for OFDM modulated two-way relay network," in *Proc. 12th IEEE Wireless Commun. Netw. Conf. (WCNC)*, Sydney, Australia, Apr. 2010.
- [53] T.-H. Pham, Y.-C. Liang, H. K. Garg, and A. Nallanathan, "Joint channel estimation and data detection for MIMO-OFDM two-way relay networks," in *Proc. IEEE Global Telecommun. Conf. (Globecom)*, Miami, Florida, Dec. 2010.
- [54] F. Gini and R. Reggiannini, "On the use of Cramer-Rao-like bounds in the presence of random nuisance parameters," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2120–2126, Dec. 2000.
- [55] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove: CA: Duxbury Press, 2002.
- [56] H. V. Poor, Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1994.
- [57] T. K. Mak, "Estimation in the presence of incidental parameters," Canad. J. Statis., vol. 10, no. 2, pp. 121–132, 1982.
- [58] J. Neyman and E. L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica: J. Econometric Society*, vol. 16, no. 1, pp. 1–32, 1948.
- [59] V. Nagesha and S. Kay, "Maximum likelihood estimation for array processing in colored noise," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 169–180, Feb. 1996.
- [60] T. Lancaster, "The incidental parameter problem since 1948," J. Econometrics, vol. 95, no. 2, pp. 391–413, 2000.

- [61] E. de Carvalho and D. T. Slock, "Asymptotic performance of ML methods for semiblind channel estimation," in *Proc. 31st Asilomar Conf. on Signals.*, Syst., Comput., vol. 2, Pacific Grove, CA, 1997, pp. 1624–1628.
- [62] C. Wu, "On the convergence properties of the EM algorithm," Ann. Statist., vol. 11, no. 1, pp. 95–103, 1983.
- [63] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1391–1399, Feb/Mar/Apr 1994.
- [64] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramer-Rao bound in vector parameter estimation," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 52–60, Jan. 1998.
- [65] A. Molisch, Wireless Communications. Chichester, England: Wiley, 2009.
- [66] "Ets 300 401 (final edition): Radio broadcasting systems; digital audio broadcasting (DAB) to mobile, portable and fixed receivers," *Europ. Commun. Stand. Inst., Tech. Rep.*, 2006.
- [67] "Etsi en 302 755: Digital video broadcasting (DVB) frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB- T2)," Europ. Commun. Stand. Inst., Tech. Rep., 2009.
- [68] S. Srikanth, P. Murugesa Pandian, and X. Fernando, "Orthogonal frequency division multiple access in WiMAX and LTE: a comparison," *IEEE Commun. Mag.*, vol. 50, no. 9, pp. 153–161, Sep. 2012.
- [69] A. Furuskar, S. Mazur, F. Muller, and H. Olofsson, "EDGE: Enhanced data rates for GSM and TDMA/136 evolution," *IEEE Person. Commun.*, vol. 6, no. 3, pp. 56–66, Jun. 1999.
- [70] D. Martín-Sacristán, J. F. Monserrat, J. Cabrejas-Penuelas, D. Calabuig, S. Garrigas, and N. Cardona, "On the way towards fourth-generation mobile: 3GPP LTE and LTE-advanced," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, Aug. Arcticle ID 354089, 2009.
- [71] S. Abdallah and I. N. Psaromiligkos, "Blind channel estimation for MPSK-based amplify-and-forward two-way relaying," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011.
- [72] —, "Blind channel estimation for amplify-and-forward two-way relay networks employing M-PSK modulation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3604– 3615, Jul. 2012.

- [73] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 543–551, Jul. 1983.
- [74] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, UK, 2004.
- [75] J. Nocedal and S. Wright, Numerical Optimization. Berlin, Germany: Springer-Verlag, 1999.
- [76] E. de Carvalho and D. T. Slock, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *Proc. IEEE Int. Conf. Acoust.*, *Speech, Signal Process. (ICASSP)*, vol. 5, Munich, Germany, Apr. 1997.
- [77] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3132–3143, Oct. 2009.
- [78] F. Hayashi, *Econometrics*, 9th ed. New Jersey: Princeton University Press, 2000.
- [79] R. Engle and D. McFadden, Eds., Handbook of Econometrics. New York: Elsevier Science, 1994, vol. 4.
- [80] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, 9th ed. New York: Dover Publications, 1970.
- [81] A. Leon-Garcia, Probability, Statistics, and Random Processes for Electrical Engineering, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 2008.
- [82] Y. Chen and C. Tellambura, "Performance analysis of l-branch equal gain combiners in equally correlated Rayleigh fading channels," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 150–152, 2004.
- [83] U. Mengali and A. N. D'Andrea, Synchronization techniques for digital receivers. Plenum, NY: Springer, 1997.
- [84] S. Abdallah and I. N. Psaromiligkos, "Semi-blind channel estimation for amplify-andforward two-way relay networks employing constant-modulus constellations," in *Proc.* 44th Ann. Conf. on Infor. Scien. and Syst. (CISS), Princeton, NJ, Mar. 2010.
- [85] —, "Partially-blind estimation of reciprocal channels for AF two-way relay networks employing M-PSK modulation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1649–1654, May 2012.
- [86] G. L. Stuber, *Principles of Mobile Communication*, 3rd ed. Norwell, MA: Kluwer Academic Publishers, 2011.

- [87] F. Bellili, A. Stephenne, and S. Affes, "Cramer-Rao lower bounds for NDA SNR estimates of square QAM modulated transmissions," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3211–3218, Nov. 2010.
- [88] A. H. Stroud, *Approximate Calculation of Multiple Integrals*. Englewood Cliffs, NJ: Prentice Hall, 1971.
- [89] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.
- [90] S. Abdallah and I. N. Psaromiligkos, "EM-based semi-blind channel estimation in amplify-and-forward two-way relay networks," *IEEE Wireless Commun. Lett.*, accepted for publication, Jun. 2013.
- [91] G. Wang, F. Gao, and C. Tellambura, "Superimposed pilot-based joint CFO and channel estimation for CP-OFDM modulated two-way relay networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Miami, Florida, Dec. 2010.
- [92] S. Zhang, F. Gao, and C. X. Pei, "Optimal training design for individual channel estimation in two-way relay networks," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4987–4991, Sep. 2012.
- [93] K. J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.
- [94] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.
- [95] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
- [96] X. Xu, J. Wu, S. Ren, L. Song, and H. Xiang, "Superimposed training design based on Bayesian optimization for channel estimation in two-way relay networks," *IET Commun.*, vol. 6, no. 18, pp. 3131–3139, Sep. 2012.
- [97] M. Zeng, R. Zhang, and S. Cui, "On design of collaborative beamforming for two-way relay networks," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2284–2295, May 2011.
- [98] L. Oberto and F. Pennecchi, "Estimation of the modulus of a complex-valued quantity," *Metrologia*, vol. 43, pp. 531–538, 2006.
- [99] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.