On MIMO Wireless Eavesdrop Information Rates

John Kitchen

MAppSc(Communications)
BSc(Physics & Electronics)(Hons)

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy
February 2011

Department of Electrical and Electronic Engineering

The University of Melbourne
Victoria 3010, Australia

Dedicated to my Family and Friends

Abstract

Recent research in Multiple-Input Multiple-Output (MIMO) wireless communications techniques promises both opportunities and difficulties for receiving systems. Robustness and high data rates are offered at the cost of increased system complexity. This increased complexity presents new challenges for surveillance or passive eavesdropping receivers. This thesis addresses theory and issues arising in communications eavesdropping, with a particular emphasis on the recovery of data streams produced by a MIMO wireless transmission array.

An information theory for MIMO eavesdroppers is developed based on standard communications information theory and estimation techniques for informed and partially-informed MIMO wireless reception.

Existing literature, dealing with blind source separation and communications secrecy, is drawn upon to provide a context and background theory for the MIMO wireless communications eavesdropping problem. During the development of the background theory some deficiencies were identified and are addressed in this thesis. As a consequence, a number of original contributions have been made.

Expressions required for theoretical blind source separation performance bounds for a complex-valued channel and complex-valued sources did not exist in the literature and so a derivation of the Cramér-Rao Bound (CRB), for blind separation of complex sources linearly mixed by a complex channel matrix, is included here. The CRB is also derived as a function of the source probability density function (pdf) where the generalised gaussian distribution is employed to represent the source pdf.

In our source recovery model the two primary parameters are the complex-valued channel mixing matrix and the complex-valued sources. Use of maximum likelihood techniques, where either the source or channel is known, is compared with application of independent component analysis techniques, where neither

the source nor the channel are known. The theory and simulation results show that Blind Source Separation (BSS) is not possible when the sources are independent and identically distributed (i.i.d.) proper complex Gaussian, in which case an eavesdropper would obtain no additional information about the sources given only observations on the source mixture. These results provide a benchmark for the source recovery performance obtainable by the intended receiver and the eavesdropper respectively.

To model source dependence effects that may exist in the propagation channel, Copula theory is employed and an approach derived that incorporates fading effects as well as control over the type and level of dependence in the channel.

Finally the theory and techniques, developed in this work, are brought together to provide a method for channel-independent, complex symbol stream recovery for orthogonal Space-Time (ST) block-coded signals that shows how the observed signals may be transformed to improve the results of subsequent source separation processing.

Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the PhD except where indicated,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is less than 100,000 words in length, exclusive of tables, bibliographies, appendices and footnotes.

John Kitchen

Acknowledgements

Some time ago, in the course of discussion with Bill and Stephen, who were subsequently to become two of my supervisors, the suggestion that I might study for a PhD arose. Were it not for them I might never have taken this path. I would like to thank Professor Bill Moran for his motivation and guidance as principal supervisor over the past six years. I would like to extend my sincere gratitude to my two associate supervisors: Dr Stephen D. Howard and Dr Sofia Suvorova for their support.

I am grateful for the vision and leadership of Dr Warren Marwood, as Research Leader, and Dr John Asenstorfer, in Intelligence, Surveillance & Reconnaissance Division (ISRD), for allowing me the opportunity to take on this part-time PhD candidature whilst working at the Defence Science & Technology Organisation (DSTO). Curiously I now find myself in a different group/discipline, in a different Division, within DSTO but my physical location hasn't changed. Now under new management I would like to thank Dr Mike Davies and Dr Gareth Parker, in their roles as Research Leader and Head, respectively, for allowing me to continue on and complete this project.

I would like to thank my friends and colleagues for sharing thoughts, problems and ideas. In particular, I would like to thank Dr Darren Bachmann and Peter Sarunic.

John Kitchen,
Adelaide, South Australia,
February 2011

Publications

During the course of this project, a number of publications and public presentations have been made which are based on the work presented in this thesis. They are listed here for reference.

1. **J. Kitchen, W. Moran and S.D. Howard**, *Performance Bounds for Blind MIMO Estimation*. In Defence Applications of Signal Processing (DASP), Queensland, Australia, 10–14 December 2006.

2. **J. Kitchen**, *Intercept Capacity: unknown pre-processing*. In AMSI-MASCOS, Melbourne, 11th December, 2007.

3. **J. Kitchen, W. Moran and S.D. Howard**, *Intercept Capacity: Unknown Unitary Transformation*. Entropy, 10(4):722–735, November 2008.

4. **J. Kitchen, W. Moran**, *Effect of Source Kurtosis on MIMO Intercept Rate*. In Defence Applications of Signal Processing (DASP), Lihue, Hawaii, USA, September 2009.

5. **J. Kitchen, W. Moran**, *Copula Techniques for Modelling Signal Dependence in Wireless Communications*. In 9th Engineering Mathematics and Applications Conference, EMAC2009, Adelaide, Australia, 6th–9th December 2009.

6. **J. Kitchen, W. Moran**, *Copula techniques in wireless communications*. In Proceedings of the 9th Biennial Engineering Mathematics and Applications Conference, EMAC-2009, vol.51 of ANZIAM J., pages C526–C540, August 2010.

7. **J. Kitchen, W. Moran**, *Effect of source kurtosis on MIMO information rate*. Accepted for publication in the *D*igital Signal Processing Journal, 4th February 2011.

8. **J. Kitchen**, *Channel-independent symbol stream recovery for orthogonal space-time block-coded signals*. In *4th International Conference on Signal Processing and Communication Systems ICSPCS'2010*, Gold Coast, Australia, 13–15th December 2010.

9. **J. Kitchen**, *Cramér-Rao Bounds for Complex Linear Independent Component Analysis*, to be submitted.

Contents

List of Figures

List of Algorithms

Glossary

Alice	Conventional cryptographic label for message source
Bob	Conventional cryptographic label for intended message recipient
Cryptography	The discipline concerned with communication security
Equivocation	A theoretical secrecy index, a measure of conditional entropy
Eve	Conventional cryptographic label for message eavesdropper
Gaussianity	How well the pdf of a r.v. may be represented by a Normal pdf
Negentropy	Difference, in differential entropy, from a Gaussian density
Perfect secrecy	Information-theoretic notion of secrecy

Acronyms and Abbreviations

FASTICA	Fast ICA
JADE	Joint Approximate Diagonalization of Eigenvalues
MIBS	Mutual Information Between Sources
RADICAL	RADICAL algorithm
2-D	2-Dimensional
3-D	3-Dimensional
ACMA	Algebraic Constant Modulus Algorithm
AJD	Approximate Joint Diagonalization
AM	Amplitude Modulation
AMUSE	Algorithm for Multiple Unknown Signals Extraction
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keyed
BSC	Binary Symmetric Channel
BSS	Blind Source Separation
CCA	Canonical Correlation Analysis
cdf	cumulative distribution function
cf.	Latin: compare
CG	Conjugate Gradient
CM	Constant Modulus
CMA	Constant Modulus Algorithm

CRB	Cramér-Rao Bound
CRLVB	Cramér-Rao Lower Variance Bound
dB	decibel
DMC	Discrete Memoryless Channel
DSTO	Defence Science and Technology Organisation
e.g.	exempli gratia (Latin: for example)
EJD	Exact Joint Diagonalization
EM	Expectation-Maximization
EPI	Entropy Power Inequality
et al.	et alii (Latin: and others)
etc.	et cetera (Latin: and the others)
EVD	Eigenvalue Decomposition
FICA	Fast Independent Component Analysis
FicaCPLX	Complex Fast ICA
FIM	Fisher Information Matrix
GG	Generalised Gaussian
GM	Gaussian Mixture
HOS	Higher Order Statistics
Hz	Hertz
i.e.	id est (Latin: that is)
i.i.d.	independent identically distributed
ICA	Independent Component Analysis
KL	Kullback-Liebler
LHS	left hand side

LPD	Low Probability of Detection
LPI	Low Probability of Interception
MCRB	Modified Cramér-Rao Bound
MEA	Multiple Element Antenna
MI	Mutual Information
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
MUK	Multi User Kurtosis
N-D	N-Dimensional
OFDM	Orthogonal Frequency Division Multiplexing
OSTBC	Orthogonal Space Time Block Code
OSTBCG	Orthogonal Space Time Block Coding
pdf	probability density function
PSK	Phase Shift Keyed
QAM	Quadrature Amplitude Modulation
QOSTBC	Quasi-Orthogonal Space Time Block Code
QPSK	Quadrature Phase Shift Keyed
r.v.	Random variable
RF	Radio Frequency
RHS	right hand side
Rx	Receiver
RxCU	Receiver Channel Unknown

RxCUMSU	Receiver Channel Unknown, Message Set Unknown
RxCUMSUU	Receiver Channel and Message Set known, Unknown Unitary transformation
RxCURU	Receiver Channel Unknown, Rotation Unknown
RxFI	Receiver Fully Informed
s.t.	such that
SCORE	Self-Coherence Restoral
SER	Symbol Error Rate
SIMO	Single-Input Multiple-Output
SISO	Single-Input Single-Output
snr	signal to noise ratio
SOBI	Second Order Blind Identification
SOI	Signal Of Interest
SOS	Second Order Statistics
ST	Space-Time
STBC	Space-Time Block Code
STBCD	Space-Time Block Coded
STBCG	Space-Time Block Coding
STBD	Space-Time Block Decoder
STBE	Space-Time Block Encoder
STC	Space-Time Coding
STE	Space-Time Eavesdropper
STT	Space-Time Trellis
STWIC	Space-Time Wireless Intercept Channel
SVD	Singular Value Decomposition
Tx	Transmitter
VBLAST	Vertical Bell Labs Layered Space-Time
w.r.t.	with respect to
WTC	Wire-Tap Channel

Symbols

Δ	Equivocation
α	parameter in Generalised Gaussian distribution
$\hat{\mathbf{X}}$	Estimate of the matrix $\mathbf{X}$
$\hat{\mathbf{x}}$	Estimate of the vector $\mathbf{x}$
$\hat{x}$	Estimate of the scalar x
κ, κ_α	Kurtosis, kurtosis parameterised by α
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{I}_{mm}$	$p \times p$ identity matrix, where $p = m \times m$
$\mathbf{I}_{mn}$	$q \times q$ identity matrix, where $q = m \times n$
A_b	The channel matrix between Alice and Bob
A_e	The channel matrix between Alice and Eve
A_{ica}	ICA estimate of channel matrix A_e
A_{mle}	MLE of channel matrix A_e
I_b or I_B	Bob's mutual information
I_B-MLE	Simulation mutual information obtained by Bob, using MLE
I_B-Theory	Theoretical mutual information attainable by Bob
I_e or I_E	Eve's mutual information
I_E-ICA	Simulation mutual information obtained by Eve, using ICA
I_E-Theory	Theoretical mutual information attainable by Eve
X_{ica}	ICA estimate of source matrix X
X_{mle}	MLE of source matrix X

Notation

$\mathbb{C}$	Field of complex numbers
$\mathbb{R}$	Field of real numbers
x	Scalar quantity (real or complex)
!	Factorial
$G(\cdot,\cdot)$	Grassmann manifold
$H(\cdot)$	Shannon or discrete entropy
$S(\cdot,\cdot)$	Stiefel manifold
$[\mathbf{X}]_{i,j}$ or $x_{i,j}$	Element i,j of the matrix $\mathbf{X}$
$\Gamma(\cdot)$	Gamma function
$\Im\{\cdot\}$	Imaginary part of a complex variable
$\Re\{\cdot\}$	Real part of a complex variable
$\angle x$	Angle of x
$\approx$	approximately equal to
δ_{jk}	Kronecker delta: $\delta_{jk} = \begin{cases} 1 & j = k, \\ 0 & j \neq k. \end{cases}$
$\delta_{k_1,\ldots,k_n}$	Multidimensional Kronecker delta: $\delta_{k_1,\ldots,k_n} = \begin{cases} 1 & k_1 = k_2 \ldots = k_n, \\ 0 & \text{otherwise.} \end{cases}$
$\forall$	for all
$\hat{x}$	Estimate of x
$\in$	is a member of
$[\mathbf{X}]_{i,j}$	Element i,j of the matrix $\mathbf{X}$
$\ln(\cdot)$	Natural logarithm

$\log_b(\cdot)$	Logarithm in base b
$\mathbf{X}$	Matrix quantity (real or complex)
$\mathbf{x}$	Vector quantity (real or complex)
$\mathcal{CN}(\cdot,\cdot)$	Complex normal distribution
$\mathcal{E}\left\{\cdot\right\}$	Expectation
$\mathcal{N}(\cdot,\cdot)$	Normal distribution
$\mathcal{S}$	A set
$\mathcal{U}$	The universal set
$\mathbf{A}\odot\mathbf{B}$	Hadamard product between matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A}\otimes\mathbf{B}$	Kronecker product between matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{X}^*$	Conjugate of matrix $\mathbf{X}$
$\mathbf{X}^T$	Transpose of matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	Conjugate or Hermitian transpose of matrix $\mathbf{X}$
$\phi_\alpha\left(\cdot\right)$	Score function, parameterised by α
$\propto$	Proportional to
$\sim$	is distributed according to
vec$(\cdot)$	Vec-operator: if $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]$, then vec$(\mathbf{X})=\left[\mathbf{x}_1^T,\ldots,\mathbf{x}_n^T\right]^T$
$\triangleq$	Defined as
$h\left(\cdot\right)$	Differential entropy
$p(x)$	Probability density function of x
$p(x;a)$	Probability density function of x parameterised by a deterministic a
$p(xy)$	Joint probability density function of x and y
$p(x\vert a)$	Conditional probability density function of x given a
$p_a(x)$	Value of the probability density function of x at $x=a$
x^*	Conjugate of x
$\vert\mathcal{S}\vert$	Cardinality of $\mathcal{S}$
$\vert\mathbf{X}\vert$ or det$(\mathbf{X})$	Determinant of matrix $\mathbf{X}$
$\vert x\vert$	Absolute value of x
$\vert\vert\mathbf{X}\vert\vert_F$	Frobenius norm of matrix $\mathbf{X}$: $\vert\vert\mathbf{X}\vert\vert_F=\sqrt{\mathrm{tr}\left(\mathbf{X}^\dagger\mathbf{X}\right)}$
$\vert\vert\mathbf{x}\vert\vert$	Euclidean norm of vector $\mathbf{x}$: $\vert\vert\mathbf{x}\vert\vert=\sqrt{\mathbf{x}^\dagger\mathbf{x}}$

diag($\{\mathbf{X}_k\}$)	Block-diagonal matrix with diagonal blocks given by the set $\{\mathbf{X}_k\}$
diag($\mathbf{x}$)	Matrix with entries from vector $\mathbf{x}$ on the main diagonal
I($\cdot;\cdot$)	Mutual information
rk($\mathbf{X}$)	Rank of matrix $\mathbf{X}$
sign(x)	Sign of the real-valued x
tr($\mathbf{X}$)	Trace of matrix $\mathbf{X}$

Part I

Preliminaries

Chapter 1

Introduction

1.1 Motivation for the Research Detailed in this Thesis

The recent, and ongoing, development of techniques and theory in the field of MIMO wireless communications systems for increasing the reliability and rates of data transfer has been, to a relatively small extent, paralleled by the development of theory which addresses the security of such systems. Also commonly referred to as secrecy techniques, these ideas have a grounding in the mathematics of information theory and communications secrecy first developed by Shannon in 1949 [88].

We are motivated by the need to understand and quantify the information rates that are obtainable by a passive receiving system which does not cooperate with the transmitting system, in a MIMO wireless communications scenario. Cooperation in a communications link might include an exchange of information that could be used for channel estimation, exempli gratia (e.g.) a known sequence of training symbols. However, for a communications receiver being employed for surveillance or eavesdropping purposes, such information may not be available and the receiver is faced with the task of jointly estimating the channel coefficients and the source data streams.

Factors which affect the ability of an eavesdropping system to recover the individual signal streams that originated from a MIMO wireless transmitter include: knowledge of the propagation channel, channel fading, correlations in the channel, knowledge of the source distribution or symbol set and encoding scheme in use. The eavesdropper and communicator interference environments are also factors that affect eavesdropping. Understanding how the information rate, available to an informed receiver, is affected by such factors and other parameter errors, provides a basis both for improving proposed communications secrecy techniques and, from the opposite perspective, for improving eavesdropping information rates.

1.1.1 Problem statement

In this thesis we are concerned with the passive signal recovery problem. In particular, we are interested in the recovery of MIMO digital communication symbol streams that have been linearly mixed in an unknown multipath wireless propagation channel. The motivation for this research is to determine performance limits for symbol stream recovery and develop techniques that can achieve those limits. To this end we need to develop an understanding of BSS theory and how this theory can be best exploited in the MIMO wireless communications scenario.

1.1.2 Research methodology

The research described in this thesis addresses aspects of the above problem statement. A mathematical model that represents the physical MIMO wireless communications eavesdrop scenario is employed and is first described in Chapter 3. Appropriate literature has been identified, in Chapter 2, that presents concepts and background theory required for the development and analysis of the problem at hand. Fundamental background knowledge requirements include: communications techniques, information theory, blind source separation theory, communications secrecy concepts, Copula theory and calculus of complex linear algebra. The mathematical theory and tools, considered essential for analysing the eavesdrop problem, were first developed and some deficiencies were identified. Rectifying these deficiencies forms a large part of this thesis. To test the validity of the theory, Monte Carlo computer simulation exercises were developed and the results compared with theoretically derived expressions. Many of these simulation exercises involved the use of existing code to perform the task of BSS.

1.2 Structure of the Thesis

This thesis is presented in four parts.

Part I comprises the introduction and literature review.

Part II is formed from a number of chapters concerning the research into the theory and techniques required for analysing source recovery performance in the MIMO wireless communications scenario. Chapter 3 introduces the model and assumptions used to represent the problem. Expressions for passive eavesdropper Mutual Information (MI) are developed for the MIMO scenario. In Chapter 4 source and channel estimation performance theory, for the uninformed receiver, are derived. Chapter 5 demonstrates how Copula theory may be adapted to account for channel dependence and incorporate suitable fading distributions in a MIMO channel model.

Part III comprises two chapters that involve the application and analysis of the theory developed in Part II. Chapter 6 analyses the discrete source recovery problem for the MIMO model. The effect of source kurtosis on information rate is also studied. Many communication systems currently under development propose to use Space-Time Block Code (STBC) schemes to exploit the diversity offered by MIMO configurations. In Chapter 7 an approach is presented that demonstrates how linearly mixed Orthogonal Space Time Block Code (OSTBC) symbol streams may be transformed to suit application of BSS techniques.

Part IV contains the thesis conclusions and describes potential further work.

1.3 Summary of Novel Contributions

The chapters of Part II and Part III contain the novel contributions of this thesis. Part I comprises background material that is original only in the manner of its presentation.

The chapters forming Parts II and III are summarised here and original contributions are highlighted in each chapter summary.

Chapter 3 When an array of signals are transformed by an unknown unitary matrix we would like to know how this affects the information capacity at the receiver. We derive relationships for differential entropy, employing the concept of a hypersphere, of a multidimensional array. These expressions allow us to compare the fully informed capacity (channel known) with the partially informed capacity (amplitude known). This analysis does not appear to have been previously performed and has been published as [49, 54].

Chapter 4 Blind Source Separation is an important tool for an eavesdropping sensor array and in this chapter we derive the Cramér-Rao lower variance bounds for source and channel estimation, where both the source and the channel are complex-valued. The derivations involve a Modified CRB [37]. The CRB for the case when both the source and channel are unknown and the source is complex Gaussian is also derived. This derivation is based on a method presented by Villares [104] which involves fixing one of the variables to obtain the Fisher Information Matrix (FIM) for the other variable. In this chapter we also derive FIM and CRB expressions for the complex-valued source and complex-valued channel case and where the source distribution is the Generalised Gaussian (GG). A similar result was derived in [99] for the real-valued source and channel matrix case. For comparison derivations of Maximum Likelihood Estimator (MLE) expressions for complex-valued source and channel estimation, together with their respective Cramér-Rao variance bounds, are included. These derivations have been presented and published [51].

Chapter 5 To study the performance of BSS techniques, when there exists some correlation or dependence in the channel, the need for a simple and intuitive method for introducing source dependence whilst including different fading distributions, was identified. In this chapter an approach, based on Copula techniques, is presented. Previous efforts have only considered the use of Copula techniques for simple cases such as a bivariate pdf [31]. In this chapter we combine MIMO methods with wireless communications fading distributions and implement channel dependence for a complex-valued system to obtain simulated data that may be used to exercise BSS algorithms. This is new work that has been published [52, 53].

Chapter 6 In this chapter we perform computer simulations to study information rates as a function of source kurtosis and a number of other system parameters: signal to noise ratio (snr), observation data length, channel dimensions.

Since common digital modulation schemes have distinct values of kurtosis, this study gives an indication of eavesdropper and intended receiver performance as those system parameters are varied. The simulation results and theoretical predictions show that BSS is not possible when the sources are i.i.d. proper complex Gaussian, in which case an eavesdropper obtains no additional information about the sources given only observations on the source mixture. These results provide a benchmark for the source recovery performance obtainable by the intended receiver and the eavesdropper respectively.

Chapter 7 In this chapter an optimization algorithm, for finding a unitary mixing matrix, has been adopted and further developed so that the cost function could be varied. In particular the Joint Approximate Diagonalization of Eigenvalues (JADE) cost function, Mutual Information Between Sources (MIBS) cost function and their complex-valued gradients were incorporated. The gradient of the complex-valued JADE cost function stated in [2, 3, 4] was found to be incorrect and no derivation has been found in the literature. A complete derivation of the correct expression for the complex-valued gradient of the complex JADE cost is presented here. A new approach that exploits the structure of OSTBC signals shows the effective channel that results is unitary and therefore amenable to BSS using the JADE-based optimization algorithm. Simulations and analysis indicate the benefits of this new technique and the information rates attainable by an eavesdropper intercepting a digitally modulated MIMO transmission. The findings in this chapter have been published [50].

Chapter 2

Literature Review

The recent past has witnessed considerable activity in wireless communications research involving the use of antenna arrays for transmission and reception. This activity has resulted in an increasing body of literature dealing with improvements to MIMO communications links; particularly in terms of increasing information rates or capacity and tradeoffs with robustness to propagation fading effects. Whilst capacity and robustness issues appear to form the bulk of recent MIMO literature, other interesting issues and applications have arisen. Communication link security has always been of concern but now, with the application of Multiple Element Antenna (MEA) systems, link users have available more degrees of freedom that may be exploited to provide increased communications security. Such systems have an inherent physical security provided, primarily, by the more complex propagation channel. We note that the extra security offered by MIMO systems does not preclude the use of traditional data encryption techniques so that the simultaneous use of MIMO security and cryptographic security would seem to provide a powerful combination. The field of cryptographic research is already a significant and well established area; it is considered to be outside the scope of the present study. In this thesis we concentrate our attention on issues arising through the use of MIMO systems, from an eavesdropper's perspective and, in particular, the information rates that are achievable given different states of knowledge.

In the sections that follow we briefly review the history and literature pertaining to the problem of MIMO eavesdropping. This review serves two main purposes. The first is to provide a context for MIMO eavesdropping; bringing together the theory and tools that are required to analyse and understand this particular problem. The second purpose is to identify and shortlist any deficiences in the current literature that might require further study.

2.1 Development of MIMO Techniques and Theory

In the early 1990's Winters, Salz and Gitlin provided theoretical and experimental confirmation that multiuser interference and signal fading, in wireless communication systems, can be reduced through the use of multiple antennae and optimal signal combining at the receiver. In [108, 107, 109] Winters et alii (et al.) prove theoretically that an adaptive antenna array can achieve both interference nulling and path diversity against multipath fading. In effect they demonstrated that the information capacity of a wireless communication system may be increased by exploiting the spatial dimension. In 1999 Telatar [97] derived capacity expressions for single-user MIMO links, in an additive Gaussian channel, with and without fading. Telatar confirmed the increase in information capacity through the use of multiple antennae, noting the need to know the channel parameters at the receiver and the requirement that the path gains between different antenna pairs are independent.

In 1993 Wittneben [110] proposed a form of ST modulation employing multiple transmit antennae that takes advantage of spatial diversity without increasing system bandwidth. Wittneben's scheme achieves modulation diversity by filtering the input symbols so that the information is spread over the transmitted symbols and equalisation is applied by the single-antenna receiver. Another five years passed before Alamouti [8] introduced his, now famous, coding scheme which has become known as Alamouti Space-Time Coding (STC). This particular coding scheme is known to be the simplest of the orthogonal ST block codes, requiring two transmit antennae and a single receive antenna. In Alamouti's scheme two information-bearing symbols are encoded as a 2×2 ST block code and requires two time slots for transmission and reception; thus acheiving full system information rate whilst reducing its sensitivity to propagation fading.

Since these earlier publications there has been significant research activity into the understanding, use and improvement of MIMO techniques. There have been many and varied schemes and architectures proposed, e.g. Foschini [35] introduced a layered ST architecture for Rayleigh fading environments, where the transmitter and receiver have the same number of antennae. In this architecture the transmitter does not know the channel matrix and capacity increases linearly with the number of antennae, with a fixed bandwidth and fixed transmit power.

The information-theoretic aspect of MIMO techniques has naturally attracted researchers in the field of information theory producing many works in both single user and multi-user wireless communication link theory. A new type of coding has arisen to take advantage of such systems forming the field of STC. Thus as well as coding to combat transmission errors caused by channel fading or noise, STC exploits the extra diversity available in the MIMO channel. For example, Tarokh et al. [96], in 1999, introduced STBC, based on orthogonal code blocks, for use with multiple transmit antennae. Several books describing STC have been published e.g. "Space-Time Block Coding for Wireless Communications" by Larsson & Stoica [57], "Space-Time Coding" by Vucetic & Yuan [106] and "Space-Time Codes and MIMO Systems" by Jankiraman [47].

Any recently published textbook dealing with wireless communications theory would be incomplete without a description of MIMO techniques e.g. "Wireless Communications" by Goldsmith [38], "Fundamentals of Wireless Communications" by Tse & Viswanath [102] and "MIMO Wireless Communications" by Biglieri et al. [15].

2.2 Communications Security

Paralleling the development of MIMO techniques for MEA systems, at a somewhat less frenetic pace, has been the emergence of the concept of secrecy for such systems. Although Shannon introduced the concept of secrecy systems in his 1949 paper on the theory of communications secrecy [88], this theory was, for some time, only of interest to practitioners in the field of Cryptography. However, in 1975 Wyner [111] introduced a mathematical model for the Wire-Tap Channel (WTC), where digital data is to be reliably transmitted over a Discrete Memoryless Channel (DMC) which is being intercepted by an eavesdropper via a second DMC that taps into the first DMC. Whereas Shannon did not consider channel noise in his secrecy system, Wyner included noise in his channel thus allowing information rates and secrecy to be determined by both encoding and channel statistics.

One of the first to address security issues in a MIMO context, Hero [42] provides an analysis of security in ST communications. In "Secure Space-Time Communication" he introduces ideas about Low Probability of Detection (LPD) and Low Probability of Interception (LPI) from the point of view of communication between a transmitter and intended receiver attempting to be undetectable or denying information leakage to a possible eavesdropper. In particular Hero finds that perfect secrecy signalling may be acheived if the transmitter uses a codeword set $\mathcal{S}$ of block codes that have a constant spatial inner product id est (i.e.) $\mathcal{S} = \{S : SS^{\dagger} = A\}$, where A is a nonrandom square matrix. Examples of such secrecy-achieving codes include unitary codes and Constant Modulus (CM) codes.

In 1949, following soon after his famous work "A Mathematical Theory of Communication" [87], Shannon published the treatise "Communication Theory of Secrecy Systems" [88] where he developed the basic mathematical structure of communication secrecy systems. Shannon's schematic for a general secrecy system is reproduced in Figure 2.1.

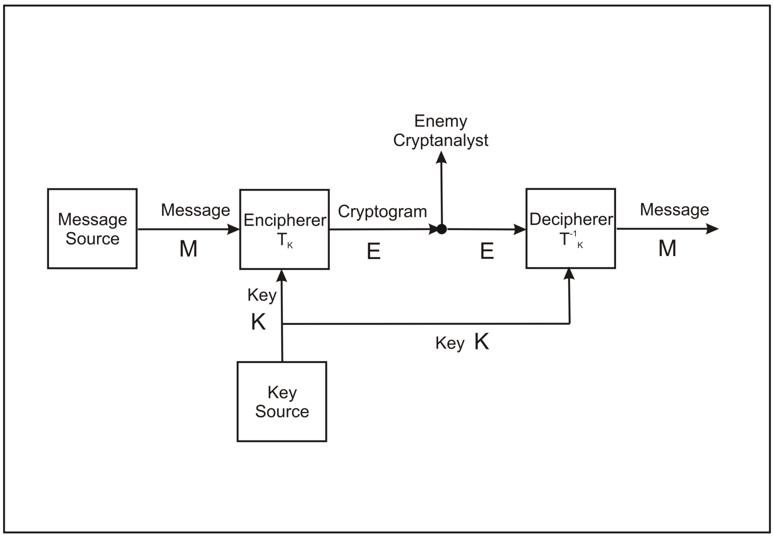

Figure 2.1: Shannon's general secrecy system.

Of particular interest is the notion of **perfect secrecy** where a cryptanalyst is unable to recover an intercepted message even if he had unlimited time, resources and encrypted data i.e. after a cryptogram has been intercepted the a posteriori probabilities of this cryptogram representing various messages are the same as the a priori probabilities of the same messages before the interception. Shannon shows that perfect secrecy is possible but requires, if the number of messages is finite, the same number of possible keys. A quantity $H(N)$ is defined, called the **equivocation**, which measures in a statistical way how near the average cryptogram of N letters is to a unique solution; that is, how uncertain the enemy is of the original message after intercepting a cryptogram of N letters. In standard information theory terminology, equivocation is equivalent to conditional entropy and quantifies the remaining uncertainty of a random variable given knowledge of another random variable. Shannon's message equivocation is given by $H(M|E) = \sum_{E,M} P(E,M) \log P_E(K)$, where E, M and K are the cryptogram, message and key, respectively, and $P(E,K)$ is the probability of key and cryptogram. $P_E(K)$ is the a posteriori probability of key K if cryptogram E is intercepted. $P(E,M)$ and $P_E(M)$ are the similar probabilities for the message.

Information-theoretic relationships can also be understood through the use of Venn diagrams and this representation is used in Appendix A as an aid to understanding Wyner's wire-tap channel.

Maurer [73] noted that Shannon's assumption that an enemy receives the same message as the legitimate receiver is motivated by considering error-free communication channels. However, most real communication channels are noisy and noisy channels are especially relevant in MIMO wireless communications.

2.3 The Wiretap Channel

Motivated by secrecy considerations, Wyner [111] considered a communications scenario in which Alice can send information to Bob over a DMC such that a wire-tapper Eve can receive Bob's channel output only through an additional cascaded independent channel, reducing the capacity of Eve's channel. Wyner proved that in such a setting Alice can send information to Bob in virtually perfect secrecy without sharing a secret key with Bob initially. Wyner's model is illustrated in Figure 2.2.

Wyner showed that it is possible to encode the transmitted data in a manner that maximises the uncertainty, or equivocation, of the data as observed by the eavesdropper or wire-tapper. When the wire-tapper's equivocation becomes equal to the entropy of the data source, then transmission to the intended receiver

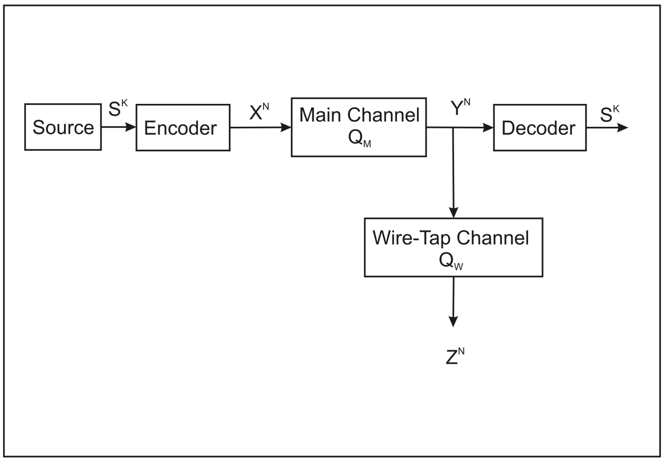

Figure 2.2: Wyner's Wire-Tap Channel

is considered to occur in perfect secrecy. A secrecy capacity C_S was defined as the maximum rate that allowed reliable transmission in perfect secrecy. This approach differs from encryption methods and relies on the snr observed by the wire-tapper being greater than the snr at the intended receiver. A simplified and more intuitive explanation of Wyner's WTC is presented in Appendix A.

Wyner's WTC has been further analysed as "The Gaussian Wire-Tap Channel" by Leung-Yan-Cheong & Hellman [59] where the main and wire-tap channels are modelled as additive Gaussian noise channels. Leung-Yan-Cheong & Hellman show that the secrecy capacity for this model equates to the difference between the capacities C_M and C_{MW} of the main and cascaded main plus wire-tap channels i.e. $C_S = C_M - C_{MW}$. Although developed for single channel communication links, these works, together with Shannon's 1949 paper on the theory of communications secrecy, have lain the foundations for further adaptations of the concept.

2.4 The Wireless Broadcast Channel

The WTC may be considered as a degraded broadcast channel where one information rate is to be maximised and the other minimised. Consider now the model introduced by Csiszár and Körner [27] where Eve's received message is not necessarily a degraded version of the legitimate receiver's message, Figure 2.3. The main channel and Eve's channel have a common input with the channel behaviour specified by the conditional joint probability $P_{YZ|X}$. The main channel is a Binary Symmetric Channel (BSC) with Bit Error Rate (BER) ϵ and Eve's channel is a BSC with BER δ.

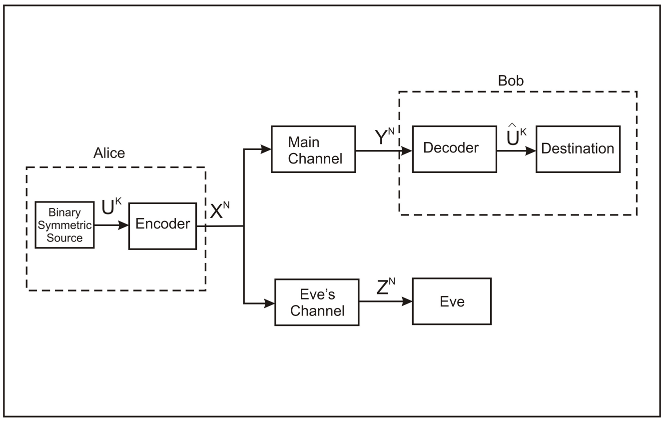

Figure 2.3: Model of DMC broadcast channel.

The secrecy capacity, from Alice to Bob, is shown to be

$$C_S = \begin{cases} h(\delta) - h(\epsilon), & \text{if} \quad \delta > \epsilon \\ 0, & \text{else,} \end{cases} \tag{2.1}$$

where $h(x)$ denotes the binary entropy function, i.e. the difference between the two channel entropies.

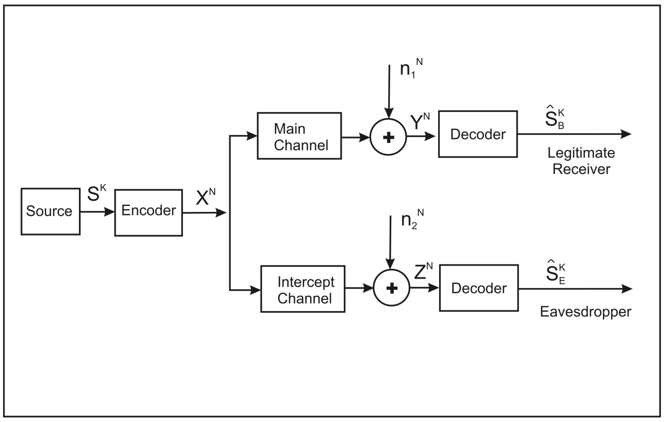

Figure 2.4: Parallel intercept channel model.

At this point a more general model might be considered for the parallel intercept wireless channel which is very similar to the wireless broadcast channel of Csiszár and Körner [27] but without restricting the channels to be BSC. This model is represented in Figure 2.4 and will form the basis for further analysis involving BSS techniques. In the figure there is a main channel between the source and the intended receiver and an intercept, or eavesdrop channel, that is independent of the main channel. Unlike the wiretap channel however, the eavesdrop channel may be better, in some sense, than the main channel. Noise or interference vectors $\mathbf{n}_i^N$ are shown in each of the two channels and these will be treated as additive random noise vectors i.e. we shall not be considering spatially coherent interference in this thesis.

2.5 Modelling Channel Dependence

Figure 2.5 gives an abstract illustration of the MIMO wireless Radio Frequency (RF) scenario where we have a multi-element source (Tx) transmitting an RF waveform to a multi-element receiver (Rx1) over an RF propagation path (shown in green). A number of scattering elements (S) are present in the RF environment. The whole RF environment is represented in grey. RF propagation between Tx and Rx2 is shown in light blue. This picture represents a communications broadcast scenario, where both of the receivers are intended to receive the signals from the transmitter and a surveillance scenario, where one of the receivers is not the intended recipient of the signals. In this study we consider a point-to-point, or single node, surveillance scenario in which a single eavesdropper is observing the communication link. In particular the model is constrained to the full-rank interference-free MIMO environment. Model constraints are:

- All transmitter and receiver antennae have the same polarization.
- Diversity is provided by the propagation environment.
- The channel has full rank - all the modes of the Singular Value Decomposition (SVD) of the channel response are nonzero.
- Transmission power is equally spread across the channel modes.
- Interference at the receiver is modelled as additive white noise i.e. spatially coherent interference is not represented here.

Whilst we only consider a full-rank, low mode spread environment, in practice any of the propagation environments represented in 2.5 could have a non-full rank MIMO channel response and therefore reduce the channel capacity. Interference is modelled as additive white noise, which simplifies later derivations of entropy and mutual information. However a more realistic model would consider local polarization/spatially coherent interference at the receiver array.

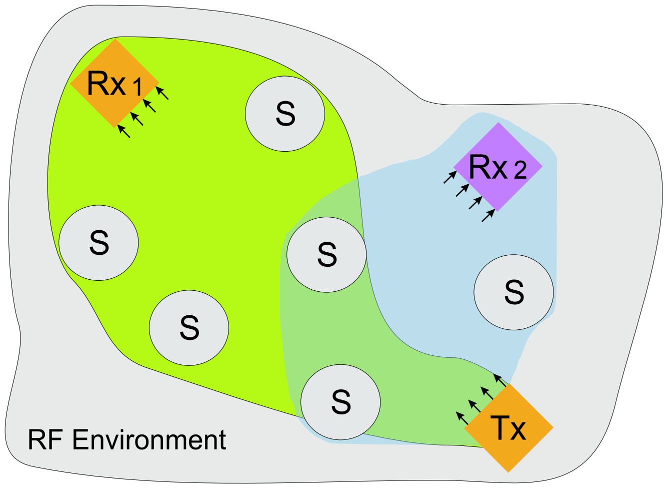

Figure 2.5: MIMO RF scenario.

A simple linear mixing model, described in Chapter 3, is commonly prescribed to represent such a scenario for analysis and simulation purposes. Dependence may be introduced at the transmitter array, the receiver array, within the propagation channel or any combination of these. In Figure 2.6 a block model for a point-to-point, or single node, MIMO wireless model is presented. An input message bit stream b_i is mapped to a symbol vector $\mathbf{s}$ and then encoded through a space-time encoder. The encoded blocks $\mathbf{X}$ may then be pre-processed before transmission through the wireless channel $\mathbf{A}$. At the receiver array noise or interference $\mathbf{W}$ is added. Following post-processing the data matrix $\mathbf{Y}$ is decoded and the estimated symbol vector $\hat{\mathbf{s}}$ demapped to retrieve an estimate of the input message bit stream.

To develop an understanding of channel effects that introduce dependence between the transmitted sources and for different propagation fading distributions, a suitable mathematical model is required. This model must allow flexibility in the types of dependence and fading distributions so that it provides a good representation of physical wireless propagation phenomena. These requirements have, to a large extent, been individually addressed by various authors. Whilst the

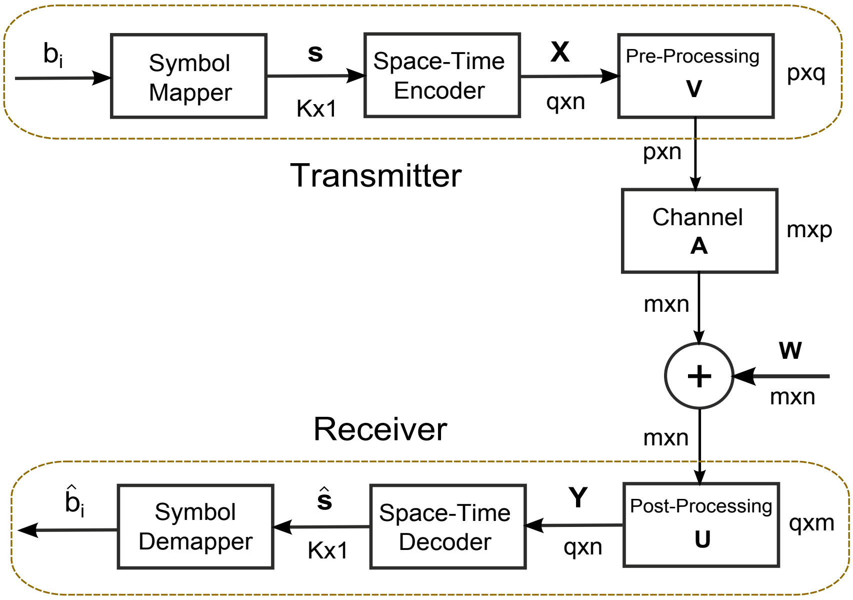

Figure 2.6: Block model of point-to-point MIMO link.

Rayleigh distribution has been employed for some time to represent fading, the Nakagami-m distribution has recently become popular for wireless fading simulation e.g. Alouini et al. [9], Beaulieu and Cheng [12]. In 2007 Ritcey [84] proposed the use of Copula theory to model multivariate fading distributions in wireless communications. Ritcey illustrates the concept through the simulation of a bivariate Rayleigh-Nakagami Copula with Nakagami marginal distributions to represent amplitude fading effects in a wireless scenario. Song et al. [92] and Zhang et al. [119, 117] derived correlated Nakagami fading models for wireless communications, with an arbitrary covariance matrix and distinct real fading parameters. These methods implement only a single distribution for the dependence and only consider amplitude fading. In 2005 Yacoub et al. [113] presented the exact expression for the joint phase-envelope distribution for the Nakagami-m distribution. This was subsequently used by authors such as: Ma and Zhang [66, 67], Santos Filho & Yacoub [85], to develop techniques for simulating a complex Nakagami fading channel with the correct amplitude and phase distributions.

2.6 Blind Source Separation

The problem of recovering signals that have been transformed through an unknown mixing process, commonly known as *blind source separation* (BSS) is a topic of wide interest in signal processing applications. The term blind refers to the fact that no explicit knowledge of the source signals or mixing system is available to an observer. Many approaches and algorithms have been developed, based on different statistical properties of the source signals. Higher Order Statistics (HOS) based algorithms, such as JADE and Fast Independent Component Analysis (FASTICA), are restricted to cases where only one of the sources may be Gaussian. Mixtures of multiple Gaussian sources may be treated through the use of Second Order Statistics (SOS) based methods and exploitation of different temporal structure in the sources. Both the HOS and SOS approaches share a common preprocessing step i.e. prewhitening of the observation data. This step reduces the search space by transforming the subsequent mixing matrix estimation step to a search for a unitary matrix.

Typical digital communication source signals, such as Phase Shift Keyed (PSK) and Quadrature Amplitude Modulation (QAM), have a significant kurtosis value and so are well suited to BSS approaches based on HOS. For this reason the principle algorithms selected for use in this study are: FASTICA and JADE. Recognising the potential for near Gaussian distributed sources, to reduce the intercept capacity of an eavesdropper in a MIMO wireless scenario, some SOS based alternatives are described here. However the implementation and analysis of these SOS methods is outside the scope of the current study. These approaches are discussed here simply to highlight the fact that alternatives to HOS do exist and should be considered if they are appropriate for the problem under study.

2.6.1 BSS Based on Higher Order Statistics

HOS based algorithms do not exploit possible temporal structure of the sources, relying instead on two primary assumptions: the source samples are identically distributed and the sources are independent of each other i.e. source sequences are i.i.d. and the different sources may have different distributions. Since Gaussian distributed sources have SOS only, HOS based algorithms are unable to separate mixtures of i.i.d. Gaussian sources. Statistical methods for performing BSS, such as Comon's 1994 description of Independent Component Analysis (ICA) [25], have resulted in popular algorithms such as FASTICA [44, 43, 46, 16, 55] and Cardoso & Souloumiac's 1993 JADE [19] method. Both algorithms are available as Matlab code, which has undoubtedly aided their popularity. It is also well known that there are ambiguities in the estimated sources that result from BSS methods and this has been addressed by Hyvärinen [45] & Oja and by Eriksson & Koivunen [32].

Abrudan et al. [2, 3, 4] have demonstrated how to incorporate the JADE cost and gradient into alternative optimization algorithms. This is particularly useful since the authors have developed a method for optimization under a unitary constraint which is ideally suited to the BSS problem. However the expression for the gradient of the JADE cost function, that was presented in [2, 3], has been found to be incorrect [50].

Tichavský et al. [99, 100] have derived estimation error performance bounds for source and mixing matrix estimation for the noiseless linear mixing model using real-valued source and real-valued mixing matrix variables. The authors employed the GG distribution to provide a random source where the kurtosis of the distribution could be continuously varied. This feature aids understanding of source separability and provides a useful reference when considering separation of digital sources with known kurtosis values. Similar source and channel estimation error bounds for complex-valued source and complex-valued mixing matrix variables have not been found in the literature.

In section 4.5 the derivation of the CRB for the mixing matrix is described. In section 6.2 we compare the theoretical CRB with simulation results using the complex variant of the FASTICA algorithm [55], in conjunction with an algorithm for resolving the permutation problem [98] associated with BSS, for mixing matrix estimation.

2.6.2 BSS Based on Second Order Statistics

After the prewhitening step, SOS based algorithms typically apply a separation technique such as diagonalizing a covariance matrix [101], or by jointly diagonalizing a number of covariance matrices [13]. SOS approaches rely on the presence of time structure in the sources and so, for cases where this condition is not satisfied, SOS alone may be insufficient for successful BSS. In particular, if the each of the sources has an i.i.d. time structure, then SOS is only useful for spatial whitening, after which a solution may be found using a HOS method. It is clear and well-known that Gaussian i.i.d. sources cannot be separated by SOS, HOS or even a combination of these approaches. However, when the sources do have different temporal covariance structures, then SOS based approach may be considered for BSS.

In the late 1980's Agee et al. [6, 7] introduced the Self-Coherence Restoral (SCORE) approach to blind adaptive signal extraction. In [6] an approach is presented that addresses communication signals extraction, through blind adaptation of an antenna array, in co-channel interference environments, using only known spectral correlation properties of the signals.

At the end of the 1980's Tong et al. [101] developed the Algorithm for Multiple Unknown Signals Extraction (AMUSE) which assumes temporally coloured sources and relies on Exact Joint Diagonalization (EJD) of the observation correlation matrix. Molgedey et al. [74] applied a similar method as in [101] using time delayed correlations. In [13] Belouchrani et al. introduced the Second Order Blind Identification (SOBI) algorithm. Unlike AMUSE, SOBI doesn't perform EJD of a matrix pair but derives the Approximate Joint Diagonalization (AJD) on a set of more than two correlation matrices. In [13] a simulation example, where the separation of two complex circular Gaussian source signals in the presence of stationary complex white noise, is studied.

In the early 1990's a Maximum Likelihood (ML) based technique was developed by Belouchrani and Cardoso in [14] for discrete source separation, with known source probability distributions. ML approaches were also investigated by: Harroy and Lacoume [41], Pearlmutter, Parra [79], Pham and Cardoso [81]. More recently, Yeredor [115] considered the separation of Gaussian sources exhibiting general, arbitrary covariance structures and derived the ML estimate of the separation matrix.

SOS techniques have also been proposed based on the Canonical Correlation Analysis (CCA) approach [36, 17, 116]. In this approach, the demixing matrix is found by maximizing the autocorrelation of each of the recovered signals. This approach relies on the idea that the sum of any uncorrelated signals has an autocorrelation whose value is less or equal to the maximum value of individual signals, as proved in [17]. Liu et al. [62] generalised the CCA approach to address the source separation problem for noisy mixtures.

2.7 Blind Separation of Space-Time Encoded Sources

One of the main aims in this study is to quantify the information rate available to a passive eavesdropper intercepting a MIMO wireless digital communications transmission. Previous work in the literature has addressed aspects of this problem. In 1998 Grellier & Comon [40, 39] considered the problem and performance of the blind separation of discrete sources, in particular for PSK sources. In [39] Grellier & Comon derive some performance bounds for Binary Phase Shift Keyed (BPSK) and 4-PSK, utilising error probability functions for these signal types. Kurtosis, defined in Appendix C, has been found to be a very useful parameter for comparing different digital modulation schemes. In his 2001 paper Mathis [72], recognising that the kurtosis of the source provides an indication of the separation difficulty faced by BSS techniques, studied the effects of timing offsets on the kurtosis of digitally modulated signals. Mathis was able to derive expressions that give the output source kurtosis as a function of input kurtosis, timing offset and pulse shaping.

In 2002 Swindlehurst [94] showed how knowledge of the structure of ST block coded signals could be combined with the Algebraic Constant Modulus Algorithm (ACMA) for blind source separation of Space-Time Block Coded (STBCD) data using a modulation symbol set that has a constant modulus such as PSK. Swindlehurst notes that his algorithm is unable to perform BSS when Alamouti Space-Time Block Coding (STBCG) is employed.

In their 2003 paper Rinas & Kammeyer [82] describe a hybrid MIMO system that uses the JADE algorithm for BSS and the Vertical Bell Labs Layered Space-Time (VBLAST) algorithm for symbol detection. The authors utilise BSS to facilitate channel estimation and then apply the VBLAST algorithm for improved symbol detection with knowledge of the finite symbol set. Performance is demonstrated by way of constellation plots showing the observed signals before BSS and the estimated constellations, after BSS.

liu et al. [61], in 2005, described the use of two iterative algorithms to find the source separating matrix, recognising that this is an orthogonal matrix when Orthogonal Space-Time Block-Coding is being employed by the transmitter. BER performance is assesed via simulations using the Alamouti code. However no comparison is made with direct application of the two well-known BSS algorithms: FASTICA and JADE.

2.8 Summary

In this review the wireless Space-Time intercept channel is considered in a context analogous the broadcast channel that evolved from the wire-tap channel model, leading to the simplifying notion that communications information secrecy is a function of the difference between mutual information for the intended channel and the mutual information for an eavesdropper's channel. Techniques for analysing this problem are drawn from the field of BSS. From the preceding survey of available literature, some deficiencies have become evident. These are summarised as follows:

- A thorough theoretical analysis of the information rates achievable by a MIMO eavesdropper, given different states of knowledge and for a complex-valued source and complex-valued channel model, has not been undertaken.
- A practical method for readily modelling dependence in a MIMO channel does not appear to be available.
- An incorrect expression has been stated in the literature for the gradient of the complex-valued JADE cost function, used in BSS optimization algorithms.
- Adaptation of Space-Time Block-Coded signals to suit BSS algorithms such as ICA or JADE has not been adequately addressed.

Part II

Theory and Techniques

Chapter 3

Information Theory for Eavesdroppers

In this chapter we are interested in differential entropy and mutual information as they apply to wireless communication links employing antenna arrays at both the transmission and receiving sites. Systems of this type are more commonly known as MIMO wireless communication systems. MIMO wireless communication techniques are known to provide increased information transfer rates, or channel capacities, over those obtainable using single transmit antenna to single receive antenna links [97, 34]; however this extra capacity comes at the expense of increased system complexity and additional processing for both the transmitter and the receiver. To correctly receive and detect the transmitted message, the receive system must know the channel, or mixing matrix, as well as the message symbol set being used. The channel matrix may be estimated when a predetermined, known message sequence is incorporated into the transmitted message and the receiver knows where in the message this sequence occurs. However the training sequence may not always be available and this presents a blind source estimation problem where neither the message nor the channel matrix are known to the receiver.

The following list defines the main system parameters that are considered in this study:

Synchronization parameters: the parameter set $\mathcal{P}$ required for the receiver to correctly synchronize with the transmitted waveform e.g.: carrier frequency, timing offset, symbol rate.

Unitary Transformation Matrix: a unitary pre-processing transformation $\mathbf{R}$ applied by the transmitter.

Channel matrix: the complex channel matrix $\mathbf{A} \in \mathbb{C}^{m \times p}$, where m is the number of receive antennae and p is the number of transmit antennae.

Message set: the symbol set $\mathcal{S}$, for discrete messages or the message covariance in the continuous case. We assume a zero mean for the message set or distribution.

Message: the message matrix $\mathbf{X}$.

Interference Covariance Matrix: the covariance matrix $\mathbf{\Sigma_w}$ of the additive noise or interference; assuming a zero mean distribution.

We may construct a set of receiver knowledge states as a function of the knowledge state of the individual system parameters. Using the function $t(\theta) \in \{0, 1\}$ to indicate if the parameter θ is unknown ($t(\theta) = 0$) or known ($t(\theta) = 1$) and the function $T(t(\mathbf{A}), t(\mathbf{X}), t(\mathbf{R}), t(\mathbf{\Sigma_w}), t(\mathcal{S})) \in \{0, 1\}$ to indicate if a set of parameter knowledge states is false ($T(t(\mathbf{A}), t(\mathbf{X}), t(\mathbf{R}), t(\mathbf{\Sigma_w}), t(\mathcal{S})) = 0$) or true ($T(t(\mathbf{A}), t(\mathbf{X}), t(\mathbf{R}), t(\mathbf{\Sigma_w}), t(\mathcal{S})) = 0$), then we can define a truth table for the states of the main system parameters and the combination of parameter states that form the receiver knowledge states of interest. The parameter $\mathcal{P}$ is assumed to be always known or knowable. To facilitate further analysis a number of receiver knowledge states are defined as follows:

State-I: The channel between the transmitter and the receiver is known. The message is known. Any unitary transformation applied by the receiver is known. The noise covariance is known. The message set or covariance is known but not necessary since the message is known. $t(\mathbf{A}) = 1$, $t(\mathbf{X}) = 1$, $t(\mathbf{R}) = 1$, $t(\mathbf{\Sigma_w}) = 1$, $t(\mathcal{S}) = 1$, $T(1, 1, 1, 1, 1) = 1$.

State-II: The channel between the transmitter and the receiver is known. The message is unknown. Any unitary transformation applied by the receiver is known. The noise covariance is known. The message set or covariance is known. $t(\mathbf{A}) = 1$, $t(\mathbf{X}) = 0$, $t(\mathbf{R}) = 1$, $t(\mathbf{\Sigma_w} = 1)$, $t(\mathcal{S}) = 1$, $T(1, 0, 1, 1, 1) = 1$.

State-III: The channel between the transmitter and the receiver is unknown. The message is known. Any unitary transformation applied by the receiver is known. The noise covariance is known. The message set or covariance is known but not necessary since the message is known. $t(\mathbf{A}) = 0$, $t(\mathbf{X}) = 1$, $t(\mathbf{R}) = 1$, $t(\mathbf{\Sigma_w} = 1)$, $t(\mathcal{S}) = 1$, $T(0, 1, 1, 1, 1) = 1$.

State-IV: The channel between the transmitter and the receiver is unknown. The message is unknown. Any unitary transformation applied by the receiver is known. The noise covariance is known. The message set or covariance is known. $t(\mathbf{A}) = 0$, $t(\mathbf{X}) = 1$, $t(\mathbf{R}) = 1$, $t(\mathbf{\Sigma_w} = 1)$, $t(\mathcal{S}) = 1$, $T(0, 0, 1, 1, 1) = 1$.

State-V: The channel between the transmitter and the receiver is unknown. The message is unknown. Any unitary transformation applied by the receiver is unknown. The noise covariance is unknown. The message set or covariance is unknown. $t(\mathbf{A}) = 0$, $t(\mathbf{X}) = 0$, $t(\mathbf{R}) = 0$, $t(\mathbf{\Sigma_w} = 0)$, $t(\mathcal{S}) = 0$, $T(0, 0, 0, 0, 0) = 1$.

State-VI: The channel between the transmitter and the receiver is known. The message is unknown. Any unitary transformation applied by the receiver is unknown. The noise covariance is known. The message set or covariance is known. $t(\mathbf{A}) = 1$, $t(\mathbf{X}) = 0$, $t(\mathbf{R}) = 0$, $t(\mathbf{\Sigma_w} = 1)$, $t(\mathcal{S}) = 1$, $T(1, 0, 0, 1, 1) = 1$.

The resulting truth table, and when combinations of parameter knowledge states satisfy the receiver knowledge states I to VI, may be represented by the Karnaugh map shown in Figure 3.1.

In the eavesdropping scenario we shall assume that the intended receiver is in the fully informed state I. The eavesdropping receiver's knowledge may be in any of the above states but is generally assumed to be in one of the partial knowledge states II to VI. States I to IV are used to derive MI expressions for the eavesdropping scenario in section 3.2.

States IV and VI represent the assumptions of the hypersphere model for mutual information derived later in this chapter. For standard communication links information theoretic derivations, such as entropy and MI, typically assume the fully informed state. In this study the partially informed states are of greater

$\mathbf{R}\Sigma_{\mathbf{w}}\mathcal{S}$

AX	000	001	011	010	100	101	111	110
00	V	0	0	0	0	0	IV	0
01	0	0	0	0	0	0	III	III
11	0	0	0	0	0	0	I	I
10	0	0	VI	0	0	0	II	0

Figure 3.1: Karnaugh map showing when defined receiver knowledge states are satisfied.

interest and so, in this chapter, entropy and MI derivations are presented to enable further study and understanding of the consequences of reducing the knowledge available to a passive eavesdropping receiver. When considered in the context of information rates or channel capacity, the reduction in MI caused by reducing the eavesdropping receiver's state of knowledge, may be considered as a measure of secrecy available to the intended communication link pair (Alice and Bob). Secrecy capacity was discussed in Chapter 2 where it was found, for a wireless broadcast channel, that it could be quantified in terms of the difference in MI between two communication links i.e. the difference between the intended link and an unintended link. Figure 3.2 shows a simple diagram that is commonly used

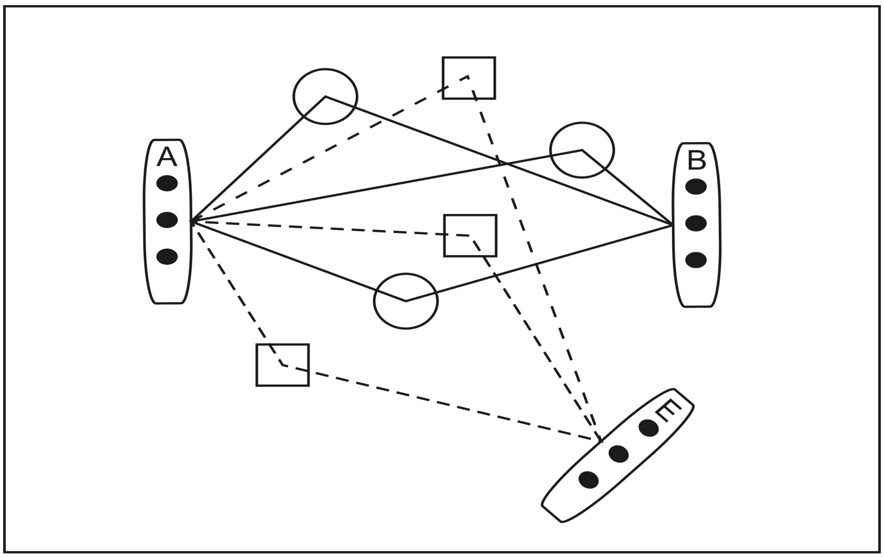

Figure 3.2: MIMO Wireless Intercept Model.

in the literature to represent the MIMO wireless broadcast scenario. By a well-known cryptographic convention, introduced by Maurer in [73], the transmission source array is labelled A (for Alice), the intended cooperative receiver array is labelled B (for Bob) and the unintended, passive intercept receiver is labelled E (for Eve). Solid lines are used to represent paths of RF propagation from Alice's antennae to Bob's antennae, dotted lines represent paths of RF propagation from Alice's antennae to Eve's antennae. In the RF environment there are objects that reflect or scatter the RF and these are represented by the squares and circles in the figure. An important point to realise here is that the paths (channel AB) between A and B are different to those between A and E (channel AE).

3.1 Model and Assumptions

As a mathematical representation of a single MIMO wireless communications link, the following simple linear transformation

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W} \tag{3.1}$$

is employed, where $\mathbf{Y}$ is the received signal matrix, $\mathbf{X}$ is the transmitted source matrix, $\mathbf{W}$ is an additive receiver noise matrix and $\mathbf{A}$ is the channel gain or mixing matrix between the transmitter and receiver. We make use of the following notational conventions:

- When the vector $\mathbf{y}$ has a real multivariate normal distribution, written as $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{y}, \boldsymbol{\Sigma}_\mathbf{y})$, with mean $\boldsymbol{\mu}_\mathbf{y}$ and covariance matrix $\boldsymbol{\Sigma}_\mathbf{y}$, the pdf for $\mathbf{y}$ is

$$p(\mathbf{y}) = |2\pi\boldsymbol{\Sigma}_\mathbf{y}|^{-1/2} \exp\left\{-[\mathbf{y} - \boldsymbol{\mu}_\mathbf{y}]^T \boldsymbol{\Sigma}_\mathbf{y}^{-1} [\mathbf{y} - \boldsymbol{\mu}_\mathbf{y}]\right\}. \tag{3.2}$$

- If the real-valued scalar y has a normal distribution, written as $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, with mean μ_y and variance σ_y^2, the pdf for y is

$$p(y) = (2\pi\sigma_y^2)^{-1/2} \exp\left\{-\frac{[y - \mu_y]^2}{2\sigma_y^2}\right\}. \tag{3.3}$$

- When $\mathbf{y}$ is a proper complex Gaussian random vector, written as $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}_\mathbf{y}, \boldsymbol{\Sigma}_\mathbf{y})$, with mean $\boldsymbol{\mu}_\mathbf{y}$ and Hermitian covariance matrix $\boldsymbol{\Sigma}_\mathbf{y}$, the pdf for $\mathbf{y}$ is

$$p(\mathbf{y}) = |\pi\mathbf{\Sigma_y}|^{-1}\exp\left\{-[\mathbf{y}-\boldsymbol{\mu}_\mathbf{y}]^\dagger\mathbf{\Sigma_y}^{-1}[\mathbf{y}-\boldsymbol{\mu}_\mathbf{y}]\right\}. \quad (3.4)$$

A proper complex random variable is uncorrelated with its complex conjugate.

- If the complex-valued scalar y has a proper complex normal distribution, written as $y \sim \mathcal{CN}(\mu_y, 2\sigma_y^2)$, with mean μ_y and variance $2\sigma_y^2$ i.e. the sum of the variances in the real and imaginary parts of y, which are assumed to both equal σ_y^2. The pdf for y is

$$p(y) = (2\pi\sigma_y^2)^{-1}\exp\left\{-\frac{|y-\mu_y|^2}{2\sigma_y^2}\right\}. \quad (3.5)$$

The MIMO channel is commonly modelled using proper complex-valued random variables for the channel components. For example Marzetta and Hochwald [70] assume i.i.d., frequency-flat Rayleigh amplitude fading between the transmit and receive antennae. Consequently the components $a_{i,j}$ of $\mathbf{A}$ are typically modelled as i.i.d. proper complex Normal: $a_{i,j} \sim \mathcal{CN}(0, 2\sigma_a^2)$. This model may be used to represent either the intended link (between Alice and Bob) or the unintended link (between Alice and Eve).

The following general assumptions are made:

- $\mathbf{X} \in \mathbb{C}^{p\times n}$ is an i.i.d. proper complex random source matrix with zero mean and component variance $var\{x_{i,j}\} = 2\sigma_x^2$.

- $\mathbf{W} \in \mathbb{C}^{m\times n}$ is an i.i.d. proper complex random Gaussian noise matrix i.e. its components are distributed as $w_{i,j} \sim \mathcal{CN}(0, 2\sigma_w^2)$ and the i^{th} column of $\mathbf{W}$, $\mathbf{w}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma_w})$. $\mathbf{W}$ does not model spatially coherent interference with a non unity covariance matrix. $\mathbf{\Sigma_w}$ is assumed known for all the receiver knowledge states used in the MI derivations.

- $\mathbf{Y} \in \mathbb{C}^{m\times n}$ is an i.i.d. proper complex observation matrix which, since $\mathbf{X}$ and $\mathbf{W}$ are zero mean, is also zero mean and its component variance is $var\{y_{i,j}\} = 2\sigma_y^2$.

- The intended channel AB is known only to Bob. Alice therefore adopts a simple transmission scheme where equal power is output from each antenna and the antenna outputs are mutually independent. If Alice knew channel AB then she could preprocess the data via SVD and apply a waterfilling technique, which is described by Tse and Viswanath in [102, Ch.7], for channel power allocation to optimize channel capacity.

- Eve does not know the intercept channel AE or the intended channel.
- The channels $\mathbf{AB} = \mathbf{A}_b \in \mathbb{C}^{m_b \times p}$ and $\mathbf{AE} = \mathbf{A}_e \in \mathbb{C}^{m_e \times p}$ vary slowly with time and may be assumed constant for the observation block lengths under consideration. Over a longer time period the components of $\mathbf{A}_b$ and $\mathbf{A}_e$ are assumed to be i.i.d., have a zero mean and component variance $var\{[A_b]_{i,j}\} = var\{[A_e]_{i,j}\} = 2\sigma_a^2$.

3.2 Mutual Information

To proceed with the derivations of MI we first recall a few definitions from Cover and Thomas [26, Chs.2,9]. The differential entropy $h(Y)$ of a continuous random variable Y, with a probability density $p(y)$ is defined as

$$h(Y) \triangleq -\int_{\mathbb{Y}} p(y) \ln(p(y)) dy = -\mathbb{E}\left\{\ln(p(y))\right\}, \tag{3.6}$$

where $\mathbb{Y}$ is the support set of the random variable Y. Y may be a scalar, a vector or a matrix and may be real or complex. When we have two random variables Y, X with joint probability density $p(y, x)$, the conditional differential entropy is defined as

$$h(Y|X) \triangleq -\int_{\mathbb{Y},\mathbb{X}} p(y, x) \ln(p(y|x)) dy\, dx, \tag{3.7}$$

where $\mathbb{X}$ is the support set of the random variable X. X may be a scalar, a vector or a matrix and may be real or complex. The MI between the two random variables Y and X is defined as

$$\begin{aligned} I(Y;X) &\triangleq \int_{\mathbb{Y},\mathbb{X}} p(y, x) \ln \frac{p(y, x)}{p(y)p(x)} dy\, dx \\ &= h(Y) - h(Y|X) = h(X) - h(X|Y), \end{aligned} \tag{3.8}$$

and represents the reduction in the uncertainty of the source variable X given knowledge of the observed variable Y. The capacity C is then obtained by maximizing the mutual information over all probability distributions for the source i.e. over $p(x)$:

$$C = \sup_{p(\mathbf{x})} I(Y;X). \tag{3.9}$$

It is well known, for example see Cover and Thomas [26], that a Gaussian source distribution is an entropy maximizer (for a given variance) and is therefore commonly used to determine channel capacities.

Turning now to the MIMO link represented by equation 3.1, where $\mathbf{Y} \in \mathbb{C}^{m\times n}$, $\mathbf{A} \in \mathbb{C}^{m\times p}$, $\mathbf{X} \in \mathbb{C}^{p\times n}$ and $\mathbf{W} \in \mathbb{C}^{m\times n}$, the MI for the legitimate user, Bob, given knowledge of the channel matrix $\mathbf{A}$, is

$$I_b = I(\mathbf{Y}|\mathbf{A};\mathbf{X}) = h(\mathbf{Y}|\mathbf{A}) - h(\mathbf{Y}|\mathbf{A},\mathbf{X}) \tag{3.10}$$

and the mutual information for the eavesdropper, Eve, is

$$I_e = I(\mathbf{Y};\mathbf{X}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}). \tag{3.11}$$

In terms of the receiver knowledge states that were defined earlier

$$I_b = h(II) - h(I) \tag{3.12}$$

and

$$I_e = h(IV) - h(III), \tag{3.13}$$

where $h(I)$, $h(II)$, $h(III)$ and $h(IV)$ are the entropies of being in states I-IV, respectively.

We wish to study how I_b and I_e are affected by different source distributions and states of knowledge. To model the source distribution we shall make use of the *Generalised Gaussian* (GG) distribution, described by Tichavský et al. in [99], and also described in Appendix C for reference. The GG distribution is a family of symmetric probability distributions with a parameter $\alpha > 0$ that determines the shape of the pdf. As α increases from zero to infinity, the shape of the pdf varies from sharply peaked to uniform on a bounded interval. Special cases occur when $\alpha = 1$ and $\alpha = 2$. In the first case the Laplace pdf is obtained and the second case yields a normal distribution. The advantage of using the GG is that it allows us to smoothly vary the shape of the distribution to represent a wide range of source pdfs. Changing the shape of the pdf changes the source differential entropy, which is maximised when the distribution is normal; using a fixed source variance for each α. Also, since we are particularly interested in BSS performance for distributions that are close to Gaussian, the GG will yield a distribution that may be approximated by a Gaussian. We shall show later how entropy power may be substituted for the variance in the Gaussian approximation. Under the assumptions stated in section 3.1 $p(\mathbf{Y}|\mathbf{X})$ and $p(\mathbf{Y}|\mathbf{A},\mathbf{X})$ are Gaussian. For small array dimensions $p(\mathbf{Y})$ is not Gaussian but, under the central limit theorem, $p(\mathbf{Y})$ becomes more Gaussian as m increases. We shall require the differential entropy for a complex Gaussian random vector. Let $\mathbf{y}$ be an $m \times 1$ complex Gaussian random vector with $\mathbf{y} \sim \mathcal{CN}(0, \mathbf{\Sigma_y})$, where $\mathbf{\Sigma_y}$ is the covariance matrix for $\mathbf{y}$. The pdf for $\mathbf{y}$ is, for example see Kay [48] or Scharf [86],

$$p(\mathbf{y}) = |\pi\mathbf{\Sigma_y}|^{-1} \exp\left\{-\mathbf{y}^{\dagger}\mathbf{\Sigma_y}^{-1}\mathbf{y}\right\}. \tag{3.14}$$

The differential entropy for this distribution is found to be

$$h(\mathbf{y}) = \ln(|\pi e\mathbf{\Sigma_y}|) \text{ nats}, \quad \text{or} \quad h(\mathbf{y}) = \log_2(|\pi e\mathbf{\Sigma_y}|) \text{ bits}. \tag{3.15}$$

There are four different differential entropies and hence four different covariance matrix estimates required in the expressions for the two mutual information cases I_b and I_e. The GG distribution will be used to generate the real and imaginary parts of the message matrix $\mathbf{X}$ and we shall use the term Gaussianity as an indication of how close the parameter α brings the GG distribution to a Gaussian distribution. The Gaussianity of the message matrix $\mathbf{X}$ will be varied from super-Gaussian (positive kurtosis) to sub-Gaussian (negative kurtosis) through application of the GG distribution. We shall use the entropy power of the source distribution as a substitute for the Gaussian variance in our covariance calculations because this allows us to treat the message as if it had a Gaussian distribution, where the reduction in entropy is due to a reduction in the source variance. The covariances that we require are: $\mathbf{\Sigma_{y|A,X}}$, $\mathbf{\Sigma_{y|A}}$, $\mathbf{\Sigma_{y|X}}$ and $\mathbf{\Sigma_y}$.

In the derivations that follow we employ the vectorisation operator $\mathbf{vec}(\cdot)$ which stacks the columns of a matrix in a column vector, i.e., for an $(m \times n)$ matrix $\mathbf{Y} = [\mathbf{y}_1\, \mathbf{y}_2\, \dots\, \mathbf{y}_n]$, where $\mathbf{y}_i$ is the ith column of $\mathbf{Y}$,

$$\mathbf{vec}(\mathbf{Y}) \triangleq [\mathbf{y}_1^T\, \mathbf{y}_2^T\, \dots\, \mathbf{y}_n^T]^T. \tag{3.16}$$

The relationship

$$\mathbf{vec}\,(\mathbf{AX}) = (\mathbf{I}_n \otimes \mathbf{A})\mathbf{vec}\,(\mathbf{X})\,, \tag{3.17}$$

which may be found in [63], for $\mathbf{A}\,(m \times p)$ and $\mathbf{X}\,(p \times n)$, is also required; $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\otimes$ is the Kronecker matrix product. We make the following definitions

$$\begin{aligned}
\mathbf{y} &\triangleq \mathbf{vec}\,(\mathbf{Y})\,, && (3.18)\\
\mathbf{a} &\triangleq \mathbf{vec}\,(\mathbf{A})\,, && (3.19)\\
\mathbf{x} &\triangleq \mathbf{vec}\,(\mathbf{X})\,, && (3.20)\\
\mathbf{w} &\triangleq \mathbf{vec}\,(\mathbf{W})\,, && (3.21)\\
\boldsymbol{\Sigma}_\mathbf{y} &\triangleq \mathbb{E}\left\{[\mathbf{y} - \mathbb{E}\{\mathbf{y}\}][\mathbf{y} - \mathbb{E}\{\mathbf{y}\}]^\dagger\right\} = 2\sigma_y^2\mathbf{I}_{mn}, && (3.22)\\
\boldsymbol{\Sigma}_\mathbf{a} &\triangleq \mathbb{E}\left\{[\mathbf{a} - \mathbb{E}\{\mathbf{a}\}][\mathbf{a} - \mathbb{E}\{\mathbf{a}\}]^\dagger\right\} = 2\sigma_a^2\mathbf{I}_{mp}, && (3.23)\\
\boldsymbol{\Sigma}_\mathbf{x} &\triangleq \mathbb{E}\left\{[\mathbf{x} - \mathbb{E}\{\mathbf{x}\}][\mathbf{x} - \mathbb{E}\{\mathbf{x}\}]^\dagger\right\} = 2\sigma_x^2\mathbf{I}_{pn}, && (3.24)\\
\boldsymbol{\Sigma}_\mathbf{w} &\triangleq \mathbb{E}\left\{[\mathbf{w} - \mathbb{E}\{\mathbf{w}\}][\mathbf{w} - \mathbb{E}\{\mathbf{w}\}]^\dagger\right\} = 2\sigma_w^2\mathbf{I}_{mn}, && (3.25)\\
\boldsymbol{\Sigma}_\mathbf{A} &\triangleq \mathbb{E}\left\{\mathbf{A}\mathbf{A}^\dagger\right\} = 2p\sigma_a^2\mathbf{I}_{m}, && (3.26)
\end{aligned}$$

where $\mathbf{I}_{mn}$ is the $q \times q$ identity matrix, where $q = m \times n$, $\mathbf{I}_{mp}$ is the $q \times q$ identity matrix, where $q = m \times p$ and $\mathbf{I}_{pn}$ is the $q \times q$ identity matrix, where $q = p \times n$.

The elements of $\mathbf{Y}$ are i.i.d. so that, for the purpose of differential entropy calculation, we can form the single column vector $\mathbf{y}$, derive the covariance matrix $\boldsymbol{\Sigma}_\mathbf{y}$ and use equation 3.15. In vector form $\mathbf{Y}$ becomes

$$\mathbf{y} = (\mathbf{I}_n \otimes \mathbf{A})\mathbf{x} + \mathbf{w}. \tag{3.27}$$

It will assist the derivations to consider the mean $\mathcal{M}$ and error $\mathcal{E}$ terms for each of the random variables, with

$$\mathbf{Y} = \mathcal{M}_\mathbf{Y} + \mathcal{E}_\mathbf{Y} = (\mathcal{M}_\mathbf{A} + \mathcal{E}_\mathbf{A})(\mathcal{M}_\mathbf{X} + \mathcal{E}_\mathbf{X}) + (\mathcal{M}_\mathbf{W} + \mathcal{E}_\mathbf{W}), \tag{3.28}$$

where the following definitions are employed:

$$\begin{aligned}
\mathcal{M}_{\mathbf{Y}} &\triangleq \mathbb{E}\{\mathbf{Y}\}, && (3.29)\\
\mathcal{M}_{\mathbf{A}} &\triangleq \mathbb{E}\{\mathbf{A}\}, && (3.30)\\
\mathcal{M}_{\mathbf{X}} &\triangleq \mathbb{E}\{\mathbf{X}\}, && (3.31)\\
\mathcal{M}_{\mathbf{W}} &\triangleq \mathbb{E}\{\mathbf{W}\}, && (3.32)\\
\mathcal{E}_{\mathbf{Y}} &\triangleq \mathbf{Y} - \mathcal{M}_{\mathbf{Y}}, && (3.33)\\
\mathcal{E}_{\mathbf{A}} &\triangleq \mathbf{A} - \mathcal{M}_{\mathbf{A}}, && (3.34)\\
\mathcal{E}_{\mathbf{X}} &\triangleq \mathbf{X} - \mathcal{M}_{\mathbf{X}}, && (3.35)\\
\mathcal{E}_{\mathbf{W}} &\triangleq \mathbf{W} - \mathcal{M}_{\mathbf{W}}. && (3.36)
\end{aligned}$$

With the preceding definitions we may now derive the covariance matrix for $\mathbf{y}$ in the four different receiver knowledge states.

State-I: When $\mathbf{A}$ and $\mathbf{X}$ are known, i.e. $\mathcal{E}_{\mathbf{A}} = \mathcal{E}_{\mathbf{X}} = 0$ and, with $\mathbf{W} \sim \mathcal{CN}(0, \mathbf{\Sigma}_{\mathbf{w}})$, we have

$$\begin{aligned}
\mathbf{Y} &= \mathcal{M}_{\mathbf{Y}} + \mathcal{E}_{\mathbf{Y}} = \mathbf{AX} + \mathcal{E}_{\mathbf{W}},\\
\mathbb{E}\{\mathbf{Y}\} &= \mathbf{AX} && (3.37)
\end{aligned}$$

and we find that

$$\begin{aligned}
\mathbb{E}\{\mathbf{y}(\mathbf{A},\mathbf{X})\} &= (\mathbf{I}_n \otimes \mathbf{A})\mathbf{x}, && (3.38)\\
\mathbf{\Sigma}_{\mathbf{y}(\mathbf{A},\mathbf{X})} &= 2\sigma_w^2 \mathbf{I}_{mn}. && (3.39)
\end{aligned}$$

State-II: If $\mathbf{A}$ is known but $\mathbf{X}$ is unknown then, with $\mathcal{E}_{\mathbf{A}} = 0$, we have

$$\begin{aligned}
\mathbf{Y} &= \mathcal{M}_{\mathbf{Y}} + \mathcal{E}_{\mathbf{Y}} = \mathbf{A}(\mathcal{M}_{\mathbf{X}} + \mathcal{E}_{\mathbf{X}}) + (\mathcal{M}_{\mathbf{W}} + \mathcal{E}_{\mathbf{W}})\\
&= \mathbf{A}(\mathcal{M}_{\mathbf{X}} + \mathcal{E}_{\mathbf{X}}) + \mathcal{E}_{\mathbf{W}}, && (3.40)
\end{aligned}$$

which leads to

$$\begin{aligned}
\mathbb{E}\{\mathbf{y}(\mathbf{A})\} &= (\mathbf{I}_n \otimes \mathbf{A})\mathbf{vec}(\mathcal{M}_{\mathbf{X}}),\\
\mathbf{\Sigma}_{\mathbf{y}(\mathbf{A})} &= (\mathbf{I}_n \otimes \mathbf{A})\mathbf{\Sigma}_{\mathbf{x}}(\mathbf{I}_n \otimes \mathbf{A})^{\dagger} + \mathbf{\Sigma}_{\mathbf{w}}. && (3.41)
\end{aligned}$$

The elements of $\mathbf{X}$ are i.i.d. so that $\mathbf{\Sigma_x} = 2\sigma_x^2\mathbf{I}_{pn}$ allowing us to write

$$\mathbf{\Sigma}_{\mathbf{y}(\mathbf{A})} = 2\sigma_x^2(\mathbf{I}_n \otimes \mathbf{A})(\mathbf{I}_n \otimes \mathbf{A})^\dagger + \mathbf{\Sigma_w} = 2\sigma_x^2(\mathbf{I}_n \otimes \mathbf{A}\mathbf{A}^\dagger) + \mathbf{\Sigma_w}. \quad (3.42)$$

If the channel matrix varies between observations then the expected value of $\mathbf{\Sigma}_{\mathbf{y}(\mathbf{A})}$ over $\mathbf{A}$ will be given by

$$\mathbb{E}_{\mathbf{A}}\left\{\mathbf{\Sigma}_{\mathbf{y}(\mathbf{A})}\right\} = 2\sigma_x^2(\mathbf{I}_n \otimes \mathbb{E}\left\{\mathbf{A}\mathbf{A}^\dagger\right\}) + \mathbf{\Sigma_w} = [4p\sigma_a^2\sigma_x^2 + 2\sigma_w^2]\mathbf{I}_{mn}. \quad (3.43)$$

State-III: If $\mathbf{X}$ is known but $\mathbf{A}$ is unknown then it is straightforward to show that

$$\begin{aligned} \mathbb{E}\{\mathbf{y}(\mathbf{X})\} &= (\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{vec}\left(\mathcal{M}_{\mathbf{A}}\right) && (3.44)\\ \mathbf{\Sigma}_{\mathbf{y}(\mathbf{X})} &= (\mathbf{X}^T \otimes \mathbf{I}_m)\mathbf{\Sigma_a}(\mathbf{X}^T \otimes \mathbf{I}_m)^\dagger + \mathbf{\Sigma_w}, && (3.45) \end{aligned}$$

since $\mathcal{M}_{\mathbf{A}} = 0$. Further, if estimation of the channel is performed to obtain $\hat{\mathbf{A}}$ or $\hat{\mathbf{a}}$ and it is known that $\mathbf{\Sigma}_{\hat{\mathbf{a}}} = 2\sigma_{\hat{a}}^2\mathbf{I}_{mp}$, then

$$\mathbf{\Sigma}_{\mathbf{y}(\mathbf{X})} = 2\sigma_{\hat{a}}^2(\mathbf{X}^T \otimes \mathbf{I}_m)(\mathbf{X}^T \otimes \mathbf{I}_m)^\dagger + \mathbf{\Sigma_w} = 2\sigma_{\hat{a}}^2(\mathbf{X}^T\mathbf{X}^* \otimes \mathbf{I}_m) + \mathbf{\Sigma_w}. \quad (3.46)$$

The message matrix varies between observations and the expected value of $\mathbf{\Sigma}_{\mathbf{y}(\mathbf{X})}$ over $\mathbf{X}$ becomes

$$\mathbb{E}_{\mathbf{X}}\left\{\mathbf{\Sigma}_{\mathbf{y}(\mathbf{X})}\right\} = 2\sigma_{\hat{a}}^2(\mathbb{E}\left\{\mathbf{X}^T\mathbf{X}^*\right\} \otimes \mathbf{I}_m) + \mathbf{\Sigma_w} = [4p\sigma_{\hat{a}}^2\sigma_x^2 + \sigma_w^2]\mathbf{I}_{mn}, \quad (3.47)$$

since $\mathbf{X}$ has i.i.d. components with zero mean and variance σ_x^2 and $\mathbf{\Sigma_w} = \sigma_w^2\mathbf{I}_{mn}$.

State-IV: When neither $\mathbf{A}$ nor $\mathbf{X}$ are known then $\mathbf{X}, \mathbf{A}, \mathbf{W}$ are treated as zero-mean, random matrices hence

$$\begin{aligned} \mathbb{E}\{\mathbf{y}\} &= 0. && (3.48)\\ \mathbf{\Sigma_y} &= \mathbb{E}_{\mathbf{A},\mathbf{X}}\left\{\mathbf{y}\mathbf{y}^\dagger\right\} = [4p\sigma_a^2\sigma_x^2 + 2\sigma_w^2]\mathbf{I}_{mn}. && (3.49) \end{aligned}$$

In deriving these covariance expressions we have assumed a Gaussian distribution for the source matrix. However, in practice, message source distributions are usually not Gaussian. Since the covariance expressions will be used to estimate differential entropies, we may substitute the variance σ_x^2 with the source entropy power $\mathcal{P}_x$, a description of entropy power and some related proofs can be found in Rioul [83], i.e.

$$\mathcal{P}_x = \frac{\exp\{2h(x)\}}{2\pi e}, \tag{3.50}$$

where $h(x)$ is the differential entropy for one of the components of $\mathbf{X}$, and then substitute $\mathcal{P}_x \mathbf{I}_{pn}$ for $\mathbf{\Sigma_x}$. In effect we substitute a Gaussian pdf, which has the same differential entropy as the actual pdf. This will allow us to vary the Gaussianity of the source distribution and study the effect on the MI as a function of covariance for an equivalent Gaussian source.

The MI for the legitimate and eavesdropper channels may now be written as

$$I_b = \ln\left(\frac{\det\left(2\sigma_x^2(\mathbf{I}_n \otimes \mathbf{A}_b\mathbf{A}_b^\dagger) + \mathbf{\Sigma_w}\right)}{\det\left(\mathbf{\Sigma_w}\right)}\right) \tag{3.51}$$

and

$$I_e = \ln\left(\frac{\det\left([4p\sigma_a^2\sigma_x^2 + 2\sigma_w^2]\mathbf{I}_{m_e n}\right)}{\det\left(2\sigma_{\hat{a}}^2[\mathbf{X}^T\mathbf{X}^* \otimes \mathbf{I}_{m_e}] + \mathbf{\Sigma_w}\right)}\right). \tag{3.52}$$

When we consider that the channel and message matrices vary with time i.e. between observation blocks, the time averaged mutual information becomes

$$\hat{I}_b = \mathbb{E}_{\mathbf{A},\mathbf{X}}\left\{I_b\right\} = \ln\left(\frac{\det\left(\mathbb{E}_{\mathbf{A}}\left\{\Sigma_{\mathbf{y}(\mathbf{A})}\right\}\right)}{\det\left(\Sigma_{\mathbf{w}}\right)}\right) = m_b n \ln\left(\frac{[2p\sigma_a^2\sigma_x^2 + \sigma_w^2]}{\sigma_w^2}\right) \tag{3.53}$$

and

$$\hat{I}_e = \mathbb{E}_{\mathbf{A},\mathbf{X}}\left\{I_e\right\} = \ln\left(\frac{\det\left(\Sigma_{\mathbf{y}}\right)}{\det\left(\mathbb{E}_{\mathbf{X}}\left\{\Sigma_{\mathbf{y}(\mathbf{X})}\right\}\right)}\right) = m_e n \ln\left(\frac{[2p\sigma_a^2\sigma_x^2 + \sigma_w^2]}{[2p\sigma_{\hat{a}}^2\sigma_x^2 + \sigma_w^2]}\right). \tag{3.54}$$

In this case it is assumed that both $\mathbf{X}$ and $\mathbf{A}$ are random variables that are independent between observation blocks.

It is clear, from the above that, when $m_e = m_b$, $\hat{I}_e \to \hat{I}_b$ as $\sigma_{\hat{a}}^2 \to 0$. Since one of our assumptions is that the elements of $\mathbf{X}$ are i.i.d. then $\mathbf{\Sigma_x}$ is a multiple of the identity, $\mathbf{\Sigma_x} = 2\sigma_x^2\mathbf{I}_{pn}$. Similarly $\mathbf{\Sigma_w} = 2\sigma_w^2\mathbf{I}_{mn}$. We have also assumed that both channels $\mathbf{A}_b$ and $\mathbf{A}_e$ are full rank. If the rank of Bob's channel, rk($\mathbf{A}_b$), reduces then $\hat{I}_b$ will decrease, which may give Eve an advantage if she can maintain a full channel rank, rk$(\mathbf{A}_e)$. If Eve finds that rk$(\mathbf{A}_e)$ is insufficient then she may be able take steps to improve her channel e.g. relocate her antennae.

In Chapter 4 we shall derive a lower variance bound for the covariance matrix of the estimated, complex mixing matrix when performing joint source and channel estimation. This value will be used to obtain an approximation for $\sigma^2_{\hat{a}}$.

3.3 Mutual Information Gradient

In studying the linear vector Gaussian channel $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, Palomar and Verdú in [77] have derived expressions that relate the gradient of mutual information, as a function of a variety of system parameters, to the error covariance matrix $\mathbf{E}$ of the estimate of the input $\mathbf{x}$ given the output $\mathbf{y}$. This will be used to give an indication as to how rapidly the information rate, to an eavesdropper, increases as the source pdf becomes less Gaussian. We shall make use the following result in [77, eqn.25]

$$([77,\text{ eqn.25}]) \quad \nabla_{\boldsymbol{\Sigma}_\mathbf{x}} I(\mathbf{x}; \mathbf{H}\mathbf{B}\mathbf{x} + \mathbf{n}) = \mathbf{B}^\dagger \mathbf{H}^\dagger \boldsymbol{\Sigma}_\mathbf{n}^{-1} \mathbf{H}\mathbf{B}\mathbf{E}\boldsymbol{\Sigma}_\mathbf{x}^{-1}, \tag{3.55}$$

where $\mathbf{H}$ is the channel matrix, $\mathbf{B}$ is a linear precoding matrix and $\mathbf{n}$ is Gaussian noise. This allows us to write the gradient of MI with respect to (w.r.t.) $\boldsymbol{\Sigma}_\mathbf{x}$ as

$$\nabla_{\boldsymbol{\Sigma}_\mathbf{x}} I(\mathbf{x}; \mathbf{A}\mathbf{x} + \mathbf{w}) = \mathbf{A}^\dagger \boldsymbol{\Sigma}_\mathbf{w}^{-1} \mathbf{A}\mathbf{E}\boldsymbol{\Sigma}_\mathbf{x}^{-1}, \tag{3.56}$$

where $\boldsymbol{\Sigma}_\mathbf{x}$ and $\boldsymbol{\Sigma}_\mathbf{w}$ are, respectively, the input and noise covariance matrices. The Minimum Mean Squared Error (MMSE) matrix $\mathbf{E}$ is defined as

$$\mathbf{E} \triangleq \mathbb{E}\left\{[\mathbf{x} - \mathbb{E}\{\mathbf{x}|\mathbf{y}\}][\mathbf{x} - \mathbb{E}\{\mathbf{x}|\mathbf{y}\}]^\dagger\right\}. \tag{3.57}$$

For the block data case, $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{W}$, that we are considering we may write

$$\mathbf{y} = \mathbf{vec}\left(\mathbf{Y}\right) = (\mathbf{I}_n \otimes \mathbf{A})\mathbf{x} + \mathbf{w} = \mathbf{B}\mathbf{x} + \mathbf{w}, \tag{3.58}$$

where $\mathbf{B} = (\mathbf{I}_n \otimes \mathbf{A})$, so that the MI gradient is

$$\nabla_{\boldsymbol{\Sigma}_\mathbf{x}} I(\mathbf{x}; \mathbf{B}\mathbf{x} + \mathbf{w}) = \mathbf{B}^\dagger \boldsymbol{\Sigma}_\mathbf{w}^{-1} \mathbf{B}\mathbf{E}\boldsymbol{\Sigma}_\mathbf{x}^{-1}. \tag{3.59}$$

Definition 1 (Fully Informed Mutual Information Gradient)**.** *The MI gradient for the fully informed receiver Bob is defined as*

$$MIG_b \triangleq \mathbf{B}_b^\dagger \mathbf{\Sigma}_\mathbf{w}^{-1} \mathbf{B}_b \mathbf{E}_b \mathbf{\Sigma}_\mathbf{x}^{-1}, \tag{3.60}$$

where $\mathbf{E}_b$ is the fully informed receiver's source estimation error covariance matrix and $\mathbf{B}_b \triangleq (\mathbf{I}_n \otimes \mathbf{A}_b)$. In terms of the receiver knowledge states defined earlier, Bob's mutual information is given by

$$I_b = h(II) - h(I). \tag{3.61}$$

Definition 2 (Partially Informed Mutual Information Gradient)**.** *The MI gradient for the partially informed receiver Eve is defined as*

$$MIG_e \triangleq \mathbf{B}_e^\dagger \mathbf{\Sigma}_\mathbf{w}^{-1} \mathbf{B}_e \mathbf{E}_e \mathbf{\Sigma}_\mathbf{x}^{-1}, \tag{3.62}$$

where $\mathbf{E}_e$ is the eavesdropper's source estimation error covariance matrix and $\mathbf{B}_e \triangleq (\mathbf{I}_n \otimes \mathbf{A}_e)$. In terms of the receiver knowledge states defined earlier, Eve's mutual information is given by

$$I_b = h(IV) - h(III). \tag{3.63}$$

Definition 3 (Mutual Information Gradient Ratio)**.** *The partially informed to fully informed MI gradient ratio is defined as*

$$\begin{aligned} MIGR &\triangleq tr\left([MIG_e]\,[MIG_b]^{-1}\right), & (3.64)\\ &= tr\left(\left[\mathbf{E}_e\mathbf{E}_b^{-1}\right]\left[\mathbf{B}_b^\dagger \mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{B}_b\right]^{-1}\left[\mathbf{B}_e^\dagger \mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{B}_e\right]\right). & (3.65)\end{aligned}$$

Now, since the noise and source vectors are i.i.d. we have $\mathbf{\Sigma}_\mathbf{w} = 2\sigma_w^2 I_{mn}$ and $\mathbf{\Sigma}_\mathbf{x} = 2\sigma_x^2 I_{mn}$ so that

$$\begin{aligned} \text{MIGR} &= \text{tr}\left(\left[\mathbf{E}_e\mathbf{E}_b^{-1}\right]\left[\mathbf{B}_b^\dagger\mathbf{B}_b\right]^{-1}\left[\mathbf{B}_e^\dagger\mathbf{B}_e\right]\right), & (3.66)\\ &= \text{tr}\left(\left[\mathbf{E}_e\mathbf{E}_b^{-1}\right]\left[\mathbf{I}_n \otimes \mathbf{A}_b^\dagger\mathbf{A}_b\right]^{-1}\left[\mathbf{I}_n \otimes \mathbf{A}_e^\dagger\mathbf{A}_e\right]\right). & (3.67)\end{aligned}$$

If Bob and Eve's receive antennae both experience a similar RF propagation environment: sufficient multipath , no strong direct RF propagation paths and similar background noise statistics, then the channels $\mathbf{A}_b$ and $\mathbf{A}_e$ will have similar statistics. As described in section 3.1 the channel may be modelled with components $a_{i,j} \sim \mathcal{CN}(0, 2\sigma_a^2)$. Let $\mathbb{E}\left\{\mathbf{A}_b^\dagger \mathbf{A}_b\right\} = 2m_b\sigma_a^2 I_p$ and $\mathbb{E}\left\{\mathbf{A}_e^\dagger \mathbf{A}_e\right\} = 2m_e\sigma_a^2 I_p$, where the expectation is taken over a number of observation blocks.

Theorem 3.3.1 (MIGR Theorem)**.** *Consider the MIMO model given by equation 3.1, assumptions listed in section 3.1, $[\mathbf{A}_e]_{ij}$ and $[\mathbf{A}_b]_{ij}$ are both distributed as $\mathcal{CN}(0, 2\sigma_a^2)$, then the expected value of the ratio MIGR is a function of the source estimation error covariance matrices and the receive array dimensions m_b and m_e.*

Proof of MIGR Theorem. When $[\mathbf{A}_e]_{ij}$ and $[\mathbf{A}_b]_{ij}$ are both distributed as $\mathcal{CN}(0, 2\sigma_a^2)$, $\mathbb{E}\left\{\mathbf{A}_b^\dagger \mathbf{A}_b\right\} = 2m_b\sigma_a^2 I_p$ and $\mathbb{E}\left\{\mathbf{A}_e^\dagger \mathbf{A}_e\right\} = 2m_e\sigma_a^2 I_p$. From the model assumptions we also have $\mathbf{\Sigma_w} = 2\sigma_w^2 I_{mn}$, so that the expected value of MIGR, as given in Definition 3, becomes

$$\begin{aligned}
\mathbb{E}_{\mathbf{A}}\left\{\text{MIGR}\right\} &= \operatorname{tr}\left(\left[\mathbf{E}_e\mathbf{E}_b^{-1}\right]\mathbb{E}\left\{\left[\mathbf{I}_n \otimes \mathbf{A}_b^\dagger \mathbf{A}_b\right]^{-1}\right\}\mathbb{E}\left\{\left[\mathbf{I}_n \otimes \mathbf{A}_e^\dagger \mathbf{A}_e\right]\right\}\right) && (3.68)\\
&= \operatorname{tr}\left(\left[\mathbf{E}_e\mathbf{E}_b^{-1}\right]\left[\frac{1}{2m_b\sigma_a^2}\mathbf{I}_{pn}\right]\left[2m_e\sigma_a^2\mathbf{I}_{pn}\right]\right), && (3.69)\\
&= \frac{m_e}{m_b}\operatorname{tr}\left(\mathbf{E}_e\mathbf{E}_b^{-1}\right). && (3.70)
\end{aligned}$$

□

If the components of the source estimates are independent then $\mathbf{E}_b = 2\sigma_{x|y,A}^2\mathbf{I}_{pn}$ and $\mathbf{E}_e = 2\sigma_{x|y}^2\mathbf{I}_{pn}$. In this case the expected value of MIGR becomes

$$\mathbb{E}_{\mathbf{A}}\left\{\text{MIGR}\right\} = \frac{m_e}{m_b} pn \frac{\sigma_{x|y}^2}{\sigma_{x|y,A}^2}. \tag{3.71}$$

We have also assumed that both channels $\mathbf{A}_b$ and $\mathbf{A}_e$ are full rank. If the rank of Bob's channel, rk$(\mathbf{A}_b)$, reduces then $\hat{I}_b$ will decrease, similarly if Eve finds that rk$(\mathbf{A}_e)$ is reduced then $\hat{I}_e$ will decrease. However, since we are dealing with MI gradients and gradient ratios here, the results for MIG$_b$, MIG$_e$ and MIGR will be unchanged.

3.4 Unknown Unitary Transformation

Here we consider the situation where a communications receiving system has prior knowledge of the message symbol set, the channel matrix between the transmission system and the receiving system, is able to resolve the transmissions from the, assumed independent, transmitter antennae but does not know the unitary transformation that has been applied at the transmitter. The question then becomes: what is the mutual information available to the receiver when an unknown unitary transformation matrix is employed by the transmitter?

In the following sections we derive expressions for differential entropy and mutual information for a multi-element transmit array to multi-element receive array system, where the transmitter and receiver have the same number N of antennae, which we shall refer to as an N-Dimensional (N-D) system.

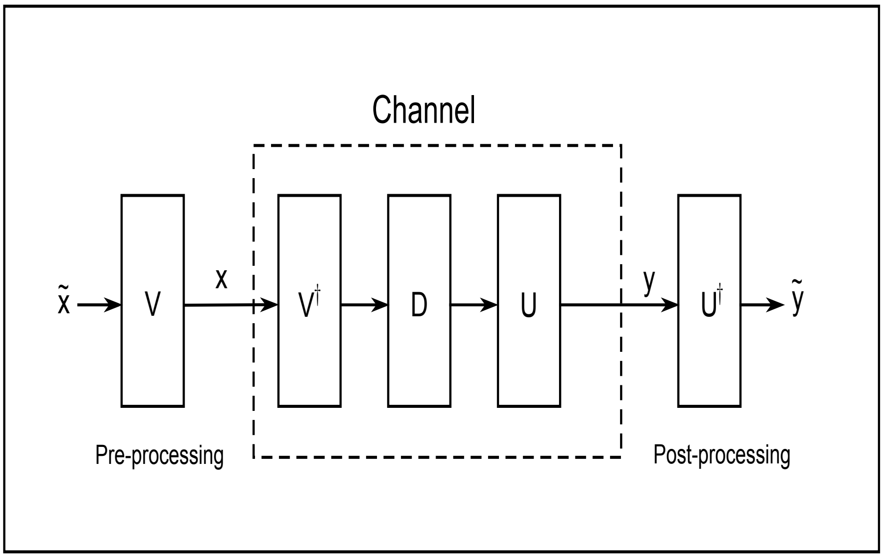

Figure 3.3: Converting a MIMO channel to a parallel channel via SVD.

The vector model that we shall base further derivations on is the simple linear transformation

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{3.72}$$

where $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is the received signal vector, $\mathbf{x} \in \mathbb{C}^{N\times 1}$ is the transmitted vector, $\mathbf{w} \in \mathbb{C}^{N\times 1}$ is additive receiver noise and $\mathbf{A} \in \mathbb{C}^{N\times N}$ is the channel gain or mixing matrix between the transmitter and receiver. A common MIMO channel model was discussed earlier in Section 3.1.

The channel matrix can be factorized using SVD as : $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger}$ and we can then use, e.g. see Tse and Viswanath [102, Ch.7]:

$$\begin{aligned} \mathbf{U}^{\dagger}\mathbf{y} &= \mathbf{D}\mathbf{V}^{\dagger}\mathbf{x} + \mathbf{U}^{\dagger}\mathbf{w} \qquad (3.73) \\ \text{or} \qquad \tilde{\mathbf{y}} &= \mathbf{D}\tilde{\mathbf{x}} + \tilde{\mathbf{w}}, \qquad (3.74) \end{aligned}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices. This allows us to view the MIMO system as if it were composed of a set of parallel channels and the input data vector can be designed with this in mind. Figure 3.3 shows how this channel, with pre and post-processing, may be configured. For such an approach to work the transmitter requires precise knowledge of the channel matrix and it is a simple matter for the intended receiver to obtain the (scaled) message, since $\mathbf{D}$ is a real diagonal matrix. However for an unintended receiver, with a different (known) channel matrix, an unknown unitary transformation has been applied. In this case we desire to know how the eavesdrop channel mutual information, is affected. We make the following assumptions:

- $\mathbf{y}$ is a proper complex $N \times 1$ observation vector.
- $\mathbf{w}$ is a proper complex $N \times 1$ random Gaussian noise vector, $w_i \sim \mathcal{CN}(0, 2\sigma_w^2)$.
- $\mathbf{x}$ is a proper complex $N \times 1$ vector that defines a set of points on the surface of an N-D hypersphere i.e. $\|\mathbf{x}\| = \sqrt{\mathbf{x}^{\dagger}\mathbf{x}} = r_0 =$ constant. This definition for $\mathbf{x}$ does not model general sources and it is assumed that σ_x is known.
- the intended channel $\mathbf{A}_b$ is known to both Alice and Bob.
- Eve knows the intercept channel $\mathbf{A}_e$ but not the intended channel.

Eve attempts to estimate the signal vector by applying the channel inverse as

$$\hat{\mathbf{x}} = \mathbf{A}_e^{-1}\mathbf{y}_e = \mathbf{V}\tilde{\mathbf{x}} + \mathbf{A}_e^{-1}\mathbf{w}_e. \qquad (3.75)$$

Eve is therefore unable to directly obtain $\tilde{\mathbf{x}}$ due to the unknown unitary matrix $\mathbf{V}$. In applying the channel inverse, the noise vector has also been scaled and the modified noise covariance term $\mathbf{A}_e^{-1}\mathbf{\Sigma}_{\mathbf{w}_e}\mathbf{A}_e^{-T}$ shows that the intercept receiver may be operating with a different snr to that of the intended receiver. This also indicates that Eve could obtain better mutual information with a better channel.

Optimal power allocation to the parallel channels between Alice and Bob would typically be implemented via a technique called waterfilling, e.g. see Tse and Viswanath [102, Ch.5] for a description, and hence lead to optimal system capacity. We have not taken waterfilling into account in this study and simply assume that equal power is assigned to each of the parallel channels. We could proceed to derive the eavesdropper MI in a cartesian or a polar coordinate system. Of course it does not matter which coordinate system we choose - we should get the same answer. It is well known that differential entropy involves a Jacobian (J) in the transformation of coordinates, e.g. see Papoulis [78], leading to a $\ln\det(J)$ term but this will cancel in the MI calculations because MI is a relative entropy i.e. the difference between two entropies. For the purpose of the current analysis our derivations will be based on a cartesian coordinate system. Since the channels are assumed known we may consider $\mathbf{y} = \mathbf{x} + \mathbf{w}$ to represent the fully informed (unitary transformation known) case and $\mathbf{y} = \mathbf{V}\mathbf{x} + \mathbf{w}$ to represent the partially informed (unitary transformation unknown) case. We can write $\mathbf{x} = \frac{\mathbf{x}}{\|\mathbf{x}\|}\|\mathbf{x}\|$ to obtain

$$\mathbf{y} = \mathbf{V}\frac{\mathbf{x}}{\|\mathbf{x}\|}\|\mathbf{x}\| + \mathbf{w} = \mathbf{v}r_0 + \mathbf{w} \tag{3.76}$$

where $r_0 = \|\mathbf{x}\|$ and $\mathbf{v} = V\frac{\mathbf{x}}{\|\mathbf{x}\|}$ is a unit vector for which we may or may not know the rotations. For the random vectors $\mathbf{y}$ and $\mathbf{x}$ the mutual information for the fully informed model is given by:

$$I_F = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}, \mathbf{V}), \tag{3.77}$$

or in terms of the receiver knowledge states defined previously

$$I_F = h(II) - h(I) \tag{3.78}$$

and for the partially informed model the mutual information is obtained from:

$$I_P = h(\mathbf{y}) - h(\mathbf{y}|r_0) \tag{3.79}$$

where the message amplitude r_0 is known but not the unitary transformation. In terms of the predefined receiver knowledge states

$$I_P = h(II) - h(VI). \tag{3.80}$$

It is well known that an $n \times n$ unitary matrix $\mathbf{V}$ (with $\mathbf{V}^{\dagger}\mathbf{V} = \mathbf{V}\mathbf{V}^{\dagger} = \mathbf{I}_n$) preserves length i.e.

$$\|\mathbf{V}\mathbf{x}\| = \|\mathbf{x}\| \tag{3.81}$$

so that, if we know the magnitudes of the x_i in $\mathbf{y} = \mathbf{V}\mathbf{x}$, then we know that the Euclidean norm of $\mathbf{x}$ is unchanged by the unitary transformation

$$(\mathbf{V}\mathbf{x})^{\dagger}(\mathbf{V}\mathbf{x}) = \mathbf{x}^{\dagger}\mathbf{x} = \|\mathbf{x}\|^2 = \sum_i |x|_i^2. \tag{3.82}$$

There are two cases that we consider where such a unitary transformation affects an eavesdropper. The first case occurs when a SVD has been applied by the transmitter and this was described earlier, the second case occurs when BSS techniques are implemented by an eavesdropper and this will be discussed later in Chapter 4.

3.5 Hypersphere Model for Mutual Information

In this section we shall derive entropy and MI expressions, for the model described in section 3.4, using the concept of a hypersphere to represent the pdfs. For the simple model

$$\mathbf{y} = \mathbf{V}\mathbf{x} + \mathbf{w}, \tag{3.83}$$

In this section we treat all of $\mathbf{y}$, $\mathbf{x}$, and $\mathbf{w}$ as real-valued random variables and $\mathbf{V}$ as a real orthogonal transformation matrix. The channel matrix $\mathbf{A}$ is also treated as real-valued. The benefit of this approach will be to simplify the derivations whilst recognising that, if their complex-valued counterparts are proper complex i.i.d. random variables, then they could be treated as real by forming composite vectors of their real and imaginary parts. We can construct the joint density function beginning with

$$p(\mathbf{y}|\mathbf{x}) = (2\pi\sigma_w^2)^{\frac{-N}{2}} \exp\left\{\frac{-[\mathbf{y}-\mathbf{x}]^T[\mathbf{y}-\mathbf{x}]}{2\sigma_w^2}\right\}. \tag{3.84}$$

To illustrate the consequence of not knowing the rotation imposed by a unitary (orthogonal in the real-valued model) transformation $\mathbf{V}$ in the 2D real-valued model, Figure 3.4 shows a message symbol set where each of the two transmitters

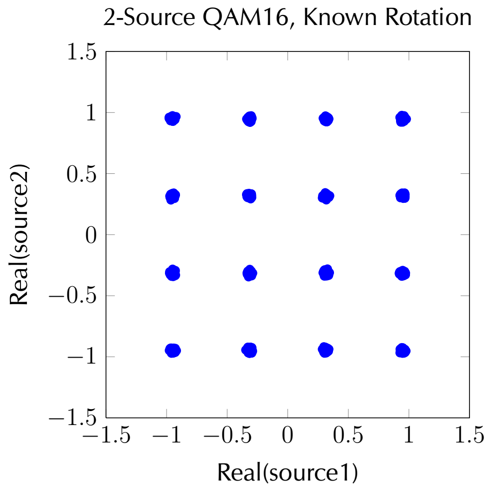

Figure 3.4: 2D Transmitter message symbol set.

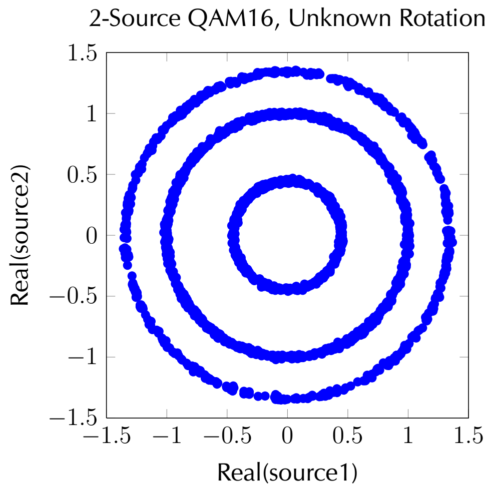

Figure 3.5: Received ring distribution caused by unknown rotation on message symbol set.

can output one of four possible values. Thus a constellation containing 16 points may be observed at the receiver and the density of these points is determined by the additive noise. If the orthogonal transformation $\mathbf{V}$ or rotation is unknown but the amplitude levels are known then the receiver might obtain a message that looks something like Figure 3.5 where the density, or thickness, of the rings is determined by the additive noise.

In this section we derive the general form for $p(\mathbf{y}|r_0)$ thus allowing us to obtain the MI for any dimension and snr. The derivation utilises a result by Vesely [103] which shows how integration to obtain the probability over an N-D spherical surface can be performed as an integral over a single sphere dimension. This result greatly simplifies the multidimensional integrals that we require to solve. The surface area, $S_N(r_0)$, of an N-D sphere, as a function of radius $r_0 = \|\mathbf{x}\|$, may be represented by

$$[103, \text{eqn.3.15}] \quad S_N(r_0) = \int_{-r_0}^{r_0} \frac{r_0 S_{N-1}(r_2)}{r_2} dx_1 \tag{3.85}$$

where $r_2 = \sqrt{r_0^2 - x_1^2}$. We can rewrite the above as

$$1 = \int_{-r_0}^{r_0} \frac{r_0 S_{N-1}(r_2)}{r_2 S_N(r_0)} dx_1 = \int_{-r_0}^{r_0} p_N(x_1) dx_1 \tag{3.86}$$

so that, for points $\mathbf{x} = [x_1 \ldots x_N]^T$, which are homogeneously distributed on an N-D spherical surface, a single x_i occurs with probability $p_N(x_1)$. Now

$$p_N(x_1) = \frac{r_0 S_{N-1}(r_2)}{r_2 S_N(r_0)} = \frac{(N-1)C_{N-1}r_2^{N-3}}{NC_N r_0^{N-2}} = \frac{(N-1)C_{N-1}}{NC_N}\frac{1}{r_0}\left[1 - \frac{x_1^2}{r_0^2}\right]^{\frac{N-3}{2}}, \tag{3.87}$$

where

$$C_N = \frac{2\pi^{N/2}}{N\Gamma(N/2)}. \tag{3.88}$$

The pdf in equation 3.84 may be written in the form

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= (2\pi\sigma_w^2)^{-N/2} \exp\left\{\frac{-\|\mathbf{y}\|^2 - \|\mathbf{x}\|^2}{2\sigma_w^2}\right\} \exp\left\{\frac{\sum_{i=1}^N x_i y_i}{\sigma_w^2}\right\} \\ &= (2\pi\sigma_w^2)^{-N/2} \exp\left\{\frac{-\|\mathbf{y}\|^2 - \|\mathbf{x}\|^2}{2\sigma_w^2}\right\} \exp\left\{\frac{\mathbf{x}\,\boldsymbol{.}\,\mathbf{y}}{\sigma_n^2}\right\}, \end{aligned} \tag{3.89}$$

from which we wish to obtain $p(\mathbf{y}|\, r_0)$. Assuming now that $\|\mathbf{x}\| = r_0$ is given we obtain $p(\mathbf{y}|r_0)$ by integrating over $\mathbf{x}$ as follows

$$\begin{aligned} p(\mathbf{y}|r_0) &= \int_{\|\mathbf{x}\|=r_0} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x} \\ &= (2\pi\sigma_w^2)^{-N/2} \exp\left\{\frac{-\|\mathbf{y}\|^2 - r_0^2}{2\sigma_w^2}\right\} \int_{\|\mathbf{x}\|=r_0} \exp\left\{\frac{\mathbf{x}\,\boldsymbol{.}\,\mathbf{y}}{\sigma_w^2}\right\} p(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{3.90}$$

We proceed to calculate this integral by first noting that, since the points $\mathbf{x}$ are uniformly distributed over the surface of an N-D sphere, we only need to perform the integral along a single dimension, e.g. x_1 and replace $p(\mathbf{x})$ with $p_N(x_1)$ using equation 3.87 derived earlier. To better understand this, consider the dot product $\mathbf{x}\,\boldsymbol{.}\,\mathbf{y}$. The dot product will be unchanged if both vectors are operated on by the same orthogonal transformation. Let the orthogonal transformation matrix be $\mathcal{R} \in \mathbb{R}^{N\times N}$, then

$$(\mathcal{R}\mathbf{x})\,\boldsymbol{.}\,(\mathcal{R}\mathbf{y}) = (\mathcal{R}\mathbf{x})^T(\mathcal{R}\mathbf{y}) = \mathbf{x}^T\mathcal{R}^T\mathcal{R}\mathbf{y} = \mathbf{x}^T\mathbf{y} = \mathbf{x}\,\boldsymbol{.}\,\mathbf{y}, \tag{3.91}$$

since $\mathcal{R}\mathcal{R}^T = \mathcal{R}\mathcal{R}^{-1} = I$. So we are free to choose any orthogonal transformation matrix and the integral will be unaffected. Let us choose $\mathcal{R}$ such that $\mathcal{R}\mathbf{y} = \|\mathbf{y}\|[1, 0, \ldots, 0]^T = \|\mathbf{y}\|\mathbf{e}$, where $\mathbf{e}$ is a unit vector, i.e. the vector $\mathbf{y}$ is rotated to lie along the y_1 axis. Let $\mathbf{x}' = (\mathcal{R}\mathbf{x})$ then we have

$$\mathbf{x}' \cdot (\mathcal{R}\mathbf{y}) = \mathbf{x}' \cdot \|\mathbf{y}\|\mathbf{e} = \|\mathbf{y}\|(\mathbf{x}')^T\mathbf{e} = \|\mathbf{y}\|x_1'. \tag{3.92}$$

Hence

$$\begin{aligned}\int_{\|\mathbf{x}\|=r_0} \exp\left\{\frac{\mathbf{x} \cdot \mathbf{y}}{\sigma_w^2}\right\} p(\mathbf{x})d\mathbf{x} &= \int_{-r_0}^{r_0} p_N(x_1') \exp\left\{\frac{\|\mathbf{y}\|x_1'}{\sigma_w^2}\right\} dx_1' \\ &= \frac{(N-1)C_{N-1}}{NC_N}\frac{1}{r_0}\int_{-r_0}^{r_0}\left[1 - \frac{x_1'^2}{r_0^2}\right]^{\frac{N-3}{2}} \exp\left\{\frac{\|\mathbf{y}\|x_1'}{\sigma_w^2}\right\} dx_1'.\end{aligned} \tag{3.93}$$

We may make a change of variable by letting $z = \frac{x_1'}{r_0}$ to get

$$\int_{\|\mathbf{x}\|=r_0} \exp\left\{\frac{\mathbf{x} \cdot \mathbf{y}}{\sigma_w^2}\right\} p(\mathbf{x})d\mathbf{x} = \frac{(N-1)C_{N-1}}{NC_N}\int_{-1}^{1}\left[1 - z^2\right]^{\frac{N-3}{2}} \exp\left\{\frac{\|\mathbf{y}\|r_0 z}{\sigma_w^2}\right\} dz. \tag{3.94}$$

We make use of an integral form of the modified Bessel function of the first kind [1]:

$$I_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma(\nu + 1/2)}\int_{-1}^{1}\left(1 - t^2\right)^{\nu - 1/2} e^{\pm zt}dt, \qquad \Re(\nu) > \frac{-1}{2}. \tag{3.95}$$

So that

$$\int_{-1}^{1}\left[1 - z^2\right]^{\frac{N-3}{2}} \exp\left\{\frac{\|\mathbf{y}\|r_0 z}{\sigma_w^2}\right\} dz = \frac{\pi^{1/2}\Gamma\left(\frac{N-1}{2}\right) I_{\frac{N}{2}-1}(\lambda)}{\left(\frac{\lambda}{2}\right)^{\frac{N}{2}-1}}, \tag{3.96}$$

where $\lambda = \frac{\|\mathbf{y}\|r_0}{\sigma_w^2}$ and since

$$\frac{(N-1)C_{N-1}}{NC_N} = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)\pi^{1/2}}, \tag{3.97}$$

then equation 3.94 may be written as

$$\int_{\|\mathbf{x}\|=r_0} \exp\left\{\frac{\mathbf{x}\cdot\mathbf{y}}{\sigma_w^2}\right\} p(\mathbf{x})d\mathbf{x} = \frac{\Gamma\left(\frac{N}{2}\right)2^{\frac{N}{2}-1}I_{\frac{N}{2}-1}(\lambda)}{\lambda^{\frac{N}{2}-1}}. \tag{3.98}$$

The general form for the density, given r_0, is therefore

$$p(\mathbf{y}|r_0) = (2\pi\sigma_w^2)^{-\frac{N}{2}} \exp\left\{\frac{-{r_0}^2 - \|\mathbf{y}\|^2}{2\sigma_w^2}\right\} \frac{\Gamma\left(\frac{N}{2}\right)2^{\frac{N}{2}-1}I_{\frac{N}{2}-1}(\lambda)}{\lambda^{\frac{N}{2}-1}}. \tag{3.99}$$

The entropy calculation involves a multidimensional integration over the components in $\mathbf{y}$:

$$h(\mathbf{y}|r_0) = -\int_{\mathbf{y}} p(\mathbf{y}|r_0) \ln p(\mathbf{y}|r_0) d\mathbf{y}. \tag{3.100}$$

It has not been possible to find a closed form solution for the integral in equation 3.100 and so it was necessary to calculate it numerically. $h(\mathbf{y}|r_0)$ is the entropy in the observed data given knowledge only of the source magnitudes and is important because it is required for the calculation of eavesdropper MI in section 3.6. Therefore a receiver, that has prior knowledge of the message symbol set but is unable to resolve a unitary transformation that has been applied by the transmitter, may be expected to reduce the uncertainty in their observations to $h(\mathbf{y}|r_0)$, at best. This situation also occurs when the receiver applies a BSS algorithm, discussed later in Chapter 4, where ambiguity in the resolved sources takes the form of a unitary transformation.

3.6 Numerical Calculations and High SNR Approximation

Now that we have a general form for $p(\mathbf{y}|r_0)$ we may proceed to derive the MI for both the fully informed case and the partially informed (amplitude only) case. For the orthogonal transformation model given by equation 3.83, the differential entropies in the fully informed case are:

$$\begin{aligned} h(\mathbf{y}|\mathbf{x}) &= \frac{N}{2}\ln(2\pi e\sigma_w^2), \\ h(\mathbf{y}) &= \frac{N}{2}\ln(2\pi e\sigma_y^2), \end{aligned} \tag{3.101}$$

where $\sigma_y^2 = \sigma_x^2 + \sigma_w^2$. Hence the the fully informed mutual information is

$$I_F = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) = \frac{N}{2} \ln \left(\frac{\sigma_y^2}{\sigma_w^2} \right) = \frac{N}{2} \ln (\rho + 1) \, \text{Nats s}^{-1}. \tag{3.102}$$

Alternatively

$$I_F = \frac{N}{2} \log_2 (\rho + 1) \, \text{Bits s}^{-1}. \tag{3.103}$$

When only the signal amplitude is given, the partially informed MI is

$$I_P = h(\mathbf{y}) - h(\mathbf{y}|r_0), \tag{3.104}$$

where $h(\mathbf{y}|r_0)$ is given by equation 3.100.

At high snr $\|\mathbf{y}\| \approx r_0$ so that $\lambda \approx \frac{r_0^2}{\sigma_w^2} = \rho$ and, when ρ is sufficiently large

$$I_\nu(\lambda) \approx \frac{e^\lambda}{(2\pi\lambda)^{\frac{1}{2}}}. \tag{3.105}$$

So, at high snrs,

$$\begin{aligned} p(\mathbf{y}|r_0) &\approx (2\pi\sigma_w^2)^{-\frac{N}{2}} \exp\left\{ \frac{-{r_0}^2 - \|\mathbf{y}\|^2}{2\sigma_w^2} \right\} \frac{\Gamma\left(\frac{N}{2}\right) 2^{\frac{N-3}{2}} e^\lambda}{\lambda^{\frac{N-1}{2}} \pi^{\frac{1}{2}}} && (3.106) \\ &= (2\pi\sigma_w^2)^{-\frac{1}{2}} \exp\left\{ \frac{-{r_0}^2 - \|\mathbf{y}\|^2 + 2r_0\|\mathbf{y}\|}{2\sigma_w^2} \right\} \frac{\Gamma\left(\frac{N}{2}\right)}{2\pi^{\frac{N}{2}} r_0^{N-1}} && (3.107) \\ &= (2\pi\sigma_w^2)^{-\frac{1}{2}} \exp\left\{ \frac{-(\|\mathbf{y}\| - r_0)^2}{2\sigma_w^2} \right\} \frac{\Gamma\left(\frac{N}{2}\right)}{2\pi^{\frac{N}{2}} r_0^{N-1}}. && (3.108) \end{aligned}$$

Now the surface area, $S_N(r_0)$, of an N-dimensional sphere, with radius r_0, is given by Sommerville [91] as

$$S_N(r_0) = \frac{2\pi^{\frac{N}{2}} r_0^{N-1}}{\Gamma\left(\frac{N}{2}\right)} \tag{3.109}$$

and we note that, at high snrs, $p(\mathbf{y}|r_0)$ factors as the product of two distributions: a normal distribution for the magnitude and a uniform distribution on the surface of an N-sphere

$$p(\mathbf{y}|r_0) \approx \mathcal{N}(r_0, \sigma_w^2) \left(\frac{1}{S_N(r_0)} \right). \tag{3.110}$$

Imagining an N-D "fuzzy" shell, we might therefore interpret $p(\mathbf{y}|r_0)$ as

$$p(\mathbf{y}|r_0) \approx p(\text{normal})p(\text{surface}), \tag{3.111}$$

where $p(\text{normal})$ = probability of position normal to shell surface and $p(\text{surface})$ = probability of position on shell surface. At high snr the differential entropy for the distribution $p(\mathbf{y}|r_0)$ is therefore approximately equal to the sum of the entropies for the two factored distributions $p(\text{normal})$ and $p(\text{surface})$:

$$h(\mathbf{y}|r_0) \approx \ln\left(\frac{2\pi^{\frac{N}{2}} r_0^{N-1}}{\Gamma\left(\frac{N}{2}\right)}\right) + \frac{1}{2}\ln\left(2\pi e \sigma_w^2\right), \tag{3.112}$$

The partially informed mutual information may now be approximated as

$$\begin{aligned} I_P &= h(\mathbf{y}) - h(\mathbf{y}|r_0) \\ &\approx \ln\left(\frac{\sigma_y^N}{r_0^{N-1}\sigma_w}\right) + \frac{1}{2}\ln\left(\pi^{-1}2^{N-3}e^{N-1}\right) + \ln\left(\Gamma\left(\frac{N}{2}\right)\right) \text{ Nats s}^{-1}. \end{aligned} \tag{3.113}$$

In Figure 3.6 some high snr estimates for $h(\mathbf{y}|r_0)$ are compared with their numerically calculated equivalents, showing an improving fit as the snr increases. In all cases the error improves as the snr increases. As the dimensionality increases the estimate requires a higher snr to achieve a smaller error.

The fully informed MI, equation 3.102, for dimensions two to five has been calculated and the results are presented in Figure 3.7. snr is shown as $10\log_{10}(\rho)$ and MI values have been converted to $\log_2$ values using $\log_2(x) = \log_2(e)\log_e(x)$. As a result MI is shown in Bits s^{-1}. Increasing the snr increases I_F with the logarithm of snr and increasing the dimensionality N of the signal vector scales I_F by N for any value of snr. Furthermore, since a Gaussian source distribution has been used, these results represent the Shannon capacity limits for this model.

Similarly, for the partially informed case, the MI in equation 3.104, for dimensions two to five has been calculated numerically and the results are presented in Figure 3.8 as a function of $10\log_{10}(\rho)$. Again, increasing the snr increases I_P with the logarithm of snr. However the slope of I_P is significantly less than the corresponding slope for I_F. This means that increasing the snr is more beneficial to the intended receiver when the eavesdropper is only able to observe amplitude values. We note that this is the case when both Bob and Eve observe the same snr. In practice Eve may be be able to reduce the difference $I_S = I_F - I_P$ for example by improving her channel or array gain and effectively operate at a higher snr. However this really only leads to a better estimate of signal amplitude levels. Increasing the dimensionality N of the signal vector scales I_P for any value of snr but the scaling relationship is more complicated in this case.

3.7 Summary

Expressions for fully and partially informed MI have been derived employing simplifying approximations to enable tractability. These expressions allow a comparison between the intended link MI and the MI available to an eavesdropper.

A relationship between the MI gradients for Bob and Eve has been investigated allowing a comparison of the rate of change of MI as the Gaussianity of the source distribution is varied or as the source estimation error changes.

The problem of determining the intercept MI, available to a receiving system which knows its channel matrix but has no prior knowledge of an orthogonal transformation that has been applied at the transmitter, has been analysed. Entropy derivations were performed giving some insight to the general multidimensional, high snr case. The exact MI for the N-D case has been obtained but requires numerical integration to derive the differential entropy for the partially informed case. The fully informed MI may be likened to the difference in entropy between two N-D probability spheres: the larger sphere, representing the distribution of the signal plus noise vector, and the smaller sphere, representing the distribution of the noise vector. At high snr, the partially informed MI was found to be equal to the difference in entropy between an N-D probability sphere, representing the distribution of the signal plus noise vector, and an N-D probability shell, representing the distribution of the amplitude plus noise vector.

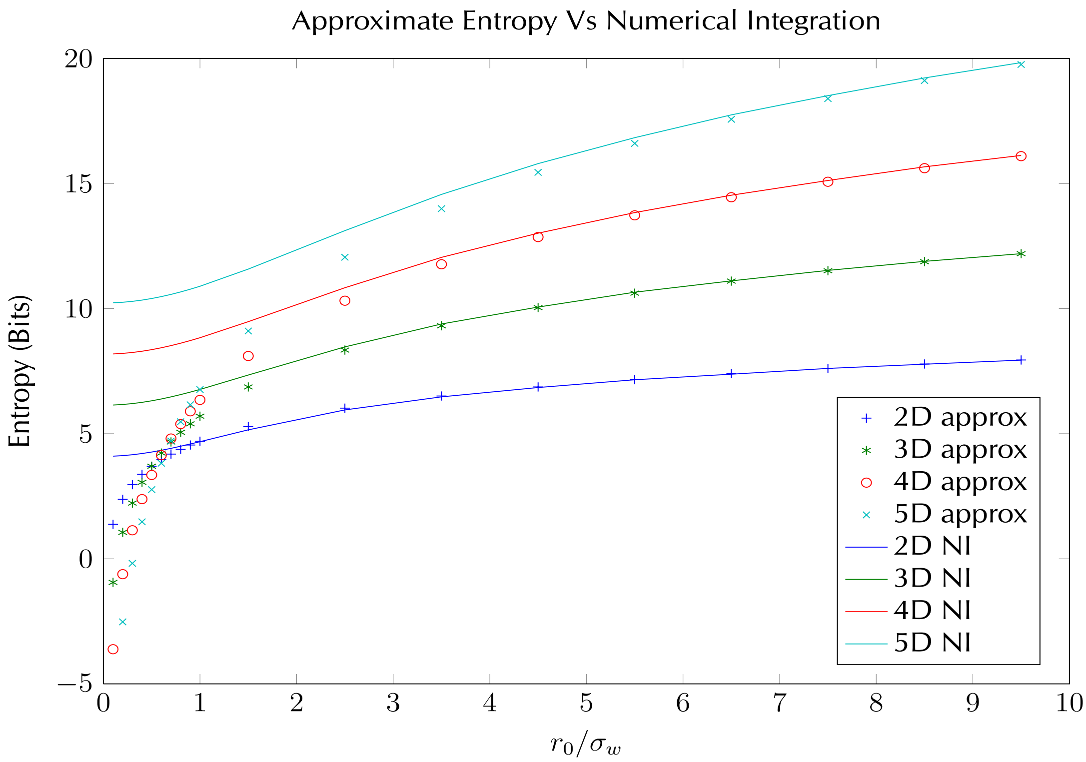

Figure 3.6: $h(\mathbf{y}|r_0)$ Vs SNR. Comparing high snr approximation with numerically integrated values.

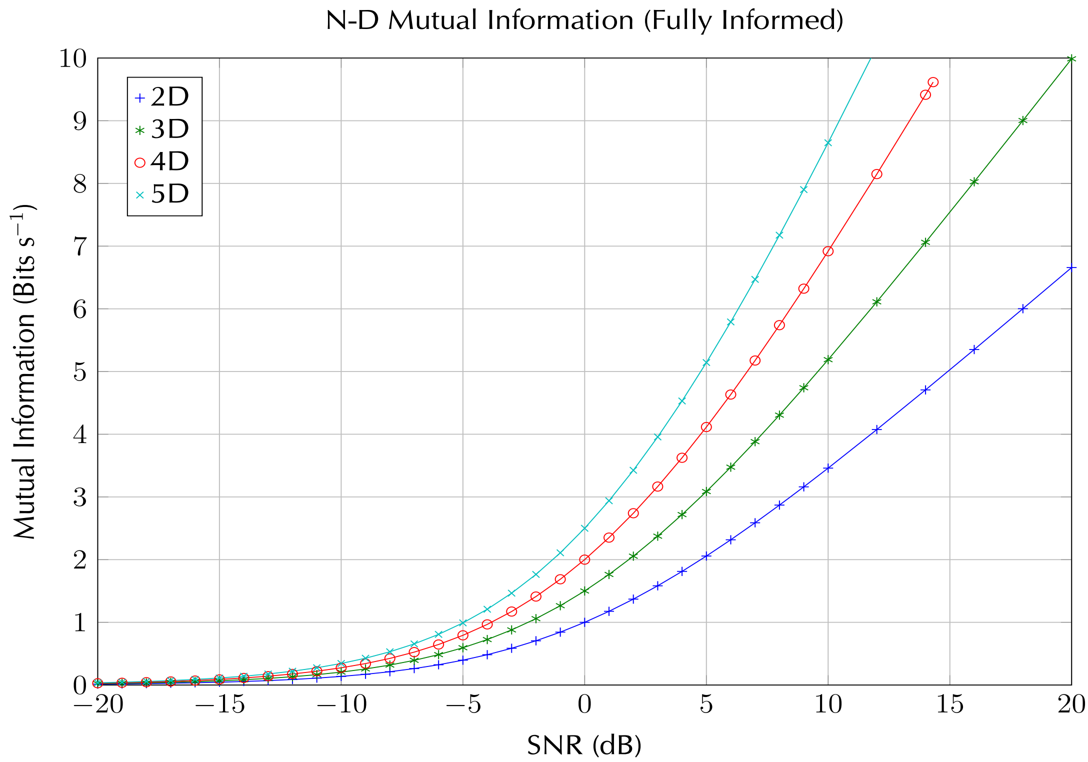

Figure 3.7: Mutual Information Vs SNR for fully informed case.

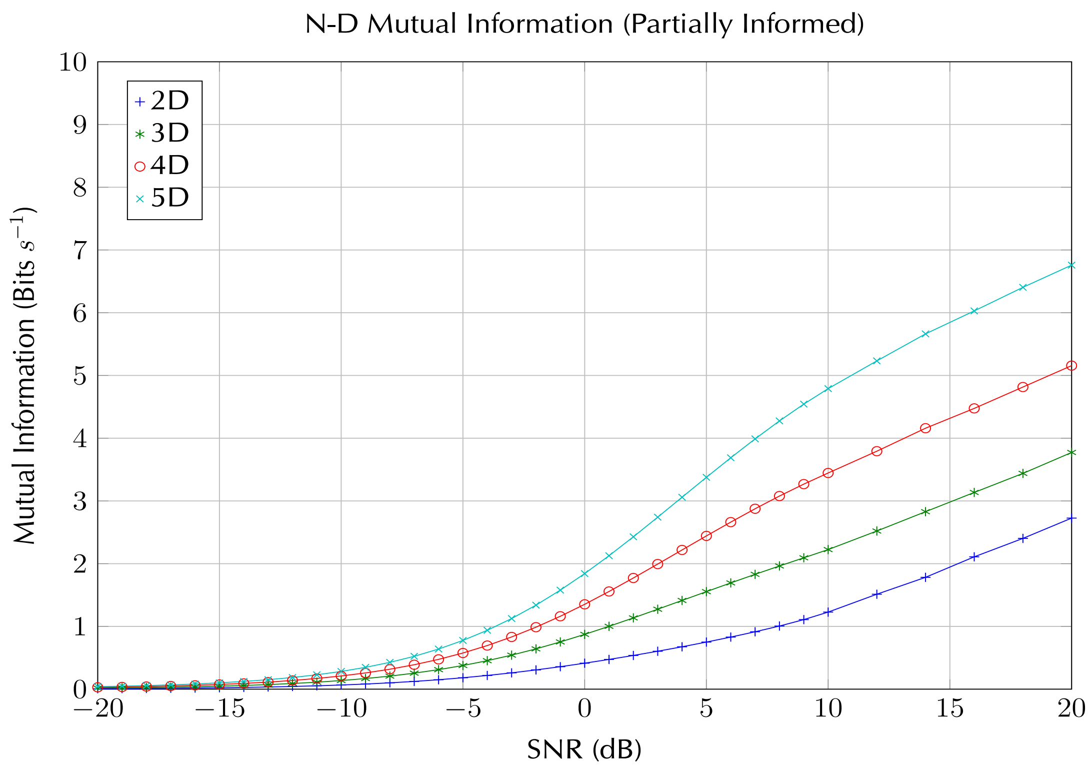

Figure 3.8: Mutual Information Vs SNR for amplitude informed case.

Chapter 4

Source and Channel Estimation

This chapter is concerned with determining the performance limits for MIMO channel and transmitter source estimation. These are the two variables of most interest to a MIMO eavesdropper. Clearly the primary objective, for the eavesdropper, is to obtain the source message and so it might seem that source estimation is the only variable of interest. However, in practice, both are required as the channel coefficients may vary due to a changing RF propagation environment and channel tracking becomes an important part of the source estimation procedure. We begin with MLE source estimation in section 4.1 and MLE channel estimation in section 4.2. Derivations of this kind may be found in the literature. e.g. expressions for channel estimation, assuming that the noise covariance matrix is an identity matrix, are provided by Larsson & Stoica in [57, Ch.9], Scharf [86, Ch.6] derives the parameter estimate and FIM for a real-valued multivariate linear model and Kay [48, Ch.15] provides MLE derivations for complex data. The derivations given here address the complex-valued linear normal model, are for completeness, and are used for subsequent comparisons with BSS performance results. An approach described by Villares [104] has been adapted to obtain some insight to the problem of jointly estimating the source and channel matrices using MLE techniques.

We utilise the block complex data model of equation 3.1 and described in Section 3.1. A channel estimator for this model may be derived from the likelihood function:

$$f(\mathbf{Y}|\mathbf{X},\mathbf{A}) = \frac{1}{|\pi\mathbf{\Sigma}_{\mathbf{w}}|^n} \exp\{-\mathrm{tr}\left([\mathbf{Y}-\mathbf{A}\mathbf{X}]^\dagger \mathbf{\Sigma}_{\mathbf{w}}^{-1}[\mathbf{Y}-\mathbf{A}\mathbf{X}]\right)\}, \tag{4.1}$$

where $\mathbf{\Sigma_w}$ is the covariance matrix for one column of $\mathbf{W}$. Let the score function, for estimating the complex valued $\mathbf{A}$, be defined as

$$s_{\mathbf{A}}(\mathbf{Y};\mathbf{A}) \triangleq \frac{\partial \ln f(\mathbf{Y};\mathbf{A})}{\partial \mathbf{A}^*}, \tag{4.2}$$

then the FIM may be obtained from one of two possible forms:

$$\mathbf{J_A} \triangleq \mathbb{E}_{\mathbf{Y};\mathbf{A}} \left\{ s_{\mathbf{A}}(\mathbf{Y};\mathbf{A}) s^{\dagger}_{\mathbf{A}}(\mathbf{Y};\mathbf{A}) \right\} \tag{4.3}$$

or

$$\mathbf{J_A} \triangleq \mathbb{E}_{\mathbf{Y};\mathbf{A}} \left\{ \frac{\partial s_{\mathbf{A}}(\mathbf{Y};\mathbf{A})}{\partial \mathbf{A}} \right\} = \mathbb{E}_{\mathbf{Y};\mathbf{A}} \left\{ \frac{\partial^2 \ln f(\mathbf{Y};\mathbf{A})}{\partial \mathbf{A} \partial \mathbf{A}^*} \right\} \tag{4.4}$$

and the CRB is given by the inverse of the FIM.

The Modified Cramér-Rao Bound (MCRB), described by Gini et al. in [37], may be used in cases where we are dealing with unknown nuisance parameters, such as the parameter $\mathbf{X}$ here, and is obtained from the modified FIM defined by Villares as [104]

$$\mathbf{J_A} \triangleq -\mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \left\{ \frac{\partial^2 f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X};\mathbf{A})}{\partial \mathbf{A} \partial \mathbf{A}^*} \right\}. \tag{4.5}$$

4.1 Source Estimation, Channel Known

When the channel $\mathbf{A}$ is already known and $\mathbf{X}$ is an unknown constant, the likelihood function for the observed $\mathbf{Y}$ is

$$f(\mathbf{Y}|\mathbf{A};\mathbf{X}) = \frac{1}{|\pi \mathbf{\Sigma_w}|^n} \exp\{-\mathrm{tr}\left([\mathbf{Y}-\mathbf{AX}]^{\dagger} \mathbf{\Sigma}_{\mathbf{w}}^{-1} [\mathbf{Y}-\mathbf{AX}]\right)\}. \tag{4.6}$$

If we define $\mathbf{T} \triangleq [\mathbf{Y}-\mathbf{AX}]^{\dagger} \mathbf{\Sigma}_{\mathbf{w}}^{-1} [\mathbf{Y}-\mathbf{AX}]$ then the log likelihood function $\mathcal{L}$ is

$$\mathcal{L}_{\mathbf{Y}|\mathbf{A};\mathbf{X}} = -n \ln(|\pi \mathbf{\Sigma_w}|) - \mathrm{tr}\left(\mathbf{T}\right). \tag{4.7}$$

We shall also use the definition, given by Lütkepohl [63] and Magnus and Neudecker [68],

$$D_{\mathbf{X}^*}(\mathbf{Z}) = \frac{\partial \mathbf{vec}\left(\mathbf{Z}\right)}{\partial \mathbf{vec}^T\left(\mathbf{X}^*\right)}, \tag{4.8}$$

which is the complex derivative of the complex matrix $\mathbf{Z}$ w.r.t. the complex matrix $\mathbf{X}^*$. The derivative w.r.t. the complex conjugate is necessary to obtain the correct Hessian matrix. Now, making use of the following matrix relationships, which can be found in [63]:

$$\begin{aligned}
\frac{\partial \text{tr}\left(\mathbf{A}\mathbf{X}^T\mathbf{B}\right)}{\partial \mathbf{X}} &= \mathbf{B}\mathbf{A}, &(4.9)\\
\frac{\partial \text{tr}\left(\mathbf{X}^T\mathbf{A}\right)}{\partial \mathbf{X}} &= \frac{\partial \text{tr}\left(\mathbf{A}\mathbf{X}^T\right)}{\partial \mathbf{X}} = \mathbf{A}, &(4.10)\\
\frac{\partial \text{tr}\left(\mathbf{A}\mathbf{X}\mathbf{B}\right)}{\partial \mathbf{X}} &= \mathbf{A}^T\mathbf{B}^T, &(4.11)\\
\mathbf{vec}\left(\mathbf{A}\mathbf{B}\mathbf{C}\right) &= \left(\mathbf{C}^T \otimes \mathbf{A}\right)\mathbf{vec}\left(\mathbf{B}\right), &(4.12)
\end{aligned}$$

we obtain

$$D_{\mathbf{X}^*}(\mathcal{L}_{\mathbf{Y}|\mathbf{A};\mathbf{X}}) = \mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{Y} - \mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}\mathbf{X}. \tag{4.13}$$

The MLE for the source is obtained when $D_{\mathbf{X}^*}(\mathcal{L}_{\mathbf{Y}|\mathbf{A};\mathbf{X}}) = 0$, resulting in:

$$\hat{\mathbf{X}}_{ML} = \left(\mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{Y}. \tag{4.14}$$

The FIM is given by

$$\begin{aligned}
\mathbf{J}_{\mathbf{X}|\mathbf{A}} = -D_{\mathbf{X}}D_{\mathbf{X}^*}(\mathcal{L}_{\mathbf{Y}|\mathbf{A};\mathbf{X}}) &= D_{\mathbf{X}}\left(\mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}\mathbf{X}\right)\\
&= D_{\mathbf{X}}\left(\mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}\mathbf{X}\mathbf{I}_n\right)\\
&= \mathbf{I}_n \otimes \mathbf{A}^\dagger\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{A}. \qquad (4.15)
\end{aligned}$$

Therefore the CRB for the source estimate is found to be:

$$\text{CRB}(\mathbf{X}|\mathbf{A}) = \mathbf{I}_n \otimes \mathbf{A}^{-1}\mathbf{\Sigma}_{\mathbf{w}}\mathbf{A}^{-\dagger}. \tag{4.16}$$

The modified CRB may be employed to derive an estimate of the source CRB when only the channel covariance is known. The modified CRB is described and derived by Villares in [104]. Thus, with $\mathbf{\Sigma}_{\mathbf{w}} = \sigma_w^2\mathbf{I}_m$,

$$\text{MCRB}(\mathbf{X}) = \mathbf{I}_n \otimes \sigma_w^2(\mathbb{E}\left\{\mathbf{A}^\dagger\mathbf{A}\right\})^{-1} \tag{4.17}$$

which becomes, with $\mathbf{\Sigma_A} \triangleq \mathbb{E}\left\{\mathbf{A}^\dagger\mathbf{A}\right\}$,

$$\text{MCRB}(\mathbf{X}) = \mathbf{I}_n \otimes \sigma_w^2\mathbf{\Sigma_A}^{-1}, \tag{4.18}$$

provided that $\mathbf{\Sigma_A}^{-1}$ is invertible. Since $\mathbf{\Sigma_A} = m\sigma_a^2\mathbf{I}_p$ in our model we obtain

$$\text{MCRB}(\mathbf{X}) = \frac{\sigma_w^2}{m\sigma_a^2}\mathbf{I}_{pn}. \tag{4.19}$$

4.2 Channel Estimation, Source known

If we are given the source symbols, the likelihood function for the observed $\mathbf{Y}$ is

$$f(\mathbf{Y}|\mathbf{X};\mathbf{A}) = \frac{1}{|\pi\mathbf{\Sigma_w}|^n}\exp\{-\text{tr}\left([\mathbf{Y}-\mathbf{AX}]^\dagger\mathbf{\Sigma_w}^{-1}[\mathbf{Y}-\mathbf{AX}]\right)\}. \tag{4.20}$$

The derivation in section 4.1 can be conveniently reused here to obtain the channel estimator and CRB by considering $\mathbf{Y}^T = \mathbf{X}^T\mathbf{A}^T + \mathbf{W}^T$. Let $\mathbf{Y}_1 = \mathbf{Y}^T$, $\mathbf{W}_1 = \mathbf{W}^T$, $\mathbf{B} = \mathbf{X}^T$ and $\mathbf{C} = \mathbf{A}^T$, then we have

$$f(\mathbf{Y}_1|\mathbf{B};\mathbf{C}) = \frac{1}{|\pi\mathbf{\Sigma}_{\mathbf{w}_1}|^m}\exp\{-\text{tr}\left([\mathbf{Y}_1-\mathbf{BC}]^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}[\mathbf{Y}_1-\mathbf{BC}]\right)\}. \tag{4.21}$$

Defining $\mathbf{T} \triangleq [\mathbf{Y}_1-\mathbf{BC}]^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}[\mathbf{Y}_1-\mathbf{BC}]$ then the log likelihood function is

$$\mathcal{L}_{\mathbf{Y}_1|\mathbf{B};\mathbf{C}} = -m\ln(|\pi\mathbf{\Sigma}_{\mathbf{w}_1}|) - \text{tr}\left(\mathbf{T}\right) \tag{4.22}$$

and we find that

$$D_{\mathbf{C}^*}(\mathcal{L}_{\mathbf{Y}_1|\mathbf{B};\mathbf{C}}) = \mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{Y}_1 - \mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{BC}, \tag{4.23}$$

which is zero when

$$\mathbf{C} = \left(\mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{Y}_1, \tag{4.24}$$

or when

$$\hat{\mathbf{A}}_{ML} = \mathbf{Y}\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{X}^\dagger\left(\mathbf{X}\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{X}^\dagger\right)^{-1}. \tag{4.25}$$

The FIM for $\mathbf{C}$ is

$$\begin{aligned} -D_{\mathbf{C}}D_{\mathbf{C}^*}(\mathcal{L}_{\mathbf{Y}_1|\mathbf{B};\mathbf{C}}) &= D_{\mathbf{C}}\left(\mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{B}\mathbf{C}\right) \\ &= D_{\mathbf{C}}\left(\mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{B}\mathbf{C}\mathbf{I}_m\right) \\ &= \mathbf{I}_m \otimes \mathbf{B}^\dagger\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{B}. \end{aligned} \tag{4.26}$$

This is the FIM for $\mathbf{A}^T$ but, since $\mathbf{A}$ is i.i.d., it is also the FIM for $\mathbf{A}$. So we may write

$$\mathbf{J}_{\mathbf{A}|\mathbf{X}} = \mathbf{I}_m \otimes \left(\mathbf{X}\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{X}^\dagger\right)^T, \tag{4.27}$$

and inverting $\mathbf{J}_{\mathbf{A}|\mathbf{X}}$ gives the CRB for the channel estimate:

$$\text{CRB}(\mathbf{A}|\mathbf{X}) = \mathbf{I}_m \otimes \left(\mathbf{X}\mathbf{\Sigma}_{\mathbf{w}_1}^{-1}\mathbf{X}^\dagger\right)^{-T}. \tag{4.28}$$

The MCRB may be employed to derive an estimate of the channel CRB when only the source covariance is known, Villares [104]. Thus, with $\mathbf{\Sigma}_{\mathbf{w}_1} = \sigma_w^2\mathbf{I}_n$,

$$\text{MCRB}(\mathbf{A}) = \mathbf{I}_m \otimes \sigma_w^2(\mathbb{E}\left\{\mathbf{X}^*\mathbf{X}^T\right\})^{-1}, \tag{4.29}$$

which becomes, with $\mathbf{\Sigma}_{\mathbf{X}} \triangleq \mathbb{E}\left\{\mathbf{X}^*\mathbf{X}^T\right\} = \mathbb{E}\left\{\mathbf{X}\mathbf{X}^\dagger\right\}$,

$$\text{MCRB}(\mathbf{A}) = \mathbf{I}_m \otimes \sigma_w^2\mathbf{\Sigma}_{\mathbf{X}}^{-1}. \tag{4.30}$$

Given $\mathbf{\Sigma}_{\mathbf{X}} = n\sigma_x^2\mathbf{I}_p$, then

$$\text{MCRB}(\mathbf{A}) = \frac{\sigma_w^2}{n\sigma_x^2}\mathbf{I}_{mp}. \tag{4.31}$$

For the model considered, the two modified bounds result in similar forms i.e. MCRB($\mathbf{X}$) is a diagonal matrix with entries $\frac{\sigma_w^2}{m\sigma_a^2} = \frac{1}{m\rho_a}$ and MCRB($\mathbf{A}$) is diagonal with entries $\frac{\sigma_w^2}{n\sigma_x^2} = \frac{1}{n\rho_x}$. As we might have expected, the bounds are inversely proportional to the channel-to-noise or signal-to-noise power ratios ρ_a and ρ_x respectively. Increasing the dimension m of the channel improves the source estimate and increasing the length n of the source matrix improves the channel estimate.

4.3 Channel and Source Unknown

It is difficult to derive a CRB for arbitrary source distributions in the noisy linear model and here we shall only consider the case where the source has a Gaussian distribution and follow the derivation of the unconditional CRB given by Villares in [104] to obtain the FIM for channel or source estimation. If we treat the channel matrix as fixed for an $m \times n$ block of observed data then we need only consider the likelihood function for a single observed $m \times 1$ vector $\mathbf{y}$ since we know that the CRB is additive in this case, and hence the total CRB will be given by $n \cdot$ CRB for n consecutive and independent observed vectors. The likelihood function for the observed $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ is

$$f(\mathbf{y}; \mathbf{\Theta}) = \frac{1}{|\pi \mathbf{\Sigma}_{\mathbf{y}}|} \exp\{-\mathbf{y}^{\dagger} \mathbf{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}\}, \tag{4.32}$$

where $\mathbf{\Sigma}_{\mathbf{y}} = \mathbf{A}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{A}^{\dagger} + \mathbf{\Sigma}_{\mathbf{w}}$, $\mathbf{\Sigma}_{\mathbf{x}}$ is the covariance matrix for the transmitted vector $\mathbf{x}$, $\mathbf{\Sigma}_{\mathbf{w}}$ is the covariance matrix for the noise vector $\mathbf{w}$ and where $\mathbf{\Theta}$ may be either $\mathbf{A}$ or $\mathbf{x}$. The log likelihood function, for the parameter $\mathbf{\Theta}$, is

$$\mathcal{L}_{\mathbf{y};\mathbf{\Theta}} = -\ln(|\pi \mathbf{\Sigma}_{\mathbf{y}}|) - \mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \mathbf{y}\mathbf{y}^{\dagger}\right). \tag{4.33}$$

Using the following relationships and since $\mathbf{y}^{\dagger}\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y} = \mathrm{tr}\left(\mathbf{y}^{\dagger}\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}\right) = \mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}\mathbf{y}^{\dagger}\right)$ (see also Appendix G):

$$\begin{aligned} \frac{\partial}{\partial \theta_k^*} \ln|\mathbf{\Sigma}_{\mathbf{y}}| &= \mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \theta_k^*}\right), \\ \frac{\partial}{\partial \theta_k^*} \mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}\mathbf{y}^{\dagger}\right) &= -\mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \theta_k^*} \mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}\mathbf{y}^{\dagger}\right), \end{aligned} \tag{4.34}$$

where θ_k^* is the k^{th} scalar component of $\mathbf{\Theta}^*$, then the score function for the parameter θ_k^* is

$$\frac{\partial \mathcal{L}_{\mathbf{y};\mathbf{\Theta}}}{\partial \theta_k^*} = \mathrm{tr}\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \theta_k^*} \mathbf{\Sigma}_{\mathbf{y}}^{-1} [\mathbf{y}\mathbf{y}^{\dagger} - \mathbf{\Sigma}_{\mathbf{y}}]\right). \tag{4.35}$$

Applying the trace relationship given by equation G.26 we get

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathbf{y};\mathbf{\Theta}}}{\partial \theta_k^*} &= \mathbf{vec}^T\left([\mathbf{y}\mathbf{y}^{\dagger} - \mathbf{\Sigma}_{\mathbf{y}}]^T\right)\left(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \otimes \mathbf{\Sigma}_{\mathbf{y}}^{-1}\right) \mathbf{vec}\left(\frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \theta_k^*}\right) \\ &= \left[\mathbf{vec}\left(\mathbf{y}^*\mathbf{y}^T\right) - \mathbf{vec}\left(\mathbf{\Sigma}_{\mathbf{y}}\right)\right]^T \left(\mathbf{\Sigma}_{\mathbf{y}}^{-1} \otimes \mathbf{\Sigma}_{\mathbf{y}}^{-1}\right) \mathbf{vec}\left(\frac{\partial \mathbf{\Sigma}_{\mathbf{y}}}{\partial \theta_k^*}\right), \end{aligned} \tag{4.36}$$

then by defining

$$\begin{aligned} D_{\theta_k^*}\left(\boldsymbol{\Sigma}_{\mathbf{y}}\right) &\triangleq \mathbf{vec}\left(\frac{\partial \boldsymbol{\Sigma}_{\mathbf{y}}}{\partial \theta_k^*}\right), \\ \hat{\mathbf{r}} &\triangleq \mathbf{vec}\left(\mathbf{y}^* \mathbf{y}^T\right), \\ \mathbf{r} &\triangleq \mathbf{vec}\left(\boldsymbol{\Sigma}_{\mathbf{y}}\right), \end{aligned} \tag{4.37}$$

we find that

$$\frac{\partial \mathcal{L}_{\mathbf{y};\boldsymbol{\Theta}}}{\partial \theta_k^*} = \left[(\hat{\mathbf{r}} - \mathbf{r}\right]^T \left[\boldsymbol{\Sigma}_{\mathbf{y}} \otimes \boldsymbol{\Sigma}_{\mathbf{y}}\right]^{-1} D_{\theta_k^*}\left(\boldsymbol{\Sigma}_{\mathbf{y}}\right). \tag{4.38}$$

Villares [104] defines $[D_{\mathbf{R}}(\boldsymbol{\Theta})]_p \triangleq \mathbf{vec}\left(\frac{\partial \mathbf{R}(\boldsymbol{\Theta})}{\partial \theta_p}\right)$ as the p^{th} column of $D_{\mathbf{R}}(\boldsymbol{\Theta})$ where θ_p is the p^{th} scalar component of $\boldsymbol{\Theta}$. This means that $D_{\mathbf{R}}(\boldsymbol{\Theta}) \equiv D_{\boldsymbol{\Theta}^*}(\mathbf{R})$ though Villares does not appear to have identified $D_{\mathbf{R}}(\boldsymbol{\Theta})$ as the derivative of $\mathbf{R}(\boldsymbol{\Theta})$ w.r.t. the matrix $\boldsymbol{\Theta}$. Now we may write

$$D_{\boldsymbol{\Theta}^*}\left(\mathcal{L}_{\mathbf{y};\boldsymbol{\Theta}}\right) = \left[\hat{\mathbf{r}} - \mathbf{r}\right]^T \left[\boldsymbol{\Sigma}_{\mathbf{y}} \otimes \boldsymbol{\Sigma}_{\mathbf{y}}\right]^{-1} D_{\boldsymbol{\Theta}^*}\left(\boldsymbol{\Sigma}_{\mathbf{y}}\right) \tag{4.39}$$

and the score function $\mathbf{s}_{\boldsymbol{\Theta}^*}$ for the matrix parameter $\boldsymbol{\Theta}^*$ is

$$\mathbf{s}_{\boldsymbol{\Theta}^*} = D_{\boldsymbol{\Theta}^*}^T\left(\boldsymbol{\Sigma}_{\mathbf{y}}\right) \left[\boldsymbol{\Sigma}_{\mathbf{y}} \otimes \boldsymbol{\Sigma}_{\mathbf{y}}\right]^{-1} \left[\hat{\mathbf{r}} - \mathbf{r}\right]. \tag{4.40}$$

Hence the FIM, defined as

$$\mathbf{J}_{\Theta} \triangleq \mathbb{E}_{\mathbf{y}}\left\{\mathbf{s}_{\boldsymbol{\Theta}^*} \mathbf{s}_{\boldsymbol{\Theta}^*}^{\dagger}\right\}, \tag{4.41}$$

becomes

$$\mathbf{J}_{\Theta} = D_{\Theta^*}^T(\boldsymbol{\Sigma}_{\mathbf{y}})[\boldsymbol{\Sigma}_{\mathbf{y}} \otimes \boldsymbol{\Sigma}_{\mathbf{y}}]^{-1} D_{\Theta^*}(\boldsymbol{\Sigma}_{\mathbf{y}}), \tag{4.42}$$

where we have used

$$\mathbb{E}_{\mathbf{y}}\left\{(\hat{\mathbf{r}} - \mathbf{r})(\hat{\mathbf{r}} - \mathbf{r})^{\dagger}\right\} = \boldsymbol{\Sigma}_{\mathbf{y}} \otimes \boldsymbol{\Sigma}_{\mathbf{y}}. \tag{4.43}$$

For channel estimation we require

$$\begin{aligned} D_{\mathbf{A}^*}(\boldsymbol{\Sigma}_{\mathbf{y}}) &= \frac{\partial \mathbf{vec}\left(\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{A}^{\dagger} + \boldsymbol{\Sigma}_{\mathbf{w}}\right)}{\partial \mathbf{vec}^T\left(\mathbf{A}^*\right)} \\ &= \mathbf{K}_{mm}\left[\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{x}} \otimes \mathbf{I}_m\right], \end{aligned} \tag{4.44}$$

where $\mathbf{K}$ is a $(pq \times pq)$ commutation matrix ($\mathbf{K}$ is also described in [63]) such that (s.t.) $\mathbf{K}_{pq}\mathbf{vec}\,(\mathbf{B}) = \mathbf{vec}\left(\mathbf{B}^T\right)$, for any $(p \times q)$ matrix $\mathbf{B}$. The CRB is then found from the inverse of $\mathbf{J_A}$ as

$$\mathbf{J_A} = \left[\mathbf{A}\boldsymbol{\Sigma}_\mathbf{x} \otimes \mathbf{I}_m\right]^T \mathbf{K}_{mm}^T \left[\boldsymbol{\Sigma}_\mathbf{y} \otimes \boldsymbol{\Sigma}_\mathbf{y}\right]^{-1} \mathbf{K}_{mm} \left[\mathbf{A}\boldsymbol{\Sigma}_\mathbf{x} \otimes \mathbf{I}_m\right]^* . \tag{4.45}$$

The covariance matrices for this model are: $\boldsymbol{\Sigma}_\mathbf{x} = \sigma_x^2\mathbf{I}_p$, $\boldsymbol{\Sigma}_\mathbf{y} = \sigma_y^2\mathbf{I}_m = (p\sigma_x^2\sigma_a^2 + \sigma_w^2)\mathbf{I}_m$ and $\boldsymbol{\Sigma}_\mathbf{A} = m\sigma_a^2\mathbf{I}_p$, so

$$\mathbf{J_A} = \frac{\sigma_x^4}{\sigma_y^4}\left[\mathbf{A}^T\mathbf{A}^* \otimes \mathbf{I}_m\right], \tag{4.46}$$

which clearly depends on a particular value for $\mathbf{A}$. Using the MCRB method we can then obtain $\text{MCRB}_\mathbf{A}$ as the average value for $\text{CRB}_\mathbf{A}$ over $\mathbf{A}$:

$$\text{MCRB}_\mathbf{A} = \frac{\sigma_y^4}{m\sigma_x^4\sigma_a^2}\mathbf{I}_{mp}. \tag{4.47}$$

For an $m \times n$ data block then

$$\text{MCRB}_\mathbf{A} = \frac{n\sigma_y^4}{m\sigma_x^4\sigma_a^2}\mathbf{I}_{mp}. \tag{4.48}$$

In a similar manner we may also derive a CRB for blind estimation of the sources. In this case

$$\begin{aligned} D_{\mathbf{x}^*}(\boldsymbol{\Sigma}_\mathbf{y}) &= \frac{\partial\mathbf{vec}\left(\mathbf{A}\mathbf{x}\mathbf{x}^\dagger\mathbf{A}^\dagger\right)}{\partial\mathbf{vec}^T\left(\mathbf{x}^*\right)} \\ &= \left[\mathbf{A}^* \otimes \mathbf{A}\mathbf{x}\right], \end{aligned} \tag{4.49}$$

so that the FIM for estimating $\mathbf{x}$ is

$$\mathbf{J_x} = \left[\mathbf{A}^* \otimes \mathbf{A}\mathbf{x}\right]^T \left[\boldsymbol{\Sigma}_\mathbf{y} \otimes \boldsymbol{\Sigma}_\mathbf{y}\right]^{-1} \left[\mathbf{A}^* \otimes \mathbf{A}\mathbf{x}\right]^* \tag{4.50}$$

and the resulting CRB, averaged over $\mathbf{A}$ and $\mathbf{x}$, is

$$\text{MCRB}_\mathbf{x} = \frac{n\sigma_y^4}{m^2p\sigma_a^4\sigma_x^2}\mathbf{I}_p. \tag{4.51}$$

When the noise power is small i.e. $\sigma_w^2 \approx 0$, then

$$\text{MCRB}_{\mathbf{A}} = \frac{np^2}{m}\sigma_a^2 \mathbf{I}_{mp} \tag{4.52}$$

$$\text{MCRB}_{\mathbf{x}} = \frac{np}{m^2}\sigma_x^2 \mathbf{I}_p. \tag{4.53}$$

These last two expressions highlight the fact that, when neither the source nor the channel are known, the uncertainty, or entropy, in the estimate is directly proportional to the variance in the parameter itself and employing a larger array dimension or observing longer data sequences will only increase the variance in the estimate.

4.4 Blind Source Separation

The problem of recovering signals that have been transformed through an unknown mixing process, more commonly known as BSS, arises in a broad range of signal processing applications. The term blind refers to the fact that no explicit knowledge of the source signals or the mixing system is available to an observer. Statistical methods for performing BSS, such as ICA, described by Comon in [25], have resulted in popular algorithms such as FASTICA developed by Hyvärinen et al. in [44, 16].

We have assumed that the signals from each source transmitter are complex-valued, statistically independent, and the observed data is a linear combination of the source waveforms with Additive White Gaussian Noise (AWGN). As stated previously this is a model that has been studied in the field of ICA and what we now require is a suitable algorithm for estimating both the complex-valued channel matrix and the complex-valued sources. Such an algorithm is described in [16], though we do note however that the FASTICA algorithm was intended for use with the standard linear model: $\mathbf{Y} = \mathbf{AX}$, not the noisy linear model: $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$.

The question arises as to the applicability of CRB analysis for the blind estimation problem described here. Our problem involves trying to simultaneously estimate two unknown subspaces: source matrix and mixing matrix. The estimation problem of $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$ is invariant to the transformations: $\mathbf{A} \mapsto \mathbf{AU}$, $\mathbf{X} \mapsto \mathbf{U}^{-1}\mathbf{X}$, $\mathbf{\Sigma}_x \mapsto \mathbf{U}^{-1}\mathbf{\Sigma}_x\mathbf{U}^{-\dagger}$, where $\mathbf{U}$ is a unitary matrix. The only invariant of $\mathbf{A} \mapsto \mathbf{AU}$ is the column span of $\mathbf{A}$ and the Hermitian structure of $\mathbf{\Sigma}_x \mapsto \mathbf{U}^{-1}\mathbf{\Sigma}_x\mathbf{U}^{-\dagger}$ is also invariant. Therefore only the column span of $\mathbf{A}$ and the covariance matrix of $\mathbf{Y}$ may be measured. Consider the QR decomposition $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is a unitary matrix and $\mathbf{R}$ is upper triangular then the invariant part of $\mathbf{A}$ is seen to be given by $\mathbf{R}$. The ambiguity in the product $\mathbf{AX}$ results in a singularity in the FIM and the CRB is therefore not defined.

Problems of this nature have been addressed by Smith in [90] and Xavier and Barroso in [112] and form a part of the rapidly evolving and increasingly popular area of information geometry e.g. Amari et al. [10, 11]. Information geometry assigns families of probability distributions to a differentiable manifold, where properties of the family are represented by geometric relations such as distance (Kullback-Leibler divergence) and curvature (Fisher information).

Ambiguities that result through the use of ICA techniques, discussed by Davies in [28] : scale, phase and permutation, do not necessarily represent a serious problem for discrete communication signal types. Permutation means that we may have to keep track of the individual sources and phase rotations for PSK or QAM signals can be estimated and corrected. Scaling issues are avoided by normalising the observed signal powers to unity. Restrictions usually applied in this method are: 1) at most one of the source signals has a Gaussian distribution 2) the mixing matrix $\mathbf{A}$ should be full rank. The first restriction does not present a problem for MIMO wireless communications as the pdfs of the digital modulation schemes that are employed are not Gaussian. The FASTICA algorithm employs a measure of kurtosis for its contrast function and this is known to be appropriate for digital modulation schemes. Therefore, after successful BSS processing, all that remains is to deduce the correct ordering of the separated sources and correct any phase rotation that might have occurred.

4.5 Derivation of Mixing Matrix CRB

The performance of the FASTICA algorithm, for real-valued signals and a real-valued mixing matrix, has been studied by Tichavský et al. in [99, 100]. Whereas Tichavský et al. [99] derived the CRB for the real-valued linear ICA model, our purpose here is to derive the CRB for linear ICA, with complex-valued signals, a complex-valued mixing matrix and for a general source distribution. In this section we make use of a result by Brandwood that simplifies the calculation of complex gradients. Brandwood proved the following theorem [18, Thm.1]:

Theorem 4.5.1 (Brandwood)**.** *Let $g : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ be a function of a complex number z and it's conjugate z^*, and let g be analytic w.r.t. each variable (z and z^*) independently. Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ be the function of the real variables x and y s.t. $g(z, z^*) = f(x, y)$, where $z = x + jy$. Then the partial derivative $\frac{\partial g}{\partial z}$, treating z^* as a constant in g, gives the same result as $\frac{1}{2}\left(\frac{\partial f}{\partial x} - j\frac{\partial f}{\partial y}\right)$. Similarly, $\frac{\partial g}{\partial z^*}$ is equivalent to $\frac{1}{2}\left(\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y}\right)$.*

If g is analytic on z^*, when considering z as a constant, then we say that g satisfies Brandwood's analyticity condition. Similarly, if g is analytic on z, when considering z^* as a constant, then we also say that g satisfies Brandwood's analyticity condition. This result also applies to vector and matrix expressions. Brandwood's Theorem allows us to directly calculate derivatives w.r.t. a complex argument, which may be simpler than calculating the gradients for the real-valued components that form the complex argument.

For the noiseless ICA model $\mathbf{Y} = \mathbf{AX}$ a lower bound for $\mathbf{\Sigma_A}$ may be obtained as the inverse of the FIM $F_{\mathbf{A}}$ of the complex-valued mixing matrix e.g. Carvalho et al. [29]:

$$F_{\mathbf{A}} = \mathbb{E}\left\{\left(\frac{\partial \ln p(\mathbf{Y}|\mathbf{A})}{\partial \mathbf{A}^*}\right)\left(\frac{\partial \ln p(\mathbf{Y}|\mathbf{A})}{\partial \mathbf{A}^*}\right)^{\dagger}\right\}, \tag{4.54}$$

where the complex derivative, defined by Brandwood in [18], is defined as $\frac{\partial}{\partial \mathbf{A}^*} \triangleq \frac{1}{2}\left[\frac{\partial}{\partial \mathbf{A}_r} + j\frac{\partial}{\partial \mathbf{A}_i}\right]$ and $\mathbf{A}_r$, $\mathbf{A}_i$ are, respectively, the real and imaginary parts of $\mathbf{A}$. Since $\mathbf{X}$ is i.i.d. $\mathbf{Y}$ is composed of n independent observations of a random vector with the same distribution. Because of this $F_{\mathbf{A}}$ is n times the FIM obtained from using a single column of $\mathbf{Y}$ and $\mathbf{X}$. The pdf for the column vector $\mathbf{y} = \mathbf{Ax}$ is

$$p_{\mathbf{y}}(\mathbf{y}) = |\det(\mathbf{AA}^*)|^{-1} p_{\mathbf{x}}(\mathbf{A}^{-1}\mathbf{y}), \tag{4.55}$$

where we have used the Jacobian for a complex linear transformation $J = |\det(\mathbf{AA}^*)|$, as proved by Mathai in [71]. The derivative of the log-likelihood, or score function is

$$\mathcal{L} = \frac{\partial \ln p_{\mathbf{x}}(\mathbf{A}^{-1}\mathbf{y})}{\partial \mathbf{A}^*} - \frac{\partial \ln|\det(\mathbf{AA}^*)|}{\partial \mathbf{A}^*}. \tag{4.56}$$

Letting $\mathbf{u} = (\mathbf{u}^r + j\mathbf{u}^i) = \mathbf{A}^{-1}\mathbf{y}$ and assuming that a function $f(\cdot, \cdot)$ exists s.t. $p_{\mathbf{x}}(\mathbf{u}^r, \mathbf{u}^i) = f(\mathbf{u}, \mathbf{u}^*)$ and satisfies the Brandwood analyticity condition, then we may write

$$\frac{\partial \ln f(\mathbf{u},\mathbf{u}^*)}{\partial A^*} = -\mathbf{A}^{-\dagger}\frac{\partial \ln f(\mathbf{u},\mathbf{u}^*)}{\partial \mathbf{u}^*}\mathbf{u}^\dagger = \mathbf{A}^{-\dagger}\phi(\mathbf{u})\mathbf{u}^\dagger, \tag{4.57}$$

where $\phi(\mathbf{u}) \triangleq -\frac{\partial \ln f(\mathbf{u},\mathbf{u}^*)}{\partial \mathbf{u}^*}$. The score function is therefore

$$\mathcal{L} = \mathbf{A}^{-\dagger}\phi(\mathbf{u})\mathbf{u}^\dagger - \mathbf{A}^{-\dagger} = \mathbf{A}^{-\dagger}[\phi(\mathbf{u})\mathbf{u}^\dagger - \mathbf{I}_m] = \mathbf{A}^{-\dagger}\mathcal{F}(\mathbf{u}), \tag{4.58}$$

since $\frac{\partial \ln|\det(\mathbf{A}\mathbf{A}^*)|}{\partial \mathbf{A}^*} = \mathbf{A}^{-\dagger}$ and defining $\mathcal{F}(\mathbf{u}) \triangleq \phi(\mathbf{u})\mathbf{u}^\dagger - \mathbf{I}_m$. To calculate the FIM for $\mathbf{A}$ we must first convert $\mathcal{L}$ to vector form. We vectorise $\mathcal{L}$ by using $\mathbf{vec}\,(\mathbf{AB}) = (\mathbf{I}_p \otimes \mathbf{A})\mathbf{vec}\,(\mathbf{B})$, where $\mathbf{A} : m \times n$ and $\mathbf{B} : n \times p$, this definition can be found in [63], so that

$$\mathbf{vec}\,(\mathcal{L}) = [\mathbf{I}_m \otimes \mathbf{A}^{-\dagger}]\mathbf{vec}\,(\mathcal{F}(\mathbf{u}))\,. \tag{4.59}$$

The FIM becomes

$$\mathbf{F}_\mathbf{A} = [\mathbf{I}_m \otimes \mathbf{A}^{-\dagger}]\mathbf{F}_\mathbf{I}[\mathbf{I}_m \otimes \mathbf{A}^{-1}], \tag{4.60}$$

where $\mathbf{F}_\mathbf{I} = \mathbb{E}\left\{\mathbf{vec}\,(\mathcal{F}(\mathbf{u}))\,\mathbf{vec}\,(\mathcal{F}(\mathbf{u}))^\dagger\right\}$ and the covariance matrix is lower bounded by $\mathbf{F}_\mathbf{A}^{-1}$ as

$$\mathbf{\Sigma}_\mathbf{a} \geqslant (\mathbf{I}_m \otimes \mathbf{A})\mathbf{F}_I^{-1}(\mathbf{I}_m \otimes \mathbf{A}^\dagger). \tag{4.61}$$

We find that elements of the matrix $\mathbf{F}_\mathbf{I}$ are given by

$$\begin{aligned}
[\mathbf{F}_\mathbf{I}]_{ij,kl} &= \mathbb{E}\left\{[\phi_i u_j^* - \delta_{ij}][\phi_k^* u_l - \delta_{kl}]\right\} \\
&= \delta_{ij}\delta_{kl} - \delta_{ij}\mathbb{E}\left\{\phi_k^* u_l\right\} - \delta_{kl}\mathbb{E}\left\{\phi_i u_j^*\right\} + \mathbb{E}\left\{\phi_i u_j^* \phi_k^* u_l\right\}.
\end{aligned} \tag{4.62}$$

The complex score function may be written [5, 18]

$$\phi(\mathbf{u}) \triangleq -\frac{\partial \ln f(\mathbf{u},\mathbf{u}^*)}{\partial \mathbf{u}} = -\frac{1}{2}\left[\frac{\partial \ln p_\mathbf{x}(\mathbf{u}^r,\mathbf{u}^i)}{\partial \mathbf{u}^r} + j\frac{\partial \ln p_\mathbf{x}(\mathbf{u}^r,\mathbf{u}^i)}{\partial \mathbf{u}^i}\right], \tag{4.63}$$

where $\mathbf{u}^r$ is the real part of $\mathbf{u}$ and $\mathbf{u}^i$ is the imaginary part of $\mathbf{u}$. Since we treat the real and imaginary parts of the sources as independent then $p_\mathbf{x}(\mathbf{u}^r,\mathbf{u}^i) = p_\mathbf{x}(\mathbf{u}^r)p_\mathbf{x}(\mathbf{u}^i)$ and so

$$\phi(\mathbf{u}) = -\frac{1}{2}\left[\frac{\partial \ln p_\mathbf{x}(\mathbf{u}^r)}{\partial \mathbf{u}^r} + j\frac{\partial \ln p_\mathbf{x}(\mathbf{u}^i)}{\partial \mathbf{u}^i}\right] = -\frac{1}{2}\left[\phi^r + j\phi^i\right], \tag{4.64}$$

where ϕ^r is the real part of $\phi(\mathbf{u})$ and ϕ^i is the imaginary part of $\phi(\mathbf{u})$. Thus the derivations for the complex FIM can be performed using the real source distribution results from [99, 114] in the real and imaginary parts of $\phi(\mathbf{u}) = -\frac{1}{2}\left[\phi^r + j\phi^i\right]$

and $\mathbf{u} = [\mathbf{u}^r + j\mathbf{u}^i]$. The following assumptions and definitions are used:

1. $\mathbf{A} \in \mathbb{C}^{m \times m}$ is nonsingular.
2. The source signals x_i are mutually independent and identically distributed.
3. $\mathbb{E}\{x_i^r\} = \mathbb{E}\{x_i^i\} = 0$. The real and imaginary components of $\mathbf{x}$ are i.i.d. with zero mean.
4. $\mathbb{E}\{u_i^r\} = \mathbb{E}\{u_i^i\} = 0$. The real and imaginary components of $\mathbf{u}$ are i.i.d. with zero mean, when the previous three conditions are satisfied.
5. $\mathbb{E}\{\phi_i^r\} = \mathbb{E}\{\phi_i^i\} = 0$, when $\mathbf{u}$ has a zero mean and a symmetric pdf.
6. $\mathbb{E}\{(u_i^r)^2\} = \mathbb{E}\{(u_i^i)^2\} = 1$.
7. $\kappa \triangleq \mathbb{E}\{(\phi_i^r)^2\} = \mathbb{E}\{(\phi_i^i)^2\}$.
8. $\eta \triangleq \mathbb{E}\{(\phi_i^r u_i^r)^2\} = \mathbb{E}\{(\phi_i^i u_i^i)^2\}$.
9. $\mathbb{E}\{\phi_i^r u_i^r\} = \mathbb{E}\{\phi_i^i u_i^i\} = \delta_{ij}$. See Appendix G.
10. $\mathbb{E}\left\{\phi_i^r u_j^i\right\} = \mathbb{E}\left\{\phi_i^i u_j^r\right\} = 0,\ i \neq j$.

We can now derive the terms in $[\mathbf{F_I}]_{ij,kl}$ as follows:

$$\begin{aligned}
\mathbb{E}\{\phi_k^* u_l\} &= \mathbb{E}\left\{\frac{1}{2}[\phi_k^r - j\phi_k^i][u_l^r + ju_l^i]\right\} = \frac{1}{2}\left[\delta_{kl} + \delta_{kl} - j\delta_{kl} + j\delta_{kl}\right] = \delta_{kl}, \\
\mathbb{E}\left\{\phi_i u_j^*\right\} &= \mathbb{E}\left\{\frac{1}{2}[\phi_i^r + j\phi_i^i][u_j^r - ju_j^i]\right\} = \frac{1}{2}\left[\delta_{ij} + \delta_{ij} - j\delta_{ij} + j\delta_{ij}\right] = \delta_{ij}.
\end{aligned} \tag{4.65}$$

$\mathbb{E}\left\{\phi_i u_j^* \phi_k^* u_l\right\}$ is non-zero when:

1. $i = j = k = l$, or $\mathbb{E}\{\phi_i u_i^* \phi_i^* u_i\} = \frac{1}{2}[\eta + \kappa]\delta_{ijkl}$,
2. $i = l,\ j = k,\ i \neq j$, or $\mathbb{E}\left\{\phi_i u_i \phi_j^* u_j^*\right\} = \delta_{il}\delta_{jk} - \delta_{ijkl}$,
3. $i = j,\ k = l,\ i \neq k$, or $\mathbb{E}\{\phi_i u_i^* \phi_k^* u_k\} = \delta_{ij}\delta_{kl} - \delta_{ijkl}$,
4. $i = k,\ j = l,\ i \neq j$, or $\mathbb{E}\left\{\phi_i \phi_i^* u_j u_j^*\right\} = \kappa[\delta_{ik}\delta_{jl} - \delta_{ijkl}]$.

Hence the general form for the entries in $\mathbf{F_I}$ is given by

$$[\mathbf{F_I}]_{ij,kl} = \delta_{il}\delta_{jk} + \frac{1}{2}[\eta - \kappa - 4]\delta_{ijkl} + \kappa\delta_{ik}\delta_{jl}, \tag{4.66}$$

where δ_{ijkl} is defined as

$$\delta_{ijkl} \triangleq \begin{cases} 1 & \text{if } i = j = k = l, \\ 0 & \text{otherwise.} \end{cases} \tag{4.67}$$

We may rewrite this to obtain the mn^{th} element of $\mathbf{F_I}$ as

$$[\mathbf{F_I}]_{m,n} = \delta_{il}\delta_{jk} + \left[\frac{1}{2}(\eta - \kappa) - 2\right]\delta_{ijkl} + \kappa\delta_{ik}\delta_{jl}, \tag{4.68}$$

where $m = (i-1)d + j$ and $n = (k-1)d + l$.

In the real case, Tichavský et al. [99] found that

$$[\mathbf{F_I}]_{m,n} = \delta_{il}\delta_{jk} + [\eta - \kappa - 2]\,\delta_{ijkl} + \kappa\delta_{ik}\delta_{jl}, \tag{4.69}$$

where m and n are as defined here. The difference between the real and complex $\mathbf{F_I}$ is therefore $\frac{1}{2}(\eta - \kappa)\delta_{ijkl}$. In Appendix E we provide a proof showing that $\eta = \alpha + 1$ for the Generalised Gaussian distribution.

To obtain $\mathbf{\Sigma_a}$, equation 4.61, it is necessary to first invert $\mathbf{F_I}$; the derivation of $\mathbf{F_I}^{-1}$ is provided in Appendix D. As discussed by Tichavský et al. in [99], $\mathbf{F_I}^{-1}$ can be used to obtain the CRB for source estimation since $\mathbf{F_I}$ represents the gain matrix $\mathbf{G}$, which is independent of the mixing matrix i.e. the CRB for $\mathbf{G}$ is found as

$$\mathbf{\Sigma_g} = \left[\mathbf{F_I}^{-1}\right]_{mm}, \tag{4.70}$$

where $m = (i-1)d + j$ and $i \neq j$ and, in Appendix D, we find that

$$[\mathbf{F_I}^{-1}]_{mm} = \frac{\kappa}{\kappa^2 - 1}, \tag{4.71}$$

which is the same as for the real-valued case derived in [99].

In the simulation studies that follow we require the mean value of $\mathbf{\Sigma_a}$ taken over a number of repetitions, where $\mathbf{A}$ is randomly generated at each simulation instance. With the components $a_{i,j} \sim \mathcal{CN}(0, 2\sigma_a^2)$, the diagonal elements of

$\mathbb{E}\left\{\boldsymbol{\Sigma}_{\mathbf{a}}\right\}$ are

$$[\mathbb{E}\left\{\boldsymbol{\Sigma}_{\mathbf{a}}\right\}]_{ii} = 2\sigma_a^2 \text{tr}\left(\mathbf{F}_{\mathbf{I}}^{-1}\right). \tag{4.72}$$

Since we have assumed a noiseless model, this $\boldsymbol{\Sigma}_{\mathbf{a}}$ is only valid, in practice, for a high snr. The resulting $\mathbf{F}_{\mathbf{A}}^{-1}$ has some non-zero, off-diagonal elements and the diagonal elements are not all the same. A useful simplification may be achieved by assuming that $\boldsymbol{\Sigma}_{\mathbf{a}} = 2\sigma_a^2\mathbf{I}_{mm}$; substituting $2\sigma_a^2$ as the mean of the diagonal elements of $\mathbf{F}_{\mathbf{A}}^{-1}$ i.e. $2\sigma_a^2 \approx \frac{1}{m}\text{tr}\left(\mathbf{F}_{\mathbf{A}}^{-1}\right)$. We have calculated $\boldsymbol{\Sigma}_{\mathbf{a}}$ for the GG distribution since this gives us a means to continuously vary the source Gaussianity. This result will provide a useful comparison for the digital source distributions especially since the parameter α may be converted to a kurtosis value as shown in Appendix C.

4.6 Simulation Results

The FASTICA algorithm, described and developed by Koldovský and Tichavský in [55] and available as Matlab code, was employed to perform blind source separation and obtain estimates of the source and mixing matrices. Pseudocode for the FASTICA algorithm is listed in Appendix M. As can be seen the algorithm first performs a whitening of the observation data, which is common to many BSS methods. The prewhitening has the effect of reducing the mixing matrix search space, for contrast function optimization, to a search for an optimal unitary matrix. The core of the FASTICA algorithm is the fixed-point ICA stage, where a unitary matrix is found that optimizes a contrast function. This amounts to maximizing the resultant kurtosis in the separated source estimates. Finally the algorithm returns the source and channel estimates, taking account of the initial whitening that was applied. However the algorithm returns a mixing matrix estimate which has an unknown scale and permutation. To compare the mixing matrix estimates with the original matrix we must first determine what the permutation is and adjust the mixing matrix accordingly. We use the optimal pairing technique described by Tichavský and Koldovský in [98] which finds the nearest matrix (in the Frobenius norm sense) to the original matrix with the same rows (up to the signs and order). Rescaling occurs in FASTICA when the observed data $\mathbf{Y}$ matrix is whitened as $\mathbf{Y} = \mathbf{QY}$, where $\mathbf{Q}$ is obtained from the diagonal matrix of eigenvalues $\mathbf{D}$ and the matrix of eigenvectors $\mathbf{V}$ of the covariance matrix of $\mathbf{Y}$ as $\mathbf{Q} = [\mathbf{D}^{-1}]^{\frac{1}{2}}\mathbf{V}^{\dagger}$. Thus to correctly estimate the resulting Mean Squared Error (MSE) results we need to scale by tr $(|\mathbf{Q}|^2)$. Hence the results represent the best that could be achieved using the FASTICA algorithm. As described previously, HOS approaches such as FASTICA are constrained to BSS cases where at most one of the sources has a Gaussian distribution i.e. zero kurtosis. However they would appear to be entirely suitable for BSS of digital communications waveforms which have significant kurtosis values. Our simulations and analysis of results were implemented using the free Matlab alternative: GNU Octave, for numerical computations.

4.6.1 BSS of Discrete sources

Simulations were performed by generating a complex random pulse stream, one for each of four source transmitters, with Quadrature Phase Shift Keyed (QPSK) modulation and square root, raised-cosine pulse shaping. The pulse streams were linearly transformed by a complex-valued random mixing matrix (zero-mean, unit variance; $a_{i,j} \sim \mathcal{CN}(0,1)$) and AWGN noise, for a range of snrs, added to the resulting matrix. Both the phase rotation and the separated signal ordering were corrected by correlating each original source signal with all of the separated signals. The highest correlation magnitude indicates which separated signal is the best estimate for each source and the phase is simply found as the mean phase value of the complex cross correlation.

Figure 4.1 shows the input symbols for each QPSK source, prior to mixing and without noise. This may be compared with the FASTICA output (for a random channel and 10decibel (dB) snr), shown in Figure 4.2, after the phase rotation has been corrected. It is interesting to note that each output constellation appears to have a different snr which is a result of the way in which the FASTICA algorithm operates, e.g. see Bingham and Hyvärinen [16].

Figure 4.3 shows the simulation results for estimating the Symbol Error Rate (SER), where each symbol stream consisted of 1000 symbols. In this case the closest symbols were found for each separated symbol stream and the number of errors counted. The symbol error rate is shown compared with the modified CRB for separation error MCRB$_{\mathbf{x}}$, equation 4.51, which has been scaled by $\sqrt{1000}$ to account for the number of symbols used.

Figure 4.4 shows the MSE between the source waveforms and the BSS estimated waveforms. This is compared with the modified CRB for source separation error $\text{MCRB}_{\mathbf{x}}$, equation 4.51, derived in section 4.3. The simulation results appear to match quite well with the theoretical values even though the theory assumed a Gaussian distribution for the sources and the sources generated here are not Gaussian distributed.

Figure 4.5 compares the channel estimation error power, obtained from the simulations and is compared with the modified CRB for channel estimation $\text{MCRB}_{\mathbf{a}}$, equation 4.47 , derived earlier in section 4.3. The two plots are similar in trend with the FASTICA simulation results being greater than the theoretical values. This is, to some extent, due to the Gaussian assumption for the source distribution used in the theoretical derivations.

The theoretical CRB plots shown in figures 4.4 and 4.5 are the same because we used $\sigma_a^2 = \frac{1}{2}$ and $\sigma_x^2 = \frac{1}{2}$ leading to the same values for $\text{MCRB}_{\mathbf{x}}$ and $\text{MCRB}_{\mathbf{a}}$ as a function of snr.

4.6.2 BSS of Generalised Gaussian Sources

The Generalised Gaussian distribution (Appendix C) was used to generate random instances of the source matrix $\mathbf{X}$. The Gaussianity or Kullback-Liebler (KL) divergence, of this distribution can be controlled through the parameter α. The real and imaginary parts of $\mathbf{X}$ were generated independently with the same value for α. For each parameter set: {Gaussianity (α), source block length (n), number of sources (2)}, up to 1000 repetitions were performed, the MSE in the mixing matrix estimates were calculated at each iteration and the mean of those results taken. For each repetition a different complex-valued mixing matrix, with elements $a_{i,j} \sim \mathcal{CN}(0, 1)$, was randomly generated. In each of the figures that follow, the theoretical CRB has a large peak at $\alpha = 2$, highlighting the fact that Gaussian sources cannot be blindly separated using a HOS based technique such as FASTICA. The FIM provides a measure of the information that the random vector $\mathbf{y}$ carries about the unknown $\mathbf{A}$ so the theoretical CRB plots indicate that the Fisher Information is zero when the sources have a Gaussian distribution.

Figure 4.6 shows the results for MSE in the mixing matrix and the theoretical values given by equation 4.72, with $\mathbf{F_I}$ calculated for the GG distribution, as α is varied, for block length $n = 100$. This is a short block length and the FASTICA algorithm does not perform well, returning an estimation error variance that does not achieve the theoretical CRB.

When the block length is increased to 1000, the FASTICA estimation error improves for values of α that are far from Gaussian i.e. $|\alpha - 2| > 0.5$, as shown in Figure 4.7. As n increases the simulation results provide a increasingly better fit with the theoretical CRB, as can be expected in Figures 4.8 and 4.9.

Figures 4.6 to 4.9 highlight the fact that HOS based BSS is not possible when the sources have a Gaussian distribution. This suggests the possibility of constructing a Gaussian based source signal, with hidden structure, that would deny separability to an eavesdropper whilst allowing the intended receiver to still be able to perform source separation.

4.7 Summary

In this chapter we derived theoretical variance bounds for a MIMO link represented mathematically as a linear block complex data model. Using a ML approach and assuming a Gaussian distributed source, a CRB for source estimation, given knowledge of the channel was derived. Similarly a ML CRB for channel estimation, given knowledge of the source was derived. In the case where both the source and channel are unknown, modified CRB estimates were derived.

Simulations were developed and performed with a non Gaussian distributed source to compare the performance of a popular BSS algorithm with the modified CRB bounds for source and channel estimation. The simulation results show that the separation performance and channel estimation performance (scale, phase and permutation accounted for) can be usefully compared with these bounds.

We also derived analytic CRB expressions for the noiseless complex linear ICA model and a general source distribution. The CRB for source estimation was found to be the same as for the real-valued case and the CRB for estimation of the complex-valued mixing matrix was found to be similar to its real-valued counterpart. Simulations produced results that indicate good agreement between the performance of the FASTICA algorithm and the theoretical CRB for complex-valued mixing matrix estimation. The theoretical variance bounds developed in this chapter may be applied in entropy calculations leading to MI estimates such as those derived in Chapter 3.

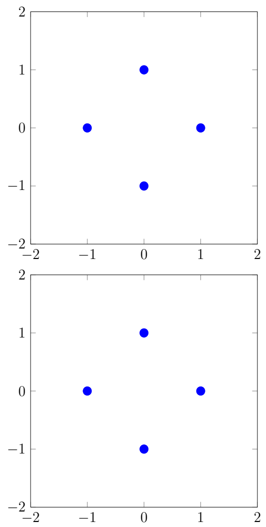

Figure 4.1: QPSK Input Symbols, no mixing or additive noise.

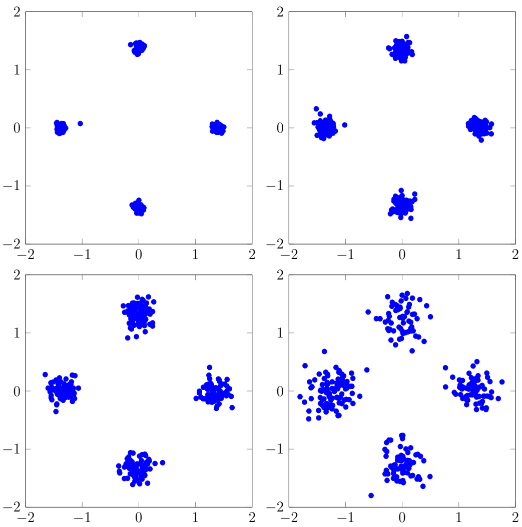

Figure 4.2: Estimated Output Symbols after BSS and phase correction.

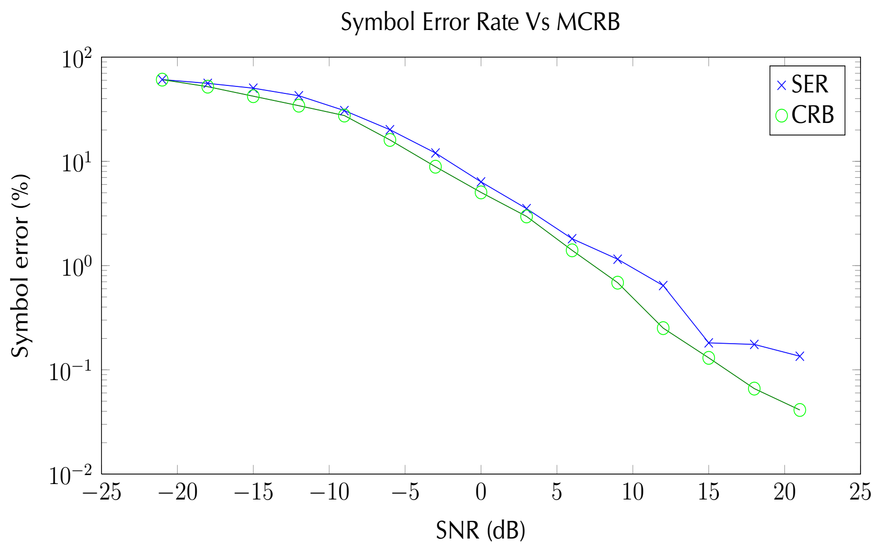

Figure 4.3: Symbol Error Rate Vs MCRB for source separation error.

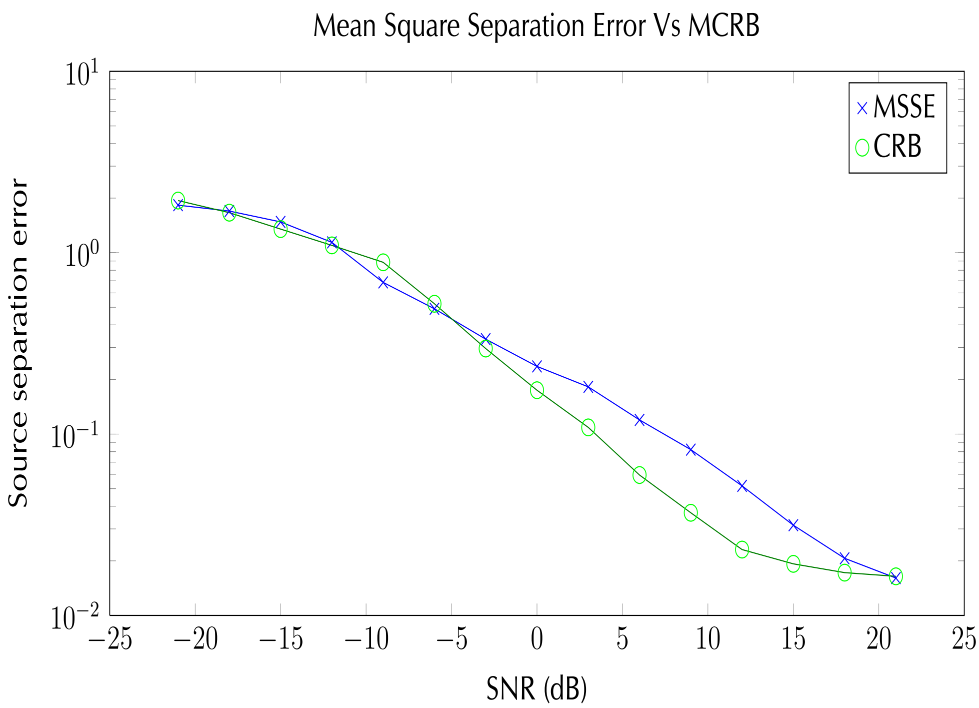

Figure 4.4: Mean Square Separation Error Vs MCRB for source separation error.

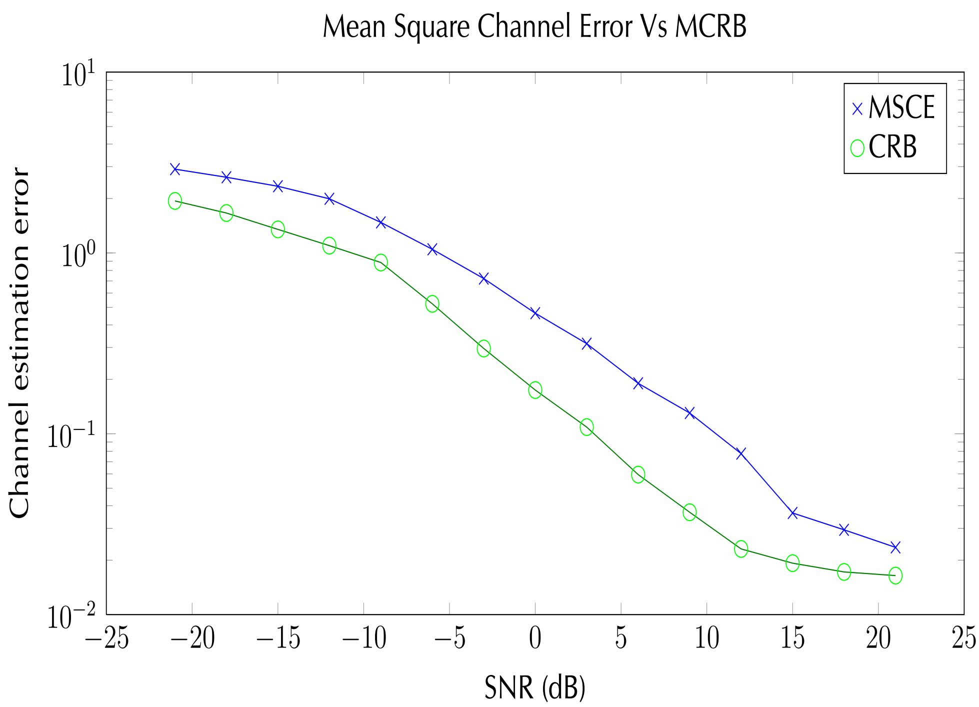

Figure 4.5: Mean Square Channel Error Vs MCRB for channel estimation.

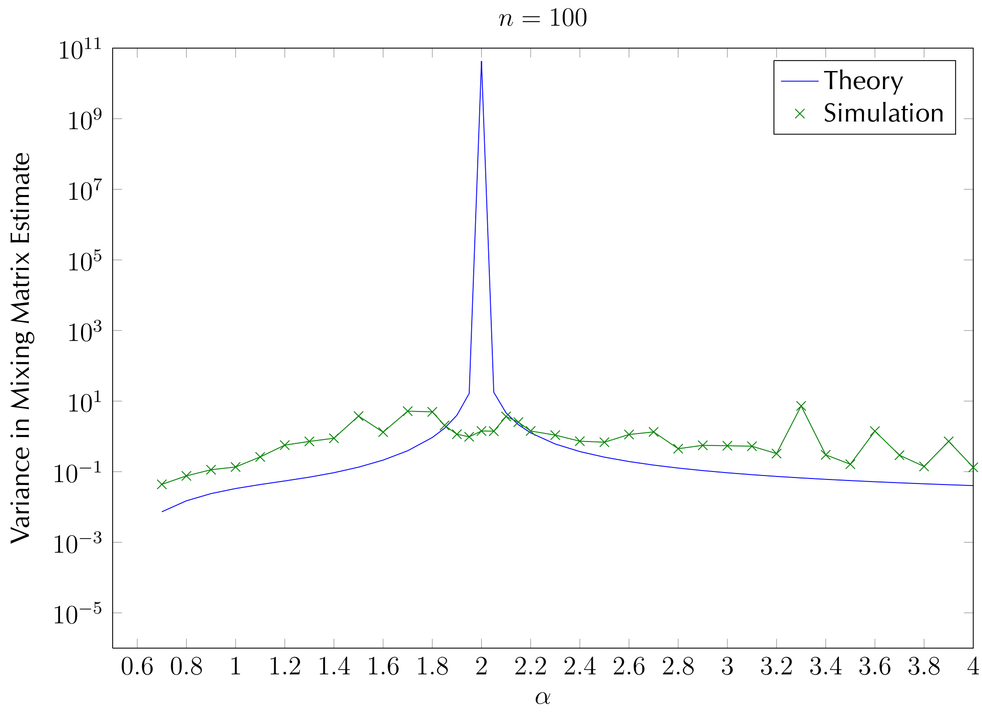

Figure 4.6: Mean squared error in mixing matrix estimate $\hat{A}$, as a function of α and data block length $n = 100$.

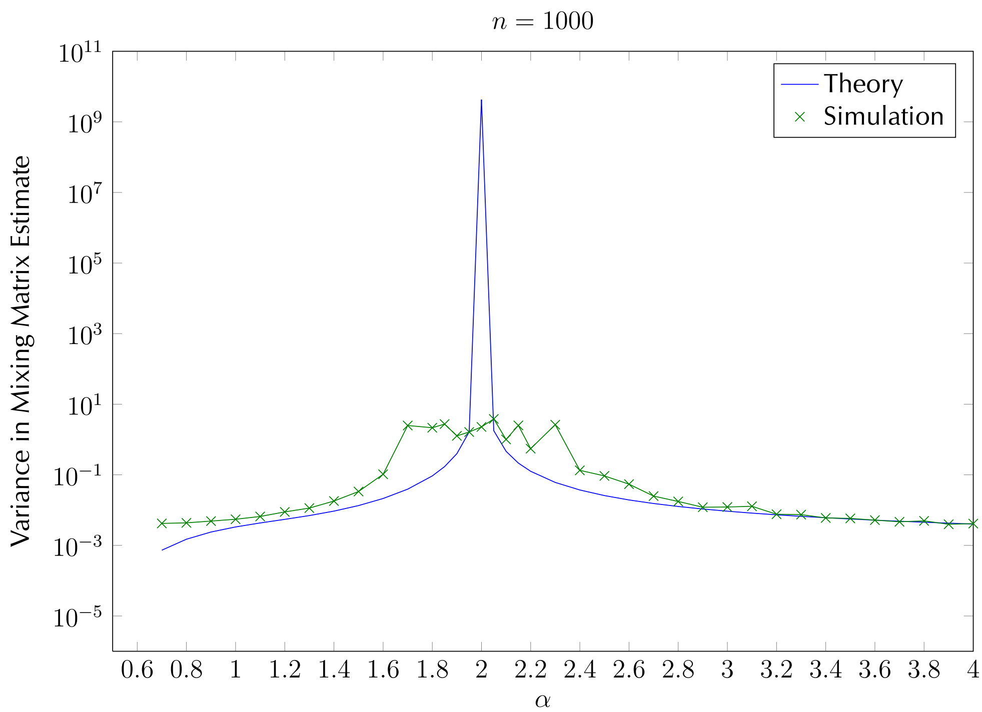

Figure 4.7: Mean squared error in mixing matrix estimate $\hat{A}$, as a function of α and data block length $n = 1000$.

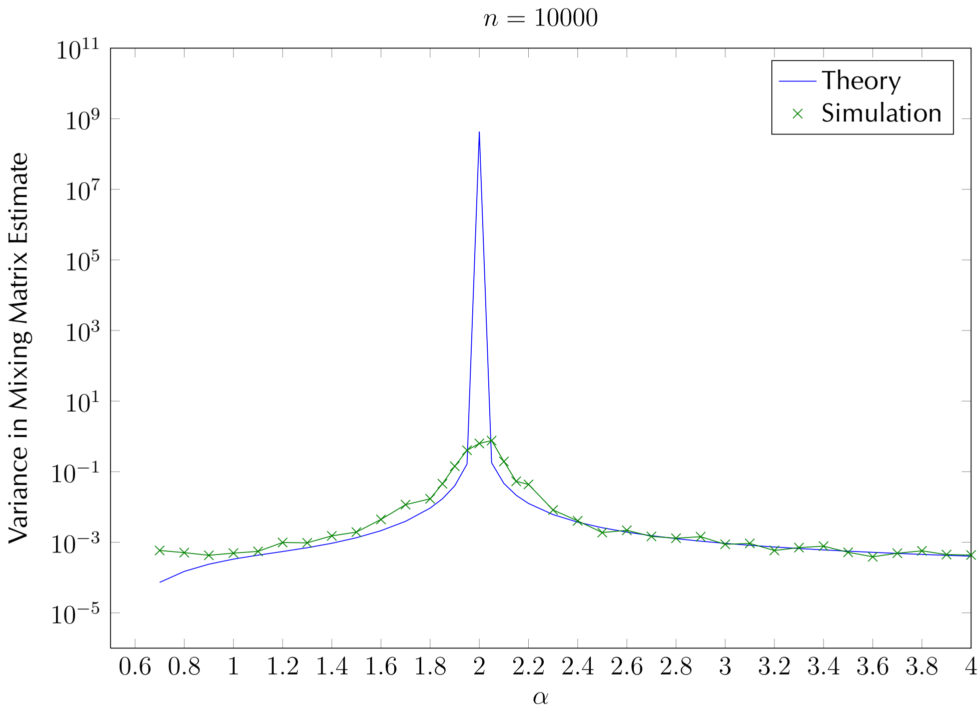

Figure 4.8: Mean squared error in mixing matrix estimate $\hat{A}$, as a function of α and data block length $n = 10000$.

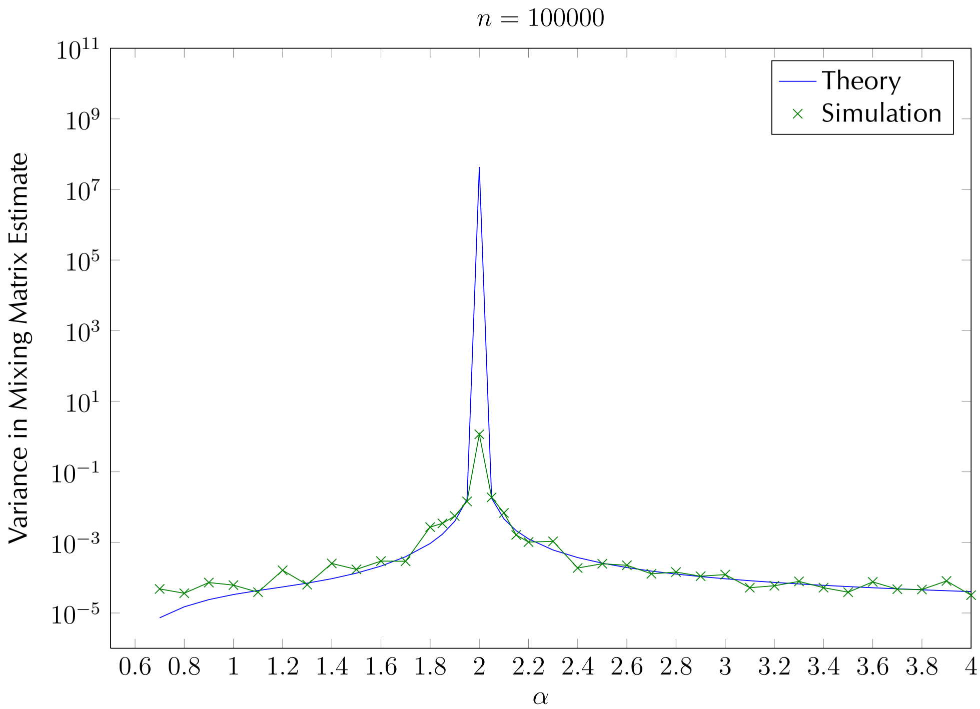

Figure 4.9: Mean squared error in mixing matrix estimate $\hat{A}$, as a function of α and data block length $n = 100000$.

Chapter 5

Copula Techniques for Modelling Channel Dependence

5.1 Introduction

The theory of copulas was originally developed as a means of incorporating dependence between random variables in the field of finance and there are many useful introductory texts on the subject; for example see Nelsen [76]. Two MIMO wireless communications challenges are suggested in this chapter as potential candidates for the application of copula techniques: modelling signal correlation and propagation effects, separation of mixed sources. The ability to model correlation and propagation effects in the MIMO wireless scenario will facilitate an analysis and understanding of these effects so that BSS approaches might be developed to overcome them.

The Rayleigh distribution has been a long-term standard for modelling RF propagation fading effects in wireless communications scenarios, however, thanks to its wide versatility and analytic tractability, the Nakagami–m distribution has recently become popular for modelling fading effects, e.g. see Alouini et al. [9] and Beaulieu and Cheng [12]. Modelling data correlation in the MIMO wireless propagation scenario is a complicated and computationally demanding problem and so a simple, intuitive approach is desirable. In this chapter we develop a complex Nakagami distribution for inclusion in a correlated fading channel model, which is implemented using the copula method.

There are several HOS based algorithms that have been developed for the purpose of separating mixtures of independent sources and for many of these algorithms to be successful, the original independent sources must have non-Gaussian marginal probability densities. One of these algorithms, the FASTICA algorithm, has already been discussed in Chapter 4 where it was found to perform well for the model and assumptions used. Copula methods also appear to offer an alternative approach to BSS that does not depend on the source distributions but which instead exploits the structure of the dependence between the sources. This suggests that some of the limitations in ICA techniques may be overcome through the use of copulas. For example a copula based approach may be able to separate mixtures of Gaussian sources, which HOS based BSS methods fail to.

In this chapter copula techniques are adapted to simplify the modelling of signal dependence for MIMO wireless communication simulation purposes and hence enable a study of the effects of dependence on information rates. The suitability of copula methods for BSS in MIMO applications is also briefly discussed.

5.2 Copula theory

A copula can be briefly described as a function that connects one-dimensional marginal probability distributions through a single multivariate probability distribution and may therefore be used as a means for deriving multivariate distributions with any desired dependence incorporated. There are several well-known copula function families which are described in the literature, for example see Nelsen [76]. The basis for the theory of copulas stems from Sklar's Theorem [76] which states that an m-dimensional copula is a function C from the unit m-cube $[0,1]^m$ to the unit interval $[0,1]$ and satisfies certain conditions. In other words, an m-copula is an m-dimensional cumulative distribution function (cdf) where all m marginal distributions are uniform. To understand the relationship between distribution functions and copulas, consider a continuous, real-valued, m-variate distribution function $F(\mathbf{y}) = F(y_1, \ldots, y_m)$ with univariate marginal distributions $F_1(y_1), ..., F_m(y_m)$ and inverse functions $F_1^{-1}, \ldots, F_m^{-1}$. Then $y_1 = F_1^{-1}(u_1) \sim F_1, \ldots, y_m = F_m^{-1}(u_m) \sim F_m$ where $u_1, \ldots, u_m$ are uniformly distributed variates. Hence

$$\begin{aligned} F(\mathbf{y}) &= F(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)) \\ &= Pr[U_1 \leq u_1, \ldots, U_m \leq u_m] \\ &= C(u_1, \ldots, u_m) \\ &= C(\mathbf{u}) \end{aligned} \tag{5.1}$$

is the copula associated with the distribution function. That is, if $\mathbf{y} \sim F$, and F is continuous then

$$(F_1(y_1), \ldots, F_m(y_m)) \sim C, \tag{5.2}$$

and if $\mathbf{u} \sim C$, then

$$(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)) \sim F. \tag{5.3}$$

The copula function is frequently written as $C(F_1(y_1), \ldots, F_m(y_m); \theta)$, where θ is a parameter of the copula called the dependence parameter and measures dependence between the marginal distributions. One of the advantages of using copulas is that the marginal distributions can be from different distribution families. We may therefore treat marginal distributions and dependence separately. In short the copula method involves specifying the marginal distributions of each random variable together with a function that links them together and a parameter that controls the level of dependence between the marginals. We shall make use of two copulas in this study: multivariate Gaussian for modelling dependence and the independent (or product copula) for BSS purposes.

The *product copula*, also known as the independent copula, has no dependence between variates. Its density function is unity everywhere. For independent random variables $y_1, \ldots, y_m$ the cdf is

$$F(y_1, \ldots, y_m) = \Pi_{k=1}^m F_k(y_k) \tag{5.4}$$

and the product copula is

$$C(u_1, \ldots, u_m) = \Pi_{k=1}^m u_k. \tag{5.5}$$

The real-valued, *multivariate Gaussian copula* takes the form

$$\begin{aligned} C(\mathbf{u};\boldsymbol{\Theta}) &= \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_m)); \boldsymbol{\Theta}), \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_m)} \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Theta}|^{1/2}} \times \exp\left\{-\frac{1}{2}\mathbf{y}^T\boldsymbol{\Theta}^{-1}\mathbf{y}\right\} dy_1 \ldots dy_m, \end{aligned} \tag{5.6}$$

where Φ_G is the real-valued, multivariate Gaussian distribution with correlation matrix $\boldsymbol{\Theta}$. $\boldsymbol{\Theta}$ is a symmetric, positive definite matrix with all ones on the main diagonal. If the marginals are standard real-valued, normal distributions then the Gaussian copula generates the standard real-valued, joint normal distribution function. The corresponding density is

$$c(\Phi(y_1), \Phi(y_2), \ldots, \Phi(y_m); \boldsymbol{\Theta}) = \frac{\frac{1}{(2\pi)^{m/2}|\boldsymbol{\Theta}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^T\boldsymbol{\Theta}^{-1}\mathbf{y}\right\}}{\prod_{k=1}^{m}\left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_k^2\right\}\right)}. \tag{5.7}$$

Let $u_k = \Phi(y_k)$, so that $y_k = \Phi^{-1}(u_k)$, then the density may be written as [21]

$$c(u_1, u_2, \ldots, u_m) = \frac{1}{|\boldsymbol{\Theta}|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{\Psi}^T(\boldsymbol{\Theta}^{-1} - \mathbf{I})\boldsymbol{\Psi}\right\}, \tag{5.8}$$

where $\boldsymbol{\Psi} = [\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_m)]^T$.

5.3 Correlated fading

In this section we develop the MIMO wireless propagation model and show how a copula may be employed to account for dependence or correlation in the propagation channel. In wireless communications, fading is the attenuation that a signal experiences when passing through a propagation medium and is often modelled as a random process. Reflectors in the environment surrounding a transmitter and receiver create multiple paths that a transmitted signal may follow. As a result, a receiver sees the superposition of multiple copies of the transmitted signal. Each copy will experience differences in attenuation, delay and phase shift. The most commonly employed probability distributions for modelling such multipath fading effects are:

Rayleigh fading. The Rayleigh fading model is used when there is no line of sight signal. This model assumes that the magnitude of the signal varies randomly according to a Rayleigh distribution, i.e. the magnitude of the sum of two uncorrelated Gaussian random variables. Rayleigh fading has been used to model the effect of urban environments on radio signals when there is no dominant line of sight propagation between the transmitter and receiver.

Rician fading. Rician fading is used to model the case where there is a dominant propagation path. The signal arrives at the receiver via different paths and the signal from one of the paths, usually the line of sight path, is much stronger than the others.

Nakagami fading. The sum of multiple i.i.d. Rayleigh fading signals have a Nakagami distributed signal amplitude. Nakagami fading occurs for multipath scattering with large time delay spreads, with different groups of scattered signals. Within any one group, the phases of individual scattered signals are random but the time delays are approximately equal for all signals. The magnitude of the sum of signals in a group is Rayleigh distributed. The average time delay is assumed to be significantly different between groups.

The MIMO wireless RF scenario was presented earlier in Section 2.5 and a mathematical model commonly employed for MIMO wireless simulation purposes was described in Section 3.1. To include dependence, or correlation, we may write

$$\mathbf{Y} = f(\mathbf{A})\mathbf{X} + \mathbf{W}, \tag{5.9}$$

where $f(\mathbf{A})$ is a function that imposes dependence on the channel matrix. For example, in a spatially correlated wireless channel, we might have

$$f(\mathbf{A}) = \mathbf{R}^{1/2}\mathbf{A}\mathbf{T}^{1/2}, \tag{5.10}$$

where $\mathbf{R}$ and $\mathbf{T}$ are, respectively, an $m \times m$ receive correlation matrix and an $m \times m$ transmit correlation matrix. Dependence may be introduced at the transmitter array, the receiver array, within the propagation channel or any combination of these. Alternatively we may introduce dependence in the channel matrix by treating the matrix as a vector of i.i.d. components and then applying the copula technique for multivariate dependence. This provides a flexible approach which allows us to use channel coefficients with different distributions, if required, and introduce dependence between these coefficients using one of the many multivariate copula distribution functions that are available. To obtain a sequence of random fading channel coefficients that are dependent we may take the following approach

- Let $\mathbf{a} = \mathbf{vec}\,(\mathbf{A})$ that is $\mathbf{a}$ is an $m^2 \times 1$ vector with elements $a_1, a_2, \ldots, a_{m^2}$ that are independent random variables distributed according to whatever fading type we require; this could be a mix of Rayleigh, Rician or Nakagami-m random variables.

- Let $a_i \sim P_i(a_i)$ that is a_i is distributed with distribution function $P_i(a_i)$ and let the desired joint distribution for $\mathbf{a}$ be $P(\mathbf{a}) = P(a_1, a_2, \ldots, a_{m^2})$. The copula is defined for $P(\mathbf{a})$ as $C(\mathbf{u}) = P(\mathbf{a})$, where $u_i = P_i(a_i)$ or, alternatively, $a_i = P_i^{-1}(u_i)$ and the u_i are uniformly distributed variates.

- The inverse functions of the marginal distributions are $P_1^{-1}, \ldots, P_{m^2}^{-1}$ so that $a_1 = P_1^{-1}(u_1), a_2 = P_2^{-1}(u_2), \ldots, a_{m^2} = P_{m^2}^{-1}(u_{m^2})$, where $u_1, \ldots, u_{m^2}$ are uniformly distributed variates. Hence we have

$$P(\mathbf{a}) = P(P_1^{-1}(u_1), \ldots, P_{m^2}^{-1}(u_{m^2})) = C(u_1, \ldots, u_{m^2}) = C(\mathbf{u}), \qquad (5.11)$$

  where $C(\mathbf{u})$ is the copula that must be chosen to link the marginals.

In short the procedure for generating a channel matrix sequence, with dependence, is as follows

1. Choose an appropriate multivariate copula and generate a matrix, which is $m^2 \times n$, of dependent uniformly distributed random variables, where each row corresponds to one of the channel matrix coefficients.

2. Choose a fading distribution for each of the elements (rows) of the matrix and apply the inverse function so that a matrix of dependent random variables with the desired distributions is obtained.

3. Convert the matrix to a sequence of $m \times m$ matrices using the inverse of the $\mathbf{vec}\,(\cdot)$ operation for each column of the $m^2 \times n$ matrix.

To obtain a complex-valued mixing matrix we assume that the real components are independent of the imaginary components and repeat the above procedure to obtain two real matrices. This procedure yields the correct random matrix for the fading amplitude, however we need to also consider the distribution of the phase of the coefficients. Let $\mathbf{A}_R$ and $\mathbf{A}_I$ represent, respectively, the real and imaginary parts of $\mathbf{A}$, obtained from two repetitions of the above procedure, then we obtain the correct phase distribution (for the Nakagami distribution [113]) in forming the complex mixing matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{A}_R \odot sign(\mathbf{B}_R) + j\mathbf{A}_I \odot sign(\mathbf{B}_I), \tag{5.12}$$

where $\mathbf{B}_R$ and $\mathbf{B}_I$ are two matrices with i.i.d. normally distributed components and which are the same size as $\mathbf{A}_R$ and $\mathbf{A}_I$ repectively. Complex values are formed using the imaginary unit $j = \sqrt{-1}$ and $\odot$ is the Hadamard or elementwise product. A similar method is described by Ma & Zhang [66], who generated two independent random Gaussian input sequences, transformed these into Nakagami sequences, multiplied each by the sign of its original Gaussian input sequence then combined as a single complex sequence. Here we have simply multiplied the Nakagami sequences, representing the real and imaginary parts of the complex sequence, by the signs of two independent Gaussian sequences.

5.4 Blind source separation

Many BSS techniques have been developed that attempt to separate a multivariate signal into subcomponents that are mutually independent. BSS techniques typically rely on objective function tests for non-Gaussianity in the estimated components and the independent components are identifiable only up to a permutation and scaling of the sources. Copula based approaches have been previously proposed by Chen et al. [20] and Ma and Sun [64], which have preprocessing steps in common with the ICA algorithms. In the case of a multivariate Gaussian copula the copula parameter is the correlation matrix so that the sources will have been resolved when there is zero correlation between separated components. Alternative tests for independence include Kendall's τ, described by Christensen in [23], the Robust, Accurate, Direct, ICA, aLgorithm (RADICAL) algorithm, developed by Learned-Miller and Fisher in [58] and correlation between marginals. A pseudocode listing for RADICAL is provided in Appendix M. RADICAL first whitens the observation data and then proceeds to optimize the spacings entropy (a measure of MI) of the estimated sources, by a brute-force testing over all possible Jacobi rotation angles. Once the optimal Orthogonal (rotation in 2-D) matrix has been found the algorithm returns the source and channel estimates; taking into account the prewhitening. RADICAL estimates the entropy of the marginals, and hence dependence, as a function of the order statistics of the marginals. This contrasts with FASTICA which uses a cumulant (kurtosis) based method. Kendall's τ [23] or the Kendall rank correlation coefficient is a statistic that measures the association between pairs of random sequences. All of the algorithms proceed with the following steps

- Center the observed data - subtract the mean and normalise (unit power) the observed mixture power.
- Whiten observations - via Eigenvalue Decomposition (EVD). This procedure converts the observation covariance matrix to an identity matrix and reduces the channel search space to a search for a unitary transformation.
- Find a unitary (complex data) or orthogonal (real data) transformation that minimises an objective function : kurtosis, negentropy, correlation, copula parameter.

5.5 Simulation Results

5.5.1 Correlated Fading

A GNU Octave implementation for dependent fading channel generation, in the complex-valued model, i.e. $\mathbf{A}, \mathbf{X}, \mathbf{Y}, \mathbf{W}$ are all complex-valued, has been developed and is listed in Appendix L for reference. The code allows for marginal channel distributions to be chosen from either the Rayleigh or Nakagami-m distributions. The multivariate channel copula may be selected from the Normal, Student-t, Clayton, Frank or Gumbel distributions. The Nakagami-m distribution can be obtained in two different ways: inverse distribution approximation, e.g. see Beaulieu and Cheng [12] or inverse gamma distribution, e.g. see Zhang [118] then take the square root of the result. The latter method is attractive since the inverse function for the gamma distribution already exists in Octave and Matlab. The first method is based on calculating coefficients for particular function values and its use requires a look up table, with interpolation to obtain points not previously calculated.

Simulations have been performed to demonstrate the utility of this method and the results are shown in Figures 5.1 and 5.2. In the simulations two elements of a channel matrix are studied. A sequence of 1000 instances of the pair of elements is generated where the elements follow a Nakagami-m distribution and a Gaussian copula is utilised, with a correlation matrix where the cross-correlation terms $= 0.9$, i.e. they are highly dependent. The true phase and amplitude expressions for the Nakagami-m distribution were obtained from Yacoub et al. [113] and are used in the simulations for comparison.

Figure 5.1 shows a scatter plot of the correlated amplitudes of the two elements. The associated histogram plots compare the amplitude distributions with the theoretical Nakagami amplitude distribution and show a good agreement between simulation results (shown in blue) and theory (shown as green curves).

Figure 5.2 shows the correlated phases of the two elements. The associated histogram plots compare the phase distributions with the theoretical Nakagami phase distribution and show a good agreement between simulation results (shown in blue) and theory (shown as green curves).

5.5.2 BSS, Real Model

In this section we compare the BSS performance of: MLE, FASTICA, copula using Kendall's τ [23], copula using cross correlation, RADICAL [58]. We only consider a real-valued model here, i.e. $\mathbf{A}, \mathbf{X}, \mathbf{Y}, \mathbf{W}$ are all real-valued, because the RADICAL algorithm has only been implemented for real-valued data. We have simulated the MIMO scenario where there is a transmitter array with two elements and a receiver array with two elements so that the channel $\mathbf{A}$ is represented by a 2×2 matrix. Random message blocks $\mathbf{X}$, of size 2×500, were generated under the assumption that the channel matrix remained constant for this block length. The distribution of the independent sources was controlled by employing the GG distribution, described in Appendix C, parameterised by α, where the distribution is Gaussian when $\alpha = 2$. When $\alpha < 2$ the distribution has a positive kurtosis and when $\alpha > 2$ the distribution has a negative kurtosis. For each value of α, 100 instances of the 2×500 message block and the 2×2 channel matrix were generated. A Gaussian noise matrix $\mathbf{W}$ was added so that the input snr was $10\,dB$. The performance of the separation algorithms, for each repetition, was calculated as

$$\text{output snr} = 10\log_{10}\left(\frac{\sum_{i,j}|x_{ij}|^2}{\sum_{i,j}|x_{ij}-\hat{x}_{ij}|^2}\right). \tag{5.13}$$

The average performance of the separation algorithms was then calculated as the mean output snr over the 100 repetitions. Ambiguities in scale and permutation have also been taken into account in the simulations. In the real-valued channel and real-valued data case, after prewhitening, the algorithms must find an orthogonal 2×2 matrix that maximises the estimated source independence. This is equivalent to finding the optimum angle for a 2D rotation matrix.

Figure 5.3 compares the simulation results for the real-valued channel and real-valued data case. The MLE assumes that the mixing matrix is known a priori and therefore performs better than the other algorithms. The MLE results are also seen to be independent of the source distribution. The FASTICA results have a minimum when the GG distribution parameter $\alpha = 2$, confirming the well-known fact that this algorithm has difficulty in separating a mixture of Gaussian sources. However, as $|\alpha - 2|$ increases, FASTICA is better able to separate the sources. Results from the RADICAL algorithm are similar to those from FASTICA for $\alpha < 2$ but degrade when $\alpha > 2$. Results from the copula-based approach using Kendall's τ and correlation, are poor but are clearly independent of the source distribution. These last two methods appear to be no better than the poorest results from FASTICA. This seems to indicate that copula-based techniques for BSS may not be useful for practical separation of digital communication waveforms that typically have a non-Gaussian distribution.

5.5.3 BSS, Complex Model

We have simulated the complex-valued MIMO scenario, i.e. $\mathbf{A}, \mathbf{X}, \mathbf{Y}, \mathbf{W}$ are all complex-valued, where there is a transmitter array with two elements and a receiver array with two elements so that the channel $\mathbf{A}$ is represented by a 2×2 matrix. Random message blocks $\mathbf{X}$, of size 2×500, were generated under the assumption that the channel matrix remained constant for this block length. The distribution of the independent sources was controlled by employing the GG distribution. The real and imaginary parts of $\mathbf{X}$ were generated using the same value of α. For each value of α, 100 instances of the 2×500 message block and the 2×2 channel matrix were generated. A Gaussian noise matrix $\mathbf{W}$ was added so that the input snr was $10\,dB$. The performance of the separation algorithms, for each repetition, was calculated using equation 5.13. The average performance of the separation algorithms was then calculated as the mean output snr over the 100 repetitions. Ambiguities in scale and permutation have also been taken into account in the simulations. After observation data prewhitening, the algorithms must find a unitary 2×2 matrix that maximises the estimated source independence. In the 2D complex-valued channel and complex-valued data case, the unitary matrix is formed from a rotation angle and three phases and so there are four degrees of freedom that must be optimized, as described by Dita in [30].

Figure 5.4 compares the results for the complex-valued channel and complex-valued data case. As for the real-valued case the FASTICA results have a minimum when $\alpha = 2$. Results from the Kendall's τ and correlation algorithms are poor but again are independent of the source distribution. We note that, even when the sources are close to Gaussian, the Kendall's τ and correlation algorithms seem to perform no better than FASTICA.

5.5.4 BSS, Correlated Complex Channel

In the following simulations the copula technique described in section 5.3 is employed to introduce channel correlations into the point-to-point MIMO model, i.e. we do not attempt copula-based BSS in these simulations. A 2×2 wireless link is envisaged where the source distribution is controlled via the parameter α in the GG distribution. The complex model is assumed here, i.e. all of $\mathbf{A}, \mathbf{X}, \mathbf{Y}, \mathbf{W}$ are complex-valued. The values of α were converted to kurtosis (see Appendix C) for future comparison with digital source kurtosis values. Theoretical MI values, I_B-Theory, for a channel-informed receiver (Bob) were calculated using equation 3.53 and theoretical MI estimates, I_E-Theory, for an uninformed receiver (Eve), were calculated using equation 3.54. Eve's ability to estimate the channel is reflected in the variance $\sigma^2_{\hat{a}}(\kappa)$, equation 4.72, which is a function of source kurtosis and is included in I_E-Theory. ML estimation was used to obtain I_B-MLE and the FASTICA algorithm [55] was employed to obtain estimates of I_E-ICA. Both I_B and I_E were scaled by $\frac{1}{mn}$ so that the units are shown as Bits sec^{-1} ant^{-1}.

For each parameter set: {kurtosis (κ), correlation (ρ), simulation iteration}, a sequence of complex-valued channel matrices was generated. The components $a_{i,j}$ of the channel matrices had a Nakagami fading distribution and correlation between the components was introduced using the method described in Section 5.3. In all cases the blocklength is $N = 1000$ and the snr is 20dB.

Figure 5.5 shows the results for the case where $\rho = 0$, i.e. the channel components are i.i.d.. The I_B-Theory and I_B-MLE plots follow each other quite closely. A notable feature of these two plots is the peak when $\kappa = 0$, i.e. Gaussian sources. At this point I_B is maximised and hence determines the channel capacity. The I_B plot shows the reduction in channel capacity as a result of decreasing the source Gaussianity. The plot for I_E-Theory shows a sharp dip when $\kappa = 0$, highlighting the inability of BSS techniques to separate Gaussian distributed sources. The results for I_E-ICA provide a reasonable match to I_E-Theory except that the width of the dip for I_E-ICA is much wider than that for I_E-Theory and does not reduce as much at $\kappa = 0$.

In Figure 5.6 some channel correlation has been introduced, i.e. $\rho = 0.3$. The only differences between the $\rho = 0.3$ and $\rho = 0$ results seem to occur for large positive values of kurtosis $\kappa > 3$. For these values of kurtosis both the I_B-MLE and I_E-ICA plots appear to noticeably fall below their respective theoretical plots.

In Figure 5.7 the channel correlation is $\rho = 0.6$ with the same features as for $\rho = 0.3$ but now the simulation results for I_B-MLE and I_E-ICA more noticeably fall below their theoretical counterparts.

In Figure 5.8 the channel correlation is $\rho = 0.9$. Now the gap between I_B-MLE and I_B-Theory has increased but is still less than 1 Bit sec^{-1} ant^{-1}.

The results above clearly demonstrate that a reduction in I_B occurs as the source kurtosis departs from zero. A further reduction in I_B is incurred as the channel coefficients becomes more correlated. However the situation for Eve is quite different. As the source kurtosis departs from zero Eve is better able to separate the mixed sources using ICA. Increasing the channel correlation degrades I_E at high values of kurtosis. As the sources distributions approach zero kurtosis, or as they become more Gaussian, the difference between I_B and I_E increases, reaching its maximum at zero-kurtosis. This clearly indicates that maximum secrecy capacity, which was stated earlier in section 2.3 as $C_S = C_M - C_{MW}$ or, in this case $I_S = I_B - I_E$, is attained using Gaussian sources and for an eavesdropper employing a HOS based method for BSS.

5.6 Summary

A method has been developed to combine RF propagation fading effects with channel dependence for modelling complex-valued MIMO wireless channel scenarios. This has been made possible through the use of copula theory and has resulted in a practical approach which has proved useful in the study of channel dependence effects. A MIMO wireless channel simulator, with channel dependence, has been implemented and allows for a selection of the fading distribution as well as the level of dependence required. Simulations were performed to demonstrate a Nakagami fading channel with various degrees of correlation between the channel coefficients.

The performance of a selection of BSS techniques was evaluated, for both real-valued and complex-valued models, with no channel dependence. Simulation results provided a comparison of these techniques and showed that the FASTICA algorithm performed better than the copula-based techniques for BSS. However we note that the i.i.d. GG distribution, used to generate the sources in these simulations, had no temporal structure. The copula-based methods may perform better with sources that have some temporal structure.

A simulation exercise was performed to study how MI is affected by both source kurtosis and channel correlation. For this exercise the FASTICA algorithm was employed. A number of observations were made regarding the degradation of MI attainable by a channel-informed receiver or an uninformed receiver. In short using a Gaussian distributed source (zero kurtosis) is optimal for a channel-informed link both for maximising channel capacity and for minimising the ability of an eavesdropper, who is using a HOS based BSS technique, to resolve the sources.

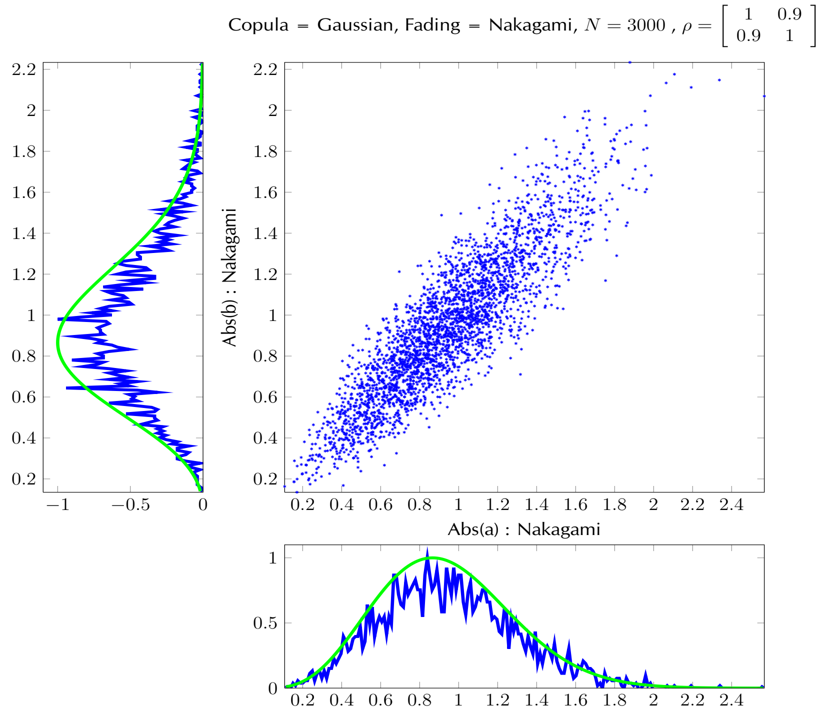

Figure 5.1: Amplitude distributions, Nakagami fading, Gaussian copula.

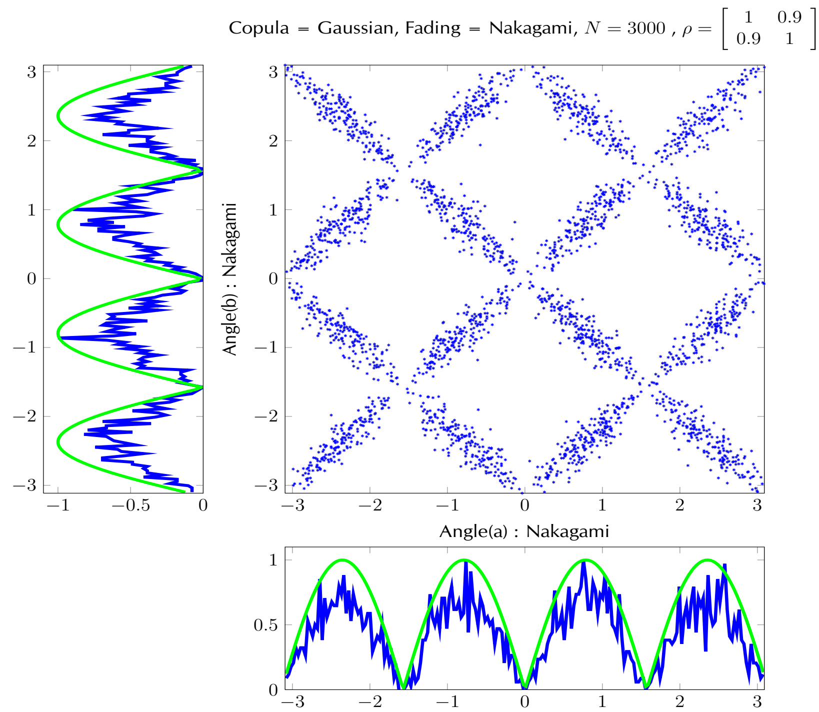

Figure 5.2: Phase distributions, Nakagami fading, Gaussian copula.

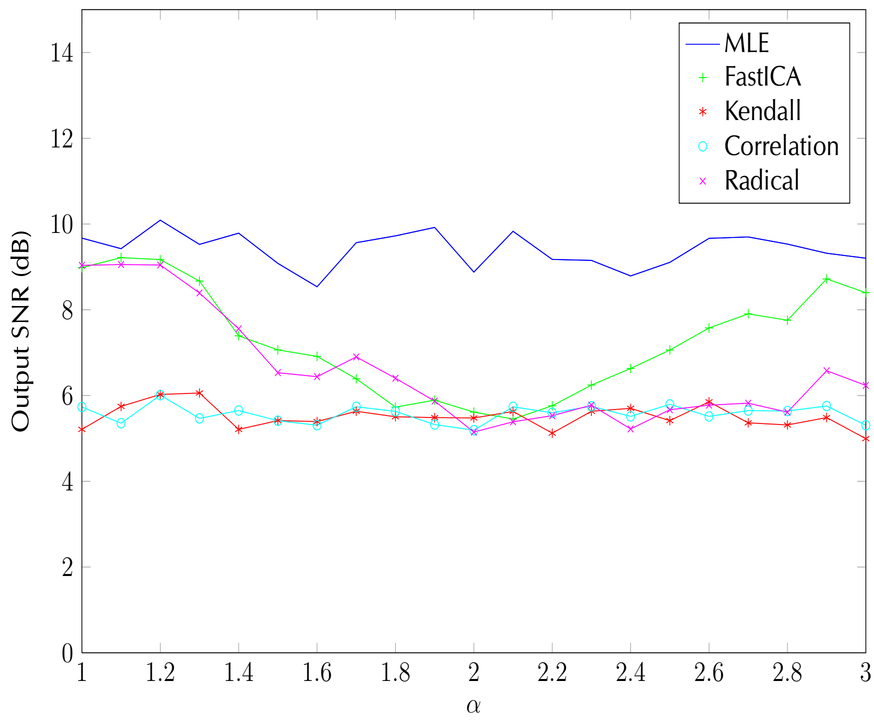

Figure 5.3: Separation performance, real-valued data, SNR = 10dB, N = 500, Array = 2.

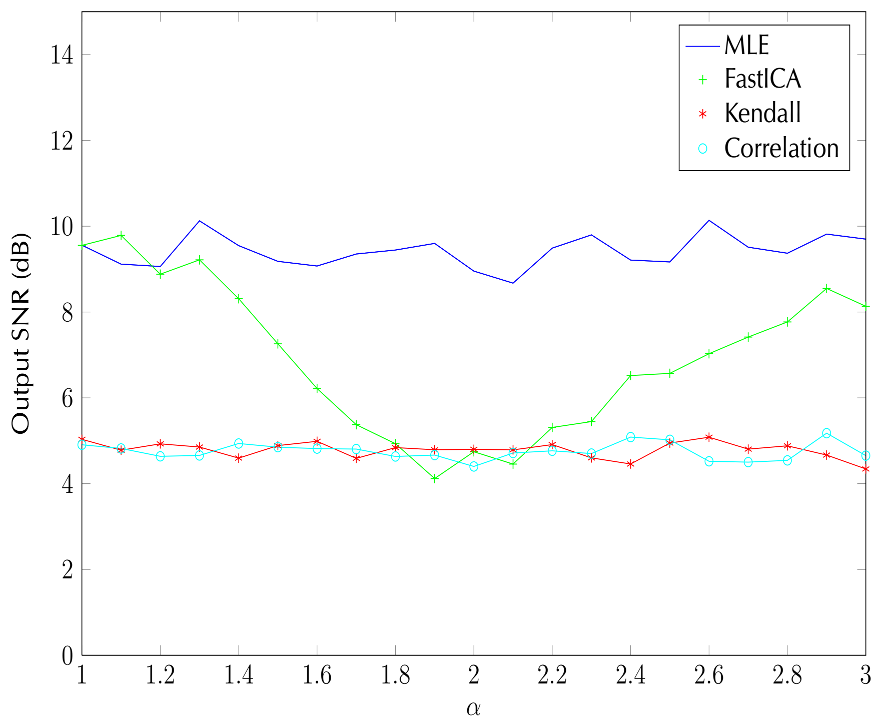

Figure 5.4: Separation performance, complex-valued data, SNR = 10dB, N = 500, Array = 2.

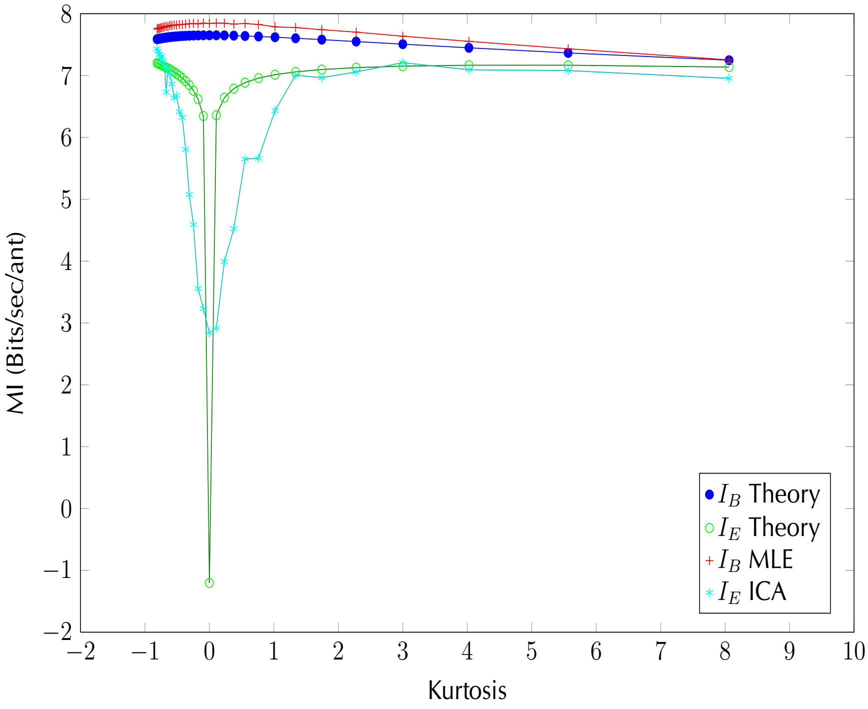

Figure 5.5: MI Vs Kurtosis, SNR = 20dB, N = 1000, Array = 2, corr. = 0.

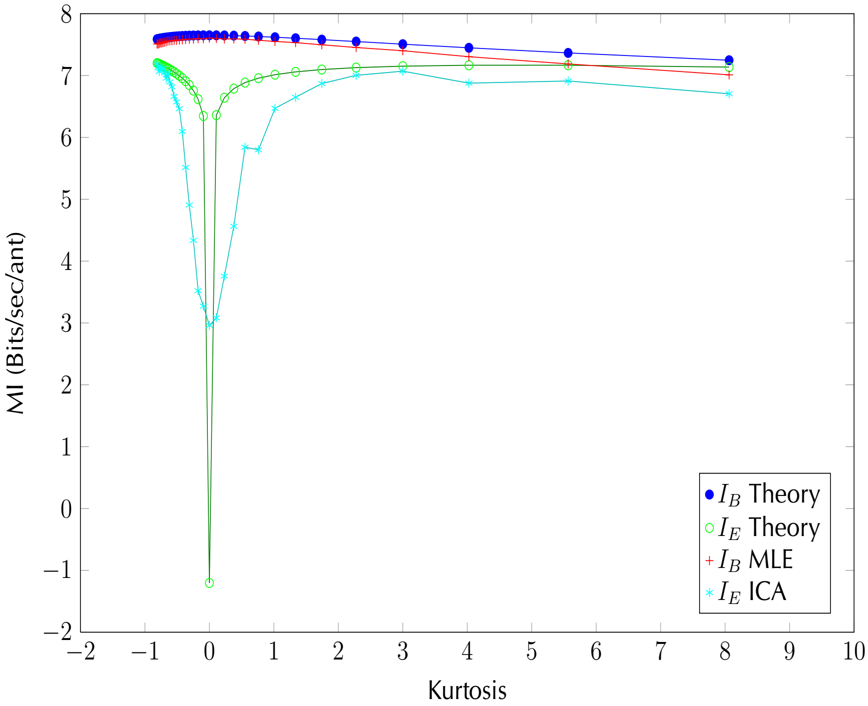

Figure 5.6: MI Vs Kurtosis, SNR = 20dB, N = 1000, Array = 2, corr. = 0.3.

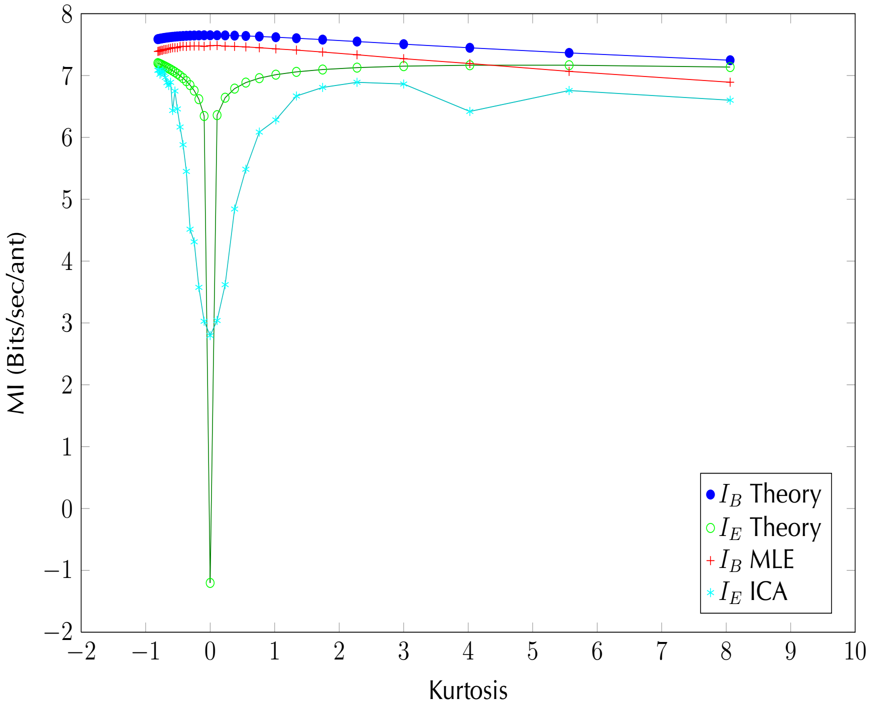

Figure 5.7: MI Vs Kurtosis, SNR = 20dB, N = 1000, Array = 2, corr. = 0.6.

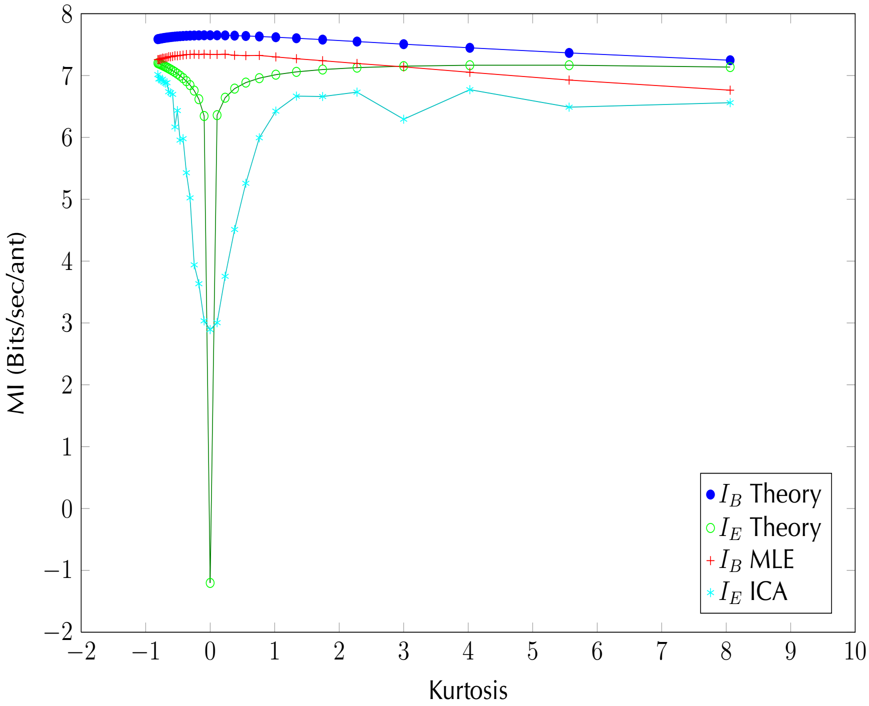

Figure 5.8: MI Vs Kurtosis, SNR = 20dB, N = 1000, Array = 2, corr. = 0.9.

Part III

Discrete Source Recovery

Chapter 6

Source Recovery Versus System Parameters

6.1 Introduction

In Part II performance measures for MLE and BSS techniques were derived, resulting in MLE expressions for source and channel estimation and approximate covariance bounds for estimating these parameters. The performance of BSS techniques was established in the form of CRB expressions for joint source and channel estimation. These measures provide performance bounds that may be compared with the results of Monte Carlo computer simulations of a MIMO wireless communications link or eavesdrop scenario.

In this chapter we examine the results of a set of simulation exercises which were designed to test source estimation performance as the source distribution is smoothly varied in terms of Gaussianity, or more appropriately for digital communication signals, in terms of source kurtosis. This is achieved by employing the GG to represent the pdf of the sources. The GG and the relationship between kurtosis and the GG parameter α is described in Appendix C. Whereas the simulation exercise carried out in section 5.5 studied the effects of channel dependence, the analysis presented in this chapter compares MI performance across a range of snrs, blocklengths and array sizes.

The underlying MIMO wireless communications model and assumptions that have been used here are the same as those listed in section 3.1. The simulations in this chapter represent the complex MIMO model, where $\mathbf{A}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$ are all complex-valued.

When using the FASTICA or JADE algorithms for BSS, scale and permutation issues are the same as described in Section 4.6.

A number of parameters are studied in this chapter and are defined here for reference

Definition 4 (A_b)**.** *The channel matrix, between Alice and Bob.*

Definition 5 (A_e)**.** *The channel matrix, between Alice and Eve.*

Definition 6 (I_B)**.** *Bob's mutual information*

Definition 7 (I_E)**.** *Eve's mutual information*

Definition 8 (I_B-Theory)**.** *Theoretical value for the mutual information attainable by Bob, where the channel A_b is assumed known. This is obtained from equation 3.53.*

Definition 9 (I_B-MLE)**.** *Simulation value for the mutual information obtained by Bob, using knowledge of the channel A_b. This is obtained from maximum likelihood estimation of the source and the mutual information expression given by equation 3.53.*

Definition 10 (I_E-Theory)**.** *Theoretical value for the mutual information attainable by Eve, where the channel A_e is unknown. This is obtained from theoretical values for the channel estimation variance, equation 4.72, and the mutual information expression given by equation 3.54.*

Definition 11 (I_E-ICA)**.** *Simulation value for the mutual information obtained by Eve. This is obtained from the source and channel estimation errors using the FASTICA algorithm substituted into the mutual information expression given by equation 3.54.*

Definition 12 (I_E-JADE)**.** *Simulation value for the mutual information obtained by Eve. This is obtained from the source and channel estimation errors using the JADE algorithm substituted into the mutual information expression given by equation 3.54.*

Definition 13 (A_{mle})**.** *Maximum likelihood estimator of channel matrix A_e, where the source is given.*

Definition 14 (A_{ica})**.** *Blind estimation of the channel matrix A_e, using the FASTICA algorithm.*

Definition 15 (X_{mle})**.** *Maximum likelihood estimator of source matrix X, where the channel is given.*

Definition 16 (X_{ica})**.** *Blind estimation of the source matrix X, using the FASTICA algorithm.*

6.2 Generalised Gaussian Simulation Analysis

The following Monte Carlo computer simulation studies were designed to compare previously derived theoretical expressions with the performance of the FASTICA algorithm for blind source estimation, allowing us to compare the information rate reduction in the legitimate system with the information rate increase in the eavesdropper channel.

The matrix dimensions for the channel $\mathbf{A}$ are $m \times p$. For the cases where the channel matrix is square, the FASTICA algorithm was used. However, for the overdetermined cases, i.e. where $m > p$, the JADE algorithm was used. This was necessary because the FASTICA algorithm assumes the same observation matrix size as the source matrix, i.e. size$(\mathbf{Y}) = p \times n =$size$(\mathbf{X})$, which means that the size of $\mathbf{A}$ must be $p \times p$. The JADE algorithm finds the p most significant eigenmatrices in its calculations and so only assumes that $m \geq p$. For the source signals that have been used here, no discernible difference in BSS performance between FASTICA and JADE was noted in square channel simulations; otherwise it would have been necessary to compare results obtained from both algorithms.

A pseudocode listing for the JADE algorithm is given in Appendix M. The JADE algorithm first performs a whitening operation on the observed data. Next a set of cumulant matrices is calculated from the whitened observations and the most significant set of eigenpairs (corresponding to the number of sources) is identified. The cumulant matrix set is jointly diagonalized by finding the optimal Jacobi rotations. A unitary unmixing matrix is found as the product of all the Jacobi rotations performed. The source and channel matrix estimates are returned, taking into account the prewhitening.

The parameter α in the generalised Gaussian distribution has been converted to the kurtosis equivalent value, as described by Cichocki and Amari in [24]. In the simulations a different channel realisation was generated for each of the legitimate and eavesdropper channels: $\mathbf{A}_b$ and $\mathbf{A}_e$ respectively, with the same distribution for each i.e. $[\mathbf{A}]_{ij} \sim \mathcal{CN}(0,1)$. For each parameter set: {kurtosis, snr, block length, channel dimensions}, 100 repetitions were performed and the mean of those results taken.

Figure 6.1 compares the simulation results for Bob's MI, I_B-MLE, and Eve's MI, I_E-ICA, with their theoretically predicted values, I_B-Theory (equation 3.53) and I_E-Theory (equation 3.54), derived earlier in Chapter 3. The parameter set used to obtain these results is: {snr(dB), blocklength (n), channel dimensions $(m \times p)$ }$=$ $\{10, 100, 2 \times 2\}$. For this small blocklength and low snr, I_B for both theory and MLE match closely but I_E-ICA only indicates a broad dip around zero kurtosis. Otherwise the I_E-Theory results lie approximately 1 Bit below the theoretical results.

Figure 6.2 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 100, 2 \times 2\}$. The increase in snr clearly raises the I_B values by approximately 3dB. The increase in I_E is not as great.

Figure 6.3 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{30, 100, 2 \times 2\}$. Once again the I_B results have increased by approximately another 3dB. However the increase in snr has had little effect on the I_E results.

Figure 6.4 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{10, 1000, 2 \times 2\}$. Increasing the blocklength at this snr has brought the results for I_B and I_E closer together, with only the dip at $\kappa = 0$ evident for I_E.

Figure 6.5 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 1000, 2 \times 2\}$. For this blocklength, increasing the snr, has led to an increase in the I_B and I_E results.

Figure 6.6 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{30, 1000, 2 \times 2\}$. Further increasing the snr has increased the I_B results but the increase in the I_E results is not as great. However I_E in this case is higher than I_E when the blocklength was 100.

Figure 6.7 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{10, 1000, 4 \times 4\}$. In this case the array size was increased and, after normalising by the number of antennae, are similar to those obtained for an array size of 2. The width of the dip in I_E-ICA is perhaps broader than for the two antenna case.

Figure 6.8 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 1000, 4 \times 4\}$. Increasing the snr increases the I_B results and the I_E results though not as much as in the 2-antenna case.

Figure 6.9 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{30, 1000, 4 \times 4\}$. Further increasing the snr to 30dB again increases I_B but the gap between I_B and I_E has increased. This gap is greater than the gap for the 2-antenna case.

Figure 6.10 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{10, 3000, 2 \times 2\}$. The I_B and I_E results are indistinguishable for this snr and blocklength, except for the small dip at $\kappa = 0$ for I_E.

Figure 6.11 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 3000, 2 \times 2\}$. Raising the snr to 20dB increases both I_B and I_E and these are still quite close together except for the dip in I_E at $\kappa = 0$.

Figure 6.12 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{30, 3000, 2 \times 2\}$. Further increasing the snr to 30dB raises both I_B and I_E with a small gap between the two sets of results. This gap is noticeably smaller than when the blocklength was 1000.

Figure 6.13 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 1000, 4 \times 2\}$. The is no obvious difference between this plot and the 2×2 channel case with the other parameters the same.

Figure 6.14 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{10, 3000, 8 \times 2\}$. Comparing this plot with the 2×2 channel case. We now observe that Eve's MI is reducing for the non-zero kurtosis values.

Figure 6.15 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 3000, 16 \times 2\}$. Comparing this plot with the 2×2 channel case. We observe that Eve's MI further reduces as the source for all kurtosis values.

Figure 6.16 compares the simulation results for MI with the theoretically predicted values using the simulation parameter set $\{20, 3000, 16 \times 4\}$. Clearly, as the receiver array dimension increases, the eavesdropper's performance decreases. This is a consequence of having to estimate a larger mixing matrix, where the errors in the estimate become larger.

In all of these figures there is a dip in eavesdropper MI ($\mathbf{I}_E$-Theory and $\mathbf{I}_E$-ICA) at zero kurtosis (Gaussian source), for the eavesdroppper using ICA, which is zero in theory. However the eavesdropper MI $\mathbf{I}_E$ rises rapidly as the source distribution departs from Gaussianity and approaches the MI ($\mathbf{I}_B$ theory and MLE) obtained via MLE for the legitimate receiver.

Some other features to note are:

- As the block length increases, the width of the dip in I_E-ICA reduces.
- As the block length increases, I_E improves.
- As the snr increases the gap between I_B and I_E increases.
- As the number of antennae is increased the width of the dip in I_E-Theory increases.

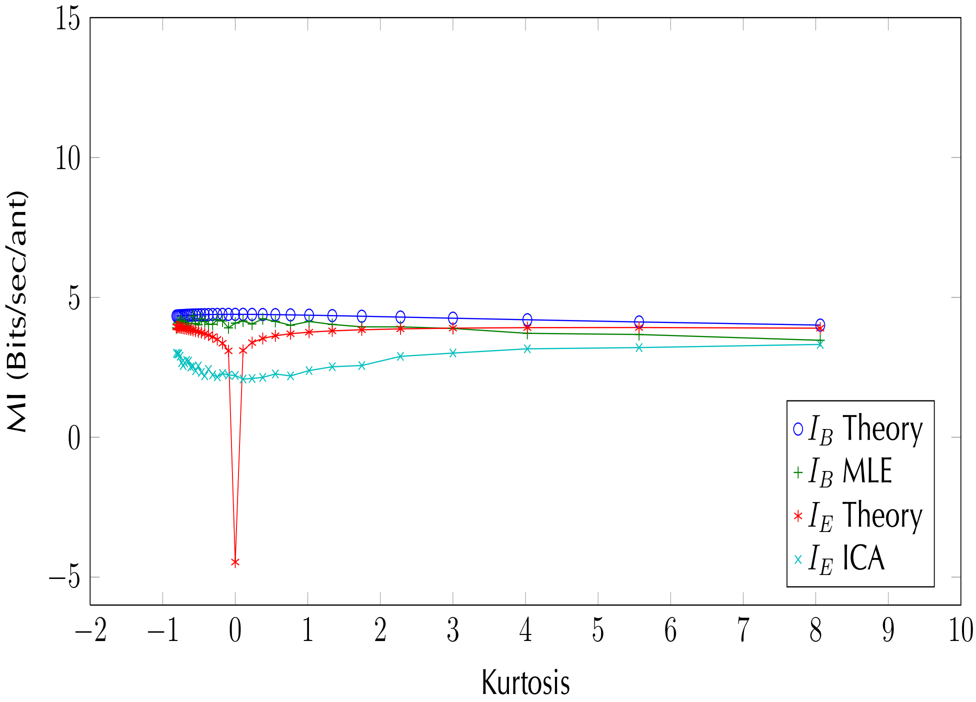

Figure 6.1: MI Vs Kurtosis, SNR = 10dB, N = 100, channel = 2×2.

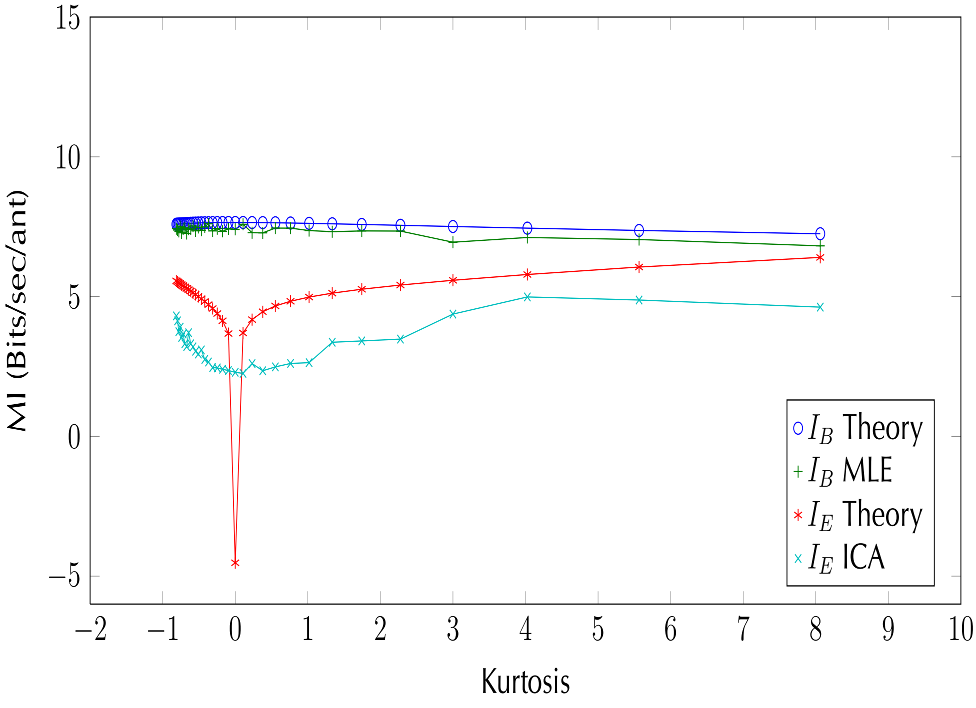

Figure 6.2: MI Vs Kurtosis, SNR = 20dB, N = 100, channel = 2×2.

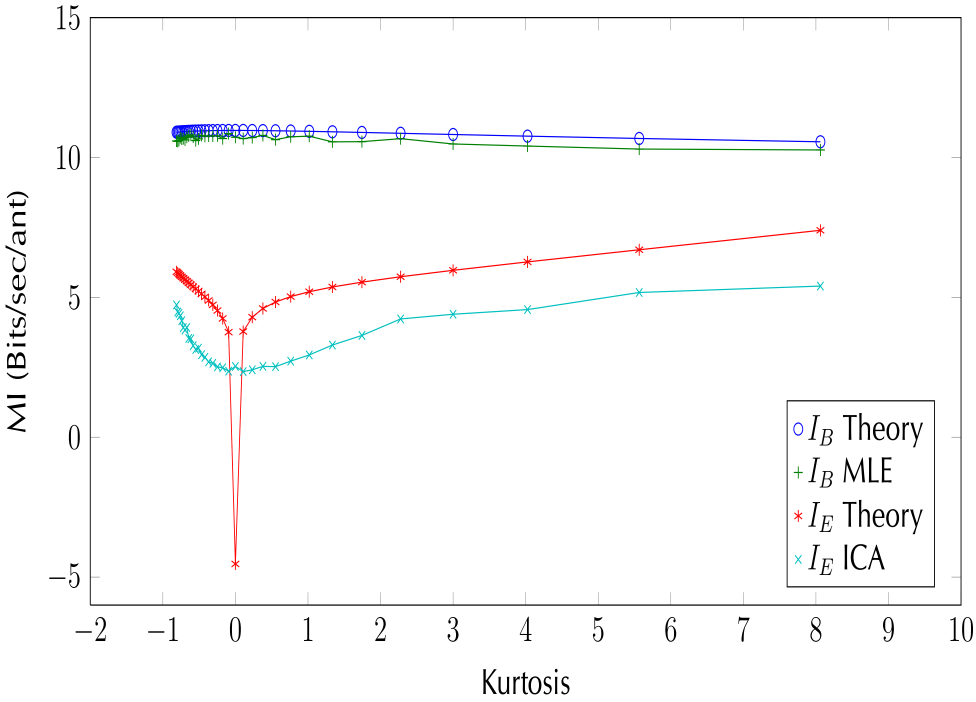

Figure 6.3: MI Vs Kurtosis, SNR = 30dB, N = 100, channel = 2 × 2.

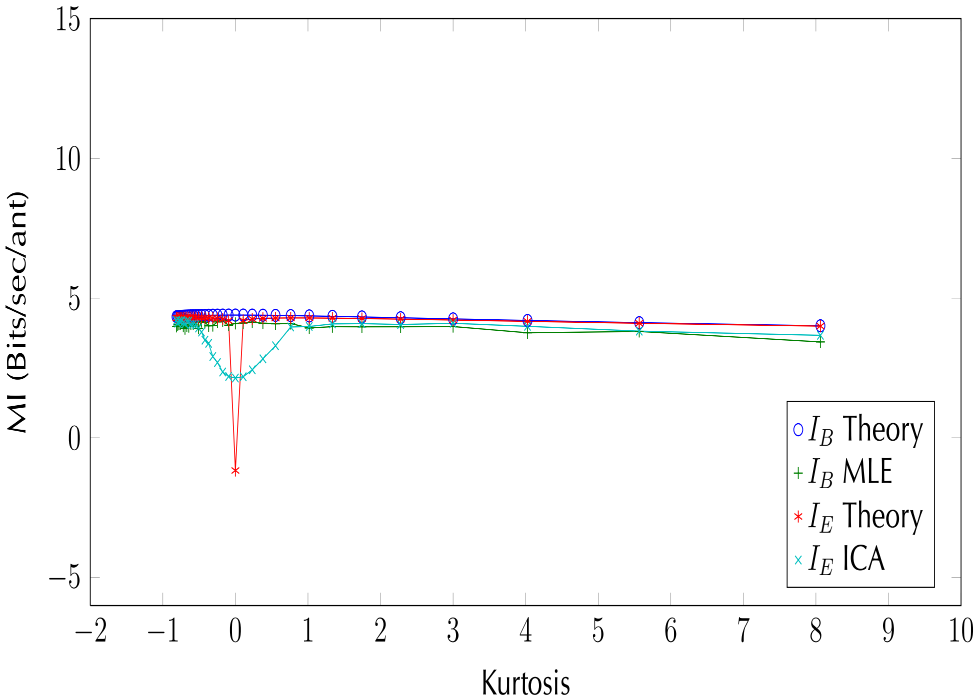

Figure 6.4: MI Vs Kurtosis, SNR = 10dB, N = 1000, channel = 2 × 2.

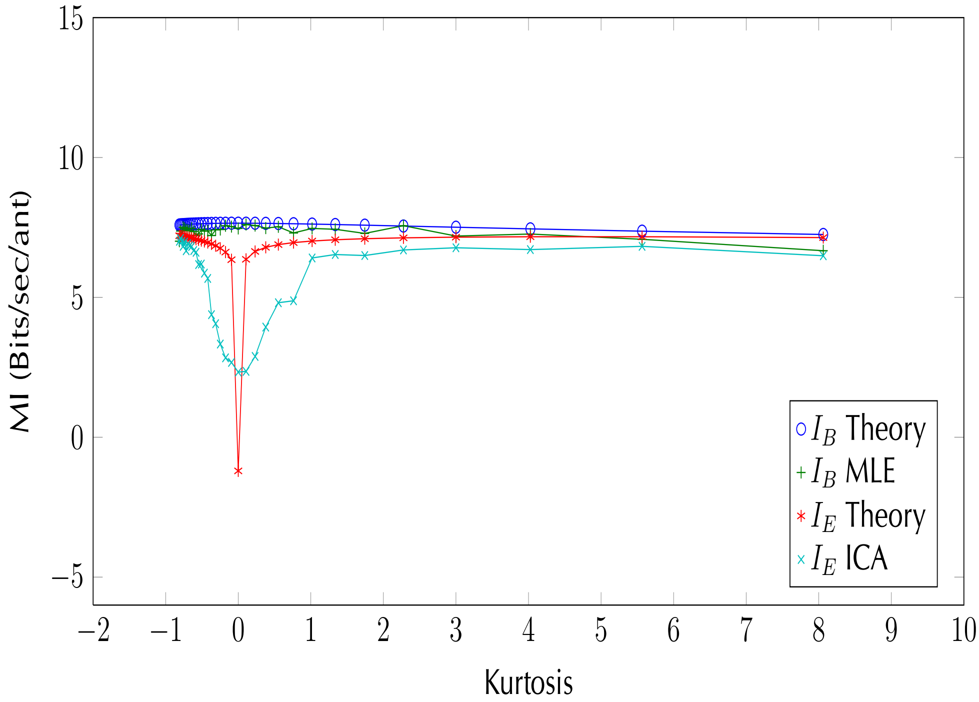

Figure 6.5: MI Vs Kurtosis, SNR = 20dB, N = 1000, channel = 2×2.

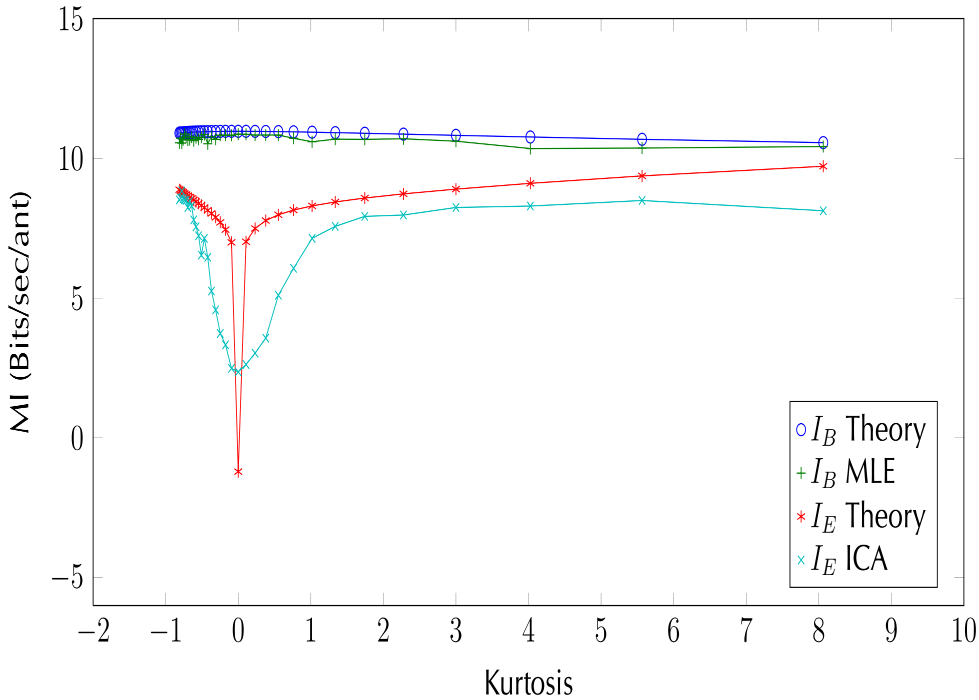

Figure 6.6: MI Vs Kurtosis, SNR = 30dB, N = 1000, channel = 2×2.

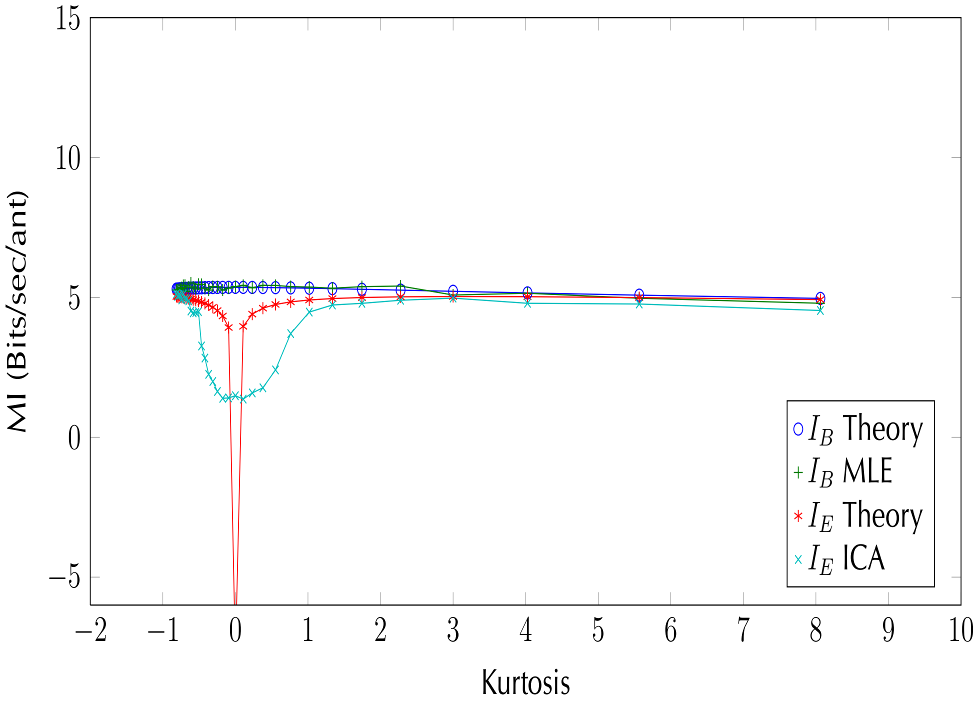

Figure 6.7: MI Vs Kurtosis, SNR = 10dB, N = 1000, channel = 4×4.

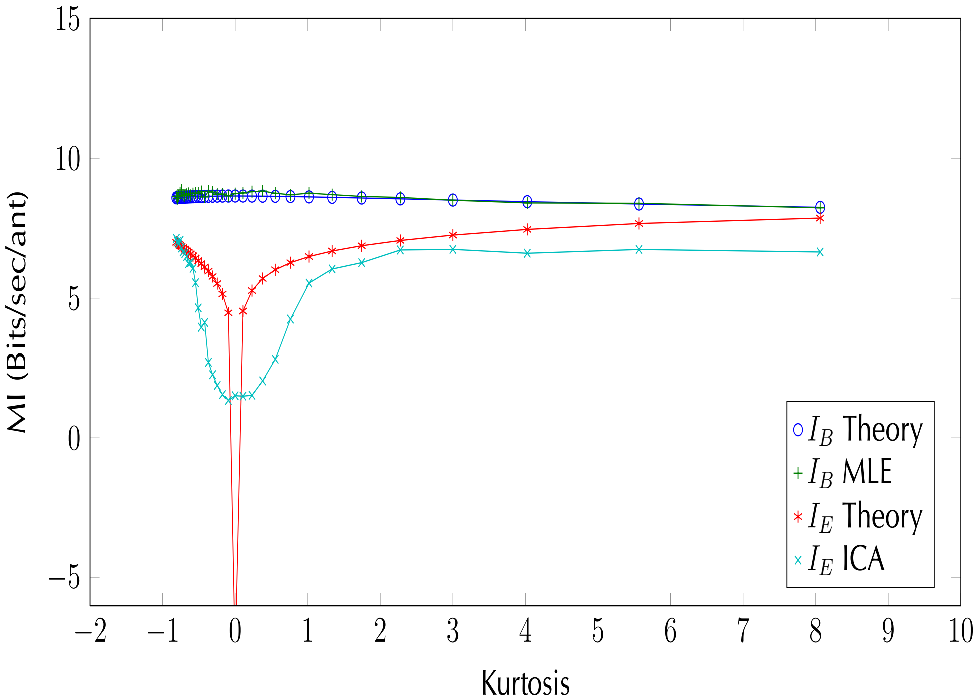

Figure 6.8: MI Vs Kurtosis, SNR = 20dB, N = 1000, channel = 4×4.

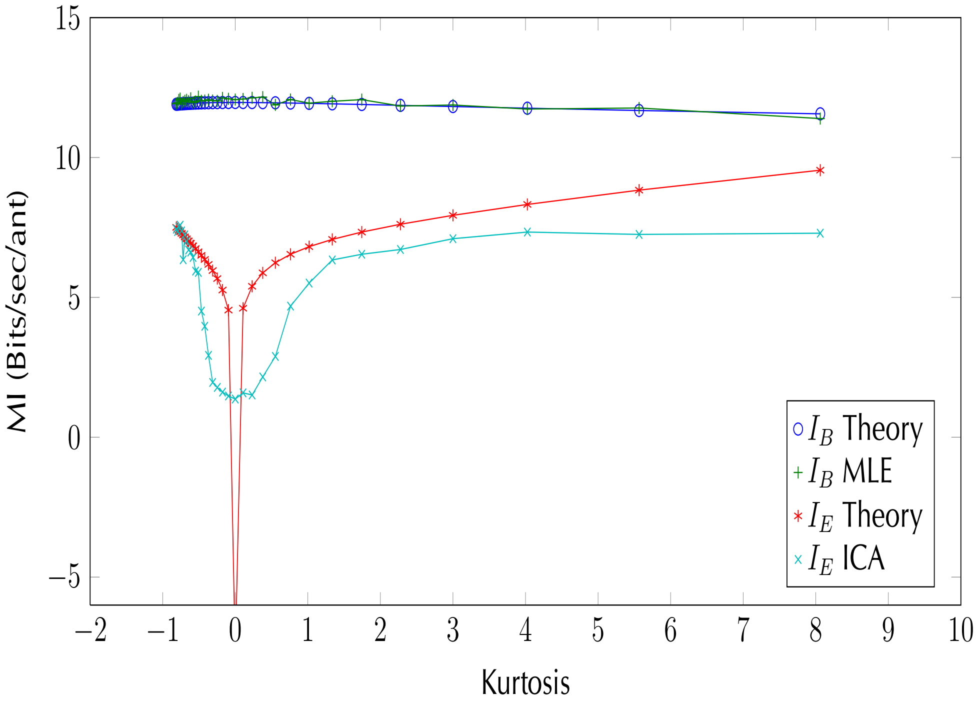

Figure 6.9: MI Vs Kurtosis, SNR = 30dB, N = 1000, channel = 4×4.

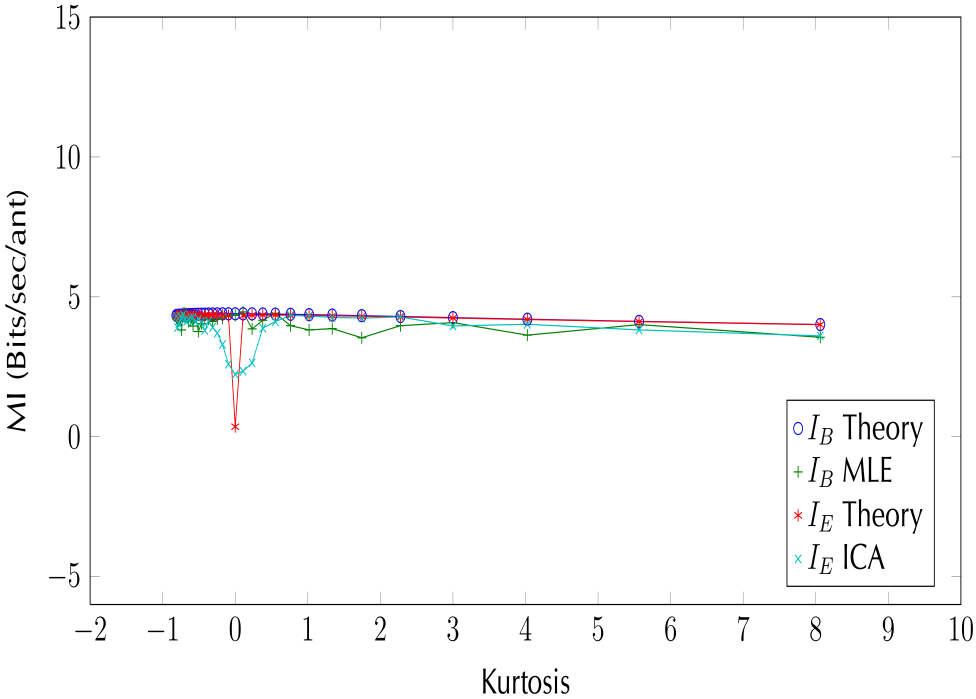

Figure 6.10: MI Vs Kurtosis, SNR = 10dB, N = 3000, channel = 2×2.

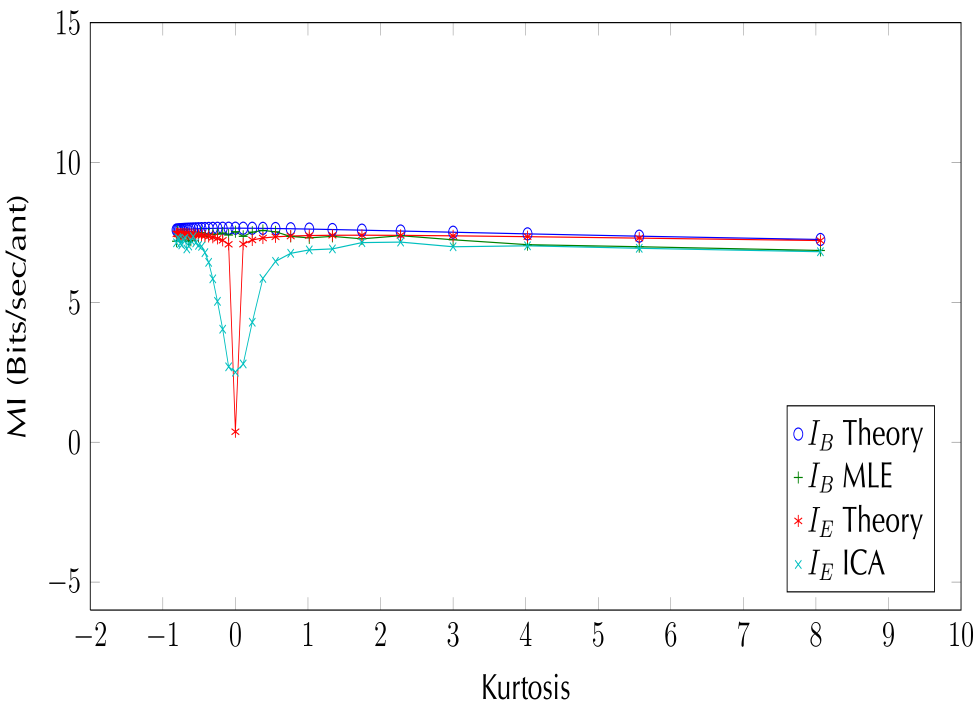

Figure 6.11: MI Vs Kurtosis, SNR = 20dB, N = 3000, channel = 2×2.

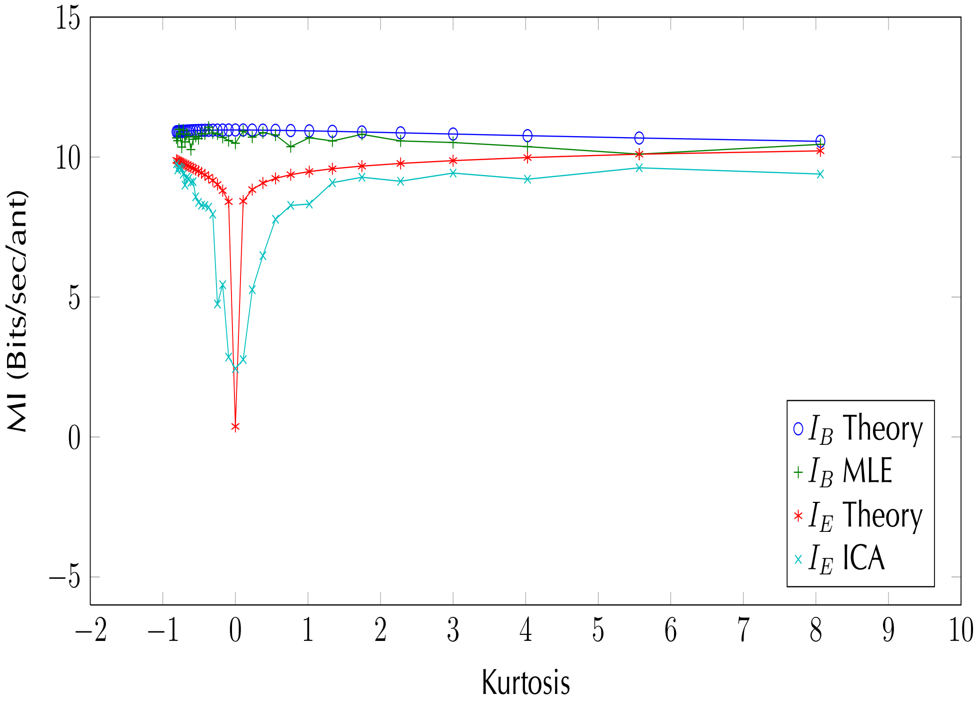

Figure 6.12: MI Vs Kurtosis, SNR = 30dB, N = 3000, channel = 2×2.

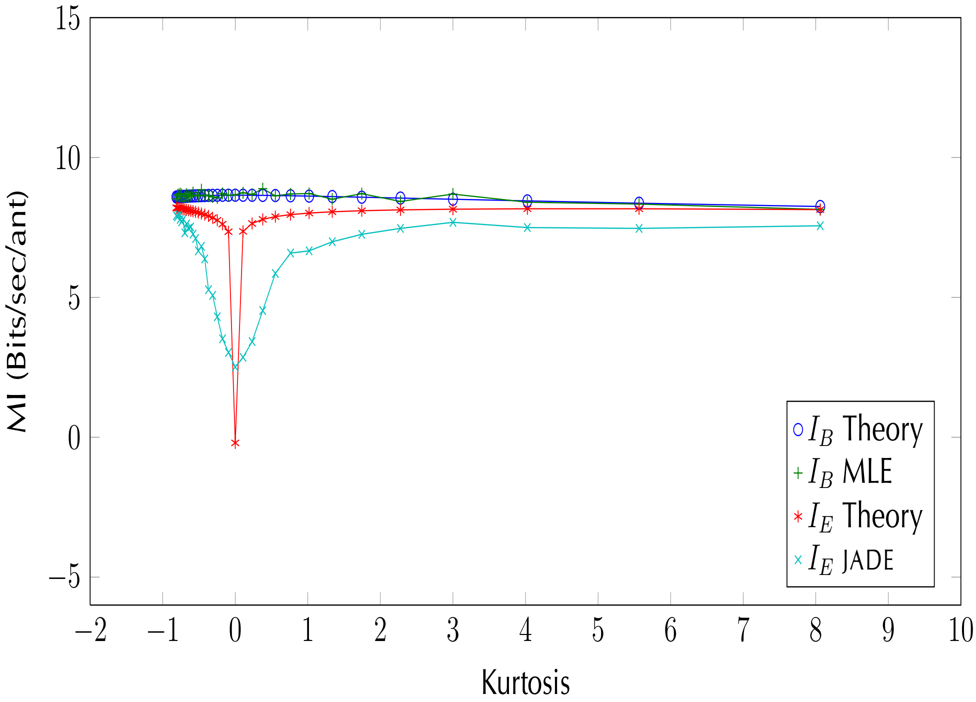

Figure 6.13: MI Vs Kurtosis, SNR = 20dB, N = 1000, channel = 4×2.

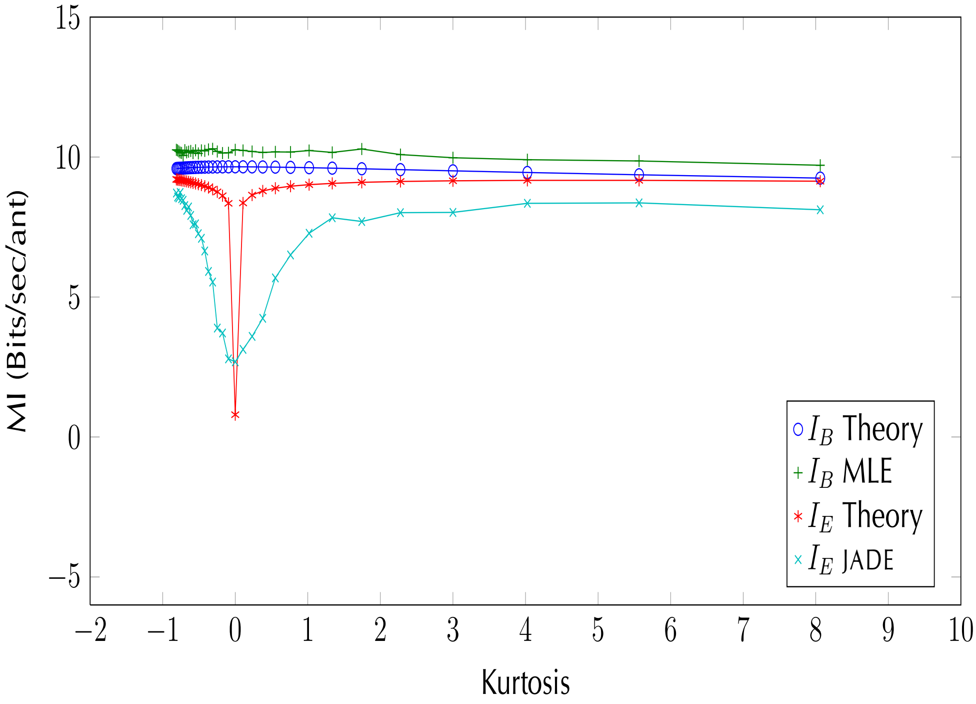

Figure 6.14: MI Vs Kurtosis, SNR = 20dB, N = 1000, channel = 8×2.

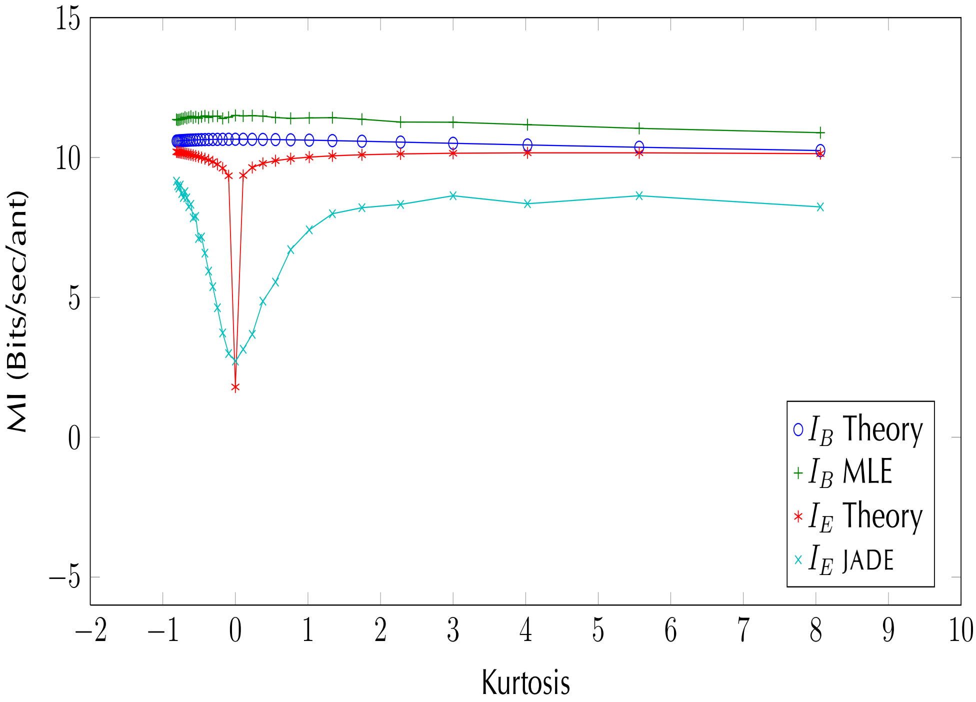

Figure 6.15: MI Vs Kurtosis, SNR = 20dB, N = 1000, channel = 16×2.

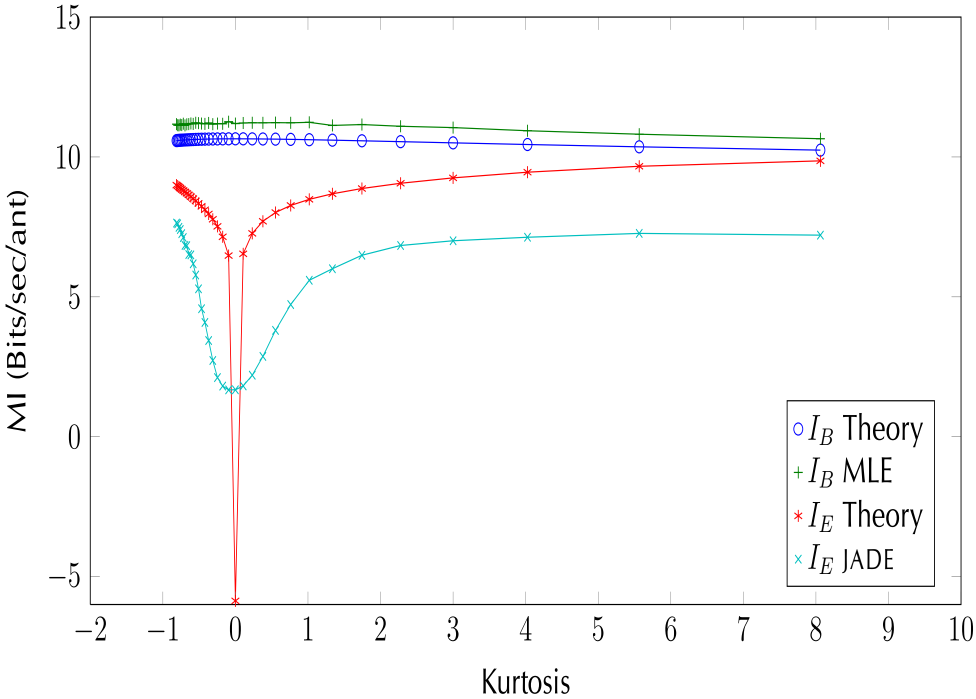

Figure 6.16: MI Vs Kurtosis, SNR = 20dB, N = 1000, channel = 16×4.

6.3 Summary

The expressions that were derived in chapter 3, for mutual information, allow us to compare the information rates achievable by a channel-informed receiver with the information rates obtainable by a passive eavesdropping receiver. The simulation exercise carried out in this chapter enabled an analysis of the effect of varying a range of system parameters: snr, blocklength, array size, source kurtosis. These results indicate that practical communication waveforms, which have high kurtosis values, may be vulnerable to blind interception and provide an eavesdropper with a useable information rate.

Once again we note that as the sources distributions approach zero kurtosis, or as they become more Gaussian, the difference between I_B and I_E increases, reaching its maximum at zero-kurtosis. This clearly indicates that maximum secrecy capacity: $I_S = I_B - I_E$, is attained using Gaussian sources and for an eavesdropper employing a HOS based method for BSS, regardless of channel dimensions and variation in the other system parameters. To maximise communications secrecy, Alice and Bob could employ the following strategies:

- Use short data block lengths. The simulations have shown that the BSS algorithms do not perform well for short observation block lengths.

- Use more transmit and receive antennae. This requires an arbitrarily rich multipath environment. There may also be physical constraints on the space available for more antennae and reducing the antenna spacing will introduce dependence between them and reduce the channel rank. Channel dependence will be caused by antenna cross-coupling and, for closely spaced antennae, the independent background noise assumption is no longer valid.

- Minimise the magnitude of the source kurtosis, i.e. manipulate the source distributions so that they become proper complex-Gaussian. Simulation results have shown that creates the worst possible problem for an eavesdropper employing HOS methods for BSS. However this will also affect other receivers so that the manipulation must occur in a manner that can be undone by the intended receiver.

How the above may be achieved is outside the scope of the current study but may be suitable for future research.

Chapter 7

Symbol Stream Recovery for OSTBC

7.1 Introduction

This chapter addresses the problem of resolving the symbol streams transmitted from the antennae of a MIMO wireless digital communications array. This scenario may be posed as a BSS problem thereby invoking the use of BSS techniques which have been under development since the 1990's. The principle assumption, in this model, is that the sources are statistically independent and that a linear mixture of the sources can therefore be separated via an optimization method that utilises a cost function which estimates the dependence between the unmixed source estimates. Independence may be determined as the MI between the resolved source estimates and is known to be attained when the joint source density function is equal to the product of the marginal source densities. To avoid confusion with the channel MI derived earlier as the entropy reduction between observed data entropy and estimated source entropy, we shall refer to the MI between the resolved source estimates as MIBS. Although this would seem to provide the best measure of independence, HOS algorithms for BSS typically rely on methods that lead to an approximation for mutual information such as kurtosis (4th order cumulants) e.g. JADE, developed by Cardoso and Souloumiac in [19], to simplify the algorithms and reduce computation times.

There are several BSS algorithms e.g. FASTICA, JADE (described in sections 4.6 and 6.2 respectively) that we could employ at this stage but we shall make use of the popular JADE algorithm [19]. The FASTICA and JADE algorithms are known to perform similarly for blind separation of the type of signal studied here but JADE was chosen due to the apparent availablity of both the cost and gradient functions that are required for incorporation in a more general optimization framework together with its ability to handle overdetermined observation data. However the complex gradient of the JADE cost function stated, without proof, by Abrudan et al. in [2] and [3] appears to be incorrect; necessitating the derivation provided in Appendix H.

Recent research into steepest descent and conjugate gradient techniques [2, 3, 4] has provided a means for readily estimating the demixing matrix, with a unitary matrix constraint. These techniques are desirable because they provide a generic optimization algorithm ideally suited to the BSS problem and which allow us to change the cost function to suit the problem or take advantage of known properties of the sources. In Appendix J we provide the Matlab code for the optimization algorithm that was used in the simulation exercises.

A method for estimating MIBS based on estimating Shannon entropy was developed by Kraskov et al. in [56] and Stögbauer et al. in [93]. To use MIBS in an optimization algorithm we also require the gradient as a function of the demixing matrix and this is derived in Appendix I.

The mathematical model and associated assumptions are the same as those described in Section 3.1 , with the exception that the source distribution is now discrete; resulting from the use of complex signalling constellations such as PSK or QAM, and the receiver is able to synchronize correctly with the observed signals.

A MIMO system employing space-time block-coding techniques is not ideally suited to BSS; the assumption of mutual source independence may be violated. However, as we shall find in section 7.5, direct application of the BSS algorithms can still provide useful results. The purpose of this chapter is to develop a method that exploits knowledge of the STBC scheme, in particular the OSTBC scheme, that a MIMO link may be using.

7.2 Space-Time Block Codes

We now consider the generation of linear space-time block-codes and how we might exploit knowledge of their properties as an aid to the source separation problem. Various authors e.g. Swindlehurst and Leus [95], Ma [65] and Choqueuse et al. [22], have given general representations for a linear Space-Time Block Encoder (STBE), where a vector of n_s symbols $\mathbf{s} = [s_1 s_2 \ldots s_{n_s}]^T$ is encoded as a space-time block for transmission over n_t parallel signal streams of length n_b. Swindlehurst and Leus [95] give a general form for the encoded symbol sequence transmitted from individual antennae and in a similar vein we shall consider how the columns of the STBC are formed. In the following derivations we assume that $n_t = n_r$, where n_t, n_r are the number of transmit antennae and the number of receiver antennae, respectively. Let $\mathbf{B}$ be an $n_t \times n_b$ STBC formed from an input data sequence of n_s symbols, or $\mathbf{B} \in \text{STBC}(n_t, n_s, n_b)$, and where the symbols are taken from a vector symbol set $\mathcal{S}$ of size n_{ss}. Assuming perfect synchronization, the k^{th} observed signal block is given by

$$\mathbf{Y}_k = \mathbf{A}\mathbf{B}_k + \mathbf{W}_k. \tag{7.1}$$

We can apply a complex-to-real conversion to equation 7.1 by stacking the real and imaginary parts as follows:

$$\begin{bmatrix} \Re(\mathbf{Y}_k) \\ \Im(\mathbf{Y}_k) \end{bmatrix} = \begin{bmatrix} \Re(\mathbf{A}) & -\Im(\mathbf{A}) \\ \Im(\mathbf{A}) & \Re(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \Re(\mathbf{B}_k) \\ \Im(\mathbf{B}_k) \end{bmatrix} + \begin{bmatrix} \Re(\mathbf{W}_k) \\ \Im(\mathbf{W}_k) \end{bmatrix}. \tag{7.2}$$

Letting $\boldsymbol{\Psi}_k = \left[\Re(\mathbf{Y}_k^T)\Im(\mathbf{Y}_k^T)\right]^T$, $\boldsymbol{\Theta}_k = \left[\Re(\mathbf{B}_k^T)\Im(\mathbf{B}_k^T)\right]^T$, $\boldsymbol{\Omega}_k = \left[\Re(\mathbf{W}_k^T)\Im(\mathbf{W}_k^T)\right]^T$ and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \Re(\mathbf{A}) & -\Im(\mathbf{A}) \\ \Im(\mathbf{A}) & \Re(\mathbf{A}) \end{bmatrix}, \tag{7.3}$$

$$\text{then we may write} \quad \boldsymbol{\Psi}_k = \boldsymbol{\Lambda}\boldsymbol{\Theta}_k + \boldsymbol{\Omega}_k. \tag{7.4}$$

Vectorising the k^{th} observed block gives

$$\boldsymbol{\psi}_k = \left[\mathbf{I}_{n_b} \otimes \boldsymbol{\Lambda}\right] \boldsymbol{\theta}_k + \boldsymbol{\omega}_k, \tag{7.5}$$

where $\boldsymbol{\psi}_k = \mathbf{vec}\,(\boldsymbol{\Psi}_k)$, $\boldsymbol{\theta}_k = \mathbf{vec}\,(\boldsymbol{\Theta}_k)$ and $\boldsymbol{\omega}_k = \mathbf{vec}\,(\boldsymbol{\Omega}_k)$.

With $\boldsymbol{\sigma} = \left[\Re(\mathbf{s}^T)\Im(\mathbf{s}^T)\right]^T$, the formation of the columns $\mathbf{c}_i$ of $\boldsymbol{\Theta}$, by the STBE, may be represented by

$$\mathbf{c}_i = \mathbf{U}_i\boldsymbol{\sigma}, \tag{7.6}$$

where the $\mathbf{U}_i$ are orthogonal permutation and sign change matrices. In this representation the block code is given by $\boldsymbol{\Theta} = [\mathbf{U}_1\boldsymbol{\sigma} \quad \mathbf{U}_2\boldsymbol{\sigma} \quad \ldots \quad \mathbf{U}_{n_b}\boldsymbol{\sigma}]$ and $\boldsymbol{\theta} = \mathbf{U}\tilde{\boldsymbol{\sigma}}$, where we have made the following definitions: $\mathbf{U} = \text{diag}(\mathbf{U}_1, \ldots, \mathbf{U}_{n_b})$, $\tilde{\boldsymbol{\sigma}} = [\boldsymbol{\sigma}\,\boldsymbol{\sigma}\,\ldots\,\boldsymbol{\sigma}]^T$. The k^{th} observed signal vector may now be written as

$$\boldsymbol{\psi}_k = [\mathbf{I}_{n_b} \otimes \boldsymbol{\Lambda}]\,\mathbf{U}\tilde{\boldsymbol{\sigma}}_k + \boldsymbol{\omega}_k. \tag{7.7}$$

If we apply the decoding operation to the observed signal vectors, assuming perfect synchronization, then this is represented by

$$\boldsymbol{\phi}_k = \mathbf{U}^T\boldsymbol{\psi}_k = \mathbf{U}^T[\mathbf{I}_{n_b} \otimes \boldsymbol{\Lambda}]\,\mathbf{U}\tilde{\boldsymbol{\sigma}}_k + \mathbf{U}^T\boldsymbol{\omega}_k. \tag{7.8}$$

If we ignore the noise term then $\boldsymbol{\phi}_k$ may be written as

$$\boldsymbol{\phi}_k = \text{diag}\left(\mathbf{U}_1^T\boldsymbol{\Lambda}\mathbf{U}_1\boldsymbol{\sigma}_k, \ldots, \mathbf{U}_{n_b}^T\boldsymbol{\Lambda}\mathbf{U}_{n_b}\boldsymbol{\sigma}_k\right) \tag{7.9}$$

and we make the observation that summing across the rows of $\boldsymbol{\Phi}_k$, where $\boldsymbol{\phi}_k = \mathbf{vec}\,(\boldsymbol{\Phi}_k)$, is equivalent to forming the sum

$$\boldsymbol{\Sigma}_k = \sum_i^{n_b} \mathbf{U}_i^T\boldsymbol{\Lambda}\mathbf{U}_i\boldsymbol{\sigma}_k = \mathcal{H}\boldsymbol{\sigma}_k, \tag{7.10}$$

$$\text{where} \quad \mathcal{H} = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}. \tag{7.11}$$

$\mathbf{H}$ is the new effective channel for the transmitted symbol vector $\mathbf{s}_k$ i.e. we have converted the observed signal block $\mathbf{Y}_k = \mathbf{A}\mathbf{B}_k$ into the vector $\mathbf{g}_k = \mathbf{H}\mathbf{s}_k$, where $\boldsymbol{\Sigma}_k = \left[\Re(\mathbf{g}_k^T)\Im(\mathbf{g}_k^T)\right]^T$. For the orthogonal space-time block codes that we consider here we find that the coding has the effect of making $\mathbf{H}$ unitary. This shows that, given an appropriate array configuration, the observation matrix $\mathbf{Y}$ resulting from an OSTBC and channel $\mathbf{A}$ may be reconstructed as the product of a unitary matrix $\mathbf{H}$, which is now effectively the channel matrix, and the transmitted symbol vectors $\mathbf{s}_k$. Reforming the observed data in this fashion has a two-fold benefit in terms of BSS: the effective channel matrix is already unitary which simplifies the optimization process and the symbol stream is now independent i.e. the redundancy in the code has been removed. To illustrate the above derivations we shall now consider two examples of OSTBC schemes.

7.2.1 OSTBC(2,2,2)

A single block using the Alamouti code, developed by Alamouti in [8], which is an OSTBC$(2, 2, 2)$ code, is given by

$$\mathbf{B}_A = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix} \tag{7.12}$$

and uses

$$\mathbf{U}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{U}_2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{7.13}$$

and $\mathbf{s} = [s_1 s_2]^T$. Suppose we have a receive array with two antennae (the Alamouti STC is designed for a 2×1 link) then, when this block is transmitted via a channel represented by the complex matrix $\mathbf{A}$ the observed (noiseless) matrix $\mathbf{Y}$ is

$$\mathbf{Y} = \mathbf{A}\mathbf{B}_A = \begin{bmatrix} (a_{11}s_1 + a_{12}s_2) & (-a_{11}s_2^* + a_{12}s_1^*) \\ (a_{21}s_1 + a_{22}s_2) & (a_{22}s_1^* - a_{21}s_2^*) \end{bmatrix}. \tag{7.14}$$

Now if we decode $\mathbf{Y}$ to obtain the matrix $\boldsymbol{\psi}$, sum to get $\boldsymbol{\Sigma}$,

$$\text{then} \quad \mathbf{g} = \begin{bmatrix} (a_{11} + a_{22}^*)s_1 + (a_{12} - a_{21}^*)s_2 \\ (a_{21} - a_{12}^*)s_1 + (a_{11}^* + a_{22})s_2 \end{bmatrix} \tag{7.15}$$

$$\text{or} \quad \mathbf{g} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \mathbf{H}\mathbf{s}. \tag{7.16}$$

7.2.2 OSTBC(4,4,8)

Another example, taken from Tarokh et al [96], is the $\frac{1}{2}$-rate Orthogonal Space-Time Block-Code OSTBC$(4, 4, 8)$ designed for 4 transmit antennae:

$$\mathbf{B}_T = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 & s_1^* & -s_2^* & -s_3^* & -s_4^* \\ s_2 & s_1 & s_4 & -s_3 & s_2^* & s_1^* & s_4^* & -s_3^* \\ s_3 & -s_4 & s_1 & s_2 & s_3^* & -s_4^* & s_1^* & s_2^* \\ s_4 & s_3 & -s_2 & s_1 & s_4^* & s_3^* & -s_2^* & s_1^* \end{bmatrix}. \tag{7.17}$$

Let

$$\begin{aligned}\mathbf{A}_1 &= \mathbf{I}_4, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{A}_3 &= \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \mathbf{A}_4 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},\end{aligned} \tag{7.18}$$

$$\text{so that} \quad \mathbf{U}_i = \begin{cases} \begin{bmatrix} \mathbf{A}_i & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{A}_i \end{bmatrix}, & \text{i = 1...4;} \\ \\ \begin{bmatrix} \mathbf{A}_{i-4} & \mathbf{0}_4 \\ \mathbf{0}_4 & -\mathbf{A}_{i-4} \end{bmatrix}, & \text{i = 5...8.} \end{cases} \tag{7.19}$$

and the columns of $\boldsymbol{\Theta}$ are formed using equation (7.6). Proceeding as before to decode blocks of observed data, we again obtain the representation $\mathbf{g} = \mathbf{H}\mathbf{s}$, where $\mathbf{H}$ is a 4×4 orthogonal matrix and $\mathbf{s} = [s_1 s_2 s_3 s_4]^T$. In this case 4×8 blocks of observed data are reconstructed as 4×1 vectors $\mathbf{g}$ and the effective channel is a real orthogonal matrix. From the BSS perspective this leads to a reduction of $8 : 1$ in the length of the observed data which could lead to difficulties in separating very small data lengths. However, as previously noted, the BSS algorithms benefit from the effective channel matrix being already unitary and the effective symbol streams are now independent, assuming the original message symbol streams input to the STBE were independent.

For this technique to be applied to source separation, synchronization and knowledge of block starting times are clearly essential.

7.3 OSTBC Complex Representation

An alternative and perhaps more elegant representation for OSTBC, which avoids the complex-to-real conversion in the previous section, is described here. Let the augmented input symbol vector be $\tilde{\boldsymbol{\sigma}} = \left[\mathbf{s}^T \mathbf{s}^\dagger\right]^T$, then vectorising the code block B_k may be written as

$$\mathbf{b}_k = \mathbf{U}\tilde{\boldsymbol{\sigma}}_k, \tag{7.20}$$

where $\mathbf{U}$ is a scaled, orthogonal permutation and sign change matrix, i.e. $\mathbf{U}^T\mathbf{U} = c\mathbf{I}$, for some constant c. Letting $\boldsymbol{\psi}_k$ be the vectorised k^{th} observed block, then

$$\boldsymbol{\psi}_k = \left[\mathbf{I}_{n_b} \otimes \mathbf{A}\right] \mathbf{U}\tilde{\boldsymbol{\sigma}}_k + \boldsymbol{\omega}_k. \tag{7.21}$$

If we apply the decoding operation then

$$\boldsymbol{\phi}_k = \mathbf{U}^T\boldsymbol{\psi}_k = \mathbf{U}^T \left[\mathbf{I}_{n_b} \otimes \mathbf{A}\right] \mathbf{U}\tilde{\boldsymbol{\sigma}}_k + \mathbf{U}^T\boldsymbol{\omega}_k. \tag{7.22}$$

If we ignore the noise term then $\boldsymbol{\phi}_k$ may be written as

$$\boldsymbol{\phi}_k = \mathbf{M}\tilde{\boldsymbol{\sigma}}_k, \tag{7.23}$$

where the matrix M takes the form

$$\mathbf{M} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}. \tag{7.24}$$

Let $\tilde{\boldsymbol{\phi}}_k$ represent the matrix formed from the vector $\boldsymbol{\phi}_k$ when the lower half of $\boldsymbol{\phi}_k$ is conjugated, i.e.

$$\tilde{\boldsymbol{\phi}}_k = \begin{bmatrix} \phi_{k_1} \\ \phi_{k_2} \\ \vdots \\ \phi_{k_{ns}} \\ \phi_{k_1}^* \\ \phi_{k_2}^* \\ \vdots \\ \phi_{k_{ns}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12}^* \\ \mathbf{M}_{21} & \mathbf{M}_{22}^* \end{bmatrix} \begin{bmatrix} \mathbf{s}_k \\ \mathbf{s}_k \end{bmatrix}. \tag{7.25}$$

Finally we obtain the vector $\mathbf{g}_k$ as follows

$$\begin{aligned}
\mathbf{g}_k &= [\mathbf{I}_{ns}\,\mathbf{I}_{ns}]\,\tilde{\boldsymbol{\phi}}_k \\
&= [\mathbf{M}_{11} + \mathbf{M}_{12}^* + \mathbf{M}_{21} + \mathbf{M}_{22}^*]\,\mathbf{s} \\
&= \mathbf{H}\mathbf{s}, \qquad (7.26)
\end{aligned}$$

where $\mathbf{H}$ is the effective channel matrix for the transmitted symbol vector $\mathbf{s}_k$.

To illustrate the above derivations we shall now consider some examples of OSTBC schemes.

7.3.1 OSTBC(2,2,2)

A single block using the Alamouti code [8], which is an OSTBC$(2, 2, 2)$ code, is given by equation 7.12 and uses

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad (7.27)$$

and $\tilde{\boldsymbol{\sigma}} = [s_1\, s_2\, s_1^*\, s_2^*]^T$. Suppose we have a receive array with two antennae then, when this block is transmitted via a channel represented by the complex matrix $\mathbf{A}$ the observed (noiseless) matrix $\mathbf{Y}$ is

$$\mathbf{Y} = \mathbf{A}\mathbf{B}_A = \begin{bmatrix} (a_{11}s_1 + a_{12}s_2) & (-a_{11}s_2^* + a_{12}s_1^*) \\ (a_{21}s_1 + a_{22}s_2) & (a_{22}s_1^* - a_{21}s_2^*) \end{bmatrix}. \qquad (7.28)$$

Now if we decode $\mathbf{Y}$ to obtain the matrix $\boldsymbol{\psi}$, sum to get $\boldsymbol{\Sigma}$, then

$$\mathbf{g} = \begin{bmatrix} (a_{11} + a_{22}^*)s_1 + (a_{12} - a_{21}^*)s_2 \\ (a_{21} - a_{12}^*)s_1 + (a_{11}^* + a_{22})s_2 \end{bmatrix} \qquad (7.29)$$

or

$$\mathbf{g} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \mathbf{H}\mathbf{s}. \qquad (7.30)$$

7.3.2 OSTBC(3,4,8)

Another example, taken from Tarokh et al [96], is the $\frac{1}{2}$-rate OSTBC$(3,4,8)$ designed for 3 transmit antennae:

$$\mathbf{B}_T = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 & s_1^* & -s_2^* & -s_3^* & -s_4^* \\ s_2 & s_1 & s_4 & -s_3 & s_2^* & s_1^* & s_4^* & -s_3^* \\ s_3 & -s_4 & s_1 & s_2 & s_3^* & -s_4^* & s_1^* & s_2^* \end{bmatrix}. \tag{7.31}$$

$$\mathbf{U}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{7.32}$$

$$\mathbf{U}_2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \tag{7.33}$$

$$\mathbf{U}_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \tag{7.34}$$

$$\mathbf{U}_4 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{7.35}$$

so that

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & 0 \\ \mathbf{U}_2 & 0 \\ \mathbf{U}_3 & 0 \\ \mathbf{U}_4 & 0 \\ 0 & \mathbf{U}_1 \\ 0 & \mathbf{U}_2 \\ 0 & \mathbf{U}_3 \\ 0 & \mathbf{U}_4 \end{bmatrix}, \tag{7.36}$$

which leads to

$$\tilde{\phi}_k = \begin{bmatrix} \mathbf{M}_{11} & 0 \\ 0 & \mathbf{M}_{11}^* \end{bmatrix} \begin{bmatrix} \mathbf{s}_k \\ \mathbf{s}_k \end{bmatrix} \tag{7.37}$$

and

$$\mathbf{g}_k = 2\Re\{\mathbf{M}_{11}\}\mathbf{s}_k. \tag{7.38}$$

The effective channel, in this case, is almost orthogonal.

7.3.3 OSTBC(4,4,8)

Another example, taken from Tarokh et al [96], is the $\frac{1}{2}$-rate OSTBC$(4,4,8)$ designed for 4 transmit antennae as shown in equation 7.17. Let $\mathbf{U}_1 = \mathbf{I}_4$,

$$\mathbf{U}_2 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{7.39}$$

$$\mathbf{U}_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \tag{7.40}$$

$$\mathbf{U}_4 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \tag{7.41}$$

so that

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & 0 \\ \mathbf{U}_2 & 0 \\ \mathbf{U}_3 & 0 \\ \mathbf{U}_4 & 0 \\ 0 & \mathbf{U}_1 \\ 0 & \mathbf{U}_2 \\ 0 & \mathbf{U}_3 \\ 0 & \mathbf{U}_4 \end{bmatrix}, \tag{7.42}$$

which leads to

$$\tilde{\phi}_k = \begin{bmatrix} \mathbf{M}_{11} & 0 \\ 0 & \mathbf{M}_{11}^* \end{bmatrix} \begin{bmatrix} \mathbf{s}_k \\ \mathbf{s}_k \end{bmatrix} \tag{7.43}$$

and

$$\mathbf{g}_k = 2\Re\{\mathbf{M}_{11}\}\mathbf{s}_k. \tag{7.44}$$

7.4 Demixing Matrix Estimation

Abrudan, Eriksson and Koivunen [2, 3, 4] have recently developed efficient algorithms for optimization under unitary matrix constraint. These algorithms provide us with a means for implementing a variety of cost functions in order to obtain demixing matrix estimates. Given a real cost function which is a function of a complex parameter (the demixing matrix in this case) we also require the associated complex gradient. We derive the gradient for the JADE cost function in Appendix H and implement this in a conjugate gradient optimization algorithm based on the algorithm described in [2]. In section 7.5 we refer to the conjugate gradient algorithm with the JADE cost function as simply the JADE algorithm. Matlab code for the conjugate gradient optimization algorithm, with the JADE cost and gradient functions, is provided in Appendix J.

Next we consider the MIBS cost and derive the complex gradient as a function of the demixing matrix. The derivation of the MIBS gradient is provided in Appendix I. In section 7.5 we refer to the conjugate gradient algorithm with the MIBS cost function as simply the MIBS algorithm. The MIBS algorithm also requires estimates of the source score function and we have employed a method similar to that described by Vlassis and Motomura in [105] that uses data histogramming and Gaussian Mixture (GM) modelling to estimate both the source pdf $p(x)$ and the source score function $-\frac{\partial \log p(x)}{\partial x}$. We have found however that similar results may be achieved by smoothing the data histogram and simply taking the derivative of its logarithm. Matlab code for the conjugate gradient optimization algorithm using the MIBS cost and gradient functions is provided in Appendix K.

To find the unitary demixing matrix for the BSS problem we have chosen to utilise the Conjugate Gradient (CG) optimization algorithm described by Abrudan et al. in [4, Table 3], where step 4 is implemented using steps 5-7 from Abrudan et al. in [3, Table 2].

The main benefit of this approach is in the flexibility that it allows: the cost function costf(W,Y) and complex gradient gradf(W,Y) may be for JADE, MIBS or any other suitable cost and gradient pair.

7.4.1 JADE Cost and Gradient

Defining $\mathbf{\Phi} \triangleq \mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W}$, where $\mathbf{W}$ is unitary, then the JADE algorithm minimises the sum of the squared magnitudes of the off-diagonal elements of $\mathbf{\Phi}$. Alternatively JADE minimises the following cost function [3]

$$C_{JADE} = \sum_{i=1}^{m} tr\left\{\mathbf{\Phi}\mathbf{\Phi}^{\dagger} - \mathbf{\Phi} \odot \mathbf{\Phi}^{\dagger}\right\}. \tag{7.45}$$

The eigenmatrices $\hat{\mathbf{M}}_i$ are estimated from the fourth order cumulants of the whitened observations and these are described by Cardoso and Souloumiac in [19]. The cost function is therefore used to diagonalize the eigenmatrices, with respect to $\mathbf{W}^*$.

The Euclidean gradient of the JADE cost function, w.r.t. $\mathbf{W}^*$, obtained from

$$\mathbf{G}_{JADE} = 2\frac{\partial C_{JADE}}{\partial \mathbf{W}^*}, \tag{7.46}$$

is derived in Appendix H, where it is shown that

$$\mathbf{G}_{JADE} = 2\sum_{i}^{m}\{\hat{\mathbf{M}}_i\hat{\mathbf{M}}_i^{\dagger}\mathbf{W} - \hat{\mathbf{M}}_i\mathbf{W}[\mathbf{I}\odot\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i^{\dagger}\mathbf{W}] - \hat{\mathbf{M}}_i^{\dagger}\mathbf{W}[\mathbf{I}\odot\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W}]\}. \tag{7.47}$$

In [3] it is stated, without proof, that the gradient of the JADE cost function is

$$\begin{aligned} \Gamma_{\mathbf{W}} &= 2\sum_{i}^{m}\hat{\mathbf{M}}_i\mathbf{W}\left[\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W} - \mathbf{I}\odot\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W}\right] \\ &= 2\sum_{i}^{m}\hat{\mathbf{M}}_i\hat{\mathbf{M}}_i\mathbf{W} - \hat{\mathbf{M}}_i\mathbf{W}\left[\mathbf{I}\odot\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W}\right]. \end{aligned} \tag{7.48}$$

Clearly $\Gamma_{\mathbf{W}} \neq \mathbf{G}_{JADE}$. The consequences of using the incorrect gradient are demonstrated by the simulation results shown in Figure 7.7. In this example a QAM source symbol set and Alamouti coding was implemented, the observation block size is 1000, the transmitter and receiver array size is 2 and the snr is 10dB. Results using the incorrect gradient are labelled JADE$_{inc}$, results using the correct gradient are labelled JADE$_{cor}$ and, for comparison, maximum likelihood source estimation (channel assumed known) results are labelled ML. The results show that frequent large errors occur when using the incorrect gradient expression; indicating the failure of the optimization algorithm to find a suitable demixing matrix. When the correct gradient expression was applied no such failures occurred in the simulation examples.

7.4.2 Mutual Information Cost and Gradient

For an m-dimensional random vector $\mathbf{u}$, the MIBS of its components is defined as

$$I(\mathbf{u}) \triangleq \mathbb{E}\left\{\log \frac{p(\mathbf{u})}{\prod_{i=1}^{m} p(u_i)}\right\}. \tag{7.49}$$

The MIBS can also be written in terms of entropy

$$I(\mathbf{u}) = \sum_{i=1}^{m} H(u_i) - H(\mathbf{u}), \tag{7.50}$$

where $H(u) = -\mathbb{E}\left\{\log p(u)\right\}$. The source separation problem may be solved by minimising the MIBS, acting as the cost function in an optimization algorithm. In Appendix I we show that the gradient, w.r.t. $\mathbf{W}^*$, of this cost function is

$$\frac{\partial I(\mathbf{u})}{\partial \mathbf{W}^*} = \left[\boldsymbol{\psi}(\mathbf{u}, \mathbf{u}^*)\mathbf{u}^\dagger - \mathbf{I}\right]\mathbf{W}. \tag{7.51}$$

7.5 Simulation Results

Application of the theory developed in this chapter is demonstrated by way of Monte Carlo simulations implemented in Matlab. The following subsections compare source separation performance in two ways: quality of estimated constellation and equivalent channel capacity as a function of snr. The decoding process mentioned here refers to equations 7.8 and 7.22, either of which may be applied.

7.5.1 Example Constellation Results

A 4×4 MIMO passive intercept scenario is envisaged where the complex OS-TBC$(4, 4, 8)$ coding scheme described in Section 7.2 forms the signal of interest. The symbol set used is the QAM constellation, with four complex-valued symbols.

A number of blocks (500) were generated, forming the 4×4000 complex-valued matrix $\mathbf{X}$, linearly mixed by a randomly generated 4×4 complex-valued channel matrix $\mathbf{A}$ and the observation matrix formed as $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$, where $\mathbf{W}$ is a 4×4000 AWGN noise matrix. The snr power ratio for each row of $\mathbf{Y}$, i.e. each receive antenna input, is 10dB.

Figure 7.1 shows the effect of channel mixing, for an OSTBC, applied to the four transmitter QAM sources. The figure shows each of the four receive antenna inputs which are linear mixtures of the four transmitter outputs. The channel mixing has obscured the original sources.

Direct application of the JADE algorithm to the mixed data, without pre-processing with the decoding procedure described previously, yields separated source results such as those shown in Figure 7.2. This algorithm can usually obtain a channel estimate, via optimized unitary matrix estimation, that is close to the true channel. However some rotational ambiguities clearly remain. There is also a scale ambiguity but the observed data is normalised so that it has unit power and all the sources are assumed to have the same power.

Figure 7.3 shows results typically obtained by the MIBS algorithm, without any prior decoding. However, for the MIBS algorithm to successfully converge to the global minimum, it was found necessary to preprocess the observed data with the JADE algorithm. So this is effectively a hybrid algorithm. The figure shows that the MIBS algorithm was able to obtain a slightly better estimate of the mixing matrix by successfully derotating the sources in this particular example; this was not observed to occur in every simulation case. The MIBS algorithm has the advantage of knowing the source pdfs and is therefore better able to minimise the mutual information between the separated sources. Implementing the MIBS method proved to be computationally demanding, particularly in the calculation of mutual information at each iteration of the optimization process.

In Figure 7.4 the decoding procedure has been applied directly to the mixed data to demonstrate that it provides no obvious benefit to do so.

Figure 7.5 provides an example of applying the decoding procedure to the mixed data then applying JADE to estimate the sources. It is clear that the symbol constellations are better aligned with the plot axes and are slightly less noisy than those shown in Figure 7.2.

In Figure 7.6 decoding followed by the JADE/MIBS combination has been applied. The results do not appear to be significantly different to those obtained in Figure 7.5 so it becomes questionable as to whether the small gains offered by the MIBS stage are worth the high computional overhead.

7.5.2 Equivalent Channel Capacity

To study performance as a function of snr we utilise Slimane's symmetric capacity [89], also described in Appendix F. In the simulations that follow the source estimation error for: MLE, JADE without decoding, JADE with decoding, is converted to a received snr and then fed, along with the symbol constellation points, to Slimane's symmetric capacity estimation calculation. The resulting capacity estimates are compared with the estimate based on the original snr.

In the first example a QAM source symbol set and Alamouti coding was implemented. The observation block size was 1000, the transmitter and receiver array size was 2. The results in Figure 7.8 show that the capacity estimate for MLE closely follows the ideal symmetric capacity curve. The capacity estimates for both JADE with and without decoding are very similar. In the region 0 to 10dB the JADE capacities are a little less than the MLE values. At high snrs all the capacity estimates converge to the maximum value of 2 Bits/sec/antenna.

In the second example a QAM source symbol set and OSTBC3 coding was implemented. The observation block size was 1000, the transmitter and receiver array size was 3. Once again the results in Figure 7.9 show that the capacity estimate for MLE closely follows the ideal symmetric capacity curve. The results for JADE with and without decoding are very similar but the decoding appears to have provided a small improvement. The JADE results fall short of the ideal values, except at high snr, where they converge.

A third simulation example represented a system using a QAM source symbol set and OSTBC4 coding. The observation block size was 1000, the transmitter and receiver array size was 4. In Figure 7.10 the ideal and MLE results closely match. The JADE results are poorer than the ideal estimates and the decoding appears to have had a more significant benefit than that demonstrated in Figure 7.9.

7.6 Summary

We have developed an optimization algorithm for source separation where the cost function can be easily changed to suit the problem. This has been demonstrated by the implementation of two different cost functions: JADE and MIBS. STBC signals are not ideally suited for BSS techniques because of the dependence introduced by the coding process. Despite this the JADE algorithm produces reasonable results which may be corrected by post processing with the MIBS algorithm. The MIBS based method makes use of prior knowledge of the source pdfs and can minimise the mutual information, between source estimates, better than JADE. However MIBS is less able to find a global minimum and is computationally demanding.

Using knowledge of the OSTBC encoding scheme an eavesdropper can decode the observed data to improve the performance of their BSS algorithm and, in turn, increase the mutual information I_E. This means that the channel secrecy capacity $I_S = I_B - I_E$ is effectively reduced. Simulation results showing equivalent channel capacity demonstrate the potential gain from exploiting knowledge of the OSTBC encoding scheme and also show the relative capacities I_B and I_E.

The approach described in this chapter simplifies the source separation problem since the effective linear mixing matrix is unitary and the effective symbol streams are independent, assuming the symbol sequence input to the STBE is independent to begin with. A source separation algorithm such as JADE provides a means for blind channel estimation when there is no prior knowledge of either the channel or the sources and the sources are not Gaussian distributed. If the source distributions are known then the MIBS method may be employed to further improve the quality of the separation but comes with an increased computation cost.

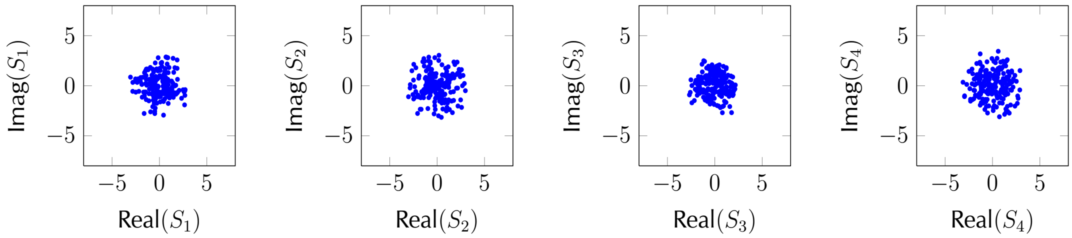

Figure 7.1: 4×4 Tx/Rx simulation using STBC with a QAM symbol set. Four sources are linearly mixed by the channel matrix. No processing has been applied. Each graph shows one of the receiver input streams.

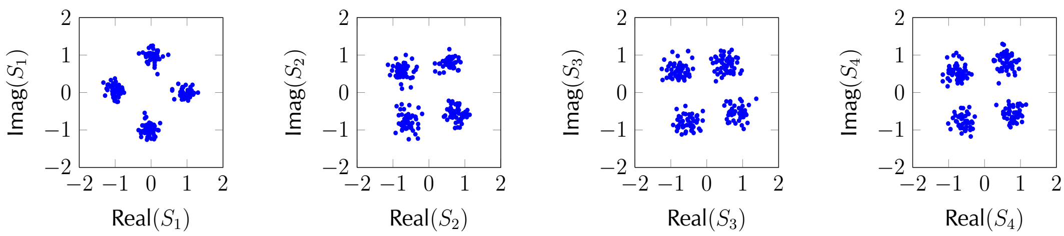

Figure 7.2: 4×4 Tx/Rx simulation using STBC with a QAM symbol set. Four sources are linearly mixed by the channel matrix. JADE has been applied to estimate the sources. Each graph shows one of the estimated source symbol streams.

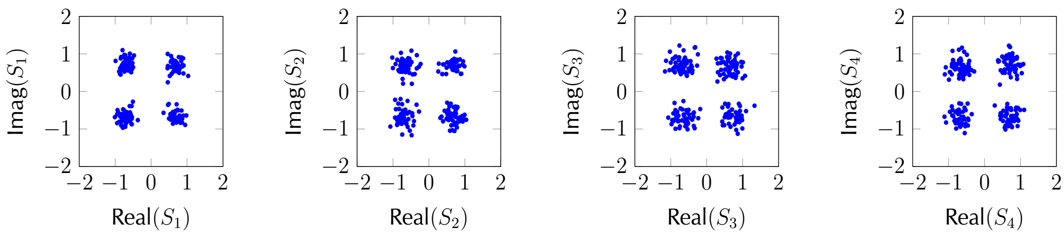

Figure 7.3: 4×4 Tx/Rx simulation using STBC with a QAM symbol set. Four sources are linearly mixed by the channel matrix. JADE/MIBS has been applied to estimate the sources. Each graph shows one of the estimated source symbol streams.

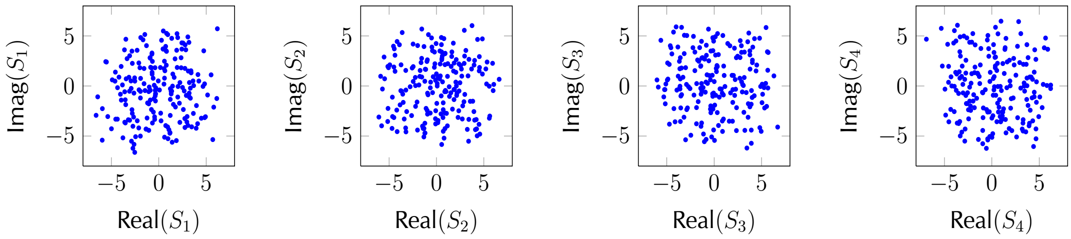

Figure 7.4: 4×4 Tx/Rx simulation using STBC with a QAM symbol set. Four sources are linearly mixed by the channel matrix. A decoding process has been applied. Each graph shows one of the estimated source symbol streams.

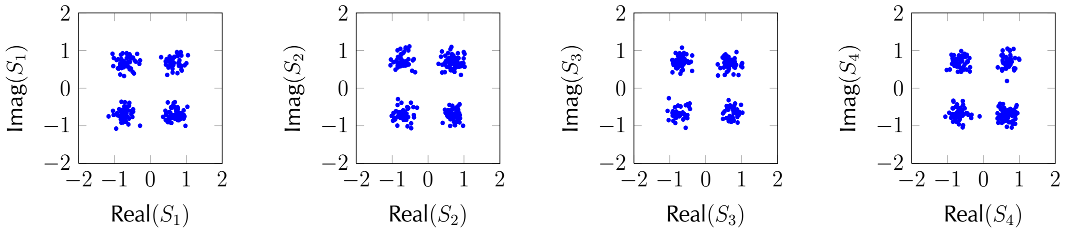

Figure 7.5: 4×4 Tx/Rx simulation using STBC with a QAM symbol set. Four sources are linearly mixed by the channel matrix. Decoding has been applied, followed by JADE to estimate the sources. Each graph shows one of the estimated source symbol streams.

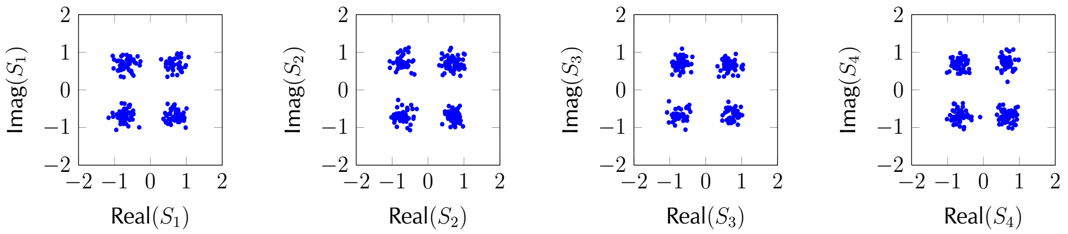

Figure 7.6: 4×4 Tx/Rx simulation using STBC with a QAM symbol set. Four sources are linearly mixed by the channel matrix. Decoding has been applied, followed by JADE/MIBS to estimate the sources. Each graph shows one of the estimated source symbol streams.

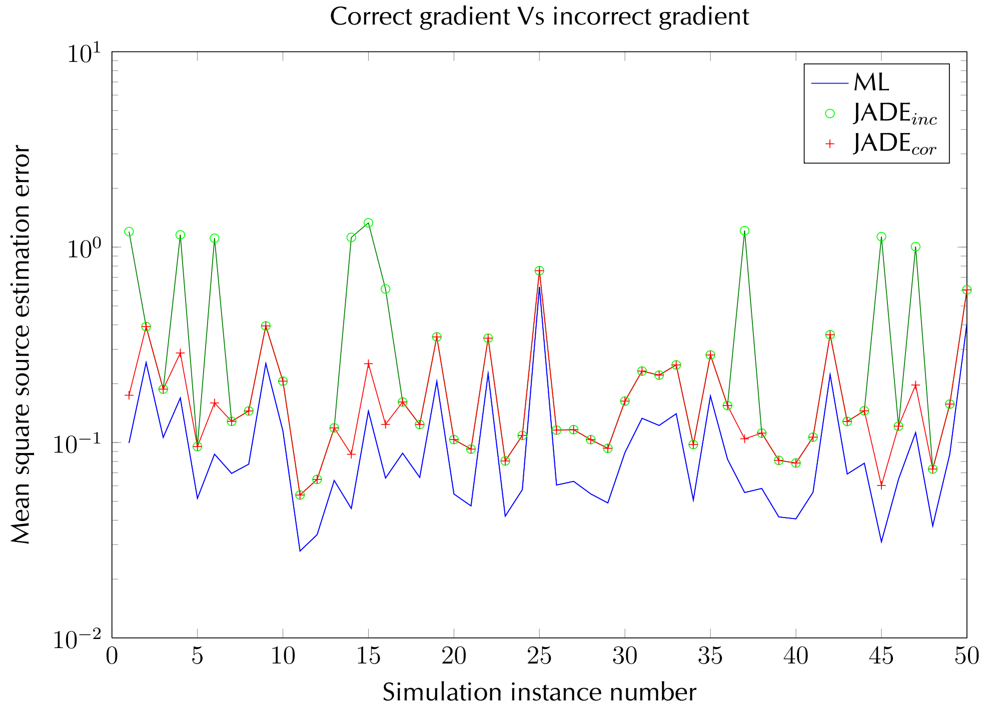

Figure 7.7: JADE source estimation error. Comparing use of incorrect gradient expression (JADE$_{inc}$) with correct gradient expression (JADE$_{cor}$).

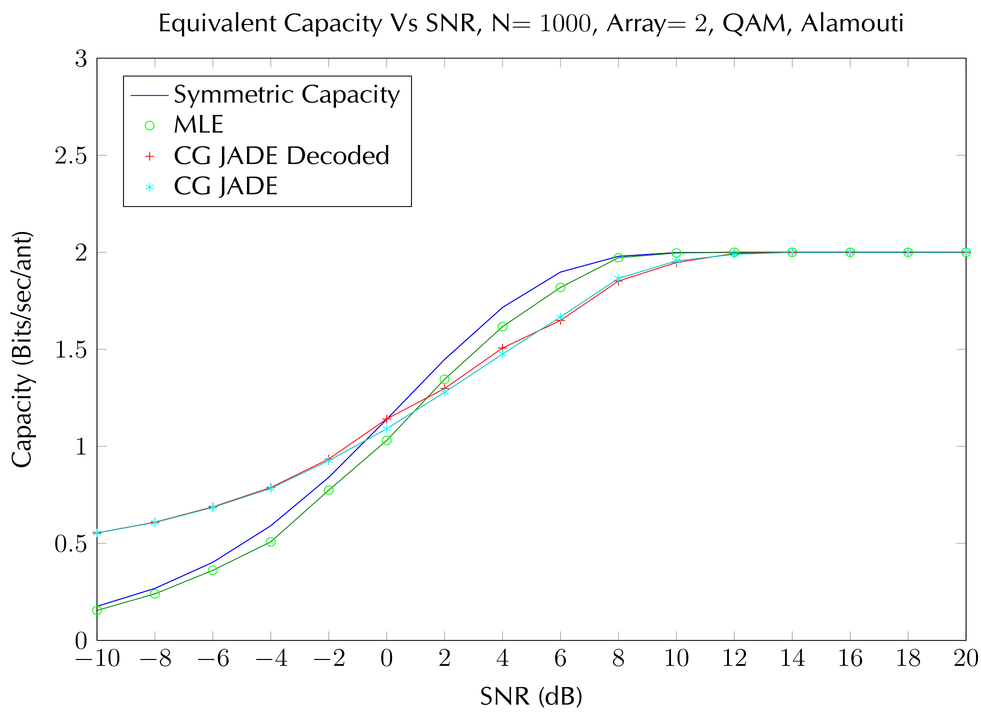

Figure 7.8: Symmetric Capacity using Decode/JADE for QAM/Alamouti.

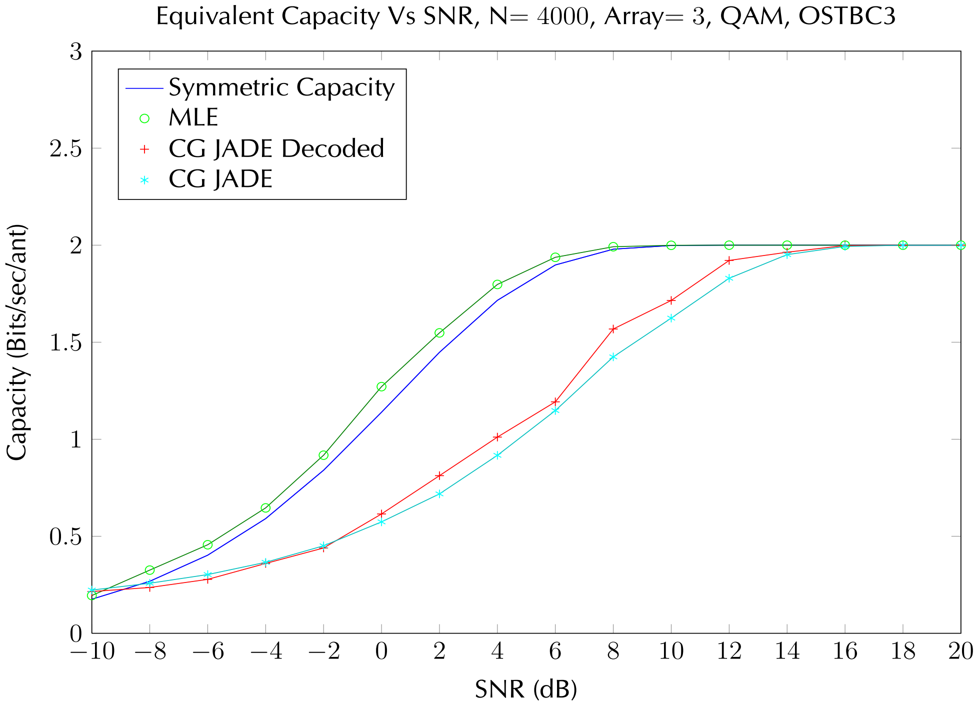

Figure 7.9: Symmetric Capacity using Decode/JADE for QAM/OSTBC3.

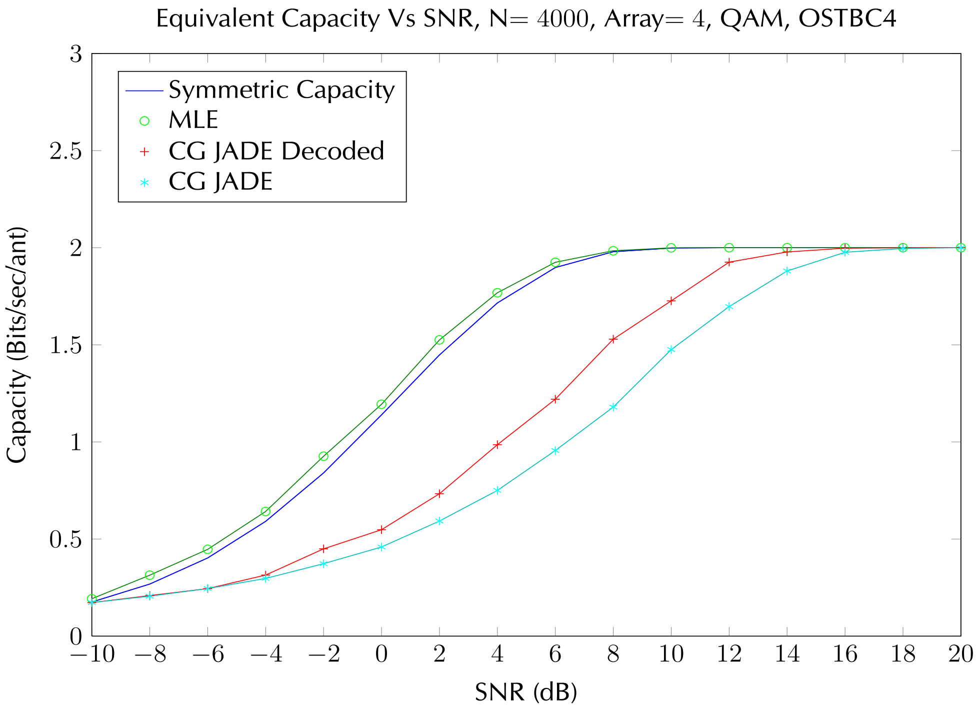

Figure 7.10: Symmetric Capacity using Decode/JADE for QAM/OSTBC4.

Part IV

Epilogue

Chapter 8

Conclusions

This project set out to advance the field of communications surveillance theory and techniques. A particular focus was established for the problem of MIMO wireless communications eavesdropping, with a view to determining the information rates that might be achievable by a passive eavesdropping receiver. A number of original contributions, in the form of mathematical tools and techniques that enable the study and analysis of the eavesdropping problem, have been presented and several of these have resulted in the publications listed in the front matter of this thesis.

The thesis has been presented in a number of parts. In Part I *Preliminaries*, preliminary material was presented to provide a context for the remainder of the thesis, to highlight relevant previous literature and identify areas of deficiency that are addressed in subsequent chapters. Chapter 1 *Introduction* and Chapter 2 *Literature Review* together form Part I.

Part II *Theory and Techniques*, comprising Chapters 3 *Information Theory for Eavesdroppers*, 4 *Source and Channel Estimation* and 5 *Copula Techniques for Modelling Channel Dependence*, concerned the development of the theory required for undestanding and analysing MIMO wireless communications eavesdropping information rates.

In Chapter 3 expressions for MI are derived, with some simplifying assumptions, so that a comparison between the MI available to an intended receiver may be compared to the MI available to an unintended receiver or eavesdropper who, it is generally assumed, has less prior information available. The case where there has been an unknown unitary precoding, such as for SVD processing, or through the use of a BSS technique, such as ICA, was also considered in Chapter 3. An alternative model for analysing the unknown unitary transformation problem was described in Chapter 3. This model employed the concept of a hypershere and the resulting expressions provided some insight into the relationship between the partially informed (eavesdropper) receiver and the fully informed (intended) receiver, for a general array dimension. Chapter 4 considered the set of fundamental states of knowledge for an eavesdropping system and derived MLE expressions for source and channel estimation and performance bounds in the form of a Cramér-Rao Lower Variance Bound (CRLVB) for parameter estimation. A CRLVB for joint complex-valued source and complex-valued channel estimation, using BSS techniques and the GG pdf, was also derived in Chapter 4. The results of Part II are used in subsequent chapters to provide theoretical performance curves for comparison with Monte Carlo simulation results.

Copula techniques were introduced in Chapter 5 as a technique for modelling source dependence introduced by the propagation channel or through cross-coupling inherent in the transmit or receive antenna arrays. A method was described for modelling a complex-valued source and channel MIMO link where the sources could experience a range of different fading distributions e.g. Nakagami fading and the structure of the dependence could be modelled as a chosen multivariate pdf e.g. multivariate Normal. Monte Carlo simulations were designed and performed to analyse the performance of BSS, representing an eavesdropping receiver, as the channel dependence is increased. As expected, the information rate obtained was found to reduce as the channel correlation increased.

In Part III *Discrete Source Recovery*, the theory developed in Part II was put to use in analysing the performance of a receiving system intercepting MIMO wireless digital communication signals. Whereas in previous chapters the sources were modelled using the GG distribution, this chapter considered the more realistic scenario where the sources were streams of symbols taken from discrete constellation sets such as those used for PSK and QAM modulations. For the purpose of comparison and analysis Monte Carlo simulations were performed covering a range of snrs and source data block lengths as a function of source kurtosis.

In chapter 7 a specific type of digital communications signal was studied i.e. the OSTBC signal, a coding scheme commonly described in the literature, for application in MIMO communication systems. It was shown how knowledge of the structure of this signal could be exploited so that a properly configured receiving system could process the observed data to provide a suitable input to a BSS algorithm and hence improve the eavesdropper's source recovery performance.

In short this thesis has combined theory and techniques to provide a toolbox for analysing the MIMO wireless communications eavesdrop problem.

Chapter 9

Further Work

Throughout the work for this thesis some ideas were considered, explored and then either rejected as being too difficult, required too much time to develop or perhaps unlikely to contribute directly to the direction that the thesis seemed to be taking at the time. Here we briefly summarise those problems or ideas considered worthy of further analysis and proposals for potentially interesting future research.

- Information geometry has recently become a topic of some interest for signal processing applications. A particularly relevant example is described by Zheng and Tse in [120], where a geometric interpretation for multi-antenna channel capacities was described. If time had permitted then this model would have been explored further as an alternative method for deriving eavesdropper information rates.

- Copula theory was introduced in Chapter 5 primarily as a means for modelling channel dependence. It was also briefly considered as an alternative to the BSS algorithms that have been studied. However some drawbacks were discovered in implementing this approach: the independence tests were inefficient and no more successful than second order tests e.g. correlation. A study using sources that have some temporal structure or dependence is required to establish if the copula based approach is more appropriate for source signals of this type.

- Some consideration of the design of Gaussian sources with hidden structure could lead to waveforms or distributions that provide enhanced secrecy.

- Complex variants of the copula families do not appear to be available. Copulas for complex-valued data could prove to be useful in a range of applications.
- The present study has focussed on the application of HOS based approaches for solving the BSS problem. However these approaches are unable to separate mixtures of Gaussian sources. An investigation into the application of SOS based methods is warranted and, most importantly, could provide a means for separating Gaussian sources.
- In Chapter 7 a method for preprocessing the observed signals was developed for the specific case of OSTBC signals. While this is considered to be the type of coding most amenable to such a technique, some benefit might be gained by examining other STC schemes.
- The implications for an eavesdropper when the intended users employ secrecy techniques in a MIMO wireless communications link, have not been addressed here. Examples of such secrecy techniques have been described by Negi & Goel [75], Li & Ratazzi [60]. This might form a substantial project in its own right.

References

[1] Milton Abramowitz and Irene Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55 of *NBS Applied Mathematics*. US. Nat. Bureau Stand., Washington, D.C., 10th edition, December 1972.

[2] T. Abrudan, J. Eriksson, and V. Koivunen. Efficient Riemannian Algorithms for Optimization under Unitary Matrix Constraint. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 2353 –2356, March 2008.

[3] T.E. Abrudan, J. Eriksson, and V. Koivunen. Steepest Descent Algorithms for Optimization Under Unitary Matrix Constraint. *Signal Processing, IEEE Transactions on*, 56(3):1134 –1147, March 2008.

[4] Traian Abrudan, Jan Eriksson, and Visa Koivunen. Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Processing*, 89(9):1704 – 1714, March 2009.

[5] T. Adali and Hualiang Li. A Practical Formulation for Computation of Complex Gradients and its Application to Maximum Likelihood ICA. *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2:II–633 – II–636, April 2007.

[6] B.G. Agee, S.V. Schell, and W.A. Gardner. The SCORE Approach to Blind Adaptive Signal Extraction: An Application of the Theory of Spectral Correlation. In *Spectrum Estimation and Modeling, 1988., Fourth Annual ASSP Workshop on*, pages 277 –282, August 1988.

[7] B.G. Agee, S.V. Schell, and W.A. Gardner. Spectral Self-Coherence Restoral: A New Approach to Blind Adaptive Signal Extraction Using Antenna Arrays. *Proceedings of the IEEE*, 78(4):753 –767, April 1990.

[8] Siavash M. Alamouti. A Simple Transmit Diversity Technique for Wireless Communications. *IEEE Journal on Select Areas in Communications*, 16(8):1451–1458, October 1998.

[9] Mohamed-Slim Alouini, Ali Abdi, and Mostafa Kaveh. Sum of Gamma Variates and Performance of Wireless Communication Systems over Nakagami-Fading Channels. *Vehicular Technology, IEEE Transactions on*, 50(6):1471 –1480, November 2001.

[10] S.-I. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen, and C.R. Rao. *Differential Geometry in Statistical Inference*, volume 10 of *Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, 1987.

[11] Shun-ichi Amari and Hiroshi Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society and Oxford University Press, 2000.

[12] N.C. Beaulieu and C. Cheng. An Efficient Procedure for Nakagami-m Fading Simulation. In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 6, pages 3336–3342, November 2001.

[13] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A Blind Source Separation Technique Using Second-Order Statistics. *Signal Processing, IEEE Transactions on*, 45(2):434 –444, February 1997.

[14] Adel Belouchrani and Jean-François Cardoso. Maximum Likelihood Source Separation for Discrete Sources. In *in Proc. EUSIPCO*, pages 768–771, 1994.

[15] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, and H. Vincent Poor. *MIMO Wireless Communications*. Cambridge University Press, January 2007.

[16] Ella Bingham and Aapo Hyvärinen. A Fast Fixed-Point Algorithm for Independent Component Analysis of Complex Valued Signals. *International Journal of Neural Systems*, 10(1):1–8, February 2000.

[17] Magnus Borga and Hans Knutsson. A Canonical Correlation Approach to Blind Source Separation. Technical Report LiU-IMT-EX-0062, Department of Biomedical Engineering, Linkoping University, Sweden, June 2001.

[18] D.H. Brandwood. A complex gradient operator and its application in adaptive array theory. *Communications, Radar and Signal Processing, IEE Proceedings F*, 130(1):11–16, February 1983.

[19] J.F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *Radar and Signal Processing, IEE Proceedings F*, 140(6):362–370, December 1993.

[20] R.-B. Chen, M. Guo, W. Härdle, and S.-F. Huang. Independent Component Analysis Via Copula Techniques. SFB 649 Discussion Paper SFB649DP2008-004, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany, January 2008.

[21] Umberto Cherubini, Elisa Luciano, and Walter Vecchiato. *Copula Methods in Finance*. Wiley finance series. Wiley, 2004.

[22] V. Choqueuse, K. Yao, L. Collin, and G. Burel. Hierarchical Space-Time Block Code Recognition Using Correlation Matrices. *Wireless Communications, IEEE Transactions on*, 7(9):3526 –3534, September 2008.

[23] David Christensen. Fast algorithms for the calculation of Kendall's τ. *Computational Statistics*, 20(1):51–62, March 2005.

[24] A. Cichocki and S. Amari. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley and Sons, 2002.

[25] Pierre Comon. Independent component analysis, A new concept? *Signal Processing*, 36:287–314, 1994.

[26] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[27] Imre Csiszár and János Körner. Broadcast Channels with Confidential Messages. *IEEE Transactions on Information Theory*, IT-24(3):339–348, May 1978.

[28] M. Davies. Identifiability Issues in Noisy ICA. *Signal Processing Letters, IEEE*, 11(5):470–473, May 2004.

[29] E. de Carvalho, J. Cioffi, and D. Slock. Cramér-Rao Bounds for Blind Multichannel Estimation. *Global Telecommunications Conference, 2000. GLOBECOM'00. IEEE*, 2:1036–1040, November 2000.

[30] P Dita. Factorization of unitary matrices. *Journal of Physics A: Mathematical and General*, 36(11):2781–2789, March 2003.

[31] T.S. Durrani and X. Zeng. Copulas for bivariate probability distributions. *Electronics Letters*, 43(4):248–249, February 2007.

[32] J. Eriksson and V. Koivunen. Complex Random Vectors and ICA Models: Identifiability, Uniqueness, and Separability. *Information Theory, IEEE Transactions on*, 52(3):1017–1029, March 2006.

[33] Jan Eriksson, Anne-Mari Seppola, and Visa Koivunen. Complex ICA for Circular and Non-Circular Sources. In *Proceedings of the 13th European Signal Processing Conference (EUSIPCO '05), Antalya, Turkey*, September 2005.

[34] G. J. Foschini and M. J. Gans. On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas. *Wireless Personal Communications*, 6:311–335, 1998.

[35] Gerard J. Foschini. Layered Space-Time Architecture for Wireless Communication in a Fading Environment when Using Multi-Element Antennas. *Bell Labs Technical Journal*, pages 41–59, Autumn 1996.

[36] J. Galy and C. Adnet. Canonical Correlation Analysis: A Blind Source Separation Using Non-Circularity. In *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, volume 1, pages 465 –473 vol.1, December 2000.

[37] Fulvio Gini, Ruggero Reggiannini, and Umberto Mengali. The Modified Cramér-Rao Bound in Vector Parameter Estimation. *Communications, IEEE Transactions on*, 46(1):52 –60, January 1998.

[38] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[39] O. Grellier and P. Comon. Performance of Blind Discrete Source Separation. In *Eusipco*, volume IV, pages 2061–2064, Rhodes, Greece, September 1998.

[40] Olivier Grellier and Pierre Comon. Blind Separation of Discrete Sources. *Signal Processing Letters, IEEE*, 5(8):212 –214, August 1998.

[41] Franck Harroy and Jean-Louis Lacoume. Maximum likelihood estimators and Cramer-Rao bounds in source separation. *Signal Processing*, 55(2):167 – 177, July 1996.

[42] Alfred O. Hero. Secure Space-Time Communication. *IEEE Transactions on Information Theory*, 49(12):3235–3249, December 2003.

[43] Aapo Hyvärinen. Survey on Independent Component Analysis. *Neural Computing Surveys*, 2(94):128, 1999.

[44] Aapo Hyvärinen and Erkki Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[45] Aapo Hyvärinen and Erkki Oja. Independent Component Analysis: A Tutorial. *Neural Computing Surveys*, 2:94–128, April 1999.

[46] Aapo Hyvärinen and Erkki Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411–430, 2000.

[47] Mohinder Jankiraman. *Space-Time Codes and MIMO Systems*. Artech House, 2004. ISBN 1580 538657.

[48] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.

[49] J. Kitchen. Intercept Capacity: Unknown Pre-Processing. In *AMSI-MASCOS*, Melbourne, Australia, December 2007.

[50] J. Kitchen. Channel-Independent Symbol Stream Recovery for Orthogonal Space-Time Block-Coded Signals. In *4th International Conference on Signal Processing and Communication Systems ICSPCS'2010*, Gold Coast, Australia, December 2010.

[51] J. Kitchen and W. Moran. Performance Bounds for Blind MIMO Estimation. In *Defence Applications of Signal Processing (DASP)*, Queensland, Australia, December 2006.

[52] J. Kitchen and W. Moran. Copula Techniques for Modelling Signal Dependence in Wireless Communications. In *9th Engineering Mathematics and Applications Conference, EMAC2009*, December 2009.

[53] J. Kitchen and W. Moran. Copula techniques in wireless communications. In P. Howlett, M. Nelson, and A. J. Roberts, editors, *Proceedings of the 9th Biennial Engineering Mathematics and Applications Conference, EMAC-2009*, volume 51 of *ANZIAM J.*, pages C526–C540, August 2010.

[54] John Kitchen, Bill Moran, and Stephen D. Howard. Intercept Capacity: Unknown Unitary Transformation. *Entropy*, 10(4):722–735, November 2008.

[55] Zbyněk Koldovský and Petr Tichavský. Blind Instantaneous Noisy Mixture Separation with Best Interference-Plus-Noise Rejection. In M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, editors, *ICA'07: Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation*, volume 4666 of *Lecture Notes in Computer Science*, pages 730–737. Springer-Verlag, December 2007.

[56] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating Mutual Information. *Phys. Rev. E*, 69(6):066138, June 2004.

[57] Erik G. Larsson and Petre Stoica. *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003. ISBN 0521 824567.

[58] Erik G. Learned-Miller and John W. Fisher III. ICA Using Spacings Estimates of Entropy. *Journal of Machine Learning Research*, 4(7-8):1271–1295, December 2003.

[59] S. K. Leung-Yan-Cheong and M. E. Hellman. The Gaussian Wire-Tap Channel. *IEEE Transactions on Information Theory*, IT-24(4):451–456, July 1978.

[60] Xiaohua Li and E.P. Ratazzi. MIMO Transmissions with Information-Theoretic Secrecy for Secret-Key Agreement in Wireless Networks. *Military Communications Conference, 2005. MILCOM 2005. IEEE*, pages 1353–1359 Vol. 3, October 2005.

[61] Ju Liu, Bo Gu, Hongji Xu, and Jianping Qiao. Blind Detection of Orthogonal Space-Time Block Coding Based on ICA Schemes. In Jun Wang, Xiaofeng Liao, and Zhang Yi, editors, *Advances in Neural Networks - ISNN 2005*, volume 3498 of *Lecture Notes in Computer Science*, pages 309–314. Springer Berlin / Heidelberg, 2005.

[62] Wei Liu, Danilo P. Mandic, and Andrzej Cichocki. Blind Source Separation Based on Generalised Canonical Correlation Analysis and its Adaptive Realization. In *Image and Signal Processing, 2008. CISP '08. Congress on*, volume 5, pages 417 –421, May 2008.

[63] H. Lütkepohl. *Handbook of Matrices*. Wiley, 1996.

[64] Jian Ma and Zengqi Sun. Copula Component Analysis. In *ICA'07: Proceedings of the 7th international conference on Independent Component Analysis and Signal Separation*, pages 73–80, Berlin, Heidelberg, 2007. Springer-Verlag. isbn = 3-540-74493-2, 978-3-540-74493-1.

[65] Wing-Kin Ma. Blind ML Detection of Orthogonal Space-Time Block Codes: Identifiability and Code Construction. *Signal Processing, IEEE Transactions on*, 55(7):3312–3324, July 2007.

[66] Yao Ma and Dongbo Zhang. Complex Nakagami Channel Simulator with Accurate Phase and Auto-Correlation Properties. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 3847–3851, November 2007.

[67] Yao Ma and Dongbo Zhang. A Method for Simulating Complex Nakagami Fading Time Series with Nonuniform Phase and Prescribed Autocorrelation Characteristics. *Vehicular Technology, IEEE Transactions on*, 59(1):29 –35, January 2010.

[68] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Economics*. John Wiley & Sons, July 2002.

[69] A. Mansour, C. Jutten, and N. Ohnishi. Kurtosis: Definition and Properties. In *International Conference on Multisource Multisensor Information Fusion*, pages 40–46, July 1998.

[70] Thomas L. Marzetta and Bertrand M. Hochwald. Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading. *IEEE Transactions on Information Theory*, 45(1):139–157, January 1999.

[71] A.M. Mathai. *Jacobians of Matrix Transformations and Functions of Matrix Argument*. World Scientific, June 1997.

[72] H. Mathis. On the Kurtosis of Digitally Modulated Signals with Timing Offsets. In *Wireless Communications, 2001. (SPAWC '01). 2001 IEEE Third Workshop on Advances in Signal Processing*, pages 86–89, Taiwan, March 2001.

[73] U.M. Maurer. Secret Key Agreement by Public Discussion from Common Information. *IEEE Transactions on Information Theory*, 39(3):733–742, May 1993.

[74] L. Molgedey and H. G. Schuster. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Physical Review Letters*, 72:3634–3637, 1994.

[75] R. Negi and S. Goel. Secret Communication using Artificial Noise. *Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd*, 3:1906–1910, September 2005.

[76] R.B. Nelsen. *An Introduction to Copulas*. Springer-Verlag New York, Inc., 1999.

[77] D. P. Palomar and S. Verdú. Gradient of Mutual Information in Linear Vector Gaussian Channels. *IEEE Transactions on Information Theory*, 52(1):141–154, January 2006.

[78] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1989.

[79] Barak A. Pearlmutter and Lucas C. Parra. Maximum Likelihood Blind Source Separation: A Context-Sensitive Generalization of ICA. In *Advances in Neural Information Processing Systems 9*, pages 613–619. MIT Press, 1997.

[80] K.B. Petersen and M.S. Pedersen. The Matrix Cookbook, November 2008.

[81] Dinh-Tuan Pham and J.-F. Cardoso. Blind Separation of Instantaneous Mixtures of Nonstationary Sources. *Signal Processing, IEEE Transactions on*, 49(9):1837 –1848, September 2001.

[82] J. Rinas and K.D. Kammeyer. MIMO Measurements of Communication Signals and Application of Blind Source Separation. In *Signal Processing and Information Technology, 2003. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on*, pages 94 – 97, December 2003.

[83] O. Rioul. Information Theoretic Proofs of Entropy Power Inequalities. *eprint arXiv: 0704.1751*, April 2007.

[84] J.A. Ritcey. Copula Models for Wireless Fading and their Impact on Wireless Diversity combining. In *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pages 1564 –1567, November 2007.

[85] J.C.S. Santos Filho and M.D. Yacoub. On the Simulation and Correlation Properties of Phase-Envelope Nakagami Fading Processes. In *Microwave and Optoelectronics Conference, 2007. IMOC 2007. SBMO/IEEE MTT-S International*, pages 558–562, November 2007.

[86] Louis L. Scharf. *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Addison-Wesley, 1991.

[87] C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, July 1948.

[88] C. E. Shannon. Communication Theory of Secrecy Systems. *Bell Syst. Tech. J.*, 29:656–715, 1949.

[89] S.B. Slimane. Approximation to the Symmetric Capacity of Rayleigh Fading Channels with Multi-Level Signals. *Communications Letters, IEEE*, 10(3):129–131, March 2006.

[90] S. T. Smith. Covariance, Subspace, and Intrinsic Cramér-Rao Bounds. *Signal Processing, IEEE Transactions on*, 53(5):1610–1630, May 2005.

[91] Duncan M'Laren Young Sommerville. *An Introduction to the Geometry of N Dimensions*. New York, E. P. Dutton and company, inc., 1929.

[92] Zhefeng Song, Keli Zhang, and Yong Liang Guan. Generating Correlated Nakagami Fading Signals with Arbitrary Correlation and Fading Parameters. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 3, pages 1363–1367 vol.3, 2002.

[93] Harald Stögbauer, Alexander Kraskov, Sergey A. Astakhov, and Peter Grassberger. Least-dependent-component analysis based on mutual information. *Phys. Rev. E*, 70(6):066123, December 2004.

[94] A.L. Swindlehurst. Blind Separation of Space-Time Block Coded Signals via the Analytic Constant Modulus Algorithm. In *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002*, pages 447 – 451, August 2002.

[95] A.L. Swindlehurst and G. Leus. Blind and Semi-Blind Equalization for Generalized Space-Time Block Codes. *Signal Processing, IEEE Transactions on*, 50(10):2489 – 2498, October 2002.

[96] V. Tarokh, H. Jafarkhani, and A.R. Calderbank. Space-Time Block Codes from Orthogonal Designs. *Information Theory, IEEE Transactions on*, 45(5):1456 –1467, July 1999.

[97] İ. Emre Telatar. Capacity of Multi-Antenna Gaussian Channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999.

[98] Petr Tichavský and Zbyněk Koldovský. Optimal Pairing of Signal Components Separated by Blind Techniques. *IEEE Signal Processing Letters*, 11(2):119–122, February 2004.

[99] Petr Tichavský, Zbyněk Koldovský, and Erkki Oja. Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis. *IEEE Trans. on Signal Processing*, April 2006.

[100] Petr Tichavský, Zbyněk Koldovský, and Erkki Oja. Corrections to: Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis. *IEEE Transactions on Signal Processing*, 56(4):1715–1716, April 2008.

[101] L. Tong, V.C. Soon, Y.F. Huang, and R. Liu. AMUSE: A New Blind Identification Algorithm. In *Circuits and Systems, IEEE International Symposium on*, volume 3, pages 1784 –1787, May 1990.

[102] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.

[103] Franz Vesely. Statistical physics. Web Tutorial, January 2005.

[104] J. Villares. *Sample Covariance Based Parameter Estimation for Digital Communications*. PhD thesis, UPC, May 2005.

[105] N. Vlassis and Y. Motomura. Efficient Source Adaptivity in Independent Component Analysis. *Neural Networks, IEEE Transactions on*, 12(3):559 –566, May 2001.

[106] Branka Vucetic and Jinhong Yuan. *Space-Time Coding*. John Wiley and Sons, 2003. ISBN 0470 847573.

[107] Jack H. Winters, Jack Salz, and Richard D. Gitlin. Adaptive Antennas for Digital Mobile Radio. *Proceedings of the Adaptive Antenna Systems Symposium*, November 1992.

[108] Jack H. Winters, Jack Salz, and Richard D. Gitlin. The Capacity of Wireless Communication Systems Can Be Substantially Increased by the Use of Antenna Diversity. *Proc. 1st International Conf. on Universal Personal Communications*, September 1992.

[109] Jack H. Winters, Jack Salz, and Richard D. Gitlin. The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems. *IEEE Transactions on Communications*, 42(2/3/4):1740–1751, February 1994.

[110] Armin Wittneben. A New Bandwidth Efficient Transmit Antenna Modulation Diversity Scheme for Linear Digital Modulation. *Communications, 1993. ICC 93. Geneva. Technical Program, Conference Record, IEEE International Conference on*, 3:1630–1634 vol.3, May 1993.

[111] A. D. Wyner. The Wire-Tap Channel. *The Bell System Technical Journal*, 54(8):1355–1387, October 1975.

[112] João Xavier and Victor Barroso. Intrinsic Variance Lower bound (IVLB): An Extension of the Cramér-Rao Bound to Riemannian Manifolds. *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05).IEEE International Conference on*, 5, 2005.

[113] M.D. Yacoub, G. Fraidenraich, and J.C.S. Santos Filho. Nakagami-m phase-envelope joint distribution. *Electronics Letters*, 41(5):259–261, March 2005.

[114] H. H. Yang. Serial Updating Rule for Blind Separation Derived from the Method of Scoring. *Signal Processing, IEEE Transactions on*, 47(8):2279–2285, August 1999.

[115] A. Yeredor. Blind Separation of Gaussian Sources With General Covariance Structures: Bounds and Optimal Estimation. *Signal Processing, IEEE Transactions on*, 58(10):5057 –5068, October 2010.

[116] Li Yi-Ou, T. Adali, Wei Wang, and V.D. Calhoun. Joint Blind Source Separation by Multiset Canonical Correlation Analysis. *Signal Processing, IEEE Transactions on*, 57(10):3918 –3929, October 2009.

[117] Keli Zhang, Zhefeng Song, and Yong Liang Guan. Simulation of Nakagami Fading Channels with Arbitrary Cross-Correlation and Fading Parameters. *Wireless Communications, IEEE Transactions on*, 3(5):1463–1468, September 2004.

[118] Q.T. Zhang. A Decomposition Technique for Efficient Generation of Correlated Nakagami Fading Channels. *Selected Areas in Communications, IEEE Journal on*, 18(11):2385–2392, November 2000.

[119] Q.T. Zhang. A Generic Correlated Nakagami Fading Model for Wireless Communications. *Communications, IEEE Transactions on*, 51(11):1745–1748, November 2003.

[120] Lizhong Zheng and David N. C. Tse. Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel. *Information Theory, IEEE Transactions on*, 48(2):359–383, February 2002.

Appendices

Appendix A

The Wire-Tap Channel

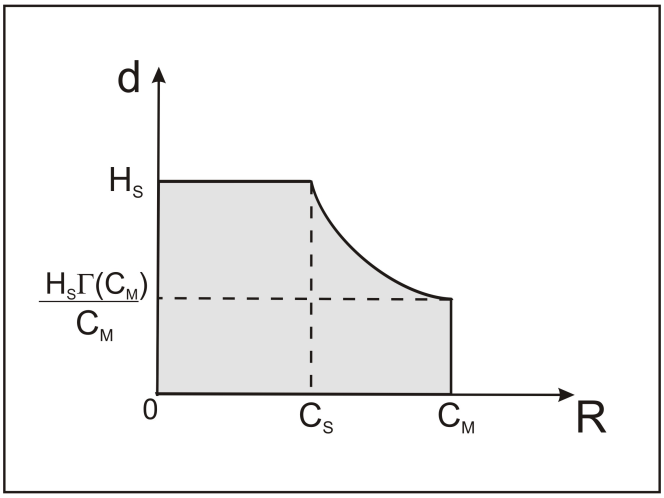

Figure A.1: Wyner's achievable (R,d) region.

Wyner showed that is possible to encode data in such a way that the wire-tapper's level of confusion will be as high as possible and characterised the set of achievable transmission rate / wire-tapper equivocation pairs (R, d) as shown in figure A.1. He defines the equivocation $\Delta = \frac{1}{K}H(\mathbf{S}^K|\mathbf{Z}^N)$ and proves the following theorems:

Wyner Theorem: For the set $\mathcal{R}$ of achievable (R, d) pairs is equal to

$$\bar{\mathcal{R}} \triangleq \{(R, d) : \quad 0 \leqq R \leqq C_M, \quad 0 \leqq d \leqq H_S, \quad Rd \leqq H_S\Gamma(R)\} \tag{A.1}$$

Wyner's WTC has subsequently been recognised as a form of degraded broadcast channel (Leung-Yan-Cheong & Hellman [59]), the difference being that one information rate is to be maximised and the other minimised. (Leung-Yan-Cheong & Hellman) Wyner determined that the achievable (R, d) region when both channels are DMC.

A.1 Simplified Explanation of the WTC

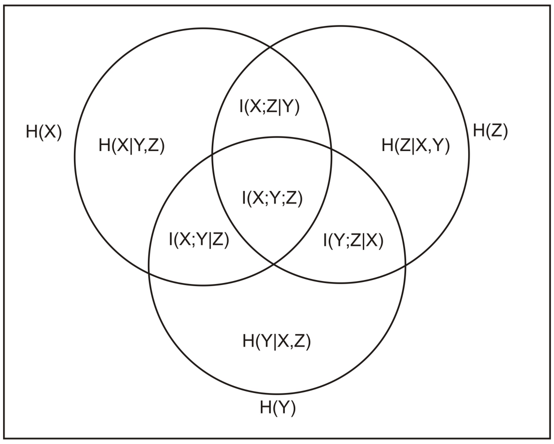

Figure A.2: Three Random Variables

Figure A.2 shows the most general representation for three random variables. We first note that the WTC is a Markov process: $X \to Y \to Z$ and its Venn diagram representation is shown in figures A.3 and A.4.

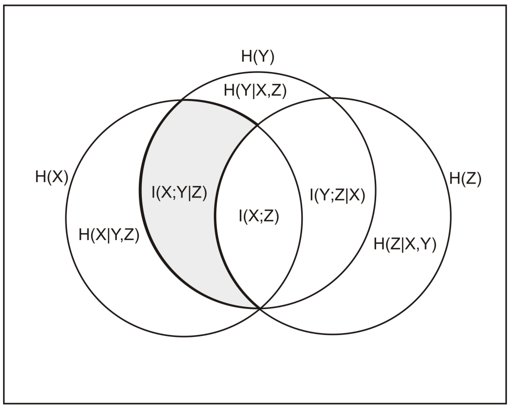

Figure A.3: MC/WTC - 3 variables.

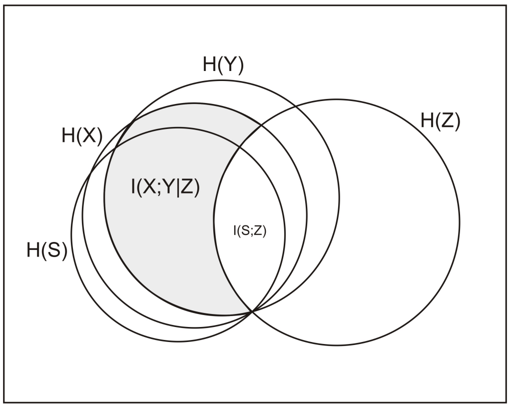

Figure A.4: MC/WTC - 4 variables.

Using the figures we may make the following observation: for a large enough value of N (yet to be determined) we may say that:

$$H(S|Z) \leqslant NI(X;Y|Z), \tag{A.2}$$

which may be written as:

$$K\left(\frac{1}{K}H(S|Z)\right) \leqslant NI(X;Y|Z) \tag{A.3}$$
$$\left(\frac{H_S K}{N}\right)\left(\frac{1}{K}H(S|Z)\right) \leqslant H_s I(X;Y|Z) \tag{A.4}$$
$$R\Delta \leqslant H_S I(X;Y|Z). \tag{A.5}$$

This very nearly equates to the result obtained by Wyner with the differences being Δ in place of d and $I(X;Y|Z)$ in place of $\Gamma(R)$. However we may now interpret the theorem as follows: it is possible, by choosing N large enough, to communicate reliably and ensure that the wire-tapper's equivocation is close to H_S.

Another way to understand the wire-tap channel (and perhaps the more general case) is given in figure A.5. Consider the following:

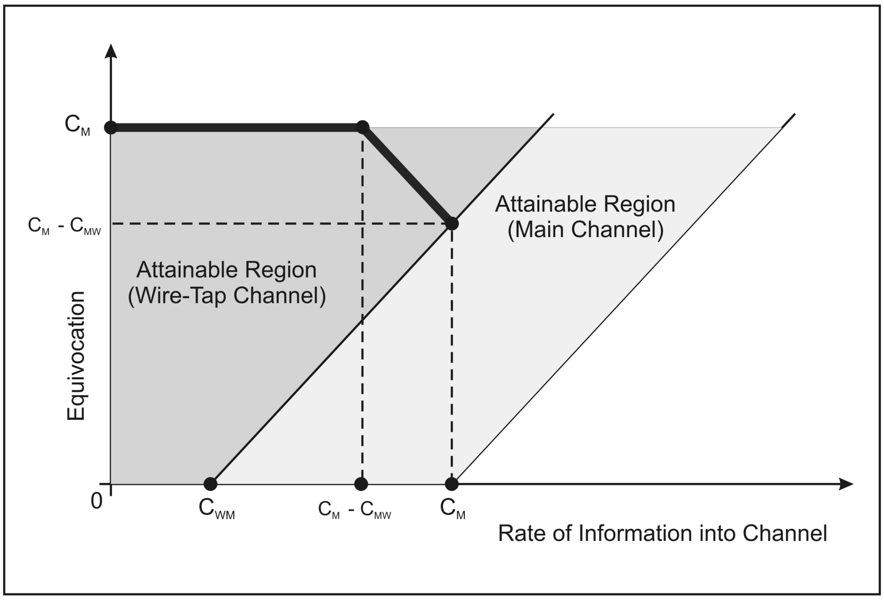

Figure A.5: Wire-Tap Channel - Capacity Diagram.

- The main channel input rate can take any value up to the main channel capacity C_M with zero equivocation i.e. error free.

- If the input rate to the main channel is C_M the input rate to the wire-tap channel will also be C_M but the equivocation at the WTC output will be $C_M - C_{MW}$. At an input rate of C_M the main channel has no freedom for further coding to increase the WTC equivocation.
- If the input rate is decreased then the main channel will have some freedom to code the message such that the WTC equivocation is increased.
- when the input rate is reduced to $C_M - C_{MW}$ it is possible to code the message such that the WTC equivocation is increased to the source entropy H_S or, if the channel was initially matched to the source, C_M.

Hence the bold curve obtained helps to understand Wyner's rate/equivocation region.

Appendix B

Rotation Entropy

The FastICA algorithm returns a channel matrix estimate which has an unknown scale and permutation. To compare the source estimates with the original sources we must first determine what the scale and permutation is and adjust the mixing matrix accordingly. We employ the "nearest2.m" algorithm provided by Tichavsky and Koldovsky [98] which finds the nearest matrix (in the Frobenius norm sense) to the original matrix with the same rows (up to the signs and order). Of course we have not yet factored in the uncertainty due to the unknown rotation (unitary transformation) incurred by the ICA technique. This uncertainty, or entropy, leads to a further decrease in the blind information. In [120] the authors show that the entropy for an unknown unitary matrix $\mathbf{U} \in \mathcal{C}^{M\times T}$, where the row vectors span the same subspace, can be determined from the logarithm of the volume of a Grassmann manifold as

$$h(\mathbf{U}) = \log|G(T,M)| \tag{B.1}$$

and where

$$|G(T,M)| = \frac{|S(T,M)|}{S(M,M)} = \frac{\prod\limits_{i=T-M+1}^{T} \frac{2\pi^i}{(i-1)!}}{\prod\limits_{i=1}^{M} \frac{2\pi^i}{(i-1)!}}. \tag{B.2}$$

$S(T,M)$ is the stiefel manifold, defined as the set of all unitary $M\times T$ matrices i.e.

$$S(T,M) = \left\{Q \in \mathcal{C}^{M\times T} : QQ^{\dagger} = \mathbf{I}_M\right\} \tag{B.3}$$

Appendix C

The Generalised Gaussian Distribution

In many of the simulations that have been performed the sources were obtained from the GG distribution, with zero-mean, unit variance, and parameterised by α. For a scalar random variable which takes values $x \in \mathbb{R}$, the GG distribution is given by:

$$f_\alpha(x) = \frac{\alpha\beta}{2\Gamma\left(\frac{1}{\alpha}\right)} \exp\left\{-[\beta|x|]^\alpha\right\}, \tag{C.1}$$

where

$$\beta = \sqrt{\frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}} \tag{C.2}$$

and $\alpha > 0$ determines the Gaussianity of the distribution. When $\alpha = 2$ the distribution is Gaussian, $\alpha = 1$ produces a Laplacian distribution and when $\alpha \to \infty$ the distribution becomes uniform. The k^{th} absolute moment for the distribution is given by

$$\mathbb{E}\left\{|x|^k\right\}_\alpha = \int_{-\infty}^{\infty} f_\alpha(x)dx = \frac{1}{\beta^k}\frac{\Gamma\left(\frac{k+1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \tag{C.3}$$

and the score function for the distribution is

$$\phi_\alpha(x) = \frac{|x|^{\alpha-1}sign(x)}{\mathbb{E}\left\{|x|^\alpha\right\}}. \tag{C.4}$$

Thus we find that

$$\begin{aligned}\kappa_\alpha &= \mathbb{E}\left\{\phi_\alpha^2(x)\right\} = \frac{\mathbb{E}\left\{|x|^{2\alpha-2}\right\}}{\left[\mathbb{E}\left\{|x|^\alpha\right\}\right]^2} \\ &= \begin{cases} \frac{\Gamma\left(2-\frac{1}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)}{\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2} & \alpha > \frac{1}{2}, \\ +\infty & \text{otherwise.} \end{cases}\end{aligned} \tag{C.5}$$

With some further calculation it is possible to also show that $\eta_\alpha = \alpha + 1$. We employ the cumulant-based definition of kurtosis where it is defined as the normalised (by the square of the second cumulant κ_2) fourth cumulant κ_4 [69]:

$$\text{kurtosis} \triangleq \frac{\kappa_4}{\kappa_2^2}. \tag{C.6}$$

For a distribution with a zero-mean this becomes

$$\text{kurtosis} = \frac{m_4}{m_2^2} - 3, \tag{C.7}$$

where m_4, m_2 are the 4th and 2nd–order moments respectively. The conversion from α to kurtosis is derived from the moments of the distribution [24]:

$$\text{kurtosis} = \frac{\Gamma\left(\frac{5}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma^2\left(\frac{3}{\alpha}\right)} - 3. \tag{C.8}$$

Figure C.1 shows a plot of kurtosis versus α, where we note that the kurtosis is positive for $\alpha < 2$ and the kurtosis is negative when $\alpha > 2$. Figure C.2 shows |kurtosis| versus α on a log scale to emphasize the negative kurtosis values.

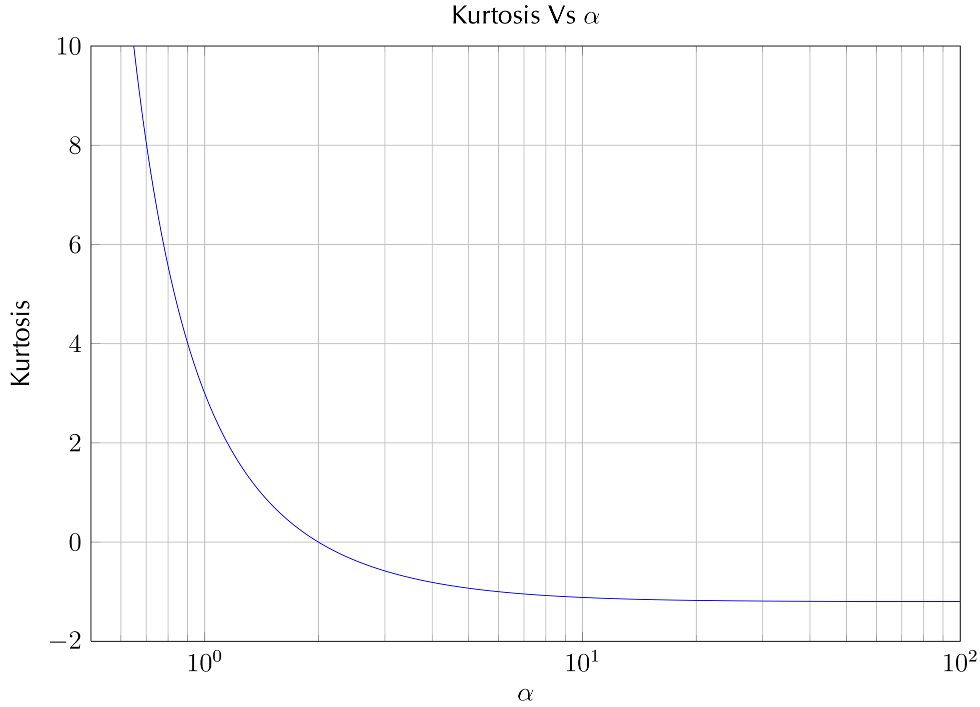

Figure C.1: Kurtosis Versus Alpha.

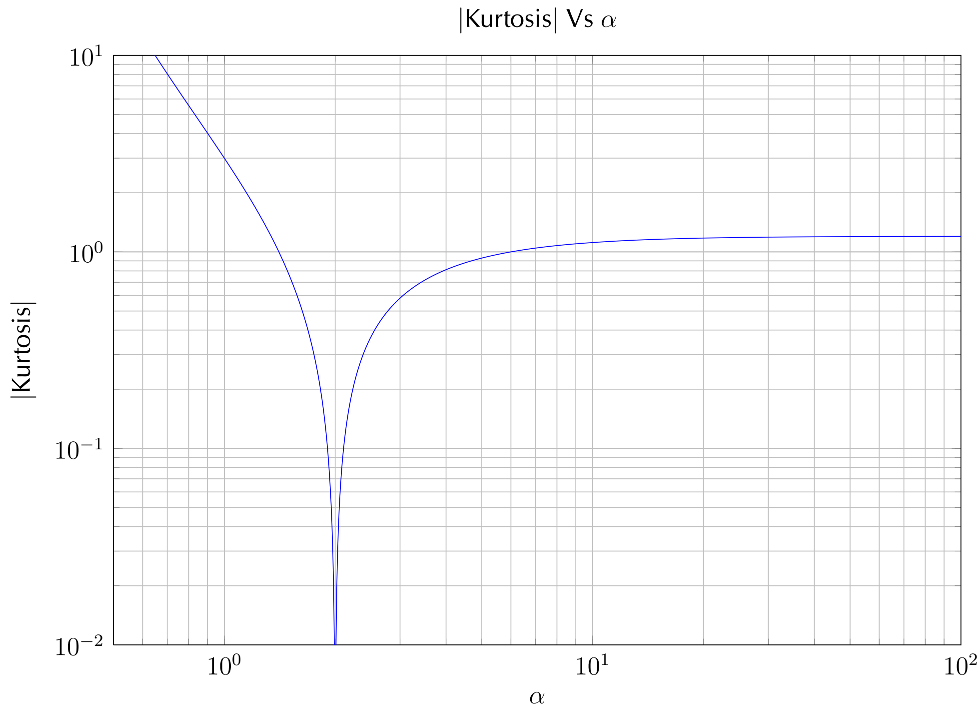

Figure C.2: Abs(Kurtosis) Versus Alpha.

Appendix D

Inverse of FIM

The elements of $\mathbf{F_I}$ were found to be given by

$$[\mathbf{F_I}]_{ij,kl} = \delta_{il}\delta_{jk} + \left[\frac{1}{2}(\eta - \kappa) - 2\right]\delta_{ijkl} + \kappa\delta_{ik}\delta_{jl}. \tag{D.1}$$

$\mathbf{F_I}$ is therefore a square matrix of dimension $d^2 \times d^2$ with entries that take one of four values:

$$[\mathbf{F_I}]_{m,n} = \begin{cases} \frac{1}{2}(\eta + \kappa - 2) & \text{if } i = j = k = l, \\ \kappa & \text{if } i = k, j = l, i \neq j, \\ 1 & \text{if } i = l, j = k, i \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{D.2}$$

The first two values: $\frac{1}{2}(\eta + \kappa - 2)$ and κ occur on the main diagonal of $\mathbf{F_I}$. Since $\mathbf{F_I}$ is a real, symmetric square matrix it may be written as an eigendecomposition i.e.

$$\mathbf{F_I} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \tag{D.3}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with entries $\in \{\frac{1}{2}(\eta + \kappa - 2), \kappa - 1, \kappa + 1\}$, $\mathbf{\Lambda} = diag\{\frac{1}{2}(\eta+\kappa-2), \frac{1}{2}(\eta+\kappa-2), \ldots, \kappa-1, \kappa-1, \ldots, \kappa+1, \kappa+1, \ldots\}$ i.e. the first d entries are $\frac{1}{2}(\eta+\kappa-2)$, the next $n = \sum_{k=1}^{d} k$ entries are $\kappa-1$ and the last n entries are $\kappa + 1$. $\mathbf{Q}$ is a real square orthonormal matrix with entries $\in \{1, -1, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\}$. Each row of $\mathbf{Q}$ contains a 1 or -1 and a further two entries taken from $\{\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\}$. The effect of $\mathbf{Q}$ is to permute the values $\frac{1}{2}(\eta+\kappa-2)$ on the main diagonal and take sums and differences of $\kappa - 1$ and $\kappa + 1$ to obtain entries, with the value κ, also on the main diagonal and off-diagonal entries with the value 1. The determinant of $\mathbf{F_I}$ is

$$\det \mathbf{F_I} = \det \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T = \det \boldsymbol{\Lambda} = \prod_{k=1}^{d^2} \lambda_{kk} = 2^{-d}(\eta + \kappa - 2)^d(\kappa^2 - 1)^n, \tag{D.4}$$

where $n = \sum_{k=1}^{d} k$. The inverse of $\mathbf{F_I}$ is given by

$$\mathbf{F_I}^{-1} = \mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^T \tag{D.5}$$

where $[\boldsymbol{\Lambda}^{-1}]_{kk} = \frac{1}{\lambda_{kk}}$. We find that the diagonal entries for $\mathbf{F_I}^{-1}$ are given by

$$[\mathbf{F_I}^{-1}]_{ii} = \begin{cases} \frac{2}{\eta+\kappa-2} & \text{for } 1 + (i-1)(d+1), i = 1, 2, \ldots, d, \\ \frac{\kappa}{\kappa^2-1} & \text{otherwise.} \end{cases} \tag{D.6}$$

Off-diagonal entries for $\mathbf{F_I}^{-1}$ are $\in \{0, \frac{1}{\kappa^2-1}\}$.

Appendix E

Useful Result for GG Distribution

For the GG distribution we show that $\eta = \alpha+1$. For real or imaginary components η is defined as

$$\eta \triangleq \mathbb{E}\left\{(\phi u)^2\right\}, \tag{E.1}$$

where

$$\phi \triangleq \frac{-1}{f(x)}\frac{\partial f(x)}{\partial x}. \tag{E.2}$$

$$\eta = \int_R t^2\phi^2(t)f(t)dt. \tag{E.3}$$

For the GG distribution η is found from

$$\begin{aligned}
\eta &= \int_R t^2\left[\frac{|t|^{\alpha-1}sign(t)}{\mathbb{E}_\alpha^2\left\{|t|^\alpha\right\}}\right]^2 f(t)dt = \frac{1}{\mathbb{E}_\alpha^2\left\{|t|^\alpha\right\}}\int |t|^{2\alpha}dt \\
&= 2\left[\frac{\beta^\alpha\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(1+\frac{1}{\alpha}\right)}\right]^2\int_0^\infty t^{2\alpha}f(t)dt \\
&= 2\left[\alpha\beta^\alpha\right]^2\int_0^\infty t^{2\alpha}\left(\frac{\alpha\beta}{2\Gamma\left(\frac{1}{\alpha}\right)}\exp\left\{-\beta^\alpha|t|^\alpha\right\}\right)dt \\
&= \frac{\alpha^3\beta^{2\alpha+1}}{\Gamma\left(\frac{1}{\alpha}\right)}\int_0^\infty t^{2\alpha}\exp\left\{-\beta^\alpha t^\alpha\right\}dt \\
&= \frac{\alpha^3\beta^{2\alpha+1}}{\Gamma\left(\frac{1}{\alpha}\right)}\frac{\Gamma\left(2+\frac{1}{\alpha}\right)}{\alpha}\beta^{-(2\alpha+1)} = \alpha+1.
\end{aligned} \tag{E.4}$$

Appendix F

Symmetric Capacity

Symmetric channel capacity is defined as having a channel probability transition matrix whose rows are permutations of each other and the columns are permutations of each other [26, chapt.8]. Slimane [89] has derived an expression that provides an approximation for the symmetric capacity, for a single fading channel and finite input alphabet. Slimane eqn(12) :

$$C^*(a) \approx -\frac{1}{q}\sum_{i=1}^{q}\log\left(\frac{1}{q}\sum_{j=1}^{q}e^{-a^2|\Delta_{ij}|^2}\right) - \frac{1}{q}\sum_{i=1}^{q}\sum_{j=1}^{q}\frac{\left(e^{-a^2|\Delta_{ij}|^2\frac{1-\alpha}{2-\alpha}}-e^{-a^2|\Delta_{ij}|^2}\right)}{\sum_{l=1}^{q}e^{-a^2|\Delta_{il}|^2(1-\alpha)}}, \tag{F.1}$$

where :

q is the number of levels or points in the signal constellation.

a is the fading amplitude,

$\alpha = \frac{a^2\gamma_o}{2(1+a^2\gamma_o)}$,

γ_o is the average received snr per transmitted symbol, $\gamma_o = \frac{E_s}{N_o}$,

N_o is the noise power spectral density,

E_s is the average energy per symbol,

$\Delta_{ij} \triangleq \frac{s_i - s_j}{\sqrt{N_o}}$,

s_i is a transmitted symbol.

For the AWGN model used in the present study, $a = 1$ and we obtain the approximate symmetric capacity values shown in figures F.1 and F.2, for PSK and QAM modulation constellation types respectively.

We may use the symmetric capacity $C(a)$ approximation to obtain an estimate for the source entropy, for the case $y = ax + w$, as

$$h(x) \approx C(a) + h(x|y, a), \tag{F.2}$$

where a is the fading amplitude and $h(x|y, a)$ is derived from the covariance matrix of the MLE for x given y and a. If the fading amplitude $a = 1$ then $h(x|y, a) = h(w)$ and $h(w) = \log_2(\pi e \sigma_w^2)$ bits. $C(a)$ indicates the capacity or information rate that we can expect for the timing offset simulations that we perform, where we study 16QAM and 16PSK source types with AWGN and varying data lengths.

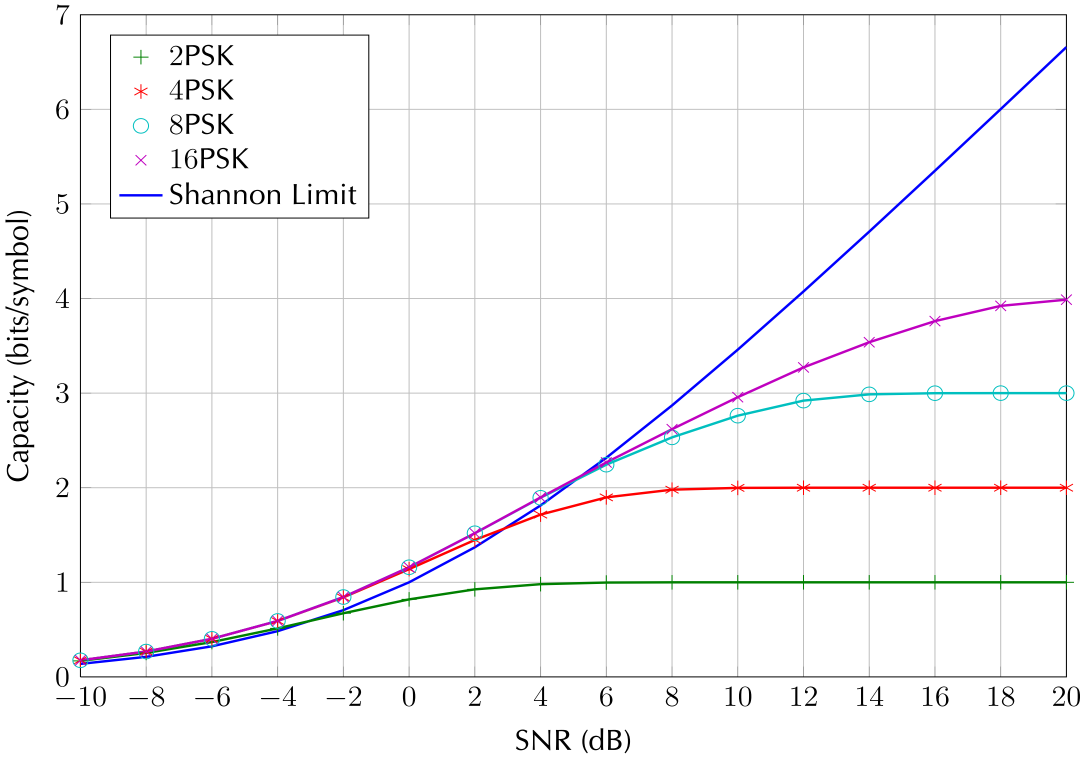

Figure F.1: Symmetric Capacity for PSK Signals.

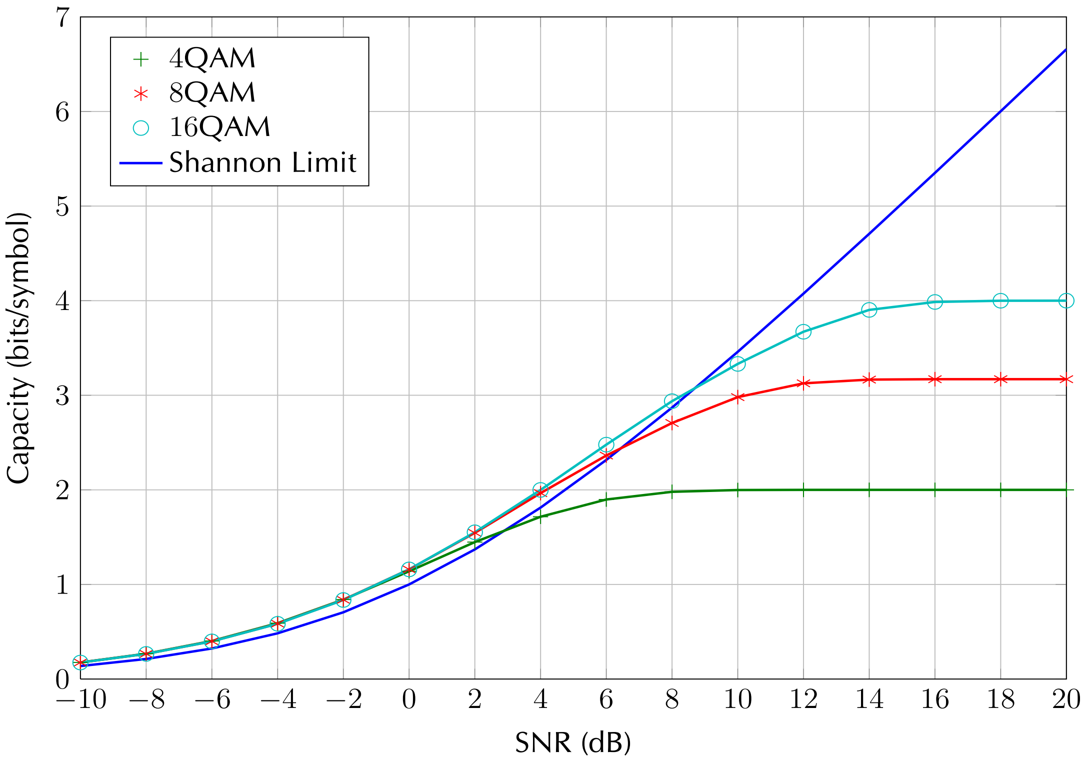

Figure F.2: Symmetric Capacity for QAM Signals.

Appendix G

Matrix Relationships

Several matrix relationships are required in this thesis. These relationships have been gathered from the Matrix Cookbook [80], the *Handbook of Matrices* [63] and *Matrix Differential Calculus with Applications in Statistics and Economics* [68].

G.1 Matrix Derivatives

$a\,(1 \times 1)$ and $\mathbf{B}\,(n \times p)$:

$$\frac{\partial a}{\partial \mathbf{B}} = \begin{bmatrix} \frac{\partial a}{\partial b_{1,1}} & \frac{\partial a}{\partial b_{1,2}} & \cdots & \frac{\partial a}{\partial b_{1,p}} \\ \frac{\partial a}{\partial b_{2,1}} & \frac{\partial a}{\partial b_{2,2}} & \cdots & \frac{\partial a}{\partial b_{2,p}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a}{\partial b_{n,1}} & \frac{\partial a}{\partial b_{n,2}} & \cdots & \frac{\partial a}{\partial b_{n,p}} \end{bmatrix}. \tag{G.1}$$

$\mathbf{a}\,(m \times 1)$ and $\mathbf{B}\,(n \times p)$:

$$\frac{\partial \mathbf{a}}{\partial \mathbf{B}} \equiv \frac{\partial \mathbf{vec}\,(\mathbf{a})}{\partial \mathbf{vec}^T\,(\mathbf{B})} = \begin{bmatrix} \frac{\partial a_{1,1}}{\partial b_{1,1}} & \frac{\partial a_{1,1}}{\partial b_{2,1}} & \cdots & \frac{\partial a_{1,1}}{\partial b_{n,p}} \\ \frac{\partial a_{2,1}}{\partial b_{1,1}} & \frac{\partial a_{2,1}}{\partial b_{2,1}} & \cdots & \frac{\partial a_{2,1}}{\partial b_{n,p}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{m,1}}{\partial b_{1,1}} & \frac{\partial a_{m,1}}{\partial b_{2,1}} & \cdots & \frac{\partial a_{m,1}}{\partial b_{n,p}} \end{bmatrix}. \tag{G.2}$$

$\mathbf{A}\,(m \times n)$ and $\mathbf{B}\,(n \times p)$:

$$\frac{\partial \mathbf{A}}{\partial \mathbf{B}} \equiv \frac{\partial \mathbf{vec}\,(\mathbf{A})}{\partial \mathbf{vec}^T\,(\mathbf{B})} = \begin{bmatrix} \frac{\partial a_{1,1}}{\partial b_{1,1}} & \frac{\partial a_{1,1}}{\partial b_{2,1}} & \cdots & \frac{\partial a_{1,1}}{\partial b_{n,p}} \\ \frac{\partial a_{2,1}}{\partial b_{1,1}} & \frac{\partial a_{2,1}}{\partial b_{2,1}} & \cdots & \frac{\partial a_{2,1}}{\partial b_{n,p}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{m,n}}{\partial b_{1,1}} & \frac{\partial a_{m,n}}{\partial b_{2,1}} & \cdots & \frac{\partial a_{m,n}}{\partial b_{n,p}} \end{bmatrix} . \tag{G.3}$$

$$\begin{aligned} \partial(\mathbf{XY}) &= (\partial\mathbf{X})\mathbf{Y} + \mathbf{X}(\partial\mathbf{Y}) && \text{(G.4)} \\ \partial(\ln(\det(\mathbf{X}))) &= \text{tr}\left(\mathbf{X}^{-1}\partial\mathbf{X}\right) && \text{(G.5)} \\ \partial(\text{tr}\,(\mathbf{X})) &= \text{tr}\,(\partial\mathbf{X}) && \text{(G.6)} \\ \partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && \text{(G.7)} \\ \partial(\text{tr}\left(\mathbf{X}^{-1}\right)) &= -\text{tr}\left(\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1}\right) . && \text{(G.8)} \end{aligned}$$

$\mathbf{X}\,(m \times n)$, $\mathbf{A}\,(p \times n)$ and $\mathbf{B}\,(m \times p)$:

$$\frac{\partial \text{tr}\left(\mathbf{A}\mathbf{X}^T\mathbf{B}\right)}{\partial \mathbf{X}} = \mathbf{BA} . \tag{G.9}$$

$\mathbf{X}\,(m \times n)$, $\mathbf{A}\,(p \times m)$ and $\mathbf{B}\,(n \times p)$:

$$\frac{\partial \text{tr}\,(\mathbf{AXB})}{\partial \mathbf{X}} = \mathbf{A}^T\mathbf{B}^T . \tag{G.10}$$

$\mathbf{X}\,(m \times n)$, $\mathbf{A}\,(m \times n)$:

$$\frac{\partial \text{tr}\left(\mathbf{X}^T\mathbf{A}\right)}{\partial \mathbf{X}} = \frac{\partial \text{tr}\left(\mathbf{A}\mathbf{X}^T\right)}{\partial \mathbf{X}} = \mathbf{A} . \tag{G.11}$$

$\mathbf{X}\,(m \times n)$, $\mathbf{A}\,(m \times n)$:

$$\frac{\partial \text{tr}\left(\mathbf{X}^\dagger\mathbf{A}\right)}{\partial \mathbf{X}^*} = \mathbf{A} . \tag{G.12}$$

$\mathbf{X}\,(m \times n)$, $\mathbf{A}\,(n \times m)$:

$$\frac{\partial \text{tr}\,(\mathbf{AX})}{\partial \mathbf{X}} = \frac{\partial \text{tr}\,(\mathbf{XA})}{\partial \mathbf{X}} = \mathbf{A}^T . \tag{G.13}$$

$\mathbf{X}\,(m \times m)$:

$$\frac{\partial \text{tr}\,(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \text{tr}\,(\mathbf{X}^T)}{\partial \mathbf{X}} = \mathbf{I}_m\,. \tag{G.14}$$

$\mathbf{X}\,(m \times n)$, $\mathbf{A}\,(p \times m)$ and $\mathbf{B}\,(n \times q)$:

$$\frac{\partial \mathbf{AXB}}{\partial \mathbf{X}} = \mathbf{B}^T \otimes \mathbf{A}\,. \tag{G.15}$$

$\mathbf{A}\,(m \times n)$, $\mathbf{B}\,(m \times n)$, $\mathbf{C}\,(m \times n)$:

$$\frac{\partial \text{tr}\,(\mathbf{A}^T\mathbf{B} \odot \mathbf{A}^T\mathbf{C})}{\partial \mathbf{A}} = \mathbf{B}[\mathbf{I}_n \odot \mathbf{A}^T\mathbf{C}] + \mathbf{C}[\mathbf{I}_n \odot \mathbf{A}^T\mathbf{B}]\,. \tag{G.16}$$

$\mathbf{A}\,(m \times n)$, $\mathbf{B}\,(n \times m)$, $\mathbf{C}\,(n \times m)$:

$$\frac{\partial \text{tr}\,(\mathbf{AB} \odot \mathbf{AC})}{\partial \mathbf{A}} = [\mathbf{I}_m \odot \mathbf{AC}]\mathbf{B}^T + [\mathbf{I}_m \odot \mathbf{AB}]\mathbf{C}^T\,. \tag{G.17}$$

$\mathbf{A}\,(m \times n)$, $\mathbf{B}\,(m \times n)$, $\mathbf{C}\,(m \times n)$:

$$\frac{\partial \text{tr}\,(\mathbf{A}^\dagger\mathbf{B} \odot \mathbf{A}^\dagger\mathbf{C})}{\partial \mathbf{A}^*} = \mathbf{B}[\mathbf{I}_n \odot \mathbf{A}^\dagger\mathbf{C}] + \mathbf{C}[\mathbf{I}_n \odot \mathbf{A}^\dagger\mathbf{B}]\,. \tag{G.18}$$

G.2 Kronecker Products

$\mathbf{A}\,(m \times n)$, $\mathbf{B}\,(p \times q)$, $\mathbf{C}\,(n \times r)$ and $\mathbf{D}\,(q \times s)$:

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}\,. \tag{G.19}$$

$\mathbf{A}\,(m \times n)$, $\mathbf{B}\,(n \times p)$ and $\mathbf{C}\,(p \times q)$:

$$\mathbf{vec}\,(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathbf{vec}\,(\mathbf{B})\;. \tag{G.20}$$

$\mathbf{A}\,(m \times n)$ and $\mathbf{B}\,(n \times p)$:

$$\begin{aligned}
\mathbf{vec}\,(\mathbf{AB}) &= (\mathbf{I}_p \otimes \mathbf{A})\mathbf{vec}\,(\mathbf{B}) && \text{(G.21)}\\
&= (\mathbf{B}^T \otimes \mathbf{I}_m)\mathbf{vec}\,(\mathbf{A}) && \text{(G.22)}\\
&= (\mathbf{B}^T \otimes \mathbf{A})\mathbf{vec}\,(\mathbf{I}_n)\;. && \text{(G.23)}
\end{aligned}$$

$\mathbf{A}\,(m \times m)$ and $\mathbf{B}\,(n \times n)$:

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \,. \tag{G.24}$$

$\mathbf{A}\,(m \times n)$ and $\mathbf{B}\,(p \times q)$:

$$(\mathbf{A} \otimes \mathbf{B})^{T} = \mathbf{A}^{T} \otimes \mathbf{B}^{T} \,. \tag{G.25}$$

$\mathbf{A}\,(m \times n)$, $\mathbf{B}\,(n \times p)$, $\mathbf{C}\,(p \times q)$ and $\mathbf{D}\,(q \times m)$:

$$\begin{aligned} \mathrm{tr}\,(\mathbf{ABCD}) &= \mathbf{vec}\left(\mathbf{D}^{T}\right)^{T}(\mathbf{C}^{T} \otimes \mathbf{A})\mathbf{vec}\,(\mathbf{B}) \\ &= \mathbf{vec}\,(\mathbf{D})^{T}(\mathbf{A} \otimes \mathbf{C}^{T})\mathbf{vec}\left(\mathbf{B}^{T}\right) \,. \end{aligned} \tag{G.26}$$

$\mathbf{A}\,(m \times n)$ and $\mathbf{B}\,(n \times m)$:

$$\mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,(\mathbf{BA}) \,, \tag{G.27}$$

$$= \mathbf{vec}^{T}\left(\mathbf{A}^{T}\right)\mathbf{vec}\,(\mathbf{B}) \,, \tag{G.28}$$

$$= \mathbf{vec}^{T}\left(\mathbf{B}^{T}\right)\mathbf{vec}\,(\mathbf{A}) \,. \tag{G.29}$$

$\mathbf{A}\,(m \times m)$ and $\mathbf{B}\,(n \times n)$:

$$\mathrm{tr}\,(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}\,(\mathbf{A})\,\mathrm{tr}\,(\mathbf{B}) \,. \tag{G.30}$$

G.3 Miscellaneous

$\mathbf{A}\,(m \times m)$, $c \in \mathbb{C}$:

$$\det(c\mathbf{A}) = c^{m}\det(\mathbf{A}) \,. \tag{G.31}$$

$\mathbf{I}_m\,(m \times m)$:

$$\det(\mathbf{I}_m) = 1 \,. \tag{G.32}$$

$\mathbf{A}\,(m \times m)$ and $\mathbf{B}\,(n \times n)$:

$$\det(\mathbf{A} \otimes \mathbf{B}) = \det\left[\det(\mathbf{A})\right]^{n}\left[\det(\mathbf{B})\right]^{m} \,. \tag{G.33}$$

For an $(m \times n)$ matrix $\mathbf{A}$, $\mathbf{K}_{mn}$ is an $mn \times mn$ commutation matrix such that

$$\mathbf{K}_{mn}\mathbf{vec}\left(\mathbf{A}\right) = \mathbf{vec}\left(\mathbf{A}^T\right) . \tag{G.34}$$

In section 4.5 it is stated that

$$\mathbb{E}\left\{\phi_i^r u_i^r\right\} = \mathbb{E}\left\{\phi_i^i u_i^i\right\} = \delta_{ij}. \tag{G.35}$$

This may be derived via integration by parts, as follows

$$\begin{aligned}
\mathbb{E}\left\{\phi_i^r u_i^r\right\} &= -\int p_x(u_i^r)\frac{\partial \ln p_x(u_i^r)}{\partial u_i^r}u_i^r du_i^r \\
&= -\int p_x(u_i^r)\frac{1}{p_x(u_i^r)}\frac{\partial p_x(u_i^r)}{\partial u_i^r}u_i^r du_i^r \\
&= -\int \frac{\partial p_x(u_i^r)}{\partial u_i^r}u_i^r du_i^r \\
&= -p_x(u_i^r)u_i^r|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} p_x(u_i^r)\frac{\partial u_i^r}{\partial u_i^r}du_i^r \\
&= -p_x(u_i^r)u_i^r|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} p_x(u_i^r)du_i^r \\
&= 0 + 1,
\end{aligned} \tag{G.36}$$

assuming $p_x(u_i^r)$ and u_i^r are continuously differentiable and $p_x(u_i^r)$ vanishes to zero at $\pm\infty$.

Appendix H

Derivation of JADE Gradient

Defining $\mathbf{\Phi} = \mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W}$, where $\mathbf{W}$ is unitary i.e. $\mathbf{W} \in U(n)$, then the JADE algorithm minimises the following cost function [3]

$$C_{JADE} = \sum_{i=1}^{m} \mathrm{tr}\left(\mathbf{\Phi}\mathbf{\Phi}^{\dagger} - \mathbf{\Phi} \odot \mathbf{\Phi}^{\dagger}\right) \tag{H.1}$$

$$= \sum_{i=1}^{m} \mathrm{tr}\left(\mathbf{\Phi}\mathbf{\Phi}^{\dagger}\right) - \sum_{i=1}^{m} \mathrm{tr}\left(\mathbf{\Phi} \odot \mathbf{\Phi}^{\dagger}\right), \tag{H.2}$$

with respect to $\mathbf{W}^*$. The eigenmatrices $\hat{\mathbf{M}}_i$ are estimated from the fourth order cumulants of the whitened observations [19]. The cost function is then used to diagonalize the eigenmatrices.

The Euclidean gradient of the JADE cost function, w.r.t. $\mathbf{W}^*$, is obtained from

$$\begin{aligned}\mathbf{G}_{JADE} &= 2\frac{\partial C_{JADE}}{\partial \mathbf{W}^*} \\ &= 2\sum_{i=1}^{m} \frac{\partial \mathrm{tr}\left(\mathbf{\Phi}\mathbf{\Phi}^{\dagger}\right)}{\partial \mathbf{W}^*} + 2\sum_{i=1}^{m} \frac{\partial \mathrm{tr}\left(\mathbf{\Phi} \odot \mathbf{\Phi}^{\dagger}\right)}{\partial \mathbf{W}^*} \\ &= 2\sum_{i=1}^{m} \frac{\partial \mathrm{tr}\left(\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\hat{\mathbf{M}}_i^{\dagger}\mathbf{W}\right)}{\partial \mathbf{W}^*} + 2\sum_{i=1}^{m} \frac{\partial \mathrm{tr}\left(\mathbf{W}^{\dagger}\hat{\mathbf{M}}_i\mathbf{W} \odot \mathbf{W}^{\dagger}\hat{\mathbf{M}}_i^{\dagger}\mathbf{W}\right)}{\partial \mathbf{W}^*}\end{aligned} \tag{H.3}$$

Making use of the matrix relationship:

$$\frac{\partial \text{tr}\left(\mathbf{A}^\dagger \mathbf{B}\right)}{\partial \mathbf{A}^*} = \mathbf{B}, \tag{H.4}$$

the first term on the right hand side (RHS) of equation H.3 is

$$2\sum_{i=1}^{m} \frac{\partial \text{tr}\left(\mathbf{W}^\dagger \hat{\mathbf{M}}_i \hat{\mathbf{M}}_i^\dagger \mathbf{W}\right)}{\partial \mathbf{W}^*} = 2\sum_{i=1}^{m}\left\{\hat{\mathbf{M}}_i \hat{\mathbf{M}}_i^\dagger \mathbf{W}\right\}. \tag{H.5}$$

The second term on the RHS of equation H.3 is found, using the relationship

$$\frac{\partial \text{tr}\left(\mathbf{A}^\dagger \mathbf{B} \odot \mathbf{A}^\dagger \mathbf{C}\right)}{\partial \mathbf{A}^*} = \mathbf{B}[\mathbf{I} \odot \mathbf{A}^\dagger \mathbf{C}] + \mathbf{C}[\mathbf{I} \odot \mathbf{A}^\dagger \mathbf{B}], \tag{H.6}$$

which is derived in Section H.1 below. Hence

$$\frac{\partial \text{tr}\left(\mathbf{W}^\dagger \hat{\mathbf{M}}_i \mathbf{W} \odot \mathbf{W}^\dagger \hat{\mathbf{M}}_i^\dagger \mathbf{W}\right)}{\partial \mathbf{W}^*} = \left\{\hat{\mathbf{M}}_i \mathbf{W}[\mathbf{I} \odot \mathbf{W}^\dagger \hat{\mathbf{M}}_i^\dagger \mathbf{W}] + \hat{\mathbf{M}}_i^\dagger \mathbf{W}[\mathbf{I} \odot \mathbf{W}^\dagger \hat{\mathbf{M}}_i \mathbf{W}]\right\}. \tag{H.7}$$

Therefore we find that

$$\mathbf{G}_{JADE} = 2\sum_{i=1}^{m}\left\{\hat{\mathbf{M}}_i \hat{\mathbf{M}}_i^\dagger \mathbf{W} - \hat{\mathbf{M}}_i \mathbf{W}[\mathbf{I} \odot \mathbf{W}^\dagger \hat{\mathbf{M}}_i^\dagger \mathbf{W}] - \hat{\mathbf{M}}_i^\dagger \mathbf{W}[\mathbf{I} \odot \mathbf{W}^\dagger \hat{\mathbf{M}}_i \mathbf{W}]\right\}. \tag{H.8}$$

H.1 Derivation of Equation I.6

The relationship shown in equation H.6 does not appear to exist in the open literature and so a derivation is presented here. We proceed by finding

$$\frac{\partial \text{tr}\left(\mathbf{A}^T \mathbf{B} \odot \mathbf{A}^T \mathbf{C}\right)}{\partial a_{m,n}} \equiv \left[\frac{\partial \text{tr}\left(\mathbf{A}^T \mathbf{B} \odot \mathbf{A}^T \mathbf{C}\right)}{\partial \mathbf{A}}\right]_{m,n}. \tag{H.9}$$

This allows us to use the known relationships

$$\frac{\partial (\text{tr}\left(\mathbf{X}\right))}{\partial z} = \text{tr}\left(\frac{\partial \mathbf{X}}{\partial z}\right) \tag{H.10}$$

and

$$\frac{\partial(\mathbf{X} \odot \mathbf{Y})}{\partial z} = \frac{\partial \mathbf{X}}{\partial z} \odot \mathbf{Y} + \mathbf{X} \odot \frac{\partial \mathbf{Y}}{\partial z}, \tag{H.11}$$

where z is a scalar, so that

$$\begin{aligned} \frac{\partial \text{tr}\left(\mathbf{A}^T\mathbf{B} \odot \mathbf{A}^T\mathbf{C}\right)}{\partial a_{m,n}} &= \text{tr}\left(\frac{\partial(\mathbf{A}^T\mathbf{B})}{\partial a_{m,n}} \odot \mathbf{A}^T\mathbf{C} + \mathbf{A}^T\mathbf{B} \odot \frac{\partial(\mathbf{A}^T\mathbf{C})}{\partial a_{m,n}}\right) \\ &= \text{tr}\left(\frac{\partial(\mathbf{A}^T\mathbf{B})}{\partial a_{m,n}} \odot \mathbf{A}^T\mathbf{C}\right) + \text{tr}\left(\mathbf{A}^T\mathbf{B} \odot \frac{\partial(\mathbf{A}^T\mathbf{C})}{\partial a_{m,n}}\right). \end{aligned} \tag{H.12}$$

Now

$$\frac{\partial[\mathbf{A}^T\mathbf{B}]_{i,j}}{\partial a_{m,n}} = \delta_{i,n} b_{m,j}, \tag{H.13}$$

which yields a matrix, of the same dimensions as $\mathbf{A}$, with all rows zero except the nth row. The action of the trace operator then is to select the non-zero term in position n, n of its matrix argument. Therefore

$$\frac{\partial \text{tr}\left(\mathbf{A}^T\mathbf{B} \odot \mathbf{A}^T\mathbf{C}\right)}{\partial a_{m,n}} = b_{m,n}[\mathbf{A}^T\mathbf{C}]_{n,n} + c_{m,n}[\mathbf{A}^T\mathbf{B}]_{n,n}, \tag{H.14}$$

from which we deduce that

$$\frac{\partial \text{tr}\left(\mathbf{A}^T\mathbf{B} \odot \mathbf{A}^T\mathbf{C}\right)}{\partial \mathbf{A}} = \mathbf{B}[\mathbf{I}_n \odot \mathbf{A}^T\mathbf{C}] + \mathbf{C}[\mathbf{I}_n \odot \mathbf{A}^T\mathbf{B}], \tag{H.15}$$

or

$$\frac{\partial \text{tr}\left(\mathbf{A}^\dagger\mathbf{B} \odot \mathbf{A}^\dagger\mathbf{C}\right)}{\partial \mathbf{A}^*} = \mathbf{B}[\mathbf{I}_n \odot \mathbf{A}^\dagger\mathbf{C}] + \mathbf{C}[\mathbf{I}_n \odot \mathbf{A}^\dagger\mathbf{B}]. \tag{H.16}$$

Appendix I

Mutual Information Gradient

For an m-dimensional random vector $\mathbf{u}$, the MI of its components is defined as

$$I(\mathbf{u}) \triangleq \mathbb{E}\left\{\log \frac{p(\mathbf{u})}{\prod_{i=1}^{m} p(u_i)}\right\}. \tag{I.1}$$

The MI can also be written in terms of entropy

$$I(\mathbf{u}) = \sum_{i=1}^{m} H(u_i) - H(\mathbf{u}), \tag{I.2}$$

where $H(u) = -\mathbb{E}\{\log p(u)\}$. The source separation problem may be obtained by minimising the MI. We employ a gradient-based method which requires differentiating $I(\mathbf{Wy})$ w.r.t. $\mathbf{W}^*$ and we find that this is a function of the score function. So with

$$I(\mathbf{u}) = \mathbb{E}\{\log p(\mathbf{u})\} - \sum_{i=1}^{m} \mathbb{E}\{\log p(u_i)\}, \tag{I.3}$$

the demixing process requires $\mathbf{u} = \mathbf{Wy}$ for source estimation. Since

$$p(\mathbf{u}) = \frac{p(\mathbf{y})}{|\det \mathbf{WW}^*|}, \tag{I.4}$$

then

$$I(\mathbf{u}) = \mathbb{E}\{\log p(\mathbf{y})\} - \log|\det \mathbf{WW}^*| - \sum_{i} \mathbb{E}\{\log p(u_i)\}. \tag{I.5}$$

The gradient of $I(\mathbf{u})$ w.r.t. $\mathbf{W}^*$ is

$$\frac{\partial I(\mathbf{u})}{\partial \mathbf{W}^*} = -\frac{\partial \log |\det \mathbf{W}\mathbf{W}^*|}{\partial \mathbf{W}^*} - \sum_i \frac{\partial \mathbb{E}\left\{\log p(u_i)\right\}}{\partial \mathbf{W}^*}, \tag{I.6}$$

since $\mathbb{E}\left\{\log p(\mathbf{y})\right\}$ does not involve $\mathbf{W}^*$. The first term on the RHS of the above equation is

$$\frac{\partial \log |\det \mathbf{W}\mathbf{W}^*|}{\partial \mathbf{W}^*} = \mathbf{W}^{-\dagger}. \tag{I.7}$$

The second term requires first rewriting $p(\mathbf{u})$ as $p(\mathbf{u}, \mathbf{u}^*)$, via the Brandwood analyticity condition [5], so that

$$\begin{aligned} \sum_i \frac{\partial \mathbb{E}\left\{\log p(u_i)\right\}}{\partial \mathbf{W}^*} &= \sum_i \frac{\partial \mathbb{E}\left\{\log p(u_i, u_i^*)\right\}}{\partial \mathbf{u}_i^*} \frac{\partial u_i^*}{\partial \mathbf{W}^*} \\ &= -\boldsymbol{\psi}(\mathbf{u}, \mathbf{u}^*)\mathbf{y}^\dagger, \end{aligned} \tag{I.8}$$

where $\boldsymbol{\psi}(\mathbf{u}, \mathbf{u}^*) = \frac{1}{2}\left[\frac{\mathbb{E}\{\partial p_x(\mathbf{u}_R, \mathbf{u}_I)\}}{\partial \mathbf{u}_R} + j\frac{\mathbb{E}\{\partial p_x(\mathbf{u}_R, \mathbf{u}_I)\}}{\partial \mathbf{u}_I}\right]$ is a vector of complex score functions [33]. The score functions can be calculated if the source distributions are known or estimated directly from observed data using a method such as that described in [105]. The MI gradient may now be written as

$$\frac{\partial I(\mathbf{u})}{\partial \mathbf{W}^*} = \boldsymbol{\psi}(\mathbf{u}, \mathbf{u}^*)\mathbf{y}^\dagger - \mathbf{W}^{-\dagger}. \tag{I.9}$$

Since $\mathbf{W}$ is unitary, $\mathbf{W}^\dagger\mathbf{W} = \mathbf{I}$, the matrix inversion can be avoided by writing:

$$\frac{\partial I(\mathbf{u})}{\partial \mathbf{W}^*}\mathbf{W}^\dagger\mathbf{W} = \left[\boldsymbol{\psi}(\mathbf{u}, \mathbf{u}^*)\mathbf{u}^\dagger - \mathbf{I}\right]\mathbf{W}, \tag{I.10}$$

so that

$$\frac{\partial I(\mathbf{u})}{\partial \mathbf{W}^*} = \left[\boldsymbol{\psi}(\mathbf{u}, \mathbf{u}^*)\mathbf{u}^\dagger - \mathbf{I}\right]\mathbf{W}. \tag{I.11}$$

Appendix J

Source and Channel Estimation Code

Listed below is a Matlab implementation of the CG algorithm using the JADE cost function and gradient derived in Chapter 7. The conjugate gradient optimization algorithm is based on the algorithm described in [2].

```
%========================================
function [Ahat,Shat,W]=cgjade(Yo,m);
%========================================
% Yo = mixed observed data
% N = number of sensors
% M = number of sources
% Ahat = estimated mixing matrix
% Shat = estimated sources
% W = estimated demixing matrix
n=m; N=length(X);
%========================================
% Whitening
%========================================
IWht=sqrtm((Yo*Yo')/N);
Wht=inv(IWht);   Y=Wht*Yo;
%=======================================
% Estimate Cumulants
%========================================
Q=estcum(Y,m,N);
%========================================
% Estimate Eigenmatrices
```

```
%========================================
Mhat=esteig(Q,m);
%========================================
% Conjugate Gradient search
%========================================
W=cgrad(Mhat,m);
Ahat=IWht*W;
Shat=W'*Y;
W=W'*Wht;  return
%========================================

%========================================
function W=cgrad(Mhat,m)
%========================================
% Based on algorithm described in
% abrudanmarch2009, abrudanmarch2008:
% Efficient Riemannian Algorithms for
% Optimization Under Unitary Matrix
% Constraint.
% Uses Armijo type step size from abrudanmarch2008:
% Steepest Descent Algorithms for Optimization
% Under Unitary Matrix Constraint
%========================================
asmax=20;
kmax=100;
tol=1e-12;
gprod=1;
W=eye(m);
mu=0.1;
k=0;
while (k<kmax)&&(gprod>tol)
if (mod(k,m*m)==0)
    gam=jadegrad(Mhat,W,m);
    G=gam*W'-W*gam';
    H=G;
end
    gprod=gg(G,G);
    if (gprod<tol)
```

```
       return;
    else
       P=expm(-mu*G);
       Q=P*P;
       as=0;
       while (jc(W,Mhat,m)-jc(Q*W,Mhat,m)>=mu*gg(G,G))
             P=Q;
             Q=P*P;
             mu=2*mu;
             as=as+1;if (as>=asmax),return;end
       end
       as=0;
       while (jc(W,Mhat,m)-jc(P*W,Mhat,m)<0.5*mu*gg(G,G))
             P=expm(-mu*G);
             mu=0.5*mu;
             as=as+1;if (as>=asmax),return;end
       end
       W=P*W;
       gam=jadegrad(Mhat,W,m);
       G0=G;
       G=gam*W'-W*gam';
       prgam=gg(G-G0,G)/gg(G0,G0);
       H=G+prgam*H;
       if gg(H,G)<0
          H=G;
       end
    end
    k=k+1;
end;  return
%========================================

%========================================
function g=gg(G1,G2)
%========================================
g=0.5*real(trace(G1'*G2));  return
%========================================

%========================================
```

```
function c=jc(W,Mhat,m)
%========================================
c=0;
for k=1:m:m*m
    Mh=Mhat(:,k:k+m-1);
    F=abs(W'*Mh*W).^2;
    f1=sum(sum(F));
    f2=trace(F);
    c=c+f1-f2;
end;  return
%========================================

%========================================
function G=jadegrad(Mhat,W,m)
%========================================
I=eye(m); G=zeros(m);
for k=1:m:m*m
    Mh=Mhat(:,k:k+m-1);
    g1=Mh*Mh'*W;
    g2=Mh*W*(I.*(W'*Mh'*W));
    g3=Mh'*W*(I.*(W'*Mh*W));
    G=G+g1-g2-g3;
end
G=2*G;  return
%========================================

%========================================
function Q=estcum(Y,m,N)
%========================================
% From "jade.m" by J.F. Cardoso, Nov. 1997
%========================================
R=(Y*Y')/N;
C=(Y*Y.')/N;
Yl=zeros(1,N);
Ykl=zeros(1,N);
Yjkl=zeros(1,N);
Q=zeros(m*m*m*m,1) ;
indx=1;
```

```
for lx=1:m; Yl=Y(lx,:);
for kx=1:m; Ykl=Yl.*conj(Y(kx,:));
for jx=1:m; Yjkl=Ykl.*conj(Y(jx,:));
for ix=1:m;
q1=(Yjkl*Y(ix,:).')/N;
q2=R(ix,jx)*R(lx,kx);
q3=R(ix,kx)*R(lx,jx);
q4=C(ix,lx)*conj(C(jx,kx));
Q(indx)=q1-q2-q3-q4;
indx=indx+1;
end;end;end;end;  return
%========================================

%========================================
function M=esteig(Q,m)
%========================================
% From "jade.m" by J.F. Cardoso, Nov. 1997
%========================================
[U,D]= eig(reshape(Q,m*m,m*m));
[la,K]=sort(abs(diag(D)));
M=zeros(m,m*m);
Z=zeros(m);
h=m*m;
for u=1:m:m*m,
Z(:)=U(:,K(h));
M(:,u:u+m-1)=la(h)*Z;
h=h-1;
end;  return
%========================================
```

Appendix K

MI Source and Channel Estimation Code

Listed below is a Matlab implementation of the CG algorithm using the MI cost function and gradient. The conjugate gradient optimization algorithm is based on the algorithm described in [2].

```
%================================================
function [pdfs,scores,SXS]=smf(X)
%================================================
% Est score function using Savitzky-Golay filter
%================================================
%
%================================================
dofigs=0;
i=sqrt(-1);
[m,n]=size(X);
if n>m
    X=X.';
end
cmplx=0;
if ~isreal(X);
   X=[real(X) imag(X)];
   cmplx=1;
end
[m,n]=size(X);
```

```
SXS=sort(X);
pdfs=zeros(m,n);
scores=zeros(m,n);
for kk=1:n
    SX=SXS(:,kk);
%=================================================
% histogram parameters
%=================================================
k=300; a=min(SX); b=max(SX);
%=================================================
% inter-bin distance
%=================================================
d=(b-a)/(k-3);
%=================================================
%
%=================================================
M=zeros(k,1);
M=linspace(a-d,b+d,k)';
C=zeros(k,1);
C=histc(SX,M)/m;
%=================================================
%
%=================================================
pord=51;
ford=9;
sord=ford-3;;
pdf=savgol(C,pord,1,0);
score=savgol(log(pdf),ford,2,1)/d;
pdf=pdf(sord:end-sord);
score=score(sord:end-sord);
M=M(sord:end-sord);
%=================================================
% interpolate for the density at points X
%=================================================
probdens=zeros(m,1);
probdens=interp1(M,pdf,SX,'cubic','extrap');
scr=zeros(m,1);
scr=interp1(M,score,SX,'cubic','extrap');
```

```
%==================================================
%
%==================================================
pdfs(:,kk)=probdens;
scores(:,kk)=scr;
end % kk
%==================================================
%
%==================================================
if cmplx
   n=floor(n/2);
   pdfs=pdfs(:,1:n)+i*pdfs(:,n+1:2*n);
   scores=scores(:,1:n)+i*scores(:,n+1:2*n);
   SXS=SXS(:,1:n)+i*SXS(:,n+1:2*n);
end
%==================================================
return
%==================================================


%==================================================
function [Ahat,Shat]=cgmi(X,A,scores,scsupport);
%==================================================
global scrs scsup
scrs=scores; scsup=scsupport;
[m,n]=size(X); i=sqrt(-1);
%==================================================
% Whitening
%==================================================
IWht=sqrtm((X*X')/n);
Wht=inv(IWht);
Y=Wht*X;
%==================================================
% Conjugate Gradient search
%==================================================
W=cgrad(Y);
Ahat=IWht*W;
Shat=W'*Y;
%==================================================
```

```
return
%================================================

%================================================
function g=gg(G1,G2)
%================================================
g=0.5*real(trace(G1'*G2));
%================================================
return
%================================================

%================================================
function c=micost(W,Y)
%================================================
Zr=real(W'*Y);
Zi=imag(W'*Y);
Z=[Zr;Zi];
%================================================
% Calls MIhigherdim.m by
% A. Kraskov, H. Stogbauer, and P. Grassberger,
% Estimating mutual information.
% Phys. Rev. E 69 (6) 066138, 2004
%================================================
c1=MIhigherdim(Z,1,1,1);
%================================================
c=max(0,c1);
%================================================
return
%================================================

%================================================
function G=migrad(W,Y)
%================================================
global scrs scsup
[m,n]=size(Y);
Z=W'*Y;
yscores=fitscore(Z,scrs,scsup);
G=-(yscores*Z'/n-eye(m))*W;
```

```
%=================================================
return
%=================================================

%=================================================
function W=cgrad(Y)
%=================================================
% Based on algorithm described in
% abrudanmarch2008:
% Efficient Riemannian Algorithms for Optimization
% Under Unitary Matrix Constraint.
% Uses Armijo type step size from abrudanmarch2008:
% Steepest Descent Algorithms for Optimization
% Under Unitary Matrix Constraint
%=================================================
[m,n]=size(Y);
asmax=10;
tol=1e-12;
gprod=1;
W=eye(m);
mu=0.1;
%=================================================
k=0;
while (gprod>tol)
if (mod(k,m*m)==0)
    gam=migrad(W,Y);
    G=gam*W'-W*gam';
    H=G;
end
    gprod=gg(G,G);
    if (gprod<tol) return;
    else
       P=expm(-mu*G);
       Q=P*P;
       as=0;
       while (micost(W,Y)-micost(Q*W,Y)>=mu*gg(G,G))
             P=Q;
             Q=P*P;
```

```
            mu=2*mu;
            as=as+1;if (as>=asmax),return;end
        end
        as=0;
        while (micost(W,Y)-micost(P*W,Y)<0.5*mu*gg(G,G))
            P=expm(-mu*G);
            mu=0.5*mu;
            as=as+1;if (as>=asmax), return; end
        end
        W=P*W;
        gam=migrad(W,Y);
        G0=G;
        G=gam*W'-W*gam';
        prgam=gg(G-G0,G)/gg(G0,G0);
        H=G+prgam*H;
        if gg(H,G)<0,H=G;end
    end % else
    k=k+1;
end % while
%================================================
return
%================================================


%================================================
function [yscores,SX]=fitscore(X,xscores,xsupport);
%================================================
[m,n]=size(X);
if n>m
    X=X.';
end
i=sqrt(-1);
cmplx=0;
if ~isreal(X);
   X=[real(X) imag(X)];
   xscores=[real(xscores) imag(xscores)];
   xsupport=[real(xsupport) imag(xsupport)];
   cmplx=1;
end
```

```
[m,n]=size(X);
yscores=zeros(size(X));
SX=zeros(size(X));
for k=1:n
    yscores(:,k)=interp1(xsupport(:,k),xscores(:,k),X(:,k),'linear','extrap');
end
if cmplx
   n=floor(n/2);
   yscores=0.5*(yscores(:,1:n)+i*yscores(:,n+1:2*n));
end
yscores=yscores.';
%=================================================
return
%=================================================
```

Appendix L

Copula Correlated Channel Code

Listed below is an Octave implementation of the technique for generating a correlated fading channel, described in Chapter 5.

```
%========================================================
function u=GenCopMIMO
%========================================================
% John Kitchen 16/July/2009
%
% calls "mvcoprnd.m" by Robert Kopocinski,
% Master Thesis: "Simulating dependent random variables using copulas.
% Applications to Finance and Insurance".
% matlab code available in "copula_functions.zip" from MatlabCentral
% http://www.mathworks.com/matlabcentral/fileexchange/15449
%========================================================
msize=2; % figure marker size
lwidth=2; % figure linewidth
i=sqrt(-1);
alpha=0;
rho=0;
omega=1;
nbin=200;
%fading='n';
%========================================================
disp('========================================================');
disp(['Copula families available are :']);
```

```
disp(['Clayton(C), Frank(F), Gumbel(G), Normal(N), Student-t(T)']);
disp('==========================================================');
disp(['Fading Distributions available are :']);
disp(['Rayleigh(R), Nakagami(N)']);
disp('==========================================================');
%disp(['Gaussian(G), Rayleigh(R), Nakagami(N)']);
%=======================================================
% Set Parameters
%=======================================================
family =  input('Choose Copula Family  [N] : ','s');
if (isempty(family))    family='n';       end
fading =  input('Choose Fading Distribution  [N] : ','s');
if (isempty(fading))    fading='n';       end
    N = input('Enter Number of Data Samples to Input      [1e4] :');
if (isempty(N))    N=1e4;      end
    M = input('Enter Array Size      [2] :');
if (isempty(M))    M=2;      end
switch lower(family)
       case 'c',
       alpha = input('Enter value for alpha (alpha>=0) [0] :');
       if (isempty(alpha)) alpha=0; end
       case 'f',
       alpha = input('Enter value for alpha (-infty<alpha<infty) [0] :');
       if (isempty(alpha)) alpha=0; end
       case 'g',
       alpha = input('Enter value for alpha (alpha>=1) [1] :');
       if (isempty(alpha)) alpha=1; end
       case 'n',
       rho = input('Enter rho  [eye(M)] :');
   if (isempty(rho)) rho=eye(M); end
       case 't',
       rho = input('Enter rho  [eye(M)] :');
   if (isempty(rho)) rho=eye(M); end
end
%=======================================================
% Verify Chosen Parameters. Quit if not happy!
%=======================================================
[msr,nsr]=size(rho);
```

```
vr=reshape(rho,1,msr*nsr);
disp('=======================');
disp('Your selection is  : ');
disp(['Copula Family     : ',family]);
disp(['Fading            : ',fading]);
disp(['Number of samples : ',num2str(N)]);
disp(['Array size        : ',num2str(M)]);
disp(['Alpha             : ',num2str(alpha)]);
disp(['Rho               : ',num2str(vr)]);
disp('=======================');
proceed='y';
proceed =  input('Do You Wish to Continue?    [(y)/n]  : ','s');
if (proceed~='y')& (proceed~='Y')
                     disp('No :-( ');
return;
else
                     disp('Yes! :-) ');
end
%=========================================================
%=========================================================
switch lower(family)
case 'n'
             cname='Gaussian'
             u1 = mvcoprnd(cname,rho,N,M);
             u2 = mvcoprnd(cname,rho,N,M);
  case 't'
             cname='T';
             dof=1;
             u1 = mvcoprnd(cname,rho,N,M,dof);
             u2 = mvcoprnd(cname,rho,N,M,dof);
case {'c', 'f', 'g'}
             switch lower(family)
               case 'c', cname='Clayton';
               case 'f', cname='Frank';
               case 'g', cname='Gumbel';
             end
             u1 = mvcoprnd(cname,alpha,N,M);
             u2 = mvcoprnd(cname,alpha,N,M);
```

```
otherwise
     error('Unrecognized copula type: ''%s''',family);
end
%=========================================================
%=========================================================
switch lower(fading)
       case 'n'
           fname='Nakagami';
           X1 = gaminv(u1,M/2,omega/M); X1=sqrt(X1);
           X2 = gaminv(u2,M/2,omega/M); X2=sqrt(X2);
       case 'r'
           fname='Rayleigh';
           X1 = raylinv(u1,omega/M);
           X2 = raylinv(u2,omega/M);
end
corrcoef(X1,X1)
corrcoef(X2,X2)
corrcoef(X1,X2)
%=========================================================
%=========================================================
Y11=sign(randn(N,1)).*X1(:,1);
Y21=sign(randn(N,1)).*X2(:,1);
Y12=sign(randn(N,1)).*X1(:,2);
Y22=sign(randn(N,1)).*X2(:,2);
Ya=Y11+i*Y21;
Yb=Y12+i*Y22;
ANGYa=angle(Ya); AYa=abs(Ya);
ANGYb=angle(Yb); AYb=abs(Yb);
[naa,ctraa] = hist(ANGYa,nbin);
[nampa,ctrampa] = hist(AYa,nbin);
[nab,ctrab] = hist(ANGYb,nbin);
[nampb,ctrampb] = hist(AYb,nbin);
npa=nakphase(ctraa,M);
npb=nakphase(ctrab,M);
nakampa=nakpdf(ctrampa,M,omega);
nakampb=nakpdf(ctrampb,M,omega);
%=========================================================
%=========================================================
```

```
[n1,ctr1] = hist(X1(:,1),nbin);
[n2,ctr2] = hist(X2(:,1),nbin);
n1max=max(n1); n1=n1/n1max;
n2max=max(n2); n2=n2/n2max;
x1max=max(X1(:,1));
x2max=max(X1(:,2));
x1min=min(X1(:,1));
x2min=min(X1(:,2));
[p1,nvar]=nakpdf(ctr1,M/2,omega/2); p1=p1/max(p1);
[p2,nvar]=nakpdf(ctr2,M/2,omega/2); p2=p2/max(p2);
%=========================================================
%=========================================================
switch lower(family)
case {'c', 'f', 'g'}
        parext=['Alpha = ',num2str(alpha)];
    case {'n','t'}
        parext=['Rho = [',num2str(vr),']'];
end
titletext=['Copula=',cname,', Fading=',fname,', N=',num2str(N),' , ',parext];
%=========================================================
%=========================================================
figure;
subplot(2,2,2);
      plot(X1(:,1),X2(:,1),'*','markersize',msize);
      axis([x1min x1max x2min x2max]);
      h1 = gca;
      xlabel(['X1  :  ',fname]);
      ylabel(['X2  :  ',fname]);
      title(titletext);
subplot(2,2,4);
      plot(ctr1,n1,'b','linewidth',lwidth); hold on;
      plot(ctr1,p1,'g','linewidth',lwidth);
      axis([x1min x1max 0 max(n1)*1.1]);
      h2 = gca;
subplot(2,2,1);
      plot(-n2,ctr2,'b','linewidth',lwidth); hold on;
      plot(-p2,ctr2,'g','linewidth',lwidth);
      axis([-max(n2)*1.1 0 x2min x2max]);
```

```
      h3 = gca;
set(h1,'Position',[0.35 0.35 0.55 0.55]);
set(h2,'Position',[.35 .05 .55 .15]);
set(h3,'Position',[.1 .35 .15 .55]);
colormap([.8 .8 1]);
%=========================================================
%=========================================================
[n1,ctr1] = hist(AYa,nbin);
[n2,ctr2] = hist(AYb,nbin);
n1max=max(n1); n1=n1/n1max;
n2max=max(n2); n2=n2/n2max;
x1max=max(AYa);
x2max=max(AYb);
x1min=min(AYa);
x2min=min(AYb);
p1=nakpdf(ctr1,M,omega);
p2=nakpdf(ctr2,M,omega);
p1=p1/max(p1);
p2=p2/max(p2);
%=========================================================
%=========================================================
figure;
subplot(2,2,2);
      plot(AYa,AYb,'*','markersize',msize);
      axis([x1min x1max x2min x2max]);
      h1 = gca;
      xlabel(['Abs(a)  :  ',fname]);
      ylabel(['Abs(b)  :  ',fname]);
      title(titletext);
subplot(2,2,4);
      plot(ctr1,n1,'b','linewidth',lwidth); hold on;
      plot(ctr1,p1,'g','linewidth',lwidth);
      axis([x1min x1max  0 max(n1)*1.1]);
      h2 = gca;
subplot(2,2,1);
      plot(-n2,ctr2,'b','linewidth',lwidth); hold on;
      plot(-p2,ctr2,'g','linewidth',lwidth);
      axis([-max(n2)*1.1 0 x2min x2max]);
```

```
      h3 = gca;
set(h1,'Position',[0.35 0.35 0.55 0.55]);
set(h2,'Position',[.35 .05 .55 .15]);
set(h3,'Position',[.1 .35 .15 .55]);
colormap([.8 .8 1]);
%=========================================================
%=========================================================
[n1,ctr1] = hist(ANGYa,nbin);
[n2,ctr2] = hist(ANGYb,nbin);
n1max=max(n1); n1=n1/n1max;
n2max=max(n2); n2=n2/n2max;
x1max=max(ANGYa);
x2max=max(ANGYb);
x1min=min(ANGYa);
x2min=min(ANGYb);
p1=nakphase(ctr1,M);
p2=nakphase(ctr1,M);
p1=p1/max(p1);
p2=p2/max(p2);
%=========================================================
%=========================================================
figure;
subplot(2,2,2);
      plot(ANGYa,ANGYb,'*','markersize',msize);
      axis([x1min x1max x2min x2max]);
      h1 = gca;
      xlabel(['Angle(a)  :  ',fname]);
      ylabel(['Angle(b)  :  ',fname]);
      title(titletext);
subplot(2,2,4);
      plot(ctr1,n1,'b','linewidth',lwidth); hold on;
      plot(ctr1,p1,'g','linewidth',lwidth);
      axis([x1min x1max 0 max(n1)*1.1]);
      h2 = gca;
subplot(2,2,1);
      plot(-n2,ctr2,'b','linewidth',lwidth); hold on;
      plot(-p2,ctr2,'g','linewidth',lwidth);
      axis([-max(n2)*1.1 0 x2min x2max]);
```

```
      h3 = gca;
set(h1,'Position',[0.35 0.35 0.55 0.55]);
set(h2,'Position',[.35 .05 .55 .15]);
set(h3,'Position',[.1 .35 .15 .55]);
colormap([.8 .8 1]);
%========================================================
%========================================================

%========================================================
function np=nakphase(theta,m)
%========================================================
% John Kitchen 26/June/2009
%========================================================
% see Yacoub et al "Nakagami-m phase-envelope
% joint distribution"
% IEE Electronics Letters, vol.41, pp.259-261, March 2005.
%========================================================
%========================================================
p1=gamma(m)*abs(sin(2*theta)).^(m-1);
p2=2^m*gamma(m/2)*gamma(m/2);
np=p1./p2;
%========================================================
return
%========================================================
```

Appendix M

Blind Source Separation Algorithms

Algorithm 1 RADICAL Algorithm

Pseudocode for RADICAL based on Learned-MIller and Fisher [58].
Robust, **A**ccurate, **D**irect **ICA** a**L**gorithm (RADICAL)

Require: $\mathbf{Y}$ is the $p \times n$ observation matrix.
Model is $\mathbf{Y} = \mathbf{AX}$.
$\mathbf{A}$ is an unknown $m \times p$ full rank matrix.
$\mathbf{X}$ is a $p \times p$ source matrix.
For each k, components of $\mathbf{X}(:, k)$ are statistically independent.
For each i, $\mathbf{X}(i, :)$ is a zero-mean source signal.
$s =$ spacing size.
$a =$ determines angular resolution for Jacobi rotations.

procedure WHITENING($\mathbf{Y}$)
    Whiten the observation data
    $\mathbf{R_y} = \frac{1}{n}\mathbf{Y}\mathbf{Y}^\dagger$
    EVD: $\mathbf{R_y} = \mathbf{EDE}^\dagger$
    Whitening matrix: $\mathbf{\Omega} = \mathbf{D}^{-1/2}\mathbf{E}^\dagger$
    $\mathbf{Z} = \mathbf{\Omega Y}$
end procedure

procedure JACOBI ROTATIONS($\mathbf{Z}$)
    $\mathbf{V} = \mathbf{I}_p$
    for all pairs (i, j) **do**
        find $2-$D Jacobi rotation for $Z(i, :), Z(j, :)$
        s.t. $\theta^* = \arg\min($spacings-entropy$(\theta))$
        $\mathbf{V} = \mathbf{V} \times$ 2-D-Rotation(θ^*)
    end for
end procedure

procedure ESTIMATES($\mathbf{V}$,$\mathbf{\Omega}$,$\mathbf{Y}$)
    $\hat{\mathbf{X}} = \mathbf{V}^\dagger\mathbf{\Omega Y}$
    $\hat{\mathbf{A}} = \mathbf{\Omega}^{-1}\mathbf{V}^\dagger$
end procedure

Algorithm 2 JADE Algorithm

Pseudocode for JADE based on Cardoso and Souloumiac [19].

Require: $\mathbf{Y}$ is the $m \times n$ observation matrix
Model is $\mathbf{Y} = \mathbf{AX} + \mathbf{W}$.
$\mathbf{A}$ is an unknown $m \times p$ full rank matrix.
$\mathbf{X}$ is a $p \times n$ source matrix.
For each k, components of $\mathbf{X}(:, k)$ are statistically independent.
For each i, $\mathbf{X}(i, :)$ is a zero-mean source signal.
At most one source has a vanishing 4^{th}-order cumulant.
W is a $m \times n$ matrix of spatially white noise. $\mathbf{\Sigma_w} = \sigma_w^2 \mathbf{I}_m$

procedure WHITENING($\mathbf{Y}$)
    Whiten the observation data
    $\mathbf{R_y} = \frac{1}{n}\mathbf{YY}^\dagger$
    EVD: $\mathbf{R_y} = \mathbf{EDE}^\dagger$
    Whitening matrix $\mathbf{\Omega} = \mathbf{D}^{-1/2}\mathbf{E}^\dagger$
    $\mathbf{Z} = \mathbf{\Omega Y}$
end procedure
procedure CUMULANTS($\mathbf{Z}$)
    Estimate the set of cumulant matrices from the whitened observations
    $\mathbf{Q_z} = Cum\left(z_i, z_j^*, z_k^*, z_l\right)$
    EVD: $\mathbf{Q_z} = \mathbf{EDE}^\dagger$
    find p most significant eigen pairs: $\left\{\hat{\lambda}_r, \hat{\mathbf{M}}_r | 1 \le r \le p\right\}$
    set $\mathcal{N} = \left\{\hat{\lambda}_r \hat{\mathbf{M}}_r | 1 \le r \le p\right\}$
end procedure
procedure DIAGONALIZATION($\mathcal{N}$)
    Jointly diagonalize the set $\mathcal{N}$ by a unitary matrix $\mathbf{V}$,
    equivalent to finding $\mathbf{V} = \arg\min \sum_i \mathrm{Off}\left(\mathbf{V}^\dagger \mathbf{Q}_{\mathbf{z}_i} \mathbf{V}\right)$
    repeat
        for each pair of rows i and j, $i \neq j$ **do**
            Find the Jacobi rotation that will minimize the sum of the ij and ji elements in all cumulant matrices
            if the rotation angle is above some threshold **then**
                perform the rotation
            end if
        end for
    until no rotations or maximum iterations performed
    The unmixing matrix is the product of all the Jacobi rotations performed
end procedure
procedure ESTIMATES($\mathbf{V}$,$\mathbf{\Omega}$,$\mathbf{Y}$)
    $\hat{\mathbf{X}} = \hat{\mathbf{V}}^\dagger \mathbf{\Omega Y}$
    $\hat{\mathbf{A}} = \mathbf{\Omega}^+ \mathbf{V}^\dagger$
end procedure

Algorithm 3 FASTICA Algorithm

Pseudocode for FASTICA based on Bingham and Hyvärinen [16].

Require: $\mathbf{Y}$ is the $p \times n$ observation matrix
Model is $\mathbf{Y} = \mathbf{AX}$.
$\mathbf{A}$ is an unknown $p \times p$ full rank matrix.
$\mathbf{X}$ is a $p \times n$ source matrix.
For each k, components of $\mathbf{X}(:, k)$ are statistically independent.
For each i, $\mathbf{X}(i, :)$ is a zero-mean source signal.
At most one source has a vanishing 4^{th}-order cumulant.

procedure WHITENING($\mathbf{Y}$)
    Whiten the observation data
    $\mathbf{R_y} = \frac{1}{n}\mathbf{YY}^\dagger$
    EVD: $\mathbf{R_y} = \mathbf{EDE}^\dagger$
    Whitening matrix $\mathbf{\Omega} = \mathbf{D}^{-1/2}\mathbf{E}^\dagger$
    $\mathbf{Z} = \mathbf{\Omega Y}$
end procedure

procedure FIXED POINT ICA($\mathbf{Z}$)
    Initialise: $\mathbf{V} = \mathbf{I}$, $count = 0$
    repeat
        for $k = 1$ to p **do**
            $\mathbf{v} = \mathbf{V}(:, k)$
            $\mathbf{b} = \mathbf{v}^\dagger \mathbf{z}$
            $\mathbf{V}(:, k) = \mathbb{E}\left\{\mathbf{z}\mathbf{b}^* g(|\mathbf{b}|^2)\right\} - \mathbb{E}\left\{g(|\mathbf{b}|^2) + |\mathbf{b}|^2 g'(|\mathbf{b}|^2)\right\}\mathbf{v}$
        end for
        $\mathbf{V} = \mathbf{V}(\mathbf{V}^\dagger\mathbf{V})^{-1/2}$ ▷ Symmetric decorrelation
        count=count+1
    until (count>maxcount) or $\mathbf{V}$ converges
end procedure

procedure ESTIMATES($\mathbf{V}$,$\mathbf{\Omega}$,$\mathbf{Y}$)
    $\hat{\mathbf{X}} = \mathbf{V}^\dagger\mathbf{\Omega Y}$
    $\hat{\mathbf{A}} = \mathbf{\Omega}^{-1}\mathbf{V}^\dagger$
end procedure