A study on the application of independent component analysis to *in vivo* ^{1}H magnetic resonance spectra of childhood brain tumours for data processing

by

Jie Hao

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Electronic, Electrical and Computer Engineering
College of Engineering and Physical Sciences
The University of Birmingham
31st March 2010

ABSTRACT

Independent component analysis (ICA) has the potential of automatically determining metabolite, macromolecular and lipid (MMLip) components that make up magnetic resonance (MR) spectra. However, the realiability with which this is accomplished and the optimal ICA approach for investigating *in viv*o MR spectra, have not yet been determined.

A wavelet shrinkage de-noising based enhancement algorithm, utilising a newly derived relationship between the real and imaginary parts of the MR spectrum, is proposed. This algorithm is more robust compared with conventional de-noising methods.

The two approaches for applying ICA, blind source separation (BSS) and feature extraction (FE), are thoroughly examined. A feature dimension selection method, which has not been adequately addressed, is proposed to set a theoretical guideline for ICA dimension reduction.

Since the advantages and limitations of BSS-ICA and FE-ICA are different, combining them may compensate their disadvantages and lead to better results. A novel ICA approach involving a hybrid of the two techniques for automated decomposition of MRS dataset is proposed. It has been demonstrated that hybrid ICA provides more realistic individual metabolite and MMLip components than BSS-ICA or FE-ICA. It can aid metabolite identification and assignment, and has the potential for extracting biologically useful features and discovering biomarkers.

ACKNOWLEDGEMENTS

I would like to express my love and gratitude to all of my family for their unconditional love and support throughout my research. I would especially like to thank Zou Xin, who was always there for me during the hard and happy times.

I would like to thank all my colleagues from the Electrical, Electronic and Computer Engineering and Birmingham Children's Hospital. I would especially like to thank my supervisors, Dr Theo Arvanitis and Dr Andrew Peet for their invaluable assistance and patience during this study. I would also like to thank Nigel Davies for much advice and collecting all the data for this thesis. I would like to thank Martin Wilson and Greg Reynolds for providing the simulator and suggestions on many parts of this study. Additional thanks goes to Yu Sun, John Easton, Tristan Payne and Markus Saleh. Additional thanks goes to the clinicians who have helped in this project and particularly to Dr Lesley MacPherson (Radiology Department at Birmingham Children's Hospital) and Rachel Grazier (data manager, Birmingham Children's Hospital).

I'd like to thank the ORSAS for funding the three years of this work.

I'd also like to thank Nigel Davies, Alex Upton, David and Valerie Hornsby for proofreading my thesis draft.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Ala	Alanine
BSS	Blind source separation
CHESS	Chemical shift selective
Cho	Choline
Cr	Creatine
CRLB	Cramer–Rao lower bounds
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
emf	Electromotive force
FE	Feature extraction
FFT	Fast Fourier transform
FID	Free induction decay
FT	Fourier transform
FWHM	Full width at half maximum
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
GPC	Glycerophosphorylcholine
Gua	Guanidoacetate
IC	Independent component
ICA	Independent component analysis
IFT	Inverse Fourier transform
Lac	Lactate

LCModel	Linear combination model
LDA	Linear discriminant analysis
m-Ins	Myo-Inositol
MMLip	Macromolecular and lipid
MR	Magnetic resonance
MRA	Multiresolution Analysis
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
MRSI	Magnetic resonance spectroscopic imaging
MSE	Mean-square error
NAA	N-acetyl aspartate
NAAG	N-acetylaspartylglutamate
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PCh	Phosphorylcholine
pdf	Probability density function
ppm	Parts per million
PRESS	Point resolved spectroscopy
RF	Radio frequency
RMS	Root mean square
SNR	Signal-to-noise ratio
STEAM	Stimulated echo acquisition mode
SURE	Stein's Unbiased Risk Estimate
SVD	Singular value decomposition
SVM	Support vector machine
Tau	Taurine

tCho	Total choline
TE	Echo time
TM	Mixing time
TR	Repetition Time
WSD	Wavelet shrinkage de-noising

LIST OF SYMBOLS

Symbol	Meaning
$\mathbf{A}$	The mixing matrix
$\mathbf{A}^{BSS}$	The superscription ^{BSS} means that matrix **A** is used in BSS-ICA
$\mathbf{A}^{FE}$	The superscription ^{FE} means matrix **A** is used in FE-ICA
$\mathbf{A}_{L\times M}$	The matrix **A** with dimension $L\times M$
$\mathbf{B}$	The unmixing matrix
$\mathbf{B}_0$	The strong magnetic field
$\mathbf{B}_1$	The weaker oscillating magnetic field
$\mathbf{B}_{\text{eff}}$	The effective field
ΔB	The inhomogeneities of the magnetic field
$\mathbf{C}_{\mathbf{x}}$	The covariance matrix of $\mathbf{x}$.
$\mathbf{D}$	The diagonal matrix of the eigenvalues
$\mathbf{E}$	The orthogonal matrix with the column vectors being eigenvectors
ΔE	The energy difference between adjacent energy levels
$G^1(\cdot)$	The odd function,
$G^2(\cdot)$	The even function
H(x)	The Entropy of x
h	Planck's constant
I	The spin quantum number
J_{MSE}^{PCA}	The value of the minimum MSE of PCA
$\boldsymbol{J}$	The spin-spin coupling constant
$\mathbf{J}$	The Jacobian matrix

$J(\mathbf{y})$	The negnetropy of y
J_{MSE}^{PCA}	The value of the minimum MSE of PCA
k	The Boltzmann constant
$\mathbf{M}$	The magnetization
$\mathbf{M}_{xy,0}$	The initial transverse component of $\mathbf{M}$
$\mathbf{M}_z$	The longitudinal component of the magnetization
$\mathbf{M}_{xy}$	The transverse component of the magnetization
m_{I}	The magnetic quantum number of I
$\mathbf{m}_{\mathbf{x}}$	The mean vector of $\mathbf{x}$
$p_x(x)$	The probability density function of x
$\mathbf{R}$	The set of real numbers
S_0	The maximum signal amplitude
$S_x(t)$	The real part of a complex signal $S(t)$
$S_y(t)$	The imaginary part of a complex signal $S(t)$
SNR_{E}	A general definition of SNR as the ratio of the clean signal power over noise power
SNR_{LC}	The LC Model defined SNR
T_1	The spin-lattice relaxation time
T_2	The spin-spin relaxation time
T_2^*	The effective decay time
$\mathbf{V}$	The whitening matrix
v	The precession frequency in Hertz
v_{ref}	The prcession frequency of the reference molecule
$\mathbf{W}$	The orthonormal ICA matrix

γ	The gyromagnetic ratio
γ_s	The step size in ICA estimation
θ	The percentage threshold
κ_k	The k^{th} cumulant of x
$\mathbf{\Lambda}$	An invertible diagonal matrix with the diagonal components $\lambda_i \neq 0$.
$\boldsymbol{\mu}$	The magnetic moment
σ	The shielding constant
σ_{ref}	The shielding constant of the reference molecule
ϕ	The phase shift
$\phi(t)$	The father wavelet
$\psi(t)$	The mother wavelet function
ω_L	The Larmor frequency
ω_{RF}	The rotating frequency of $\mathbf{B}_1$ around the z axis.
ω_0	The frequency with which $\mathbf{M}_{xy}$ appears to precess around the original point in the rotating frame

CHAPTER 1

INTRODUCTION

Magnetic resonance spectroscopy (MRS) has been an important analytical tool in organic chemistry, biology, and materials science for more than a half-century (Salibi and Brown, 1998). MRS applied to living animals and human beings is called *in vivo* MRS, and can obtain metabolic information about various tissues in a non-invasive way. ^{1}H MRS is a clinically useful diagnostic tool for the *in vivo* assessment of brain tumours (Preul et al., 1996, Burtscher and Holtas, 2001, Howe and Opstad, 2003). Therefore, it can be used to complement the magnetic resonance imaging (MRI) and computed tomography (CT) information to give an additional dimension to predict the tumour aggressiveness (Ketonen et al., 2004). Many studies on adults have shown a high accuracy in the non-invasive diagnosis of brain tumours by MRS and evidence is emerging that the technique is also valuable in children (Peet et al., 2008).

^{1}H MRS is based on a well known property that a proton will resonate at a slightly different frequency depending on its molecular environment (Reiser et al., 2008). This phenomenon, known as ‘chemical shift’, is due to the fact that electrons surrounding a proton produce a

shielding effect, which influences the local magnetic field experienced by the proton. The concentrations of certain metabolites or macromolecular and lipid (MMLip) components can be calculated by measuring the area under the corresponding peaks in a magnetic resonance (MR) spectrum. As the proton has a high MR sensitivity and a high natural abundance, a large number of brain metabolites containing protons can generate signals at specific frequencies in the MR spectrum.

The appearance of the MR spectra varies depending on the patient's clinical presentation and the various metabolites' concentrations. Although there exists a high degree of variability in human brain tumour types, some general characteristic differences between normal and brain tumour tissues have been reported (Govindaraju et al., 2000). Several metabolite concentrations have been found to be decreased or increased when measured in spectra of tumours compared to those within normal tissue (Reiser et al., 2008, Panigrahy et al., 2006, Barker and Lin, 2006, Gillard et al., 2004, Moreno-Torres et al., 2004). Furthermore, this change in concentration can be used for specific tumour determination. In tumour diagnosis, accurate determination of the grade of malignancy is one of the most important factors for some types of tumours. Generally, fast-growing tumours are much more likely to emerge again after treatment than slow-growing tumours (Pizzo and Poplack, 2006). On the other hand, high-grade tumours do respond better to radiotherapy and chemotherapy (Pizzo and Poplack, 2006). Certain metabolite and MMLip components are considered to be metabolic markers of tumour malignancy (Howe et al., 2003).

The quantification of spectral peaks plays an important role in the diagnostic capacity of MRS compared to MRI. The reason is that MRI relies on the detection of spatial or tissue abnormalities as a result of disease conditions, while MRS reveals the differences in relative

proportions of metabolites peaks at a tissue voxel (volume of interest). However, metabolite concentration measurement in *in vivo* MRS is complicated by resonance overlap, baseline distortions, non-Lorentzian lineshapes and low signal-to-noise ratio (SNR) problems. It is hard for an untrained user to interpret these spectra visually. There is a need for an optimal data processing technique for ^{1}H MR spectra to aid the characterisation of spectra and, therefore, support the identification of useful diagnostic and prognostic biomarkers for tumours.

Various methods have been used to estimate the concentrations, ranging from simple integration of the spectra to fitting algorithms in the time or frequency domains (Joliot et al., 1991). Although there are a number of methods available for analysing spectra, most of these still need interaction by the user in order to determine which metabolites are present in the spectrum. A commonly used method to decompose the MR spectrum to its individual components is by fitting the spectrum to some predetermined basis set of simulated or measured metabolites and MMLip components, such as LCModel (Provencher, 1993, Provencher, 2001, Provencher, 2009), AQSES (Poullet et al., 2007) and QUEST (Ratiney et al., 2005). These methods, however, require a prior knowledge of the MR spectrum to choose which basis set spectra is to be included and to set the fitting parameters. The residual, which cannot be fitted by the basis set, is treated as noise (Provencher, 2009).

A recent study showed that fitting spectra to yield metabolite concentrations and using these in a linear discriminant analysis (LDA) produced better results than pattern recognition performed directly on the spectra (Opstad et al., 2007). Highly accurate results were obtained when a similar approach was applied to MRS of childhood cerebellar tumours (Davies et al., 2008). However, these approaches suffer from the disadvantage that metabolites will be

quantitated only if they were included in the basis set used to determine the metabolite concentrations. Furthermore, other metabolites will be overestimated to account for signals which have not been included. Evidence that this is a practical limitation of the method is available from studies, which have identified metabolites that are not in commonly used basis sets (Opstad et al., 2003). A definitive assignment of metabolite signals is not always possible from *in vivo* MRS (Panigrahy et al., 2006), a fact that hampers the construction of more complete basis sets.

An alternative approach, which would automatically decompose a dataset of spectra into their metabolite signals, could be a potentially useful advance in the classification of tumours based on their MRS profiles. Independent component analysis (ICA) (Hyvärinen et al., 2001, Hyvarinen and Oja, 2000) is a statistical technique that reveals hidden factors within a dataset of signals and requires no prior knowledge of that dataset. This makes ICA a strong candidate for automated decomposition of MR spectra.

There are two main approaches to the application of ICA, and it is important to compare these methods in order to understand their strengths and discover their limitations. Originally, ICA was developed to solve the blind source separation (BSS) (Cardoso, 1998) problem. In this problem, the method separates the unknown independent sources of signals from their experimentally observed mixtures. In parallel to the study of BSS, ICA can also be applied for extracting specific features (Feature Extraction - FE) from observed signals (Hoyer and Hyvärinen, 2000). The latter approach achieves dimensionality reduction, where features are extracted from the observed data for compression, de-noising and pattern recognition purposes (Guyon, 2006). Since the ^{1}H MRS signal can be considered as a linear combination

of metabolite components and noise (Keeler, 2005), ICA may provide an automated method for identifying these signals.

Tumours that occur in children are mainly different to those occurring in the adult population, with some tumour types being unique to the paediatric population. These are usually classified by the WHO classification system of 2007 (Louis et al., 2007). Astrocytomas are the most common type of glioma in both adults and children. They develop from cells called astrocytes. Astrocytomas can vary in aggression and prognosis, ranging from slow growing to very fast growing, highly invasive. In children, most of these tumours are considered low-grade, while in adults most are high-grade (Burkhard et al., 2009). Medulloblastoma is the most common primary malignant central nervous system tumour arising in childhood. It accounts for 20-25% percent of paediatric brain tumours (Paulino and Melian, 1999). It is a highly invasive embryonal neuroepithelial tumour that arises in the cerebellum and has a tendency to disseminate throughout the CNS early in its course. Ependymomas are the third most common type of brain tumour in children (Pizzo and Poplack, 2006). It account for 5-10% of paediatric brain tumours. The underlying cause and basic biology of these tumours is not fully understood, and the optimal staging and treatment of ependymomas in children requires more study.

Childhood brain tumours are highly variable and particularly difficult to diagnose using MRI alone, and the MRS has lots of potential to provide addition useful information for this application. Although MRS has been investigated for the diagnosis of adult brain tumours, little work has been done on the clinically important area of the paediatric case. Therefore, the techniques investigated in this thesis will facilitate the use of MRS for the analysis and diagnosis of childhood brain tumour. Nevertheless, the techniques developed in this thesis are

not limited to childhood brain tumour. They should be applicable to the general clinical MRS data with certain threshold adjustments depending on cases.

1.1 Aim and objectives

Although the subjects of MRS data analysis (Henriksen, 1995, Mierisova and Ala-Korpela, 2001, Barker et al., 2001) and ICA have been researched for decades, there are very few studies using ICA on MRS data for automated metabolite and MMLip components decomposition. It is hoped that this research will provide a thorough investigation on this topic.

The previous studies of the ICA technique on *in vivo* MRS datasets did not clearly determine under what conditions ICA could extract individual metabolite and MMLip components, and, therefore, did not fully expose what the precise benefits of ICA were. The aim of the work described in this thesis is to investigate the ability of ICA for the automated decomposition of *in vivo* MRS datasets, and to realise its full potential as a supportive biomarker identification tool for diagnostic and prognostic purposes within the clinical setting.

Four objectives are set for this study. The first objective is to reduce the noise effect on the MR spectra before the ICA decomposition. The MR spectra are usually corrupted by noise from various sources, hence de-noising should be performed as a pre-processing step of the data analysis. Noise reduction based on the mathematical property of Lorentzian lineshape and wavelet shrinkage de-noising (WSD) is investigated.

The second objective is to search for a guideline for the feature dimension reduction in the ICA method. The actual number of dimensions of a dataset is usually unobservable, but a

truthful estimate is important for the ICA process. For MRS datasets, the dimension being reduced to should be the same as the number of individual metabolite and MMLip components contained in that dataset. A good estimate of the actual dimension in the dataset is the starting point for a reliable ICA decomposition.

The third objective is to investigate the ability of ICA when applied to varied *in vivo* MR spectra. It is essential to have a sound knowledge of the ability of ICA to perform the automated decomposition of MRS data and the conditions under which it is likely to fail. Since a number of factors can potentially influence the performance of ICA, it is important to systematically evaluate each one individually.

The fourth and final objective of this research is to develop a new and more efficient ICA-based method for the automatic decomposition of the MR spectra into meaningful metabolite and MMLip components. Automation is not only important for the processing of a large dimensional dataset, but also for ensuring that results are reproducible and independent of a user and the bias of his prior knowledge of the character of the spectra. As ICA requires no prior information and knowledge about the observed dataset, it is hoped that this new method can reveal some hidden information about the MR spectra.

1.2 Thesis original contributions and organisation

This thesis is divided into nine chapters. The original contributions arise mainly in chapters five, six, seven and eight.

The problem of how many ICs should be chosen in an ICA application (for an MRS dataset or any other type of datasets) has not been adequately addressed in previous ICA studies. Part of

Chapter five proposes a feature dimension reduction criterion for the ICA method. The overall SNR in the MRS dataset is used to estimate the percentage of the effective information content in the noisy dataset; hence the number of ICs to be used in the ICA process can be obtained.

Chapter six is devoted to a more robust de-noising of the MR spectra for the purpose of supporting the ICA application. In previous studies, only the real part of the MR spectrum was analysed, the imaginary part was usually ignored. In this chapter, a WSD-based MRS enhancement algorithm is proposed, where both the real and imaginary parts of MR spectrum are employed. First, the relationship between the real and imaginary parts of the MR spectrum is derived, where the real part can be calculated from the imaginary part, and vice versa. Then, the WSD is applied on the imaginary part of the spectrum and an estimate of the real part is derived from the enhanced imaginary part according to the relationship between them. The final de-noised real part of the MR spectrum is obtained by averaging the directly WSD-enhanced real part and the estimate obtained from the de-noised imaginary part. The noise reduction by combining the derived relationship and WSD has shown improvement in the de-noised SNR value compared to the results by applying WSD on the free induction decay (FID) signal and on the real part of the MR spectrum. This work has been presented in the 20th Annual British Chapter of the ISMRM MR Symposium for Post-docs and PhD students 2010 (Hao et al., 2010).

Chapter seven provides a systematic comparison of the ability of ICA on both simulated and experimental *in vivo* ^{1}H MRS datasets of childhood brain tumours. The study compared the performance of two ICA-based algorithms, FE and BSS algorithms. It was found that the FE-based ICA method was limited in that a combination of metabolite and MMLip components

commonly appear in the same IC, and a large sample size is required, whereas most of the problems with the BSS-based ICA method were caused by the overlapping peaks. This work has been presented in the 13th Annual Meeting of the British Chapter of the ISMRM 2008 (Hao et al., 2007) and published in the Journal of NMR in Biomedicine (Hao et al., 2009b).

Chapter eight relates to the development of a novel hybrid ICA method for the automatic decomposition of *in vivo* ^{1}H MR spectra of childhood brain tumours. This approach combines the advantages of both BSS-ICA and FE-ICA techniques. Experiments were performed on synthesised and experimental MRS datasets. The hybrid ICA method showed obvious improvements in its ability to decompose spectra compared with that of the FE-ICA or BSS-ICA methods, with an increased correlation between the obtained ICs and simulated metabolite and MMLip components. The results demonstrated that the proposed hybrid ICA method provided more realistic individual metabolite and MMLip components than the BSS-ICA or FE-ICA methods alone for both synthesised and experimental datasets. It can aid metabolite identification and assignment. The hybrid ICA method has the potential for extracting biologically useful features and discovering biomarkers. This work is currently under peer review to the Journal of NMR in Biomedicine and part of the work has been presented in European Society for Magnetic Resonance in Medicine and Biology (ESMRMB) 2008 Congress (Hao et al., 2008) and ISMRM 17th Scientific Meeting and Exhibition 2009 (Hao et al., 2009a).

The rest of the thesis is organised as follows.

Chapter two introduces the concepts of magnetic resonance and its applications in living systems. After a brief explanation of the physical principles that govern the technique of MRS, the information content of ^{1}H MR spectra of the brain is presented. The potential benefits of

MRS in brain tumour diagnosis are outlined. The last section of this chapter contains a short review of the commonly used MRS metabolite component quantitation tool, namely the LCModel (Provencher, 2009).

Chapter three discusses the two main approaches in solving the linear transformation problem in MRS, namely the second-order and higher-order statistical methods. The limitations of the second-order statistical methods, when decomposing the MR spectra, are addressed. The higher-order statistical methods make use of the higher-order information, which were ignored by the second-order statistical methods, and is considered to present a closer modelling of the MRS data.

Chapter four reviews the theoretical and mathematical background of the ICA method. First, the derivation of the ICA technique is introduced, and then the assumptions and uncertainties of the ICA model are discussed. The relationship between the ICA and the well known second-order statistical method, principal component analysis (PCA), is also explained. Several representative ICA algorithms are briefly explained. The applications of ICA in MRS are reviewed at the end of this chapter.

The rest of Chapter five addresses the materials and methods used for the completion of the research work discussed in this thesis. The different implementation aspects of FE- and BSS-based ICA methods for the MRS dataset are presented. The details of constructing the simulated datasets and collecting the experimental dataset used in the experiments of this thesis are described.

Chapter nine summarises this thesis and provides some concrete conclusions. Suggestions are given for possible future extension of the research, in the context of general digital signal

processing and for its application to MRS, specifically.

CHAPTER 2

MAGNETIC RESONANCE SPECTROSCOPY

2.1 Introduction

Nuclear magnetic resonance (NMR), or magnetic resonance (MR)[1] in short, is a phenomenon that relies on the magnetic properties of the atomic nucleus (Bloch, 1946, Purcell et al., 1946). Magnetic resonance spectroscopy (MRS) (Gillard et al., 2004, Vo-Dinh and Gauglitz, 2003, de Graaf, 2007, Reiser et al., 2008, Hoch and Stern, 1996, Jacobsen, 2007, Hore, 1995, Keeler, 2005, Macomber, 1998) is a non-invasive analytical and diagnostic technique that exploits the NMR phenomenon and can be performed with a standard clinical magnetic resonance imaging (MRI) scanner. It can reveal the molecular structure of a tissue voxel of interest, through the measurement of the interaction between certain nuclei in a strong magnetic field and an oscillating radio frequency electromagnetic field (Keeler, 2005).

There has been more interest in the clinical usefulness of proton MRS in recent years. The

[1] The same phenomenon can occur with electrons, e.g. electron spin resonance (ESR), but here MR refer to NMR.

proton is the most sensitive nucleus for NMR, in terms of intrinsic NMR sensitivity (high gyromagnetic ratio), high natural abundance (>99.9 %) and high biological abundance of the ^{1}H isotope of hydrogen (Gillard et al., 2004). The positively charged spinning proton generates a magnetic field and possesses a magnetic moment. The NMR experiment can exploit the magnetic properties of ^{1}H to provide detailed information on molecular structure.

The basic principles of the MR spectroscopy are explained in section 2.2. Section 2.3 focuses on the information content of the proton MR spectra of the brain. The last section in this chapter briefly reviews the LCModel (Provencher, 2009), which is a quantitation tool commonly used in the further processing and analysis of *in vivo* MR spectra.

2.2 Fundamentals of magnetic resonance spectroscopy

2.2.1 Nuclear magnetic resonance

Subatomic particles (electrons, protons and neutrons) behave as if they were spinning on their own axes, as shown in Figure 2.1(a). This spin is an inherent property of the nucleus. A spinning charge creates a small magnetic field aligned with the axis of spinning, and the spinning nucleus possesses a nuclear magnetic moment. When placed within an external static magnetic field, the nuclei can either be aligned along or opposed to it. In many nuclei (such as ^{12}C), these spins are paired against each other, such that the nucleus of the atom has no overall spin. Nuclei with zero spin have zero nuclear magnetic moment and cannot be detected by the NMR method. However, in some nuclei (such as ^{1}H and ^{13}C), the nucleus does possess an overall spin. It is observed that only isotopes with an odd number of protons and/or an odd number of neutrons possess non-zero nuclear spin. This is because a proton spin can only pair with another proton spin, but not a neutron spin, and vice versa (Macomber, 1998). The

different nuclear spin states of each nucleus can be characterised by the spin quantum number *I*. The rules for determining the spin quantum number *I* are as follows:

1. I is zero for a nucleus with both even numbers of neutrons and protons.
2. I is an integer for a nucleus with both odd numbers of neutrons and protons.
3. I is a half-integer for a nucleus with one odd number for either neutrons or protons.

It is clear from the above rules that the spin quantum number of hydrogen is $I = 1/2$. From quantum mechanics (Gillespie, 1974), it is known that a nucleus with spin quantum number *I* adopts $2I+1$ possible orientations in a magnetic field. Each of these $2I+1$ orientations has its own magnetic quantum number m_I in the range $m_I = -I : 1 : I$. The spinning nucleus generates a small magnetic field, therefore, possesses a magnetic moment $\boldsymbol{\mu}$. Its magnitude can be expressed as (Reiser et al., 2008):

$$\mu = \gamma \frac{h}{2\pi} m_I \tag{2.1}$$

where γ is the gyromagnetic ratio, which is a unique value for each type of nucleus, and h is Planck's constant $h = 6.626 \times 10^{-34}\,\mathrm{J \cdot s}$.

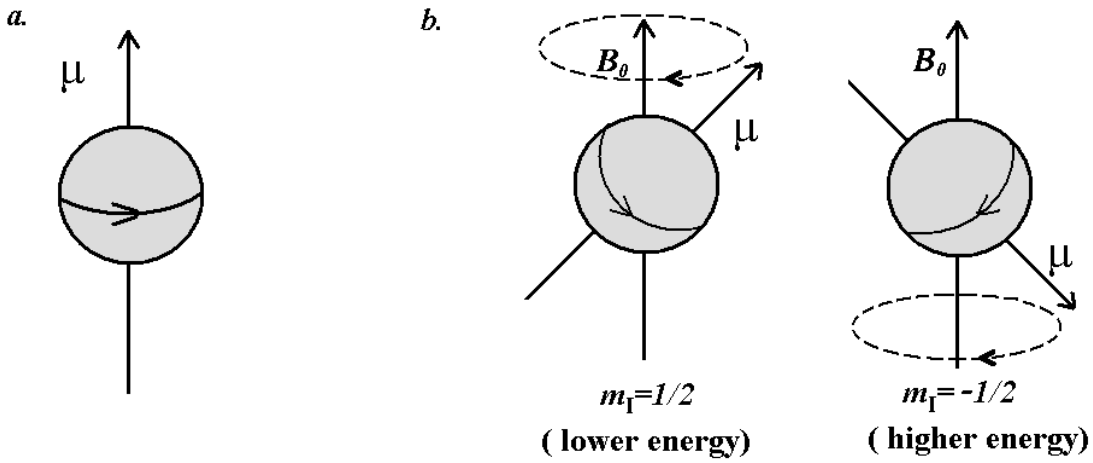


Figure 2.1: (a) The spin of a nucleus can be visualized as a rotational motion around its own axis, corresponding to the axis of its magnetic moment. (b) When applying an external static magnetic field $\mathbf{B}_0$, the spins are precessing around $\mathbf{B}_0$. The spin of an I =1/2 nucleus can be aligned along (low energy) or opposite (high energy) to the magnetic field.

When a strong magnetic field $\mathbf{B}_0$ is applied along a direction designated as the z axis, there is a slight tendency for magnetic moments to align in the lower energy direction. For nuclei with positive gyromagnetic ratios, such as protons, the lower energy direction is the general direction of the magnetic field $\mathbf{B}_0$. The alignment where it is opposed to the field $\mathbf{B}_0$ is less stable, as this requires a higher energy state. It is possible to make the nuclei flip from the more stable alignment to the less stable one by supplying exactly the right amount of energy. This will be explained later in this section.

The separation of spin energy levels in a magnetic field is called the Zeeman effect (Macomber, 1998). The nuclear magnetic moments are not actually lined up parallel to the z direction. Instead, they move in a circular fashion about the z axis, a motion called precession, as shown in Figure 2.1(b). The frequency of that precession is proportional to the gyromagnetic ratio γ and to the applied magnetic field strength $\mathbf{B}_0$, in units of radians per

second,

$$\omega_L = \gamma B_0 = 2\pi v \tag{2.2}$$

The above equation is known as the Larmor equation (McRobbie, 2007) and ω_L is termed as the Larmor frequency, where B_0 is the magnitude of $\mathbf{B}_0$ and v is the precession frequency in Hertz.

In a strong magnetic field $\mathbf{B}_0$, the $2I+1$ states for a spin-I nucleus are equally spaced. The energy difference (ΔE) between adjacent energy levels is,

$$\Delta E = \gamma \frac{h}{2\pi} B_0 \tag{2.3}$$

There are two possible spin states for a spin-1/2 nucleus, the lower energy spin state (m_I =1/2) and the higher energy spin state (m_I = –1/2). Their relative energies with an external magnetic field $\mathbf{B}_0$ are illustrated in Figure 2.2.

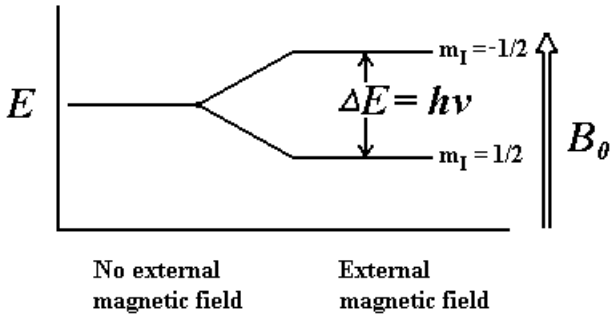


Figure 2.2: Proton ^{1}H energy level changing with and without an external magnetic field $\mathbf{B_0}$.

At thermal equilibrium, the ratio of the populations between the higher energy state and the lower energy state is given by the Boltzmann distribution (Jacobsen, 2007) as,

$$\frac{P_{higher}}{P_{lower}} = e^{-\Delta E/kT} \tag{2.4}$$

where P is the population of the particles in each state, ΔE is the difference in energy between the two states, T is the absolute temperature in Kelvin (K), and k is the Boltzmann constant $k = 1.381 \times 10^{-23}\,\mathrm{J \cdot K^{-1}}$.

The population difference between the higher and lower energy states is the most interesting thing for the MR phenomenon, because the magnetism of every "up" (higher energy) nuclear magnet cancels the magnetism of every "down" (lower energy) nuclear magnet, and it is only the difference in population that results in a "net magnetization" of the sample.

It is possible for a nucleus in the lower energy state to absorb a photon of electromagnetic energy and be promoted to the adjacent higher energy state. A photon possesses a discrete amount of energy that is directly proportional to its frequency,

$$E = h\nu \tag{2.5}$$

The energy of the photon must exactly match the energy difference between adjacent energy levels to be absorbed by a particle, hence transferring energy. The frequency ν of the photon depends on both $\mathbf{B}_0$ and γ, and is usually in the radio frequency (RF) range for ^{1}H in clinical MR acquisition environment, from equations (2.3, 2.5) we get,

$$\nu = \gamma \frac{B_0}{2\pi} \tag{2.6}$$

which indicates that ν is equal to the Larmor frequency. At this condition, the system is said to be in resonance and ν is called the resonant frequency. The amount of energy absorbed by the nuclei (and emitted at a later stage) depends on the population difference between the two energy states.

In order to measure the resonant frequency of each nucleus within a molecule, we need to have some way of getting the nuclei to absorb or emit RF energy. If a weaker oscillating magnetic field $\mathbf{B}_1$ is applied perpendicular to $\mathbf{B}_0$ and at a frequency that exactly equals to the Larmor frequency ν of the nucleus, absorption of energy will occur. The nucleus will flip from its lower energy state to the higher energy state. At the same time, there is another process that is equally likely, called "stimulated emission" (Jacobsen, 2007), where the photon can also be absorbed by a spin in the higher energy state and flip it down to the lower energy state with the emission of two photons. The rate of these processes is proportional to the population of spins in each of the two energy states. According to the Boltzmann distribution, there will be a slight preference to the lower energy state at thermal equilibrium. So when we turn on the oscillating magnetic field $\mathbf{B}_1$ at the Larmor frequency, a net absorption of RF energy will be observed.

After the $\mathbf{B}_1$ irradiation is switched off, the system will emit the absorbed energy and return to its initial equilibrium. The emitted signal is called the free induction decay (FID) signal and corresponds to an exponentially decaying sinusoid in the time domain (Reiser et al., 2008). The initial intensity of the signal is proportional to the number of nuclei that contributes to it. The precession of the magnetization vector to its equilibrium is what we actually detect in an

NMR experiment, which will induce an electromotive force (emf) in a receiving coil positioned in the transverse plane (McRobbie, 2007). This will be explained in detail in the next few sections.

2.2.2 Chemical shift

At any given field strength, each nucleus has a characteristic resonant frequency. It cannot be emphasized too much that the resonant frequency of a nucleus is proportional to the gyromagnetic ratio, γ, and to the magnetic field it experiences. This relationship forms the basis of nearly every experiment observed by NMR. But if all protons in a molecule had exactly the same resonant frequency, the technique would be useless because we would only see a single peak in the spectrum representing all of the protons. In fact, there are slight differences in resonant frequencies of different chemical elements depending on their chemical environment of the nucleus within a molecule.

The nucleus is located at the centre of a cloud of electrons. From Lenz's law (Reiser et al., 2008), when the sample is inserted in the magnetic field, the electrons begin to rotate around the nucleus (the induced current) and produce a magnetic field that is proportional and opposed to the field $\mathbf{B_0}$ at the centre of the rotation. The induced magnetic field modifies the external magnetic field in the surrounding area of the nucleus. This subtle variation, on the order of one part in a million, is called the chemical shift, which can provide detailed information about the structure of molecules. Therefore, the resonant frequency is not only a characteristic of the type of nucleus but also varies slightly depending on the position of that atom within a molecule (molecular structure). The relationship that the resonant frequency is exactly proportional to external field strength still holds, but it is the local magnetic field strength at the position of the nucleus that is important. The magnitude of the effective field

(Hore, 1995), , constituted by the external magnetic field $\mathbf{B_0}$ and the induced magnetic field, can be expressed as,

$$B_{\text{eff}} = B_0(1-\sigma) \tag{2.7}$$

where σ is a shielding constant, in units of parts per million, which reflects the extent to which the electron cloud around the nucleus "shields" it from the external magnetic field.

As a result of nuclear shielding, the resonance frequency becomes,

$$\nu = \frac{\gamma B_{\text{eff}}}{2\pi} = \frac{\gamma B_0(1-\sigma)}{2\pi} \tag{2.8}$$

The chemical shift is defined in terms of the difference in resonance frequencies between the nucleus of interest and a reference nucleus. To make chemical shifts the same regardless of magnet strength, we use the δ scale in parts per million (ppm), where the proportionality to $\mathbf{B_0}$ is already taken into account.

$$\delta = 10^6 \frac{\nu - \nu_{ref}}{\nu_{ref}} = 10^6 \frac{\sigma_{ref} - \sigma}{1-\sigma_{ref}} \tag{2.9}$$

For ^{1}H NMR, the reference molecule used is usually that of tetramethylsilane, $Si(CH_3)_4$.

Nuclei in a different chemical environment emit signals with different frequencies. These differences in the chemical shifts are miniscule: for a ^{1}H nucleus the "spread" of resonant frequencies around the fundamental frequency is only about 10 ppm. However, it is this minute frequency variation which makes MRS a very attractive tool, since it allows the differentiation between molecular structures, thus greatly aiding the unambiguous detection

and assignment of compounds.

2.2.3 Spin-spin coupling

Another valuable piece of information about molecular structure is obtained from the phenomenon called spin–spin coupling (Yan, 2002), which describes the interactions between neighbouring spins in the same molecule, and induces a further differentiation in resonance frequencies. This effect is transmitted through bonds and exists only when the two nuclei are very close (three bonds or less) in the bonding network (Jacobsen, 2007). Since the resonant frequency is always proportional to the magnetic field experienced by the nucleus, the interactions between nearby spins cause the protons to resonate at several slightly different frequencies very close to each other, shown as a doublet, triplet or multiplet in the NMR spectra. The frequency differences of the multiplets depend on the spin-spin coupling constant (Reiser et al., 2008), $\boldsymbol{J}$, usually measured in Hz. The magnitude of $\boldsymbol{J}$ is independent of the strength of the applied magnetic field. This effect is another feature of NMR spectra that can aid in the identification and characterisation of biochemical compounds.

2.2.4 Relaxation processes

Before applying the oscillating magnetic field $\mathbf{B}_1$, the nuclei in the two spin states are precessing with the Larmor frequency with random of phases. The net nuclear magnetization $\mathbf{M}$ of all the individual nuclear magnetic moments $\boldsymbol{\mu}$ is aligned with the magnetic field direction in the z axis (Figure 2.3(a)). It has no component in the xy plane due to phase cancellation. When the irradiation by $\mathbf{B}_1$ begins, all of the individual nuclear magnetic moments become phase coherent, and form a precessing “bundle”. The magnetization $\mathbf{M}$

now will precess around the *z* axis with the Larmor frequency and have a component in the *xy* plane ($\mathbf{M}_{xy,0}$) as shown in Figure 2.3(c).

$$M_{xy,0} = M \sin\alpha \tag{2.10}$$

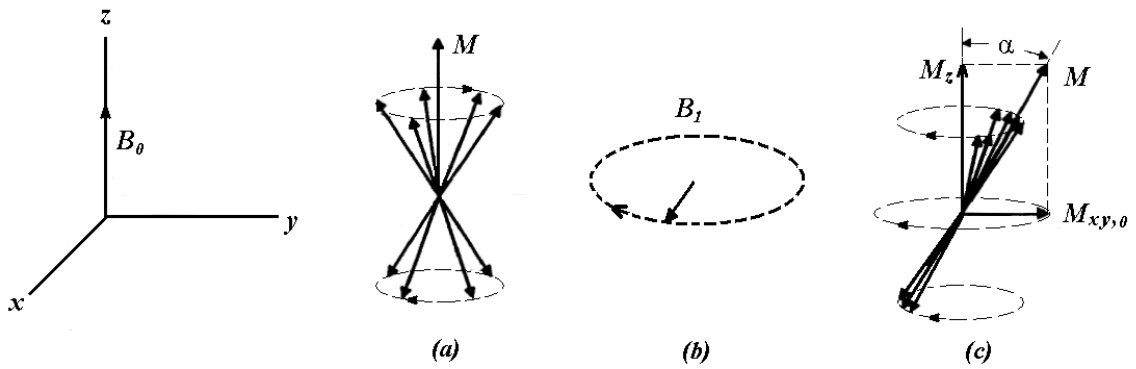


Figure 2.3: Precession of a collection of spin-1/2 nuclei around external magnetic field $\mathbf{B}_0$, the net nuclear magnetization **M** is the vector sum of all the individual nuclear magnetic moments, (a) before irradiation by $\mathbf{B}_1$, (b) the rotating magnetic field $\mathbf{B}_1$ in *x, y* plane, (c) during irradiation by $\mathbf{B}_1$. (Figure adopted from (Macomber, 1998))

The angle α is determined by the power and duration of the $\mathbf{B}_1$ irradiation. The magnetization **M** can be completely excited onto the transverse plane (α=90°) or even inverted to the −*z* axis (α=180°), giving rise to the effects of the so-called 90° or 180° pulse, respectively. After the $\mathbf{B}_1$ irradiation is switched off, the nuclear magnetization returns to its initial equilibrium state by relaxation processes (Schorn and Taylor, 2004). The individual nuclear magnetic moments will gradually return to a random arrangement around the *z* axis through spin-spin interactions with relaxation time T_2. Simultaneously, spins that were flipped to the higher energy state relax back to the lower energy state through spin-lattice interactions with relaxation time T_1. These biological parameters T_1 and T_2 are tissue-specific time constants,

which introduce the possibility to separate different tissue types.

The spin-lattice relaxation time T_1 is the time it takes for the longitudinal component of the magnetization, $\mathbf{M}_z$, to return to 63% of its Boltzmann equilibrium value $\mathbf{M}$ (Bigler, 2000),

$$M_z(t) = M(1 - e^{-t/T_1}) \tag{2.11}$$

The spin-spin relaxation time T_2 is the time it takes for the transverse component of the magnetization, $\mathbf{M}_{xy}$, to decay to 37% of its initial value after being flipped into the xy plane.

$$M_{xy}(t) = M_{xy,0}\ e^{-t/T_2} \tag{2.12}$$

In practice, the transverse decay time is also affected by the inhomogeneities of external magnetic field. The effective decay time T_2^* of $\mathbf{M}_{xy}$ is always smaller than T_2 and is defined as:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B \tag{2.13}$$

where ΔB indicates the inhomogeneities of the magnetic field (Reiser et al., 2008).

2.2.5 Free induction decay (FID) signal

The radiofrequency field $\mathbf{B}_1$ rotates around the z axis at a frequency ω_{RF}. Viewing this rotation from the fixed Cartesian coordinate's perspective, the description of the motion of the magnetization vectors can be complicated and difficult to visualize. It can be more convenient if we consider the motion of the magnetization from the point of view of an observer rotating

about the z axis, in synchronism with the rotating RF field: this is the so-called rotating frame of reference (Landini et al., 2005). This rotating frame is a coordinate system whose transverse plane is rotating at an angular frequency ω_{RF} around the longitude axis, so the field $\mathbf{B}_1$ appears stationary. To distinguish it from the conventional stationary frame, we use x', y', and z' (same as z) to denote the three orthogonal axes of this frame.

As $\mathbf{M}$ precesses around the z axis with the Larmor frequency $\omega_L = \gamma B_{\text{eff}}$, in the xy plane, the $\mathbf{M}_{xy}$ precesses around the original point (z axis) with the same frequency ω. If the precession frequency ω of $\mathbf{M}_{xy}$ was identical to the radiofrequency ω_{RF}, then $\mathbf{M}_{xy}$ would appear stationary in the rotating frame. In general, the frequency ω_0 with which $\mathbf{M}_{xy}$ appears to precess around the original point in the rotating frame is not zero and is reduced to:

$$\omega_0 = \omega - \omega_{RF} \tag{2.14}$$

If we place a receiver coil with its axis aligned in the $x'y'$ plane, the changing magnetization $\mathbf{M}_{xy}$ caused by the relaxation will induce a current in that coil. This current can be amplified and recorded, and is known as the free induction decay (FID) signal (as described before in section 2.2.1). Redraw the magnetization $\mathbf{M}_{xy}$ in the rotating $x'y'$ plane (Figure 2.4), and let the precession start from positive x towards y axis. The x and y components of the magnetization become:

$$\begin{aligned} M_x(t) &= M_{xy}(t)\cos\omega_0 t = M_{xy,0}\, e^{-t/T_2^*}\cos\omega_0 t \\ M_y(t) &= M_{xy}(t)\sin\omega_0 t = M_{xy,0}\, e^{-t/T_2^*}\sin\omega_0 t \end{aligned} \tag{2.15}$$

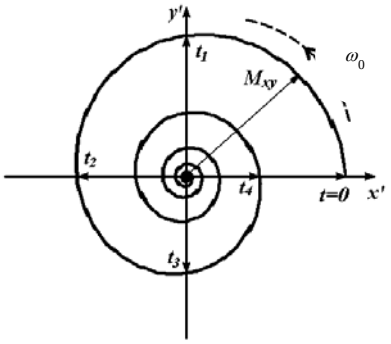


Figure 2.4 Project the magnetization **M** to the *x'y'* plane in the rotating frame.

The FID signal recorded is proportional to the transverse magnetization. Its actual relation will not be discussed here. The *x* and *y* components of the signal can be represented as,

$$\begin{aligned} S_x(t) &= S_0 e^{-t/T_2^*} \cos \omega_0 t \\ S_y(t) &= S_0 e^{-t/T_2^*} \sin \omega_0 t \end{aligned} \tag{2.16}$$

where S_0 is the maximum signal amplitude. If we regard the signal $S_x(t)$ and $S_y(t)$ as the real and imaginary part of a complex signal $S(t)$ (Keeler, 2005), we have

$$\begin{aligned} S(t) &= S_x(t) + i\, S_y(t) \\ &= S_0 e^{-t/T_2^*} \cos \omega_0 t + i\, S_0 e^{-t/T_2^*} \sin \omega_0 t \\ &= S_0 e^{-t/T_2^*} e^{i\omega_0 t} \end{aligned} \tag{2.17}$$

In practice, the recorded signal $S(t)$ may not always start at the positive *x* axis, it may have a different phase. Hence, the recorded signal $S(t)$ is phase shifted. Mathematically, the phase shifted signal becomes,

$$S(t) = S_0 e^{i\phi} e^{-t/T_2^* + i\omega_0 t} = S_0 e^{i\phi} e^{-\alpha t + i\omega_0 t} \tag{2.18}$$

where ϕ is the phase shift as shown in Figure 2.5.

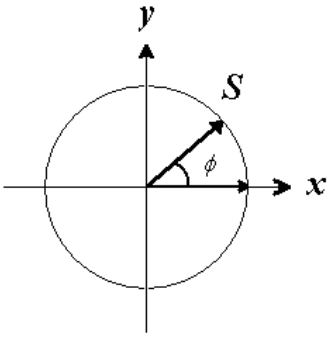


Figure 2.5 The start point of the recorded signal *S* is phase shifted by ϕ from the positive *x* axis.

The received FID signal is usually the sum of contributions from the same type of nuclei with different Larmor frequencies and phases, which is a superposition of Lorentzian lineshapes.

$$y(t) = \sum_{n=1}^{N} S_{0,n} e^{i\phi_n} e^{-\alpha_n t + i\omega_{0,n} t} \tag{2.19}$$

The MR spectrum is obtained by taking the Fourier transform on the FID signal (equation 2.19) over a certain time. The frequency domain signal is known as the spectrum. Like the time domain signal, the frequency domain signal has real and imaginary parts.

$$\begin{aligned} S(\omega) &= \int_0^{\infty} S(t)\, e^{-j\omega t} dt \\ &= S_0 \frac{\alpha \cos\phi - (\omega_0 - \omega)\sin\phi}{\alpha^2 + (\omega_0 - \omega)^2} + jS_0 \frac{\alpha \sin\phi + (\omega_0 - \omega)\cos\phi}{\alpha^2 + (\omega_0 - \omega)^2} \end{aligned} \tag{2.20}$$

At zero phase, $\phi = 0$, the real part of the spectrum is called absorption mode Lorentzian, and the imaginary part of the spectrum gives a lineshape known as the dispersion mode Lorentzian (Figure 2.6). The dispersion lineshape is broader than the absorption mode, and has positive and negative parts. It is commonly not interpreted as it appears more complex and can cause confusion.

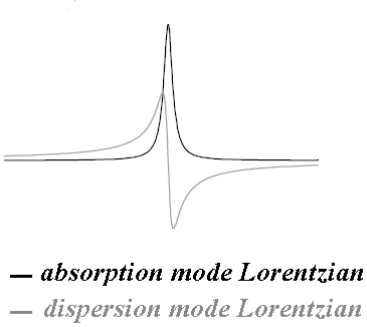


Figure 2.6: Spectrum of a Fourier transformation of an exponentially decaying FID gives rise to Lorentzian absorption and dispersion lineshapes.

Figure 2.7 shows an example of short echo time (TE)[2] proton FID signal and the real part of its MR spectrum. The peak intensities on the vertical axis of the spectrum correspond to the relative number of chemically equivalent protons, and the resonance frequencies are represented in ppm on the horizontal axis.

[2] Echo time (TE) is defined as the time between the start of the RF pulse and the maximum in the signal.

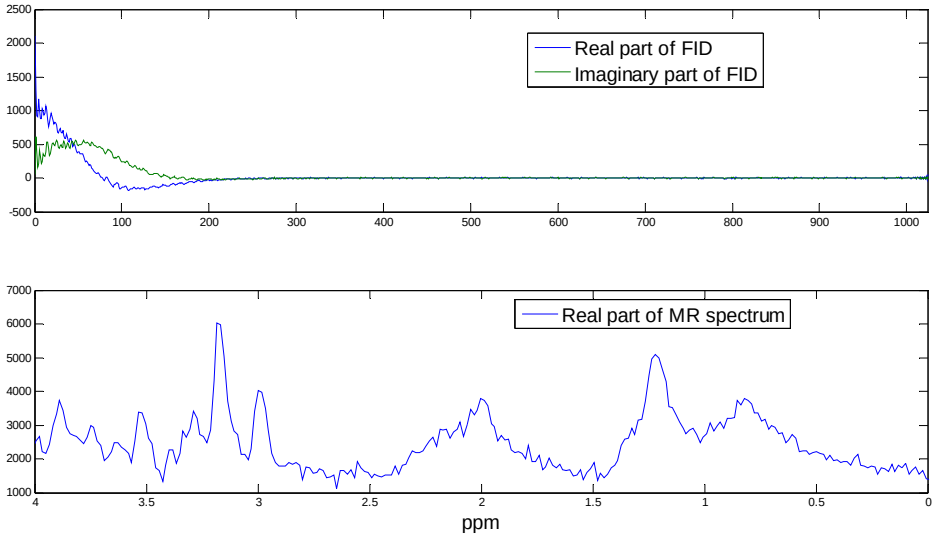


Figure 2.7: *In vivo* ^{1}H NMR spectra of the human brain.

2.2.6 Spin Echo

The spin echo can eliminate the influence of the inhomogeneity of the external magnetic field and thus allow the measurement of the T_2 relaxation times specific to the substance/tissue.

A spin echo is a sequence of a 90° pulse followed by a delay TE/2, a 180° pulse and a second delay TE/2 (Landini et al., 2005). The rotating frame view of an on-resonance magnetization subjected to this sequence is shown in Figure 2.8. The initial 90° pulse rotates the equilibrium magnetization **M** by 90° about the *x'* axis to the *y'* axis creating transverse magnetization (Figure 2.8a and 2.8b). During the first delay TE/2, this transverse magnetization decays because of spin-spin relaxation and the inhomogeneity of the $\mathbf{B}_0$ magnetic field (Figure 2.8c). Applying a 180° pulse at time TE/2 following the initial 90° pulse causes all the isochromats to rotate by 180° about the *x'* axis; this brings the isochromats to their mirror image position (Figure 2.8d). Following the 180° pulse, since the Larmor frequency of each isochromat is unchanged, the frequency and direction of precession of isochromats in the *x'y'* plane remain the same as prior to this pulse. Precession of isochromats for a period TE/2 after the 180°

pulse allows all the isochromats to refocus along the negative *y'* axis forming a “spin echo” (Figure 2.8e).

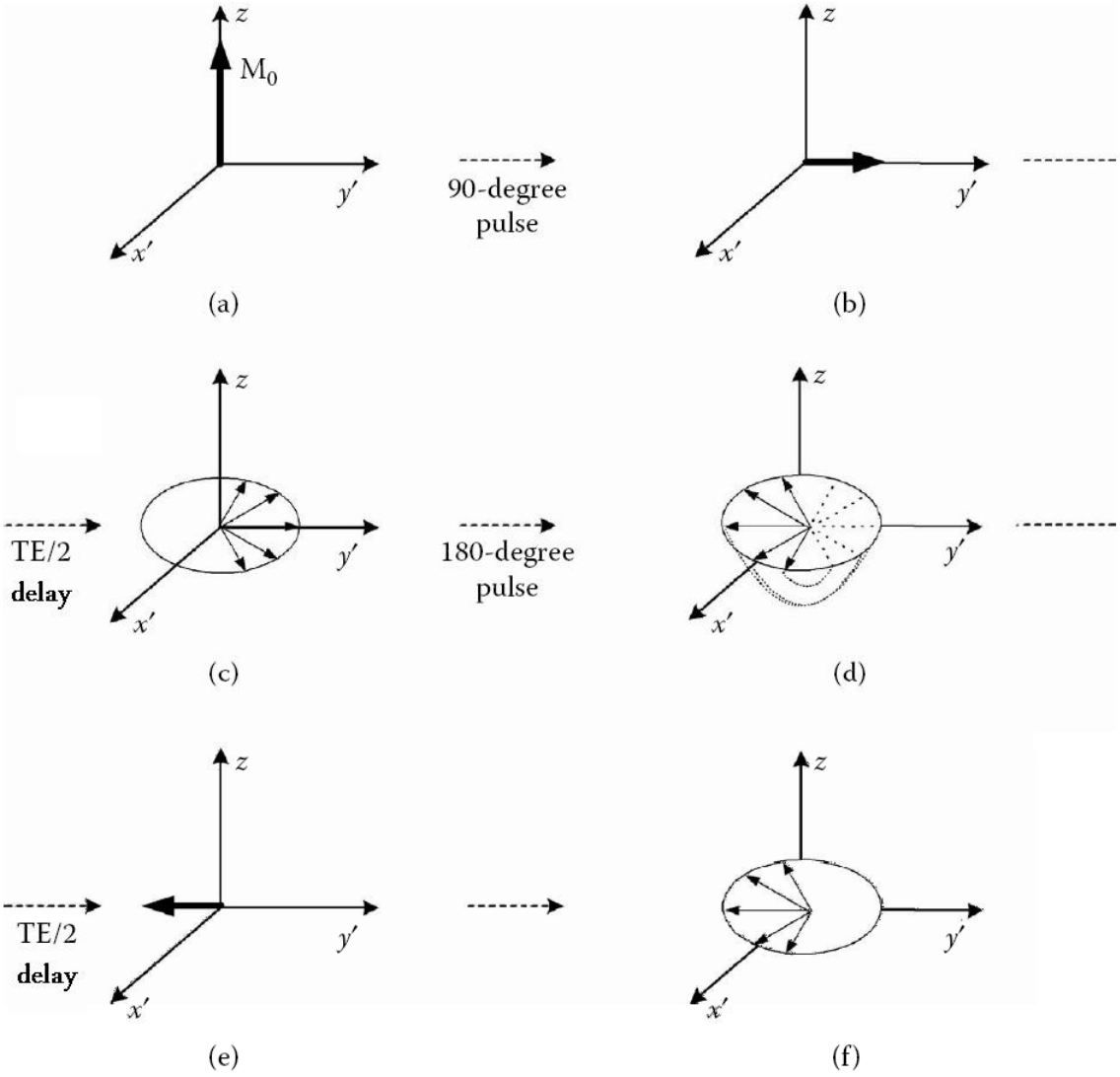


Figure 2.8: Rotating frame view of a magnetization subjected to a spin echo sequence (Figure adopted from (Landini et al., 2005)).

2.2.7 Spatial localization

Generally, two different approaches are used for collecting proton spectroscopy data from a voxel of interest: point resolved spectroscopy (PRESS) (Bottomley, 1984) and stimulated echo acquisition mode (STEAM) (Frahm et al., 1987) pulse sequence-based approaches. A single voxel is defined by the intersection of three mutually orthogonal slices. Each slice is excited by a spatially selective pulse, applied in the presence of a gradient (McRobbie, 2007).

In PRESS, a 90° pulse is followed by two 180° pulses so that the primary spin echo is refocused again by the third pulse. Each pulse has a slice-selective gradient on one of the three principle axes (Figure 2.9), so that protons within the voxel are the only ones to experience all three RF pulses. After the excitation by the first 90° pulse, transversal magnetization is produced within a slice perpendicular to the *z* axis in this case. This magnetization starts to dephase and, after a time $TE_1/2$, the magnetization of a part of the slice is refocused by the first 180° pulse in the same way as in conventional spin-echo sequences (McRobbie, 2007, Landini et al., 2005). The spin echo occurring at the time TE_1 is not evaluated, the magnetization dephases again, and the second 180° pulse is applied at the time $TE_1+TE_2/2$ to give an echo at the time TE_1+TE_2, where the data acquisition time starts.

The signal intensity depends on the time spacing of pulses and relaxation times, and is intrinsically twice as high as STEAM, so spectra can be acquired with good SNR in a relatively short time. The PRESS technique allows higher ^1H MRS signal intensity in comparison to the STEAM sequence.

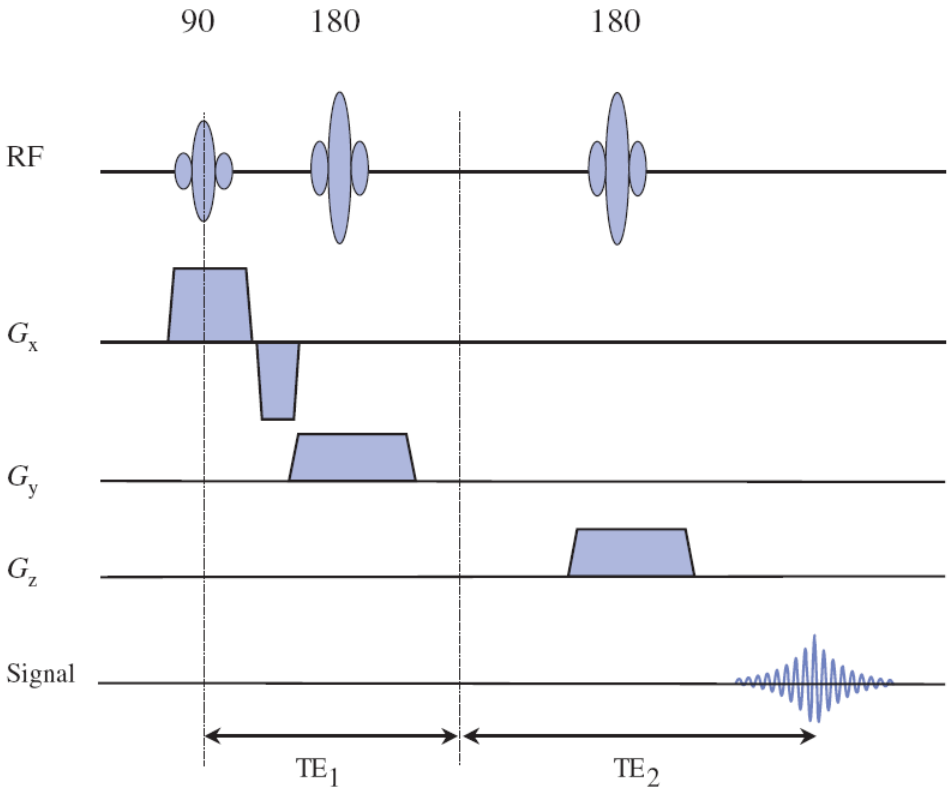


Figure 2.9: The PRESS pulse sequence (Figure adopted from (McRobbie, 2007)), TE denotes the echo time.

STEAM uses the effect of a stimulated echo occurring after the application of three successive pulses. The most intense signal strength of a stimulated echo can be obtained if all three pulses are 90° pulses (Figure 2.10). The first 90° pulse is used to produce transversal magnetization within a selected slice, as in the PRESS sequence. Although most of the spins are in phase immediately after the excitation, they begin to dephase with time, under the influence of local inhomogeneities of the static magnetic field and that of the applied field gradients. After a time TE/2, the second 90° pulse rotates the dephased magnetization within the *xy* plane into the *zy* plane. After the last 90° pulses, the remaining part of the magnetization is again flipped into the transverse plane. After the third pulse, a rephasing occurs similar to conventional spin-echo sequences and, at the time TE+TM, all spins are

again in the same phase and build a measurable net magnetization vector.

The main advantage of the STEAM sequence compared to PRESS is a reduced minimal TE, but is intrinsically only half the amplitude of a conventional echo.

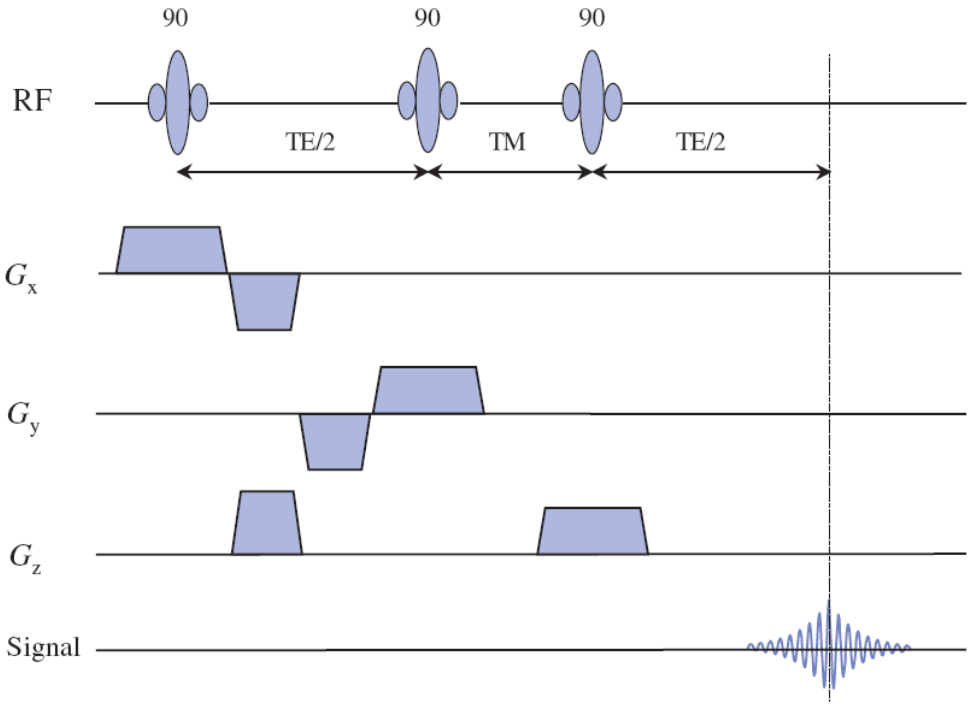


Figure 2.10: The STEAM pulse sequence (Figure adopted from (McRobbie, 2007)), TM denotes the mixing time.

2.2.8 Challenges in MR spectroscopy

The application of *in vivo* proton MRS is challenging for several reasons. *In vivo* MRS signals are characterized by a low SNR, the presence of unwanted spectral components like the residual water peak in proton MRS and distortions with respect to the expected model function. In short TE proton spectra, different peaks are strongly overlapping, making it difficult to determine the model parameters (Keshavan et al., 2000).

The FID signal contains noise from various sources. The noise is contributed by the

amplifiers and other electronics in the spectrometer, the thermal noise from the coil used to detect the signal and the thermal noise from the sample (Edelstein et al., 1986). In order to increase the SNR, a number of data acquisitions are performed consecutively and the final signal is the average of all measured signals.

Because the $\mathbf{B_0}$ magnetic field distribution in the human head is highly heterogeneous (Li et al., 1995), the spectral resolution can be decreased. To produce a more homogeneous magnetic field within the voxels, shim coils are used. The effect of the shimming process reduces spectral peak broadening and also improves SNR.

The brain metabolite levels are on the order of 10 mM or less, whereas protons in brain water are approximately 80 M, and lipids containing protons are also present in high concentrations (Barker and Lin, 2006). This makes metabolite detection difficult and ambiguous. Numerous methods for water suppression have been developed in high-resolution NMR spectroscopy, and some of these methods have been applied to *in vivo* spectroscopy. Typically a very narrow bandwidth frequency-selective pulse, often called a CHEmical Shift Selective (CHESS) pulse (McRobbie, 2007), is applied at exactly the Larmor frequency of water using a low-power Gaussian shaped pulse to give a 90° pulse, followed by gradient pulses to spoil or suppress any transverse magnetization. Lipid suppression can be performed in several different ways. One approach is to avoid exciting the lipid signal using, e.g. STEAM or PRESS localization to avoid exciting lipid-containing regions.

The phase of the raw spectrum after Fourier transformation is usually incorrect, and requires manual or automatic adjustment (Keeler, 2005). An MRS signal is typically represented by its real and imaginary part, and the process of “phasing” is required to correctly produce the real part of the spectrum for classification.

The baseline of the resulting spectrum is typically distorted, and has to be corrected. Short TE ^{1}H MRS signals are characterized by the presence of a partially unknown broad baseline underlying the sharper resonances of the metabolites of interest. For manual baseline correction, the user defines several spectral points, typically between the major metabolite peaks, as “baseline”. The computer then fits and subtracts a smooth curve through these points. The result is a spectrum with a flat baseline, which is better suited for determination of metabolite peak areas.

2.3 Information content of proton MR spectra of the brain

Since nearly all metabolites contain protons, *in vivo* ^{1}H NMR spectroscopy is in principle a powerful technique to observe, identify and quantify a large number of biologically important compounds in intact tissue. Proton MRS of the brain provides a diagnostic modality for the biochemical characterization of developmental and pathologic neurological conditions. Either short or long TE can be employed to evaluate the brain parenchyma. In short TE (around 30ms) spectra, metabolites with both short and long T_2 relaxation times are observed. Hence they are useful in evaluating complex metabolic abnormalities as more metabolites are presented for analysis. Glutamine, glutamate, myo-inositol, and most amino acids are better evaluated at short TE (Reiser et al., 2008). With a long TE, only metabolites with a long T_2 are observed. It produces spectra with N-acetyl aspartate, creatine, and choline as the dominant peaks in normal brain tissue. Evaluation for the presence of lactate can be performed more easily at long TE (135ms) since the doublet peak is inverted relative to most other peaks.

The information content of the main resonances (Gillard et al., 2004, Govindaraju et al., 2000)

appearing in *in vivo* ^{1}H MR spectra of the brain will be briefly explained in this section. A quantum mechanics based simulation program (Appendix A) (Reynolds et al., 2006) is used to generate the ^{1}H MR spectra of these metabolite and MMLip components at short TE and a field strength of 1.5T (Figure 2.11). The chemical structures of these metabolites are shown in Figure 2.12. These metabolite and MMLip components are also the main components used to construct the synthesised datasets (described in Chapter 5) for the experiments in this thesis.

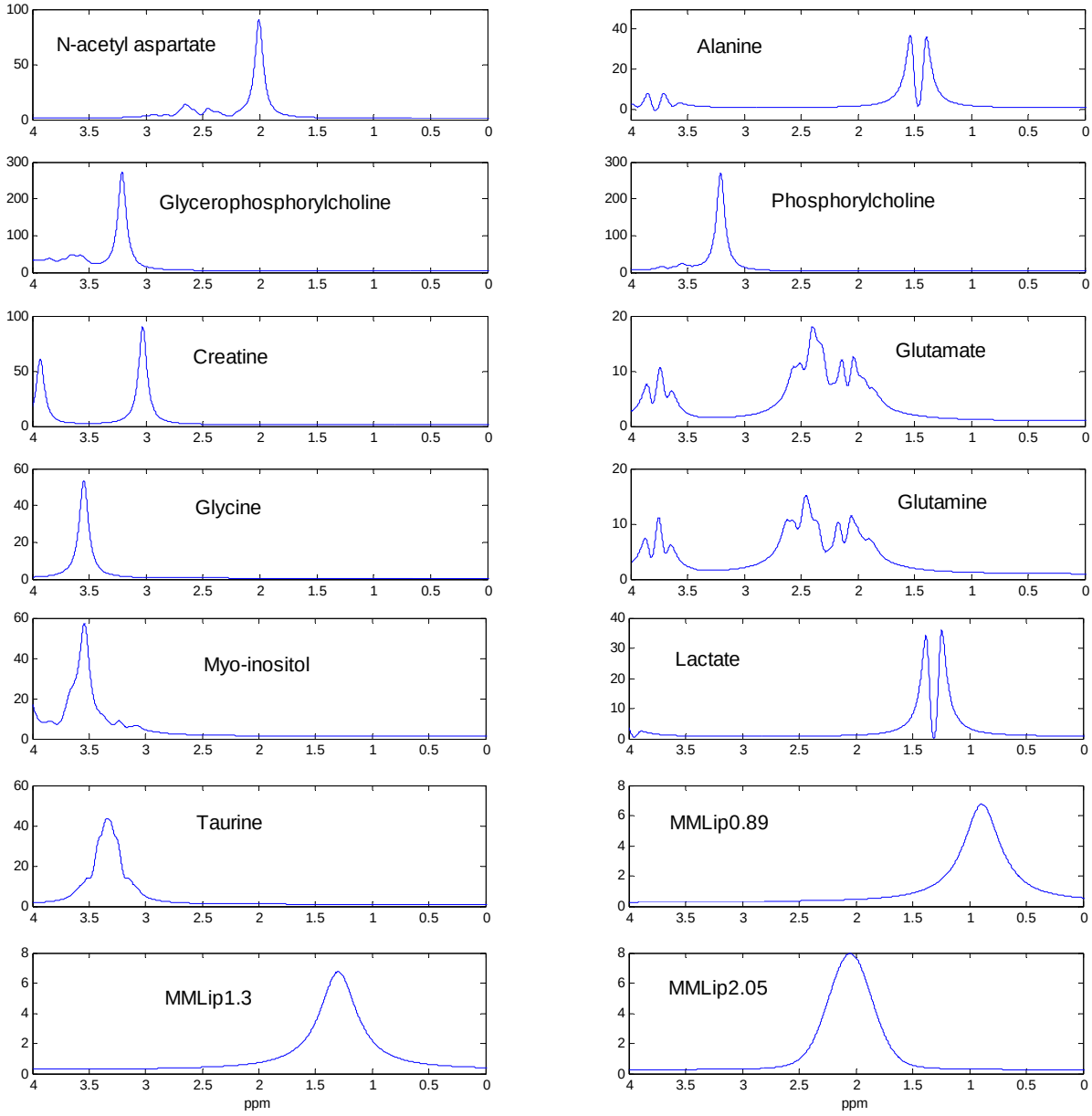
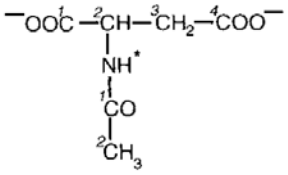
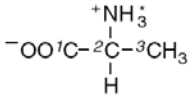


Figure 2.11 The simulated ^{1}H MR spectra for the most commonly seen metabolite and MMLip components at a field strength of 1.5T and a line width of 5.5 Hz in the range of 0 - 4.0 ppm.

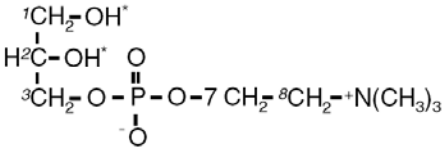
N-acetyl aspartate



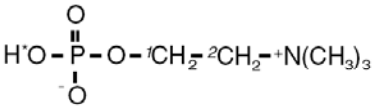
Alanine



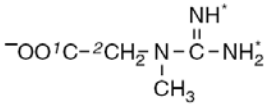
Glycerophosphorylcholine



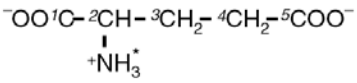
Phosphorylcholine



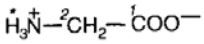
Creatine



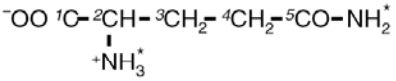
Glutamate



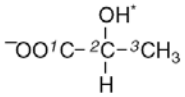
Glycine



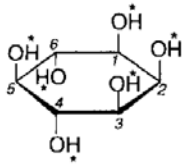
Glutamine



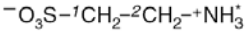
Lactate



Myo-inositol



Taurine



Choline

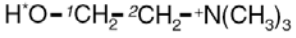


Figure 2.12 The chemical structures of metabolites (de Graaf, 2007) shown in Figure 2.11, the exchangeable protons are indicated by asterisks.

2.3.1 N-acetyl aspartate (NAA)

N-acetyl aspartate (NAA) is a free amino acid present in the brain at relatively high concentration. It has seven protons that give MR signals between 2.0 and 8.0 ppm. The most prominent resonance signal, a singlet at 2.02 ppm, occurs from the N-acetyl methyl ($-CH_3$) group of NAA. At lower field strengths, this resonance may also contain smaller contributions from N-acetylaspartylglutamate (NAAG). NAA also has three doublet-of-doublets centred at 2.49, 2.67 and 4.38 ppm that correspond to the protons of aspartate methylene ($-CH_2$) and methine (−CH) groups. At shorter echo times NAA is overlapping with glutamate and macromolecules. Its exact physiological function is poorly understood, but brain NAA is commonly believed to be a health neuronal cell marker (Gillard et al., 2004, Panigrahy et al., 2006). The presence of a large NAA peak normally indicates more normal neuronal presence and function, while diminished peaks occur in situations where neural damage (such as tumour) has occurred (Reiser et al., 2008). However this is not fully substantiated as NAA concentrations differ among neuron types.

2.3.2 Alanine (Ala)

Alanine (Ala) is a non-essential amino acid that contains a methyl group. Similar to lactate, the −CH3 and −CH protons of Ala give rise to a doublet at 1.47 ppm and a quartet at 3.77 ppm. The spectral region at 1.47 ppm is usually overlapped with lipid resonances. It can be observed with current methods only at pathologically elevated concentrations (Panigrahy et al., 2006).

2.3.3 Choline (Cho), Glycerophosphorylcholine (GPC) and Phosphorylcholine (PCh)

Besides resonances from NAA and total creatine, the most prominent resonance in ^{1}H MR spectra from brain arises from the methyl protons of choline-containing compounds. A strong singlet resonance at 3.2 ppm originating from the nine equivalent protons in trimethylamine ($-N(CH_3)_3$) groups of the free choline (Cho), glycerophosphorylcholine (GPC) and phosphorylcholine (PCh), which is often referred to as 'total choline' (tCho). Apart from the nine protons in the $-N(CH_3)_3$ group, there are a few small signals resonate from several $-CH$ and $-CH_2$ groups between 3.5 and 4.3 ppm. Increasing Cho levels seem to be a characteristic of many types of neoplasms, including high-grade brain tumours (provided that they are not necrotic), and others (Barker and Lin, 2006). At short echo times, overlapping resonances from myo-inositol, and taurine should be taken into account. Free choline is a minor contributor to the 'total choline' resonance, since the concentration is well below the NMR detection limit.

2.3.4 Creatine (Cr)

The creatine (Cr) signal contains two prominent singlet resonances from the methyl protons of creatine and phosphocreatine (PCr) at 3.03 ppm and methylene protons at 3.94 ppm for *in vivo* MR spectrum. Cr and PCr are presented in brain, muscle, and blood. The ^{1}H MR spectra of creatine and phosphocreatine are very similar. The difference between the Cr (3.027 ppm) and PCr (3.029 ppm) methyl resonances is too small to allow a reliable separation at low field *in vivo*. These compounds are a marker of metabolic brain energy. In the normal brain, the Cr levels are higher in grey matter than those in white matter (Gillard et al., 2004). Its levels

remain relatively constant in tumours. In many disease processes, a relative loss or gain of NAA and Cho can be indicated by comparison with Cr (Ross and Bluml, 2001). However, since Cr levels are subject to variability in disease process such as creatine deficiency syndromes and in the setting of stroke, trauma, and necrotic tumours (Gillard et al., 2004), care must be taken when evaluating them.

2.3.5 Glutamate (Glu)

Glutamate (Glu) is an amino acid with an acidic side chain. Glutamate and glutamine are very difficult to separate in proton spectra at lower field strengths, i.e. below 3T, and are often labelled as a composite peak, Glx. Glu has two methylene groups and a methine group that are strongly coupled. As a result of the extensive and strong scalar coupling interactions Glu has a complex MR spectrum with signal spread out over many low intensity resonances. Signal from the single methine proton appears as a doublet-of-doublets centred at 3.75 ppm, while the resonances from the other four protons of the methylene groups appear as multiplets in the 2.04 − 2.35 ppm range.

2.3.6 Glutamine (Gln)

The amino acid glutamine (Gln) is a precursor and storage form of Glu located in astrocytes. Gln is found primarily in astrocytes. Gln is structurally similar to Glu with a single methine group and two methylene groups. As a result, the chemical shifts and scalar coupling interactions are also similar. The methine proton resonates as a triplet at 3.76 ppm, while the multiplets of the four methylene protons are closely grouped between 2.12 and 2.46 ppm. In addition, Gln has two NMR detectable amide protons at 6.82 and 7.73 ppm. Although the role and significance of the different levels of Glu and Gln is unclear, the quantitation of these

metabolites proved to be useful for the separation of several tumour groups (Panigrahy et al., 2006).

2.3.7 Glycine (Gly)

Glycine (Gly) is a simple amino acid that functions as an inhibitory neurotransmitter and antioxidant, and is distributed throughout the brain and central nervous system. It has two methylene protons that co-resonate at 3.55 ppm. For *in vivo* MR spectrum, the Gly resonance overlaps with those of myo-inositol.

2.3.8 Lactate (Lac)

Lactate (Lac) is the end-product of anaerobic glycolysis. Its signal detection is commonly carried out via the doublet from the methyl group at 1.31 ppm. The single methine proton resonates as a quartet at 4.10 ppm. In normal human brain, it is not observed by *in vivo* MRS. Any detectable increase in Lac can be considered abnormal. Elevated Lac level is usually the result of deranged energy metabolism, and has been observed in brain tumours (Gillard et al., 2004), and other conditions.

2.3.9 Myo-inositol (m-Ins)

Myo-inositol (m-Ins) is the most prominent form of the nine isomers of inositol in tissue (Cerdan et al., 1985). It is a cyclic sugar alcohol that has six protons and gives four groups of resonances. The two prominent multiplets are a doublet-of-doublets centred at 3.52 ppm originates from the 1CH and 3CH protons, while the 4CH and 6CH protons give rise to a triplet at 3.61 ppm. The largest signal occurs at 3.56 ppm. A triplet at 3.27 ppm is typically overlapped with Cho, and another at 4.05 ppm is usually not observed because of water

suppression. The function of m-Ins is uncertain.

2.3.10 Taurine (Tau)

Taurine (Tau) is an amino acid, which is found at high concentration at the time of birth and decreases with age. It has two adjacent methylene groups giving two triplets at 3.25 and 3.42 ppm. At lower field *in vivo* studies, these resonances commonly overlap with the resonances from m-Ins and Cho. Tau is present in all cells of the central nervous system, its exact function is not known, however elevated concentration in medulloblastoma has been reported in the study by Panigrahy et al. (Panigrahy et al., 2006) and Moreno-Torres et al. (Moreno-Torres et al., 2004).

2.3.11 Macromolecules and lipids (MMLip)

In short TE ^{1}H MR spectra, a significant fraction of the observed signal must be attributed to macromolecular and lipid resonances underlying those of the metabolites. A minimum of 10 macromolecular resonances (Hofmann et al., 2001, Pfeuffer et al., 1999) are observed between 0.93 and 4.3 ppm. Assigning them to specific proteins is currently virtually impossible. Lipids are either presented as subcutaneous or interstitial adipose tissue. It is usually observed in short TE spectrum due to the shorter T_2 compared with metabolites. In MR spectra of human brain, the major lipid peaks occur at 1.3 ppm due to the methylene protons and at 0.89 ppm due to the methyl protons. Macromolecules and lipid components are often grouped together to account for three broad resonances at 0.89 ppm (MMLip0.89), 1.3 ppm (MMLip1.3) and 2.05 ppm (MMLip2.05).

2.4 LCModel

The linear combination model (LCModel) proposed by Provencher (Provencher, 1993, Provencher, 2001) is one of the more widely used methods, in recent years, for spectral quantitation and metabolite identification. The *in vivo* MR spectra are analysed as a linear combination of a basis set of individual *in vitro* metabolite solutions. In simple terms, it adjusts the amplitudes, frequencies, line widths and phases of the metabolite basis set to match the *in vivo* MR spectrum as closely as possible. It returns the estimated concentrations for each metabolite, lipid and macromolecular components by reference to an unsuppressed water spectrum from the same voxel (if available) and estimates the uncertainty in the fit by Cramer–Rao lower bounds (CRLB) (although systematic errors cannot be estimated in the same way) (Kay, 1993). A fitting error of 20% or greater generally indicates that the peak area determination is unreliable.

The overlapping peaks and irregular baselines are dealt with by fitting the *in vivo* spectrum as a combination of pure, model spectra from each of the expected compounds in the tissue. The basis set used consists of a separate spectrum for each of the 16 metabolites and 9 lipid and macromolecular components likely to be present in the spectrum. One of the most important aspects of the LCModel analysis is the generation of an accurate basis set. There are generally two methods of obtaining a basis set, by measurement or through simulation (Provencher, 2009). In the measurement approach, the solutions of expected metabolites are measured under similar conditions as used for the *in vivo* NMR measurement. An attractive alternative is simulating the basis spectra. By default, the LCModel simulates macromolecule and lipid basis spectra.

The spectral range for the analysis was set by default to 0.2–4.0 ppm in LCModel (Provencher, 2009), as this region contains most signals of interest. Automatic phasing of the spectrum is carried out. Correction for eddy current distortion can be made, using a water spectrum (Klose, 1990).

The irregular spectral baseline contains broad signals from macromolecules. The LCModel baseline can automatically account for a missing or incorrect macromolecule model, with good data (Pfeuffer et al., 1999). With poorer data, the baseline becomes flatter (because there is no longer sufficient information to determine the baseline detail), and the concentration estimates systematically get worse. Methods always using flat baselines can yield more reproducible (consistently wrong) results, with deceptively low coefficients of variation. Flexible baselines are needed to compensate for the inevitable variability and machine induced acquisition effects, particularly of macromolecule and lipid signals.

The LCModel estimates concentrations of even minor metabolites to high internal precision, which should be useful in physiological and clinical studies. A model of complete spectra is used rather than individual peaks, so two metabolites with very close peaks in one frequency region can still be resolved if they have different signals in other parts of the spectrum.

The LCModel has the advantage that a maximal amount of prior knowledge can be used. Moreover, model distortions due to acquisition effects can also be taken into account. Ideally, the method requires the recording of a basis set of metabolite spectra under exactly the same settings as the signal under investigation and the model order should be provided prior to quantification. This not only includes knowledge about relations between the model parameters, but also about the model order, which is often unknown in experimental situations.

The LCModel performs under the assumption that all the metabolites presented in the MR spectrum should be included in the basis set. If a metabolite is missing from the basis set, it will usually lead to a systematic bias in the estimated parameters, and in particular the concentrations of the metabolites. It is possible that if an “unknown” signal is identified in the spectra, a peak or peaks could be simulated and the spectra can be re-analysed. However, this requires manual interactions, which are both subjective and time-consuming. Furthermore, in practice, where noise is present, the accuracy of determining individual metabolite levels is reduced.

Thus, there is a need for a method that produces an automatic and objective decomposition, particularly when there is a large collection of spectra to analyse.

2.5 Summary

The concepts of MR spectroscopy and its application in living systems are briefly explained in this chapter. Since an MR spectrum can be considered as a linear combination of metabolites and MMLip components, one straight forward option to recover these components is through a linear transformation algorithm. Based on this idea, Chapters 3, 4, 5, 7 and 8 explore in more detail applying the linear transformation algorithm, specifically the ICA algorithm, for metabolites and MMLip components decomposition.

It is also explained that the MR spectrum is a complex valued signal with absorption and dispersion mode Lorentzian lineshapes as its real and imaginary parts, respectively. Although the imaginary part is not commonly used for spectrum analysis, it should contain the same amount of information as in the real part with certain phase difference. In practice, where noise is present, the imaginary part should be a useful addition to the real part in retrieving the

pure spectrum. This is investigated further in Chapter 6, where a more robust de-noising algorithm is proposed by utilising both the real and imaginary parts of an MR spectrum.

CHAPTER 3

LINEAR TRANSFORMATION OF MRS DATA

3.1 Introduction

As discussed in Chapter 2, a proton MR spectrum can be considered as a linear mixture of metabolite and MMLip components with noise. In order to recover these components, a straight forward solution would be to linearly transform the observed spectra back to its individual components. The most popular methods for finding a linear transformation of an observed dataset are the second-order statistics based methods, which only use the information contained in the covariance matrix (second cumulant) of each variable. Principal component analysis (PCA) (Jolliffe, 2002) and factor analysis (Kim and Mueller, 1978) are two classical second-order statistics based methods (Hyvarinen, 1999). As an extension to the classical methods, the ICA (Hyvärinen et al., 2001) is believed to be a more powerful technique, capable of finding the underlying factors or sources when these classic methods fail completely. The use of these techniques is to be understood in the context of the classical assumption of Gaussianity as it will be described in the sections below.

3.2 Moments and cumulants

Assume that x is a real-valued continuous scalar random variable with probability density function (pdf), $p_x(x)$. Every probability distribution is uniquely specified by its characteristic function, and vice versa (Papoulis, 1991). The first characteristic function, $f_1(\omega)$ of x, is defined as the continuous Fourier transform of the pdf (Hyvärinen et al., 2001), $p_x(x)$:

$$f_1(\omega) = E\{\exp(j\omega x)\} = \int_{-\infty}^{\infty} \exp(j\omega x)\, p_x(x) dx \tag{3.1}$$

where ω is the transformed variable corresponding to x. Expanding the characteristic function $f_1(\omega)$ into its Taylor series yields (Papoulis, 1991, Girolami, 1999),

$$f_1(\omega) = \int_{-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \frac{x^k (j\omega)^k}{k!} \right) p_x(x) dx = \sum_{k=0}^{\infty} E\{x^k\} \frac{(j\omega)^k}{k!} \tag{3.2}$$

Thus, the coefficient terms of this expansion are the moments $E\{x^k\}$ of x. So this characteristic function $f_1(\omega)$ is also called the *moment generating function*.

The second characteristic function, $f_2(\omega)$ of x, is given by the natural logarithm of the first characteristic function equation (3.1):

$$f_2(\omega) = \ln(f_1(\omega)) = \ln\left(E\{\exp(j\omega x)\}\right) \tag{3.3}$$

The cumulants κ_k of x are defined in a similar way to the respective moments as the coefficients of the Taylor series expansion of the second characteristic function equation (3.3):

$$f_2(\omega) = \sum_{k=0}^{\infty} \kappa_k \frac{(j\omega)^k}{k!} \tag{3.4}$$

where the *k*th cumulant is obtained as the derivative,

$$\kappa_k = (-j)^k \left. \frac{d^k f_2(\omega)}{d\omega^k} \right|_{\omega=0} \tag{3.5}$$

Moments are one set of descriptive constants of a distribution. Cumulants make up another set of descriptive constants. Cumulants and moments are different, though clearly related. Cumulants are not directly estimable by summitry or integrative processes (Nandi, 1999). For zero-mean distributions, the first three central moments and the corresponding cumulants are identical, but they begin to differ from order four:

$$\begin{aligned} \kappa_1 &= E\{x\} = 0, \\ \kappa_2 &= E\{x^2\}, \\ \kappa_3 &= E\{x^3\}, \text{ and} \\ \kappa_4 &= E\{x^4\} - 3\left(E\{x^2\}\right)^2 \end{aligned} \tag{3.6}$$

The fourth cumulant κ_4 is the definition of kurtosis (Hyvärinen et al., 2001).

Let $\mathbf{x} = [x_1, ..., x_L]^T$ be a zero-mean observation sample of size *L*, then a commonly used consistent estimator of kurtosis is given by:

$$\hat{\kappa}_4 = L\sum (x_i)^4 - 3\left[\sum (x_i)^2\right]^2 \tag{3.7}$$

However, the above estimator is biased. Another frequently used estimator of kurtosis is defined as:

$$\hat{\kappa}_4^U = \frac{(L-1)}{(L-2)(L-3)}\left\{(L+1)\hat{\kappa}_4 + 6\right\} \tag{3.8}$$

It has been proved that this is unbiased for normal distributions (Kendall et al., 1987).

Kurtosis is a measure of the Gaussianity of the probability distribution of a real-valued random variable. For a Gaussian variable its kurtosis is zero. Kurtosis can also be used to further differentiate non-Gaussian variables into a sub-Gaussian distribution, which has a negative kurtosis value, and a super-Gaussian distribution, which has a positive value. Furthermore, the absolute value of kurtosis will be increasing as the non-Gaussianity increases. Higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly sized deviations (Hyvärinen et al., 2001).

Higher-order cumulants measure the departure of a random process, from a Gaussian random process, with an identical mean and covariance function. Thus a Gaussian random process has higher-order cumulants which are identically zero.

Gaussian pdf is completely characterised by its first two moments, the analysis of linear systems and signals has so far been quite effective in many circumstances. It has nevertheless been limited by the assumptions of Gaussianity. When signals are non-Gaussian, the first two moments do not define their pdf and consequently higher-order statistics (of order greater than two), can reveal more information about them than the second-order statistics can do alone.

3.3 Gaussian density

For a *L*-dimensional random variable $\mathbf{x}$, it is said to be Gaussian, if the pdf of $\mathbf{x}$ has the form

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{L/2}(\det \mathbf{C}_{\mathbf{x}})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_{\mathbf{x}})^T \mathbf{C}_{\mathbf{x}}^{-1}(\mathbf{x}-\mathbf{m}_{\mathbf{x}})\right) \tag{3.9}$$

where L is the dimension of $\mathbf{x}$, $\mathbf{m}_{\mathbf{x}}$ is its mean and $\mathbf{C}_{\mathbf{x}} = E\left\{(\mathbf{x}-\mathbf{m}_{\mathbf{x}})(\mathbf{x}-\mathbf{m}_{\mathbf{x}})^T\right\}$ is the covariance matrix of $\mathbf{x}$. The notation det $\mathbf{C}_{\mathrm{x}}$ is used for the determinant of the matrix $\mathbf{C}_{\mathrm{x}}$.

The multivariate Gaussian density (equation 3.9) is completely determined by the mean vector $\mathbf{m}_{\mathbf{x}}$ and covariance matrix $\mathbf{C}_{\mathbf{x}}$ of $\mathbf{x}$. Therefore, all the higher-order moments should also depend only on $\mathbf{m}_{\mathbf{x}}$ and $\mathbf{C}_{\mathbf{x}}$. This implies that higher-order moments do not carry any novel information about the Gaussian distribution. An important consequence of the Gaussian pdf is that linear processing methods based on first- and second-order statistical information are usually optimal for Gaussian data. It is useless to include any higher order information.

Hence the second-order statistical methods, such as PCA and factor analysis, all assume the data being analysed are Gaussian. Besides, the second-order statistics based methods often require only classical matrix manipulations, which are computationally simple. ICA does not bring out anything new compared with standard PCA, if the data are Gaussian (Hyvärinen et al., 2001).

3.4 Uncorrelatedness or independence?

In reality, however, the data often do not follow a Gaussian distribution, and the situation is not as simple as those methods assume. Many experimentally observed datasets, including MR spectra, have super-Gaussian distributions (Hyvärinen et al., 2001). This can be easily verified by calculating the kurtosis of an MRS dataset; as a super-Gaussian signal will have a kurtosis value greater than zero. Data with a super-Gaussian distribution have a probability

density peaked at zero and heavy tails, when compared to a Gaussian density of the same variance.

PCA, intimately related to singular value decomposition (SVD), is widely used in statistical data analysis, feature extraction and data compression (Jolliffe, 2002, Diamantaras and Kung, 1996). The mathematical details of PCA will be explained in Chapter 4. The basic goal of PCA is to reduce the dimension L of the dataset to the dimension m (usually $m << L$), so that the m components explain the maximum amount of variance possible. Indeed, it can be proven that the representation given by PCA is an optimal linear transformation technique, in the mean-square sense (Jolliffe, 2002). However, the mean-square error (MSE) assesses the quality of an estimator in terms of its variation, which is based on the second-order information only. This is only adequate for Gaussian data. Non-Gaussian data may contain a lot of additional information in its higher-order statistics. The flexibility of the ICA approach, by incorporating higher-order statistical information, resides in transforming the PCA ill-posed problem, associated with decorrelated decompositions, into a well-posed problem of independent decompositions, that is, ICA avoids the non-uniqueness associated with PCA (Mutihac and Van Hulle, 2004). By taking the non-Gaussianity into account, Zou et al. (Zou et al., 2006) have argued that for clean non-Gaussian signals or Gaussian noise corrupted signals, ICA-based feature extraction achieves the minimum mismatch (can be evaluated in terms of a statistic distance such as the Kullback-Leibler (KL) distance (Cover and Thomas, 2006)) between the distributions of the represented features and original signals. It could be roughly described that the purpose of the second-order statistics based methods is to find a faithful representation of the data in the sense of MSE, whereas the ICA method tries to find a meaningful representation (Comon, 1994, Hyvärinen et al., 2001).

The results from PCA, when applied on MRS data, usually represent various mixtures of metabolite and MMLip components instead of single ones (Kuesel et al., 1996). It is expected that ICA can give results related more closely to the underlying metabolite and MMLip components of the spectra, compared to the second-order statistics based methods.

We want to find statistically independent components (ICs) under the general case of non-Gaussian data. The first thing to note is that independence is a much stronger property than uncorrelatedness. Two random vectors **x** and **y** are uncorrelated if:

$$E\{\mathbf{x}\mathbf{y}^T\} = E\{\mathbf{x}\}E\{\mathbf{y}^T\} \tag{3.10}$$

One way of stating how independence is stronger than uncorrelatedness is that independence implies nonlinear uncorrelatedness: If two random variables, e.g. x and y, are independent, then any nonlinear transformations of them, $g(x)$ and $h(y)$ are uncorrelated (in the sense that their covariance is zero). In contrast, if two random variables are merely uncorrelated, such nonlinear transformations do not have zero covariance in general.

A key concept that constitutes the foundation of ICA is statistical independence. For simplicity, consider first the case of two different random variables **x** and **y**. The random variable **x** is independent of **y**, if knowing the value of **y** does not give any information on the value of **x**.

Mathematically, statistical independence is defined in terms of probability densities. The random variables **x** and **y** are said to be independent if and only if

$$p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y}) = p_{\mathbf{x}}(\mathbf{x})\,p_{\mathbf{y}}(\mathbf{y}) \tag{3.11}$$

In other words, the joint density $p_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y})$ of **x** and **y** must factorize into the product of their marginal densities $p_{\mathbf{x}}(\mathbf{x})$ and $p_{\mathbf{y}}(\mathbf{y})$.

Uncorrelatedness in itself is not enough to separate the components. If two random variables are independent, they are uncorrelated, but uncorrelatedness does not imply independence. This is something that classical second-order statistics based methods cannot estimate, because they are essentially based on decorrelation (Hyvärinen et al., 2001).

3.5 Summary

Independence is stronger than uncorrelatedness. The ICA transformation should reveal more meaningful information than the PCA transformation. A detailed discussion about the ICA method compared with PCA method will be presented in the next chapter.

CHAPTER 4

INDEPENDENT COMPONENT ANALYSIS (ICA)

This chapter explains the basic concepts of independent component analysis (ICA). It starts by introducing the derivation of the ICA technique. Then the assumptions and uncertainties of the ICA model are discussed. The relationship between the ICA and the well known PCA techniques is also illustrated. The latter discussion is followed by an outline of several representative ICA algorithms. The blind source separation (BSS) and feature extraction (FE)-based ICA approaches are briefly described, with an overall review of the application of ICA to the MRS signal processing problem.

4.1 Definition of ICA

4.1.1 Problem description

A long-standing problem in statistics is how to transform the data so that its essential structure is made more visible or accessible. In neural computation, this belongs to the area of unsupervised learning. The ICA technique has attracted much attention recently in dealing

with this kind of problem and is becoming more and more popular. The idea of ICA was first described by Herault, Jutten and Ans (Ans et al., 1985). Following this work, many ICA algorithms based on various criteria were proposed (Jutten and Herault, 1991, Comon et al., 1991, Sorouchyari, 1991, Pham, 1992, Bell and Sejnowski, 1995, Cardoso and Souloumiac, 1993, Cardoso and Laheld, 1996, Comon, 1994). In this section, we will give a definition of the problem within a mathematic framework.

The ICA technique was initially researched in the context of the BSS problem, which is to find the original independent signals from linear mixtures of those signals, without knowledge of the physical mixing system. ICA could be applied to the analysis of data from many different applications, including digital images, economic indicators, psychometric measurements, etc (Hyvärinen et al., 2001). In many cases, the measurements are given as a set of parallel signals or time series. Typical examples include mixtures of simultaneous speech signals that have been picked up by several microphones, brain waves recorded by multiple sensors, interfering radio signals arriving at a mobile telephone, parallel time series obtained from some industrial process, etc.

Suppose we have L observed data, $\mathbf{x} = [x_1, ..., x_L]^T$ (the superscript ^{T} denotes the transpose), which are the linear mixtures of M independent sources, denoted by $\mathbf{s} = [s_1, ...s_M]^T$ with the linear mixing matrix $\mathbf{A}_{L\times M}$. The linear mixing model can be written as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{4.1}$$

The aim of ICA is to estimate the original independent sources **s** from **x** without knowing any information about $\mathbf{A}$. This can be done by means of estimating an unmixing matrix $\mathbf{B}$,

equal to the pseudo-inverse of **A** , so that

$$\mathbf{y} = \mathbf{Bx} \tag{4.2}$$

where **y** is the estimation of **s** . The mixing and unmixing network is depicted in Figure 4.1, where the mixing part is a weighted summation of the source signals, and the unmixing part is the reverse process of the previous step that theoretically should recover the original source signals.

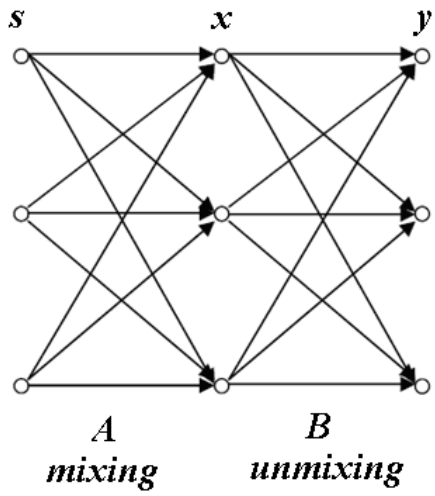


Figure 4.1 Linear mixing and unmixing network.

The ICA technique aims to estimate **B** so as to obtain **y** from **x** based on some statistical properties of the original sources.

4.1.2 Assumptions in ICA

We have to make some assumptions to guarantee that the ICA model can be estimated.

1. The original signals are assumed to be statistically independent[3].

2. The original signals must have non-Gaussian distributions.

As discussed in Chapter 3, the higher-order cumulants are zero for Gaussian distributions, but such higher-order information is essential for estimation of the ICA model. Therefore, ICA is basically impossible for Gaussian variables.

4.1.3 Uncertainties in ICA

The ICA model can estimate the original sources, however, some uncertainties will still remain due to the limitation of the available information about the original sources. The following uncertainties will inevitably hold:

1. The variances (energies) of the estimated independent components cannot be determined. Because the scales of **s** and **A** are unknown, changing the scales of **s** will not affect the model. So the unmixing matrix **B** can be described as $\mathbf{B} = \mathbf{\Lambda}^{-1}\mathbf{A}^{-1}$, where $\mathbf{\Lambda}$ is an invertible diagonal matrix with the diagonal components $\lambda_i \neq 0$.
2. The order of the estimated independent components cannot be determined. The reason is that a permutation matrix **P** and its inverse can be substituted in the model without changing it, i.e., $\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$. Since there is only one component equal to 1 in each row and column of **P** and $\mathbf{P}^{-1}$, the elements of **Ps** are the original independent signals but in another order, and $\mathbf{A}\mathbf{P}^{-1}$ can be treated as a

[3] This is the basic principle of ICA, and mostly the only assumption needed to ensure the estimation of the model. The independence is defined by the probability densities as described in Chapter 3.

new mixing matrix to be estimated by ICA.

4.2 Solution of ICA

In the ICA framework, the unmixing matrix $\mathbf{B}$ is usually decomposed into a whitening matrix $\mathbf{V}$ and an orthonormal ICA matrix $\mathbf{W}$. Hence, the problem shown in Figure 4.1 can be solved using the ICA model as depicted in Figure 4.2.

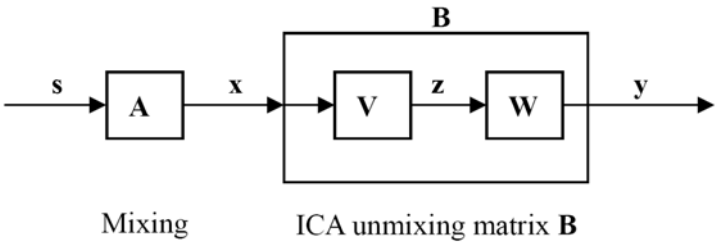


Figure 4.2: The mixing model and the two-step ICA unmixing model

The details of obtaining the whitening and orthonormal matrices are explained in the following subsections. An example of the ICA algorithm in solving a practical problem of two mixtures with two independent components is also given to illustrate, step by step, the ICA model results. The procedures are the same for any number of dimension which is greater than one, however the 2-dimensional example is chosen here as it is visually more obvious.

4.2.1 Centring the variables

Without losing generality, both the mixture variables and the independent components are assumed to have a zero mean. If this assumption is not true, some pre-processing can be

performed to make it hold. The process is called “centring the variables”. This involves subtracting the variables’ sample mean. The ICA models presented in this thesis are all based on using centred variables, where the mean of the observed variable $\mathbf{x}'$ is subtracted:

$$\mathbf{x} = \mathbf{x}' - E\{\mathbf{x}'\} \tag{4.3}$$

where $E\{\cdot\}$ is the expectation operation. The mean is estimated from the available observations $\mathbf{x}'$. The independent components $\mathbf{s}$ are also mean-centred,

$$E\{\mathbf{s}\} = \mathbf{A}^{-1}E\{\mathbf{x}\} \tag{4.4}$$

and the mixing matrix remains the same after this pre-processing, as follows:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}' - E\{\mathbf{x}'\} \\ &= \mathbf{A}\mathbf{s}' - \mathbf{A}E\{\mathbf{s}'\} \\ &= \mathbf{A}(\mathbf{s}' - E\{\mathbf{s}'\}) \\ &= \mathbf{A}\mathbf{s} \end{aligned} \tag{4.5}$$

where $\mathbf{s}'$ is the non-zero mean independent components corresponding to $\mathbf{x}'$.

So this can always be done without affecting the estimation of the mixing matrix **A**. After estimating the mixing matrix and the independent components for the zero-mean data, the subtracted mean can be simply reconstructed by adding $\mathbf{A}^{-1}E\{\mathbf{x}'\}$ to the zero-mean ICs after the estimation.

$$\begin{aligned} \mathbf{s}' &= \mathbf{s} + E\{\mathbf{s}'\} \\ &= \mathbf{s} + \mathbf{A}^{-1}E\{\mathbf{x}'\} \end{aligned} \tag{4.6}$$

4.2.2 Whitening in ICA

Unlike correlation (second-order statistics) based transformations, e.g. PCA, ICA not only makes use of second-order statistics but it also utilises higher-order statistics to make the signals as independent from each other as possible. Nevertheless, PCA is often used as the first step in ICA to transform the variables into uncorrelated ones, a process which is typically called whitening (Comon, 1994). Sometimes, dimensionality reduction (Fukunaga, 1990) can also be performed if the number of independent components is less than the number of source variables. Further details on dimensionality reduction will be discussed in Chapter 5.

Whitening is essentially a linear transformation of the observed data vector **x** into a set of uncorrelated variables, as described by the following equation:

$$\mathbf{z} = \mathbf{V}\mathbf{x} \tag{4.7}$$

where $\mathbf{V}$ is the whitening matrix, and the obtained data $\mathbf{z}$ is whitened.

A zero-mean random vector $\mathbf{z} = [z_1 \ldots z_L]^T$ is said to be white if its elements z_i are uncorrelated and have unit variances (Hyvärinen et al., 2001):

$$E\{z_i z_j\} = \delta_{ij} \tag{4.8}$$

In terms of the covariance matrix, this means that,

$$E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I} \tag{4.9}$$

where **I** is the unit matrix.

One popular method for whitening is to use the eigenvalue decomposition of the covariance matrix (Hyvärinen et al., 2001). This is straightforward in terms of the PCA expansion, as seen in the equation:

$$E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T \tag{4.10}$$

where $\mathbf{E}$ is the orthogonal matrix with the column vectors being eigenvectors $\mathbf{e}_i$ of the covariance matrix $\mathbf{C}_\mathbf{x}=E\{\mathbf{x}\mathbf{x}^T\}$, and $\mathbf{D}$ is the diagonal matrix of the corresponding eigenvalues, represented by $\mathbf{D} = \text{diag}(d_1,...,d_L)$.

The eigenvalues of $\mathbf{C}_\mathbf{x}$ also give the variances of the principal components. Note that, because the principal components have zero means, a small eigenvalue (a small variance) indicates that the value of the corresponding principal component is mostly close to zero.

An important application of PCA is data compression (Lv and Zhao, 2005, Majumdar, 2009, Jolliffe, 2002). It can be shown (Diamantaras and Kung, 1996) that the value of the minimum MSE is,

$$J_{MSE}^{PCA} = \sum_{i=m+1}^{L} d_i \tag{4.11}$$

The sum of eigenvalues corresponding to the discarded eigenvectors $\mathbf{e}_{m+1}$, …, $\mathbf{e}_L$.

The error decreases when more terms are included (m increases) in equation (4.11). The error will equal to zero when $m = L$. However, a very important practical problem is how to choose m. This is a trade-off between error and the amount of data needed. A threshold is needed below which the eigenvalues, hence the principal components, are small and can be

considered insignificant. This limit determines how many principal components are used. It is usually set by choosing the minimum number of principal components that would explain the data well enough. However, in practice, with no use of prior information about the observations, its dimension is actually chosen by trial and error with no theoretical guidelines (Hyvärinen et al., 2001). A dimension reduction method for the MRS data is proposed in this thesis, and will be explained in detail in Chapter 5.

Now the whitening transformation can be easily calculated as

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^{T} \tag{4.12}$$

Since whiteness is related to independence, one could hope that whitening solves the problem. As shown in Figure 4.3(c), now the mixtures are uncorrelated, but the distribution is clearly not the same as the original one in Figure 4.3(a). The independent components are still mixed by an orthogonal mixing matrix, which corresponds to a rotation of the plane. This shows that PCA is not enough to obtain independent sources. PCA is based on the second-order statistics (covariance), while independence is measured by higher-order statistics (such as kurtosis).

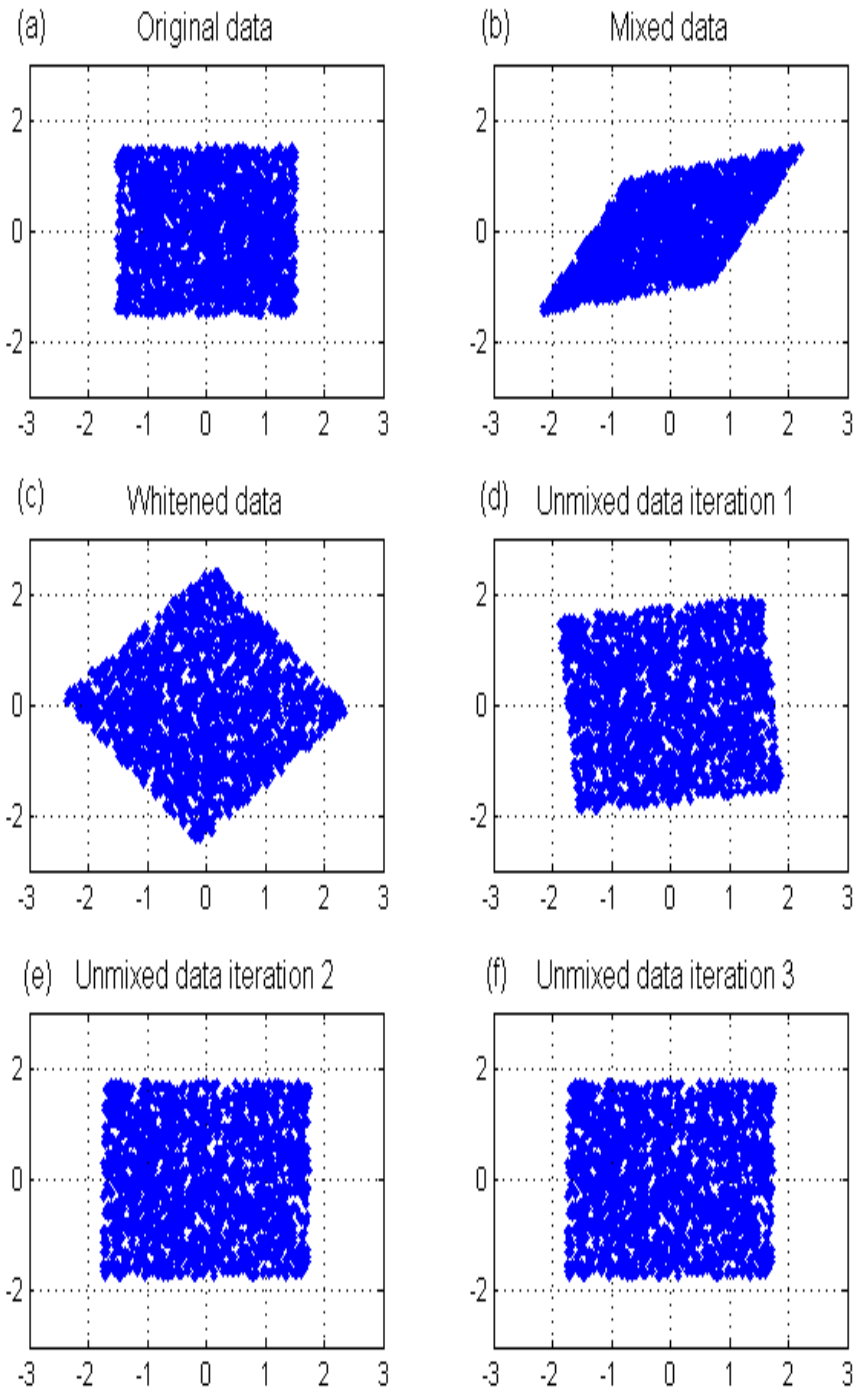


Figure 4.3: Illustration of the ICA algorithm: (a) the joint distribution of two uniform distributed original independent components s1 and s2. Horizontal axis: s1, vertical axis: s2, (b) the joint distribution of the observed mixtures x1 and x2. Horizontal axis x1, vertical axis: x2, (c) the joint distribution of the whitened mixtures of the independent components, where we can see that the whitening (PCA unmixing) is still a rotation away from its original form. (d) the joint distribution of estimated original independent components after one training iteration, (e) the joint distribution of estimated independent components after two training iterations, and (f) the joint distribution of estimated independent components after three training iterations. Gradually, the second half of ICA algorithm “rotates” the unmixed data back to its original form.

To show that the whitening is not enough for the estimation of independent components, consider an orthogonal transformation $\mathbf{U}$ of the whitened vector **z**:

$$\mathbf{y} = \mathbf{U}\mathbf{z} \tag{4.13}$$

Due to the orthogonality of $\mathbf{U}$, we have:

$$E\{\mathbf{y}\mathbf{y}^T\} = E\{\mathbf{U}\mathbf{z}\mathbf{z}^T\mathbf{U}^T\} = \mathbf{U}\mathbf{I}\mathbf{U}^T = \mathbf{I} \tag{4.14}$$

where **I** is the identity matrix. This shows that $\mathbf{y}$ is white as well; there is no unique solution for the independent components by using the whiteness property alone. Since **y** could be any orthogonal transformation of **z**, whitening gives the ICs only up to an orthogonal transformation, which is not enough in most cases.

Uncorrelatedness is weaker than independence, and is not in itself sufficient for the estimation of the ICs. On the other hand, whitening exhausts the second-order information that is contained in the covariance matrix, and makes it easier to use the higher-order information available in the whitened vector **z**. Therefore, whitening is only the first half of ICA.

4.2.3 The second half of ICA estimation

Although the PCA technique cannot solve the ICA problem, we have shown that it can significantly simplify the computation. After the data in the ICA model is whitened by the matrix **V** given in equation (4.12), the mixing matrix becomes a new one, $\tilde{\mathbf{A}}$, from equations (4.1) and (4.7),

$$\mathbf{z} = \mathbf{Vx} = \mathbf{VAs} = \tilde{\mathbf{A}}\mathbf{s} \tag{4.15}$$

The benefit of whitening resides in the fact that the new mixing matrix $\tilde{\mathbf{A}} = \mathbf{VA}$ is orthonormal. As illustrated in Figure 4.3, the transformation from **s** (Figure 4.3(a)) to **z** (Figure 4.3(c)) is just a rotation. It can also be seen from the following equation that the transpose of $\tilde{\mathbf{A}}$ is equal to its inverse, which is the definition of orthonormal matrix:

$$E\{\mathbf{z}\mathbf{z}^T\} = \tilde{\mathbf{A}}E\{\mathbf{s}\mathbf{s}^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I} \tag{4.16}$$

This means we can restrict the finding of the mixing matrix to an orthonormal matrix $\mathbf{W}$, so that

$$\mathbf{y} = \mathbf{Wz} \tag{4.17}$$

where $\mathbf{y}$ is a good estimate of the ICs in the original signal.

As it has been shown by Hyvärinen et al. (Hyvärinen et al., 2001), performing whitening first is a good way to reduce the complexity of computation. Now the ICA problem is reduced to estimate an orthonormal matrix $\mathbf{W}$ from **z**. According to the central limit theorem (Johnson, 2004), we know that the distribution of the sum of non-Gaussian random variables tends to be closer to Gaussian than that of the original variables. It means that the distribution of a sum of independent random variables tends toward a Gaussian distribution, and it will be maximally non-Gaussian if it equals one of the independent components. Therefore, a quantitative measure of independency is the degree of non-Gaussianity of the signal. The $\mathbf{W}$ can be found by maximizing the non-Gaussianity of $\mathbf{y}$, which can be measured in terms of higher-order

statistics.

4.3 ICA algorithms

Various ICA algorithms have been proposed (Hyvärinen et al., 2001, Hyvarinen and Oja, 2000, Bell and Sejnowski, 1995, Cardoso and Souloumiac, 1993, Cardoso and Laheld, 1996), most of them based on optimizing some criteria functions related to the higher-order statistics. Those algorithms can be categorized into explicit and implicit higher-order statistics algorithms. The explicit higher-order statistics algorithms directly employ the higher-order statistics, which are usually presented in terms of the fourth-order cumulant of the independent components, as the criterion function. The implicit higher-order statistics algorithms use some properties of independence in terms of mutual information (Coombs et al., 1970), likelihood (Pawitan, 2001), negentropy (Hyvarinen, 1999), etc., which are in turn fundamentally based on the higher-order statistics (Cardoso, 1999).

4.3.1 Explicit higher-order statistics algorithms

This family of ICA algorithms usually uses the fourth-order cumulant (Nandi, 1999) as the criterion function. Let y_i, y_j, y_k, y_l denote four zero mean variables, the fourth-order cumulant is defined as (Priestley, 1981):

$$\begin{aligned} &cum_4(y_i, y_j, y_k, y_l) \\ &= E\{y_i y_j y_k y_l\} - E\{y_i y_j\}E\{y_k y_l\} - E\{y_i y_k\}E\{y_j y_l\} - E\{y_i y_l\}E\{y_k y_j\} \end{aligned} \tag{4.18}$$

A cumulant involving at least two different variables is called a cross-cumulant; otherwise it is called an auto-cumulant. The fourth-order auto-cumulant is also referred to as kurtosis, as defined in Chapter 3, equation (3.6):

$$k(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{4.19}$$

which can be used as a simple measure of non-Gaussianity if the signals to be compared are of the same type.

From equation (4.8), it is obvious that for the whitened variables, the correlation terms in equations (4.18, 4.19) will be constant. This shows that whitening is an essential step in reducing the complexity of calculating the fourth-order cumulant and consequently the ICA estimation.

The EASI (Cardoso and Laheld, 1996) and JADE (Cardoso and Souloumiac, 1993) (Cardoso, 1999) are two widely used cumulant-based ICA algorithms. The EASI algorithm uses the kurtosis as its criterion function, where the gradient method is applied to optimize the kurtosis of **y**. In the JADE algorithm, a set of fourth-order cumulant matrices are constructed, whose diagonal components are the auto-cumulants, and the off-diagonal components are the cross-cumulants. The Jacobi optimization algorithm (Golub and Van Loan, 1989) is applied to jointly diagonalise the cumulant matrices, i.e., make the off-diagonal components close to zero. The JADE algorithm can be summarized as:

1. Initialization. Estimate a whitening matrix $\hat{\mathbf{W}}$ and set $\mathbf{Z} = \hat{\mathbf{W}}\mathbf{X}$.
2. Form statistics. Estimate a maximal set $\left\{\hat{\mathbf{Q}}_i^{\mathbf{Z}}\right\}$ of cumulant matrices.
3. Optimize an orthogonal contrast. Find the rotation matrix $\hat{\mathbf{V}}$ such that the cumulant matrices are as diagonal as possible, that is, solve $\hat{\mathbf{V}} = \arg\min \sum_i off\,(\mathbf{V}^T \hat{\mathbf{Q}}_i^{\mathbf{Z}} \mathbf{V})$.
4. Separate. Estimate $\mathbf{A}$ as $\hat{\mathbf{A}} = \hat{\mathbf{V}}\hat{\mathbf{W}}^{-1}$ and/or estimate the components as $\hat{\mathbf{S}} = \hat{\mathbf{A}}^{-1}\mathbf{X} = \hat{\mathbf{V}}^T\mathbf{Z}$.

4.3.2 Implicit higher-order statistics algorithms

For this group of algorithms, the objective functions used can be any suitable nonlinear functions as measures of non-Gaussianity. They are normally selected based on the information-theoretic quantity of differential entropy (Cover and Thomas, 2006). The most representative algorithms are infomax[4] (Bell and Sejnowski, 1995, Linsker, 1988), which minimises mutual information, and negentropy-based FastICA (Hyvärinen et al., 2001), which maximises non-Gaussianity.

Infomax approach

According to (Bell and Sejnowski, 1995, Cardoso, 1997), the application of the infomax principle to source separation actually rests with minimizing the differential entropy of the output independent components. Entropy (Shannon, 1948), denoted by *H*, is the basic concept of information theory as a measure of the uncertainty associated with a random variable, and is defined as,

$$H(\mathbf{x}) = -E\{\ln p(\mathbf{x})\} \tag{4.20}$$

where $p(\mathbf{x})$ is the distribution density function of $\mathbf{x}$. However, simple minimization would be inappropriate because the entropy of $\mathbf{y} = \mathbf{W}\mathbf{z}$ diverges to infinity for an arbitrary $\mathbf{W}$. As such, the real axis $(-\infty, +\infty)$ will be mapped to the interval (0,1) by using some nonlinear monotonously increasing functions *g*.

The infomax principle is implemented by minimizing the entropy of nonlinear mapping

[4] Infomax is an optimization principle for neural networks principle of maximization of information flow.

output $\mathbf{u} = g(\mathbf{y}) = g(\mathbf{Wz})$. Hence the contrast function Ψ will be defined as:

$$\Psi = H(\mathbf{u}) = H(g(\mathbf{Wz})) \tag{4.21}$$

where H is the entropy as defined in equation (4.20), and g is the nonlinear function that maps the real axis $(-\infty, +\infty)$ to the interval (0,1). One option of g is the distribution probability function as the range of the probability values fall into that interval. In this case, the differential of g will be the distribution density function of y_i, denoted by $p(y_i)$. We will have $p(y_i) = \dfrac{dg(y_i)}{dy_i}$ and $g(y_i) = \int_{-\infty}^{y_i} p(u)du$.

The multivariate probability density function of $\mathbf{y}$ can be written as:

$$p(\mathbf{y}) = \frac{p(\mathbf{z})}{|\mathbf{J}|} \tag{4.22}$$

where $\mathbf{J}$ is the Jacobian of the transformation. The Jacobian is the determinant of the matrix of partial derivatives:

$$\mathbf{J} = \det \begin{bmatrix} \dfrac{\partial y_1}{\partial z_1}, \cdots \dfrac{\partial y_1}{\partial z_N} \\ \vdots \\ \dfrac{\partial y_N}{\partial z_1}, \cdots \dfrac{\partial y_N}{\partial z_N} \end{bmatrix} \tag{4.23}$$

From equations (4.20, 4.22), the entropy can be written as

$$H(\mathbf{u}) = E\{\ln |\mathbf{J}|\} - E\{\ln p(\mathbf{z})\} \tag{4.24}$$

Because $p(\mathbf{z})$ has no relation to the unmixing matrix **W**, in order to minimize the entropy, we need only concentrate to minimize the first term $E\{\ln|\mathbf{J}|\}$. So that the contrast function Ψ can be written as

$$\Psi = E\{\ln|\mathbf{J}|\} \tag{4.25}$$

In the infomax algorithm, a sigmoid function (Mitchell, 1997), which has an "S" shape in the interval of (0,1), is proposed in place of the nonlinear function g. The ICA matrix **W** can be obtained by optimizing the contrast function (equation (4.25)) using a gradient algorithm.

Negentropy-based FastICA approach

The information theory illustrates that the non-Gaussianity of signals can be evaluated in terms of entropy, which is maximized by Gaussian distributed signals and monotonously decreases as the non-Gaussianity of signals increases. The ICA can be performed by optimizing the entropy-based criterion function. A normalized version of entropy is often used, which is referred to as negentropy,

$$J(\mathbf{y}) = H(\mathbf{y}_{Gauss}) - H(\mathbf{y}) \tag{4.26}$$

The negentropy is zero for Gaussian signals and always non-negative. However, it is computationally very difficult to obtain, due to the integral of the probability functions in the expectation operation in equation (4.20) for entropy. To solve this problem, some approximation of negentropy is usually used. In the negentropy-based FastICA algorithm (Hyvärinen et al., 2001), the negentropy is approximated in terms of two non-polynomial functions, one odd and one even. The following approximation can be obtained:

$$J(\mathbf{y}) = k_1(E\{G^1(\mathbf{y})\})^2 + k_2(E\{G^2(\mathbf{y})\} - E\{G^2(\mathbf{v})\})^2 \tag{4.27}$$

where k_1 and k_2 are constants, $G^1(\cdot)$ is the odd function, $G^2(\cdot)$ is the even function, and $\mathbf{v}$ is a Gaussian variable with zero mean and unit variance. If $\mathbf{y}$ has a symmetric distribution, the first term of equation (4.27) will be zero. It can be written as

$$J(\mathbf{y}) = k(E\{G(\mathbf{y})\} - E\{G(\mathbf{v})\})^2 \tag{4.28}$$

The negentropy can be maximized by using the gradient algorithm, and $\mathbf{W}$ can be obtained by

$$\mathbf{w} \leftarrow \mathbf{w} + \gamma_s E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} \tag{4.29}$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \tag{4.30}$$

where $\mathbf{w}$ is a row of $\mathbf{W}$, γ_s is step size, $g(\cdot)$ is the first derivative of $G(\cdot)$, which can be chosen from one of the following functions:

$$g_1(\mathbf{y}) = \tanh(a_1\mathbf{y}) \tag{4.31}$$

$$g_2(\mathbf{y}) = \mathbf{y}\exp(-\mathbf{y}^2/2) \tag{4.32}$$

$$g_3(\mathbf{y}) = \mathbf{y}^3 \tag{4.33}$$

The drawbacks of the gradient algorithm are that it slowly converges and is sensitive to parameters, such as γ_s and initial value of $\mathbf{W}$. As $\mathbf{W}$ is a unitary matrix, it can be proven that the gradient of $J(\mathbf{y})$ must point in the same direction as $\mathbf{W}$. Based on that fact, a faster and

more reliable ICA algorithm can be obtained to replace equation (4.29) with,

$$\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z})\} - E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{w} \tag{4.34}$$

where $g'(\cdot)$ is the first derivative of $g(\cdot)$. The corresponding functions of $g'(\cdot)$ are

$$g'_1(\mathbf{y}) = a_1(1-\tanh^2(a_1\mathbf{y})) \tag{4.35}$$

$$g'_2(\mathbf{y}) = (1-\mathbf{y}^2)\exp(-\mathbf{y}^2/2) \tag{4.36}$$

$$g'_3(\mathbf{y}) = 3\mathbf{y}^2 \tag{4.37}$$

Equations (4.34, 4.30) form a basic iteration in negentropy-based FastICA algorithm. The converged **W** is the desired orthonormal matrix. The negentropy-based FastICA algorithm, proposed by Hyvärinen et al. (Hyvärinen et al., 2001), for finding one maximally non-Gaussian direction can be described as:

---

1. Centre the dataset to make the mean zero.
2. Reduce the dimension and whiten the dataset to obtain **z**.
   - Find the eigenvectors and eigenvalues of the covariance matrix.
   - Set a limit for the cumulated energy content of the eigenvectors, and select the subset of the eigenvectors accordingly as basis vectors.
   - Compute the whitening matrix V from equation (4.12).
3. Initialize the matrix **W**, e.g. to the unit matrix.
4. Update **W** according to equation (4.34), then orthogonalize and normalize it.
5. If **W** is not converged, go to step 4.
6. The unmixing matrix **B** is calculated from the relation **B** = **WV**; consequently, the mixing matrix and ICs are obtained.

The negentropy-based FastICA in the FastICA MATLAB package (FastICApackage) is applied in all the experiments performed in this thesis.

4.4 Blind source separation (BSS) and feature extraction (FE)-based ICA

ICA can be used to extract independent components from different kinds of observed variables. The BSS separates the original independent sources from the observed variable, whereas the FE expresses the observed variable in terms of the weighted summation of some bases. The fundamental difference between BSS and FE methods when applying ICA is their assumptions regarding statistical independence. The goal of BSS-based ICA is to recover any independent sources given only the linear mixtures of these sources as observations. The observations are the weighted summation of many sources; the goal of FE-based ICA is to estimate the weighting coefficients of the sources in the mixture.

When applied to MRS data, the metabolite and MMLip components are assumed independent from each other for the BSS-based ICA, and their corresponding concentrations in the target voxel are assumed independent for FE-based ICA.

Various efforts have shown that the ICA approach can obtain certain individual metabolite and MMLip components.

In the following studies, the FE-based ICA is applied. The FE-based ICA algorithm was applied together with the SVD technique for de-noising and quantification of the MRS signals in the study by Stamatopoulos et al. (Stamatopoulos et al., 2009). The SVD algorithm was applied first to split the MRS signal into signal subspace and noise subspace. The ICA

transformation was applied in both of the subspaces. The ICA technique was used as a FE method for classifying brain tumours on *in vivo* magnetic resonance spectroscopy (MRS) data (Lee et al., 2000, Huang et al., 2003, Menze et al., 2006) and on magnetic resonance spectroscopic imaging (MRSI) data (Simonetti et al., 2005). In the study by Menze et al. (Menze et al., 2006) and Simonetti et al. (Simonetti et al., 2005), the FE-based ICA technique was compared for classification performance with PCA, simple quantitation and LCModel, but was not shown to be useful. The FE-based ICA was also applied to extract the individual metabolites from the MRS signal (Ma and Sun, 2005). The ICA technique was used as a feature extraction method for classifying low- and medium-grade astrocytic tumours (Huang et al., 2003). The ICA feature-based classification showed comparable results with using a subset of the original variables. Similarly, the FE-based ICA was applied for the discrimination of five classes of brain tumours (Lee et al., 2000). The ICA-based algorithm obtained comparable classification results against the PCA- and LDA-based algorithms. It was also applied on brain tumour MRS data for LDA and support vector machine (SVM)-based classification and achieved accurate performance (Luts et al., 2008). The FE-based ICA was applied for classifying the acceptable/unacceptable data (quality control) using SVM and obtained a classification accuracy of over 90% (Wright et al., 2008).

The BSS-based ICA is applied in the following studies. Szabo de Edelenyi et al. (Szabo de Edelenyi et al., 2005) applied the BSS-based ICA on a set of brain tumours MRSI data to reveal the necrosis, tumoural tissue and healthy tissue. The results showed the reliability of the independent component obtained by ICA. This approach was applied for decomposition of *in vivo* MR spectra by Ladroue et al. (Ladroue et al., 2003) and it showed that ICA can extract certain biochemical components from a large dataset. In the study by Pulkkinen et al. (Pulkkinen et al., 2005), the BSS-based ICA was also used to decompose MRSI data of brain

tumours into diagnostically useful components. All these studies applied ICA to experimental data.

From these previous studies, it was unclear under what conditions ICA could extract individual metabolite and MMLip components and, therefore, what the precise benefits of ICA were. While in this chapter the principle of ICA is explained in detail, in Chapter 7, a systematic comparison of the results of ICA applied to *in vivo* MRS datasets acquired under different experimental conditions is performed to evaluate the potential benefits of ICA for the interpretation of clinical MR spectra. A detailed explanation of the two approaches to ICA outlined here will be provided in the next chapter.

4.5 Summary

The theoretical background and the solutions of ICA algorithms were reviewed in this chapter. By employing higher-order statistics, ICA acknowledges the non-Gaussian distributions of practical signals, which have been ignored by classical methods. It can be seen as an extension to PCA, but ICA is a much more powerful technique. The central limit theorem shows that non-Gaussianity is a measure of independence. The independent components can be obtained by finding directions in which the data is maximally non-Gaussian. Various methods of measuring the non-Gaussianity form the bases of several different ICA estimation algorithms. The two main approaches of the ICA applications were reviewed as well, and details of their applications on MRS data will be investigated in the next chapter.

CHAPTER 5

MATERIALS AND METHODS

5.1 Introduction

As explained in Chapter 4, the ICA algorithm can be applied in two ways on an MRS dataset. One is based on the assumption that the metabolite and MMLip components are independent from each other, which is referred to as the BSS-based ICA method. The other, namely the FE-based ICA method, is based on the assumption that the independency is only on their concentrations in the MR spectra. However, we can see that there are some metabolites and MMLip components which are un-independent from each other just by visually inspect their spectra lineshape; and there are chemical reactions between certain metabolite and MMLip components which will make their concentrations dependent from each other. Which of the above assumptions fits more closely to the true model of the MRS case will be revealed in Chapter 7, and a novel approach of combining the two assumptions to better fit the model of MRS will be explained in Chapter 8.

For a given MRS dataset, once the concentrations are known, the corresponding metabolite

and MMLip components can be uniquely determined. The methods of applying BSS-ICA and FE-ICA to an MRS dataset are presented in the first part of this chapter.

To reduce confusion in explaining the concepts of both BSS- and FE-based ICA on MRS dataset, the terms “basis” and “independent component (IC)” are referring to the individual metabolite or MMLip components in this thesis.

The datasets used in this thesis consist of both synthesised and experimental MR spectra. The synthesised MR spectra were taken to be linear combinations of several most commonly seen metabolite and MMLip components in *in vivo* MR spectra of childhood cerebellar tumours. The experimental dataset contains MR spectra obtained from brain tumours in children before, during and after treatment, the details of which are explained in the second part of this chapter.

5.2 Methods

5.2.1 Blind source separation by ICA

Since the proton MR spectra can be considered as a linear mixture of metabolite and MMLip components with noise, the retrieval of these ICs from a set of MR spectra can be considered as a BSS problem.

Suppose we have a dataset of l MR spectra, each of which has a dimension d, denoted by $\mathbf{X}_{l\times d}$, and assume there are m ($l \geq m$) independent metabolite and MMLip components, denoted by $\mathbf{S}^{BSS}_{m\times d}$. Then, equation (4.1) can be extended into the matrix notation, as follows:

$$\mathbf{X}_{l\times d} = \mathbf{A}^{BSS}_{l\times m}\ \mathbf{S}^{BSS}_{m\times d} \qquad (5.1)$$

where $\mathbf{X}_{l\times d}$ is the observed MRS dataset with MR spectra as its rows, $\mathbf{A}_{l\times m}^{BSS}$ is the matrix of the mixing coefficients for the metabolite and MMLip components and $\mathbf{S}_{m\times d}^{BSS}$ contains the metabolite and MMLip components on its rows.

Then ICA algorithm can be applied to $\mathbf{X}_{l\times d}$ to derive $\mathbf{S}_{m\times d}^{BSS}$ by maximizing the independency of $\mathbf{B}_{m\times l}^{BSS}\ \mathbf{X}_{l\times d}$, where

$$\mathbf{B}_{m\times l}^{BSS} = pinv(\mathbf{A}_{l\times m}^{BSS}) \tag{5.2}$$

The assumption taken here is that the metabolite and MMLip components are statistically independent from each other.

5.2.2 Feature extraction by ICA

The FE-ICA method is based on the idea that the observed signal is a composition of many bases with different weighting coefficients. In this case, the bases are spectra of metabolite and MMLip components and composition weighting coefficients indicate their concentrations in the MR spectra. Assuming these metabolite and MMLip components are mixed randomly, then the mixing coefficients can be considered to be independent from each other. In the context of feature extraction, the metabolite and MMLip components are indirectly achieved via estimating those independent mixing coefficients by using the ICA criterion. Once those coefficients are properly estimated, the bases of those coefficients will be the expected metabolite and lipid components.

Considering the same dataset in the previous subsection, with the l MR spectra of dimension d, and assuming the dimension is to be reduced to m ($m \leq d$), equation (4.1) can be rewritten in

matrix form as:

$$\mathbf{X}_{d\times l}^{T} = \mathbf{A}_{d\times m}^{FE}\ \mathbf{S}_{m\times l}^{FE} \tag{5.3}$$

where $\mathbf{X}_{d\times l}^{T}$ is the observed MRS dataset with each MR spectrum as its columns, $\mathbf{A}_{d\times m}^{FE}$ is the bases matrix with metabolite and MMLip components as its columns and $\mathbf{S}_{m\times l}^{FE}$ the matrix of the corresponding concentration coefficients.

The FE-based ICA algorithm will recover the concentration coefficients $\mathbf{S}_{m\times l}^{FE}$, and as such $\mathbf{A}_{d\times m}^{FE}$ can be determined.

5.2.3 Noise

In practice, the observed data matrix $\mathbf{X}$ is always corrupted by a certain amount of noise, and our mixing model needs to be rewritten as,

$$\mathbf{X}^{\text{obs}} = \mathbf{X} + \mathbf{N} \tag{5.4}$$

where $\mathbf{X}^{\text{obs}}$ is the observed data matrix, $\mathbf{X}$ is the clean data matrix and $\mathbf{N}$ is additive Gaussian noise. Any independent component estimated from the mixture $\mathbf{X}^{\text{obs}}$ is also corrupted by noise. To obtain ICs closer to their true form, we need to remove the noise from the observed data matrix before applying ICA. In practice, the additive noise $\mathbf{N}$ is removed by using signal enhancement algorithms.

A wavelet shrinkage de-noising (WSD)-based algorithm is also proposed for the de-noising of MR spectrum, where a relationship between the real and imaginary parts of the MR spectrum

is derived and applied together with WSD. The detailed de-noising algorithm will be presented in Chapter 6.

5.2.4 Feature dimension selection

As already discussed in section 4.2.2, feature dimension selection is usually performed based on the PCA eigenvalue criterion. The number of ICs is usually chosen by finding the minimum number of principal components that explain the data well enough, which is indicated by their corresponding eigenvalues in terms of containing, for example, 95% (this number is arbitrarily chosen to aid the explanation of the concept, its value varies for each case) of the variance. As discussed before in section 4.2.2, the number of dimensions is actually chosen by trial and error with no theoretical guidelines. The reason may be that the information level in the dataset cannot be estimated in most cases.

Here we propose a general guideline of dimension selection for the MRS dataset. Since the variance can be considered as the power of the signal, 95% of the variance means that the clean signal, denoted as vector **x**, has 95% of power over the noisy signal $\mathbf{x}^{obs}$ and the remaining 5% is the noise **n**. Since the power of the clean signal is closely related to signal-to-noise ratio (SNR), it could be used to estimate the percentage of the clean signal. Hence the information content in the data can be treated specifically for each dataset. A general definition of the estimated SNR_E is the ratio of the clean signal power over noise power:

$$SNR_E = \frac{P_{cleansignal}}{P_{noise}} = \frac{\sum_{i=1}^{L} x_i^2}{\sum_{i=1}^{L} n_i^2} \tag{5.5}$$

where x_i is the i^{th} element of vector **x**, and n_i is the i^{th} element of vector **n**, respectively. L is the length of these vectors. We can easily calculate the percentage θ of the clean signal variance over the noisy one by the following equation. Hence the number of informative ICs can be obtained accordingly, as follows:

$$\begin{aligned}\theta &= \frac{P_{cleansignal}}{P_{noisysignal}} \times 100\% = \frac{\sum_{i=1}^{L} x_i^{\,2}}{\sum_{i=1}^{L} (x_i^{\text{obs}})^2} \times 100\% \\ &= \frac{\sum_{i=1}^{L} x_i^{\,2}}{\sum_{i=1}^{L} (x_i + n_i)^2} \times 100\% \\ &= \frac{\sum_{i=1}^{L} x_i^{\,2}}{\sum_{i=1}^{L} x_i^{\,2} + \sum_{i=1}^{L} n_i^{\,2} + \mathbf{2}L\, E\{\mathbf{x}\}E\{\mathbf{n}\}} \times 100\% \\ &= \frac{\sum_{i=1}^{L} x_i^{\,2}}{\sum_{i=1}^{L} x_i^{\,2} + \sum_{i=1}^{L} n_i^{\,2}} \times 100\% \\ &= \frac{SNR_{\text{E}}}{SNR_{\text{E}} + 1} \times 100\%\end{aligned} \tag{5.6}$$

where θ is the percentage threshold, x_i^{obs} is the i^{th} element of vector x^{obs}, and $E\{\cdot\}$ denotes expectation operation. From the definition, we know that the expectation of white Gaussian vector, $E\{\mathbf{n}\}$, is zero. The following relationship between the powers (variances) of the noisy, clean and noise signals exists:

$$\sum_{i=1}^{L} (x_i^{\text{obs}})^2 = \sum_{i=1}^{L} x_i^{\,2} + \sum_{i=1}^{L} n_i^{\,2} \tag{5.7}$$

This thesis follows the LCModel definition of the SNR, which is the ratio of the highest signal intensity in the 0.2-4 ppm region to twice the root mean square (RMS) of the LCModel fit residuals, as in equation (5.8):

$$SNR_{\mathrm{LC}} = \frac{\max(\mathbf{x})}{2 \times \mathrm{RMS}(\mathbf{n})} \tag{5.8}$$

Since it is not obvious how to relate this to the signal power, we need to convert the SNR_{LC} to SNR_{E}, in order to guide the dimension reduction. The following equation converts the LCModel SNR to the general SNR based on equations (5.5, 5.7 and 5.8):

$$\begin{aligned} SNR_{\mathrm{E}} &= \frac{\sum_{i=1}^{L} x_i^2}{\sum_{i=1}^{L} n_i^2} = \frac{\sum_{i=1}^{L} (x_i^{\mathrm{obs}})^2 - \sum_{i=1}^{L} n_i^2}{\sum_{i=1}^{L} n_i^2} = \frac{L\, E\{(\mathbf{x}^{\mathrm{obs}})^2\}}{\sum_{i=1}^{L} n_i^2} - 1 \\ &= \frac{E\{(\mathbf{x}^{\mathrm{obs}})^2\}}{\sum_{i=1}^{L} n_i^2 \Big/ L} - 1 = \left(\frac{\max(\mathbf{x})}{2 \times \sqrt{\sum_{i=1}^{L} n_i^2 \Big/ L}} \right)^2 \frac{4 \times E\{(\mathbf{x}^{\mathrm{obs}})^2\}}{\left(\max(\mathbf{x})\right)^2} - 1 \\ &= SNR_{\mathrm{LC}}^2 \frac{4 \times E\{(\mathbf{x}^{\mathrm{obs}})^2\}}{\left(\max(\mathbf{x})\right)^2} - 1 \end{aligned} \tag{5.9}$$

The max($\mathbf{x}$) can be approximated by the maximum of the de-noised $\mathbf{x}^{\mathrm{obs}}$. The feature dimension selection will be performed based on the SNR_{E} obtained from SNR_{LC} using equation (5.9).

The percentage threshold θ is calculated using equations (5.6, 5.9). In the PCA dimension reduction, the eigenvalues are sorted in descending order. The number of relevant components, m, can be obtained when the first m eigenvalues accumulates over the threshold θ. This can be

expressed as:

$$\sum_{i=1}^{m} d_i \geq \theta, \text{ and } \sum_{i=1}^{m-1} d_i < \theta \qquad (5.10)$$

where d_i is the percentage of the i^{th} eigenvalue in descending order.

The SNR chosen for each dataset is an indication of the overall quality of MR spectra in the dataset. A few outliers with very high SNRs may bias the general quality of the spectra. Kurtosis is one of the higher-order statistical quantities, which are very sensitive to outliers. Kurtosis is a measure of the "peakedness" of the probability distribution of a real-valued random variable. Removing the outliers eventually reduces the kurtosis to a smaller value. The detailed findings will be presented in the results section in Chapter 8.

5.2.5 Bootstrap confidence intervals

The bootstrap (Johnson, 2001) was introduced by Bradley Efron (Efron and Tibshirani, 1993), mainly to calculate confidence intervals for parameters in situations where standard methods were not applicable (Efron and Gong, 1983, Zoubir and Iskander, 2004). With the bootstrap, a new set of experiments is not needed; instead, the original data is reused. Specifically, the original observations are randomly reassigned with a sample size identical to the original observations. These assignments and re-computations are done a large number of times and considered as repeated experiments.

Suppose we want to find the 95% bootstrap confidence interval of a variable mean, from an observation (size n), the values c_1 and c_2 are sought, so that

$$P(\bar{M} - m \le c_2) = 0.975 \text{ and } P(\bar{M} - m \le c_1) = 0.025 \qquad (5.11)$$

where m is the mean of the variable, and $\bar{M}$ denotes various possible mean values of an observation from the variable, which can be approximated by the original observation mean $\bar{m}$. Then

$$P(\bar{M} - c_2 \le \mu \le \bar{M} - c_1) = 0.95 \qquad (5.12)$$

So that

$$(\bar{M} - c_2, \bar{M} - c_1) \qquad (5.13)$$

is a 95% confidence interval for μ.

The procedural steps to find the approximate 95% confidence interval is (Johnson, 2001):

(i) Reassign n samples which are chosen from the original observations, compute the mean, denoted as $\bar{m}_i$, of the resample data.

(ii) Repeat step (i) a number of times, B, to come up with estimates $\bar{m}_1, \bar{m}_2, \ldots \bar{m}_B$.

(iii) Use the sample percentiles to estimate the desired population percentiles. With B = 1000, for example, sort the estimates above in ascending order as $\bar{m}_{(1)} \le \bar{m}_{(2)} \le \ldots \le \bar{m}_{(999)} \le \bar{m}_{(1000)}$ and use $\bar{m}_{(25)}$ to estimate the 2.5th percentile of $\bar{M}$ and use $\bar{m}_{(975)}$ to estimate the 97.5th percentile of $\bar{M}$. Solve c_1 and c_2 by using the equation (5.11). Finally, $(\bar{m} - c_2, \bar{m} - c_1)$ gives the desired confidence interval.

It is important, in this procedure, to produce estimates $\bar{m}_i$ with a sample size identical to the

original observations size of *n*. For a sample size less than *n*, the procedure will tend to give overestimated results; likewise, for a value more than *n*, the procedure would tend to give underestimated results.

The bootstrap re-sampling was used to produce $B = 1000$ sample sets by repeated samples with replacement from the original sample set. The 95% confidence interval was calculated by finding the 2.5th and 97.5th percentiles from the *B* re-sampled sets. The synthesised results presented in Chapter 7 were the average of 200 repeated experiments and with a 95% confidence interval calculated by the bootstrap re-sampling method.

5.3 Materials

The experiments are carried out on both synthesised and experimental MRS datasets. The synthesised MR spectra were taken to be linear combinations of the most commonly quantitated metabolite and MMLip components in *in vivo* MR spectra of childhood cerebellar tumours (Ketonen et al., 2004). The coefficients chosen are such that their means and standard deviations corresponded to those determined from LCModel fits to experimental *in vivo* MRS data from three types of childhood cerebellar tumours. The experimental dataset contains *in vivo* MR spectra of childhood brain tumours acquired before, during and after treatment.

5.3.1 Synthesised datasets

Various sets of synthesised MR spectra of three types of childhood brain (cerebellar) tumours (TE = 30 ms, 1.5T and 3T) were generated using a quantum mechanical-based simulation program (Reynolds et al., 2006), more details are given in Appendix A. Each synthesised MR spectrum was constructed as a linear combination of the most commonly quantitated

individual metabolite and lipid components.

The synthesised basis set 1 contains the following metabolite and MMLip components: Alanine (Ala) (doublet at 1.47 ppm and a quartet at 3.77 ppm), Choline (Cho) (singlet at 3.20 ppm and multiplet around 6.6 ppm), Creatine (Cr) (singlets at 3.04 ppm and 3.9 ppm), Lactate (Lac) (doublet at 1.33 ppm), N-Acetyl Aspartate (NAA) (singlet at 2.01 ppm and multiplets around 2.6 and 2.8 ppm), myo-Inositol (m-Ins) (multiplets appearing as a single peak at 3.56 ppm), Taurine (Tau) (two triplets at 3.25 and 3.42 ppm), and three lipids at 0.89, 1.30 and 2.05 ppm (Howe and Opstad, 2003, Govindaraju et al., 2000) as shown in Figure 5.1.

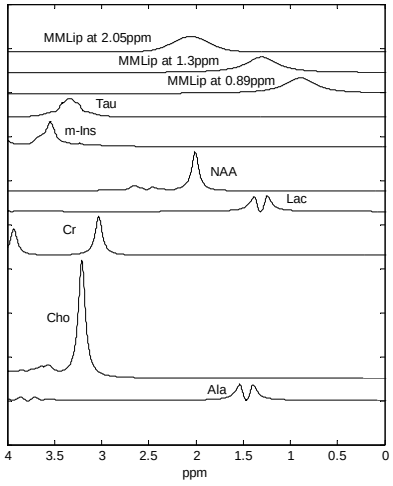


Figure 5.1: The 10 metabolites and lipids used in synthesised basis set 1.

In addition, we create a more complex synthesised basis set 2 which includes Glutamine (Gln) & Glutamate (Glu) (broad multiplet signals between 2 and 2.4 ppm and ~3.8 ppm), Glycine (Gly) (singlet at ~3.56 ppm) and use Glycerophosphorylcholine (GPC) & Phosphorylcholine (PCh) to replace Cho (Figure 5.2). The synthesised basis sets 1 and 2 described here will be

used in the experiments in both Chapters 7 and 8.

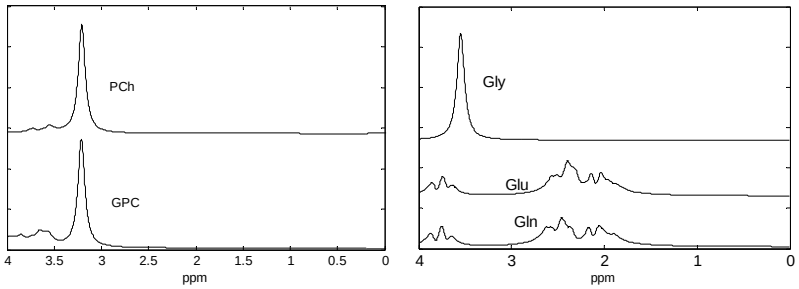


Figure 5.2: The GPC and PCh used to replace Cho, and Gln, Glu, Gly added in the synthesised basis set 2.

The mean and standard deviation for each metabolite or MMLip component concentration were set to those observed in *in vivo* MR spectra of the three classes of childhood cerebellar tumours (Davies et al., 2008), namely astrocytoma, ependymoma and medulloblastoma, as shown in Table 5.1.

Table 5.1:Metabolite and MMLip components concentrations (Davies et al., 2008) used for the simulation of astrocytoma, ependymoma, and medulloblastoma spectra.

Metabolite	Concentration (mmol/kg)					
	Astrocytoma		Ependymoma		Medulloblastoma	
	Mean	SD	Mean	SD	Mean	SD
Ala	1.2	0.4	0.5	0.5	1.4	0.9
Cho	1.4	0.4	3.1	1.5	4.1	1.7
Cr	1.2	1.2	6.3	2.3	3.7	2.1
Lac	2.6	0.9	2.8	1.1	3.5	3.6
NAA	1.5	1.5	0.5	0.3	0.5	0.4
m-Ins	2.4	1.6	14.8	5.4	6.5	3.7
Tau	0.2	0.4	1.0	1.1	3.4	2.0
MMLip0.9	9.9	6.5	9.0	3.1	14.4	5.9
MMLip1.3	21.0	14.0	24.5	18.8	42.6	33.0
MMLip2.05	9.9	6.5	9.0	3.1	14.4	5.9
PCh	0.1	0.3	0.3	0.5	2.0	1.4
GPC	1.3	0.6	2.9	1.0	2.1	1.5
Gln	3.5	1.7	5.7	2.4	2.1	1.5
Glu	2.4	1.5	2.1	0.9	3.6	1.6
Gly	0	0	0	0	3.5	1.6

The sum of correlation coefficients between the obtained ICs and the original individual metabolite and MMLip components was used as a measure of the quality of the resulting ICs for the synthesised datasets. An IC correlated more closely to its original component will have a correlation coefficient closer to 1.

5.3.2 Experimental dataset

We have used an experimental dataset, which contains *in vivo* MR spectra from children aged under 16 years with brain tumours obtained prior to treatment and in follow-up scans between 1 January 2003 and 6 May 2008 at Birmingham Children's Hospital. All studies were performed using a 1.5T scanner (Siemens Symphony Magnetom, NUM4 and GE Signa Excite & HDx). MR spectra were acquired using a point resolved single voxel spectroscopy (PRESS)

(as explained in Chapter 2) sequence (TE = 30 ms, TR[5] = 1500 ms). The voxel volume was 3.4 ml (1.5 cm sided cube) or 8 ml (2 cm sided cube) with 256 or 128 signal averages acquired respectively. The bandwidth of the GE spectra was 1.22 Hz per point, while it was 0.98 Hz per point for the Siemens spectra, and the spectral range analysed was 0.2 ppm to 4.0 ppm. The FIDs of GE spectra were re-sampled accordingly and an additional shift (0.0017 ppm) along the ppm-axis was required in the spectra domain to fit these with the Siemens spectra. Ethical approval was obtained for the study from the Local Research Ethics Committee and parental consent was obtained.

The 115 experimental spectra were obtained after applying two quality control criteria: (a) Full width at half maximum (FWHM) of the water reference peak less than 6 Hz and (b) SNR of the spectra greater than 10. The correlation coefficients between the obtained ICs and the simulated metabolites from (Reynolds et al., 2006) or basis MMLip components from LCModel are used as a measure of the quality of the resulting ICs.

To make the experimental dataset similar to the synthesised dataset, the baseline estimated from the LCModel software package (Provencher, 2009) was removed from each MR spectrum. Actually, the LCModel baseline mostly accounts for a missing or incorrect macromolecule model with good data, and becomes flatter (inadequate information to determine the baseline detail) with poorer data. ICA applied to an experimental dataset without subtracting the baseline assumed by LCModel may reveal components not included in the model. This is not presented in this thesis and will be investigated in a future study.

[5] Repetition time (TR): The amount of time that exists between successive pulse sequences applied to the same slice or voxel.

5.4 Summary

The different implementation aspects of BSS- and FE-based ICA methods are explained with the consideration of noise corruption. A feature dimension reduction guideline for using ICA on the MRS data is also proposed based on the SNR of the dataset. The bootstrap re-sampling is used when the dataset is not large enough to be considered representative. This method attempts to determine the probability distribution from the data itself.

The synthesised MRS datasets constructed in this chapter will be used in Chapters 7 and 8 for analysing the applicability of BSS- and FE-based ICA. They will also be used to test the proposed hybrid ICA method in Chapter 8. The findings will be verified with the experimental dataset.

CHAPTER 6

NOISE REDUCTION ON *IN VIVO* ^{1}H MR SPECTRA USING WAVELET SHRINKAGE DE-NOISING (WSD) AND LORENTZIAN LINESHAPE PROPERTY

6.1 Introduction

The SNR of an MR spectrum is proportional to the square root of the number of acquisitions (Kreis, 1997, Schorn and Taylor, 2004). Clearly there is a trade-off between the SNR required and the total scanning time. In clinical circumstances, long and time-consuming measurements usually are not preferred. As a consequence, *in vivo* ^{1}H MR spectra are often characterized by a low SNR. The de-noising is an important pre-processing step in the analysis of the MRS signals. Improvements have been found when the de-noised MR spectra were used for metabolite decomposition by peak picking (Dancea and Günther, 2005),

independent component analysis (ICA) (Hao et al., 2009b) or principal component analysis (PCA) (Trbovic et al., 2005). It has been shown that the wavelet shrinkage de-noising (WSD) is an efficient de-noising algorithm for ^{1}H MR spectra. Hector et al. (Cancino-De-Greiff et al., 2002) examined the WSD on free induction decay (FID) signals and Hoch et al. (Hoch and Stern, 1996) briefly showed noise reduction by WSD on MRS data.

In previous research, only the real part of the MR spectrum is analysed, the imaginary part is usually ignored. As part of this research, we will incorporate the imaginary part of MR spectrum to improve the robustness of the WSD-based MR spectrum de-noising performance. First, a novel analysis, which reveals the relationship between the real and imaginary parts of the MR spectrum, is presented. The analysis demonstrates that the real and imaginary parts of an MR spectrum can be derived from each other. Based on that conclusion, a novel WSD-based MRS enhancement algorithm is proposed. The de-noising processing, including WSD, always leads to some inevitable signal misrepresentation (Lu and Wang, 2003). To alleviate such a distortion, the WSD-enhanced real part of MR spectrum is averaged with the derived real part from the WSD-enhanced imaginary part according to the relationship obtained.

In this chapter, we examine several commonly used wavelet bases in the proposed scheme and compare the results with the conventional algorithms. The signals to be de-noised are the synthesised FID signals and their respective MR spectra. The de-noising performance is measured by SNR on the real part of the spectrum.

6.2 Theory

6.2.1 Relationship between the real and imaginary parts of the MR spectrum

We have derived a relationship between the real and imaginary parts of the MR spectrum as:

$$A = -B_I + jB_R, \quad \begin{cases} A_R = -B_I \\ A_I = B_R \end{cases} \tag{6.1}$$

$$B = A_I - jA_R, \quad \begin{cases} B_R = A_I \\ B_I = -A_R \end{cases} \tag{6.2}$$

where A and B are the inverse Fourier transform (IFT) of the real and imaginary parts of the spectrum, respectively.

The detailed derivation of the relationship between the real and imaginary parts of the MR spectrum is presented in Appendix B.

A linear combination of several Lorentzian lineshapes should still obey the above derived equations. Hence the real part of the MR spectrum can be calculated from its imaginary part, and vice versa. So for clean signals, the derived real and imaginary parts of the spectrum are exactly the same as their original ones as shown in Figure 6.1. It is clearer in Figure 6.2, where the real and derived real parts of an MR spectrum overlap with each other, similarly for the imaginary parts. When noise is present, the spectra could not satisfy the above equations, since noise does not have the Lorentzian lineshape. As illustrated in Figure 6.3 and 6.4, the derived noisy real and imaginary parts of the spectrum are very close to their original ones with slight variations due to the noise effect. Averaging the corresponding calculated and original parts of the spectrum should reduce the noise effect on the noisy signal. For the shown example, the real part of the spectrum has an SNR of 7.88, and the averaged spectrum has an SNR of 11.23. Similar effect can be observed for an experimental MR spectrum (Figure 6.6).

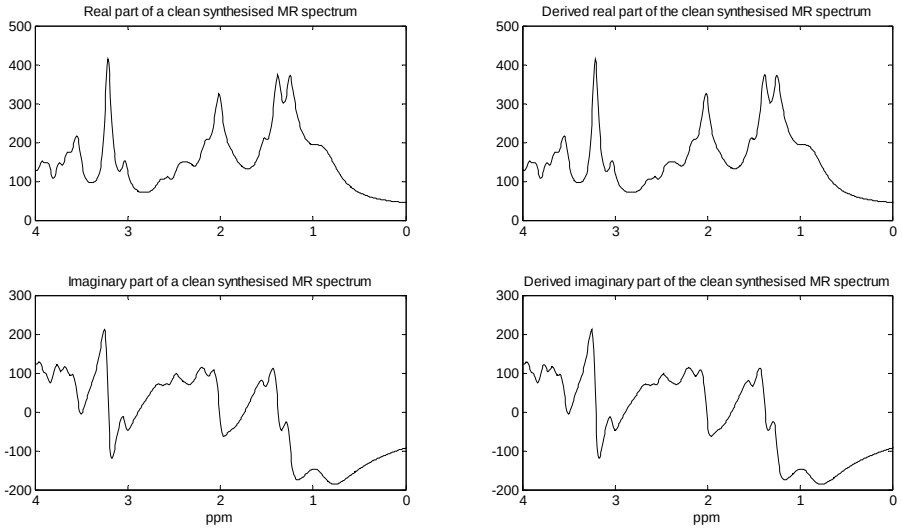


Figure 6.1: The real and imaginary parts of an example clean synthesised MR spectrum and the derived real and imaginary parts.

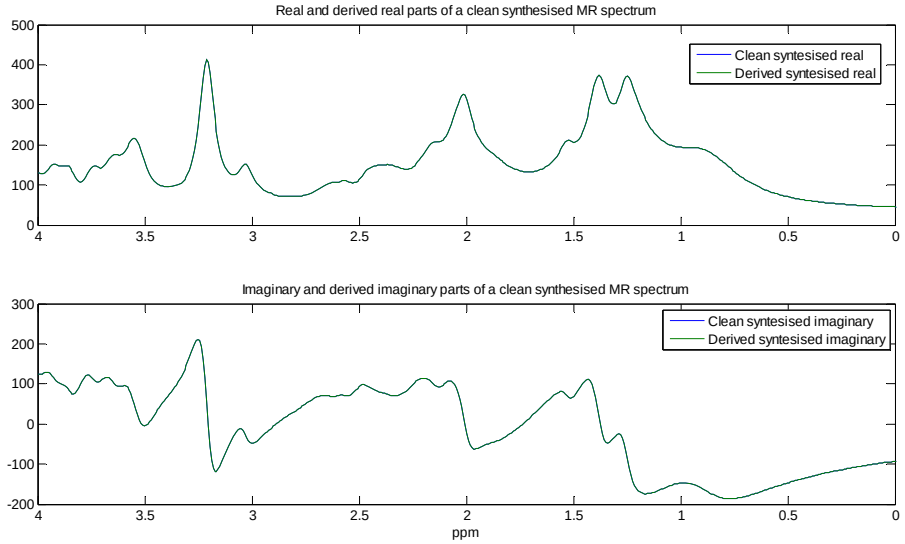


Figure 6.2: The real and the derived real, the imaginary and the derived imaginary parts of the clean synthesised MR spectrum plotted together.

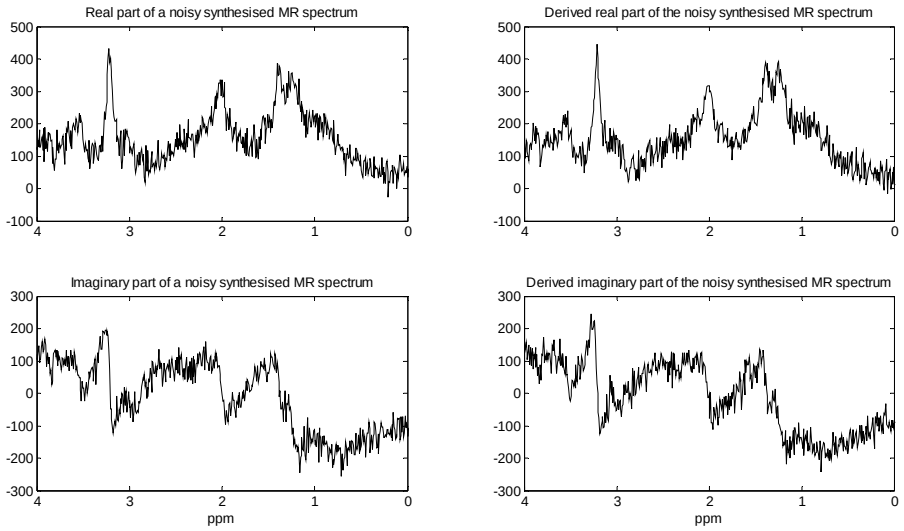


Figure 6.3: The real and imaginary parts of an example noisy synthesised MR spectrum and the derived real and imaginary parts.

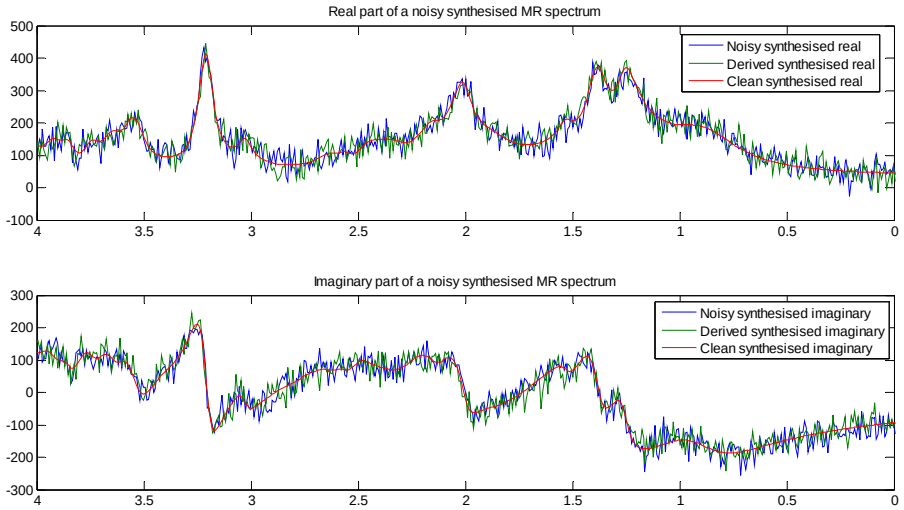


Figure 6.4: The clean, noisy and derived real, the clean, noisy and derived imaginary parts of the noisy synthesised MR spectrum plotted together.

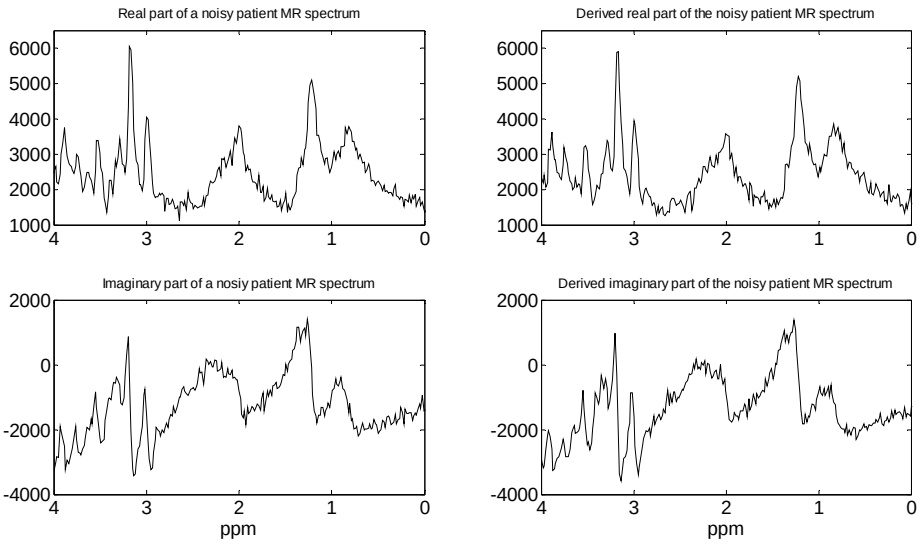


Figure 6.5: The real and imaginary parts of an experimental MR spectrum and the derived real and imaginary parts.

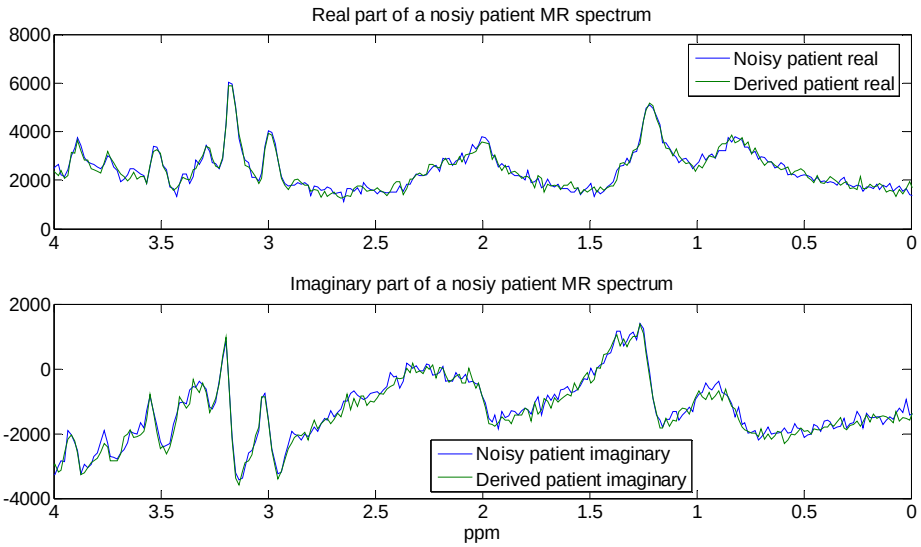


Figure 6.6: The noisy and derived real, the noisy and derived imaginary parts of the experimental MR spectrum plotted together.

6.2.2 Wavelet Transform

The family of wavelet basis functions can be obtained from the mother wavelet function $\psi(t) \in L^2(\mathbf{R})$ by dilation and translation (Daubechies, 1992),

$$\psi_{(b,a)}(t) = a^{-1/2}\psi\left[a^{-1}(t-b)\right] \tag{6.3}$$

where the dilation and translation parameters a, b ($b \in \mathbf{R}$ and $a > 0$) may be continuous or discrete.

With continuous parameters, the continuous wavelet transform (CWT) of a function $f(t) \in L^2(\mathbf{R})$ is defined as:

$$WT_f(a,b) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt = \left\langle f(t), \psi_{a,b}(t)\right\rangle \tag{6.4}$$

Through CWT analysis, a set of wavelet coefficients $\left\{WT_f(a,b)\right\}$ are obtained. These coefficients indicate how close the signal is to a particular wavelet basis function. The CWT provides a redundant representation of the signal in the sense that the entire support of $\left\{WT_f(a,b)\right\}$ need not be used to recover the $f(t)$ signal. Therefore, the dilation and translation parameters are evaluated on a discrete grid of time-scale plane, leading to a discrete set of continuous basis functions.

The discrete wavelet transform (DWT) is obtained by setting the dilation and translation parameters as,

$$a = a_0^m \quad \text{and} \quad b = n a_0^m b_0 \tag{6.5}$$

where a_0, b_0 are fixed and $m, n \in \mathbf{Z}$. The most widely used set is $a_0 = 2$, $b_0 = 1$, so the family of discrete wavelets basis becomes,

$$\psi_{m,n}(t) = 2^{-m/2} \psi\left(2^{-m} t - n\right) \tag{6.6}$$

Multiresolution Analysis (MRA) (Mallat, 1989) provides a general framework to construct the orthonormal wavelet basis [6] and efficient implementation through fast filterbank algorithms similar to fast Fourier transform (FFT) (Bracewell, 2000, Brigham, 1988) algorithms. It is based on the idea that a function or a signal can be approximated at different dilatation levels. It consists of a sequence of embedded subspaces $...V_1 \subset V_0 \subset V_{-1} \subset ...$ of $L^2(\mathbf{R})$. A scaling function (father wavelet),

$$\phi_{m,n} = 2^{-m/2} \phi(2^{-m} t - n);\ m,\ n \in \mathbf{Z} \tag{6.7}$$

is introduced here and forms a set of orthonormal basis for the reference space V_n. The wavelet function (mother wavelet) $\{\psi_{m,n};\ m,\ n \in \mathbf{Z}\}$ is a set of orthonormal basis for the orthonormal complement W_n of V_n to V_{n-1} (i.e. $V_{n-1} = V_n \oplus W_n, n \in \mathbf{Z}$).

Since $\phi \in V_0 \subset V_{-1}$, $\psi \in W_0 \subset V_{-1}$, and the $\phi_{-1,n}$ are orthonormal basis in V_{-1}, we have,

[6] An orthonormal wavelet basis is an orthonormal set of wavelet basis as described in equation (6.41). A set of vectors form an orthonormal set if all vectors in the set are mutually orthogonal and all of unit length.

$$\phi(t) = \sum_n h_n \, \phi_{-1,n}(t) = \sqrt{2} \sum_n h_n \, \phi(2t - n)$$
$$\psi(t) = \sum_n g_n \, \phi_{-1,n}(t) = \sqrt{2} \sum_n g_n \, \phi(2t - n) \tag{6.8}$$

where $h_n = \langle \phi, \, \phi_{-1,n} \rangle$, $g_n = \langle \psi, \, \phi_{-1,n} \rangle$ can be viewed as the coefficients of lowpass and highpass filters, respectively.

6.2.3 Wavelet shrinkage de-noising

The signal and noise for one-dimensional noisy model can be presented as:

$$\mathbf{x}^{\text{obs}} = \mathbf{x} + \mathbf{n} \tag{6.9}$$

where $\mathbf{x}$ is a clean signal with length L, $\mathbf{x}^{\text{obs}}$ is a noisy signal corrupted by additive white Gaussian noise $\mathbf{n}$. The de-noising objective is to suppress the noise part of the signal $\mathbf{x}^{\text{obs}}$ so as to recover $\mathbf{x}$.

The technique works in the following way. When decomposing a signal using wavelets, they act as filters that produce averaging and detail parts of the signal. If the detail parts are small, they might be omitted without substantially affecting the main features of the signal. As noise is commonly assumed to have a Gaussian distribution, most of the noise components tend to be represented by wavelet coefficients at smaller scales (higher frequencies). Removing these coefficients would result in reducing the noise without substantially affecting the original clean signal.

The general WSD procedure involves three steps and is described below:

1. Decomposing the signal by DWT with the selected mother wavelet and selected

level of wavelet decomposition.

2. Determining the threshold for each level by certain thresholding rule and apply thresholding to the detail coefficients.
3. Reconstructing the signal by inverse DWT using the thresholded transform coefficients.

The threshold determines how much noise we want to suppress, and the larger the variance of the noise, the larger the threshold should be. The effectiveness of four thresholding selection rules (Misiti et al.) was investigated in this study.

---

1) 'rigrsure' uses the principle of Stein's Unbiased Risk Estimate (SURE),
2) 'sqtwolog' uses a universal threshold, λ is the threshold value selected.

$$\lambda = \sqrt{2\log(N)} \tag{6.10}$$

3) 'heursure' is a heuristic variant of the previous two rules.
4) 'minimaxi' uses using minimax principle for selection.

These thresholding selection rules of wavelet coefficients can be applied mainly by either hard or soft thresholding (Donoho and Johnstone, 1994, Donoho, 1995). In hard thresholding, all coefficients below a threshold are zeroed, while in the soft thresholding, all the non-zero coefficients are shrunk towards zero by subtracting the threshold.

The hard thresholding function is given as:

$$z = \text{hard}(w) = \begin{cases} w, & |w| > \lambda \\ 0, & |w| \le \lambda \end{cases} \tag{6.11}$$

where w and z are the input and output wavelet coefficients, respectively.

Similarly, the soft thresholding function is given as:

$$z = \text{soft}(w) = \begin{cases} \text{sgn}(w)\max(|w| - \lambda, 0), & |w| > \lambda \\ 0, & |w| \le \lambda \end{cases} \tag{6.12}$$

Generally soft thresholding gives fewer artefacts and preserves the smoothness of the signal. The inverse wavelet transform of the thresholded coefficients is the de-noised signal.

The three multiplicative threshold rescaling methods are also examined,

(1) 'one' for no rescaling,

(2) 'sln' for rescaling using a single estimation of level noise based on first-level coefficients,

(3) 'mln' for rescaling using level-dependent estimation of level noise.

The WSD is not limited to time domain signals. Mathematically, this method tries to remove whatever noise presented in a noisy signal regardless of its physical content. As the noise is assumed Gaussian in the time domain, its Fourier transform (FT) in the frequency domain should also be Gaussian, and vice versa. The WSD is designed to remove Gaussian noise, so it could be used on either time or frequency domain signals.

6.3 Methods

6.3.1 Synthesised MR spectra

A set of synthesised *in vivo* ^{1}H MR spectra (TE = 30 ms, 1.5T) of childhood brain tumours is

generated in the experiments. The synthesised dataset contains 50 noisy MR spectra from each of the three classes of tumours, i.e. astrocytoma, ependymoma, and medulloblastoma. The MR spectra are generated using the 14 metabolite and MMLip components of basis set 2 constructed in Chapter 5 with various amount of white Gaussian noise. The mean and standard deviation of each component concentration are set to according to Table 5.1.

6.3.2 Wavelet basis

Choosing the proper wavelet basis is very important for de-noising. Different wavelets give different de-noised signals. Generally, the employed wavelet basis should possess similar properties with the original signal. The wavelet de-noising routines used in this work were based on the Matlab wavelet toolbox (Misiti et al.). The wavelet families listed in Table 6.1 are examined in this study. The de-noising effects of the listed 54 wavelet functions are compared in the next section.

Table 6.1: Wavelet families examined in this study, the superscript in front of each wavelet name is its y axis index in Figures 6.7, 6.9 and 6.12.

Wavelets families	Wavelet name
Haar	[1]Haar
Daubechies	[2]db1 [3]db2 [4]db3 [5]db4 [6]db5 [7]db6 [8]db7 [9]db8 [10]db9 [11]db10
Symlets	[12]sym2 [13]sym3 [14]sym4 [15]sym5 [16]sym6 [17]sym7 [18]sym8
Coiflets	[19]coif1 [20]coif2 [21]coif3 [22]coif4 [23]coif5
Biorthogonal	[24]bior1.1 [25]bior1.3 [26]bior1.5 [27]bior2.2 [28]bior2.4 [29]bior2.6 [30]bior2.8 [31]bior3.1 [32]bior3.3 [33]bior3.5 [34]bior3.7 [35]bior3.9 [36]bior4.4 [37]bior5.5 [38]bior6.8
Reverse Biorthogonal	[39]rbio1.1 [40]rbio1.3 [41]rbio1.5 [42]rbio2.2 [43]rbio2.4 [44]rbio2.6 [45]rbio2.8 [46]rbio3.1 [47]rbio3.3 [48]rbio3.5 [49]rbio3.7 [50]rbio3.9 [51]rbio4.4 [52]rbio5.5 [53]rbio6.8
Discrete Meyer	[54]Dmey

6.3.3 Performance measure

The performance of de-noising algorithms is quantitatively evaluated by the MSE between original and de-noised signal,

$$\mathrm{MSE} = \frac{1}{L}\sum_{i=1}^{L}\left|x_i - \hat{x}_i\right|^2 \tag{6.13}$$

and also the SNR_{LC} (as defined by equation(5.8)) of the de-noised signal,

$$\begin{aligned} SNR_{\mathrm{LC}} &= \frac{\max(\mathbf{x})}{2\times \mathrm{RMS}\left(\mathbf{x}-\hat{\mathbf{x}}\right)} \\ &= \frac{\max(\mathbf{x})}{2\times\sqrt{\mathrm{MSE}}} \end{aligned} \tag{6.14}$$

where $\hat{\mathbf{x}}$ is the estimate of the clean signal $\mathbf{x}$, L is the length of these signals, and $\hat{x}_i$ and x_i are the ith element of $\hat{\mathbf{x}}$ and $\mathbf{x}$, respectively. Equation (6.14) shows that for each signal, the square root of MSE is inversely proportional to its SNR; therefore, we will only use the SNR value for the performance measure in this study. The SNR is calculated between the de-noised real part of the MR spectrum and its original clean spectrum and follows the definition in LCModel as defined by equation (6.14).

The wavelet decomposition is performed with levels ranging from 1 to 8. The combinations from *four thresholding rules*, *two threshold applying methods*, *three rescaling options* and *8 levels of decomposition* for every wavelet basis, listed in Table 6.1, are considered as the WSD rules. They are evaluated on each of the 150 synthesised MR spectra.

Four experiments are performed in the next section. In the first two experiments, the WSD rules are applied to the FID signals, as well as, the real parts of the MR spectra. The mean de-

noised SNRs of the whole synthesised dataset are calculated for every WSD rule. In the last two experiments, the WSD rules are applied on both the real and imaginary parts of the MR spectra, and the average signal between the WSD enhanced real and derived real from the WSD enhanced imaginary parts (equation (B.36)) gives the final de-noised result.

6.4 Result

6.4.1 Experiment 1: WSD on FID signal

De-noising by the 10368 WSD rules (54 wavelet basis × 4 thresholding rules × 2 threshold applying methods × 3 rescaling options × 8 levels of decomposition) is performed on the FID signals of the 150 synthesised MR spectra. The mean de-noised SNR values of the synthesised spectra dataset by all the WSD rules are illustrated in Figure 6.7. The lighter the colour block in Figure 6.7 is, the higher the SNR values are. The best WSD rule, which produces the highest mean SNR value of 22.70 (point (184, 54) in Figure 6.7), is the wavelet basis ‘dmey’, the ‘minimaxi’, ‘hard’ and ‘sln’ thresholding rules at level 8. The 150 noisy signals’ SNR values (with mean SNR of 10.75) and the de-noised SNR values by the above mentioned WSD rule are plotted in Figure 6.8. The average improvement of this de-noising method is 111%.

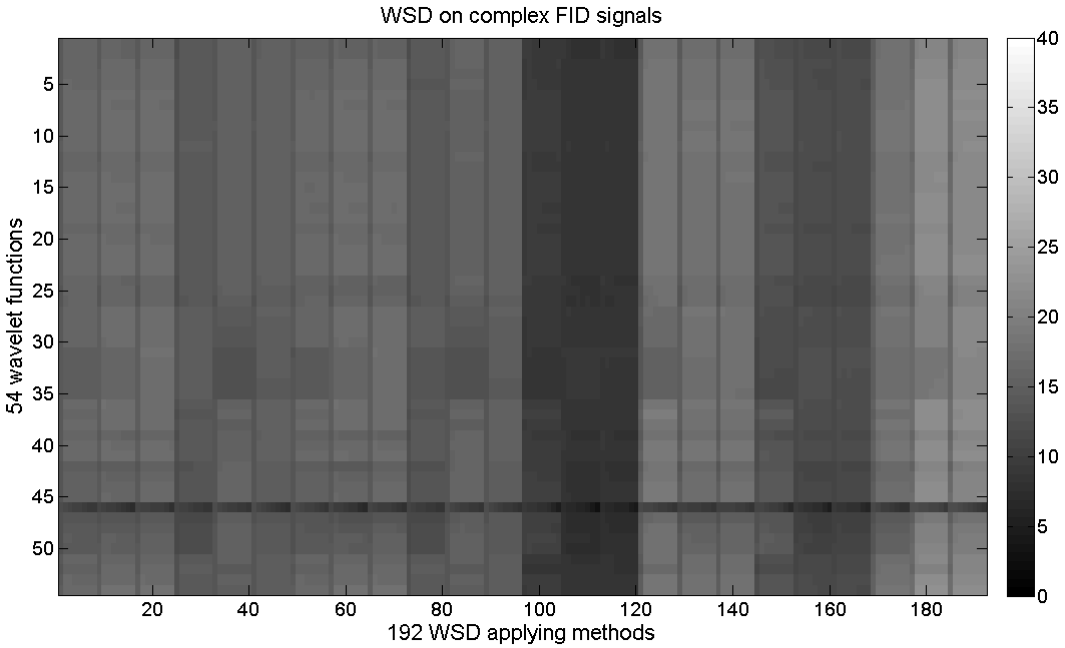


Figure 6.7: The average de-noised SNR values by the 10368 WSD rules on the FID signals of the spectra. The maximum SNR value of 22.70 occurs at point (184, 54).The 54 wavelet families in the y axis are ordered as they appeared in Table 6.1. The x axis is the combination of the WSD rule parameters from the sets: [rigrsure, heursure, sqtwolog, minimaxi], [soft, hard], [one, sln, mln], and levels [1: 8]. For example, the first 8 scales on the x axis are the WSD rules: rigrsure, soft, one, levels 1 to 8, the next 8 scales on the x axis are the WSD rules : rigrsure, soft, sln, levels 1 to 8 and so on.

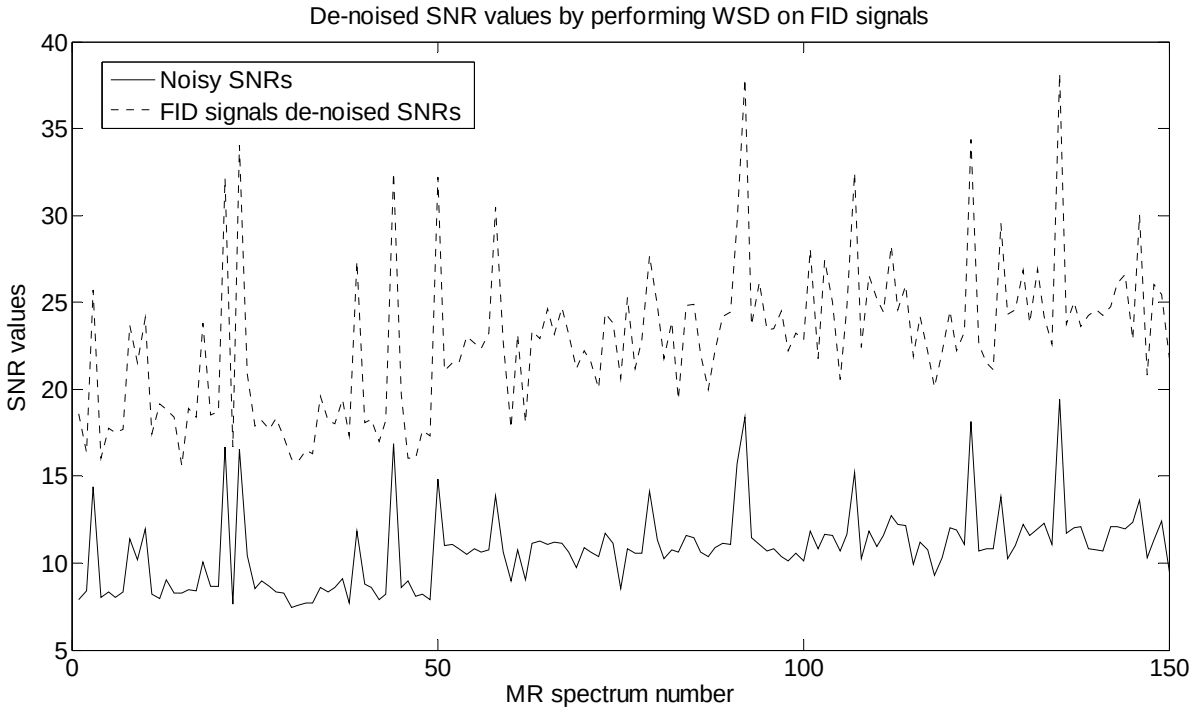


Figure 6.8: Comparing the SNRs between the noisy and the de-noised spectra by applying WSD rules on the FID signals.

6.4.2 Experiment 2: WSD on real part of the MR spectrum

Since only the real part of the MR spectrum is normally used for analysis, the de-noising effect on the noisy real part of the spectrum is examined in this experiment. Figure 6.9 shows the average de-noised SNR values by each of the WSD rules on the whole synthesised dataset. The best de-noised SNR value is 26.44 (point (60, 36) in Figure 6.9) and is produced by the wavelet basis 'bior4.4', under the thresholding rules 'heursure', 'soft' and 'sln' at level 4. The de-noised SNR values by the best WSD rule are compared with the results from experiment 1 in Figure 6.10. It is obvious that applying WSD on the real part of the spectrum generally

provides better SNR result than applying WSD on the FID signal of the same spectrum. The average improvement of the de-noised SNR values is 146% over the noisy ones, and up to 1.5 times better than the results in experiment 1.

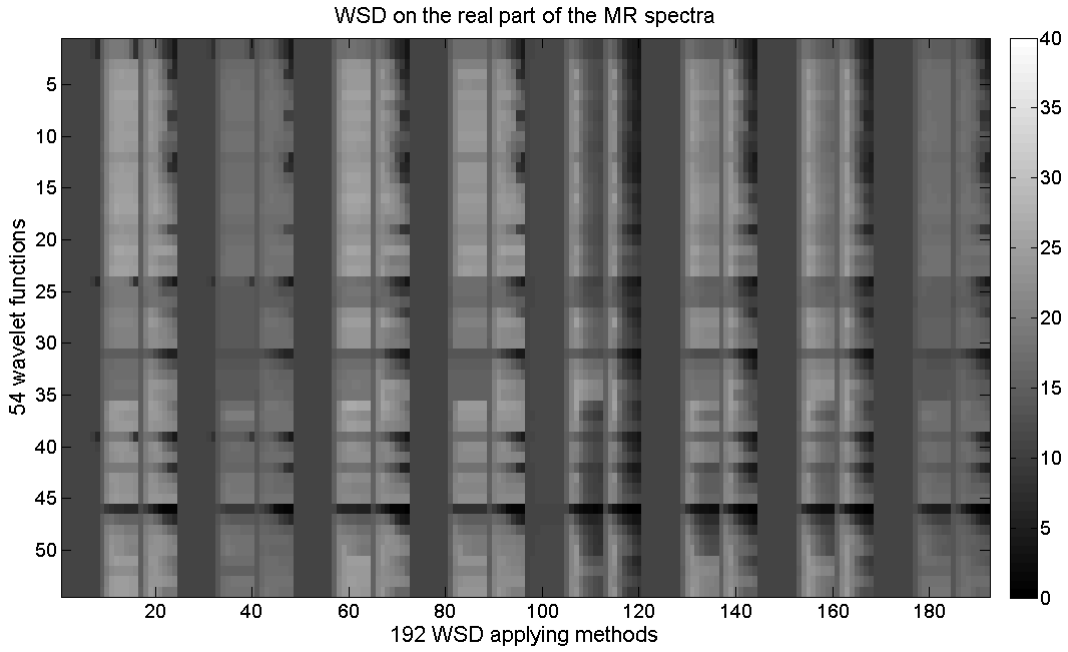


Figure 6.9: The average de-noised SNR values by the 10368 WSD rules on real part of the spectra. The maximum SNR value of 26.44 occurs at point (60, 36).The 54 wavelet families in the y axis are ordered as they appeared in Table 6.1. The x axis is the combination of the WSD rule parameters from the sets: [rigrsure, heursure, sqtwolog, minimaxi], [soft, hard], [one, sln, mln], and levels [1: 8]. For example, the first 8 scales on the x axis are the WSD rules: rigrsure, soft, one, levels 1 to 8, the next 8 scales on the x axis are the WSD rules: rigrsure, soft, sln, levels 1 to 8 and so on.

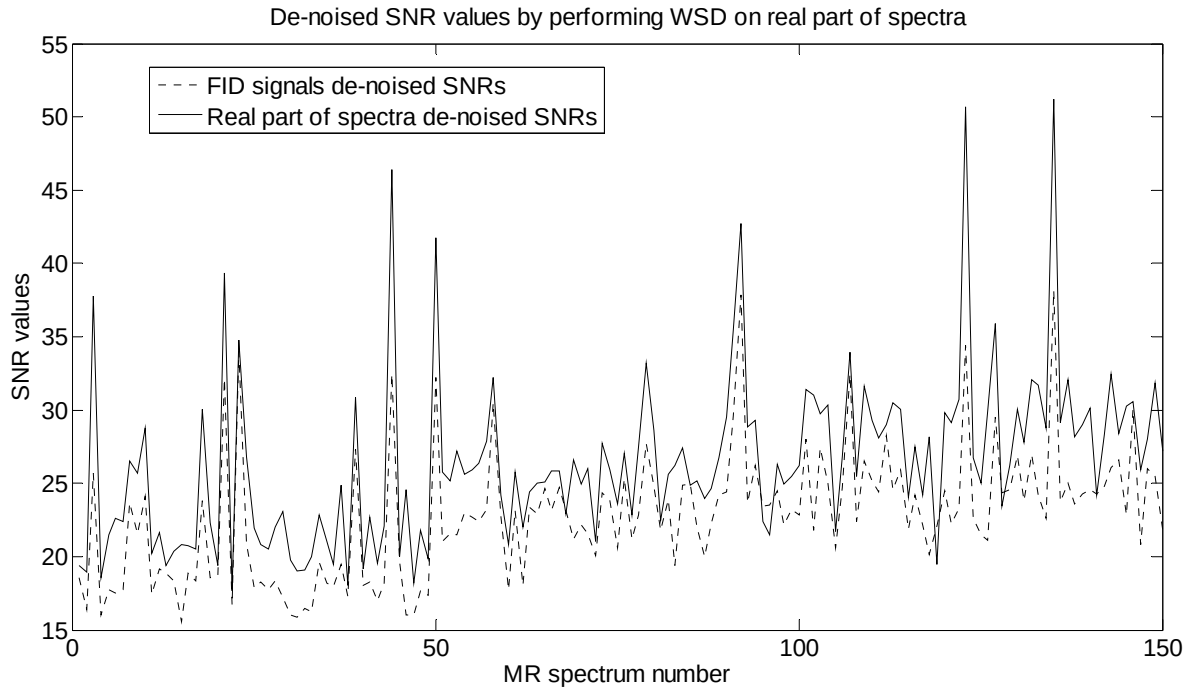


Figure 6.10: Comparing the de-noised SNRs by applying WSD on the FID signals and on the real part of the MR spectra.

6.4.3 Experiment 3: WSD on real and imaginary parts of the MR spectrum with different WSD rules

Based on the conclusions reached in section 6.2.1, a WSD-based MR spectrum enhancement algorithm is proposed. In this experiment, the WSD enhanced imaginary part of the spectrum will be used to derive another version of the real part of the spectrum. The average of the de-noised real part of the spectrum and the derived one from its de-noised imaginary part is the final de-noised signal. Two approaches of finding the best WSD rule are explored in this

experiment and the next one.

Since the best WSD rule for the real part of the spectra has already obtained in experiment 2, only the best WSD rule for the imaginary part of the spectra is searched. It is found that using wavelet basis ‘bior6.8’, under the thresholding rules ‘heursure’, ‘soft’ and ‘sln’ at level 5 works best on the imaginary part. The final de-noised SNR values obtained by averaging the two real parts of the spectrum are plotted in Figure 6.11 together with the SNR values from experiment 2 as a reference. No obvious improvement can be found as the two SNR lines are fluctuating with each other. The mean SNR value for this approach is 27.2, which is very close to 26.44 in experiment 2.

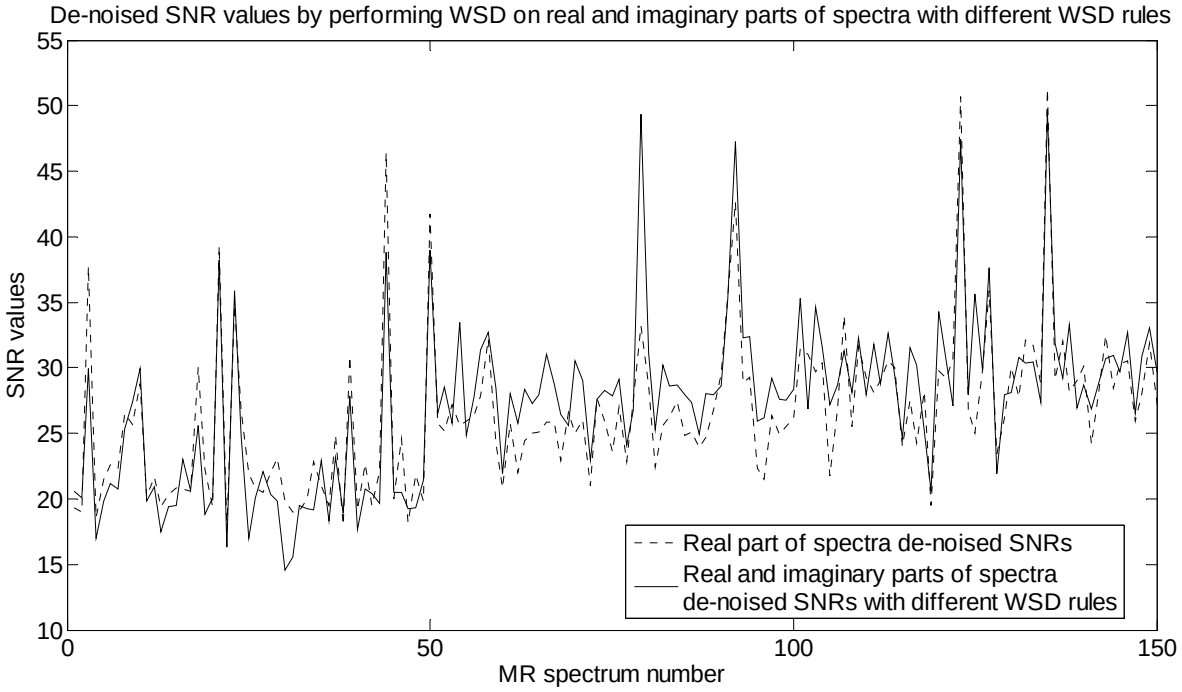


Figure 6.11: Comparing the de-noised SNRs between real and real & imaginary parts of the MR spectra with different WSD rules.

6.4.4 Experiment 4: WSD on real and imaginary parts of the MR spectrum with the same WSD rule

In this experiment, slightly different from the previous experiment, a single WSD rule is sought for both the real and imaginary parts of the spectra. The mean de-noised SNR values of the synthesised dataset obtained by the proposed algorithm with all the WSD rules are shown in Figure 6.12. The one with wavelet basis ‘bior4.4’, under the thresholding rules ‘heursure’, ‘hard’ and ‘mln’ at level 5, produces the highest mean SNR value at 38.56 (point (93, 36) in Figure 6.12). The final de-noised SNR values for all the 150 MR spectra are illustrated in Figure 6.13 and compared with the results from experiment 2. It can be seen that there is a clear gap between the enhanced SNR values from this experiment and experiment 2. The mean SNR improvement is 258% over the original noisy signal, and it is up to 2.04 times better than the results from experiment 2.

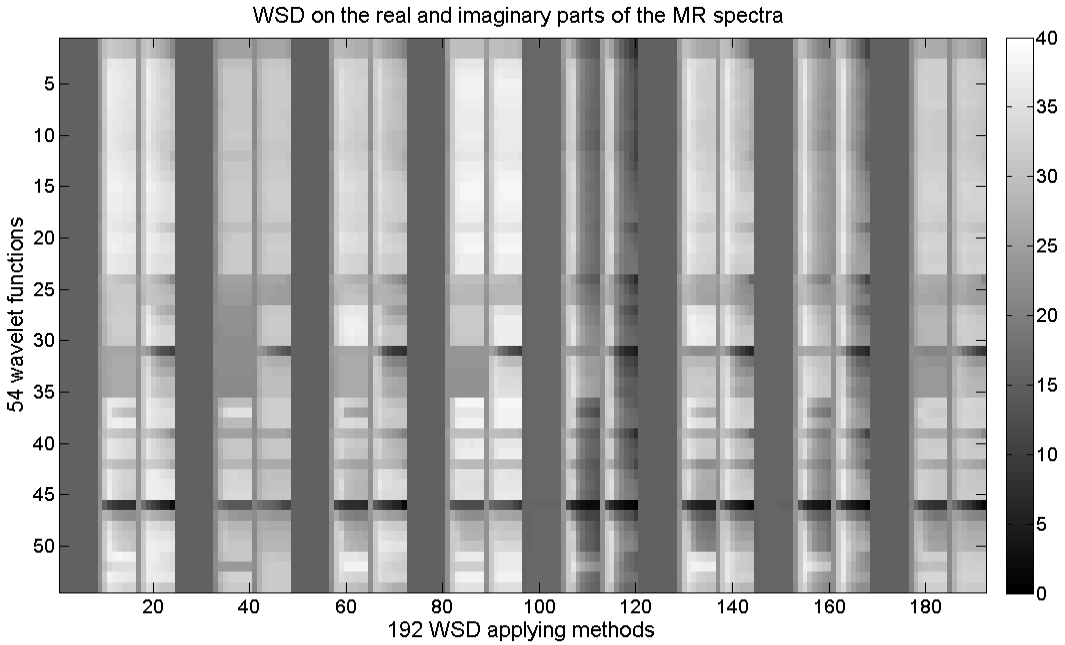


Figure 6.12: The average de-noised SNR values by the 10368 WSD rules on real and imaginary parts of the spectra. The maximum SNR value of 38.56 is at point (93, 36). The 54 wavelet families in the y axis are ordered as they appeared in Table 6.1. The x axis is the combination of the WSD rule parameters from the sets: [rigrsure, heursure, sqtwolog, minimaxi], [soft, hard], [one, sln, mln], and levels [1: 8]. For example, the first 8 scales on the x axis are the WSD rules : rigrsure, soft, one, levels 1 to 8, the next 8 scales on the x axis are the WSD rules : rigrsure, soft, sln, levels 1 to 8 and so on.

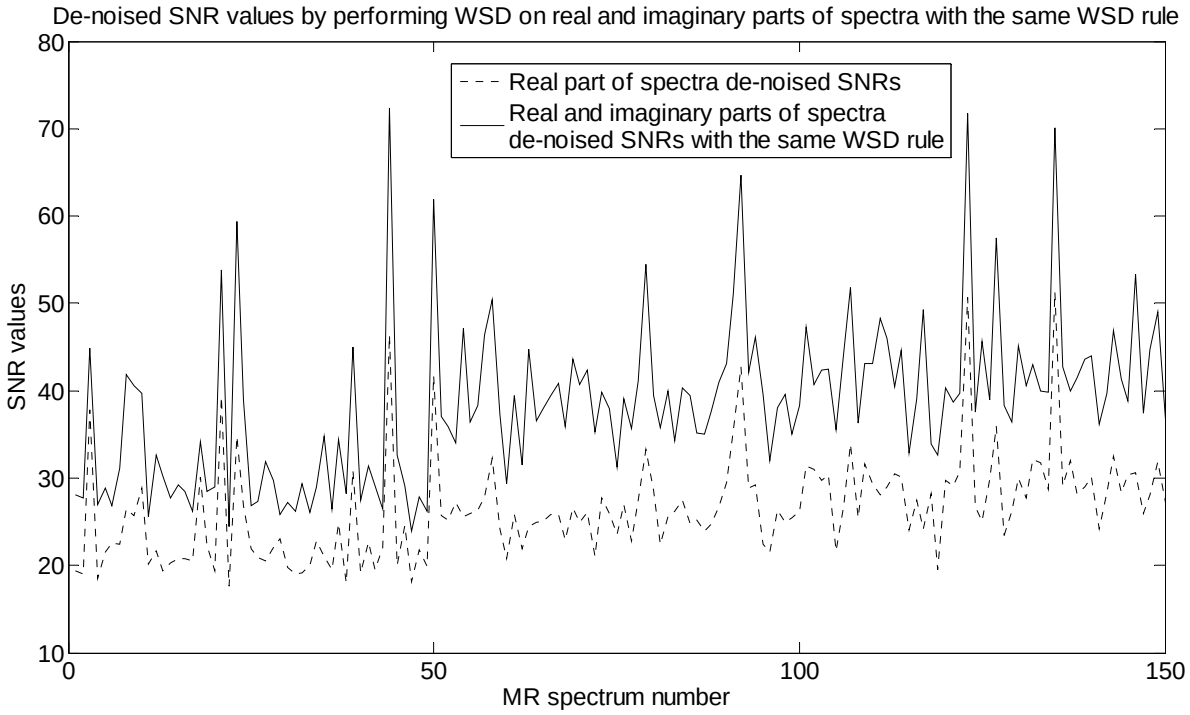


Figure 6.13: Comparing the de-noised SNRs between real and real & imaginary parts of the MR spectra.

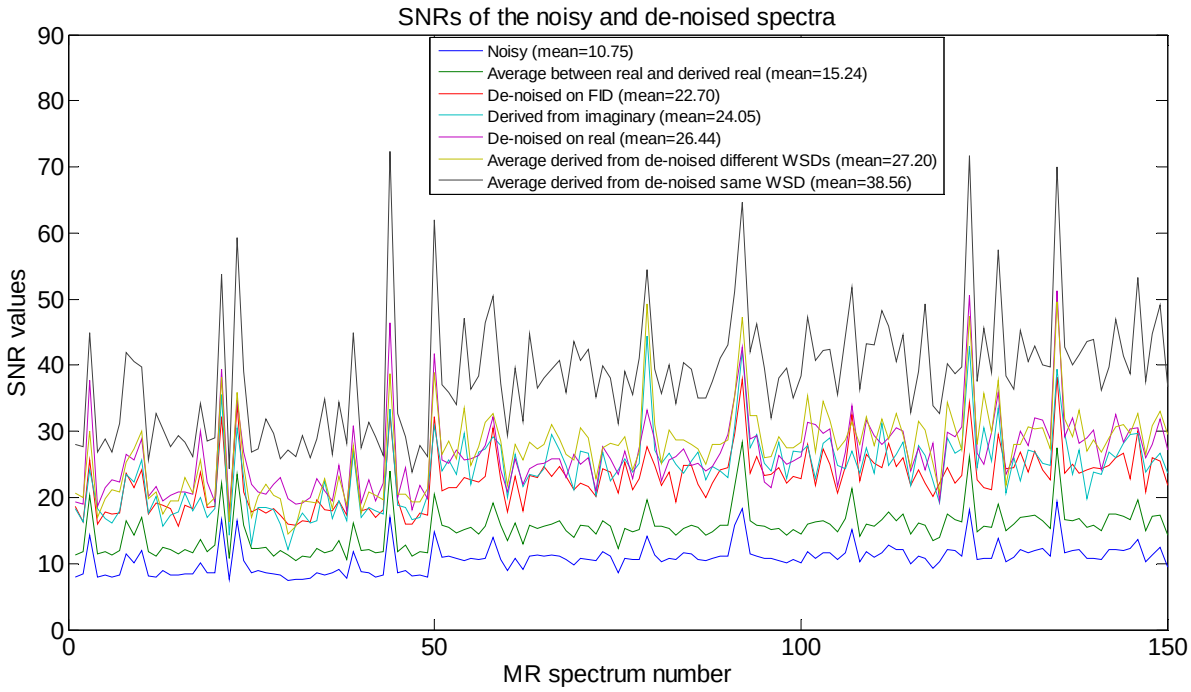


Figure 6.14: A comparison of the SNRs for noisy and de-noised spectra of the whole synthesised dataset.

The de-noised SNR values of all the 150 synthesised spectra by various methods are shown in Figure 6.14. One example of the real part of a synthesised MR spectrum and its de-noised signals from the four experiments are shown in Figure 6.15. The spectrum of the FID de-noised signal is still very fuzzy compared with the clean spectrum. Spectra from experiment 2, 3 and 4 are smoother than from experiment 1. De-noising by experiment 4 gives more detailed information and smoother spectrum than the others.

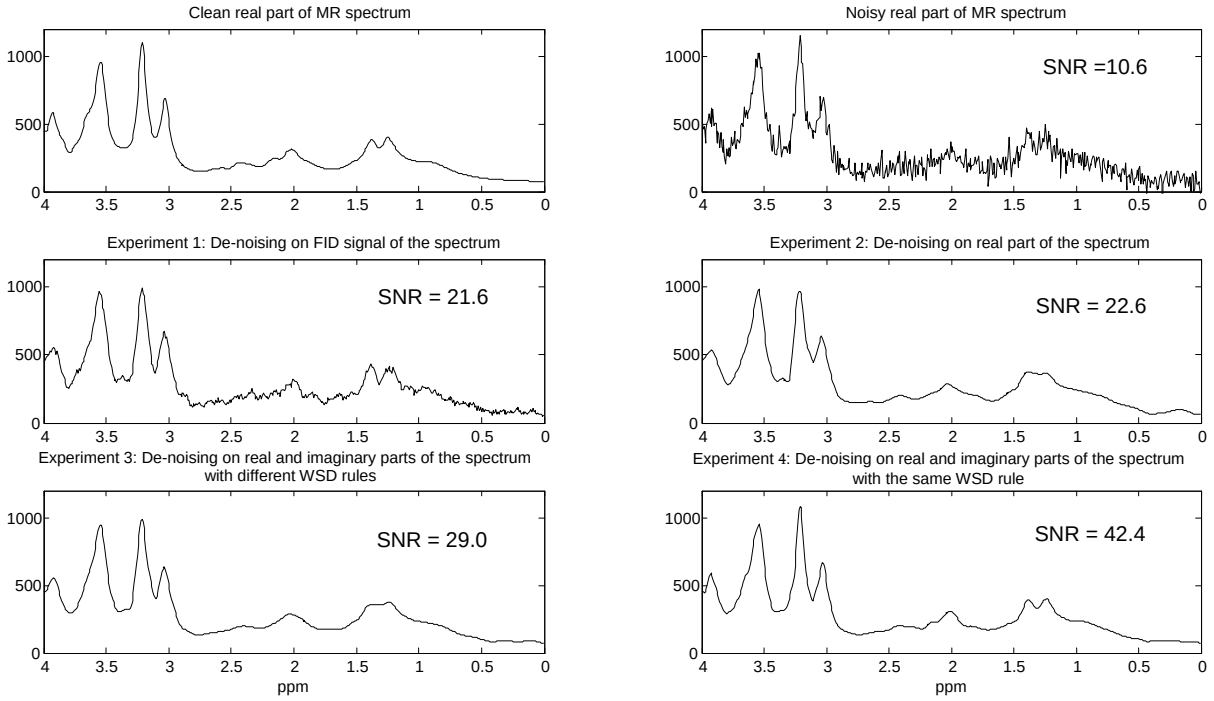


Figure 6.15: An example synthesised MR spectrum (clean and noisy) and its de-noised results in experiment 1, 2, 3 and 4.

The proposed WSD-based MRS enhancement algorithm with the best WSD rule found from experiment 4 is also applied on two experimental MR spectra and illustrated in Figure 6.16. An example of the noisy experimental MR spectrum and its de-noised results, by the proposed algorithm, are used as the inputs of the LCModel fitting (Figure 6.17), respectively. It could be seen that more metabolites with small %SD values are produced with the de-noised case. These are the estimated standard deviations expressed in percent of the estimated concentrations. A %SD < 20% has been used by many as a very rough criterion for estimates of acceptable reliability.

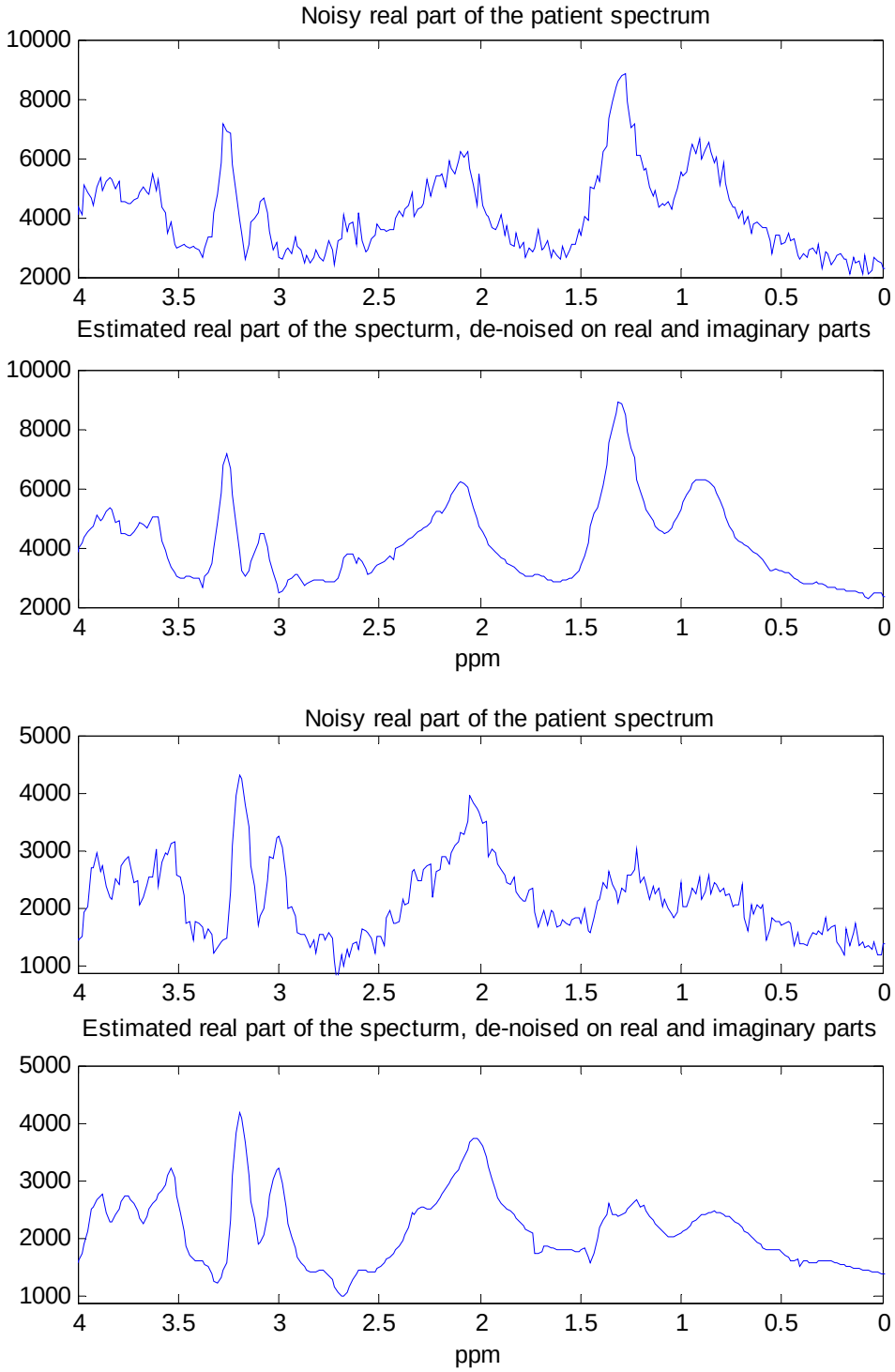


Figure 6.16: The de-noised results of two example noisy experimental MR spectrum and its de-noised signal.

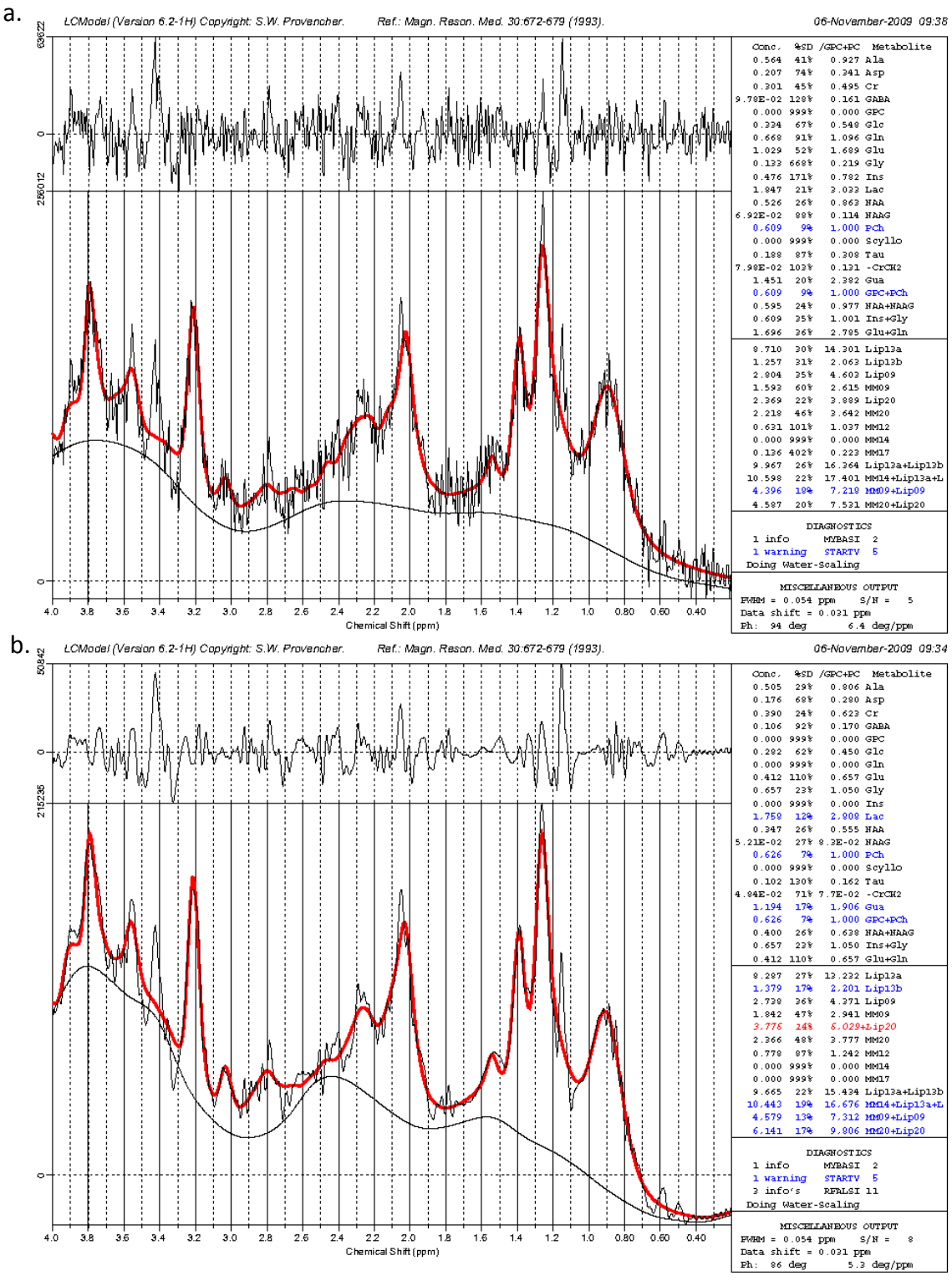


Figure 6.17: The LCModel fitting outputs of (a) the noisy MR spectrum, and (b) the de-noised MR spectrum. There are more blue coloured results on the right column of (b) than (a), which means more metabolites and MMLips with lower %SD values (less than 20%) are estimated.

6.5 Discussion

When the mean de-noised SNR values for each WSD rule for the whole synthesised dataset are plotted in each experiment, it is obvious that the overall colour of Figure 6.7 in experiment 1 is the darkest compared with Figure 6.9 in experiment 2 and Figure 6.12 in experiment 4. This implies that it is generally better to de-noise on the MR spectrum than on the FID signal. The overall grey scale in Figure 6.12 is the lightest in the three experiments. This confirms that, generally, the proposed de-noising algorithm, which utilises both the real and imaginary parts of the MR spectrum is a better approach compared to the other approaches in experiment 1 and 2.

The decomposition and reconstruction wavelet functions of bior4.4, which gives the best de-noising results is plotted in Figure 6.18. Its lineshape are very close to a single peak in the MR spectrum. There are two obvious dark horizontal lines across Figure 6.12 which correspond to the wavelet number 31, bior3.1, and wavelet number 46, rbio3.1. Figure 6.19 shows the decomposition and reconstruction wavelets of bior3.1, which are the reconstruction and decomposition wavelets of the rbio3.1 (reverse bior3.1), respectively. It is obvious that the lineshapes of the two wavelet functions are quite different from the MR spectrum examined, hence causing their poor de-noising performance. In the rest of the figure, the better performance generally comes from wavelet which has a similar lineshape as the bior 4.4. There are vertical dark columns appear in every 16 (2×8 levels) WSD rules, which correspond to the threshold rescaling method ‘one - for no rescaling’. This shows that the rescaling is necessary for the de-noising on the MR spectrum.

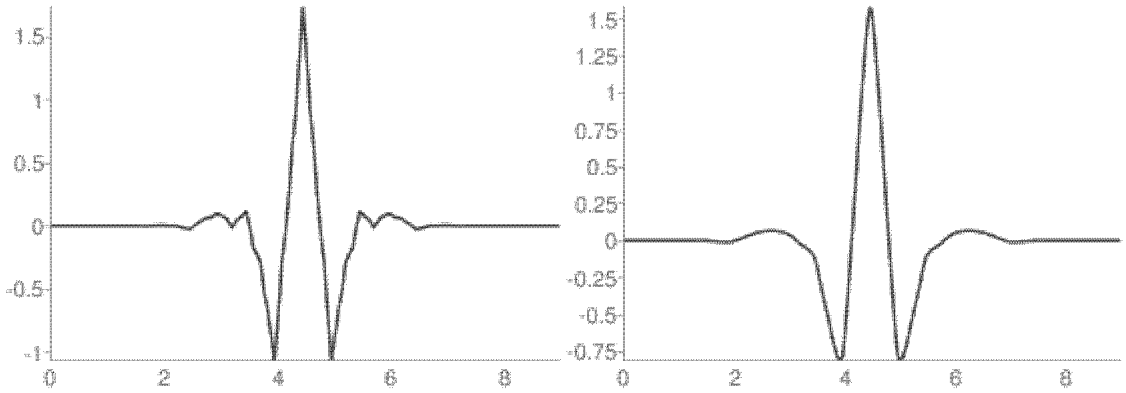


Figure 6.18: The (a) decomposition and (b) reconstruction wavelets of bior4.4.

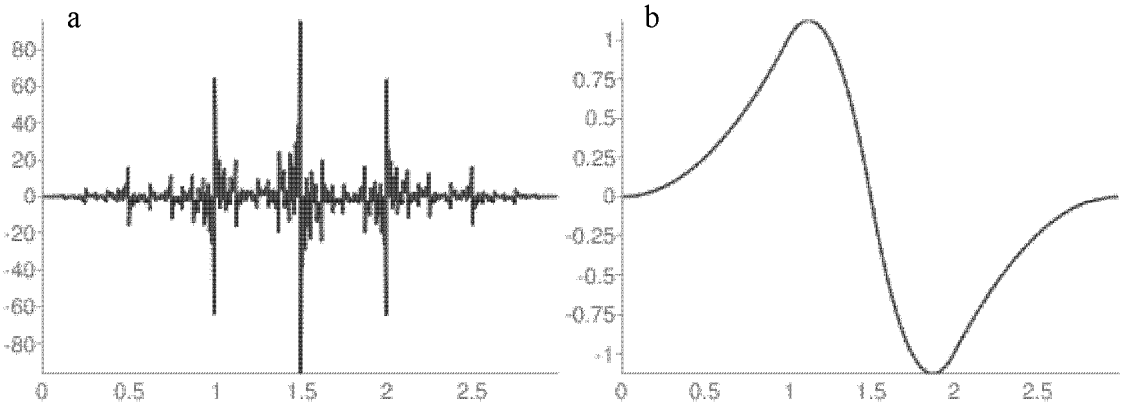


Figure 6.19: The (a) decomposition and (b) reconstruction wavelets of bior3.1, and also the (b) reconstruction and (a) decomposition wavelets of rbior3.1.

The best WSD rule for the FID signal is similar to the findings by Cancino-De-Greiff et al. (Cancino-De-Greiff et al., 2002). The best WSD rule on the real and imaginary parts of the MR spectrum in experiment 4 is different from their individual ones. Averaging the calculated real part of the spectrum and the de-noised real part of the spectrum cancels out the noise factor. The de-noised effects on the real and imaginary parts of the spectrum need to be somewhat opposite to each other in order to be cancelled out. Applying different WSD rules on the real and imaginary parts of the spectrum, such as in experiment 3, does not provide this

opposite effect on the de-noised real and imaginary parts of the spectrum. Therefore, there is not much improvement on the de-noising results compared with only apply WSD on the real part of the spectrum.

The real part of the spectrum can be derived from its corresponding imaginary part by equation (B.36), and vice versa by equation (B.37). Since the real part of the spectrum is always $\frac{\pi}{2}$ ahead of its imaginary counterpart, equation (B.36) is basically calculating the input signal at $+\frac{\pi}{2}$ phase. Similarly, equation (B.37) gives the output signal at $-\frac{\pi}{2}$ phase of its input.

De-noising using both real and imaginary parts of the spectrum utilises the embedded relation of the Lorentzian lineshape, which is the basis of the MR spectrum. This relation is not affected by the data acquisition process or environmental effects, and it is valid with any phase conditions. So the proposed enhancement algorithm can be performed at any phase condition. It can also be used for the de-noising of the imaginary part of the spectra. There may still be other applications for utilising the relationship between the real and imaginary parts of the MR spectrum, and this will be investigated in future studies.

6.6 Summary

Experiments of applying WSD for signal enhancement are performed on 150 synthesised *in vivo* ^{1}H MR spectra. It was found that the de-noised signal has higher SNR when the WSD is applied to the MR spectrum (with mean SNR of 26.44) than to its FID signal (with mean SNR of 22.7).

The best de-noising result is obtained by first applying the WSD rule (wavelet basis ‘bior4.4’, thresholding rules ‘heursure’, ‘hard’ and ‘mln’ at level 5) on both the real and imaginary parts of the MR spectrum. Then, the average of the WSD enhanced real part and derived one from the WSD enhanced imaginary part of the MR spectrum gives the final de-noised signal. The mean SNR value of the proposed algorithm is 3.58 times the noisy signal, 1.70 times the de-noised result on FID signal, and 1.46 times the de-noised result on only the real part of the spectrum.

CHAPTER 7

COMPARING FEATURE EXTRACTION AND BLIND SOURCE SEPARATION OF ICA ON CHILDHOOD BRAIN TUMOUR ^{1}H MR SPECTRA

7.1 Introduction

From the previous studies of applying ICA technique on *in vivo* MRS datasets (as reviewed in Chapter 4), it was unclear under what conditions ICA could extract individual metabolite and MMLip components and, therefore, what the precise benefits of ICA were. In this chapter, the ICA technique is first applied to synthesised MRS datasets that simulate typical childhood brain tumours, and then to the patients' experimental MRS dataset. The work presented in this chapter allows for the effects of SNR, sample size, bandwidth (measured as full-width-at-half-

maximum, FWHM), magnetic field strength and peak frequency variations to be investigated systematically (Hao et al., 2009b). The two common approaches to the application of ICA were employed with the main focus being their comparison and the identification of any critical factors of the two approaches that might affect the analysis of synthesised childhood brain tumour MRS data.

7.2 Methods

7.2.1 Synthesised datasets

Various sets of synthesised MR spectra of three classes of childhood brain tumours (TE = 30 ms, 1.5 and 3T) were generated using a quantum mechanics-based simulation program (Reynolds et al., 2006) detailed in Appendix A. Each MR spectrum was constructed as a linear combination of individual metabolite and MMLip components using the basis set 1 described in Chapter 5. A randomly distributed line-broadening (FWHM) factor was applied to each metabolite and MMLip, while a controlled amount of white noise was added. The signal-to-noise ratio (SNR) was measured linearly, as the ratio of the power (in watts) of the clean signal (the combination of individual metabolite and MMLip components) over noise.

The sum of correlation coefficients between the obtained ICs and the original individual metabolite and MMLip components was used as a measure of the quality of the resulting ICs for the synthesised datasets. ICs corresponded more closely to these components having correlation coefficients closer to 1; hence a perfectly correlated set of ICs with the original components should have a sum very close to 10 in the synthesisation experiments presented.

Each of the synthesised results, presented in the next section, was the average of 50 repeated

experiments and with a 95% confidence interval calculated by a bootstrap re-sampling method (Efron and Tibshirani, 1993, Johnson, 2001) explained in section 5.2.5.

7.2.2 Experimental dataset

Quality control criteria for the selection of experimental dataset were set based on the experiments findings from the synthesised dataset. 115 experimental MR spectra were obtained as a result of applying the following two criteria to the experimental spectra:

(1) FWHM of the water reference peak less than 6 Hz and

(2) SNR of the spectra greater than 10.

As mentioned in section 5.2.4, the SNR is defined in the LCModel as the ratio of the highest signal intensity in the 0.2–4 ppm region to twice the RMS of the LCModel fit residuals (equation (5.8)). The SNR values used in the synthesised datasets can be converted according to this definition and shown in Table 7.1.

Table 7.1: SNR values corresponding to the LCModel definition

SNRs in Synthesised Datasets	SNRs in Patient Datasets
1	3.3
4	6.6
6	8.0
10	10.4
20	14.6
40	20.8
50	23.3

7.3 Results

The synthesised datasets containing clean, noisy and de-noised signals, with all three classes (astrocytoma, ependymoma and medulloblastoma) included, were used in the generation of the results described in the following five subsections. The average spectra of all the three classes generated from the simulator are shown in Figure 7.1, with SNR equal to 6 and FWHM equal to 5.5 Hz. Figure 7.2 shows one example spectrum from each tumour class.

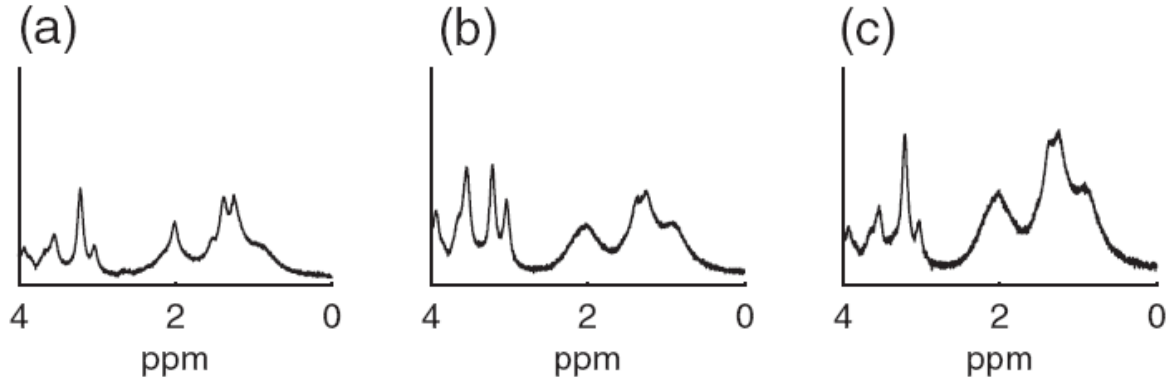


Figure 7.1: The average synthesised spectra of (a) astrocytoma, (b) ependymoma and (c) medulloblastoma

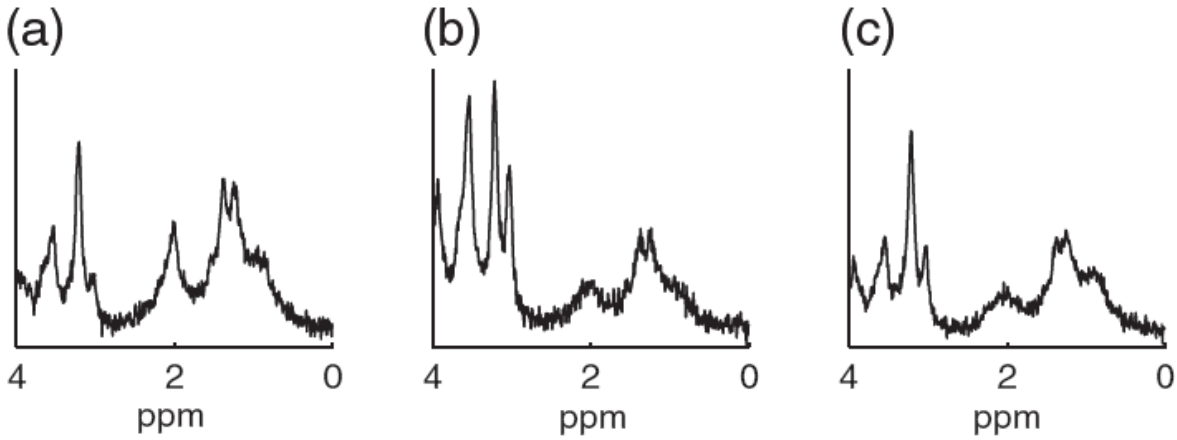


Figure 7.2: One example spectrum from each synthesised tumour class: (a) astrocytoma, (b) ependymoma and (c) medulloblastoma.

One of the sets of results for both FE-ICA and BSS-ICA methods on 600 synthesised MR spectra is illustrated in Figure 7.3. The SNR for the noisy synthesised dataset is 40, and the FWHM is 5.5 Hz. ICs from the BSS-ICA method correspond closely to individual metabolite and MMLip spectra. In contrast, for the same synthesised datasets, ICs from the FE-ICA method correspond mostly to the combination of several individual components. The MMLip components from the BSS-ICA method are sometimes affected by the metabolites appearing at the same position. It is important to note that, the BSS-ICA method treats the noise as possible independent sources, whereas the FE-ICA method removes most of the noise in its dimensionality reduction step (by PCA). Hence, when the noise level is high, more noise-related ICs appeared in the BSS-ICA method results than in the FE-ICA method. In the following experiments, the BSS-ICA algorithm is set to automatically identify the extra ICs when ICs highly related to noise are obtained in the result.

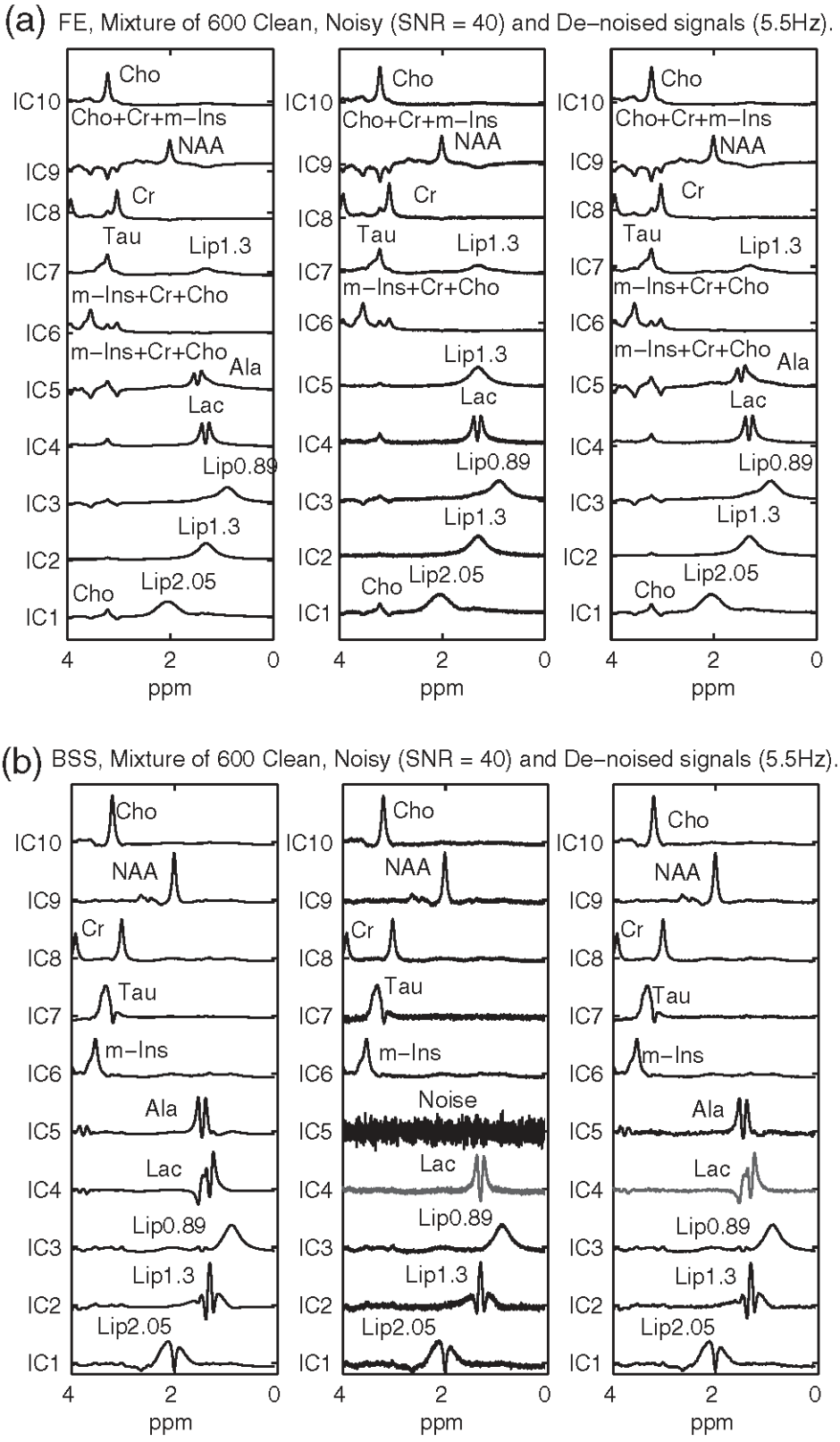


Figure 7.3: The 10 ICs obtained from the synthesized mixture of three classes by (a) the FE-ICA method and (b) the BSS-ICA method for clean, noisy and de-noised spectra labelled with the corresponding metabolite and MMLip components.

In the following subsections, the ability of ICA to extract components from the synthesised MRS datasets is examined in detail by varying the SNR, the number of spectra, the FWHM ranges, the magnetic field strength, peak frequency variations and the number of ICs in each synthesised dataset. The sum of correlation coefficients is used as a measure of the ICs' quality. In the last subsection, ICA is applied on an experimental dataset.

7.3.1 Experiments on varying SNR

To investigate the effects of SNR on the ICA methods, six different noise levels (SNR equal to 1, 6, 10, 20, 40 and 50) were used in this set of experiments. A clean synthesised dataset of 600 MR spectra including all three classes of tumour with the FWHM = 5.5 Hz was first generated, and then noise was added to obtain the required SNRs. The reason for using 600 MR spectra will be explained in the next subsection. The sum of the correlation coefficients between the obtained ICs and the independent sources is shown in Figure 7.4, where the dotted line represents the results from the BSS-ICA method, and the solid line represents the equivalent results from the FE-ICA method. The larger the value of the sum of the correlation coefficients, the greater will be the correlation between the resultant ICs and the original sources. For the de-noised synthesised dataset, both methods give better results when SNR increases from 1 to 10, and after that, relatively similar results are obtained at higher SNRs. The BSS-ICA method has a performance similar to the FE-ICA method at very low SNRs (less than about 4), whereas for the higher SNRs (greater or equal to 6), the BSS-ICA method outperforms the FE-ICA method ($p < 0.0005$). For the noisy synthesised dataset, similar results can be seen. The point when the BSS-ICA method outperforms FE-ICA appears at higher SNRs around 10 ($p < 0.0005$). Comparing the results horizontally in Figure 7.4, the BSS-ICA method has a better performance in de-noised conditions than noisy ones, while the

FE-ICA method gives similar results in both conditions. The 95% confidence interval for the BSS-ICA method is generally smaller than that of the FE-ICA method, which indicates smaller variation of the resulted ICs of the BSS-ICA method than the FE-ICA method. The detailed individual correlation coefficient between the obtained IC and its corresponding simulated metabolite/MMLip component for all the experiments performed in this section are also shown in Figure 7.5. The same experiment was performed with FWHM = 4.5, 5.0 and 6.0 Hz, respectively. Similar results as those presented in Figure 7.4 were obtained.

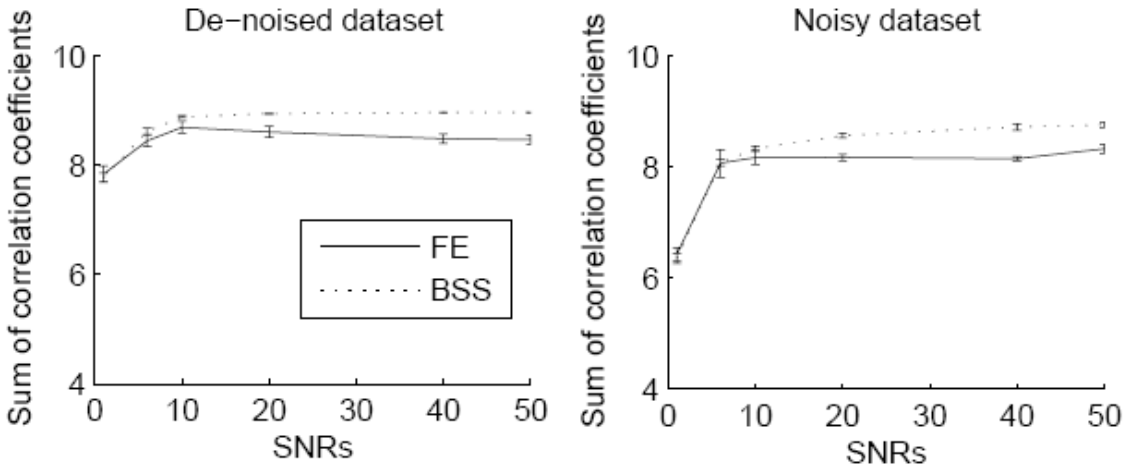


Figure 7.4: Comparison of FE-ICA and BSS-ICA methods with the SNR = 1, 6, 10, 20, 40, 50 for 600 noisy and de-noised synthesised spectral datasets at FWHM = 5.5 Hz. The error bar shows the 95% confidence interval by Bootstrap method.

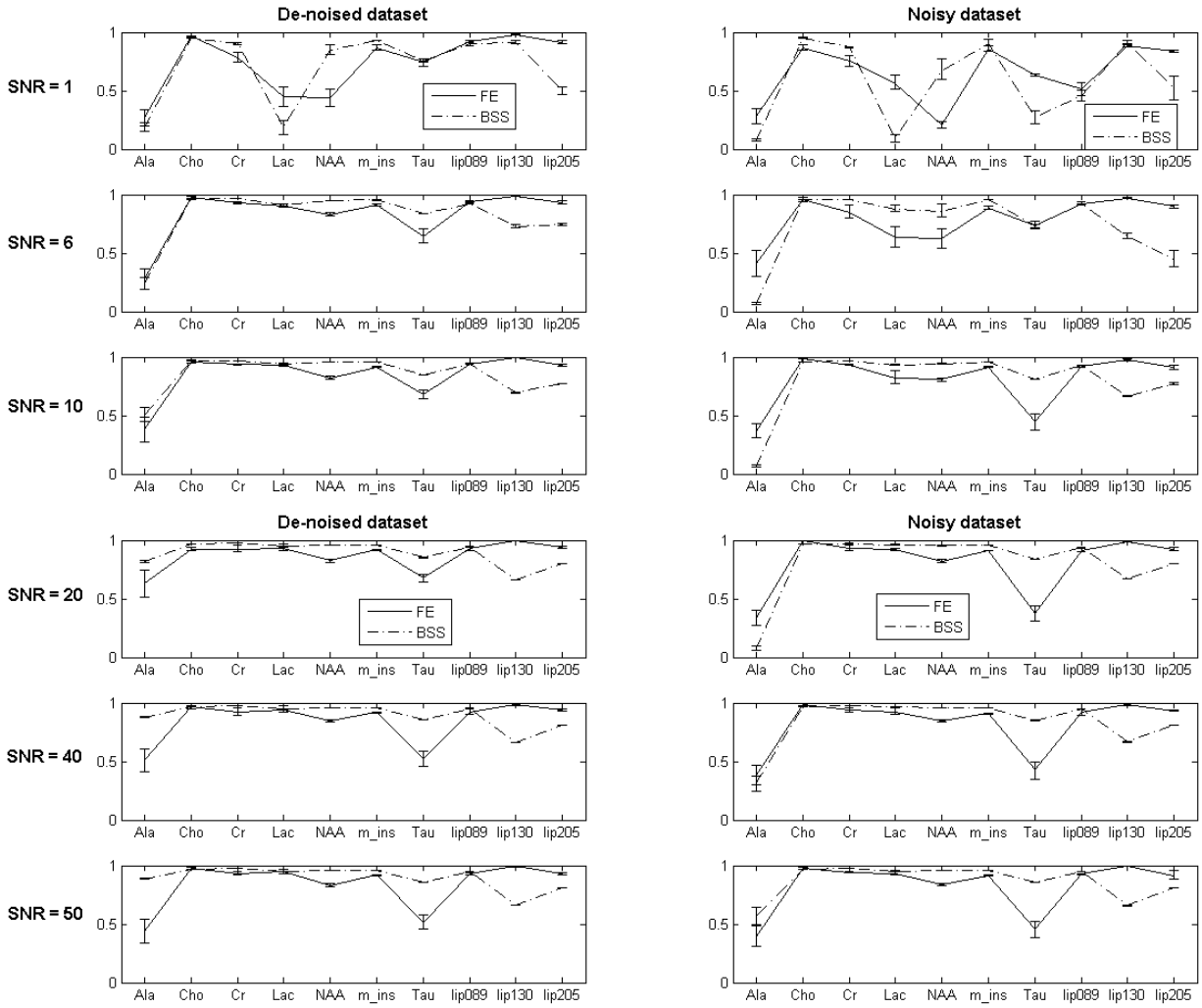


Figure 7.5: The individual correlation coefficient comparison for the FE-ICA and BSS-ICA methods with the SNR = 1, 6, 10, 20, 40, 50 for 600 noisy and de-noised synthesised spectral datasets at FWHM = 5.5 Hz. The error bar shows the 95% confidence interval by Bootstrap method.

7.3.2 Experiments on varying number of spectra

The synthesised datasets used in this set of experiments had an FWHM of 5.5 Hz, and SNR = 40. The number of spectra in each synthesised dataset was 60, 150, 300 and 600 with all three classes included. The variations in the sum of correlation coefficients between the obtained ICs and the independent sources are shown in Figure 7.6 for both the FE-ICA and BSS-ICA

methods. The performance of the BSS-ICA method is stable throughout the experiments. For the de-noised synthesised datasets, even with a small number of spectra, the BSS-ICA method can produce realistic ICs, while the results of the FE-ICA method improve with the larger number of spectra available in the set up to the inclusion of 300 spectra. For the noisy synthesised dataset, an increased number of spectra (up to 150) can improve the performance of the FE-ICA method significantly and give stable results thereafter. Performances of both the BSS-ICA and FE-ICA methods improve in de-noised conditions than in noisy ones. The 95% confidence interval for the BSS-ICA method is much smaller than that for the FE-ICA method. With 600 spectra, the results for both BSS-ICA and FE-ICA methods are stable, so when examining the effect of noise in the previous subsection, a synthesised dataset of size 600 was used to ensure that the changes in results are only caused by the SNR being the single varying parameter. The detailed individual correlation coefficient between the obtained IC and its corresponding simulated metabolite/MMLip component for all the experiments performed in this section are also shown in Figure 7.7.

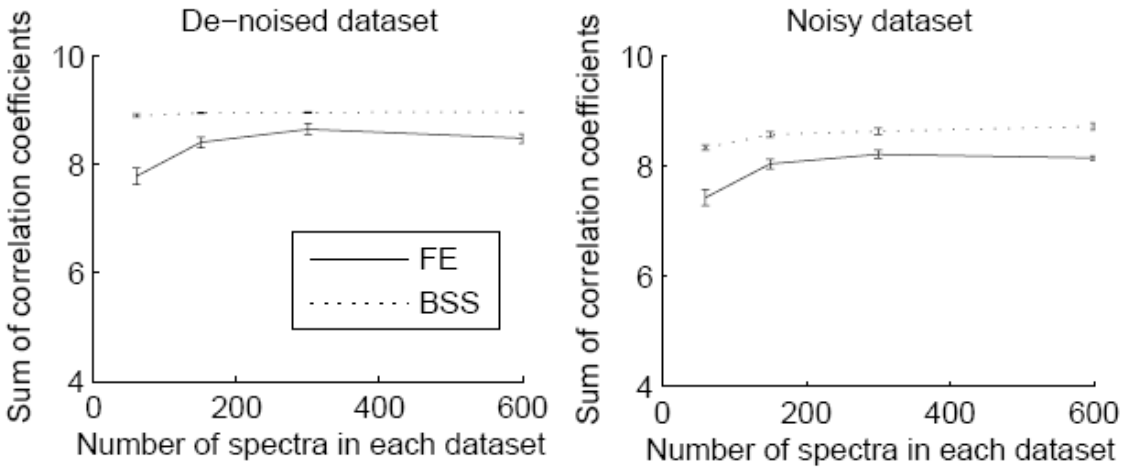


Figure 7.6: Comparison of FE-ICA and BSS-ICA methods for de-noised and noisy synthesised datasets (SNR = 40, FWHM = 5.5 Hz) with the number of spectra = 60, 150, 300, and 600. The error bar shows the 95% confidence interval by Bootstrap method.

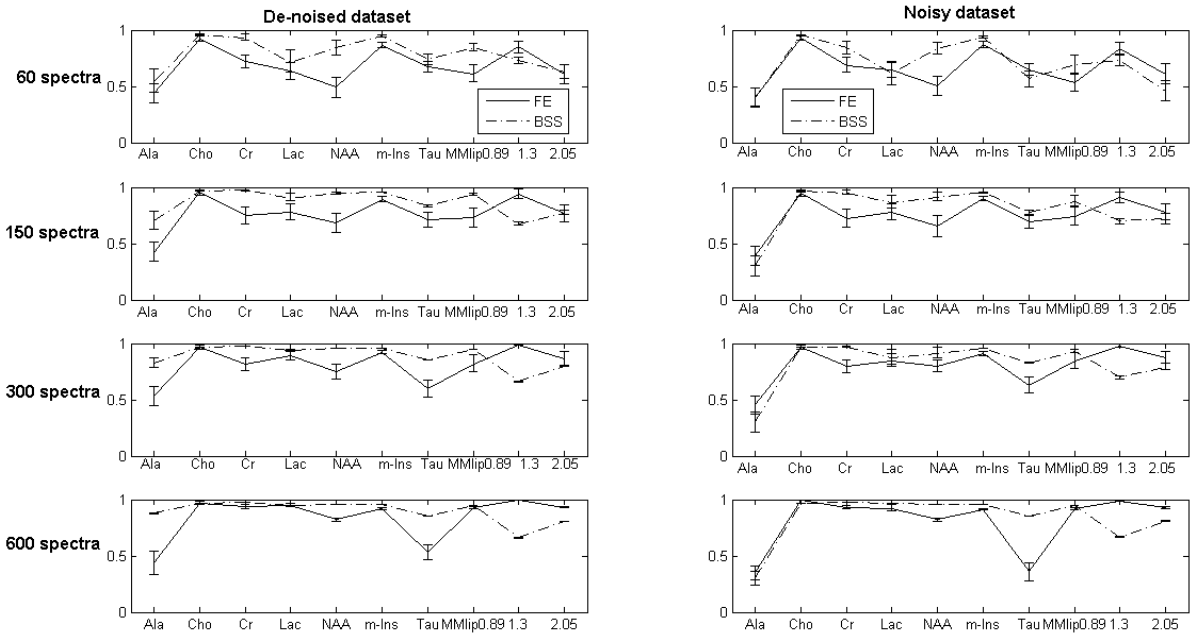


Figure 7.7: The individual correlation coefficient comparison of FE-ICA and BSS-ICA methods for de-noised and noisy synthesised datasets (SNR = 40, FWHM = 5.5 Hz) with the number of spectra = 60, 150, 300, and 600. The error bar shows the 95% confidence interval by Bootstrap method.

7.3.3 Experiments on the effect of varying FWHM

The FWHM for the synthesised tumour MRS datasets has been set to the average value 0.089±0.011 ppm (Peet et al., 2007) observed in *in vivo* MRS data of childhood cerebellar tumours. This value corresponds to approximately 5.5 (±0.6) Hz in this set of experiments. The FWHM values were varied in steps of 0.5 Hz (as seen in Figure 7.8). The synthesised dataset has 600 MR spectra when all the three tumour classes are included.

The ICs resulting from the application of both FE-ICA and BSS-ICA methods to these synthesised datasets were used to calculate the sum of correlation coefficients and are illustrated in Figure 7.8. The FE-ICA method has a better performance than the BSS-ICA method when the range of FWHM is greater than 0 Hz in the de-noised synthesised dataset

and 0.5 Hz in the noisy synthesised dataset. The performance of both FE-ICA and BSS-ICA is fairly stable until the range of FWHM is 3 Hz but decreases beyond this level. The detailed individual correlation coefficient between the obtained IC and its corresponding simulated metabolite/MMLip component for all the experiments performed in this section are also shown in Figure 7.9.

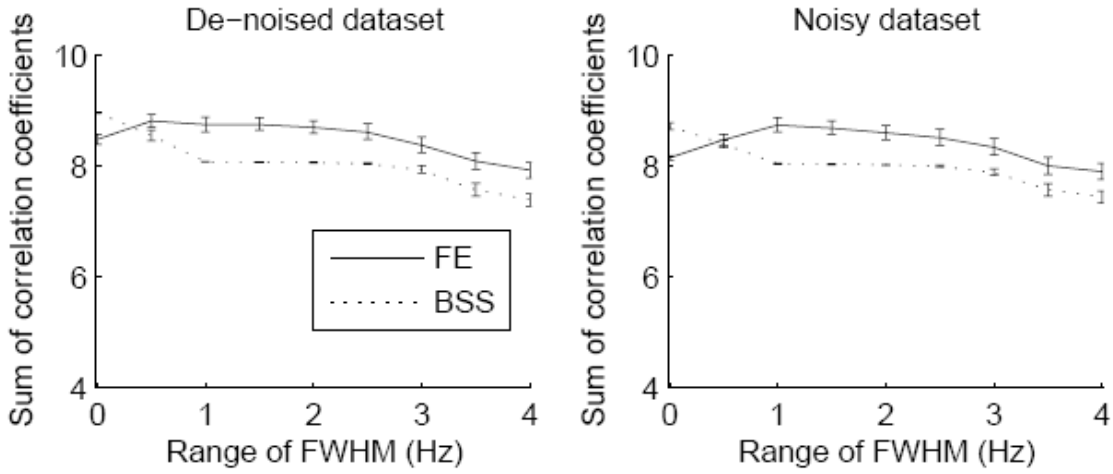


Figure 7.8: Comparison of FE-ICA and BSS-ICA methods with the FWHM = 5.5 Hz, 5.5±0.5 Hz, 5.5 ± 1 Hz, ..., 5.5 ± 4 Hz for de-noised and noisy synthesised datasets. The error bar shows the 95% confidence interval by Bootstrap method.

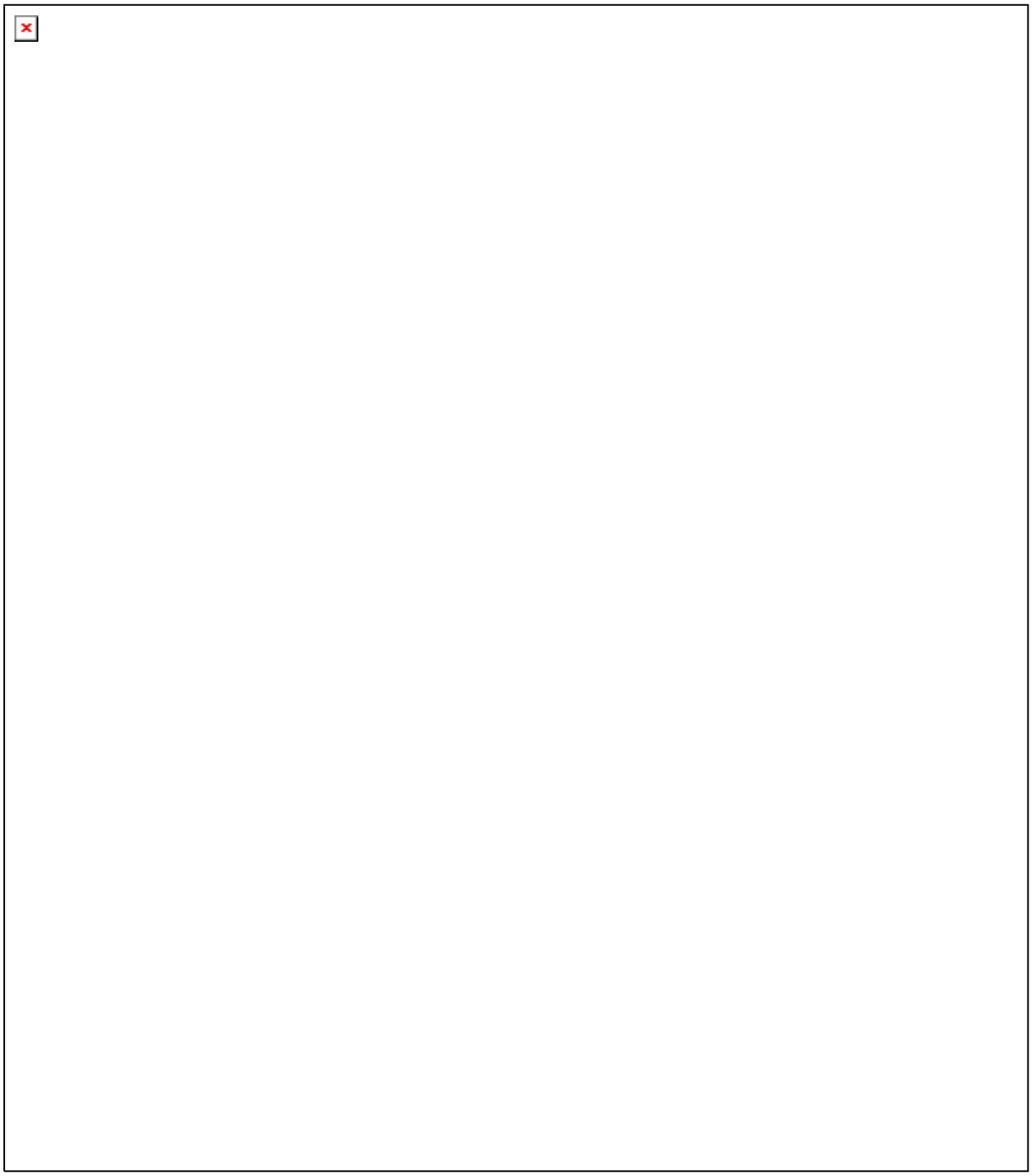


Figure 7.9: : The individual correlation coefficient comparison of FE-ICA and BSS-ICA methods with the FWHM = 5.5 Hz, 5.5±0.5 Hz, 5.5 ± 1 Hz, ..., 5.5 ± 4 Hz for de-noised and noisy synthesised datasets. The error bar shows the 95% confidence interval by Bootstrap method.

7.3.4 Experiments on varying magnetic field strength

This set of experiments examines the effect of ICA at two different magnetic field strengths, namely 1.5T and 3T for 600 spectra with FWHM = 5.5 Hz and SNR = 40. The results are shown in Figure 7.10. In the de-noised case, the BSS-ICA method produces better results at high magnetic field strengths, whereas the performance of the FE-ICA method worsens at high field strengths. In the noisy case, the BSS-ICA method performs in a consistent manner, while the FE-ICA method still performs worse than that observed at low magnetic field strengths. The detailed individual correlation coefficient between the obtained IC and its corresponding simulated metabolite/MMLip component for all the experiments performed in this section are also shown in Figure 7.11.

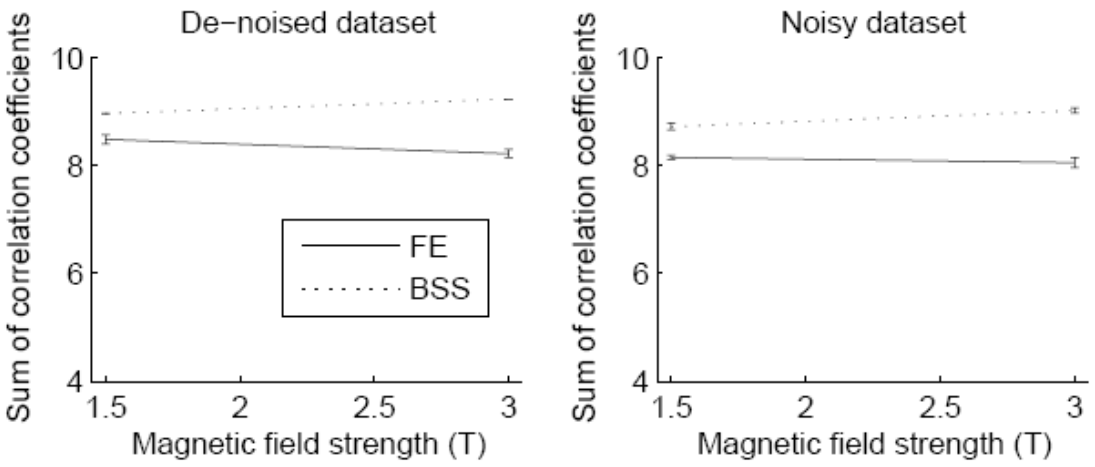


Figure 7.10: Comparison of FE-ICA and BSS-ICA methods for de-noised and noisy synthesised datasets (number of spectra = 600, FWHM = 5.5 Hz and SNR = 40) with the magnetic field strength = 1.5T and 3T. The error bar shows the 95% confidence interval by Bootstrap method.

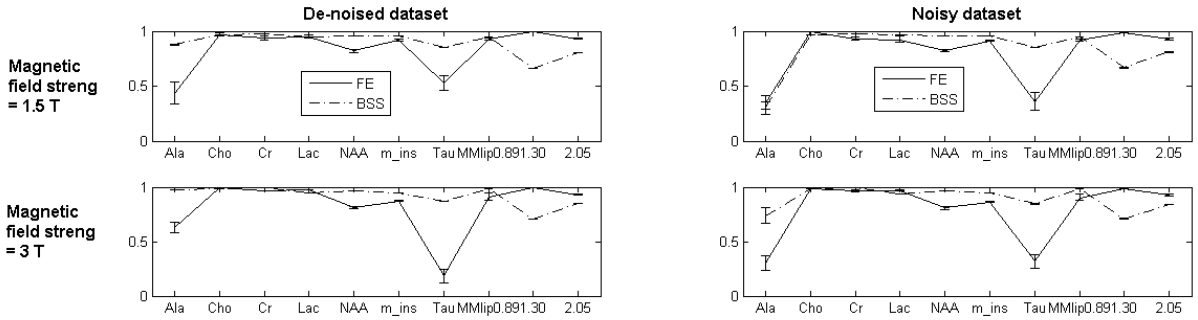


Figure 7.11: The individual correlation coefficient comparison of FE-ICA and BSS-ICA methods for de-noised and noisy synthesised datasets (number of spectra = 600, FWHM = 5.5 Hz and SNR = 40) with the magnetic field strength = 1.5T and 3T. The error bar shows the 95% confidence interval by Bootstrap method.

7.3.5 Experiments on peak frequency variations

To examine the effect of peak frequency variations on the ICA method, one of the sources (Cho) was shifted by ±0.01, ±0.03, ±0.05, ±0.06, ±0.08 and ±0.1 ppm from its original position. The Cho was chosen as it has a singlet, which does not overlap with other peaks, hence any changes in the output ICs will only be the consequences of peak frequency shifting. The synthesised datasets still contained 600 spectra of the three tumour classes with FWHM = 5.5 Hz and SNR = 40 at 1.5T. The results are shown in Figure 7.12, and the zero ppm variation case is included for comparison purpose. In the de-noised case, the results of the BSS-ICA method were not affected by the peak position variations up to ±0.05 ppm; after that point, the ICs' quality decreased and the 95% confidence interval increased. The FE-ICA method provided similar results with the zero ppm variation case up to ±0.08 ppm, and slightly worse results were obtained after that. In the noisy case, similar to the de-noised condition, the decreasing point for the BSS-ICA method was at ±0.05 ppm and for the FE-ICA method it was at ±0.08 ppm. The 95% confidence intervals for both methods increased when the frequency variations were increased. The detailed individual correlation coefficient

between the obtained IC and its corresponding simulated metabolite/MMLip component for all the experiments performed in this section are also shown in Figure 7.13.

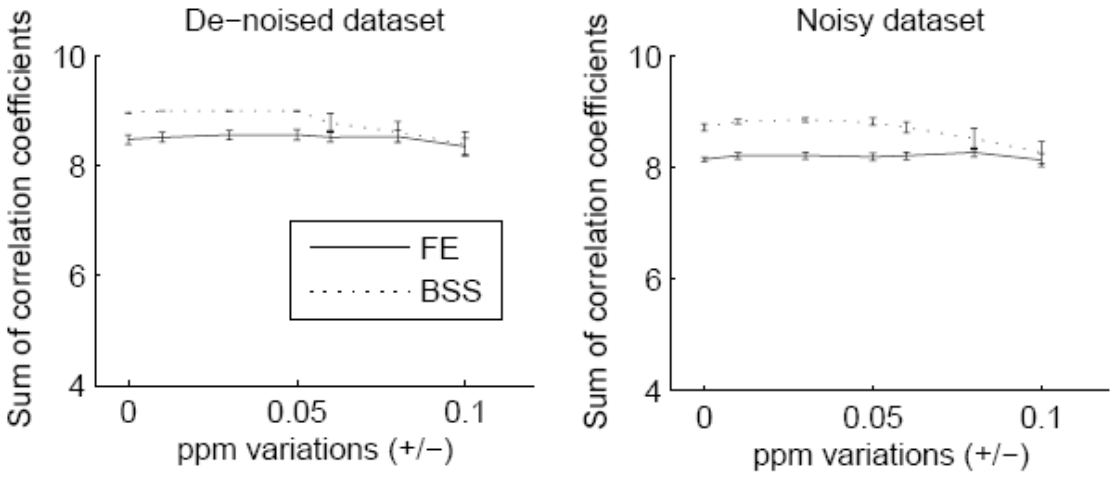


Figure 7.12: Comparison of FE-ICA and BSS-ICA methods for de-noised and noisy synthesised datasets (number of spectra = 600, FWHM = 5.5Hz, SNR = 40 and 1.5T) with the Cho peak position variations of 0, ±0.01, ±0.03, ±0.05, ±0.06, ±0.08 and ±0.1 ppm. The error bar shows the 95% confidence interval by Bootstrap method.

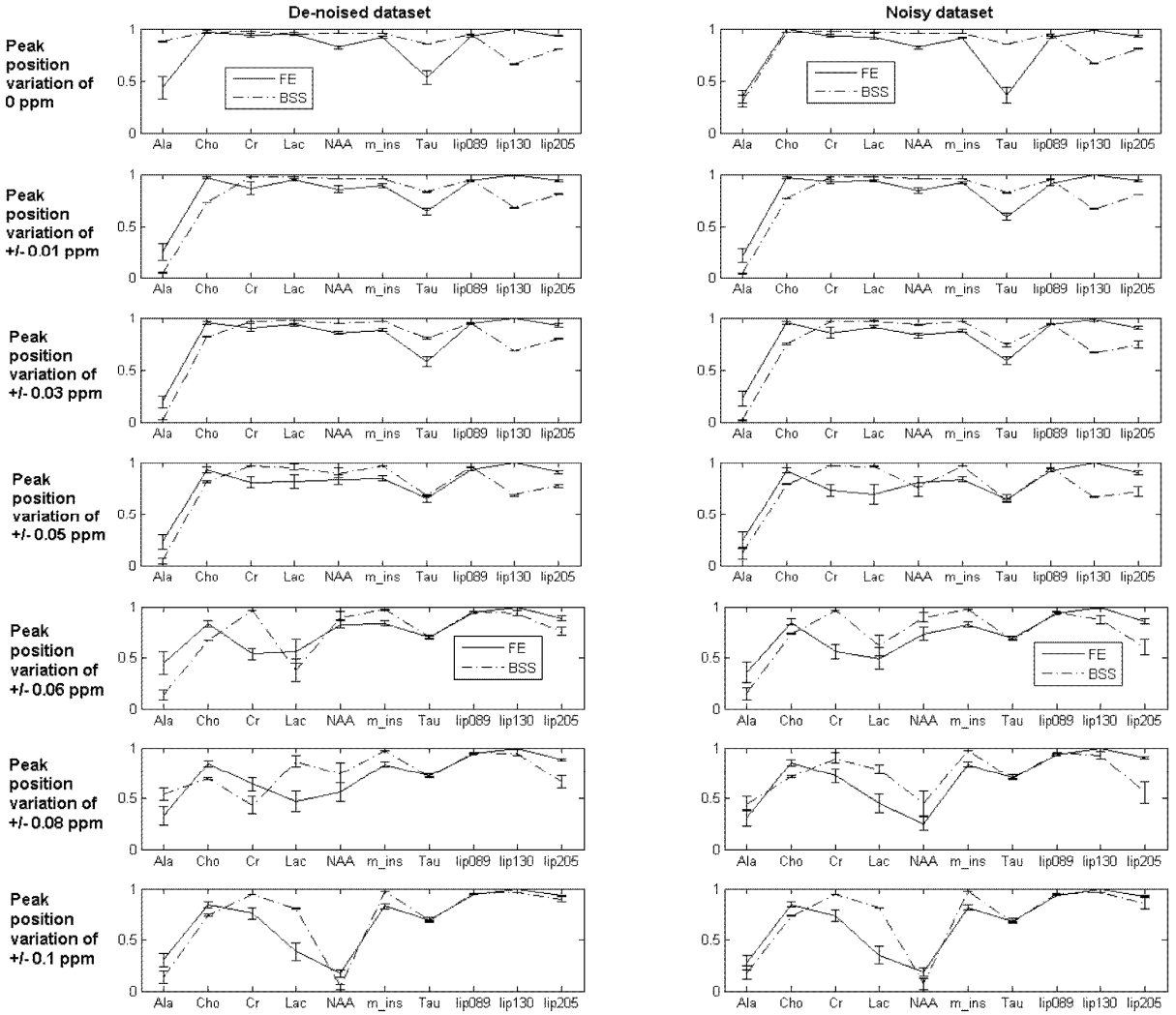


Figure 7.13: The individual correlation coefficient comparison of FE-ICA and BSS-ICA methods for de-noised and noisy synthesised datasets (number of spectra = 600, FWHM = 5.5Hz, SNR = 40 and 1.5T) with the Cho peak position variations of 0, ±0.01, ±0.03, ±0.05, ±0.06, ±0.08 and ±0.1 ppm. The error bar shows the 95% confidence interval by Bootstrap method.

7.3.6 Experiment on number of ICs

The dimensions of the synthesised datasets are all reduced from 7 to 14 to examine the sensitivity of the methods against the number of ICs. The results for one set of spectra are plotted in Figure 7.14. Varying the dimension has no observed difference for the FE-ICA method, where the ICs still contain combinations of the original metabolite and MMLip components. However, for the BSS-ICA method, when the dimension is less than 9, some

non-overlapping peaks appear in the same IC. At dimension 9, the BSS-ICA method produces 9 ICs which are very similar to the resultant ICs at dimension 10 with one of the metabolite missing. When further increasing the dimension, more noise-related ICs appear. This could be used as a rough guideline to choose the dimension of a dataset, by trial and error, for example, in the case of the experimental dataset.

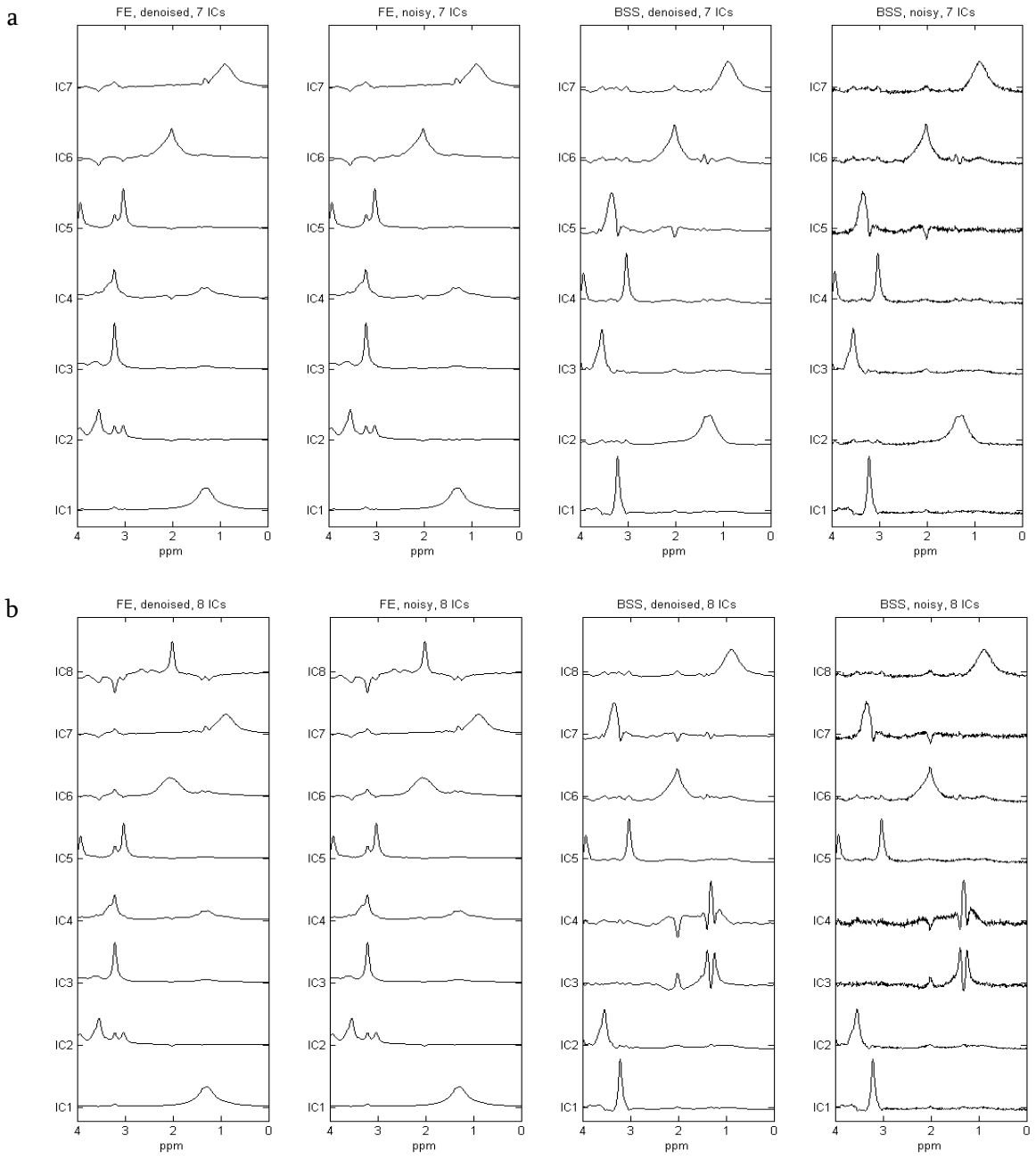
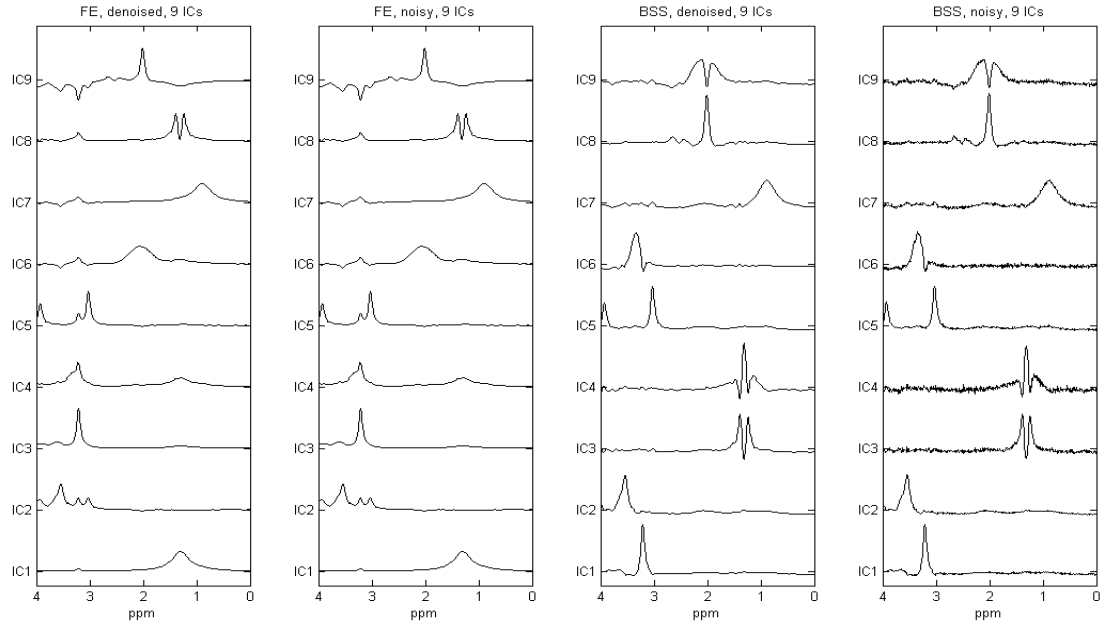


Figure 7.14: Varying the number of ICs used for the synthesised dataset (a) 7 ICs, (b) 8 ICs, (c) 9 ICs, (d) 10 ICs, (e) 11 ICs, (f) 12 ICs, (g) 13 ICs, and (h) 14 ICs. (Figure continued in the next three pages).

c



d

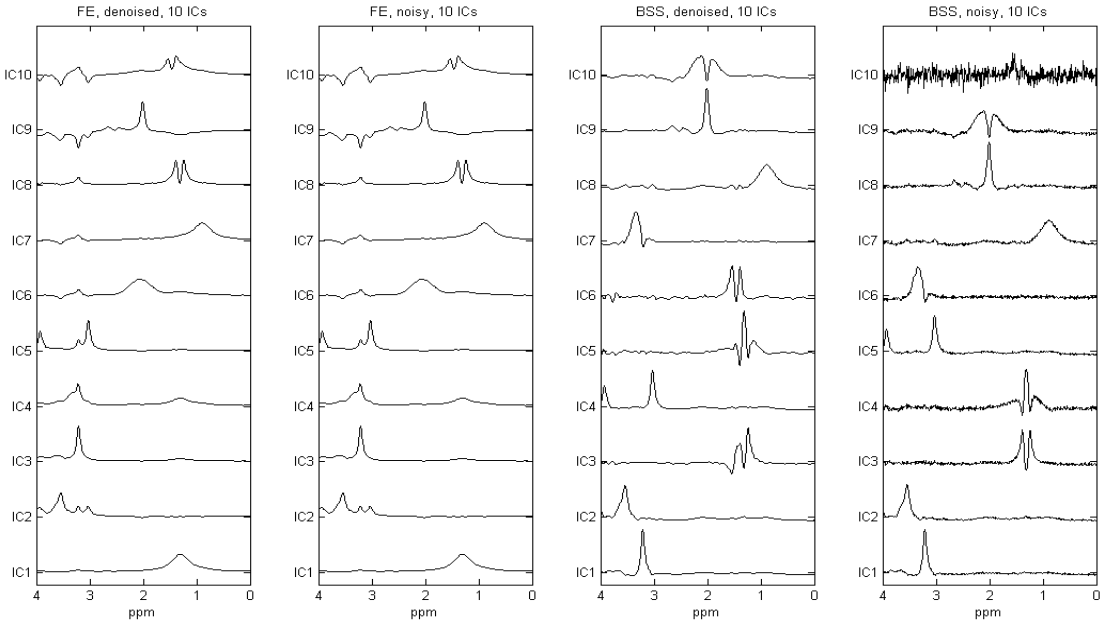


Figure 7.14: Varying the number of ICs used for the synthesised dataset (a) 7 ICs, (b) 8 ICs, (c) 9 ICs, (d) 10 ICs, (e) 11 ICs, (f) 12 ICs, (g) 13 ICs, and (h) 14 ICs. (Figure continued in the next two pages).

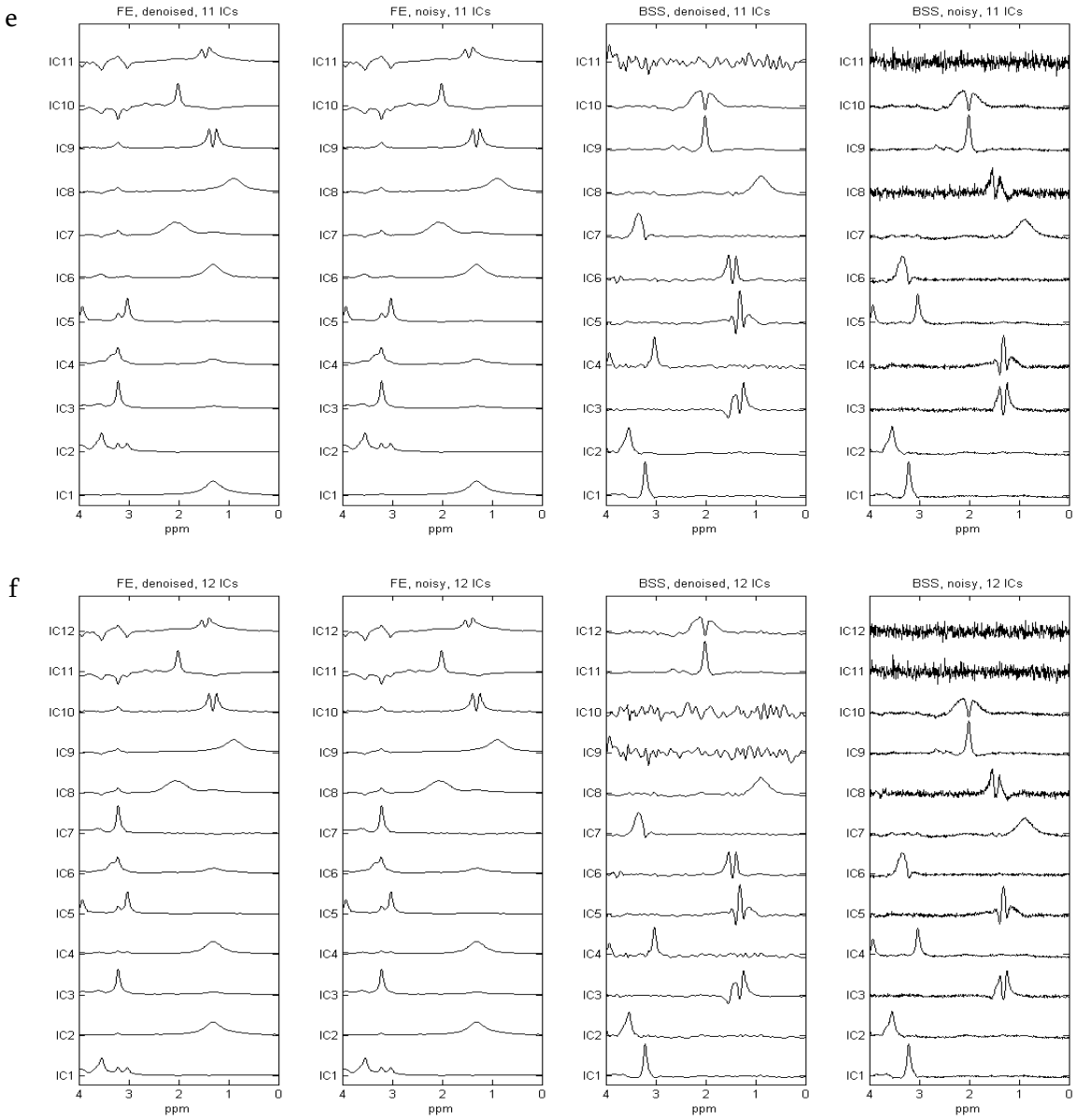
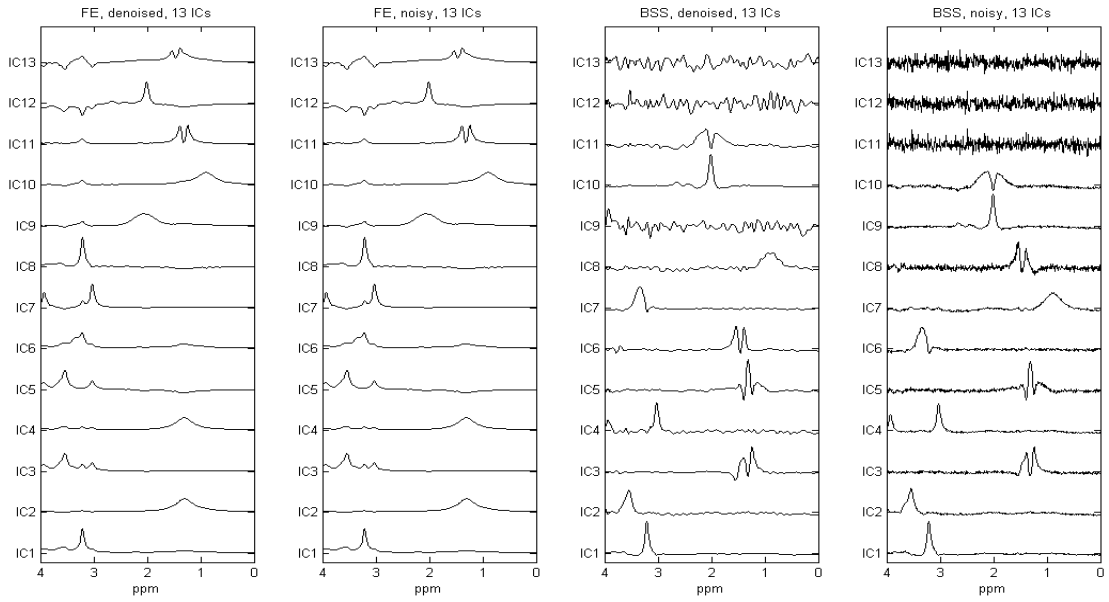


Figure 7.14: Varying the number of ICs used for the synthesised dataset (a) 7 ICs, (b) 8 ICs, (c) 9 ICs, (d) 10 ICs, (e) 11 ICs, (f) 12 ICs, (g) 13 ICs, and (h) 14 ICs. (Figure continued in the next page).

g



h

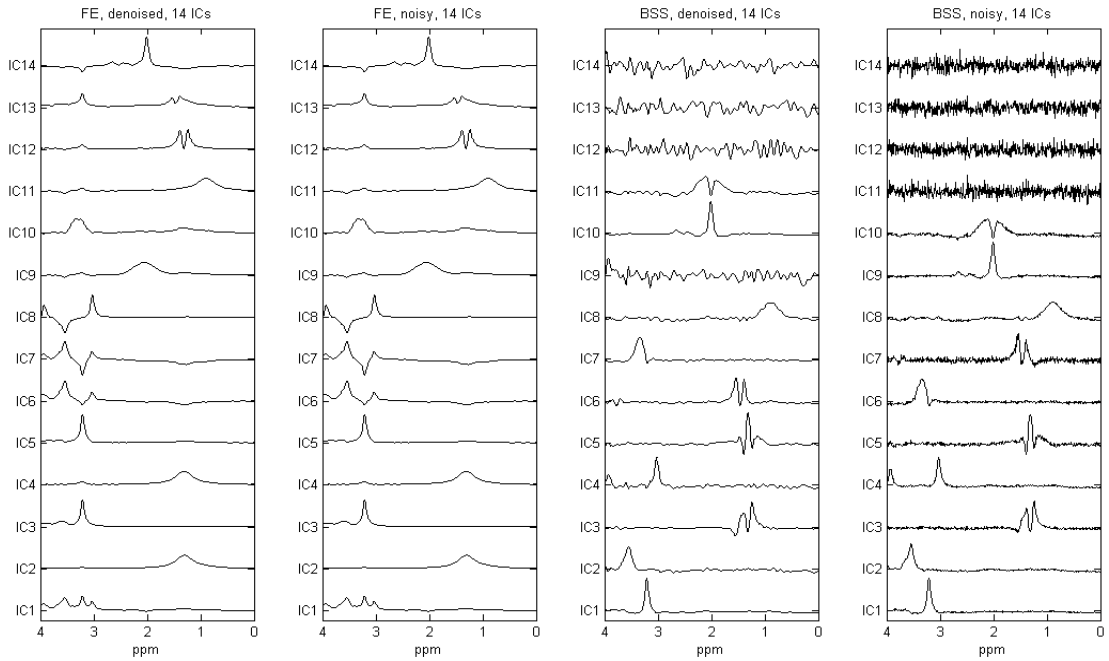


Figure 7.14: Varying the number of ICs used for the synthesised dataset (a) 7 ICs, (b) 8 ICs, (c) 9 ICs, (d) 10 ICs, (e) 11 ICs, (f) 12 ICs, (g) 13 ICs, and (h) 14 ICs.

7.3.7 Experiment on experimental dataset

Figure 7.15 and Figure 7.16 show the results from a set of 115 *in vivo* MR spectra of patients aged under 16 years. The PCA is used to reduce the experimental dataset dimension to 15. The eigenvalue sequence of the covariance matrix for the experimental dataset decreased sharply, where ten ICs can already explain more than 99.3% of the dataset. However, when ten ICs were obtained, non-overlapping peaks were seen in the same IC for the BSS-ICA method. This indicates that a sufficient number of ICs were not estimated. On increasing the number of ICs in an attempt to keep the non-overlapping peaks from appearing in one IC, it was found that the ICs were relatively stable after 15. The results were similar for both FE-ICA and BSS-ICA methods with the synthesised datasets. In the FE-ICA method (Figure 7.15), the MMLip components were mostly picked up and the three metabolites, Cr, Cho and m-Ins, were appearing in the same ICs as in the synthesised case. The overlapping peaks were not affected by each other in the FE-ICA method results. In the BSS-ICA method (Figure 7.16), several non-overlapping metabolites (IC4 m-Ins, IC10 Cr, IC14 guanidoacetate (Gua), IC15 Tau) and MMLip (IC6 MMLip at 0.9 ppm) components were obtained as expected. The overlapping peaks as shown in IC8, IC12 and IC1, IC2, IC5, IC7 were extracted similar to the synthesised results.

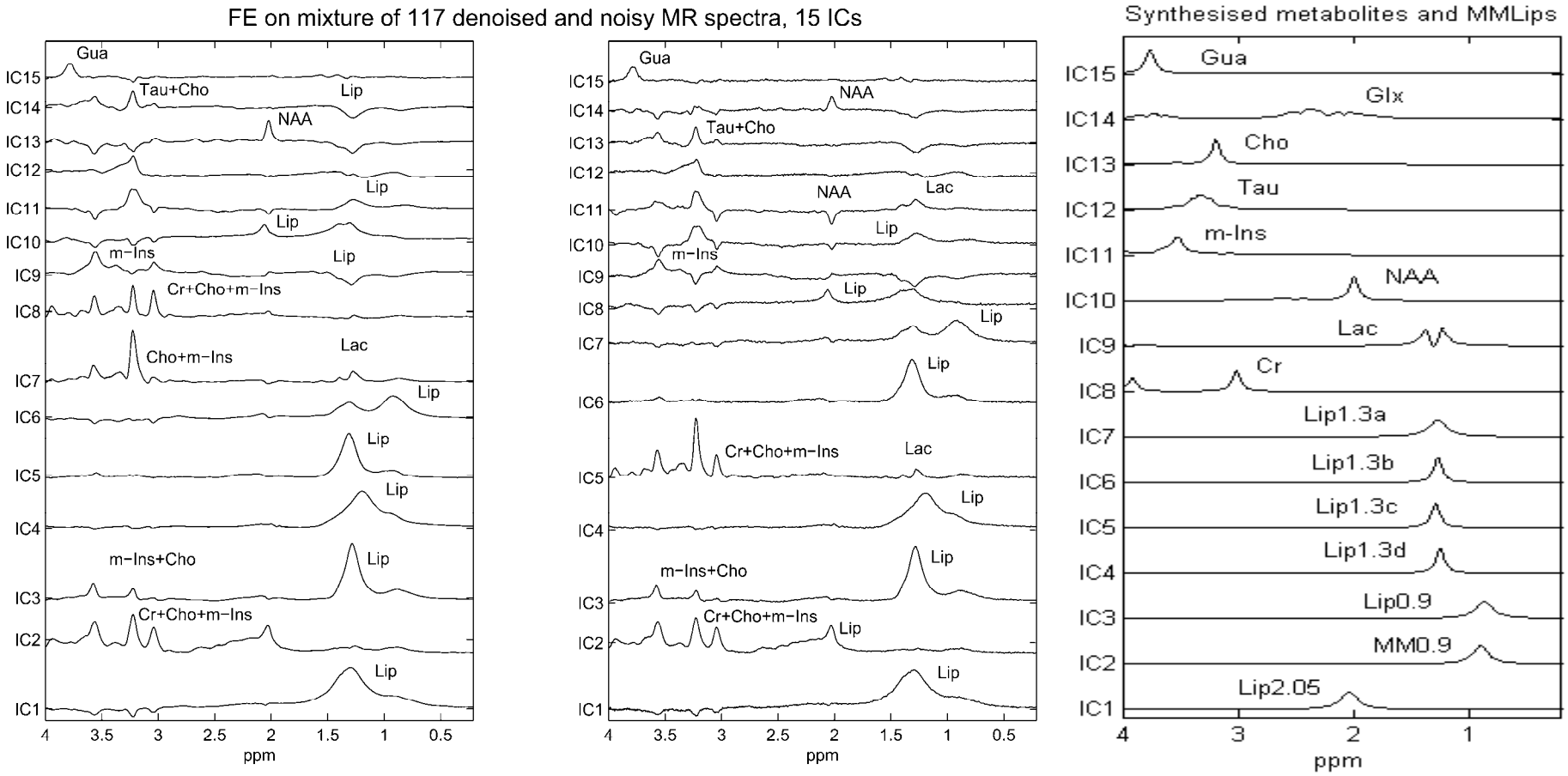


Figure 7.15: The 15 ICs obtained from the mixture of 115 experimental MR spectra by the FE-ICA method for de-noised (left) and noisy (middle) conditions labelled with the corresponding metabolites and lipids. The synthesised metabolites and MMLips spectra (right) are also plotted for comparison,

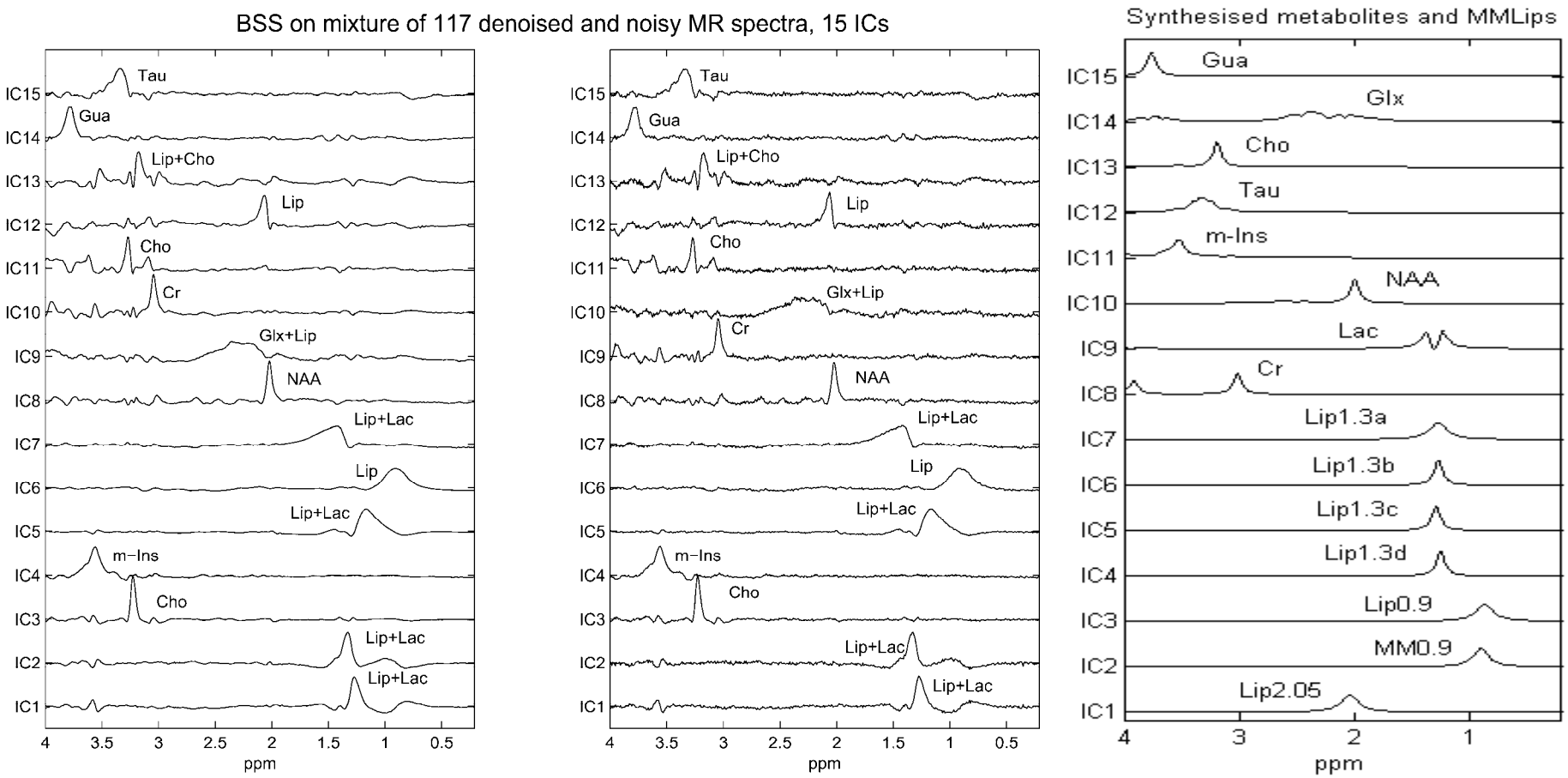


Figure 7.16: The 15 ICs obtained from the mixture of 115 experimental MR spectra by the BSS-ICA method for de-noised (left) and noisy (middle) conditions labelled with the corresponding metabolite and MMLip components. The synthesised metabolites and MMLips spectra (right) are also plotted for comparison,

7.4 Discussion

As noise was added to the synthesised datasets, the performance of the BSS-ICA method degraded with respect to that of the FE-ICA method. This is due to the BSS-ICA method treating the noise as independent sources of original signal, whereas under the FE-ICA method, it was removed by PCA at the pre-processing stage. The missing component in the BSS-ICA method may reappear when the number of ICs is increased. The automated identification of the ICs, which contain mainly noise, was implemented in the BSS-ICA method. The identification of these noise ICs is based on the fact that noise signal is more Gaussian than that of the metabolite and MMLip components.

When the SNR increases to 10 in the de-noised case, the effect of removing the noise makes less difference to the result of the FE-ICA method than those of the BSS-ICA method. It is at this level of SNR that the BSS-ICA method begins to outperform the FE-ICA method. The BSS-ICA method restores almost all the original metabolite and MMLip components except where there is significant overlapping of MMLips with metabolites. The FE-ICA method tends to be better than the BSS-ICA method where overlapping peaks are present, in particular where broad and narrow signals overlap; but it gives ICs that combine metabolite and MMLips at other frequencies. This is consistent with the assumptions taken in the BSS-ICA and FE-ICA methods by ICA. These overlapping peaks should be more correlated with each other than with the rest of the components, thus the assumption in the BSS-ICA method might not be entirely true in this condition and the assumption in the FE-ICA method seems to be more suitable for this condition. Combining the advantages of FE-ICA and BSS-ICA methods could improve the quality of the ICs, which will be investigated in details in the next chapter.

With a good SNR (greater than 10), the BSS-ICA method appears to perform better than the FE-ICA method, even with small numbers of spectra in the dataset. As can be seen in Figure 7.5, on varying the number of spectra available in the synthesised dataset, the ICs from the BSS-ICA method remain accurate. The results from the FE-ICA method improve when the number of spectra increases to 300; while further increasing the number of spectra had no marked effect on the resulting ICs. The error bars on the results of the BSS-ICA method are generally smaller than those of the FE-ICA method, indicating that BSS-ICA method performs more consistently than the FE-ICA method. Theoretically, the BSS-ICA method can be applied to any dataset with a sample number greater than or equal to the original mixing sources. So even with a dataset size smaller than 300, as long as it is greater than or equal to the number of metabolite and MMLip components contained in the mixture, the BSS-ICA method should be working, but the FE-ICA method may not.

Increasing the range of FWHM has an interesting effect on the FE-ICA method, its performance improves first before stabilizing for both the de-noised and noisy conditions. This suggests that the method is more robust and more effective on metabolite and MMLip components spectra over a wide range of frequencies compared to the BSS-ICA method. This can also be seen in the first part of Figure 7.3 where the MMLip components occupy a wider range of frequencies than the metabolite components. The experiment on magnetic field strength shows results consistent with the above. At higher field strength, the metabolite and MMLip components are generally narrower than those observed at lower field strengths. The results of the FE-ICA method are worse at 3T compared to those at 1.5T, showing that the FE-ICA method is less effective on narrow spectral components. The BSS-ICA method, on the other hand, performs better at 3T than at 1.5T, indicating that it is more effective when the components are narrower and preferably not overlapping.

The FE-ICA method has slightly higher tolerance for the peak frequency variations than the BSS-ICA method. If the frequency variations on the peak position are large, the ICA algorithm will treat them as different components. Hence, multiple ICs might be found with peaks closer to one another in the BSS-ICA method.

The correlation coefficients between the estimated mixing matrices and the original mixing matrices are calculated for the synthesised datasets. It shows that for both methods, when ICs are closely related to the metabolite or MMLip components, the corresponding correlation coefficients of the mixing matrices are very close to 1. This indicates that the mixing vectors are successfully estimated with possibly some scale differences. When the ICs are distorted by overlapping peaks or multiple components, the corresponding correlation coefficients will have a strong correlation with more than one IC or have weak correlation with any IC. The scale differences are mainly caused by the normalization in the ICA algorithm, where the matrix **S** is normalized to have the standard deviation equal to 1 in its rows. As the estimated mixing matrix is strongly correlated to the original one with only some scale differences for these well estimated ICs, it could be used for quantification and classification purposes. However, the accuracy of those applications might be limited by the quality of the estimated ICs. This requires further study which is will be dealt with in the next chapter.

The synthesised datasets are simulations of the three common tumour classes in paediatric brain tumours to demonstrate its applicability to real case data. However, the ICA method is not limited to those tumour classes, any brain tumour class or even metabolic case fulfilling the quality requirements examined in this study can be included in the dataset for analysis. As the baseline factor is not included in the synthesised dataset, when applying the ICA method on the experimental dataset, the baselines need to be removed prior to the analysis.

Similar results were obtained with the experimental patient dataset and the synthesised datasets. As in the synthesised case, for the FE-ICA method, the non-overlapping peaks of Cr, Cho, m-Ins and Lac appear in same ICs, and overlapping peaks are picked up undistorted. For the BSS-ICA method, non-overlapping metabolite and MMLip components are obtained as expected. The un-separated Cho peak in IC11 and IC13 could be explained by the peak frequency variations being greater than ±0.05 ppm.

7.5 Summary

The BSS- and FE-based ICA are capable of extracting individual metabolite and MMLip component from a set of MRS data that closely simulate *in vivo* MRS of childhood brain tumours. Similar results were obtained when the BSS-ICA and FE-ICA methods were applied to a quality controlled *in vivo* experimental MRS dataset. The BSS-ICA method is more sensitive to noise; therefore, de-noising is necessary before using ICA for the BSS-ICA method. The BSS-ICA method appears to outperform the FE-ICA method particularly with small numbers of spectra in the synthesised dataset. The FE-ICA method is more dependent on the number of spectra, so a large dataset is required when performing the FE-ICA method with ICA. We suggest that the BSS-ICA method gives better performance for the dataset with an SNR greater than or equal to 10, and a small range of FWHM and peak frequency variations for small or large number of spectra. The FE-ICA method should give better performance when a large dataset (greater or equal to 300) with low SNR (less than or approximately equal to 4) and large variations of FWHM and peak frequency are used. The optimal conditions for a reliable and repeatable experiment under either method, in the context of MRS analysis, are obtained with datasets with a minimum number of 300 spectra and an SNR greater than or equal to 10. The FE-ICA method is more robust and more

effective on metabolite and MMLip components' spectra containing a wide range of frequencies, whereas the BSS-ICA method is more effective on spectra with frequencies over a narrow range. The FE-ICA method is limited in that it performs very poorly when a combination of metabolite and MMLip components appears in the same IC, and a large sample size is required. Most of the problems with the BSS-ICA method were caused by the overlapping peaks and the existence of high levels of noise.

Since the advantages and limitations are different for both methods, combining the advantages of the two methods may compensate their disadvantages and lead to better decomposition results for revealing further hidden information in MRS. A novel ICA approach involving a hybrid of BSS and FE techniques for automated decomposition of a series of MR spectra is proposed in the next chapter, in which, the disadvantages of both methods are compensated by each other, hence a better decomposition result is generated.

CHAPTER 8

A NOVEL HYBRID METHOD OF APPLYING ICA ON *IN VIVO* ^{1}H MR SPECTRA OF CHILDHOOD BRAIN TUMOURS FOR AUTOMATIC DECOMPOSITION

8.1 Introduction

In Chapter 7, a systematic comparison of the ability of ICA was performed on both simulated and experimental *in vivo* MRS datasets of childhood brain tumours (Hao et al., 2009b). It compared the performance of two ICA-based algorithms, i.e., FE and BSS algorithms and found that the FE-based ICA method is limited in that a combination of metabolite and MMLip components commonly appears in the same IC, and a large sample size is required, whereas most of the problems with the BSS-based ICA method were caused by overlapping peaks. It concluded that stable decomposition results could be achieved with 300 MRS at an

SNR (defined in LCModel (Provencher, 2009)) greater or equal to 10.

Since both FE- and BSS-based ICA algorithms have their advantages and disadvantages, combining the two methods could potentially help compensate their problems and improve the decomposition on MRS data. The aim of this chapter is to improve the extraction of meaningful individual metabolite and MMLip components from the MRS dataset. A novel ICA approach involving a hybrid of BSS and FE techniques for automated decomposition of MR spectra is proposed. An SNR-based automatic feature dimension selection algorithm is also proposed together with the hybrid ICA algorithm. Experiments on synthesised and experimental *in vivo* childhood brain tumours MRS datasets are performed to show that the new hybrid independent components (ICs) comprise the advantages of both BSS-ICA and FE-ICA results.

8.2 Theory

8.2.1 BSS- and FE-based ICA

Recall from Chapters 3 and 4 that the ICA algorithm can be applied to a mixture observation matrix $\mathbf{X}$ to derive $\mathbf{S}$ by the following linear transformation,

$$\mathbf{X} = \mathbf{AS} \tag{8.1}$$

$$\mathbf{S} = \mathbf{BX} \tag{8.2}$$

where $\mathbf{B}$ is the unmixing matrix, $\mathbf{B} = \text{pinv}(\mathbf{A})$, which is estimated based on the maximization of independency of $\mathbf{S}$.

The unmixing processing usually consists of two steps, i.e. the whitening and separating steps. In the whitening step, the observed signal $\mathbf{X}$ is decorrelated. The signal dimension reduction can also be performed during this step based on eigenvalue feature selection criterion (Jolliffe, 2002). In the separating step, an orthogonal separation matrix is estimated based on maximization of independency criteria.

As presented in last chapter, the ICA algorithm can be applied in two ways. One is based on the independency of the metabolite and MMLip components, which is referred as the BSS-based ICA method. The other is based on the independency of their concentration coefficients, i.e. the FE-based ICA method. For a given MRS dataset, once the concentration coefficients are known, the corresponding metabolite and MMLip components can be uniquely determined.

For the BSS-ICA method, the assumption taken is that the metabolite and MMLip components are statistically independent from each other. However, this is not always the case. The experimental results from Chapter 7 (Hao et al., 2009b) showed that this assumption worked well to extract the metabolite and MMLip components that are not overlapping with each other. But when the overlapping happens, the resultant ICs are truncated by each other.

The FE-ICA method is based on the idea that the observed signal is a composition of many bases with different weighting coefficients. In this case, the bases are spectra of metabolite and MMLip components and composition weighting coefficients indicate their concentrations for the mixture. The concentration coefficients are considered independent from each other. In the FE-ICA method, the concentration coefficients matrix will be $\mathbf{S}$, the metabolite and MMLip components matrix will be $\mathbf{A}$ (with spectra of metabolite and MMLip as its column

vectors) in equation (8.1). The FE-ICA algorithm will recover the concentration coefficients $\mathbf{S}_{m\times l}^{\mathrm{FE}}$, and as such $\mathbf{A}_{d\times m}^{\mathrm{FE}}$ can be determined. The experimental results from Chapter 7 show that this method does not have a problem with overlapping peaks in the same way as the BSS-ICA method does. It can extract the overlapping peaks in an intact manner. The problem with FE-ICA though is that a single IC may contain peaks from many different metabolites.

A hybrid ICA algorithm combining the advantages of the above mentioned two approaches is needed. In the next section we present a hybrid ICA method, which incorporates the advantages of both BSS-ICA and FE-ICA methods.

8.2.2 Hybrid ICA algorithm

The proposed hybrid ICA algorithm flow chart is illustrated in Figure 8.1. The MR spectra are de-noised at the beginning of the process by wavelet shrinkage de-noising. The BSS-ICA decomposes the de-noised MRS dataset as $\mathbf{S}_{m\times d}^{\mathrm{BSS}} = \mathbf{B}_{m\times l}^{\mathrm{BSS}}\ \mathbf{X}_{l\times d}$, where dimension is reduced to m by the method described in the previous section.

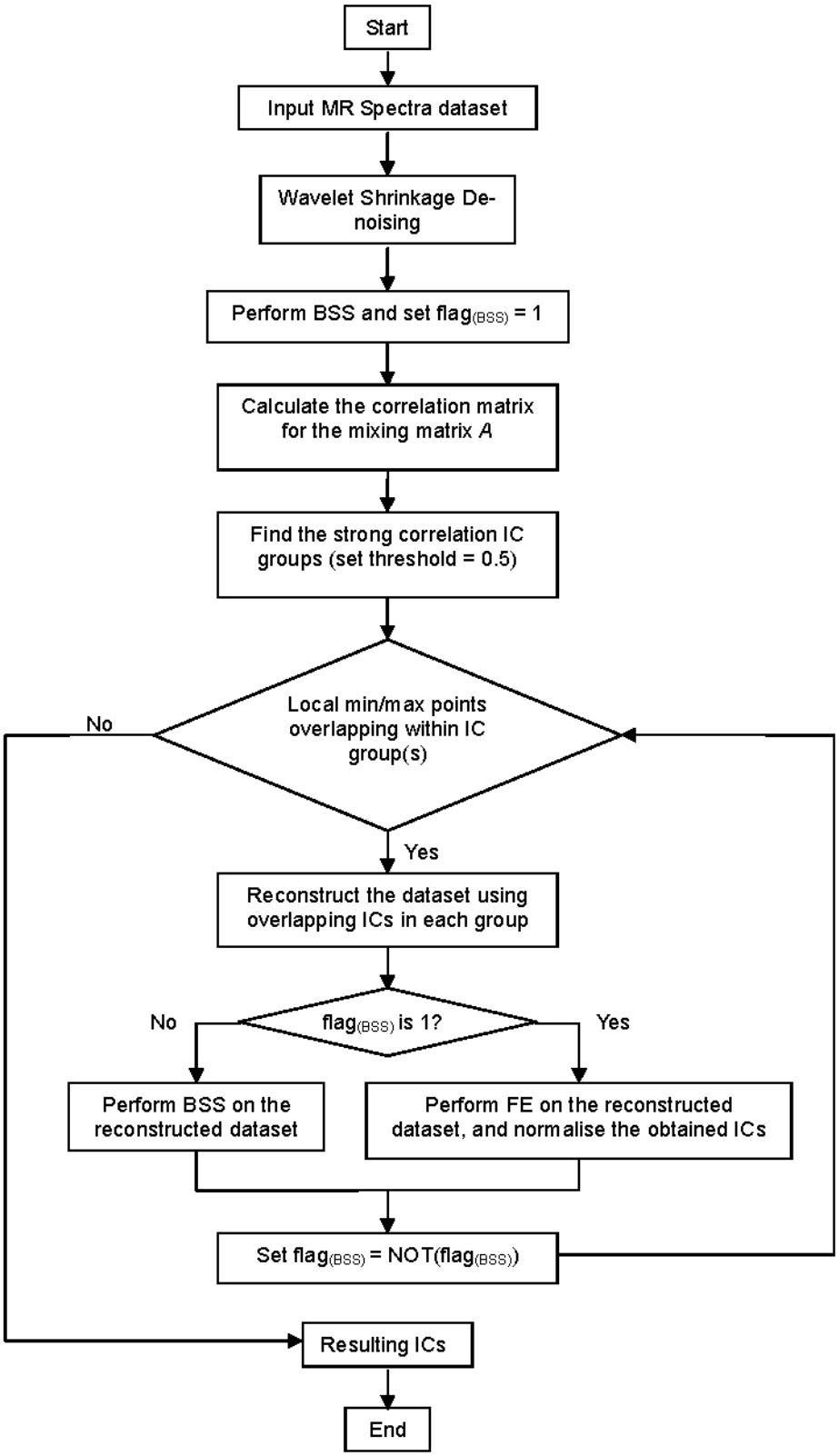


Figure 8.1: The hybrid ICA algorithm flow chart.

As the overlapping peaks produce distorted ICs in the results, they are identified by the following two steps.

First, the column vectors in the estimated mixing matrix $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$ are checked. Since $\mathbf{X}_{l\times d} = \mathbf{A}_{l\times m}^{\mathrm{BSS}}\,\mathbf{S}_{m\times d}^{\mathrm{BSS}}$, according to the ICA theory, if $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$ and $\mathbf{S}_{m\times d}^{\mathrm{BSS}}$ are the solution of the problem, then $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$ is a permuted version of $\mathbf{A}$. While if $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$ and $\mathbf{S}_{m\times d}^{\mathrm{BSS}}$ are not the solutions, some row vectors of $\mathbf{S}_{m\times d}^{\mathrm{BSS}}$ may still be the linear combinations of the original sources, and the corresponding column vectors of $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$ will be linear combinations of the column vectors of the true mixing matrix $\mathbf{A}$. In this case, the correlation between these column vectors of $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$ will be strong.

The correlation coefficient between two columns is defined as the covariance of the two columns divided by the product of their standard deviations as shown below:

$$R(\mathbf{a}_i, \mathbf{a}_j) = \frac{\mathrm{cov}(\mathbf{a}_i, \mathbf{a}_j)}{\sigma_i \sigma_j} \tag{8.3}$$

where $\mathbf{a}_i$ and $\mathbf{a}_j$ are the i^{th} and j^{th} column of matrix **A,** respectively. A threshold is set to find the strongly correlated columns of $\mathbf{A}_{l\times m}^{\mathrm{BSS}}$, hence the IC groups.

Next, within these groups, the ICs with overlapping local minimum and maximum points are identified. The differences between the adjacent points for each IC vector $[s_1, \ldots s_m]$ (row of $\mathbf{S}^{\mathrm{BSS}}$) are examined.

$$\mathbf{diff} = [s_1, ...s_{m-1}] - [s_2, ...s_m] \tag{8.4}$$

Locate the indices of the points where there is a change of sign in the vector **diff**. If it changes from positive to negative, the index corresponds to a local maximum point in the IC; otherwise, it is a local minimum point. To account for the noise factor in the ICs, we also check if the difference between adjacent local maximum and minimum is greater than certain threshold, here we set the threshold to be 1/3 of the difference between the global max/min points for each IC.

The subsets of $\mathbf{A}^{\text{BSS}}$ and $\mathbf{S}^{\text{BSS}}$ corresponding to the overlapping ICs $k_1, ...k_K$ are used to reconstruct a new dataset to the same dimension as the original one, as shown in the equation below:

$$\mathbf{X}^{\text{rec}}_{l\times d} = \mathbf{A}^{\text{BSS}}_{l\times K}\, \mathbf{S}^{\text{BSS}}_{K\times d} \tag{8.5}$$

where $\mathbf{X}^{\text{rec}}_{l\times d}$ is the reconstructed dataset from the $k_1, ...k_K$ overlapped ICs and their corresponding mixing coefficients.

After this step, the reconstructed dataset contains only the overlapping components. Then the FE-based ICA is performed on the reconstructed datasets from each of the IC groups. The FE-ICA algorithm is applied on $(\mathbf{X}^{\text{rec}}_{l\times d})^T$ to obtain $\mathbf{B}^{\text{FE}}_{K\times d}$, in this case, the dimension is reduced from l to K. The estimate of the original metabolite and MMLip components will be the column vectors of $\mathbf{A}^{\text{FE}}_{d\times K} = \text{pinv}(\mathbf{B}^{\text{FE}}_{K\times d})$. To make it more consistent with the result format from the BSS-ICA, the metabolite and MMLip components obtained from this step are normalised to have a standard deviation equal to 1, the same as in the BSS-ICA. These procedures are

repeated for each of the overlapping IC groups.

To check if the results from the FE-ICA are properly extracted, a further procedure of local min/max identification is executed. In most of the cases, the FE-ICA results will be properly extracted, and the algorithm will output its final results. However, in some complex cases, the resultant ICs from the FE-ICA still contain some multiple metabolites, and a further BSS-ICA method is needed to separate them.

8.3 Methods

8.3.1 Synthesised datasets

Various sets of synthesised MR spectra of three types of childhood brain tumours (TE = 30 ms, 1.5T and 3T) were generated using the basis sets 1 and 2 described in Chapter 5.

The mean and standard deviation of each metabolite and MMLip components concentrations were set to these observed in average *in vivo* MR spectra of the three classes of childhood cerebellar tumours (Davies et al., 2008), namely astrocytoma, ependymoma and medulloblastoma, as shown in Table 5.1.

The sum of correlation coefficients between the obtained ICs and the original individual metabolite and MMLip components was used as a measure of the quality of the resulting ICs for the synthesised datasets. IC correlated more closely to its original component will have a correlation coefficient closer to 1. Each of the synthesised results, presented in the next section, was the average of 200 repeated experiments and with a 95% confidence interval calculated by a bootstrap re-sampling method (as described in section 5.2.5).

8.3.2 Experimental dataset

The same 115 experimental spectra obtained from the two quality control criteria in Chapter 7 are used here. The correlation coefficients between the obtained ICs and the simulated metabolites from (Reynolds et al., 2006) or basis MMLip components from LCModel are used as a measure of the quality of the resulting ICs.

8.4. Results

8.4.1 Feature selection algorithm evaluation on synthesised data

The proposed SNR-based feature dimension selection algorithm was evaluated using the synthesised datasets (minimum $SNR_{LC} \geq 10$) described in section 5.3.1, and the experiments on synthesised datasets show the following:

1. If the kurtosis of the set of SNR_E for the dataset is negative or close to 0, then the average SNR_E for the dataset is used to calculate the percentage of the clean signal power over the noisy signal power in equation (5.6); therefore, the dimension to be reduced to is determined with an error of ± 3.
2. If the kurtosis of the set of SNR_E is much greater than 0, we can plot the histogram of the set of SNR_E, and exclude the outliers. Since kurtosis is very sensitive to outliers, its value may depend on only a few observations in the tails of the distribution. Removing the outliers is eventually reducing kurtosis to a smaller value. Then the maximum SNR_E is used in equation (5.6) as in the previous condition.

This method only gives an estimation of the starting point for dimension reduction, as the SNR_{LC} from LCModel fit is still an estimation of the true unknown SNR value.

8.4.2 Synthesised data basis set 1

Each synthesised dataset consists 60 spectra with an average $SNR_{LC} = 20.8$ and FWHM = 5.5 Hz. It was shown in the study by Hao et al. (Hao et al., 2009b) that the BSS-ICA method produces stable and realistic ICs with small and large datasets. Hence, small sized synthesised datasets (60 spectra) were used to make sure our experiments are robust. The ICA algorithm was applied on the wavelet de-noised dataset. One example set of the results from the BSS, FE and hybrid ICA methods are shown in Figure 8.2. It can be seen that the BSS-ICA method gives overlapping peaks in IC3, IC4 & IC6, and IC9 & IC10. Most of the FE-ICA method results have multiple metabolites or MMLip components in one IC. The hybrid ICA method restores all the overlapping ICs, which are distorted in the BSS-ICA method. The hybrid ICA results are very similar to the original individual simulated metabolites and MMLip components used. For the BSS, FE and hybrid ICA methods, the average sum of correlation coefficients (from 200 repeated experiments) with their 95% confidence intervals by bootstrapping are 8.89 (8.88 - 8.90), 7.67 (7.60 - 7.75) and 9.12 (9.06 - 9.18), respectively. The hybrid ICA method has the highest value among the three methods which indicates that the ICs from the hybrid ICA method are indeed more closely related to the original components in the basis set.

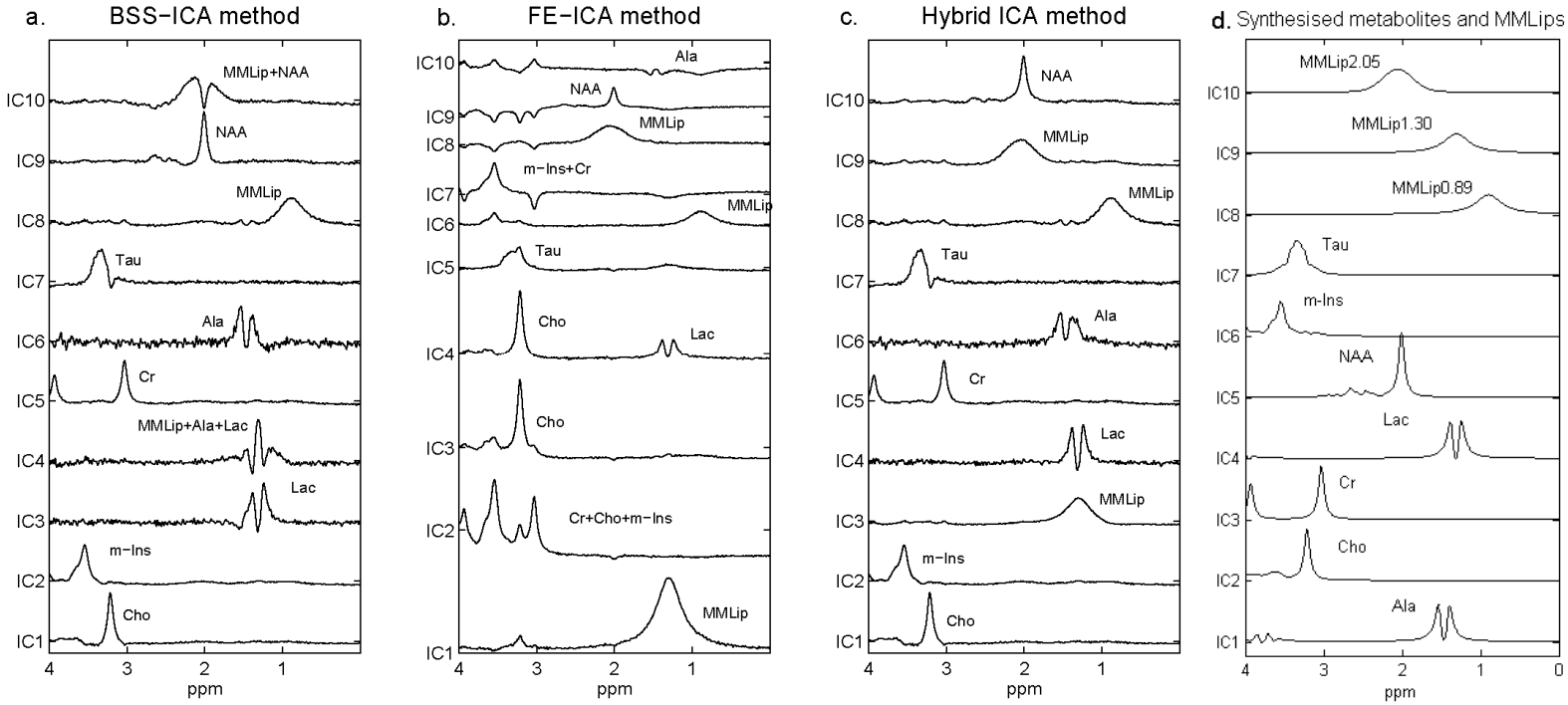


Figure 8.2: The 10 ICs obtained from 60 MRS synthesised dataset using (a) BSS, (b) FE and (c) Hybrid ICA methods labelled with the corresponding metabolite and MMLip components, (d) the synthesised metabolites and MMLips are plotted for comparison. It can be seen that the IC3 and IC9 from hybrid ICA results in (c) are improved compare to IC4 and IC10 from BSS-ICA results in (a) which correspond to IC9, MMLip1.30, and IC10, MMLip2.05 in (d). The extraction of individual metatoblites and MMLips by the FE-ICA (b) are not very successful as several combinations of metatoblites and MMLips present in most of the ICs.

8.4.3 Synthesised data basis set 2

Further experiments were performed on more complex synthesised datasets using basis set 2. The hybrid ICA method performs better than BSS-ICA and FE-ICA methods again. Similar as in the last section, the average sum of correlation coefficients for the BSS, FE and hybrid ICA methods with their 95% confidence intervals are 10.61 (10.59-10.63), 10.35 (10.27-10.44) and 11.63 (11.58-11.67), respectively. One example result is shown in Figure 8.3, and the correlation coefficients between all the resultant ICs and the simulated components are shown in Table 8.1 for the hybrid ICA method. The ICs are assigned to their most correlated metabolites or MMLip components (highlighted in grey colour). The sum of their correlation coefficients is 11.61, which lies in the 95% confidence interval for the hybrid ICA method. The corresponding concentration for each IC also has the strongest correlation with its simulated value, which confirms with the findings in Table 8.1.

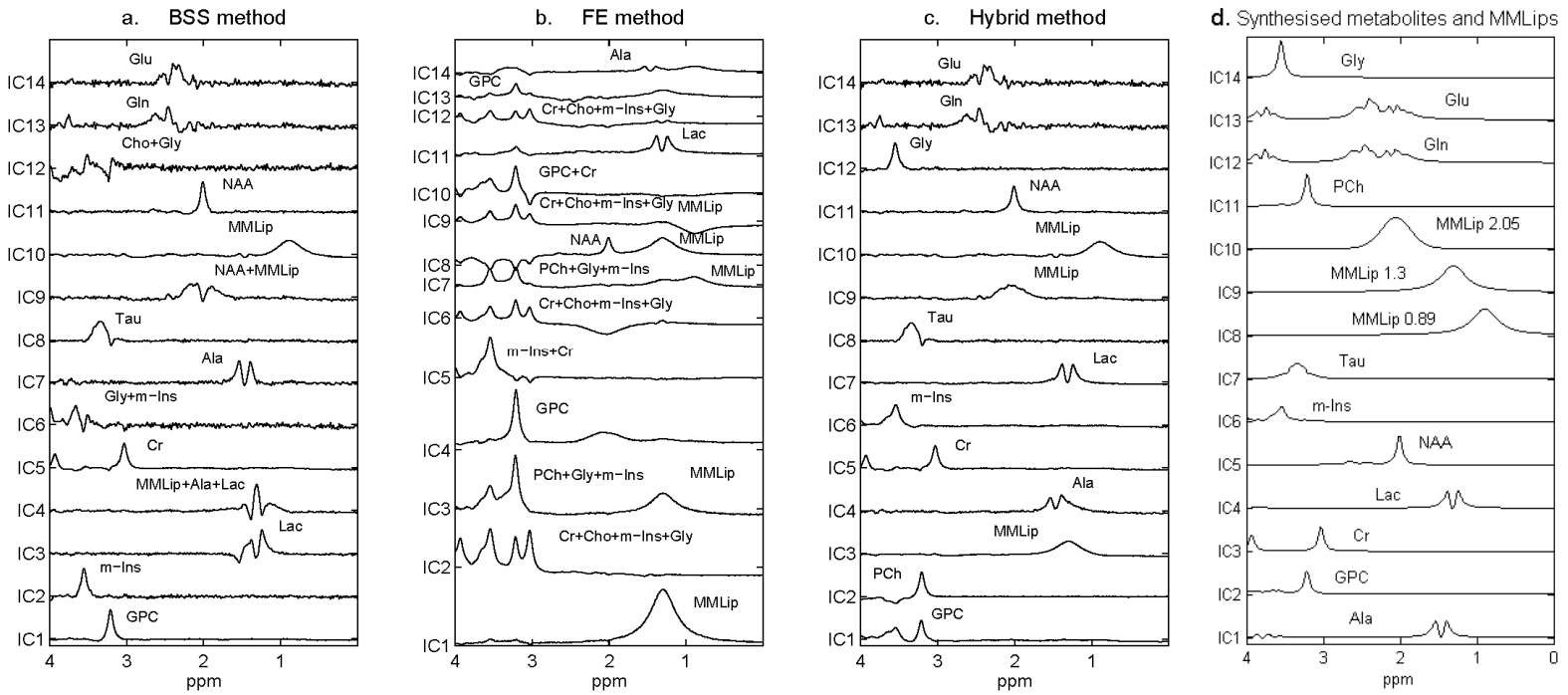


Figure 8.3: The 14 ICs obtained from 60 MRS synthesised dataset using (a) BSS, (b) FE and (c) Hybrid ICA methods labelled with the corresponding metabolite and MMLip components, (d) the synthesised metabolites and MMLips are plotted for comparison. Similar as in Figure 8.2, the ICs corresponding to overlapping metabolites or MMLips are better extracted (c) than in (a).

Table 8.1: The correlation coefficients between the resultant ICs and the simulated metabolite and MMLip components for synthesised data basis set 2.

Components	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	IC9	IC10	IC11	IC12	IC13	IC14
Ala	0.01	0.06	0.61	0.91	0.06	0.02	0.60	0.04	0.04	0.06	0.04	0.01	0.06	0.05
GPC	0.83	0.75	0.08	0.08	0.07	0.20	0.07	0.20	0.08	0.07	0.06	0.12	0.02	0.04
Cr	0.06	0.01	0.09	0.08	0.97	0.00	0.07	0.03	0.09	0.08	0.06	0.05	0.03	0.06
Lac	0.07	0.02	0.81	0.73	0.05	0.05	0.99	0.06	0.06	0.02	0.05	0.04	0.05	0.05
NAA	0.07	0.00	0.07	0.06	0.02	0.06	0.06	0.05	0.73	0.07	0.98	0.04	0.11	0.04
m-Ins	0.70	0.32	0.10	0.09	0.06	0.97	0.08	0.18	0.09	0.08	0.07	0.85	0.01	0.05
Tau	0.36	0.24	0.08	0.08	0.06	0.18	0.08	0.90	0.08	0.08	0.06	0.18	0.03	0.05
MMLip0.9	0.18	0.02	0.10	0.04	0.12	0.14	0.08	0.12	0.15	0.93	0.10	0.09	0.13	0.11
MMLip1.3	0.15	0.02	0.96	0.83	0.10	0.12	0.82	0.10	0.05	0.13	0.04	0.07	0.09	0.07
MMLip2.05	0.12	0.01	0.09	0.08	0.08	0.10	0.08	0.08	0.96	0.10	0.70	0.06	0.08	0.15
PCh	0.73	0.87	0.06	0.06	0.05	0.06	0.06	0.18	0.06	0.05	0.04	0.06	0.02	0.03
Gln	0.02	0.12	0.16	0.14	0.03	0.05	0.14	0.11	0.51	0.19	0.39	0.02	0.64	0.43
Glu	0.02	0.12	0.15	0.12	0.06	0.05	0.13	0.10	0.51	0.17	0.38	0.01	0.38	0.69
Gly	0.60	0.36	0.06	0.06	0.02	0.95	0.05	0.09	0.05	0.04	0.04	0.97	0.01	0.03

8.4.4 Experimental dataset

The kurtosis of the 115 SNR_E of the experimental dataset is 25.6. The histogram of the SNR_E is plotted and shown in Figure 8.4. The outliers are the three very high valued SNR_E near the right end of the plot. The dimension should be reduced to approximately 29 (±3) according to equation (5.6). The experimental dataset dimension is reduced to 28 as no more distinguishable new metabolite or MMLip components are shown after this value.

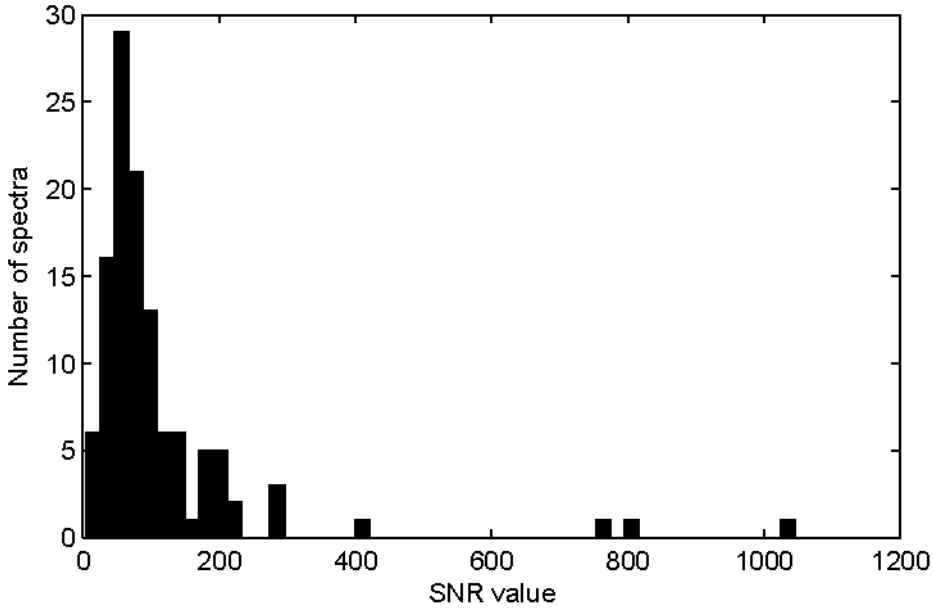


Figure 8.4: The histogram of SNR_E for the experimental dataset.

Figure 8.5 shows the resultant ICs from the set of 115 *in vivo* experimental MR spectra. The ICs are assigned to the metabolite or MMLip with the largest correlation. The sum of correlation coefficients between the resultant ICs from the experimental dataset and the simulated metabolite and MMLip components for BSS, FE and hybrid ICA methods are 17.12, 21.11 and 22.83, respectively.

For the hybrid ICA method, the correlation coefficients between the 28 ICs and the simulated components are evaluated and their three highest values are shown in Table 8.2. The ICs are assigned to their most correlated metabolite and MMLip component. The simulated metabolites are generated from the same simulator used for the synthesised dataset in this research work, and the simulated MMLip components are obtained from the LCModel basis spectra parameters (Provencher, 2009). Most of the ICs obtained from the hybrid ICA method have a strong correlation with the simulated metabolite and MMLip components.

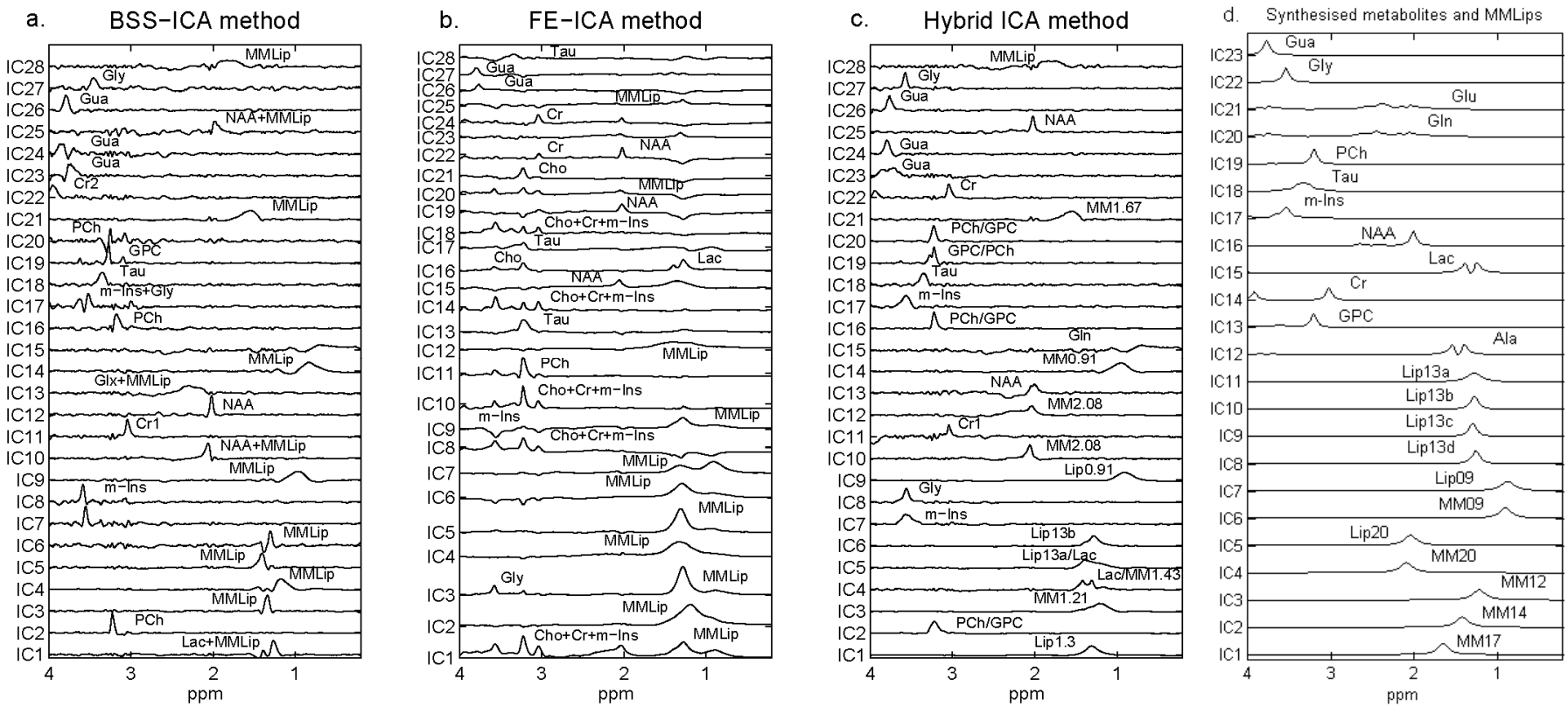


Figure 8.5: The 28 ICs obtained from 115 experimental MRS experimental dataset using (a) BSS, (b) FE and (c) Hybrid ICA methods labelled with the corresponding metabolite and MMLip components, (d) the synthesised metabolites and MMLips are plotted for comparison. Similar to the synthesised cases, in (c), the extraction of overlapping metabolites and MMLips, such as Gly, m-Ins, Cr Lac MM2.08 and MMLip at around 1.30ppm, are improved compare to the ICs in (a).

Table 8.2: The correlation coefficients between the resultant ICs and the simulated metabolite and MMLip components for the experimental dataset. The highest three correlation coefficients are shown in this table, where the first highest correlation coefficient between the metabolites and MMLips and the resultant ICs are very close to 1 in most of the cases.

	Three highest correlation coefficients (in descending order)		
Ala	IC21 (0.58)	IC5 (0.55)	IC4 (0.49)
GPC	IC2 (0.93)	IC16 (0.86)	IC20 (0.84)
Cr	IC22 (0.89)	IC11 (0.17)	IC5 (0.11)
Lac	IC5 (0.84)	IC3 (0.75)	IC6 (0.70)
NAA	IC25 (0.83)	IC12 (0.67)	IC13 (0.64)
m-Ins	IC7 (0.90)	IC8 (0.85)	IC17 (0.82)
Tau	IC18 (0.72)	IC19 (0.45)	IC2 (0.45)
PCh	IC2 (0.96)	IC16 (0.88)	IC20 (0.84)
Gln	IC15 (0.52)	IC12 (0.44)	IC25 (0.35)
Glu	IC12 (0.55)	IC15 (0.44)	IC25 (0.38)
Gly	IC7 (0.94)	IC8 (0.93)	IC17 (0.88)
Gua	IC26 (0.87)	IC24 (0.84)	IC23 (0.76)
Lip13a	IC6 (0.99)	IC1 (0.98)	IC5 (0.91)
Lip13b	IC6 (0.99)	IC1 (0.97)	IC3 (0.82)
Lip13c	IC1 (0.98)	IC6 (0.98)	IC5 (0.85)
Lip13d	IC6 (0.98)	IC1 (0.92)	IC3 (0.88)
Lip09	IC9 (0.97)	IC14 (0.87)	IC15 (0.18)
MM09	IC9 (0.97)	IC14 (0.92)	IC15 (0.09)
Lip20	IC12 (0.85)	IC10 (0.77)	IC25 (0.77)
MM20	IC12 (0.86)	IC10 (0.82)	IC25 (0.65)
MM12	IC3 (0.97)	IC6 (0.84)	IC1 (0.79)
MM14	IC4 (0.83)	IC5 (0.83)	IC1 (0.69)
MM17	IC21 (0.75)	IC28 (0.62)	IC2 (0.10)

8.5 Discussion

The hybrid ICA method was applied to both synthesised and experimental MRS datasets to decompose them into individual metabolite and MMLip components. The results show that the hybrid ICA method demonstrates the advantages of both BSS and FE-based ICA methods, while it overcomes their individual limitations.

For the BSS-ICA method, the overlapping mostly occurs when one or more narrow peaks lie in the same range with a broad peak. It can be easily shown that in this case, the metabolite and MMLip components are not statistically independent from each other by calculating the correlation coefficient between them. As shown in our synthesised data, the absolute value of correlation coefficients between the overlapped components are all greater than 0.58 (at FWHM = 5.5 Hz), which is not a weak correlation and, therefore, not statistically independent from each other. The correlation coefficients between the others are much smaller, which can be considered as weak correlation. Although independent variables are uncorrelated, uncorrelatedness does not imply independence. Here we are not trying to prove the statistical independence of those components, but to show the non-independence of the overlapping components.

The fundamental difference between the BSS-ICA and FE-ICA methods is their assumptions on statistical independence. In contrast with the BSS-ICA method, where the metabolite and MMLip components are treated as statistically independent from each other, the FE-ICA method assumes the independency of the concentration coefficients of different metabolite and MMLip components. Experiments imply that the concentrations (linear mixing) of these overlapping components are independent, or at least are more independent than the overlapping components themselves. This may explain why these overlapping components

can be better extracted from their mixture spectra by FE-ICA method.

The BSS-ICA is placed as the first step of the hybrid ICA algorithm, because it works under a smaller size of dataset. This makes the hybrid ICA method more efficient, since collecting a large dataset of MR spectra may not always be possible. We have also shown by Hao et al. (Hao et al., 2009b) that BSS-ICA method generally produces ICs more correlated to the original signals than the FE-ICA method. Another reason to use the BSS-ICA first instead of FE-ICA is that the FE-ICA result may be dominated by a few very high peak components, like the MMLip components, and combinations of several components may present throughout the resultant ICs. This will make the next stage analysis complex in terms of information to be included, and number of new ICs to be determined.

The size of the matrix $\mathbf{A}^{BSS}$ is usually much smaller than matrix $\mathbf{S}^{BSS}$, so examining $\mathbf{A}^{BSS}$ first for the overlapping ICs identification should reduce the choices of potential overlapping ICs to a smaller one. Hence, there is less computational complexity compared with examining $\mathbf{S}^{BSS}$ directly for identification.

For the synthesised datasets, the hybrid ICA method restores the same number of metabolite and MMLip components as the original synthesised components. Their mixing matrix has a very high correlation with the original one. In the synthesised data basis set 1 experiment, the problem with BSS-ICA method is extracting the overlapping components of NAA & MMLip at 2.05 ppm, and Ala, Lac & MMLip at 1.30 ppm. The FE-ICA method cannot separate Cr, Cho & m-Ins and Cho & Lac. These problems were mostly solved by the hybrid ICA method and most of the simulated basis components were restored.

The experiments on synthesised data using basis set 2 deal with a greater degree of complexity since pairs of metabolites are included, which have a similar lineshape and appear

at a similar frequency, notably PCh & GPC, Gly & m-Ins and Gln & Glu. From Figure 8.3, the hybrid ICA method produces ICs corresponding to all of the metabolite and MMLip signals. In particular, there are two ICs that correspond to each of the pairs of overlapping metabolites showing that the hybrid ICA method can detect the presence of more than one IC from signals, which have a large degree of spectral overlap. This is particularly impressive for PCh and GPC, since not only are their MRS signals very similar but the amount of PCh is small in astrocytoma and ependymoma, reducing the effective number of cases. From Table 8.1, the ICs, which correspond to the strongly overlapped metabolites, show large correlations with each of the two signals, however, this is as to be expected since the signals overlap considerably.

While the resultant ICs from the hybrid ICA method are normalised to have standard deviation equal to 1, the original mixing matrix can be easily recovered by simply dividing the mixing matrix by the true standard deviation of the ICs from theoretical values. For IC13 and IC14, they are assigned to Gln and Glu, respectively. Although, the correlation coefficients between these metabolites and their concentrations indicate that it is the best choice for assignment, their correlations are not as strong as for the rest of the ICs. This could be due to their very similar spectra shapes in low magnetic fields and the presence of noise.

We have chosen not to use guanidoacetate (Gua) in the synthesised dataset experiments, as this is not one of the most commonly quantitated metabolites *in vivo*. However, our algorithm's result is originating from the experimental data clearly identifies Gua as an independent component and verifies its existence. Therefore, we have generated a simulated Gua quantity for identification and comparison purposes within the experimental dataset.

For the experimental dataset, similar results were obtained to those in the synthesised experiments. The hybrid ICA method picks up most of the metabolite and MMLip

components within 28 ICs, whereas the BSS-ICA and FE-ICA suffer from the same problems, as revealed by the synthesised datasets. For example, the BSS-ICA results, IC17 in Figure 8.5 is very similar to the synthesised IC8 in Figure 8.3, indicating that the m-Ins IC was distorted by Gly. In Figure 8.5(c), the NAA and MMLip peaks at 2.05 ppm are restored. Several MMLip components are present in the range of 1 ppm to 1.5 ppm, with little trace of Lac in IC4 & IC5. The IC7, IC8, IC17 & IC27 are strongly correlated to Gly and m-Ins, but more to Gly in IC8 & IC27. In IC22, the two Cr peaks are shown and better than in the BSS-ICA and FE-ICA method. It showed that Gly and m-Ins were both present, and Glu & Gln could be detected. A metabolite assigned to Gua does exist and needs assigning.

For the hybrid ICA results, some ICs are very similar. This may be accounted for by the presence of metabolites with similar MRS signals such as those used in simulated bases set 2, e.g. PCh and GPC. It may also reveal the presence of other metabolites not hitherto discussed such as choline and phosphatidyl choline. The area around 3.8 ppm is particularly interesting since the assignment of a singlet in this region to Gua is controversial in tumour spectra. However, one IC corresponds to a signal, which appears as a singlet in this region but a second similar signal appears at a slightly different ppm and a more complex signal also occurs in the same region. The hybrid ICA method, therefore, starts to give some insight into complex regions of the MRS of childhood brain tumours, which has not been available previously.

Although the hybrid ICA method can reveal the presence of multiple metabolite signals, which strongly overlap, some ICs may also represent variability in the data from the known metabolites. This effect may be caused by differences in FWHM of these peaks or by referencing the spectra to the Cr peak at around 3.2 ppm or to the NAA peak at 2.01 ppm when the Cr peak is low or absent in the experimental dataset. As the experimental MR

spectra are collected under different pH and temperature environments, there might be slight variations of the metabolite and MMLip components peak positions. Increasing the number of ICs will result in more repeated components with frequency shifting and FWHM difference.

Additional confidence in the biological validity of the ICs may be obtained by comparing the various methods. Since there are several overlapping components in the ICs, the BSS-ICA results are mostly distorted. Therefore, the sum of correlation coefficients is higher for the FE-ICA method than for the BSS-ICA method. However, the Gua is successfully restored in both BSS-ICA and FE-ICA methods, suggesting that Gua is independent from the other metabolite and MMLip components in both spectrum shape and its concentration.

Another approach of estimating the dimension is examined, and the results are shown in Table 8.3. Each of the estimated dimensions for synthesised datasets shown in Table 8.3 is an average from 100 repeated experiments on the synthesised datasets (with minimum $SNR_{LC} \geq 10$ and varying mean SNRs). Four approaches of choosing the SNR values for equation (5.6) are examined in the experiment, and it shows that the mean SNR of a dataset gives closer estimate of the true dimension in all the conditions. The kurtosis is calculated as an indication of the distribution of the dataset's SNRs. The proposed method gives more accurate estimate for Gaussian distributed SNRs than for non-Gaussian (sub- and super-Gaussian) distributed ones. The estimated dimension for dataset with non-Gaussian distributed SNRs is generally smaller than its true value, and the difference between the estimated dimension and the true dimension tends to grow slowly with larger dimension values. As shown in Table 8.3, this difference is less than one for 10ICs, and is about 2 for 14ICs.

The last row of Table 8.3 gives the estimated dimensions using the experimental dataset. The

SNRs in patient dataset has a non-Gaussian distribution. The proposed method gives an estimated dimension of 15 by using the mean SNR. Based on the findings from synthesised data experiments, the estimated dimension should be smaller than its true dimension and the difference between them should be around 3. So the dimension to be reduced for patient dataset should be around 18.

Table 8.3: The estimated dimensions by the proposed feature dimension selection method for synthesised datasets with various SNR distributions.

	SNRs distributions	Kurtosis of SNRs	Estimated dimensions of the synthesised datasets from			
			Maximum SNR	Minimum SNR	Mean SNR	Median SNRs
Basis set 1 (10 metabolite and MMLip components)	SNRs with a sub-Gaussian distribution	-1.33	18.55	3.95	9.01	9.17
	SNRs with a Gaussian distribution	0.67	31.81	4.02	10.22	8.98
	SNRs with a super-Gaussian distribution	23.90	39.93	2.00	10.06	7.97
Basis set 2 (14 metabolite and MMLip components)	SNRs with a sub-Gaussian distribution	-1	27.25	6.00	12.38	12.33
	SNRs with a Gaussian distribution	0.21	34.20	4.08	14.44	10.17
	SNRs with a super-Gaussian distribution	23.85	42.22	2.00	12.52	9.70
Experimental dataset	SNRs with a super-Gaussian distribution	23.60	31	3	15	13

Figure 8.6 shows the resultant ICs from the set of 115 in *vivo* experimental MR spectra. The dimension reductions for values around 18 are also examined. For the hybrid ICA method, the difference between dimension 15 and 16 is IC18, Cr peak at 3.2ppm, in Figure 8.6(c). IC9 Lac first time appears in dimension 17. IC12 Gly is added in for dimension 18 but not for dimension 17. An IC very similar to IC18 is repeated in dimension 19. In dimension 20, a flat and broad IC appears.

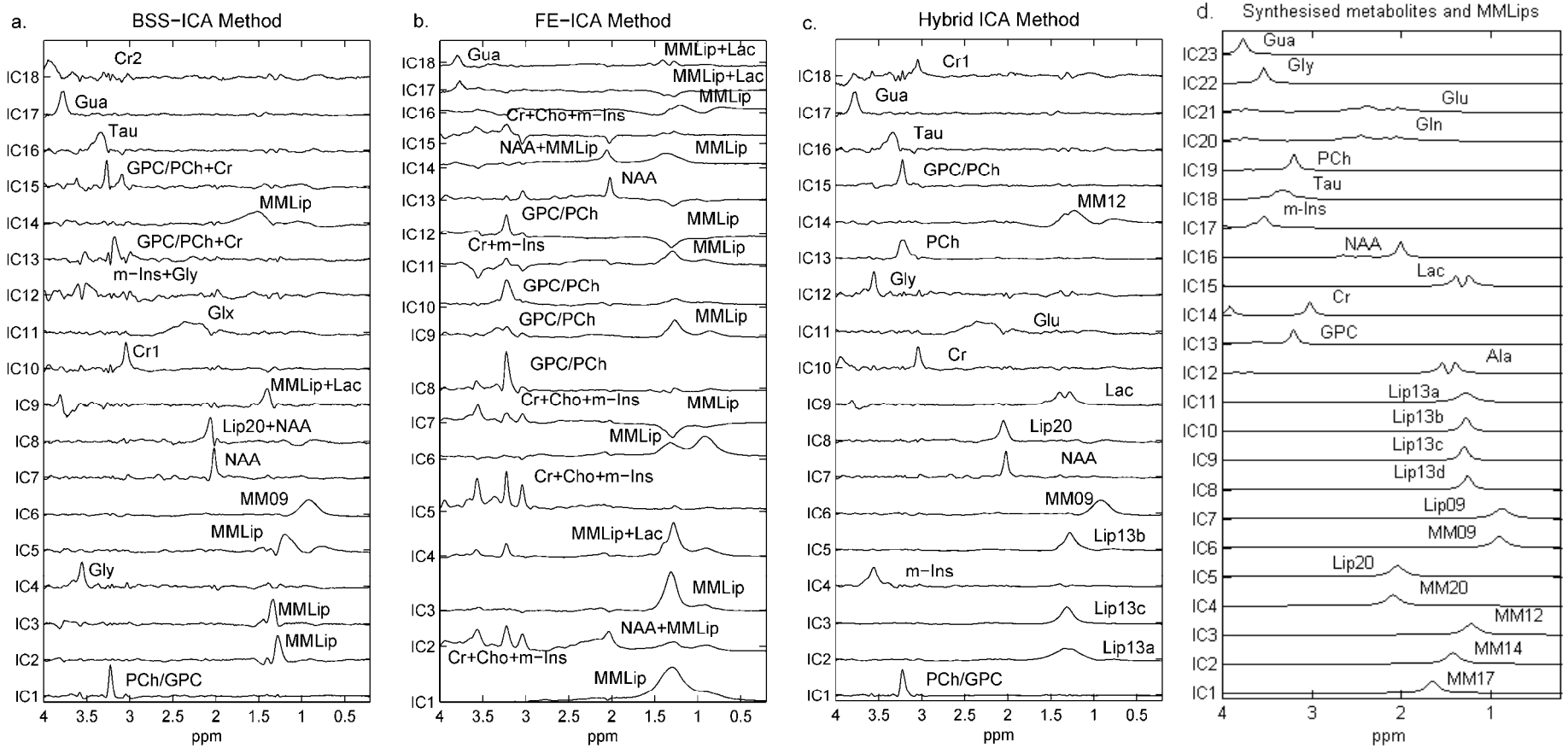


Figure 8.6: The 18 ICs obtained from 115 MR spectra experimental dataset using (a) BSS, (b) FE and (c) Hybrid ICA methods labelled with the corresponding metabolites and MMLip components, (d) the synthesised metabolites and MMLips are plotted for comparison. By choosing a smaller dimension, the problem of repeating metabolite ICs appeared in Figure 8.5(c) are eliminated. The improvement of hybrid ICA (c) when dealing with overlapping metabolites and MMLips are obvious compared to (a).

The sum of correlation coefficients between the resultant ICs from the experimental dataset and the simulated metabolites and MMLip components for BSS, FE and hybrid ICA methods are 12.96, 13.25 and 15.60, respectively. For the hybrid ICA method, the correlation coefficients between the 18 ICs and the simulated components are evaluated and their three highest values are listed in Table 8.4.

It can be seen that for this approach, the new reduced dimension of 18 gives less repeated or similar shaped ICs compared with reducing the dimension to 28. However, actually existence of those missing ICs are uncertain, and this will be investigated in future study.

Table 8.4: The correlation coefficients between the resultant ICs and the simulated metabolites and MMLip components for the experimental dataset.

	Three highest correlation coefficients (in descending order)		
Ala	IC9 (0.44)	IC2 (0.43)	IC3 (0.34)
GPC	IC15 (0.95)	IC1 (0.95)	IC13 (0.89)
Cr	IC10 (0.93)	IC18 (0.22)	IC2 (0.06)
Lac	IC2 (0.94)	IC9 (0.91)	IC3 (0.79)
NAA	IC7 (0.96)	IC8 (0.66)	IC2 (0.07)
m-Ins	IC4 (0.92)	IC12 (0.81)	IC9 (0.15)
Tau	IC16 (0.80)	IC13 (0.37)	IC15 (0.28)
PCh	IC15 (0.96)	IC1 (0.96)	IC13 (0.91)
Gln	IC2 (0.94)	IC9 (0.91)	IC3 (0.79)
Glu	IC11 (0.75)	IC8 (0.32)	IC7 (0.28)
Gly	IC4 (0.90)	IC12 (0.87)	IC9 (0.07)
Gua	IC17 (0.96)	IC18 (0.09)	IC2 (0.07)
Lip13a	IC5 (0.97)	IC3 (0.96)	IC2 (0.96)
Lip13b	IC5 (0.98)	IC3 (0.96)	IC2 (0.87)
Lip13c	IC3 (0.99)	IC5 (0.96)	IC2 (0.88)
Lip13d	IC5 (0.98)	IC3 (0.89)	IC2 (0.85)
Lip09	IC6 (0.95)	IC14 (0.17)	IC11 (0.15)
MM09	IC6 (0.95)	IC14 (0.12)	IC11 (0.12)
Lip20	IC8 (0.88)	IC7 (0.75)	IC11 (0.31)
MM20	IC8 (0.86)	IC7 (0.57)	IC11 (0.38)
MM12	IC5 (0.85)	IC2 (0.84)	IC14 (0.82)
MM14	IC2 (0.83)	IC9 (0.79)	IC3 (0.71)
MM17	IC14 (0.37)	IC2 (0.18)	IC4 (0.14)

8.6 Summary

The proposed hybrid ICA method combines the advantages of the BSS and FE-based ICA methods, whilst at the same time overcoming their disadvantages when dealing with MR spectra. The results demonstrate that the hybrid ICA method provided more realistic individual metabolites and MMLip components than the BSS-ICA and FE-ICA method alone for both synthesised and experimental datasets. It can aid metabolite identification and assignment. The hybrid ICA method has the potential for extracting biologically useful features and discovering biomarkers. However, the identification of metabolites and MMLip components with very similar peak positions and lineshapes, such as PCh and GPC, or Glu and Gln, may be confusing in some cases, and this needs further investigation. The mixing matrix, which is closely related to concentrations of the metabolites and MMLip components, could be used to construct classifiers. These will be investigated in further studies.

CHAPTER 9

CONCLUSION

9.1 Summary

This thesis presents a study of applying ICA on *in vivo* MR spectra of childhood brain tumours. To fully explore the ability of ICA on this type of data, four objectives have been set at the beginning of this study.

As the *in vivo* MR spectra are often characterized by low SNRs, a de-noising step is needed before being processed by ICA. So the first objective of this thesis was to enhance the MR spectra before decomposition. The conventional de-noising process usually only focuses on the real part of the MR spectrum and ignores the imaginary part. A novel WSD-based MRS enhancement algorithm, which utilises both the real and imaginary parts of a spectrum, has been proposed in this thesis. A relationship between the real and imaginary parts of the MR spectrum is derived, which improves the de-noising result by adding the useful information contained in the imaginary part of the spectrum. The de-noised spectra SNRs obtained by this algorithm are compared with the results from the conventional methods, which apply WSD on FID signal and on the real part of the MR spectrum. The results show that the new

approach can perform 1.46 times better than de-noising solely on the real part of the spectrum and 1.70 times better than on the FID signals.

The common approach followed in choosing dimensionality reduction in ICA is by trial and error. An accurate estimate of the dimension could aid the next stage ICA decomposition. To help make the ICA application more sensible to our MRS dataset, the second objective of the thesis was to find a dimension reduction guideline in the context of MRS. We have proposed a feature dimension reduction method based on the SNR values of the MRS dataset. The percentage of the information contained in the dataset is estimated from the overall SNR. This can be used as a guideline for the first stage ICA feature dimension reduction. Experiments on synthesised data confirmed the usefulness of this method.

ICA has the potential to automatically decompose a set of MR spectra into the individual metabolite and MMLip components. Most of current studies on the ICA application to MRS were ambiguous about how ICA works on MRS data. The third objective of this thesis was to study the ability of ICA on the MRS dataset. A systematic comparison between the two approaches of ICA, the BSS and FE methods, on synthesised *in vivo* MR spectra of childhood brain tumour datasets was performed. It has been found that applying ICA on simulated data could extract individual metabolite and MMLip components from a set of MRS data. The BSS-ICA method generates ICs with a closer correlation to the original metabolite and MMLip components than the FE-ICA method when the number of spectra in the dataset is small. The optimal conditions for a reliable and repeatable experiment under either method can be obtained with a minimum number of 300 spectra and SNR values greater than or equal to 10. The FE-ICA method is relatively insensitive to different ranges of FWHM (from 0 to 3 Hz), whereas the BSS-ICA method degrades on increasing the range of FWHM. The peak frequency variations do not affect the results within the range of ±0.08 ppm for the FE

method, and ±0.05 ppm for the BSS-ICA method. The FE-ICA method is more robust and more effective on metabolite and MMLips spectra containing a wide range of frequencies, whereas the BSS-ICA method is more effective on spectra with frequencies over a narrow range. The FE-ICA method is limited in that it performs very poorly when a combination of metabolite and MMLip components appears in the same IC, and a large sample size is required. Most of the problems with the BSS-ICA method are caused by the overlapping peaks and the existence of high levels of noise. Similar results are obtained when the BSS-ICA and FE-ICA methods are applied to a quality controlled *in vivo* MRS of experimental dataset.

By achieving the third objective, the abilities of BSS-ICA and FE-ICA on in vivo MRS dataset are revealed for the first time in the MRS research. The fourth objective of this thesis is to improve the ICA method in the automatic decomposition of the MRS dataset into meaningful metabolite and MMLip components. A novel ICA approach involving a hybrid of BSS and FE techniques for automated decomposition of a series of MR spectra has been proposed. It combines the advantages of both BSS-ICA and FE-ICA methods. Experiments were performed on synthesised and experimental in vivo MR spectra of childhood brain tumour datasets. It has been demonstrated that the proposed hybrid ICA method provided more realistic individual metabolite and MMLip components than the BSS-ICA and FE-ICA method alone for both synthesised and experimental datasets. It can aid metabolite identification and assignment. The hybrid ICA method has the potential for extracting biologically useful features and discovering biomarkers.

The techniques developed in this thesis have been used for the clinical research of Childhood brain tumour cases from Birmingham Children's Hospital and has shown some interesting results. For example, it provides further evidence for the existence of a singlet peak at 3.8

ppm in childhood brain tumours, which has tentatively been assigned to Gua in the literature (e.g. (Panigrahy et al., 2006)). The assignment of this peak requires confirmation and its significance warrants further investigation.

9.2 Future research

As the imaginary part of the spectrum is just at a different phase stage ($-\pi/2$) with the real part of the spectra, it should contain the same amount of information as the real part, but with different noise details. Using only the real part of the MR spectrum for any analysis may cause bias with a high level of noise. If we could also include the imaginary part of the MR spectrum in the same process, and compare it with the real part using their relationship derived in this thesis, it may reduce the possibility of bias by unknown effects. In this thesis, we only used this relationship to reduce the bias in de-noising, but there could be many other situations in analysing MRS when it might be useful. For example, in the hybrid ICA experiments, only the real parts of the MR spectra were examined. It could also be applied on both the real and imaginary parts of the spectra, and further improvement may be achieved by comparing the obtained ICs from both parts.

If the MR spectra for various tumour types are large enough, the mixing matrix obtained for the hybrid ICA method, which is closely related to concentrations of the metabolite and MMLip components, could be used to construct the brain tumour classifiers, and possibly, reveal new biomarkers. However, the identification of metabolite and MMLip components with very similar peak positions and lineshapes, such as PCh and GPC, or Glu and Gln, may be confusing in some cases, and this needs further investigation.

In the case where metabolites are under represented in the spectrum, they could still be extracted. One possible method is to remove all the extracted ICs from each MR spectrum

and check if there is still information left by visual inspection or by checking the non-Guassianity of the remaining spectrum, and perform accordingly. However, if the noise level is high, there is the possibility that the under represented metabolite will be buried in the noise. In that case, a better de-noising method or acquisition procedure is needed.

The fundamental difference between the traditional ICA method and the proposed hybrid ICA algorithm is in the assumption of statistical independence. The traditional ICA method assumes the statistical independence strictly in the original sources for the BSS approach and in the extracted features for the FE approach. However, the hybrid ICA method looses this assumption slightly, where the independence can be either in the original sources or in its mixing matrix. Hence, it obtains a hybrid of independent components. It has been confirmed that for the MRS data, the slightly loosed assumption is a closer fit to the true model of the MRS data. This can be explained as that for the overlapping (strongly correlated) metabolite and MMLip components, their mixing coefficients (concentrations) are more independent from each other than their individual spectra lineshapes.

In the general context of digital signal processing, there may be other types of signals, which fit better to the loosed independence assumptions. The hybrid ICA method may improve the decomposition for these types of signals compared with the traditional ICA algorithms.

Magnetic resonance spectroscopic imaging (MRSI) is a powerful technique for metabolic imaging. MRSI allows the detection of metabolic profiles from multiple spatial positions (de Graaf, 2007). The multi-voxel spectra make MRSI a perfect candidate for the hybrid ICA method. The correlated spectra in adjacent voxels should be an advantage compared to single voxel MR spectra. The decomposed ICs and their corresponding concentrations could be mapped onto the corresponding patient MRI to aid the diagnosis and prognosis for tumours. With the improved automated decomposition ability of the hybrid ICA method, there is a

possibility to reveal further hidden information in the MRSI data.

APPENDIX A

QUANTUM MECHANICAL SIMULATOR

A.1 Method

The discrete-time acquired free induction decay signal of N samples, $\mathbf{y} \in \mathbf{C}^N$, can be modelled as a linear combination of the signals from various metabolites (the basis set) and some additive noise. Writing the time-series signals from the M metabolites as the columns of a matrix $\mathbf{S} \in \mathbf{C}^{N \times M}$ and the corresponding amplitudes (and phases) as elements of a vector $\mathbf{a} \in \mathbf{C}^M$:

$$\mathbf{y} = \mathbf{S}\mathbf{a} + \mathbf{w} \tag{A.1}$$

where $\mathbf{w} \in \mathbf{C}^N$ is a vector representing noise and modelling error.

A.2 Basis set simulation

The quantum mechanical simulation of ^{1}H NMR signals from small molecules has been extensively described by Levitt et al. (Levitt, 2001). Detailed *a priori* knowledge of

metabolite chemistry makes accurate simulation possible, as shown by Govindaraju et al (Govindaraju et al., 2000). An implementation of the method in the study by Levitt et al. (Levitt, 2001) was developed by Reynolds et al. (Reynolds et al., 2006). Quantum mechanical operators are applied to the density matrix describing the spin system, in order to simulate the time domain signal of each metabolite. Each NMR experiment is modelled by the sequential application of the following operators: thermal equilibrium, RF-pulse, free-evolution and acquisition. For a simple proton one-dimensional homonuclear experiment it is assumed that the spin system begins in a state of thermal equilibrium represented by the thermal equilibrium operator. Pulse sequences are modelled by sequential application of the RF-pulse and free-evolution operators. The experiment is completed by applying the acquisition operator, ultimately producing sets of frequencies, Ω, and amplitudes.

The Hamiltonian directly dictates the free evolution dynamics of a spin system. For a motionally averaged homonuclear spin system, the Hamiltonian can be accurately represented by considering the chemical shift interaction and the J-couplings between spins in the same molecule:

$$\hat{H}^0 = \sum_{j=1}^{n} \Omega_j \hat{\mathbf{I}}_{jz} + \sum_{\forall j<k}^{n} 2\pi J_{jk} \hat{\mathbf{I}}_j \cdot \hat{\mathbf{I}}_{jz} \tag{A.2}$$

where:

$$\hat{\mathbf{I}}_j = \hat{\mathbf{I}}_{jx} e_x + \hat{\mathbf{I}}_{jy} e_y + \hat{\mathbf{I}}_{jz} e_z \tag{A.3}$$

where $\mathbf{I}_{jz}$ is the *z*-component of the j^{th} angular momentum operator, Ω_j is the chemical shift of the j^{th} spin in the rotating frame of reference, J_{jk} is the J-coupling interaction between spins *j* and *k* and *n* is the number of spins in the molecule. All of these parameters can be

represented by a lower-triangular matrix **J** and vector **Ω**. For example, lactate is completely specified for simulation purposes by data taken from (Govindaraju et al., 2000):

$$\mathbf{\Omega} = \begin{bmatrix} 4.0974 & 1.3142 & 1.3142 & 1.3142 \end{bmatrix}^T \tag{A.4}$$

$$\mathbf{J} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 6.933 & 0 & 0 & 0 \\ 6.933 & 0 & 0 & 0 \\ 6.933 & 0 & 0 & 0 \end{bmatrix} \tag{A.5}$$

APPENDIX B

DERIVATION OF THE RELATIONSHIP BETWEEN THE REAL AND IMAGINARY PARTS OF THE MR SPECTRUM

The clean free induction decay (FID) signal (Keeler, 2005) of a single isochromat can be expressed as:

$$s(t) = Ce^{j\phi}e^{-\alpha t + j\omega_0 t} \tag{B.1}$$

where C is the maximum signal amplitude. The spectrum of this FID signal is obtained by FT,

$$
\begin{aligned}
S(\omega) &= \int_0^\infty s(t)\, e^{-j\omega t} dt \\
&= Ce^{j\phi} \int_0^\infty \mathrm{e}^{-\alpha t + j\omega_0 t - j\omega t} dt \\
&= Ce^{j\phi} \left[\frac{e^{[-\alpha + j(\omega_0 - \omega)]t}}{-\alpha + j(\omega_0 - \omega)} \right]_0^\infty \\
&= \frac{-Ce^{j\phi}}{-\alpha + j(\omega_0 - \omega)} \\
&= C \frac{\alpha \cos\phi - (\omega_0 - \omega)\sin\phi}{\alpha^2 + (\omega_0 - \omega)^2} + jC \frac{\alpha \sin\phi + (\omega_0 - \omega)\cos\phi}{\alpha^2 + (\omega_0 - \omega)^2} \\
&= \mathrm{Re} + j\,\mathrm{Im}
\end{aligned}
\tag{B.2}
$$

The inverse Fourier transform (IFT) of the spectrum is,

$$
\begin{aligned}
F^{-1}\left[S(\omega)\right] &= s(t) \\
&= \frac{1}{2\pi} \int_0^\infty S(\omega)\, e^{j\omega t} d\omega \\
&= \frac{1}{2\pi} \int_0^\infty (C \frac{\alpha \cos\phi - (\omega_0 - \omega)\sin\phi}{\alpha^2 + (\omega_0 - \omega)^2} + jC \frac{\alpha \sin\phi + (\omega_0 - \omega)\cos\phi}{\alpha^2 + (\omega_0 - \omega)^2}) \mathrm{e}^{j\omega t} d\omega \\
&= \frac{1}{2\pi} \left[\int_0^\infty C \frac{\alpha \cos\phi - (\omega_0 - \omega)\sin\phi}{\alpha^2 + (\omega_0 - \omega)^2} \mathrm{e}^{j\omega t}\, d\omega + j \int_0^\infty C \frac{\alpha \sin\phi + (\omega_0 - \omega)\cos\phi}{\alpha^2 + (\omega_0 - \omega)^2} e^{j\omega t} d\omega \right]
\end{aligned}
\tag{B.3}
$$

$$
\begin{aligned}
F^{-1}\left[S(\omega)\right] &= \frac{1}{2\pi} \int_0^\infty \left(\mathrm{Re}\right) \mathrm{e}^{j\omega t}\, d\omega + j \frac{1}{2\pi} \int_0^\infty \left(\mathrm{Im}\right) e^{j\omega t} d\omega \\
&= A + jB
\end{aligned}
\tag{B.4}
$$

where A and B are the IFTs of the real and imaginary parts of the spectrum, respectively. They can be expanded as,

$$
\begin{aligned}
A / \frac{C}{2\pi} &= \int_0^\infty \frac{\alpha \cos\phi - (\omega_0 - \omega)\sin\phi}{\alpha^2 + (\omega_0 - \omega)^2} \mathrm{e}^{j\omega t}\, d\omega \\
&= \alpha \cos\phi \int_0^\infty \frac{\mathrm{e}^{j\omega t}}{\alpha^2 + (\omega_0 - \omega)^2} d\omega - \sin\phi \int_0^\infty \frac{\left(\omega_0 - \omega\right)\mathrm{e}^{j\omega t}}{\alpha^2 + (\omega_0 - \omega)^2} d\omega
\end{aligned}
\tag{B.5}
$$

Let $x = \omega_0 - \omega$, then $\omega = \omega_0 - x$,

$$\begin{aligned} &= \alpha \cos\phi \int_0^\infty \frac{e^{j(\omega_0 - x)t}}{\alpha^2 + x^2} dx - \sin\phi \int_0^\infty \frac{x e^{j(\omega_0 - x)t}}{\alpha^2 + x^2} dx \\ &= \alpha e^{j\omega_0 t} \cos\phi \int_0^\infty \frac{e^{-jxt}}{\alpha^2 + x^2} dx - e^{j\omega_0 t} \sin\phi \int_0^\infty \frac{x e^{-jxt}}{\alpha^2 + x^2} dx \end{aligned} \tag{B.6}$$

Let $x' = -xt$, so $x = -x'/t$,

$$= -\alpha t e^{j\omega_0 t} \cos\phi \int_0^\infty \frac{e^{jx'}}{(\alpha t)^2 + (x')^2} dx' + e^{j\omega_0 t} \sin\phi \int_0^\infty \frac{x' e^{jx'}}{(\alpha t)^2 + (x')^2} dx' \tag{B.7}$$

Following the results from (Abramowitz and Stegun, 1964), the above equation can be expanded as,

$$\begin{aligned} &= -\alpha t e^{j\omega_0 t} \frac{j}{2\alpha t} [e^{-\alpha t} E_1(-\alpha t - jx') - e^{\alpha t} E_1(\alpha t - jx')] \cos\phi \\ &- e^{j\omega_0 t} \frac{1}{2} [e^{-\alpha t} E_1(-\alpha t - jx') + e^{\alpha t} E_1(\alpha t - jx')] \sin\phi \end{aligned} \tag{B.8}$$

where $E_1(z)$ is the exponential integral of z, and $x' = -(\omega_0 - \omega)t$. Replace $E_1(-\alpha t - jx')$ by $E1$, and $E_1(\alpha t - jx')]$ by $E2$,

$$\begin{aligned} &= -\frac{e^{j\omega_0 t}}{2} \Big[j[e^{-\alpha t} E1 - e^{\alpha t} E2] \cos\phi - [e^{-\alpha t} E_1 1 + e^{\alpha t} E2] \sin\phi \Big] \\ &= -\frac{e^{j\omega_0 t}}{2} \Big[e^{-\alpha t} E1(-\sin\phi + j\cos\phi) - e^{\alpha t} E2(\sin\phi + j\cos\phi) \Big] \\ &= \frac{e^{j\omega_0 t}}{2} \Big[[e^{-\alpha t} E1 + e^{\alpha t} E2] \sin\phi - j[e^{-\alpha t} E1 - e^{\alpha t} E2] \cos\phi \Big] \\ &= \frac{1}{2} \big(\cos\omega_0 t + j \sin\omega_0 t \big) \Big[[e^{-\alpha t} E1 + e^{\alpha t} E2] \sin\phi - j[e^{-\alpha t} E1 - e^{\alpha t} E2] \cos\phi \Big] \\ &= \frac{1}{2} \left[\begin{array}{l} \Big([e^{-\alpha t} E1 + e^{\alpha t} E2] \sin\phi \cos\omega_0 t + [e^{-\alpha t} E1 - e^{\alpha t} E2] \cos\phi \sin\omega_0 t \Big) \\ + j \Big([e^{-\alpha t} E1 + e^{\alpha t} E2] \sin\phi \sin\omega_0 t - [e^{-\alpha t} E1 - e^{\alpha t} E2] \cos\phi \cos\omega_0 t \Big) \end{array} \right] \end{aligned} \tag{B.9}$$

so,

$$
\begin{aligned}
A &= \frac{C}{4\pi}\left[\begin{array}{l}\left([e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\cos\omega_0 t+[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\sin\omega_0 t\right)\\ +j\left([e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\sin\omega_0 t-[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\cos\omega_0 t\right)\end{array}\right] \\
&= A_R + jA_I
\end{aligned}
\tag{B.10}
$$

where A_R and A_I are the real and imaginary part of A, respectively.

$$
\begin{aligned}
A_R &= \frac{C}{4\pi}\left([e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\cos\omega_0 t+[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\sin\omega_0 t\right) \\
A_I &= \frac{C}{4\pi}\left([e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\sin\omega_0 t-[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\cos\omega_0 t\right)
\end{aligned}
\tag{B.11}
$$

Similarly, we can get,

$$
\begin{aligned}
B\Big/\frac{C}{2\pi} &= \int_0^\infty \frac{\alpha\sin\phi+(\omega_0-\omega)\cos\phi}{\alpha^2+(\omega_0-\omega)^2}\mathrm{e}^{j\omega t}\,d\omega \\
&= \alpha\sin\phi\int_0^\infty \frac{\mathrm{e}^{j\omega t}}{\alpha^2+(\omega_0-\omega)^2}\,d\omega+\cos\phi\int_0^\infty \frac{(\omega_0-\omega)\mathrm{e}^{j\omega t}}{\alpha^2+(\omega_0-\omega)^2}\,d\omega
\end{aligned}
\tag{B.12}
$$

Let $x = \omega_0-\omega$, then $\omega = \omega_0 - x$,

$$
\begin{aligned}
&= \alpha\sin\phi\int_0^\infty \frac{\mathrm{e}^{j(\omega_0-x)t}}{\alpha^2+x^2}\,dx+\cos\phi\int_0^\infty \frac{x\,\mathrm{e}^{j(\omega_0-x)t}}{\alpha^2+x^2}\,dx \\
&= \alpha\,\mathrm{e}^{j\omega_0 t}\sin\phi\int_0^\infty \frac{\mathrm{e}^{-jxt}}{\alpha^2+x^2}\,dx+\mathrm{e}^{j\omega_0 t}\cos\phi\int_0^\infty \frac{x\,\mathrm{e}^{-jxt}}{\alpha^2+x^2}\,dx
\end{aligned}
\tag{B.13}
$$

Let $x' = -xt$, $x' = -(\omega_0-\omega)t$, so $x = -x'/t$,

$$
\begin{aligned}
&= -\alpha t\, \mathrm{e}^{j\omega_0 t} \sin\phi \int_0^\infty \frac{\mathrm{e}^{jx'}}{(\alpha t)^2 + (x')^2}\, dx' - \mathrm{e}^{j\omega_0 t} \cos\phi \int_0^\infty \frac{x' \mathrm{e}^{jx'}}{(\alpha t)^2 + (x')^2}\, dx' \\
&= -\alpha t\, \mathrm{e}^{j\omega_0 t} \frac{j}{2\alpha t} [e^{-\alpha t} E_1(-\alpha t - jx') - e^{\alpha t} E_1(\alpha t - jx')] \sin\phi \\
&\quad + \mathrm{e}^{j\omega_0 t} \frac{1}{2} [e^{-\alpha t} E_1(-\alpha t - jx') + e^{\alpha t} E_1(\alpha t - jx')] \cos\phi
\end{aligned}
\tag{B.14}
$$

Replace $E_1(-\alpha t - jx')$ by $E1$, and $E_1(\alpha t - jx')$ by $E2$,

$$
\begin{aligned}
&= -\frac{\mathrm{e}^{j\omega_0 t}}{2} \Big[j[e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi + [e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \Big] \\
&= -\frac{\mathrm{e}^{j\omega_0 t}}{2} \Big[e^{-\alpha t} E1(\cos\phi + j\sin\phi) + e^{\alpha t} E2(\cos\phi - j\sin\phi) \Big] \\
&= -\frac{\mathrm{e}^{j\omega_0 t}}{2} \Big[[e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi + j[e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \Big] \\
&= -\frac{1}{2} \big(\cos\omega_0 t + j\sin\omega_0 t \big) \Big[[e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi + j[e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \Big] \\
&= -\frac{1}{2} \left[\begin{array}{l} \big([e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \cos\omega_0 t - [e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \sin\omega_0 t \big) \\ + j\big([e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \cos\omega_0 t + [e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \sin\omega_0 t \big) \end{array} \right]
\end{aligned}
\tag{B.15}
$$

And,

$$
\begin{aligned}
B &= -\frac{C}{4\pi} \left[\begin{array}{l} \big([e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \cos\omega_0 t - [e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \sin\omega_0 t \big) + \\ j\big([e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \cos\omega_0 t + [e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \sin\omega_0 t \big) \end{array} \right] \\
&= B_R + jB_I
\end{aligned}
\tag{B.16}
$$

where B_R and B_I are the real and imaginary part of *B*, respectively. They can be expressed as,

$$
\begin{aligned}
B_R &= -\frac{C}{4\pi} \big([e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \cos\omega_0 t - [e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \sin\omega_0 t \big) \\
B_I &= -\frac{C}{4\pi} \big([e^{-\alpha t} E1 - e^{\alpha t} E2] \sin\phi \cos\omega_0 t + [e^{-\alpha t} E1 + e^{\alpha t} E2] \cos\phi \sin\omega_0 t \big)
\end{aligned}
\tag{B.17}
$$

Now let us find the relationship between the real and imaginary parts of A and B,

$$\frac{A_R}{A_I}=\frac{[e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\cos\omega_0 t+[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\sin\omega_0 t}{[e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\sin\omega_0 t-[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\cos\omega_0 t} \tag{B.18}$$

$$\frac{B_I}{B_R}=\frac{[e^{-\alpha t}E1-e^{\alpha t}E2]\sin\phi\cos\omega_0 t+[e^{-\alpha t}E1+e^{\alpha t}E2]\cos\phi\sin\omega_0 t}{-[e^{-\alpha t}E1-e^{\alpha t}E2]\sin\phi\sin\omega_0 t+[e^{-\alpha t}E1+e^{\alpha t}E2]\cos\phi\cos\omega_0 t} \tag{B.19}$$

Since $E1$ and $E2$ are not dependent on the phase ϕ, they should not change with different values of ϕ. Consider a special case at zero phase. When $\phi=0$, $\sin\phi=0$, and $\cos\phi=1$, equations (B.18, B.19) become,

$$\frac{A_R}{A_I}=\frac{[e^{-\alpha t}E1-e^{\alpha t}E2]\sin\omega_0 t}{-[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\omega_0 t}=-\frac{\sin\omega_0 t}{\cos\omega_0 t} \tag{B.20}$$

$$\frac{B_I}{B_R}=\frac{[e^{-\alpha t}E1+e^{\alpha t}E2]\sin\omega_0 t}{[e^{-\alpha t}E1+e^{\alpha t}E2]\cos\omega_0 t}=\frac{\sin\omega_0 t}{\cos\omega_0 t} \tag{B.21}$$

And the relationship between equations (B.20, B.21) is

$$\frac{B_I}{B_R}=-\frac{A_R}{A_I} \tag{B.22}$$

Similarly,

$$\begin{aligned}&A_R-A_I\\&=\frac{C}{4\pi}\left([e^{-\alpha t}E1+e^{\alpha t}E2]\sin\phi\left(\cos\omega_0 t-\sin\omega_0 t\right)+[e^{-\alpha t}E1-e^{\alpha t}E2]\cos\phi\left(\sin\omega_0 t+\cos\omega_0 t\right)\right)\\&B_R+B_I\\&=-\frac{C}{4\pi}\left([e^{-\alpha t}E1+e^{\alpha t}E2]\cos\phi\left(\sin\omega_0 t+\cos\omega_0 t\right)+[e^{-\alpha t}E1-e^{\alpha t}E2]\sin\phi\left(\cos\omega_0 t-\sin\omega_0 t\right)\right)\end{aligned} \tag{B.23}$$

when $\phi=0$, we have,

$$A_R - A_I = \frac{C}{4\pi}[e^{-\alpha t}E1 - e^{\alpha t}E2](\sin\omega_0 t + \cos\omega_0 t)$$
$$B_I + B_R = -\frac{C}{4\pi}[e^{-\alpha t}E1 + e^{\alpha t}E2](\sin\omega_0 t + \cos\omega_0 t) \tag{B.24}$$

$$\frac{A_R - A_I}{B_I + B_R} = \frac{e^{-\alpha t}E1 - e^{\alpha t}E2}{-e^{-\alpha t}E1 - e^{\alpha t}E2} \tag{B.25}$$

From equations (B.22) and (B.18, B.19), we get

$$\begin{aligned}&\frac{[e^{-\alpha t}E1 - e^{\alpha t}E2]\sin\phi\cos\omega_0 t + [e^{-\alpha t}E1 + e^{\alpha t}E2]\cos\phi\sin\omega_0 t}{-[e^{-\alpha t}E1 - e^{\alpha t}E2]\sin\phi\sin\omega_0 t + [e^{-\alpha t}E1 + e^{\alpha t}E2]\cos\phi\cos\omega_0 t}\\ &= -\frac{[e^{-\alpha t}E1 + e^{\alpha t}E2]\sin\phi\cos\omega_0 t + [e^{-\alpha t}E1 - e^{\alpha t}E2]\cos\phi\sin\omega_0 t}{[e^{-\alpha t}E1 + e^{\alpha t}E2]\sin\phi\sin\omega_0 t - [e^{-\alpha t}E1 - e^{\alpha t}E2]\cos\phi\cos\omega_0 t}\end{aligned} \tag{B.26}$$

Simplifying equation (B.26) yields,

$$\begin{aligned}&[e^{-\alpha t}E1 - e^{\alpha t}E2][e^{-\alpha t}E1 + e^{\alpha t}E2]\sin^2\phi\sin\omega_0 t\cos\omega_0 t - [e^{-\alpha t}E1 - e^{\alpha t}E2]^2\sin\phi\cos\phi\cos^2\omega_0 t\\ &+[e^{-\alpha t}E1 + e^{\alpha t}E2]^2\sin\phi\cos\phi\sin^2\omega_0 t - [e^{-\alpha t}E1 - e^{\alpha t}E2][e^{-\alpha t}E1 + e^{\alpha t}E2]\cos^2\phi\sin\omega_0 t\cos\omega_0 t\\ &= [e^{-\alpha t}E1 - e^{\alpha t}E2][e^{-\alpha t}E1 + e^{\alpha t}E2]\sin^2\phi\sin\omega_0 t\cos\omega_0 t + [e^{-\alpha t}E1 - e^{\alpha t}E2]^2\sin\phi\cos\phi\sin^2\omega_0 t\\ &-[e^{-\alpha t}E1 + e^{\alpha t}E2]^2\sin\phi\cos\phi\cos^2\omega_0 t - [e^{-\alpha t}E1 - e^{\alpha t}E2][e^{-\alpha t}E1 + e^{\alpha t}E2]\cos^2\phi\sin\omega_0 t\cos\omega_0 t\end{aligned}$$

$$\begin{aligned}&-[e^{-\alpha t}E1 - e^{\alpha t}E2]^2\sin\phi\cos\phi\cos^2\omega_0 t + [e^{-\alpha t}E1 + e^{\alpha t}E2]^2\sin\phi\cos\phi\sin^2\omega_0 t\\ &= [e^{-\alpha t}E1 - e^{\alpha t}E2]^2\sin\phi\cos\phi\sin^2\omega_0 t - [e^{-\alpha t}E1 + e^{\alpha t}E2]^2\sin\phi\cos\phi\cos^2\omega_0 t\end{aligned} \tag{B.27}$$

$$\begin{aligned}&[e^{-\alpha t}E1 - e^{\alpha t}E2]^2\sin\phi\cos\phi\left(\sin^2\omega_0 t + \cos^2\omega_0 t\right)\\ &= [e^{-\alpha t}E1 + e^{\alpha t}E2]^2\sin\phi\cos\phi\left(\sin^2\omega_0 t + \cos^2\omega_0 t\right)\end{aligned} \tag{B.28}$$

$$\begin{aligned}&[e^{-\alpha t}E1 - e^{\alpha t}E2]^2 = [e^{-\alpha t}E1 + e^{\alpha t}E2]^2\\ &e^{-\alpha t}E1 - e^{\alpha t}E2 = e^{-\alpha t}E1 + e^{\alpha t}E2\\ &\text{or}\\ &-e^{-\alpha t}E1 + e^{\alpha t}E2 = e^{-\alpha t}E1 + e^{\alpha t}E2\end{aligned} \tag{B.29}$$

$$E1 = 0 \text{ or } E2 = 0 \tag{B.30}$$

But from equations (B.1, B.4, B.10, B.16), we also have

$$
\begin{aligned}
& Ce^{j\phi}e^{-\alpha t+j\omega_0 t} = A + jB \\
& = -\frac{C\,\mathrm{e}^{j\omega_0 t}}{4\pi}\begin{pmatrix}\left[e^{-\alpha t}E1(-\sin\phi + j\cos\phi) - e^{\alpha t}E2(\sin\phi + j\cos\phi)\right] \\ + j\left[e^{-\alpha t}E1(\cos\phi + j\sin\phi) + e^{\alpha t}E2(\cos\phi - j\sin\phi)\right]\end{pmatrix} \\
& = -\frac{C\,\mathrm{e}^{j\omega_0 t}}{4\pi}\begin{pmatrix}e^{-\alpha t}E1(-\sin\phi + j\cos\phi + j\cos\phi - \sin\phi) \\ + e^{\alpha t}E2(-\sin\phi - j\cos\phi + j\cos\phi + \sin\phi)\end{pmatrix} \\
& = -\frac{C\,\mathrm{e}^{j\omega_0 t}}{4\pi}\left(2e^{-\alpha t}je^{j\phi}E1\right) \\
& = Ce^{j\phi}e^{-\alpha t+j\omega_0 t}\left(-\frac{jE1}{2\pi}\right)
\end{aligned}
\tag{B.31}
$$

So $E1$ is a constant,

$$E1 = 2\pi j \tag{B.32}$$

This leaves only one option in equation (B.30), that is

$$E2 = 0 \tag{B.33}$$

Equation (B.25) now becomes,

$$\frac{A_R - A_I}{B_I + B_R} = -1 \tag{B.34}$$

Substituting the values of $E1$ and $E2$ into the original equations (B.11, B.17), we can confirm that the relationships between A_R, A_I and B_R, B_I as in equations (B.22, B.34) are still valid with any value of ϕ. So the relations between the real and imaginary parts of A and B can be obtained by solving the following equations.

$$\begin{cases} B_I + B_R = -\left(A_R - A_I\right) \\ \dfrac{B_I}{B_R} = -\dfrac{A_R}{A_I} \end{cases} \tag{B.35}$$

The solutions for A and B are

$$A = -B_I + jB_R, \begin{cases} A_R = \dfrac{\dfrac{B_I}{B_R}\left(B_I + B_R\right)}{-\dfrac{B_I}{B_R} - 1} = \dfrac{(B_I + B_R)B_I}{-B_I - B_R} = -B_I \\ A_I = \dfrac{B_I + B_R}{\dfrac{B_I}{B_R} + 1} = \dfrac{(B_I + B_R)B_R}{B_I + B_R} = B_R \end{cases} \tag{B.36}$$

$$B = A_I - jA_R, \begin{cases} B_R = \dfrac{A_R - A_I}{\dfrac{A_R}{A_I} - 1} = \dfrac{(A_R - A_I)A_I}{A_R - A_I} = A_I \\ B_I = \dfrac{\dfrac{A_R}{A_I}(A_R - A_I)}{1 - \dfrac{A_R}{A_I}} = \dfrac{(A_R - A_I)A_R}{A_I - A_R} = -A_R \end{cases} \tag{B.37}$$

The FT of A and B are the real and imaginary parts of the spectrum.

BIBLIOGRAPHY

ABRAMOWITZ, M. & STEGUN, I. A. 1964. *Handbook of mathematical functions with formulas, graphs, and mathematical tables,* Washington,, U.S. Govt. Print. Off.

ANS, B., HERAULT, J. & JUTTEN, C. Year. Adaptive neural architectures: detection of primitives. *In:* In Proc. of COGNITIVA' 85, 1985 Paris, France 593-597.

BARKER, P. B., HEARSHEN, D. O. & BOSKA, M. D. 2001. Single-voxel proton MRS of the human brain at 1.5T and 3.0T. *Magnetic Resonance in Medicine,* 45**,** 765-769.

BARKER, P. B. & LIN, D. D. M. 2006. *In vivo* proton MR spectroscopy of the human brain. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 49**,** 99-128.

BELL, A. J. & SEJNOWSKI, T. J. 1995. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation,* 7**,** 1129-1159.

BIGLER, P. 2000. *NMR spectroscopy : processing strategies,* Weinheim ; Cambridge, Wiley-VCH.

BLOCH, F. 1946. Nuclear Induction. *Physical Review* 70**,** 460-474.

BOTTOMLEY, P. A. 1984. Selective volume method for performing localizedNMRspectroscopy. *US patent,* 4**,** 228.

BRACEWELL, R. N. 2000. *The Fourier transform and its applications,* Boston, McGraw Hill.

BRIGHAM, E. O. 1988. *The fast Fourier transform and its applications,* Englewood Cliffs, N.J., Prentice Hall.

BURKHARD, T., HERZOG, C., LINZBACH, S., SPYRIDOPOULOS, I., HUEBNER, F. & VOGL, T. J. 2009. Cardiac (31)P-MRS compared to echocardiographic findings in patients with hypertensive heart disease without overt systolic dysfunction--preliminary results. *Eur J Radiol,* 71**,** 69-74.

BURTSCHER, I. M. & HOLTAS, S. 2001. Proton magnetic resonance spectroscopy in brain tumours: clinical applications. *Neuroradiology,* 43**,** 345-52.

CANCINO-DE-GREIFF, H. F., RAMOS-GARCIA, R. & LORENZO-GINORI, J. V. 2002. Signal de-noising in magnetic resonance spectroscopy using wavelet transforms. *Concepts in Magnetic Resonance,* 14**,** 388-401.

CARDOSO, J.-F. 1997. Infomax and Maximum Likelihood for Blind Source Separation. *IEEE Signal Processing Letters,* 4**,** 112 - 114.

CARDOSO, J. F. 1998. Blind signal separation: statistical principles. *Proceedings of the IEEE,* 86**,** 2009-2025.

CARDOSO, J. F. 1999. High-order contrasts for independent component analysis. *Neural Computation,* 11**,** 157-92.

CARDOSO, J. F. & LAHELD, B. H. 1996. Equivariant adaptive source separation. *Signal Processing, IEEE Transactions on,* 44**,** 3017-3030.

CARDOSO, J. F. & SOULOUMIAC, A. 1993. Blind beamforming for non-Gaussian signals. *Radar and Signal Processing, IEE Proceedings F,* 140**,** 362-370.

CERDAN, S., PARRILLA, R., SANTORO, J. & RICO, M. 1985. ^{1}H NMR detection of cerebral myo-inositol. *FEBS Letters,* 187**,** 167-72.

COMON, P. 1994. Independent component analysis, a new concept? *Signal Processing,* 36**,** 287-314.

COMON, P., JUTTEN, C. & HERAULT, J. 1991. Blind separation of sources, Part II: problems statement. *Signal Processing,* 24**,** 11-20.

COOMBS, C. H., DAWES, R. M. & TVERSKY, A. 1970. *Mathematical psychology. An elementary introduction*, Englewood Cliffs, New Jersey: Prentice-Hall.

COVER, T. M. & THOMAS, J. A. 2006. *Elements of information theory,* Hoboken, N.J., Wiley ; Chichester : John Wiley [distributor].

DANCEA, F. & G NTHER, U. 2005. Automated Protein NMR Structure Determination Using Wavelet De-noised NOESY Spectra. *Journal of Biomolecular NMR***,** 139-152.

DAUBECHIES, I. 1992. *Ten lectures on wavelets,* Philadelphia, Pa., Society for Industrial and Applied Mathematics.

DAVIES, N. P., WILSON, M., HARRIS, L. M., NATARAJAN, K., LATEEF, S., MACPHERSON, L., SGOUROS, S., GRUNDY, R. G., ARVANITIS, T. N. & PEET, A. C. 2008. Identification and characterisation of childhood cerebellar tumours by *in vivo* proton MRS. *NMR in Biomedicine,* 21**,** 908-18.

DE GRAAF, R. A. 2007. *In Vivo NMR Spectroscopy: Principles and Techniques,* Chichester, John Wiley & Sons, Ltd.

DIAMANTARAS, K. I. & KUNG, S. Y. 1996. *Principal Component Neural Networks: Theory and Applications*, Wiley.

DONOHO, D. L. 1995. Denoising via soft thresholding. *IEEE Transactions on Information Theory,* 41**,** 613-627.

DONOHO, D. L. & JOHNSTONE, J. M. 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika,* 81**,** 425-455.

EDELSTEIN, W. A., GLOVER, G. H., HARDY, C. J. & REDINGTON, R. W. 1986. The intrinsic signal-to-noise ratio in NMR imaging. *Magnetic Resonance in Medicine,* 3**,** 604-618.

EFRON, B. & GONG, G. 1983. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician,* 37**,** 36-48.

EFRON, B. & TIBSHIRANI, R. 1993. *An introduction to the bootstrap,* New York, Chapman & Hall.

FASTICAPACKAGE. *http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.shtml* [Online]. [Accessed].

FRAHM, J., MERBOLDT, K. D. & HANICKE, W. 1987. Localized proton spectroscopy using stimulated echoes. *Journal of Magnetic Resonance,* 72**,** 502-508.

FUKUNAGA, K. 1990. *Introduction to statistical pattern recognition,* Boston, Academic Press.

GILLARD, J. H., WALDMAN, A. D. & BARKER, P. B. 2004. *Clinical MR Neuroimaging: Diffusion, Perfusion and Spectroscopy,* Cambridge, Cambridge University Press.

GILLESPIE, D. T. 1974. *A quantum mechanics primer,* New York, Wiley.

GIROLAMI, M. 1999. *Self-Organising Neural Networks - Independent Component Analysis and Blind Source Separation*, Springer-Verlag.

GOLUB, G. H. & VAN LOAN, C. F. 1989. *Matrix computations*, John Hopkins University Press.

GOVINDARAJU, V., YOUNG, K. & MAUDSLEY, A. A. 2000. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine,* 13**,** 129-53.

GUYON, I. 2006. *Feature extraction : foundations and applications,* Berlin ; New York, Springer-Verlag.

HAO, J., DAVIES, N. P., WILSON, M., PEET, A. C. & ARVANITIS, T. N. 2010. Noise reduction on *in vivo* proton MR spectra using wavelet shrinkage de-noising (WSD) and the Lorentzian lineshape property. *Proceedings of British Chapter ISMRM Post-Graduate Magnetic Resonance Symposium.*

HAO, J., WILSON, M. P., DAVIES, N. P., PEET, A. C. & ARVANITIS, T. N. 2007 Application of Independent Component Analysis for Feature Extraction and Blind Source Separation on Synthesized Brain Tumour ^{1}H Magnetic Resonance Spectra. *Proceedings of 13th Annual Meeting of the British Chapter of the ISMRM.*

HAO, J., WILSON, M. P., DAVIES, N. P., PEET, A. C. & ARVANITIS, T. N. 2008. A Novel Hybrid Method of Applying Independent Component Analysis on Synthesized Brain Tumour ^{1}H Magnetic Resonance Spectra. *Proceedings of 25th Annual Meeting of the ESMRMB.*

HAO, J., WILSON, M. P., DAVIES, N. P., SUN, Y., NATARAJAN, K., MACPHERSON, L., PEET, A. C. & ARVANITIS, T. N. 2009a. A Novel Hybrid Method for Applying Independent Component Analysis to *in vivo* Paediatric Brain Tumour ^{1}H Magnetic Resonance Spectra. *Proceedings of 17th Annual Meeting of the ISMRM.*

HAO, J., ZOU, X., WILSON, M. P., DAVIES, N. P., SUN, Y., PEET, A. C. & ARVANITIS, T. N. 2009b. A comparative study of feature extraction and blind source separation of independent component analysis (ICA) on childhood brain tumour ^{1}H magnetic resonance spectra. *NMR in Biomedicine,* 22**,** 809-18.

HENRIKSEN, O. 1995. In vivo quantitation of metabolite concentrations in the brain by means of proton MRS. *NMR in Biomedicine,* 8**,** 139-148.

HOCH, J. C. & STERN, A. S. 1996. *NMR Data Processing,* New York, John Wiley & Sons, Inc.

HOFMANN, L., SLOTBOOM, J., BOESCH, C. & KREIS, R. 2001. Characterization of the macromolecule baseline in localized ^{1}H-MR spectra of human brain. *Magnetic Resonance in Medicine,* 46**,** 855-63.

HORE, P. J. 1995. *Nuclear magnetic resonance,* Oxford ; New York, Oxford University Press.

HOWE, F. A., BARTON, S. J., CUDLIP, S. A., STUBBS, M., SAUNDERS, D. E., MURPHY, M., WILKINS, P., OPSTAD, K. S., DOYLE, V. L., MCLEAN, M. A., BELL, B. A. & GRIFFITHS, J. R. 2003. Metabolic profiles of human brain tumors using quantitative *in vivo* ^{1}H magnetic resonance spectroscopy. *Magnetic Resonance in Medicine,* 49**,** 223-32.

HOWE, F. A. & OPSTAD, K. S. 2003. ^{1}H MR spectroscopy of brain tumours and masses. *NMR in Biomedicine,* 16**,** 123-31.

HOYER, P. O. & HYV RINEN, A. 2000. Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems,* 11 191-210.

HUANG, Y., LISBOA, P. J. G. & EL-DEREDY, W. 2003. Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection. *Statistics in Medicine,* 22**,** 147-64.

HYV RINEN, A., KARHUNEN, J. & OJA, E. 2001. *Independent component analysis,* New York ; Chichester, Wiley.

HYVARINEN, A. 1999. Survey on Independent Component Analysis. *Neural Computing Surveys,* 2**,** 94-128.

HYVARINEN, A. & OJA, E. 2000. Independent component analysis: algorithms and applications. *Neural Networks,* 13**,** 411-30.

JACOBSEN, N. E. 2007. *NMR spectroscopy explained : simplified theory, applications and examples for organic chemistry and structural biology,* Hoboken, N.J., Wiley-Interscience.

JOHNSON, O. 2004. *Information theory and the central limit theorem*, Imperial College Press.

JOHNSON, R. W. 2001. An Introduction to the Bootstrap. *Teaching Statistics,* 23**,** 49-54.

JOLIOT, M., MAZOYER, B. M. & HUESMAN, R. H. 1991. In vivo NMR spectral parameter estimation: a comparison between time and frequency domain methods. *Magnetic Resonance in Medicine,* 18**,** 358-70.

JOLLIFFE, I. T. 2002. *Principal component analysis,* New York, Springer.

JUTTEN, C. & HERAULT, J. 1991. Blind separation of sources, Part 1: an adaptive algorithm based on neuromimetic architecture. *Signal Processing,* 24**,** 1-10.

KAY, S. M. 1993. *Fundamentals of statistical signal processing : estimation theory*, PTR Prentice Hall.

KEELER, J. 2005. *Understanding NMR Spectroscopy,* Chichester, John Wiley & Sons Ltd.

KENDALL, M., STUART, A. & ORD, J. K. 1987. *Kendall's Advanced Theory of Statistics,* NewYork, Oxford University Press.

KESHAVAN, M. S., STANLEY, J. A. & PETTEGREW, J. W. 2000. Magnetic resonance spectroscopy in schizophrenia: methodological issues and findings--part II. *Biological Psychiatry,* 48**,** 369-380.

KETONEN, L. M., HIWATASHI, A., SIDHU, R. & WESTESSON, P.-L. 2004. *Pediatric Brain and Spine: An Atlas of MRI and Spectroscopy,* Berlin, Springer.

KIM, J.-O. & MUELLER, C. W. 1978. *Introduction to factor analysis : what it is and how to do it,* Beverly Hills, Calif., Sage Publications.

KLOSE, U. 1990. *In vivo* proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine,* 14**,** 26-30.

KREIS, R. 1997. Quantitative localized ^{1}H MR spectroscopy for clinical use. *Journal of Progress in Nuclear Magnetic Resonance Spectroscopy,* 31**,** 155-195.

KUESEL, A. C., STOYANOVA, R., AIKEN, N. R., LI, C. W., SZWERGOLD, B. S., SHALLER, C. & BROWN, T. R. 1996. Quantitation of resonances in biological ^{31}P NMR spectra via principal component analysis: potential and limitations. *NMR in Biomedicine,* 9**,** 93-104.

LADROUE, C., HOWE, F. A., GRIFFITHS, J. R. & TATE, A. R. 2003. Independent component analysis for automated decomposition of *in vivo* magnetic resonance spectra. *Magnetic Resonance in Medicine,* 50**,** 697-703.

LANDINI, L., POSITANO, V. & SANTARELLI, M. F. 2005. *Advanced image processing in magnetic resonance imaging,* New York, Marcel Dekker ; London : Taylor & Francis [distributor].

LEE, Y. Y. B., HUANG, Y., EL-DEREDY, W., LISBOA, P. J. G., ARUS, C. & HARRIS, P. 2000. Robust methodology for the discrimination of brain tumours from *in vivo* magnetic resonance spectra. *Science, Measurement and Technology, IEE Proceedings -,* 147**,** 309-314.

LEVITT, M. H. 2001. *Spin dynamics : basics of nuclear magnetic resonance,* Chichester, John Wiley & Sons.

LI, S., WILLIAMS, G. D., FRISK, T. A., ARNOLD, B. W. & SMITH, M. B. 1995. A computer simulation of the static magnetic field distribution in the human head. *Magnetic Resonance in Medicine,* 34**,** 268-275.

LINSKER, R. 1988. Self-Organization in a Perceptual Network. *Computer,* 21**,** 105-117.

LOUIS, D. N., OHGAKI, H., WIESTLER, O. D., CAVENEE, W. K., BURGER, P. C., JOUVET, A., SCHEITHAUER, B. W. & KLEIHUES, P. 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol,* 114**,** 97-109.

LU, C. T. & WANG, H. C. 2003. Enhancement of Single Channel Speech Based on Masking Property and Wavelet Transform. *Speech Communication,* 41**,** 409-427.

LUTS, J., POULLET, J.-B., GARCIA-GOMEZ, J. M., HEERSCHAP, A., ROBLES, M., SUYKENS, J. A. K. & HUFFEL, S. V. 2008. Effect of feature extraction for brain tumor classification based on short echo time ^{1}H MR spectra. *Magnetic Resonance in Medicine,* 60**,** 288-298.

LV, C. & ZHAO, Q. 2005. A Universal PCA for Image Compression.

MA, J. & SUN, Z. 2005. MRS classification based on independent component analysis and support vector machines. *Proceedings of the 5th International Conference on Hybrid Intelligent Systems***,** 509-511.

MACOMBER, R. S. 1998. *A Complete Introduction to Modern NMR Spectroscopy* New York, John Wiley & Sons, Inc.

MAJUMDAR, A. 2009. Image compression by sparse PCA coding in curvelet domain. *Signal, Image and Video Processing,* 3**,** 27-34.

MALLAT, S. G. 1989. A Theory for Multiresolution Signal Decomposition: A Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11**,** 674-693.

MCROBBIE, D. W. 2007. *MRI from picture to proton,* Cambridge, UK ; New York, Cambridge University Press.

MENZE, B. H., LICHY, M. P., BACHERT, P., KELM, B. M., SCHLEMMER, H. P. & HAMPRECHT, F. A. 2006. Optimal classification of long echo time *in vivo* magnetic resonance spectra in the detection of recurrent brain tumors. *NMR in Biomedicine,* 19**,** 599-609.

MIERISOVA, S. & ALA-KORPELA, M. 2001. MR spectroscopy quantitation: a review of frequency domain methods. *NMR in Biomedicine,* 14**,** 247-259.

MISITI, M., MISITI, Y., OPPENHEIM, G. & POGGI, J. Wavelet Toolbox User's Guid.

MITCHELL, T. M. 1997. *Machine Learning,* New York ; London, McGraw-Hill.

MORENO-TORRES, A., MARTINEZ-PEREZ, I., BAQUERO, M., CAMPISTOL, J., CAPDEVILA, A., ARUS, C. & PUJOL, J. 2004. Taurine detection by proton magnetic resonance spectroscopy in medulloblastoma: contribution to noninvasive differential diagnosis with cerebellar astrocytoma. *Neurosurgery,* 55**,** 824-9.

MUTIHAC, R. & VAN HULLE, M. 2004. Comparison of principal component analysis and independent component analysis for blind source separation. *Romanian Reports in Physics,* 56**,** 20-32.

NANDI, A. K. 1999. *Blind estimation using higher-order statistics,* Boston, MA, Kluwer Academic Publishers.

OPSTAD, K. S., LADROUE, C., BELL, B. A., GRIFFITHS, J. R. & HOWE, F. A. 2007. Linear discriminant analysis of brain tumour ^{1}H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. *NMR in Biomedicine,* 20**,** 763-70.

OPSTAD, K. S., PROVENCHER, S. W., BELL, B. A., GRIFFITHS, J. R. & HOWE, F. A. 2003. Detection of elevated glutathione in meningiomas by quantitative *in vivo* ^{1}H MRS. *Magnetic Resonance in Medicine,* 49**,** 632-7.

PANIGRAHY, A., KRIEGER, M. D., GONZALEZ-GOMEZ, I., LIU, X., MCCOMB, J. G., FINLAY, J. L., NELSON, M. D., JR., GILLES, F. H. & BLUML, S. 2006. Quantitative short echo time ^{1}H-MR spectroscopy of untreated pediatric brain tumors: preoperative diagnosis and characterization. *American Journal of Neuroradiology,* 27**,** 560-72.

PAPOULIS, A. 1991. *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill.

PAULINO, A. C. & MELIAN, E. 1999. Medulloblastoma and supratentorial primitive neuroectodermal tumors: an institutional experience. *Cancer,* 86**,** 142-8.

PAWITAN, Y. 2001. *In all likelihood : statistical modelling and inference using likelihood,* Oxford, Clarendon.

PEET, A., LATEEF, S., MACPHERSON, L., NATARAJAN, K., SGOUROS, S. & GRUNDY, R. 2007. Short echo time 1 H magnetic resonance spectroscopy of childhood brain tumours. *Child's Nervous System,* 23**,** 163-169.

PEET, A. C., ARVANITIS, T. N., AUER, D. P., DAVIES, N. P., HARGRAVE, D., HOWE, F. A., JASPAN, T., LEACH, M. O., MACARTHUR, D., MACPHERSON, L., MORGAN, P. S., NATARAJAN, K., PAYNE, G. S., SAUNDERS, D. & GRUNDY, R. G. 2008. The value of magnetic resonance spectroscopy in tumour imaging. *Archives of Disease in Childhood,* 93**,** 725-727.

PFEUFFER, J., TKAC, I., PROVENCHER, S. W. & GRUETTER, R. 1999. Toward an *in vivo* neurochemical profile: quantification of 18 metabolites in short-echo-time ^{1}H NMR spectra of the rat brain. *Journal of Magnetic Resonance,* 141**,** 104-20.

PHAM, D. T. 1992. Separation of a mixture of independent sources through a maximum likelihood approach. *In Proc. EUSIPCO***,** 771 - 774.

PIZZO, P. A. & POPLACK, D. G. 2006. *Principles and practice of pediatric oncology,* Philadelphia, Pa. ; London, Lippincott Williams & Wilkins.

POULLET, J. B., SIMA, D. M., SIMONETTI, A. W., DE NEUTER, B., VANHAMME, L., LEMMERLING, P. & VAN HUFFEL, S. 2007. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. *NMR in Biomedicine,* 20**,** 493-504.

PREUL, M. C., CARAMANOS, Z., COLLINS, D. L., VILLEMURE, J. G., LEBLANC, R., OLIVIER, A., POKRUPA, R. & ARNOLD, D. L. 1996. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Nature Medicine,* 2**,** 323-5.

PRIESTLEY, M. B. 1981. *Spectral analysis and time series,* London ; New York, Academic Press.

PROVENCHER, S. 2009. LCModel and LCMgui Users Manual. http://s-provencher.com/pub/LCModel/manual/manual.pdf.

PROVENCHER, S. W. 1993. Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra. *Magnetic Resonance in Medicine,* 30**,** 672-679.

PROVENCHER, S. W. 2001. Automatic quantitation of localized *in vivo* ^{1}H spectra with LCModel. *NMR in Biomedicine,* 14**,** 260-4.

PULKKINEN, J., HAKKINEN, A. M., LUNDBOM, N., PAETAU, A., KAUPPINEN, R. A. & HILTUNEN, Y. 2005. Independent component analysis to proton spectroscopic imaging data of human brain tumours. *European Journal of Radiology,* 56**,** 160-4.

PURCELL, E. M., TORREY, H. C. & POUND, R. V. 1946. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review,* 69**,** 37.

RATINEY, H., SDIKA, M., COENRADIE, Y., CAVASSILA, S., ORMONDT, D. V. & GRAVERON-DEMILLY, D. 2005. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR in Biomedicine,* 18**,** 1-13.

REISER, M. F., SEMMLER, W. & HRICAK, H. 2008. *Magnetic resonance tomography,* New York, Springer.

REYNOLDS, G., WILSON, M., PEET, A. & ARVANITIS, T. N. 2006. An algorithm for the automated quantitation of metabolites in *in vitro* NMR signals. *Magnetic Resonance in Medicine,* 56**,** 1211-1219.

ROSS, B. & BLUML, S. 2001. Magnetic resonance spectroscopy of the human brain. *Anatomical Record,* 265**,** 54-84.

SALIBI, N. M. & BROWN, M. A. 1998. *Clinical MR spectroscopy : first principles,* New York, Wiley-Liss.

SCHORN, C. & TAYLOR, B. F. 2004. *NMR-spectroscopy : data acquisition,* Weinheim ; [Great Britain], Wiley-VCH.

SHANNON, C. E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal,* 27**,** 379-423, 623-656.

SIMONETTI, A. W., MELSSEN, W. J., EDELENYI, F. S. D., ASTEN, J. J. A. V., HEERSCHAP, A. & BUYDENS, L. M. C. 2005. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR in Biomedicine,* 18**,** 34-43.

SOROUCHYARI, F. 1991. Blind separation of sources, Part III: stability analysis. *Signal Processing,* 24**,** 21-29.

STAMATOPOULOS, V. G., KARRAS, D. A. & MERTZIOS, B. G. 2009. On an efficient modification of singular value decomposition using independent component analysis for improved MRS denoising and quantification. *Measurement Science and Technology,* 20 9 pp.

SZABO DE EDELENYI, F., SIMONETTI, A. W., POSTMA, G., HUO, R. & BUYDENS, L. M. C. 2005. Application of independent component analysis to ^{1}H MR spectroscopic imaging exams of brain tumours *Analytica Chimica Acta,* 544**,** 36-46.

TRBOVIC, N., DANCEA, F., LANGER, T. & GUNTHER, U. 2005. Using wavelet de-noised spectra in NMR screening. *Journal of Magnetic Resonance,* 173**,** 280-7.

VO-DINH, T. & GAUGLITZ, G. 2003. *Handbook of spectroscopy,* Weinheim ; Cambridge, Wiley-VCH.

WRIGHT, A. J., AR S, C., WIJNEN, J. P., MORENO-TORRES, A., GRIFFITHS, J. R., CELDA, B. & HOWE, F. A. 2008. Automated quality control protocol for MR spectra of brain tumors. *Magnetic Resonance in Medicine,* 59**,** 1274-1281.

YAN, H. 2002. *Signal processing for magnetic resonance imaging and spectroscopy,* New York, Dekker.

ZOU, X., JANCOVIC, P. & LIU, J. 2006. The effectiveness of ICA-based representation: application to speech feature extraction for noise robust speaker recognition. *In Proc. EUSIPCO.* florence, Italy.

ZOUBIR, A. M. & ISKANDER, D. R. 2004. *Bootstrap techniques for signal processing,* Cambridge, Cambridge University Press.