

Tensor Based Approaches in Magnetic Resonance Spectroscopic Imaging and Multi-parametric MRI data Analysis

Bharath Halandur Nagaraja

Supervisor:
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. L. De Lathauwer,
(co-supervisor)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

September 2018

Tensor Based Approaches in Magnetic Resonance Spectroscopic Imaging and Multi-parametric MRI data Analysis

Bharath HALANDUR NAGARAJA

Examination committee:

Prof. dr. ir. A Bultheel, chair

Prof. dr. ir. S. Van Huffel, supervisor

Prof. dr. ir. L. De Lathauwer, co-supervisor

Prof. dr. U. Himmelreich

Prof. dr. ir. F. Maes, secretary

Prof. dr. A. Heerschap

(RUNMC, Nijmegen)

Prof. dr. ir. N. Gillis

(Université de Mons)

Dr. ir. D. Sima

(Icometrix, Leuven)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

September 2018

© 2018 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Bharath Halandur Nagaraja, Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven
(Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Sabine Van Huffel for giving me an opportunity to join her team and work on this PhD. Thank you for the academic guidance, the nice words of encouragement and warm social activities. The positive and friendly environment you have developed in BioMed kept me motivated throughout my PhD. Secondly, I would like to express my gratitude for my co-supervisor Prof. Lieven De Lathauwer. Your competence and guidance helped me develop into a better researcher.

Many thanks to my mentor Dr. Diana Sima. Thank you for always being supportive and your advice has been invaluable for shaping this PhD.

To continue, I want to thank my assessors Prof. Uwe Himmelreich and Prof. Frederik Maes for the close collaboration during our meetings. Your support and knowledge contributed heavily to my personal development and to this thesis.

I am indebted to my Examination Committee for accepting to review this thesis and helping in improving the manuscript: Prof. Adhemar Bultheel, Prof. Yves Willems, Prof. Sabine Van Huffel, Prof. Lieven De Lathauwer, Prof. Uwe Himmelreich, Prof. Frederik Maes, Prof. Arend Heerschap, Prof. Nicolas Gillis and Dr. Diana Sima.

To the best groups BioMed and BIOTENSORS: Alex, Abhi, Adrian, Alexander, Amir, Anca, Amalia, Arun, Bori, Carolina, Dzemila, Dorien, Dries, Frederik, Griet, Ignat, Jasper, John, Jonathan, Kaat, Laure, Lieven, Martijn, Mario, Margot, Matthieu, Neetha, Ninah, Nico, Nicolas, Ofelie, Otto, Rob, Simon, Sibasankar, Stijn, Tim, Thomas, Vanya, Vladimir, Ying, Yipeng and Yissel. Thank you for making my stay in Belgium enjoyable. I am happy that I could take the memories of crazy parties, dinners, volleyball, sports day, survival of students and many more with me.

Special thanks to my office buddies Adrian and Nicolas. Thanks for providing

a vibrant workplace, I really enjoyed every discussion we had. It was a great pleasure to have you as my office mates.

I am grateful to the Leuven city and Belgian beers for providing a memorable experience during my PhD. I want to thank my friends, Eshawara and Surya for their support. I really enjoyed the trips, gatherings and sports with you guys.

Finally, a big thanks to my parents Nagaraja and Shakunthala, wife Vidya and son Ankush for all the smiles you brought on my face during the difficult times. I would also like to thank all my family members: Sathia, Sandhya, Gayathri, Malatheesh, Ashoka, Rama, Karthik, Nihkil and Nishal. Your support is immensely valuable and I couldn't have reached this far without all of you.

Abstract

Accurate characterisation and localization of pathologic tissue types play a key role in diagnosis and treatment planning of brain tumors. Neuroimaging techniques such as magnetic resonance imaging (MRI), magnetic resonance spectroscopic imaging (MRSI), perfusion-weighted imaging (PWI) and diffusion weighted imaging (DWI) are being used to characterize brain tumors and detect full tumor extent. Analysing these data is both time consuming and challenging for clinicians. Automated algorithms will aid clinicians to analyse the data faster and more accurately. Blind source separation is one such technique that is commonly used to extract useful information from the data. Most of these algorithms use matrix based approaches. Working with tensor tools such as tensor decompositions can be of great benefit compared to their matrix counterpart. Tensors are applied in domains such as signal processing, biomedical engineering, statistics and machine learning. In this thesis, we aim to develop tensor based blind source separation algorithms for analysing the MRSI and multi-parametric MRI (MP-MRI) signals.

First, tensor based blind source separation methods are developed to remove artefacts. In this thesis, we focus on residual water suppression in the MRSI signal. To suppress the residual water, a Löwner/Hankel tensor is constructed from the MRSI signal. Canonical polyadic decomposition (CPD)/Multilinear singular value decomposition is applied on the tensor to extract the water component, and to subsequently remove it from the original MRSI signal. The tensor based water suppression methods show significant improvement in performance for both simulated and in-vivo MRSI signals compared to the matrix-based approaches.

Second, tensor based blind source separation is applied to differentiate various tissue types in glioma patients from MRSI/multi-parametric MRI signals. Such a tensor based tumor tissue type differentiation approach is developed which consists of building a \mathbf{xx}^T structured 3-D tensor from the MRSI spectra and then applying a non-negative CPD to extract tissue specific spectra and

its corresponding distribution in the MRSI grid. An in-vivo study shows that our tensor based approach significantly outperforms the matrix-based approaches in identifying tumor and necrotic tissue type in glioma patients. This tensor based tissue characterization approach is further extended to multi-parametric magnetic resonance imaging (MP-MRI) including conventional magnetic resonance imaging, perfusion-weighted imaging, diffusion-weighted imaging and MRSI modalities to perform tumor segmentation.

Third, we explore the applicability of tensor decompositions in supervised algorithms for voxel classification in MRSI and tumour tissue segmentation in MP-MRI. A CNN based low-rank regularized classifier is developed to classify voxels in MRSI. Multilinear singular value decomposition (MLSVD) is used to apply regularization in the convolution layer. Low-rank regularization provides slight improvement in computational complexity without degrading the classification performance. For tumour tissue segmentation, a superpixel-wise two stage random forest algorithm is developed. The whole tumor is segmented in the first stage and in the second stage sub-compartments are segmented from the whole tumor. Multilinear singular value decomposition (MLSVD) is used to extract some of the features as input to the random forest classifier. The proposed algorithm was analysed on the BRATS 2017 challenge dataset, which showed a very good performance in segmenting the whole tumor and average performance in segmenting sub-compartments. This shows that tensor based feature extraction is a viable option for tumor tissue segmentation in MP-MRI.

Beknopte samenvatting

Nauwkeurige karakterisering en lokalisatie van pathologische weefseltypen spelen een belangrijk rol bij de diagnose en behandelingsplanning van hersentumoren. Neuroimaging-technieken zoals magnetische resonantie beeldvorming (MRI), magnetische resonantie spectroscopische beeldvorming (MRSI), perfusiegewogen beeldvorming (PWI) en diffusie gewogen beeldvorming (DWI) worden gebruikt om hersentumoren te karakteriseren en de volledige tumor omtrek te detecteren. Het analyseren van deze medische beelden is zowel tijdrovend als uitdagend voor klinici. Geautomatiseerde algoritmen helpen klinici de gegevens sneller en nauwkeuriger te analyseren. Blinde bronscheiding is een dergelijke techniek die vaak wordt gebruikt om nuttige informatie uit de gegevens te extraheren. De meeste algoritmen gebruiken matrixgebaseerde benaderingen. Het werken met tensor-tools zoals tensor-decomposities kan van groot voordeel zijn in vergelijking met hun matrix-tegenhanger. Tensoren worden toegepast in domeinen zoals signaalverwerking, biomedische engineering, statistiek en machine learning. In dit proefschrift proberen we op tensor gebaseerde blind bronscheidingsalgoritmen te ontwikkelen voor het analyseren van de MRSI en multi-parametrische MRI (MP-MRI) signalen.

Ten eerste presenteren we tensor gebaseerde blinde bronscheidingsmethoden om artefacten te verwijderen. In dit proefschrift richten we ons op residuale wateronderdrukking in het MRSI-signaal. Om het resterende water te onderdrukken, wordt een Löwner / Hankel-tensor geconstrueerd uit het MRSI-signaal. Canonische polyadische decompositie (CPD) wordt toegepast op de tensor om de watercomponent te extraheren en vervolgens te verwijderen uit het oorspronkelijke MRSI-signaal. De tensor gebaseerde wateronderdrukkingsmethoden vertonen een significante verbetering in performantie voor zowel gesimuleerde als in-vivo MRSI-signalen in vergelijking met de matrix gebaseerde benaderingen.

Ten tweede stellen we een tensor gebaseerde blinde bronscheiding methode voor om verschillende weefseltypen in glioom-patiënten te onderscheiden op

basis van MRSI / multi-parametrische MRI-signalen. De methode bestaat uit het bouwen van een xxT gestructureerde 3D-tensor uit de MRSI-spectra en het toepassen van een niet-negatieve CPD om weefsel-specifieke spectra en de overeenkomstige verdeling in het MRSI-rooster te extraheren. Een in vivo onderzoek toont aan dat onze tensor gebaseerde benadering significant beter is dan de matrix gebaseerde benaderingen bij het identificeren van tumor- en necrotisch weefseltype bij glioom-patiënten. De tensor gebaseerde benadering voor weefselkarakterisering wordt verder uitgebreid tot multiparametrische magnetische resonantie beeldvorming (MP-MRI), om conventionele magnetische resonantie beeldvorming, perfusie-gewogen beeldvorming, diffusie-gewogen beeldvorming en MRSI modaliteiten samen te gebruiken voor tumorsegmentatie.

Ten derde onderzoeken we de toepasbaarheid van tensor-decomposities in gesuperviseerde algoritmen voor voxel-classificatie in MRSI en segmentatie van tumorweefsel in MP-MRI. Een convolutional neural network (CNN) gebaseerde low-rank geregulariseerde classifier is ontwikkeld om voxels in MRSI te classificeren. Multilineaire enkelvoudige waarde-decompositie (MLSVD) wordt gebruikt om regularisatie toe te passen in de convolutie-laag. Low-rank regularisatie biedt een verbetering in de computationele complexiteit zonder om de classificatieprestaties te verslechteren. Voor segmentatie van tumorweefsel is een superpixel gebaseerde methode met twee stappen van random forests classificatie ontwikkeld. De gehele tumor is gesegmenteerd in de eerste fase en de subcompartimenten van de gehele tumor worden in de tweede fase gesegmenteerd. Multilineaire singuliere waarde-decompositie (MLSVD) wordt gebruikt om een subset van kenmerken te extraheren als input voor de random forest classifier. Het voorgestelde algoritme werd geanalyseerd op de BRATS 2017 challenge-dataset, met heel goede performantie in het segmenteren van de gehele tumor en redelijke performantie in het segmenteren van subcompartimenten. Dit toont aan dat tensor gebaseerde kenmerk-extractie een haalbare optie is voor tumorweefselsegmentatie in MP-MRI.

List of Abbreviations

AD	Axial Diffusivity
ADC	Apparent diffusion coefficient
AK	Axial Kurtosis
Ala	Alanine
AQSES	Accurate Quantitation of Short Echo time domain Signals
BSS	Blind Source Separation
BTD	Block Term Decomposition
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
CHESS	Chemical Shift Selective Suppression
Cho	Choline
cMRI	Conventional Magnetic Resonance Imaging
CNN	Convolutional Neural Network
CPD	Canonical Polyadic Decomposition
Cre	Creatine
CSF	Cerebro-Spinal Fluid
CT	Computed Tomography
DC	Distribution Correlation
DCE-MRI	Dynamic Contrast-Enhanced MRI

DKI	Diffusion Kurtosis Imaging
DNN	Deep Neural Networks
DSC-MRI	Dynamic Susceptibility Contrast Magnetic Resonance Imaging
DTI	Diffusion tensor imaging
DWI	Diffusion-Weighted Imaging
ECG	Electrocardiography
ED	Edema
EEG	Electroencephalography
ET	Enhancing Tumor
FA	Fractional Anisotropy
FC	Fully Connected
FLAIR	Fluid-Attenuated Inversion-Recovery
fMRI	Functional Magnetic Resonance Imaging
fWHM	Full-Width Half-Maximum
GABA	γ -aminobutyric acid
GBM	Glioblastoma Multiforme
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
GUI	Graphical User Interface
HGG	High-grade glioma
HLSVD-PRO	Hankel Lanczos singular value decomposition with partial reorthogonalization
hNMF	Hierarchical Non-negative Matrix Factorization
HSVD	Hankel singular value decomposition
HT	Hankel-tensor

ICA	Independent Component Analysis
KA	Kurtosis Anisotropy
Lac	Lactate
LDA	Linear Discriminant Analysis
LGG	low-grade glioma
Lip	Lipid
LT	Löwner Tensor
MAD	Median Absolute Deviation
MD	Mean Diffusivity
mhinge	Multiclass hinge loss
MI	myo-Inositol
MK	Mean Kurtosis
MLSVD	Multilinear Singular Value Decomposition
MOIST	Multiply Optimized Insensitive Suppression Train
MP	Maximum Pooling
MP-MRI	Multi-parametric Magnetic Resonance Imaging
MPFIR	Maximum-Phase Finite Impulse Response
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
MT	Mixing Time
MTT	Mean Transit Time
NAA	N-Acetyl Aspartate
NCPD	Non-negative Canonical Polyadic Decomposition
NCPD-A	Non-negative Canonical Polyadic Decomposition with automatic source assignment

NCPD-M	Non-negative Canonical Polyadic Decomposition with manual source assignment
NCR	Necrotic
NET	Non-Enhancing Tumor
NMF	Non-negative Matrix Factorization
NNDSVD	Non-negative Double Singular Value Decomposition
NNLS	Non-negative least squares
PCA	Principal Component Analysis
PCh	Phosphocholine
PD	Polyadic Decomposition
PET	Positron Emission Tomography
ppm	parts-per-million
PPV	Positive Predictive Value
PRESS	Point RESolved Spectroscopy
PWI	Perfusion Weighted Imaging
QUEST	QUantitation based on quantum ESTimation
rCBV	Relative cerebral blood volume
RD	Radial Diffusivity
ReLU	Rectified Linear Unit
RF	Random Forest
RK	Radial Kurtosis
ROI	Region of Interest
SNR	Signal-To-Noise Ratio
SPC	True Negative Rate
SPECT	Single-Photon Emission Computed Tomography
std dev	Standard Deviation

STEAM	STimulated Echo Acquisition Mode
SVM	Support Vector Machines
TARQUIN	Totally Automatic Robust Quantitation in NMR
Tau	Taurine
TC	Tumor Core
TPR	True Positive Rate
TR	Repetition Time
VOI	Volume of Interest
WHO	World Health Organization
WT	Whole Tumor

Contents

Abstract	iii
Beknopte samenvatting	v
List of Abbreviations	vii
Contents	xiii
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Aim of the thesis	1
1.2 Chapter-by-chapter overview	6
2 Mathematical background: tensor decomposition and machine learning	11
2.1 Tensor decomposition	11
2.1.1 Notations and tensor preliminaries	12
2.1.2 Tensorization: Löwner and Hankel matrices/tensors . .	14
2.1.3 Canonical polyadic decomposition	16

2.1.4	Multilinear singular value decomposition	18
2.1.5	Block term decomposition	19
2.1.6	Structured tensor decompositions	21
2.1.7	Applications of tensor decomposition	21
2.2	Machine learning	22
2.2.1	Un-supervised: blind source separation	22
2.2.2	Supervised learning	30
3	Multi-parametric magnetic resonance imaging	37
3.1	Magnetic resonance imaging	37
3.1.1	Principles of MRI	37
3.1.2	Conventional MRI	40
3.1.3	Magnetic resonance spectroscopic imaging	41
3.1.4	Diffusion-weighted imaging	44
3.1.5	Perfusion-weighted imaging	44
3.2	Glioma	45
3.2.1	Neuroimaging in glioma	46
3.3	Data acquisition and pre-processing	49
3.3.1	UZ Leuven data	49
3.3.2	BRATS 2017 challenge dataset	52
4	Residual water suppression in MRSI	53
4.1	Introduction	53
4.2	Tensor based methods for water removal	55
4.2.1	MRSI and residual water	55
4.2.2	Löwner-based water suppression	56
4.2.3	Hankel-tensor based water suppression in MRSI	60
4.3	Results	61

4.3.1	Spectral variations in MRSI voxels	61
4.3.2	Simulations	62
4.3.3	In-vivo results	64
4.4	Discussion	70
4.5	Conclusion	74
5	Un-supervised tissue type differentiation in glioma	75
5.1	Introduction	75
5.2	Method	77
5.2.1	MRSI tensor construction	77
5.2.2	Non-negative CPD	79
5.2.3	Spectral recovery and non-negative least squares	81
5.2.4	Initialization	82
5.2.5	Source number estimation	82
5.2.6	Source and distribution correlation	83
5.3	Results on brain tumor dataset	86
5.4	Discussion	92
5.5	Conclusion	94
6	Supervised tumor voxel classification in MRSI	95
6.1	Introduction	95
6.2	Method	97
6.2.1	Feature extraction	97
6.2.2	Non-negative CPD for tumor classification	97
6.2.3	Random forest	98
6.2.4	Convolutional neural network	98
6.2.5	Low-rank regularization	99
6.3	Results and Discussion	104

6.4	Conclusion	109
7	Un-supervised tumor Segmentation in Multi-parametric MRI	111
7.1	Introduction	111
7.2	Method	112
7.2.1	Tensor construction	112
7.2.2	Constrained canonical polyadic decomposition	113
7.2.3	Validation	115
7.3	Results and discussion	115
7.4	Conclusion	119
8	Supervised tumor Segmentation in Multi-parametric MRI	121
8.1	Introduction	121
8.2	Method	123
8.2.1	Preprocessing	123
8.2.2	Feature extraction	123
8.2.3	Training and tissue segmentation	125
8.3	Results and Discussion	127
8.3.1	First Stage: whole tumor segmentation	127
8.3.2	High grade glioma	128
8.3.3	Validation and test dataset results	130
8.4	Conclusion	132
9	Conclusion and future perspectives	133
9.1	Conclusion	133
9.2	Future perspectives	137
A	Tissue type assignment	141

Bibliography	145
Curriculum vitae	165
List of publications	167

List of Figures

1.1	(b) T2-weighted image showing tumor around the center. (b) MRSI colour map of myoinositol. Image adapted from: [31]. . .	2
1.2	Residual water suppression from one of the voxel in the MRSI signal.	4
1.3	Nosologic images of normal, tumor and necrosis tissue types. Image adapted from: [111].	4
1.4	Schematic overview of the thesis.	6
2.1	Graphical representation of vector, matrix and third-order tensor.	12
2.2	Mode-1, mode-2 and mode-3 fibers of third-order tensor. Source: [91].	13
2.3	Graphical representation of a third-order tensor matricization in all three modes. Source: [171].	13
2.4	Graphical representation of a canonical polyadic decomposition for a third-order tensor.	17
2.5	Graphical representation of a MLSVD for a third-order tensor.	19
2.6	Graphical representation of a rank- $(L_r, L_r, 1)$ block Tensor decomposition of a third-order tensor. Source: Tensorlab documentation [185].	20
2.7	Graphical representation of NMF algorithm used for MRSI tissue differentiation.	24
2.8	Schematic representation of hNMF algorithm for MRSI tissue differentiation. Source: [111].	25

2.9	Schematic representation of hNMF algorithm for tissue differentiation in MP-MRI. Source: [152]	26
2.10	Majority voting scheme used in random forest. Image adapted from [133].	31
2.11	Model of a neuron as used in neural networks. Source: [83]	32
2.12	Hyperbolic tangent, sigmoid and ReLU activation functions from left to right, respectively. Image adapted from: [95]	33
2.13	Convolution operation on an image of size 4×4 with a 2×2 kernel. Image obtained from: [79]	34
2.14	3-D model of CNN convolution layer. Image obtained from: [153]	35
2.15	Demonstration of max-pooling operation. Image adapted from: [95]	35
2.16	Fully connected layer with one hidden layer. Image adapted from: [83]	36
3.1	Graphical representation of precessing spins under an external magnetic field B_0 and the resulting net magnetization vector M . Source: [152]	38
3.2	Illustration of change in the direction of magnetization vector M after applying the RF pulse. Source: [152].	39
3.3	PRESS and STEAM acquisition sequences for MRS. MT stands for mixing time. Source: [15].	42
3.4	(A) MRSI grid overlaid on T_1 -weighted image. (B) Spectra corresponding to voxel grid within the white box.	43
3.5	Absorption spectrum (real part) of a MRS signal obtained from a tumor region in brain.	43
3.6	T1+contrast, T2, FLAIR and expert delineation overlaid on T1+contrast of a grade IV GBM patient left to right, respectively. Active tumor, necrosis and edema are shown in red, green and blue, respectively.	46
3.7	CBV map and expert delineation overlaid on T1+contrast of a grade IV GBM patient left to right, respectively. Active tumor, necrosis and edema are shown in red, green and blue, respectively.	47

3.8	MD, FA, MK maps and expert delineation overlaid on T1+contrast of a grade IV GBM patient left to right, respectively. Active tumor, necrosis and edema are shown in red. green and blue, respectively.	47
3.9	Top row: MRSI grid overlaid on T2 image and spectrum from a voxel in necrotic region (from left to right). Bottom row: Quantified NAA and Lip maps overlaid on T2 image (from left to right).	48
4.1	Absorption spectrum from one of the voxels in the MRSI grid with a large residual water signal. The region of interest for metabolites is shown within the red box.	56
4.2	Construction of the Löwner tensor \mathcal{T} from the spectra of the different MRSI voxels	58
4.3	(a) & (c) Absorption spectra of an in-vivo signal without water suppression (red, dashed-line) overlapped with the estimated water spectra (black, solid-line) from two different voxels in the MRSI grid. (b) & (d) Individual resonance peaks used in the modelling of the water signal from (a) and (c), respectively. . .	62
4.4	(a) Absorption spectrum of a simulated in-vitro signal without water from one of the MRSI voxels. (b) Absorption spectrum of a simulated in-vitro signal with large water peak from one of the MRSI voxels. (c) Absorption spectrum of the water-suppressed signal using the Löwner method without polynomial sources. (d) Absorption spectrum of the water-suppressed signal using the Löwner method with polynomial sources.	63
4.5	Boxplot of the residual errors after water suppression using the HSVD, the Löwner (LT) and the Hankel-tensor (HT) methods on 100 simulated in-vitro MRSI signals. The error is calculated as the l_2 -norm of the difference between the water-suppressed signal and the original water-free signal (with noise).	64

- 4.6 Residual water suppression in in-vivo MRSI signals. (a-b) Absorption spectra of measured MRSI signal with large residual water peak for two voxels. (c-d) Absorption spectrum of the water-suppressed signal (blue) using the Löwner method and the quantified signal (red) in the two corresponding voxels. (e-f) Absorption spectrum of the water-suppressed signal using HSVD method and the quantified signal (red) in the two corresponding voxels. (g-h) Absorption spectrum of the water-suppressed signal using the Hankel-tensor method and the quantified signal (red) in the two corresponding voxels. 65
- 4.7 Original and residual water suppressed spectra from different locations in 16×16 voxel grid. First column: Un-suppressed original spectra, second column: water suppressed spectra using HSVD, third column: water suppressed spectra using the Hankel-tensor method, fourth column: water suppressed spectra using the Löwner tensor method. Voxel position in the 16×16 grid: Row-1: top left corner, Row-2: middle left, Row-3: bottom left corner, Row-4: middle top, Row-5: centre, Row-6: bottom middle, Row-7: top right corner, Row-8: middle right, Row-9: top right corner. 66
- 4.8 Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals. The error is calculated as the difference between the variance in water region segment and the variance from a segment in the noise region. 67
- 4.9 Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals truncated to 1024 samples per voxel. The error is calculated as the difference between the variance the in water region segment and the variance from a segment in the noise region. 68
- 4.10 Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals truncated to 512 samples per voxel. The error is calculated as the difference between the variance in the water region segment and the variance from a segment in the noise region. 68

4.11	Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals truncated to 256 samples per voxel. The error is calculated as the difference between the variance in the water region segment and the variance from a segment in the noise region.	69
4.12	Improper water suppression using the HSVD method in two voxels of the 10×10 MRSI grid. (a-b) Absorption spectra of measured MRSI signal with large residual water peak for two voxels. (c-d) Absorption spectrum of the water-suppressed signal (blue) using the Löwner method and the quantified signal (red) in the two corresponding voxels. (e-f) Absorption spectrum of the water-suppressed signal using HSVD method and the quantified signal (red) in the two corresponding voxels. (g-h) Absorption spectrum of the water-suppressed signal using the Hankel-tensor method and the quantified signal (red) in the two corresponding voxels.	71
5.1	3-way spatial tensor representation of 2-D MRSI data.	76
5.2	(a) Construction of reduced spectrum, \mathbf{x} from the pre-processed spectra. SOS: sum of squares. (b) Construction of the MRSI tensor \mathcal{T} from the reduced spectra, \mathbf{x} . K is the total number of voxels in the 2D MRSI excitation volume.	79
5.3	Non-negative CPD of MRSI tensor \mathcal{T} : MRSI tensor \mathcal{T} is decomposed into R rank-1 tensors. Common factor \mathbf{S} is used in mode-1 and mode-2 to maintain symmetry of frontal slices. Each \mathbf{s}_i gives a tissue-specific reduced spectral pattern and the corresponding \mathbf{h}_i gives the spatial distribution of the respective tissue type, upon reshaping. Non-negativity of \mathbf{S} and \mathbf{H} is imposed in the decomposition.	80
5.4	Generation of expert labeled tissue-specific (Tumor) spectrum and distribution vector. Calculation of source and distribution correlation is shown in the box. S_{ij} is the spectra at i^{th} column and j^{th} row. The tissue type T-tumor or C-control is shown between braces.	85

- 5.5 (a) Left to right, First image: T2-weighted anatomical MR image of a brain tumor with areas of necrosis. Second image: voxels within the MRSI excitation volume superimposed on anatomical image. Third image: expert labeling, where yellow (horizontal + vertical line) indicates necrotic, red (horizontal line) indicates tumor, dark blue (slanted + horizontal line) indicates normal, light blue (vertical line) indicates normal/tumor and green (slanted line) indicates spectra of poor quality. (b) Tissue distribution maps obtained from an $(L_r, L_r, 1)$ block term decomposition of spatial tensor \mathcal{P} . First three images from the left correspond to tumor, necrotic and normal tissue distribution respectively. Remaining three images correspond to bad spectra/artefact. (c) Tissue distribution maps obtained using the NCPD- l_1 algorithm on the MRSI tensor \mathcal{T} . First three images from the left correspond to tumor, necrotic and normal tissue distribution respectively. Remaining three images correspond to tumor, normal tissue and bad spectra/artefact distribution respectively. In this dataset the tumor tissue is modelled by two sources. 87
- 5.6 Box-plot of Type I source correlation, Type II source correlation and distribution correlation values obtained from NCPD without regularization, NCPD with l_1 regularization, single stage NMF and hNMF algorithms. Zero correlation values indicate that the specific tissue type was not recovered. 88
- 5.7 Estimated number of sources from the covariance matrix for 28 MRSI datasets with and without using the maximum number of tissue types as prior knowledge. The horizontal line indicates the maximum number of tissue types used as prior knowledge in the analysis ($P = 8$). 90

- 5.8 Tissue pattern differentiation using ^1H MRSI: C, T and N represent normal, tumor and necrosis, respectively. (a) First row: T2-weighted anatomical MR image of a brain tumor with areas of necrosis. Second row: voxels within the MRSI excitation volume superimposed on anatomical image. Third row: expert labeling, where yellow (horizontal + vertical line) indicates N, red (horizontal line) indicates T, magenta (no pattern) indicates T/N, dark blue (slanted + horizontal line) indicates C, light blue (vertical line) indicates C/T and green (slanted line) indicates spectra of poor quality. (b, c) results of NCPD. (b) The recovered sources from the NCPD- l_1 method are shown in black (solid line). First three rows represent C, N and T spectral sources in black (solid line), with tissue-specific spectra based on expert labeling overlaid in green (dash-dot line). The remaining four rows represent artifacts and spectra from outer edges. (c) Distribution maps corresponding to spectral profiles in (b). (d, e) Results of single-stage NMF. (d) The recovered sources are shown in black, overlaid with the expert-based tissue-specific spectra in red. First three rows show normal, necrotic and tumor spectra and the remaining rows show other spectra obtained using rank-7 NMF. (e) Distribution maps corresponding to (d). (f, g) Results of hNMF. (f) Recovered sources shown in black and expert-based tissue-specific spectra in red. First two rows show control and necrotic spectra. (g) Distribution maps corresponding to (f). . . . 91
- 6.1 Original spectra, reduced spectra and Löwner matrices belonging to three different classes. First row: real part of the spectra, second row: imaginary part of the spectra, third row: reduced spectra and fourth row: Löwner matrices. First column: tumor, second column: normal and third column: bad. 100
- 6.2 Convolutional neural network architecture used for voxel classification in MRSI. Each convolution layer shows the filter size (e.g. 3×3) and the number of output channels (e.g. 128). Maxpool/2 represents a maxpool layer with size 2×2 and stride 2, which reduces the dimension of the input image by half. 101

- 6.3 Low-rank regularized representation of convolution layer. The original convolution layer is shown in the left and the low-rank regularized representation is shown in the right. Each convolution layer shows the filter size (e.g. 3×3) and the total size of the kernel (e.g. $3 \times 3 \times I \times J$). Convolution layers in the low-rank regularized representation also show the factor matrix/core tensor corresponding to that layer 103
- 6.4 Boxplot of sensitivity, specificity, precision and F1 Score over 17 patients from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization (CNN-TR) and NCPD with automatic source assignment (NCPD-A). These values are calculated for Tumor vs others (normal + bad). 106
- 6.5 Boxplot of sensitivity, specificity, precision and F1 Score over 17 patients from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization (CNN-TR) and NCPD with automatic source assignment (NCPD-A). These values are calculated for Normal vs others (tumor + bad). 107
- 7.1 (a) Construction of feature vector \mathbf{x} from MP-MRI. (b) Construction of the MP-MRI tensor \mathcal{T} from the feature vector \mathbf{x} . K is the total number of voxels in the region of interest (ROI). 113
- 7.2 Partial non-negative CPD of MP-MRI tensor \mathcal{T} : MP-MRI tensor \mathbf{T} is decomposed into R rank-1 tensors. Common factor \mathbf{S} is used in mode-1 and mode-2 to maintain symmetry of frontal slices. Each \mathbf{s}_i gives a tissue-specific feature vector and the corresponding \mathbf{h}_i gives the spatial distribution of the respective tissue type, upon reshaping. A non-negativity constraint is imposed only on \mathbf{H} in the decomposition. 114
- 7.3 Coregistered MRI maps of several modalities and tissue abundance maps obtained from hNMF and CPD algorithms. The input maps: (A) T2, (B) T1+C, (C) MD, (D) MK, (E) rCBV, (F) Cho, (G) Cre and (H) Lip. The green box indicates the ROI. (I-L) tissue abundance maps obtained from hNMF ((I) Active tumor, (J) Necrosis, (K) Edema, (L) Cerebrospinal fluid). (M-P) tissue abundance maps obtained from constrained CPD ((M) Active tumor, (N) Necrosis, (O)) Edema, (P) Cerebrospinal fluid). 117

7.4	(A) Sources consisting of MP-MRI features obtained from CPD and hNMF algorithms. The estimated sources are shown for tumor and necrosis tissue type. (B) Comparison of the segmentation by radiologist (blue) with the segmentations obtained from CPD and hNMF (green). First two images correspond to tumor and necrosis segmentation from CPD, respectively. Last two images correspond segmentations from hNMF. Cyan indicates segmentation overlap.	118
8.1	(a) One slice of the FLAIR image. (b) Generated superpixels for the slice in (a).	123
8.2	Truncated multilinear singular value decomposition and feature extraction.	125
8.3	Demonstration of whole tumor segmentation in first stage and sub-tissue segmentation in second stage.	126
8.4	Boxplots of Dice scores and sensitivity for whole tumor (WT) obtained from first stage model on BRATS 2017 validation dataset of 46 patients.	128
8.5	Segmentation results on two slices 1-2. (a) T2 image of one slice, (b) Estimated segmentation (c) Expert segmentation. Green-Edema, Brown-enhancing tumor and Blue- Necrosis.	129
8.6	Boxplots of Dice scores for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) on BRATS 2017 training dataset of 63 patients.	130
8.7	Boxplots of Dice scores for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) on BRATS 2017 validation dataset of 46 patients.	131
A.1	Shows real part of the tissue specific spectrum ($real(W_i)$) obtained from NCPD. The peaks that are used for tissue type assignment are labelled.	142
A.2	Flowchart showing the procedure of assigning a tissue type to the source spectrum. B-bad; N-necrosis; C-control; T-tumor. FWHM stands for full width at half maximum.	143

List of Tables

4.1	Average residual water suppression error (difference in variance) obtained from HSVD, Löwner and Hankel-tensor methods over 28 in-vivo MRSI signals. The results are shown for original MRSI signal with 2048 samples for each voxel and truncated MRSI signals with 1024, 512 and 256 samples for each voxel.	69
4.2	Mean and standard deviation of Cramér-Rao bounds in % of quantified amplitude for Lipid (Lip1), Glutamate, N-acetylaspartate (NAA) and phosphocholine (PCh) metabolites over 10×10 MRSI voxel grid.	70
5.1	Mean, standard deviation, median, MAD and range of source correlation (Type I and Type II) and distribution correlation for 28 glioma datasets. The results are shown on tumor and necrotic tissue type for NCPD without regularization, NCPD with l_1 regularization, single stage NMF and hNMF algorithms. P-values indicate the statistical significance of higher median for the NCPD- l_1 algorithm (* indicates statistical significance for $p < 0.01$)	89
6.1	Number of scans along with the grid size of the MRSI signal for 17 different patients.	105

6.2	Average sensitivity, specificity, precision and F1 Score over 17 patients from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization and NCPD with automatic source assignment (NCPD-A). These values are calculated for Tumor vs others (normal + bad) and Normal vs others (tumor + bad).	108
7.1	Dice scores and source correlation of fourteen HGG patients using CPD and hNMF algorithms. Dice scores are shown for active tumor and the tumor core (tumor + necrosis) region. Source correlation is shown for active tumor tissue type. Mean and standard deviation are shown in the last two rows, respectively.	116
8.1	Features used in stage one and stage two along with their corresponding dimension.	127
8.2	Mean, standard deviation, median, 25 quantile and 75 quantile of Dice score and sensitivity for whole tumor (WT) over forty six patients from validation set using only first stage model. . .	128
8.3	Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and sensitivity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over 63 HGG patients. . . .	129
8.4	Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and sensitivity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over 46 validation dataset. .	130
8.5	Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and Hausdorff95 for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over test dataset of 146 patients.	131

Chapter 1

Introduction

1.1 Aim of the thesis

Neuroimaging techniques such as magnetic resonance imaging (MRI) and magnetic resonance spectroscopic imaging (MRSI) allow to visualize brain tissue non-invasively and are therefore very helpful in diagnosis and prognosis of various neurological diseases. In Neurooncology, these techniques are used extensively in both research as well as the clinical environment. In clinical settings these MRI techniques are being used in the characterisation and localization of pathologic tissue (Figure 1.1). Hence, they are an important aid to diagnose brain tumor and, to assist in presurgical planning and post treatment management of brain tumors. Manually characterizing/localizing tumor tissues is a tedious and time consuming job, which also suffers from inter and intra-rater variability. Developing machine learning based automated or semi-automated algorithms for pre-processing and analysing MRSI/multi-parametric MRI (MP-MRI) signals will aid clinicians to overcome those problems and results in better diagnosis and prognosis of brain tumors.

Machine learning is “the capability of the computer program to acquire or develop new knowledge or skills from existing or non existing examples for the sake of optimising performance criterion” [7]. Fundamentally, there are two different tasks within machine learning, namely supervised and un-supervised learning. In this thesis, both techniques are used with focus on un-supervised techniques, specifically blind source separation. Blind source separation (BSS) is one of the un-supervised machine learning techniques which consists of recovering the original signals from the mixture without (or as little as possible)

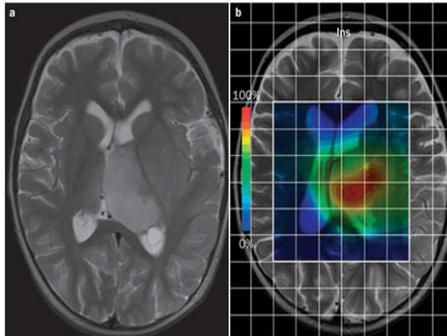


Figure 1.1: (a) T2-weighted image showing tumor around the center. (b) MRSI colour map of myoinositol. Image adapted from: [31].

prior knowledge of the mixing process or the original signals. Substantial progress in data recording technologies and information processing in recent years have enabled acquisition and analysis of large amounts of biomedical data. Extraction of the underlying health relevant information patterns, called sources, is crucial for a reliable prediction of the underlying pathology or health condition. Blind source separation provide a nice framework for such applications. BSS techniques are extensively used in biomedical applications [119]. Biomedical applications generate large amounts of data. Most of these data are not labelled as it requires a significant amount of time and domain-specific expertise. BSS methods don't require any training based on labelled data and can be applied directly on the individual data. These advantages of BSS techniques have motivated us to use it for analysing MRSI data.

Commonly used BSS techniques such as principal component analysis (PCA), non-negative matrix factorization (NMF) and independent component analysis (ICA) employ matrix based approaches. Many biomedical signals exhibit higher-order structure. Reformatting the data tensor as a matrix and using classical matrix based BSS are bounded by rigid assumptions inherent in matrix analysis and are not always a good match for higher-order data. For maintaining structural information, higher-order tensors are very attractive. They generalize vectors and matrices to higher-order tables of numbers. This has motivated researchers to use tensor based blind source separation which enables capturing multiple interactions and couplings from the higher-order data, instead of standard pairwise interactions. Also, tensors and their corresponding tools display certain properties that are not available in the matrix domain [38, 158]. Uniqueness of tensor decomposition under mild conditions is one such strong property, where additional constraints are not needed to obtain solutions as compared to the matrix case [61, 62]. Tensor based blind source

separation is solved using tensor decomposition techniques. Canonical Polyadic Decomposition (CPD) is the most well known decomposition, which decomposes a tensor in $rank - 1$ terms. CPD is unique under mild conditions, which makes it a broadly applicable key tool for BSS. In many applications $rank - 1$ is very restrictive as it cannot model variations in the source, except for strength. In tensors, uniqueness of decomposition in $rank - 1$ terms can be extended to more general or realistic terms. A block term decomposition (BTD) generalises the $rank - 1$ terms in CPD to low multilinear rank terms. When the underlying data has only second-order structure, instead of using matrix-based approaches for blind source separation (BSS), matrix data can be converted to a higher-order tensor. This transformation is called tensorization and under certain conditions tensor methods will provide advantage over matrix analysis. For example, in blind source separation (BSS) problems, provided the source can be modelled or approximated by rational functions, tensorization followed by applying tensor decompositions has better performance compared to that of the matrix based counterpart [54].

Tensor decomposition techniques such as CPD are currently emerging as a standard tool and have already been widely used in telecommunication, array processing, chemometrics, psychometrics and exploratory data analysis [99, 163]. Biomedical applications include the study of brain networks [128], brain-computer interfaces [151] seizure localization, EEG [151] and Event-Related Potentials. Applications using more generalized tensor decompositions, such as BTD are still limited [89] and have not been explored widely. On the other hand, in biomedical applications it is still common practice to store data in matrices, even when they have higher-order structure, and hence important structural information is lost. With the exception of ICA, tensorization of matrix data has been limited to a few isolated cases [125]. It is clear that tensors have a great unexplored potential in biomedical data processing. This thesis aims at exploring tensor based blind source separation techniques for processing and analysing MRSI and MP-MRI data. We focus on two main application: residual water suppression in MRSI (Figure 1.2) and tumor tissue type differentiation from MRSI/MP-MRI (Figure 1.3). The research in this thesis is part of work package six (WP6) in the BIOTENSORS project, funded by ERC Advanced Grant: BIOTENSORS (n° 339804) and meets the following objectives:

Aim1: Tensor based blind source separation techniques for MRSI pre-processing. MRSI signals contain information for estimating metabolite concentrations from in-vivo in a non-invasive fashion. Along with clinically relevant components MRSI signals also contain unwanted components such as water, baseline etc. In general, residual water is suppressed before doing any analysis in a pre-processing step (Figure 1.2). The traditional model for an MRS(I) signal is a sum of damped exponentials. This model is still widely used

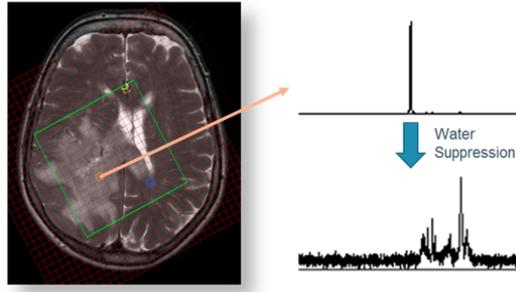


Figure 1.2: Residual water suppression from one of the voxel in the MRSI signal.

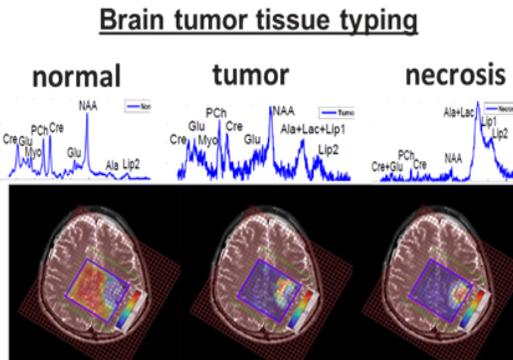


Figure 1.3: Nosologic images of normal, tumor and necrosis tissue types. Image adapted from: [111].

for filtering unwanted components (artefacts) originating from residual water [130]. Currently used matrix algorithms are based on Hankel expansions and operate on a voxel by voxel basis. Simultaneous analysis of MRSI signals, which originate from a 2D or 3D array of neighbouring voxels and share similar spectral profiles, has not been attempted. We aim to exploit the shared information present among neighbouring voxels and develop a new class of algorithms based on canonical polyadic decomposition to improve automation and enable all-at-once water suppression of the entire 2D MRSI set.

Aim2: Tensor based brain tumour tissue typing algorithms using MRSI and MP-MRI: Another major application of MRSI/MP-MRI signals we focus on is brain tumour tissue typing. The tumor region of glioblastoma multiforme (GBM) typically consists of several tissue types, which represent actively growing tumor, necrosis or normal brain tissue. In vivo MRSI has shown

an increasing power in the diagnosis of brain lesions and neurological disorders. For brain tumour patients in particular, the challenge is to classify MRSI voxels so that they recognise tumour types, grades and tissue heterogeneity. Blind source separation techniques are used to extract the tissue-specific profiles and their corresponding distribution from the MRSI data (Figure 1.3). Here, each source represent the MRS signal from a pure tissue type (healthy, actively growing tumour, necrotic tissue. . .). So far, only BSS approaches based on non-negative matrix factorization (NMF) or independent component analysis (ICA) [39, 150] have been used but they cannot handle the heterogeneity present in tissue or artefacts. Our objective is to develop CPD/block term decomposition (BTD) based algorithms to improve automation and brain tumour heterogeneity characterization.

Conventional MRI (cMRI) is a widely used imaging modality for tumor segmentation/localization. In recent years advanced modalities such as diffusion weighted imaging (DWI), perfusion weighted imaging (PWI) and MRSI are being used in Neurooncology. Studies have shown that additional structural, biological and biochemical information provided by these modalities will help in tumor characterization. In NMF based algorithms, the addition of these MRI modalities has proven to be beneficial for tumor segmentation [152]. This motivated us to also focus on extending the CPD based algorithms developed for MRSI to deal with MP-MRI data and to provide a more refined tissue characterization.

Aim3: Tensor decompositions applied to supervised tumour tissue segmentation using MRSI and MP-MRI: One of the problems with BSS techniques is the difficulty to automate the algorithm as the interpretation is left to the user. Most state-of-the-art automated algorithms for segmenting tumor regions are based on a supervised classification approach. Higher order structures occur naturally in both the data (e.g. 3D MRI scan, MP-MRI data) and in the algorithms (e.g. Convolution kernel in convolutional neural network is a 4D array). These higher order structures have not been exploited in the context of MRSI voxel classification or MP-MRI tumor segmentation. We aim to explore tensor decomposition methods such as multilinear singular decomposition (MLSVD) to exploit higher order structure in supervised MRSI voxel classification for handling over-fitting and to reduced computational complexity and in MP-MRI tumor segmentation algorithms to generate refined features.

1.2 Chapter-by-chapter overview

The schematic representation of this thesis is shown in Figure 1.4 and can be grouped into four parts: background materials, pre-processing and analysis of MRSI data, analysis of MP-MRI data and conclusion. Chapters 4, 5 and 7 resort under the umbrella of tensor based blind source separation. Chapters 6 and 8 reveal how to exploit tensor decomposition in supervised algorithms.

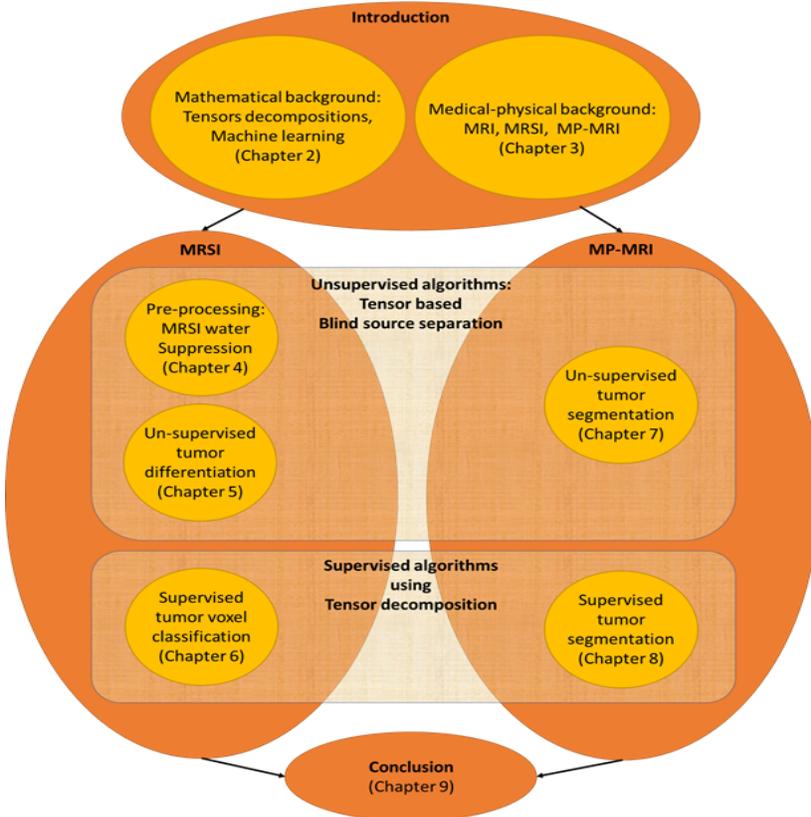


Figure 1.4: Schematic overview of the thesis.

Chapter 2 introduces the mathematical background needed for this thesis: basic concepts of tensor decompositions and machine learning. Commonly used tensor decomposition methods which form the basis of all algorithms developed in this thesis are mentioned. In the second half we introduce blind source separation as an un-supervised machine learning algorithm. Different

matrix/tensor based BSS techniques that are relevant to this thesis are explained. Finally, basic concepts of random forest and convolutional neural networks are described in a section on supervised machine learning techniques.

Chapter 3 introduces the medical-physical background: basic concepts in MRI, MRSI and MP-MRI. Different MRI modalities and their benefits in the diagnosis and management of glioma patients are described. Next, MP-MRI datasets from UZ Leuven on which tensor based BSS methods are applied are described. Acquisition parameters and pre-processing methods for all the MRI modalities are discussed. At last, the BRATS 2017 challenge dataset used in supervised brain tumor segmentation is discussed.

Chapter 4 introduces two tensor based methods for residual water suppression in MRSI. MRSI signals share common information among neighbouring voxels. Since traditional matrix-based approaches did not exploit such shared information, tensor based methods capable of exploiting such information were developed. In the first method MRSI is modelled as a sum of exponentials in the time domain and a Hankel-tensor based exponential data fitting approach is applied to water suppression. However, in the second method MRSI is modelled as a sum of rational functions in the frequency domain, where a Löwner-tensor based blind source separation technique is being developed for residual water suppression. The added value of simultaneous water suppression in MRSI using tensor decompositions over a matrix based individual voxel approach is assessed on both simulated and in-vivo MRSI datasets from UZ Leuven.

Chapter 5 focuses on tumor tissue typing in MRSI using tensor-based blind source separation. Representing the MRSI signal as a tensor in such a way that a low-rank structure can be exploited is important. To this end, a third-order \mathbf{xx}^T structured tensor construction is formulated. A non-negative canonical polyadic decomposition (NCPD) based algorithm was developed for extracting tissue specific patterns and their corresponding abundances. Schemes for initialization and automatic estimation of number of sources are developed. We hypothesize that the tensor based algorithm will be better at handling artefacts and will reveal more localized tissue distributions, thereby providing a more refined tissue characterization. The NCPD algorithm is applied to the in-vivo MRSI dataset from UZ Leuven to assess the advantages of using tensor based BSS compared to matrix based BSS methods (NMF).

Chapter 6 explores supervised algorithms for classification of voxels in MRSI. The NCPD algorithm in Chapter 5 will extract tissue specific patterns and their

corresponding abundances from MRSI. These can be further used to classify each voxel using prior knowledge. We try to improve the tumor characterization using two supervised algorithms. The first method is based on random forest, which uses a reduced spectrum from each voxel as input features. In the second approach, a convolutional neural network (CNN) architecture is developed to classify the individual voxels in MRSI into three classes: tumor, normal and bad quality. Additionally, Low-rank regularization based on multilinear singular value decomposition (MLSVD) is applied to the convolution layer of CNN for assessing the usability of tensor decomposition in supervised algorithms. The low-rank regularization will have fewer parameters to learn compared to a non-regularized convolution layer. Therefore, we expect it to handle over-fitting and result in reduced computational complexity. The advantage of using tensor based low-rank regularization in CNN is tested on an in-vivo MRSI dataset.

Chapter 7 extends the tensor based algorithm NCPD developed in Chapter 5 to handle the MP-MRI data and to perform tumor segmentation on it. The same tensor construction approach of chapter 5 is used, but now the MRSI-samples are replaced by well-chosen characteristic features quantified from a variety of MR modalities (called MP-MRI). A constrained CPD algorithm, where a non-negativity constraint is imposed on one of the factor matrices is developed for tissue characterization from MP-MRI signals. Similarly to Chapter 4, we analyse the added value of tensor based BSS over matrix based BSS using MP-MRSI datasets of high grade glioma patients from UZ Leuven.

Chapter 8 explores the applicability of tensor decomposition methods in extracting features for supervised tumor segmentation algorithms from MP-MRI data. Un-supervised algorithms developed in the previous chapter was not capable of segmenting tumor properly on a whole image. Therefore, we shifted from un-supervised to supervised algorithms for segmenting tumors using a 3D MP-MRI dataset. A superpixel wise two-stage random forest algorithm is developed. In the first stage, the whole tumor (enhancing tumor + necrosis + edema) is segmented and subsequently, in the second stage sub-compartments are segmented from the whole tumor. In both stages many features are extracted using multilinear singular value decomposition (MLSVD) to exploit the higher order structure present in the data. We expect tensor based feature extraction to be viable for tumor segmentation from MP-MRI and to provide more refined features than those obtained by averaging the matrixed parameters. The algorithms are trained and analysed using the BRATS 2017 challenge dataset.

Chapter 9 summarises the main findings of each chapter with respect to the objectives stated in the previous section. Additionally, future research directions and possible extensions of the current work are discussed.

Chapter 2

Mathematical background: tensor decomposition and machine learning

The methods developed throughout this thesis for pre-processing and analysing MP-MRI signals are all based on tensor decompositions. Therefore, the basic concept of tensors and various decomposition methods will be discussed in this chapter. In this thesis, both un-supervised and supervised machine learning algorithms are used for tumor tissue characterization and localization. Therefore, in the second half of the chapter, unsupervised blind source separation and supervised algorithms that are relevant for this thesis are discussed.

2.1 Tensor decomposition

Tensors are higher-order arrays. Vectors are first-order tensors, matrices are second-order tensors and arrays larger than second-order are called higher-order tensors. Vector, matrix and a third-order tensor are shown in Figure 2.1. In many applications the measured data contains higher-order structure, for example: 3D MRI scans, multi-lead EEG measured from different subjects. Processing the data as a tensor instead of unfolding it to a matrix (i.e. matricization) offers certain advantages such as: possibilities to obtain compact representations, the higher-order structure is preserved in the processed data, flexibility in the choice of constraints, generality of components that can be identified and

tensor decompositions can be unique under mild conditions without imposing additional constraints [61, 63, 38]. In this section we discuss the tensor tools that are used in this thesis.

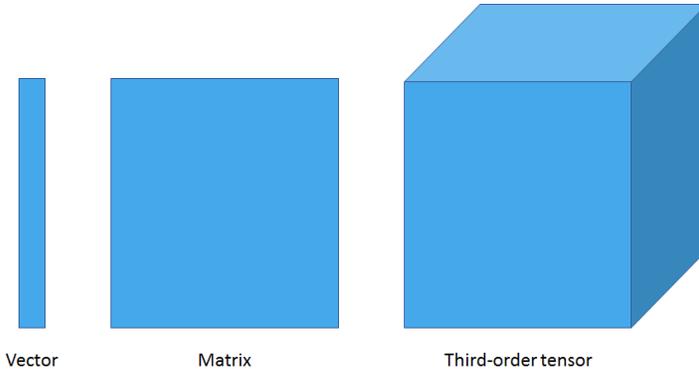


Figure 2.1: Graphical representation of vector, matrix and third-order tensor.

2.1.1 Notations and tensor preliminaries

Tensors, denoted by calligraphic letters, e.g., \mathcal{A} , are higher-order generalizations of vectors (denoted by boldface lowercase letters, e.g., \mathbf{a}) and matrices (denoted by boldface uppercase letters, e.g., \mathbf{A}). Scalars are written as italic lowercase letters, e.g., a . The entry with row index i and column index j of a matrix $\mathbf{A} \in \mathbb{C}^{I \times J}$ is denoted by a_{ij} . Likewise, the (i_1, i_2, \dots, i_N) th entry of an N th-order tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $a_{i_1 i_2 \dots i_N}$. The j th column of a matrix $\mathbf{A} \in \mathbb{C}^{I \times J}$ is denoted by \mathbf{a}_j . The superscripts \cdot^T , \cdot^H , \cdot^{-1} and \cdot^\dagger represent the transpose, complex conjugated transpose, inverse and pseudo inverse, respectively. The symbol \otimes and \circ denotes the outer product and Hadamard product, respectively. The outer product $\mathcal{A} \otimes \mathcal{B}$ of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ and a tensor $\mathcal{B} \in \mathbb{C}^{J_1 \times J_2 \times \dots \times J_M}$ is the tensor defined by: $(\mathcal{A} \otimes \mathcal{B})_{i_1 \dots i_N j_1 \dots j_M} = a_{i_1 \dots i_N} b_{j_1 \dots j_M}$.

The order of a tensor is the number of indices required to represent an element in the tensor. Vectors and matrices are tensors of order one and two, respectively. Fibers are the higher-order analogue of matrix rows and columns [99]. A mode- n fiber is defined by fixing every index except the n^{th} index. Third-order tensors have column, row, and tube fibers as shown in Figure 2.2. Similarly, slices are second-order sections of a tensor obtained by fixing all but two indices. The process of rearranging the elements of a tensor into a matrix is called

matricization or unfolding. In mode- n matricization of a tensor, the columns of the matrix contains the mode- n fibers as shown in Figure 2.3.

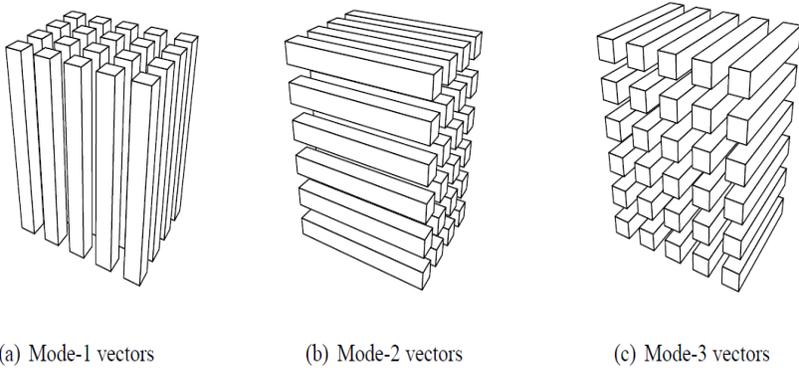


Figure 2.2: Mode-1, mode-2 and mode-3 fibers of third-order tensor. Source: [91]

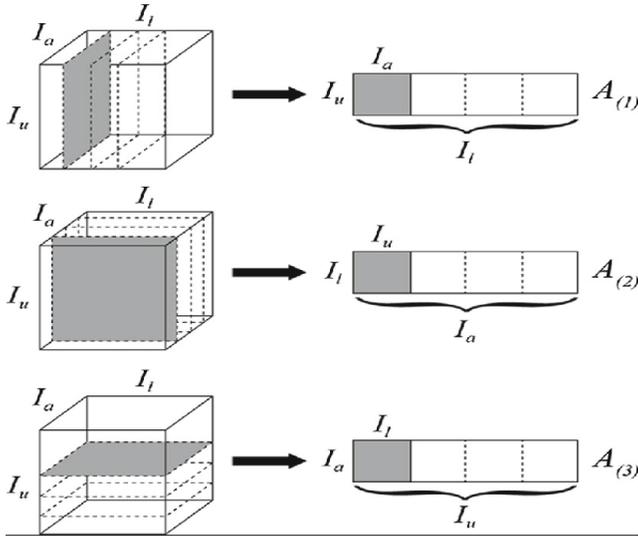


Figure 2.3: Graphical representation of a third-order tensor matricization in all three modes. Source: [171]

Tensor-matrix product

A tensor can be multiplied by a matrix in mode n , which is called tensor n -mode product. Given a tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, its n -mode product with the matrix $\mathbf{U} \in \mathbb{C}^{J \times I_n}$ is denoted as $\mathcal{T} \times_n \mathbf{U}$, will result in a tensor \mathcal{P} of size $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$, defined as

$$\begin{aligned} \mathcal{P}_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} &= (\mathcal{T} \times_n \mathbf{U})_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} \\ &= \sum_{i_n=1}^{I_n} t_{i_1 i_2 \dots i_n} u_{j i_n} \end{aligned}$$

In terms of unfolding, it can be written as $\mathbf{P}_{(n)} = \mathbf{U} \mathbf{X}_{(n)}$. When a tensor specifies a multilinear operator, the n -mode product with an invertible matrix of size $I_n \times I_n$ is related to a change of basis [99].

2.1.2 Tensorization: Löwner and Hankel matrices/tensors

When the data has higher-order structure tensor methods can be directly applied on the data. Certain manipulation on the original data can also lead to a tensor and the procedure of creating a data tensor from lower-dimensional data is referred to as tensorization. Generating a tensor from the data can be broadly classified into four groups [38]:

- From lower to higher-order structure: for example, one-way exponential signal can be arranged in a Hankel matrix or a Hankel tensor. Similarly, a Löwner tensor can be obtained by stacking Löwner matrices constructed from multiple one-way rational signals.
- Mathematical construction: for instance, N^{th} -order moments/cumulants obtained from a vector-valued random variable form an N^{th} -order tensor. Also, a multichannel EEG/ECG (*channel* \times *time*) can be transformed using time-frequency or wavelet representations into *channel* \times *time* \times *frequency* or *channel* \times *time* \times *scale* tensor, respectively.
- Experiment design: here data from different modules can be stacked into a tensor. For instance, MP-MRI images from MRI, DWI, MRSI and PWI can be stacked to form a third-order tensor.
- Naturally occurring tensor data: some of the measured/generated data exhibit higher-order structure. For example, 2D-MRSI signal (*x* – *spatial* \times *y* – *spatial* \times *spectrum*) and 3D MRI scans.

A more detailed explanation of different higher-order structures and tensorization are available in [53]. In this section two types of tensors which fall under the category of generating tensor from lower-order data (tensorization) are discussed.

Löwner matrix/tensor: while the concept of Löwner matrices is highly acknowledged in the domain of system identification [9, 10], it is not well known in other application domains. In a recent study, Löwner matrices have been used in a BSS context to separate (approximations by) rational functions [54].

Suppose a function $f(t) \in \mathbb{C}$ is given, evaluated in the point set $T = \{t_1, t_2, \dots, t_N\}$. In order to construct the Löwner matrix, the point set T is partitioned into two distinct point sets, $X = \{x_1, x_2, \dots, x_I\}$ and $Y = \{y_1, y_2, \dots, y_J\}$ with $I + J = N$. The elements of the Löwner matrix $\mathbf{L} \in \mathbb{C}^{I \times J}$ are then defined as

$$\forall i, j : l_{ij} = \frac{f(x_i) - f(y_j)}{x_i - y_j}.$$

We thus obtain the following matrix:

$$\mathbf{L} = \begin{bmatrix} \frac{f(x_1) - f(y_1)}{x_1 - y_1} & \frac{f(x_1) - f(y_2)}{x_1 - y_2} & \dots & \frac{f(x_1) - f(y_J)}{x_1 - y_J} \\ \frac{f(x_2) - f(y_1)}{x_2 - y_1} & \frac{f(x_2) - f(y_2)}{x_2 - y_2} & \dots & \frac{f(x_2) - f(y_J)}{x_2 - y_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{f(x_I) - f(y_1)}{x_I - y_1} & \frac{f(x_I) - f(y_2)}{x_I - y_2} & \dots & \frac{f(x_I) - f(y_J)}{x_I - y_J} \end{bmatrix}.$$

For partitioning, a parameter α is often used in the literature with $I = \alpha$ and $J = N - \alpha$. Matrix \mathbf{L} will be square when N is even and $\alpha = N/2$. The interleaved partitioning with $X = \{t_1, t_3, \dots\}$ and $Y = \{t_2, t_4, \dots\}$ and the block partitioning with $X = \{t_1, \dots, t_I\}$ and $Y = \{t_{I+1}, \dots, t_N\}$ are some of the commonly used partitionings. Rational function and Löwner matrix exhibit an important property, where a Löwner matrix constructed from a rational function of degree R will have a rank R [9, 116]. This property is valid for any point set partitioning.

Given K functions $f_i(t)$ evaluated on the same set of N points, a Löwner matrix \mathbf{L}_i can be computed for each function. By stacking the different matrices \mathbf{L}_i in a tensor along the third mode, a Löwner tensor $\mathcal{L} \in \mathbb{C}^{I \times J \times K}$ is obtained.

Hankel matrix/tensor: Hankel matrices are used in many applications such as system identification, coding theory. For a function $f(t) \in \mathbb{C}$ evaluated at N distinct points $T = \{t_1, t_2, \dots, t_N\}$, the elements of a $I \times J$ Hankel matrix with $I + J - 1 = N$ are defined as

$$\forall i, j : h_{ij} = f(t_{i+j-1}).$$

and in matrix form it is represented as:

$$\mathbf{H} = \begin{bmatrix} f(t_1) & f(t_2) & \dots & f(t_{J-1}) & f(t_J) \\ f(t_2) & f(t_3) & \dots & f(t_J) & f(t_{J+1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f(t_{I-1}) & f(t_I) & \dots & f(t_{I+J-3}) & f(t_{I+J-2}) \\ f(t_I) & f(t_{I+1}) & \dots & f(t_{I+J-2}) & f(t_{I+J-1}) \end{bmatrix}.$$

A Hankel matrix constructed from a sum of R complex exponentials will have a rank- R [104]. This important property is the cornerstone of Hankel based harmonic retrieval, which is used in Chapter 4. Similarly to the Löwner tensor, a Hankel tensor $\mathcal{H} \in \mathbb{C}^{I \times J \times K}$ can be constructed from K functions $f_i(t)$ by stacking the Hankel matrices \mathbf{H}_i in a tensor along the third mode.

2.1.3 Canonical polyadic decomposition

A data matrix \mathbf{X} can be decomposed into two factor matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$.

$$\mathbf{X} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \triangleq \mathbf{A}^{(1)} \mathbf{A}^{(2)T}.$$

Such a decomposition is not unique as it can have infinitely many combinations of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$:

$$\begin{aligned} \mathbf{X} &= (\mathbf{A}^{(1)} \mathbf{B})(\mathbf{B}^{-1} \mathbf{A}^{(2)T}) \\ &= \tilde{\mathbf{A}}^{(1)} \tilde{\mathbf{A}}^{(2)T}, \end{aligned}$$

where \mathbf{B} is an invertible square matrix of size $(R \times R)$. Standard matrix factorizations methods such as the QR-factorization and singular value decomposition (SVD) generate unique factors due to hard and restrictive constraints such as triangularity and orthogonality. Other factorizations methods such sparse component analysis non-negative factorization estimate factors using constraints that exploit certain properties of those factors. However, uniqueness is not always guaranteed.

Polyadic Decomposition (PD) approximates an N^{th} -order tensor as a sum of rank-1 tensors [99, 38]. For a tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ PD is defined as

$$\mathcal{T} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \otimes \dots \otimes \mathbf{a}_r^{(N)} \triangleq \left[\left[\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \right] \right].$$

where $\mathbf{A}^{(n)}$ is the n^{th} -mode factor matrix and R is the number of rank-one tensors. If R is minimal, the decomposition becomes canonical (CPD) and the rank of \mathcal{T} is defined as R . The rank of a tensor is defined as the minimum number of rank-one tensors, whose sum generate the exact tensor. Figure 2.4 shows the CPD of a third-order tensor.

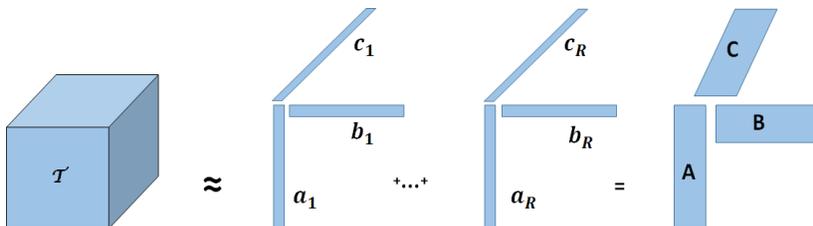


Figure 2.4: Graphical representation of a canonical polyadic decomposition for a third-order tensor.

The advantage of the CPD model is its uniqueness up to permutation and scaling under mild conditions. Several deterministic uniqueness conditions have been derived with increasing generality [61, 63, 38]. One such condition useful in this thesis is defined as [65]:

For a third-order tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$, for $K \geq R$, the CPD is unique when,

1. The factor matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ have full column rank. For factor matrices to be full column rank it is necessary for the tensor rank to be $R \leq \min(I, J)$.
2. The third factor matrix $\mathbf{A}^{(3)}$ does not contain proportional columns.

Generally, CPD is computed by minimizing the Frobenius norm of the difference between the given data tensor and its CP approximation:

$$\min_{\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}} \|\mathcal{T} - \llbracket \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket\|^2, \tag{2.1}$$

where $\|\cdot\|$ denotes the Frobenius-norms. Alternating least squares (ALS) is the simplest and most widely used algorithm for computing CPD [33, 3], where each factor matrix is computed alternately using least squares by fixing all the other factor matrices. Algorithms using simultaneous generalized Schur decomposition [49], optimization based techniques [166, 141] and many other approaches [99, 38] are also available for computing CPD.

In this thesis we have used the default CPD by nonlinear least squares (CPD-NLS) algorithm available in Tensorlab software package [185]. CPD-NLS

algorithm solves the optimization problem in (2.1.3) using nonlinear least squares methods such as Gauss–Newton or Levenberg–Marquardt. We have used the Gauss–Newton method with dogleg trust-region algorithm for our application. At each iteration, the parameters are updated and the step is calculated by solving a linearised approximation of the cost function. The cost of each iteration is reduced by exploiting the multilinear structure and using a preconditioned conjugate gradient algorithm to determine the Gauss–Newton step [166]. The computational cost of the algorithm is $\mathcal{O}(2(N + \text{it}_{\text{gn}})R \prod I_n + \text{it}_{\text{cg}}(\frac{5}{2})N^2R^2 + \frac{1}{3}NR^3 + 8R^2 \sum I_n)$ flops/iteration, where it_{gn} is the number of internal iteration per step and it_{cg} is the number of iteration of the preconditioned conjugate gradient algorithm. A more detailed explanation of this algorithm along with other optimization based approaches are available in [166].

2.1.4 Multilinear singular value decomposition

Multilinear singular value decomposition (MLSVD) can be considered as generalization of matrix singular value decomposition to higher order tensors. For an N^{th} -order $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, MLSVD is defined as [48]:

$$\mathcal{T} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \dots \times_N \mathbf{U}^{(N)}, .$$

where $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ is a unitary matrix and \mathcal{S} is $(I_1 \times I_2 \times \dots \times I_N)$ -tensor of which the subtensors $\mathcal{S}_{i_n=\alpha}$ obtained by fixing the n^{th} index to α has the following properties:

1) *all-orthogonality*: two subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all values of n , α and β subject to $\alpha \neq \beta$:

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when } \alpha \neq \beta,$$

2) *ordering*:

$$\|\mathcal{S}_{i_n=1}\| \geq \|\mathcal{S}_{i_n=2}\| \geq \dots \geq \|\mathcal{S}_{i_n=I_n}\|$$

Frobenius-norms of subtensors $\|\mathcal{S}_{i_n=i}\|$, denoted by $\sigma_i^{(n)}$ are the n -mode singular values of \mathcal{T} . Figure 2.5 shows the MLSVD of a third-order tensor. The factor matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ obtained from MLSVD of a third-order tensor span the column, row, and third-mode space, respectively. In general, the n^{th} factor matrix spans the n -mode vector space.

MLSVD can be computed using singular value decomposition. The n -mode factor matrix $\mathbf{U}^{(n)}$ can be obtained as the matrix holding the left singular vectors of $\mathbf{T}(n)$, where $\mathbf{T}(n)$ is the n -mode matrix unfolding (matricization) of the tensor \mathcal{T} . The core tensor \mathcal{S} can then be computed by applying all n -mode

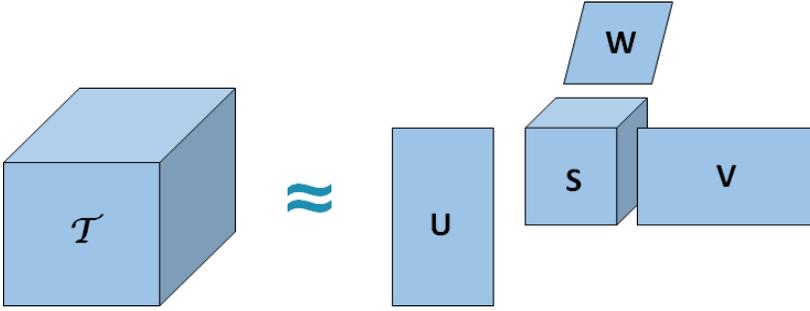


Figure 2.5: Graphical representation of a MLSVD for a third-order tensor.

products on \mathcal{T} with the inverse of corresponding factor matrix $\mathbf{U}^{(n)}$:

$$\mathcal{S} = \mathcal{T} \times_1 \mathbf{U}^{(1)H} \times_2 \mathbf{U}^{(2)H} \times_3 \cdots \times_N \mathbf{U}^{(N)H}.$$

Computing the MLSVD of an N^{th} -order tensor consists of computing N different matrix SVDs of unfolded matrices with size: $I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N$ ($1 \leq n \leq N$) and N n -mode tensor matrix multiplications. In this thesis MLSVD is computed using the *mlsvd*-function in Tensorlab toolbox [185].

A data tensor can be approximated by discarding the multilinear singular vectors and slices of the core tensor that correspond to small multilinear singular values to obtain a low multilinear rank approximation (LMLRA). Although the truncation results in a good estimate it may not be necessarily optimal in the least squares sense [38]. Algorithms based on alternating least squares or optimization techniques can be used to obtain best approximation. MLSVD has many applications, it can be considered as a multilinear extension of PCA [103]. MLSVD has been used for compression, classification, feature extraction, subspace-based harmonic retrieval, signal enhancement and many more [103, 117, 181, 140, 81]. In this thesis we have used MLSVD for harmonic retrieval, compression of fourth-order tensor and to extract features for classification.

2.1.5 Block term decomposition

Block Term Decomposition (BTD) is a recently introduced tensor decomposition method. BTD generalises CPD, where a tensor is approximated by a sum of low multilinear rank terms as opposed to rank-1 terms in CPD [47, 46, 50]. For

a N^{th} -order $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$, BTD is defined as:

$$\mathcal{T} \approx \sum_{r=1}^R \mathcal{S}^{(r)} \times_1 \mathbf{U}^{(r,1)} \times_2 \mathbf{U}^{(r,2)} \times_3 \dots \times_N \mathbf{U}^{(r,N)}.$$

where $\mathcal{S}^{(r)} \in \mathbb{C}^{J_1^{(r)} \times J_2^{(r)} \times \dots \times J_N^{(r)}}$ is the core tensor in the r^{th} term and $\mathbf{U}^{(r,n)} \in \mathbb{C}^{I_n \times J_n^{(r)}}$ is the n^{th} factor matrix in the r^{th} term of the BTD. Consider the Hankel/Löwner tensor, if the Hankel/Löwner matrix in each slice is constructed from a linear combination of degree-1 exponential/rational functions CPD will be able to extract individual components. However, for a linear combination of low degree exponential/rational functions, rank-1 term of CPD will not be sufficient to extract individual components. This requires low rank terms corresponding to the degree of exponential/rational functions for extracting individual components, in such applications BTD will be more useful than CPD.

In this thesis we have considered one particular case, decomposition in R rank- $(L_r, L_r, 1)$ block terms. In rank- $(L_r, L_r, 1)$ -BTD each of the low multilinear rank terms can be represented by a outer product of a rank- L_r matrix and a non-zero vector. For a third-order tensor $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ the rank- $(L_r, L_r, 1)$ BTD is given by:

$$\mathcal{T} \approx \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{c}_r$$

where $\mathbf{A}_r \in \mathbb{C}^{I_1 \times L_r}$ and $\mathbf{B}_r \in \mathbb{C}^{I_2 \times L_r}$ are full rank matrices. Figure 2.6 visualizes rank- $(L_r, L_r, 1)$ BTD of a third-order tensor. Algorithms such as alternating least squares [50] and optimization based methods [166] are available for computing BTD.

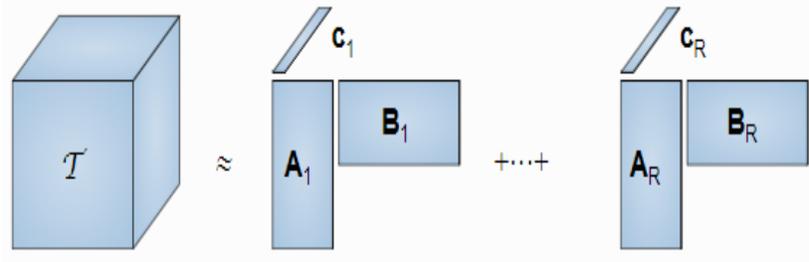


Figure 2.6: Graphical representation of a rank- $(L_r, L_r, 1)$ block Tensor decomposition of a third-order tensor. Source: Tensorlab documentation [185]

2.1.6 Structured tensor decompositions

Structured tensor decompositions are special cases where additional constraints or structure are imposed on the factor matrices [167]. Some of the common constraints/structure include orthogonality, nonnegativity, Hankel structure and polynomial structure. Imposing constraints on factor matrices can help to relax the uniqueness condition and simplify computation [168]. They also enhance the accuracy and robustness with respect to noise, help in implementing application-dependent prior knowledge and results in better interpretability. In this thesis we have imposed constraints to add application-dependent prior knowledge.

The structured data fusion is a broad framework [167] available in Tensorlab [185], which allows one to use different tensor decomposition models, perform coupled tensor decomposition with shared factors, incorporate constraints on factor matrices (e.g non-negative, Hankel structure) and apply regularization to factor matrices. In this thesis we have used structured data fusion to impose non-negative constraints, impose that a factor matrix is the same across two different modes and to apply l_1 regularization on a factor matrix in Chapter 5 and 7 for tumor tissue typing.

2.1.7 Applications of tensor decomposition

Biomedical engineering is an emerging field where applied mathematical tools are extensively used. In this thesis we use tensor decompositions as core building blocks of algorithms applied to pre-processing and analysis of Magnetic resonance spectroscopic imaging (MRSI) and multiparametric MRI (MP-MRI) signals. Tensor decompositions are also used in many other areas of signal processing and data analysis like audio and speech processing, biomedical engineering, chemometrics, bioinformatics, genetics and machine learning, to name a few. In wireless communication, it is used in sensor array processing [159], equalization [44] and space-time multiplexing codes [43]. In hyperspectral imaging, tensor decompositions, are used for compression [67], anomaly detection [110], denoising [193] and image restoration [187]. In the field of biomedical engineering, they are used to analyse Electroencephalography (EEG) [123], Electrocardiography (ECG) [126], functional magnetic resonance imaging (fMRI) [155] signals, gene expression traits [87], gait [139] and many more. For a detailed overview of tensor decompositions and its applications the reader is referred to the overview papers [99, 38, 160].

2.2 Machine learning

Machine learning is a field of computer science which deals with the design and study of algorithms that can learn from and make predictions on data. Machine learning has been used in many areas and has very wide range of applications such as medical image analysis, search engine, health care, computer vision, bioinformatics, astronomy, financial analysis to name a few. Machine learning algorithms can be grouped into supervised and un-supervised methods. Given a training set with input X and labelled response Y , a supervised algorithm learns a mapping function $h : X \rightarrow Y$ such that the predicted response is close to the real response. Un-supervised methods work only on input data to model the underlying structure or distribution in the data, which helps in interpreting and analysing the data. Supervised algorithms result in better decision boundaries and they can achieve very high performance in classification tasks. They require large training datasets and high computation time. The main advantage of un-supervised algorithms is their adaptability, where it can be applied to new data with little to no modifications. However, in supervised methods we have to start from scratch even for small modifications in the data. Also, supervised algorithms can be fully automated, whereas in un-supervised techniques the interpretation is left to the user. Supervised learning methods include support vector machines (SVM), random forest (RF), deep neural network (DNN), k-nearest neighbor algorithm, naive Bayes, linear discriminant analysis (LDA) and many more. Some of the un-supervised methods include clustering methods such as k-means, Gaussian mixture models and hierarchical clustering, blind signal separation techniques such as principal component analysis (PCA), independent component analysis (ICA) and non-negative matrix factorization (NMF), neural networks based methods such as autoencoders and deep belief nets.

2.2.1 Un-supervised: blind source separation

Blind source separation is the recovery of original signals from the mixture without (or as little as possible) prior knowledge of the mixing process or the original signals. Given a set of observed signals $\mathbf{S} \in \mathbb{C}^{N \times K}$, the BSS problem consists of identifying the mixing matrix $\mathbf{H} \in \mathbb{C}^{K \times R}$ and/or the original source signals in $\mathbf{W} \in \mathbb{C}^{N \times R}$ based on the following linear model:

$$\mathbf{S} = \mathbf{W}\mathbf{H}^T, \quad (2.2)$$

with K the number of observed signals, R the number of source signals and N the number of samples per signal. By itself, the solution cannot be uniquely identified as different working hypotheses lead to different solutions (at least for the non-trivial cases $R > 1$). Different working assumptions have been used

before such as mutual independence leading to independent component analysis, non-negativity leading to non-negative matrix factorization and so on. BSS techniques that are used in this thesis are discussed briefly in the next section.

Non-negative matrix factorization

Non-negative matrix factorization (NMF) approximates a non-negative input matrix by a product of two non-negative factor matrices. Given a non-negative matrix $\mathbf{X} \in \mathbb{R}^{N \times K}$, its NMF is defined as

$$\mathbf{X} \approx \mathbf{W}\mathbf{H}$$

where $\mathbf{W} \in \mathbb{R}_+^{N \times R}$ and $\mathbf{H} \in \mathbb{R}_+^{R \times K}$ are non-negative factor matrices. R is the rank denoting the number of columns in \mathbf{W} , typically $R \ll N$ or $R \ll K$ resulting in a low-rank (rank R) approximation. The columns of \mathbf{W} are called the NMF sources, which contains the basic/signature components. The second factor matrix \mathbf{H} contains the weights/abundances, where each columns represents the weights of R sources that approximates the corresponding column of \mathbf{X} . The factor matrices are obtained by minimizing a cost function. The Frobenius norm of the difference between the input matrix \mathbf{X} and its approximation $\mathbf{W}\mathbf{H}$ is the most commonly used cost function:

$$\min_{\mathbf{W}, \mathbf{H}} f(\mathbf{W}, \mathbf{H}) = \min_{\mathbf{W}, \mathbf{H}} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|^2 \quad \forall_{i,j} : w_{i,j} \geq 0, h_{i,j} \geq 0.$$

Other cost functions e.g. based on Kullback-Leibler divergence are also available in the literature. In general, NMF is an NP-hard problem and the solutions are not unique [182]. Therefore, practical algorithms aim at finding locally optimal solutions instead of globally optimal one. When one factor matrix is fixed NMF reduces to a convex non-negative least squares problem (NNLS) [76]. Most of the algorithms exploit this fact to estimate the factor matrices iteratively by fixing one factor matrix and solving the convex non-negative least squares problem for optimizing the other and vice versa [76]. Some of the most commonly used NMF algorithms include multiplicative update [108, 77], hierarchical alternating least squares [36, 77] and convex NMF [59]. NMF is used in many areas such as image processing [108], text mining [156], computational biology [57], brain tumor differentiation [135] and many more. In this thesis, the NMF algorithm is applied to tumor differentiation in chapter 5. A schematic representation of NMF in the context of MRSI tissue type differentiation is shown in Figure 2.7.

Hierarchical non-negative matrix factorization

Hierarchical non-negative matrix factorization (hNMF) is an extension of NMF, where NMF is applied sequentially. There are many variants of hNMF used in

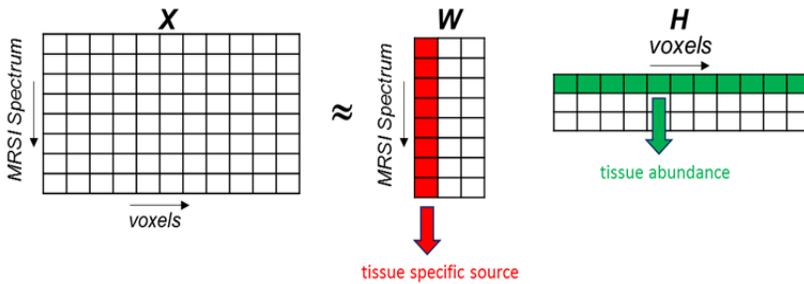


Figure 2.7: Graphical representation of NMF algorithm used for MRSI tissue differentiation.

hyperspectral imaging [78], massive data clustering [97] and text mining [174]. In this thesis we restrict our discussion to hNMF algorithm specifically designed for tissue type differentiation in MRSI [111]. hNMF consists of 3 basic steps: in the first step rank-2 NMFs is applied to obtain sources and their corresponding abundances. The sources are then assigned to normal and abnormal tissue. In the second step, several rank-2 NMF are performed on the subset of signals. For each subset, the signals are selected by applying a threshold on the abundances corresponding to the abnormal source from the first stage. The threshold is varied to obtain different subsets. Out of the several rank-2 NMFs, the best result is chosen based on the correlation between normal source from the first stage and the sources from the second stage. In the final stage, the normal source from the first stage and two abnormal sources from the second stage are combined and their corresponding abundances are estimated using non-negative least squares (NNLS). A schematic representation of the entire hNMF algorithm is shown in Figure 2.8.

The hNMF algorithm for tissue type differentiation in MRSI is used in Chapter 5 for comparison with the tensor based blind source separation algorithm. In the reminder of the thesis we refer this algorithm as hNMF. In [152], the hNMF algorithm is modified to handle multi-parametric magnetic resonance imaging (MP-MRI) data and also to provide a more refined tissue characterization. The modification is mainly done in the second and third stage as shown in Figure 2.9. In the second stage, several rank- r NMFs are applied on two sets of voxels, iteratively. At each iteration two set of voxels are selected based on the voxel-wise abundance ratio maps from the first stage. In the third stage the tissue sources obtained from both sets are re-combined to calculate the abundances over the whole ROI using non-negative least-squares fitting (NNLS). Out of the several iterations the best NMFs corresponding to the two sets is selected based on a criterion that promotes sparsity to the voxel-wise tissue

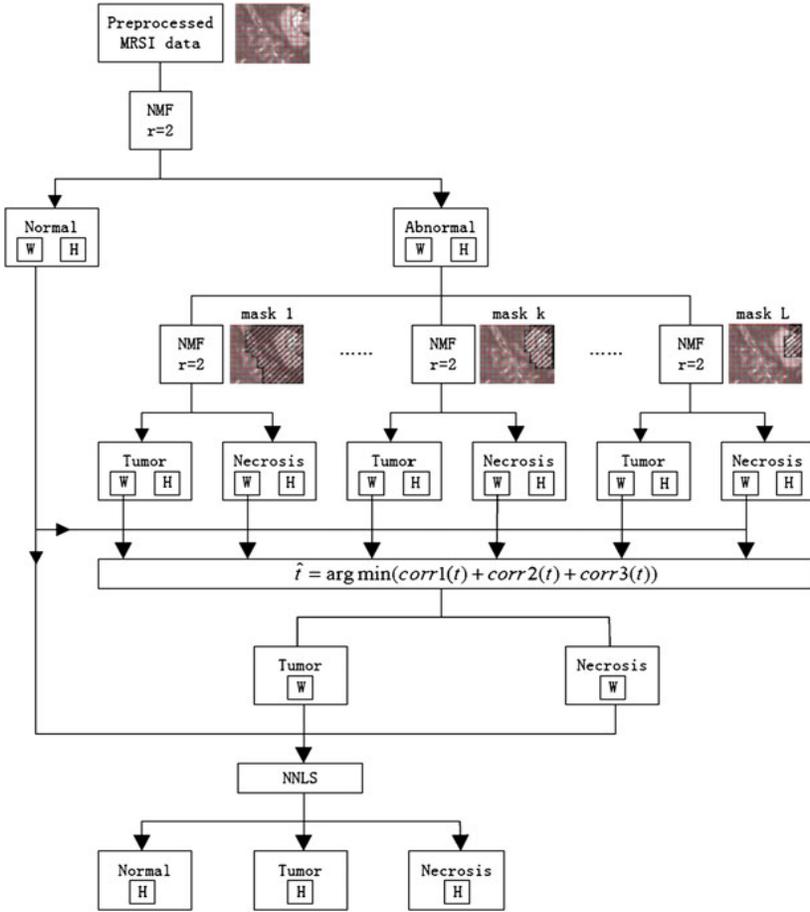


Figure 2.8: Schematic representation of hNMF algorithm for MRSI tissue differentiation. Source: [111]

abundances. This algorithm is used in Chapter 7 for comparison with the tensor based blind source separation algorithm and we refer to it as "MP-MRI hNMF" to differentiate it from the hNMF algorithm.

Löwner-based blind source separation of (approximations by) rational functions

Rational functions are formed by algebraic fractions with polynomials in the numerator and denominator. If \mathbf{L} is a Löwner matrix constructed from a

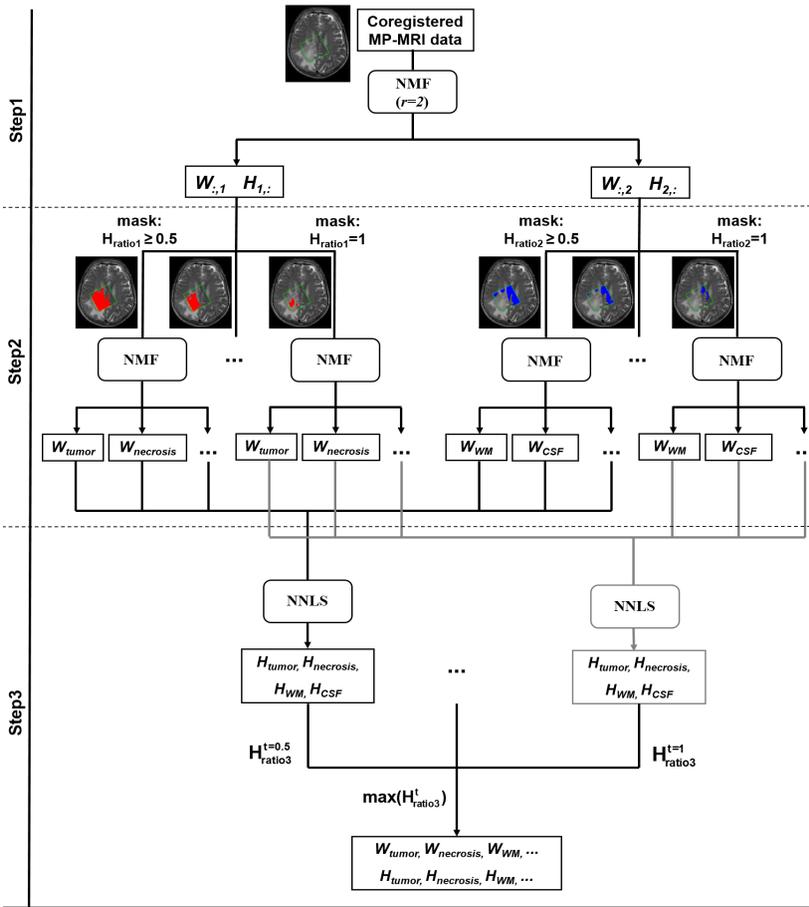


Figure 2.9: Schematic representation of hNMF algorithm for tissue differentiation in MP-MRI. Source: [152]

rational function of degree R , it has been proven that \mathbf{L} has rank R [9, 116]. The Löwner-based blind source separation technique has been developed for any arbitrary low degree rational functions using rank- $(L_r, L_r, 1)$ -BTD [54]. Since our application deals with rational functions of degree-1, we limit the discussion in this chapter to rational functions of degree 1, meaning that the numerator and denominator are linear functions. For a rational function of degree 1 its corresponding Löwner matrix will have rank-1, this is easy to verify: $f(t) = \frac{c}{t-p}$

gives $L_{i,j} = -c \cdot \frac{1}{x_i - p} \cdot \frac{1}{y_j - p}$, which is a rank-1 structure:

$$\mathbf{L} = -c \cdot \begin{bmatrix} \frac{1}{x_1 - p} \\ \vdots \\ \frac{1}{x_I - p} \end{bmatrix} \begin{bmatrix} \frac{1}{y_1 - p} & \cdots & \frac{1}{y_J - p} \end{bmatrix}.$$

Consider now the construction of two tensors $\mathcal{L}_{\mathbf{S}} \in \mathbb{C}^{I \times J \times K}$ and $\mathcal{L}_{\mathbf{W}} \in \mathbb{C}^{I \times J \times R}$. The tensors $\mathcal{L}_{\mathbf{S}}$ and $\mathcal{L}_{\mathbf{W}}$ contain Löwner matrices along the third mode constructed from the observed signals from \mathbf{S} , and the source signals from \mathbf{W} , respectively. Following the linear model (2.2.1), the tensor $\mathcal{L}_{\mathbf{S}}$ can be expressed as

$$\mathcal{L}_{\mathbf{S}} = \sum_{r=1}^R \mathbf{L}_{\mathbf{w}_r} \otimes \mathbf{h}_r. \quad (2.3)$$

If each source signal is an evaluated rational function (or can be approximated by an evaluated rational function), each matrix $\mathbf{L}_{\mathbf{w}_r}$ has (approximately) rank 1 [54]. Hence, each term in (2.2.1) has (approximately) rank 1 and the tensor $\mathcal{L}_{\mathbf{S}}$ has CPD structure with rank R .

To solve the BSS problem from (2.2.1) under the deterministic assumption of rationality, a CPD can be computed of $\mathcal{L}_{\mathbf{S}}$ with rank R . The factor matrix $\hat{\mathbf{H}}$ in the third mode is then an estimate of the mixing matrix \mathbf{H} . The source signals can be recovered as $\hat{\mathbf{W}} = \mathbf{S} \left(\hat{\mathbf{H}}^T \right)^\dagger$, provided \mathbf{H} has full column rank, which requires: $R \leq N$. It is also possible to obtain the source signals from only the estimated factor matrices $\hat{\mathbf{L}}_{\mathbf{w}_r}$ using a more complicated technique as described in [54]. Note that the two indeterminacies of the CPD are also present in the BSS: the source signals (factor vectors) are recovered up to permutation and scaling. Additionally, in BSS the source signals can only be recovered upto a constant term. In this thesis we estimate the poles of the rational function from the estimated matrices $\hat{\mathbf{L}}_{\mathbf{w}_r}$ instead of the source signals.

It remains to show that the separation is unique. This property guarantees that the recovered source signals are (estimates of) the original source signals. This uniqueness problem boils down to the CPD uniqueness given the special rational structure of the factor vectors. In [54, Theorem 4], it has been proven that the CPD is unique (up to permutation and scaling of the factor vectors) under following conditions:

1. The poles of the rational functions of the different source signals are distinct.
2. The third factor matrix \mathbf{H} does not contain proportional columns.

3. The signal length N and the rank R satisfy: $\lfloor \frac{N+1}{2} \rfloor \geq R$.

Note that uniqueness can still be guaranteed for underdetermined mixtures with fewer observed signals than source signals. Generic conditions for the BSS settings are given in [64].

In this thesis, Löwner-based blind source separation is used to develop a tensor-based residual water suppression algorithm in Chapter 4.

Hankel-based exponential data fitting

Single-channel signal In the single channel case only one signal is available (one column of \mathbf{S} in (2.2.1)). Let s_k be a time series modelled as sum of complex exponentials:

$$\mathbf{s}_{kn} = \sum_{r=1}^R c_r z_r^n, \quad \forall n = \{0, 1, \dots, N-1\}, \quad (2.4)$$

where R is the number of complex exponentials, z_r and c_r are complex pole and amplitude, respectively. Hankel singular value decomposition can be used to estimate the complex poles and amplitudes of the signal [13, 180]. A $I \times J$ Hankel matrix \mathbf{H} with $I + J = N$ constructed from the samples of \mathbf{s}_k can be decomposed as:

$$\mathcal{H} = \begin{bmatrix} 1 & \dots & 1 \\ z_1 & \dots & z_1 \\ \vdots & \ddots & \vdots \\ z_1^{I-1} & \dots & z_R^{I-1} \end{bmatrix} \begin{bmatrix} c_1 & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & c_R \end{bmatrix} \begin{bmatrix} 1 & \dots & z_1^{J-1} \\ 1 & \dots & z_2^{J-1} \\ \vdots & \ddots & \vdots \\ 1 & \dots & z_R^{J-1} \end{bmatrix} = \mathbf{PCL}^T \quad (2.5)$$

where \mathbf{P} and \mathbf{L} are Vandermonde matrices and $\mathbf{C} \in \mathbb{C}^{K \times R}$ contains the complex amplitudes. This decomposition is called a Vandermonde decomposition and the poles z_r are called generators. There is no direct way to obtain such a decomposition. However, the generators can be estimated from the truncated singular value decomposition:

$$\mathbf{H}_R = \mathbf{U}_R \Sigma \mathbf{V}_R^H \quad (2.6)$$

Because of an underlying isomorphism between equation (2.2.1) and (2.2.1), the matrix \mathbf{U}_R are related to matrices \mathbf{P} by a non-singular matrix \mathbf{Q}

$$\mathbf{U}_R = \mathbf{PQ}. \quad (2.7)$$

Signal poles z_r can be determined using the shift-invariance property of Vandermonde matrices [138]:

$$\mathbf{P}^\uparrow = \mathbf{P}_{\downarrow} \mathbf{Z}, \quad (2.8)$$

where the up and down arrow placed after a matrix stands for deleting the top and bottom row of the considered matrix, respectively and \mathbf{Z} is a diagonal matrix with signal poles z_r along the diagonal. Combining equations (2.2.1) and (2.2.1) result in the shift-invariance property of the singular matrix \mathbf{U}_R

$$\mathbf{U}_R^\uparrow = \mathbf{U}_{R\downarrow} \hat{\mathbf{Z}},$$

with $\hat{\mathbf{Z}} = \mathbf{Q}^{-1} \mathbf{Z} \mathbf{Q}$. The matrices \mathbf{Z} and $\hat{\mathbf{Z}}$ have the same eigenvalues. A least squares solution is used to estimate $\hat{\mathbf{Z}} = \mathbf{U}_{R\downarrow}^\dagger \mathbf{U}_R^\uparrow$. Finally, the signal poles z_r are obtained from the eigen-decomposition of $\hat{\mathbf{Z}}$. The complex amplitude c_r can be obtained from least squares using the Vandermonde source matrix constructed from the estimated signal poles.

This method is used for residual water suppression in Chapter 4, which is referred to as Hankel singular value decomposition (HSVD).

Multi-channel signal For the blind BSS problem in (2.2.1), if each of the observed signals is modelled as a sum of complex exponentials as in (2.2.1), a Hankel-tensor based exponential data fitting method can be applied for estimating the source and mixing matrix [138]. A tensor $\mathcal{H} \in \mathbb{C}^{I \times J \times K}$ constructed by stacking K Hankel matrices obtained from the columns of \mathbf{S} can be decomposed as follows:

$$\mathcal{H} = \sum_{r=1}^R \begin{bmatrix} 1 \\ z_r^1 \\ z_r^2 \\ \vdots \\ z_r^I \end{bmatrix} \otimes \begin{bmatrix} 1 \\ z_r^1 \\ z_r^2 \\ \vdots \\ z_r^J \end{bmatrix} \otimes \begin{bmatrix} h_{1r} \\ h_{2r} \\ h_{3r} \\ \vdots \\ h_{kr} \end{bmatrix}. \quad (2.9)$$

$$\mathcal{H} = \mathcal{I} \times_1 \mathbf{V}^{(1)} \times_2 \mathbf{V}^{(2)} \times_3 \mathbf{H},$$

where \mathcal{I} is a pseudo-diagonal $(R \times R \times R)$ -tensor with ones on its diagonal, $\mathbf{V}^{(1)} \in \mathbb{C}^{I \times R}$, $\mathbf{V}^{(2)} \in \mathbb{C}^{J \times R}$ are Vandermonde matrices and $\mathbf{H} \in \mathbb{C}^{K \times R}$ contains the complex amplitudes. In case of tensors, a Vandermonde decomposition can be obtained using CPD [168]. However, in our application of residual water suppression, the estimated generators using CPD were not good enough. Since the generators contain frequency information and are important for suppressing

water, a CPD based method was not employed in this thesis. Generators can also be obtained by applying a truncated MLSVD to the Hankel tensor \mathcal{H}

$$\mathcal{H} \approx \mathcal{H}_R = \mathcal{A} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}. \quad (2.10)$$

where \mathcal{A} is an all-orthogonal, ordered, complex-valued $(R \times R \times R)$ -tensor, $\mathbf{U}^{(1)} \in \mathbb{C}^{I \times R}$ is a complex matrix of which the orthonormal columns span the column space of $\mathbf{V}^{(1)}$, $\mathbf{U}^{(2)} \in \mathbb{C}^{J \times R}$ is a complex matrix of which the orthonormal columns span the column space of $\mathbf{V}^{(2)}$ and $\mathbf{U}^{(3)} \in \mathbb{C}^{K \times R}$ is a complex matrix of which the orthonormal columns span the column space of \mathbf{W} . Because of an underlying isomorphism between equation (2.2.1) and (2.2.1), the matrices $\mathbf{U}^{(p)}$ are related to matrices $\mathbf{V}^{(p)}$ ($p = 1, 2$) by a non-singular matrix \mathbf{Q}

$$\mathbf{U}^{(p)} = \mathbf{V}^{(p)} \mathbf{Q}.$$

A Vandermonde source matrix is constructed from the estimated signal poles and the mixing matrix \mathbf{H} can then be estimated using least squares. This method will be explored for achieving MRSI residual water suppression in Chapter 4.

2.2.2 Supervised learning

Random forest

Random Forest (RF) is an ensemble learning method where a collection of weak decision trees are used to predict the output class. Random decision forest was first proposed by Tin Kam Ho [86] and further extended by Breiman [27] and coined the term "Random Forests".

A decision tree is a non-parametric learning algorithm, which uses a model of decisions to predict the class of a target variable. It starts with single node branching into possible outcomes leading to new nodes. Each of these nodes can further branch into other possible outcomes. This process is continued till a final decision is made (Figure 2.10). In the training phase each node learns simple decision rules from the training data. A decision trees provides many advantages such as capacity to handle both numerical and categorical data, robust to outliers and simple to understand and interpret with possibility of tree visualization. The main disadvantage of decision trees is that they have low bias and very high variance, resulting in over-fitting the training data. Random forest overcomes this problem by combining bagging [26] with random selection of a feature subset at each node.

Bagging, also known as bootstrap aggregating, is one of the main parts in a random forest algorithm. Bagging selects a random sample with replacement

of the training set and each individual decision tree will learn a different classification model based on a bootstrapped sample. In each bootstrap, typically 63% of the training data is sampled with replacement and the remaining 37% from the out-of-bag data. Out-of-bag data are used for internal validation like estimating classification error and accessing feature importance. In decision trees of a random forest, at each candidate split in the learning process only a random subset is used instead of all the features. For classification problems, random forest typically uses \sqrt{d} features from a total of d features at each split. Figure 2.10 shows majority voting from a collection of decision trees used in random forest.

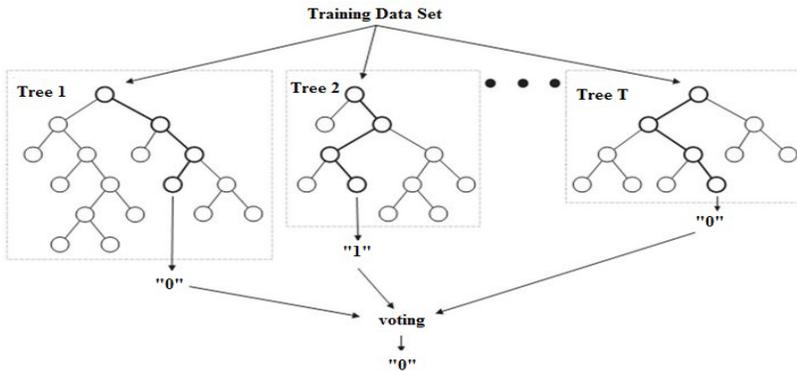


Figure 2.10: Majority voting scheme used in random forest. Image adapted from [133].

Convolutional neural network

Deep learning is the trending topic in machine learning with state-of-the-art performance in many areas. In the last decade they have grown exponentially and have made inroads into a wide range of areas ranging from healthcare to genetics to finance and to space exploration. Big companies such as Google, Microsoft, Facebook, Amazon, Instagram, Baidu, IBM, Tesla, and many more are using deep learning in many of their applications/products. Deep learning is very powerful and has huge potential, recently its performance surpassed that of human in applications like object recognition and classification [84], speech recognition [85], and defeat both European and world number 1 master in the game of Go [161]. Convolutional neural networks (CNN) are a class of deep neural networks mainly suited for image recognition and classification tasks. CNN became widely recognised after Krizhevsky et al. [101] proposed a method achieving a top-5 error of 15.3%, more than 10.8 percentage points ahead of

the runner up in the 2012 ImageNet Large-Scale Visual Recognition Challenge [149].

The most fundamental element of any neural network is a neuron as shown in Figure 2.11. A neuron consists of 1) a set of synapses or connecting links characterized by a weight, 2) an adder for summing the input signals and 3) an activation function. Each neuron will perform the dot product between the input and the weights of the synapses, adds the bias to it and performs non-linear activation to generate the output. In CNN, hyperbolic tangent (tanh), sigmoid and Rectified Linear Unit (ReLU) are commonly used activation functions, which are shown in Figure 2.12. ReLU is the most widely used activation function, which is also employed in this thesis since it is about six times faster than tanh [101].

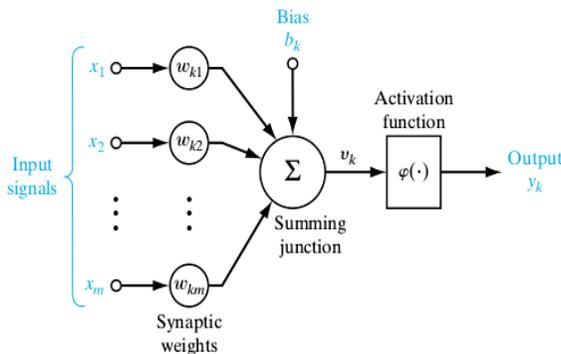


Figure 2.11: Model of a neuron as used in neural networks. Source: [83]

Most CNN architectures are usually made of the following types of layers: convolutional (conv), pooling (either maximum or average), fully connected (FC), and activation (e.g. rectified linear unit (ReLU)). Convolution and the fully connected layers are the most important ones. Weights in these layers are learned during training and are called weight layers. Several of these layers are usually connected in a sequential manner between the input and the output layer.

In 2D-convolution, the convolution kernel slides through the image and performs the dot product to generate the output as shown in Figure 2.13. In CNN, convolution layers apply the convolution operation to the input, passing the result to the next layer. The convolution filter is defined by its receptive field (filter size, width and height) with the depth of the filter equal to the depth of the input. Therefore, each filter consists of 3-D (width, height and input depth) weights. A single convolution layer can have multiple filters, generating a set

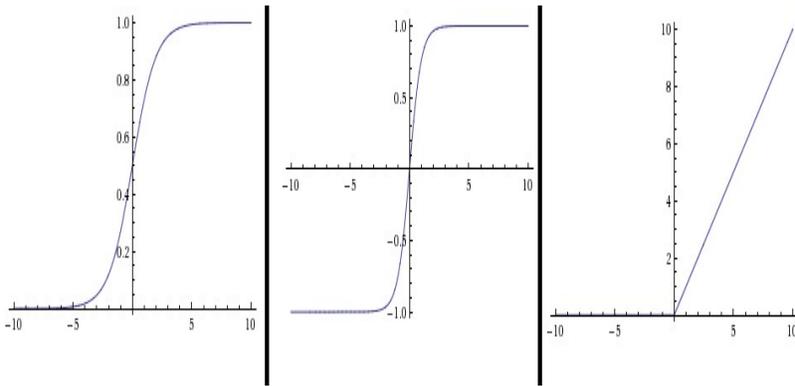


Figure 2.12: Hyperbolic tangent, sigmoid and ReLU activation functions from left to right, respectively. Image adapted from: [95]

of feature maps. The number of filters defines the output depth. Stacking the feature maps from all filters along the depth dimension forms the full output volume of the convolution layer. The kernel of the convolution layer can be represented by 4-D weights consisting of width, height, input depth and output depth. Figure 2.14 shows a convolution layer with receptive files 2×2 , input depth of three and output depth of five.

Other parameters of the convolution layer include stride and zero padding. Stride defines the amount of pixels that the convolution filter (or kernel) slides between two operations. A stride of one retains the input matrix size and a value greater than one reduces the dimension of input matrix. Zero padding adds zeros at the outer edges of the input matrix. It is mostly used to adjust or to preserve the input dimension. For example consider an input volume of size $50 \times 50 \times 20$ and a convolution layer with filter size/receptive field of 3×3 , output depth (number of filters) of 25, stride of 1 and zero pad of 1 on all sides. The convolution kernel is of size $3 \times 3 \times 20 \times 25$ and will generate an output volume of size $50 \times 50 \times 25$.

After each convolution layer, CNN architectures typically have a ReLU activation layer followed by a maximum pooling (MP) layer. In the max-pooling layer, the dimensionality of an input feature map is reduced by retaining only maximum values as shown in Figure 2.15.

Fully connected (FC) layers connect every neuron in one layer to every neuron in another layer as shown in Figure 2.16. A convolution layer can be converted to a FC layer by making the kernel size equal to the input image size, with no padding applied to the input image. In the end an activation layer will be used

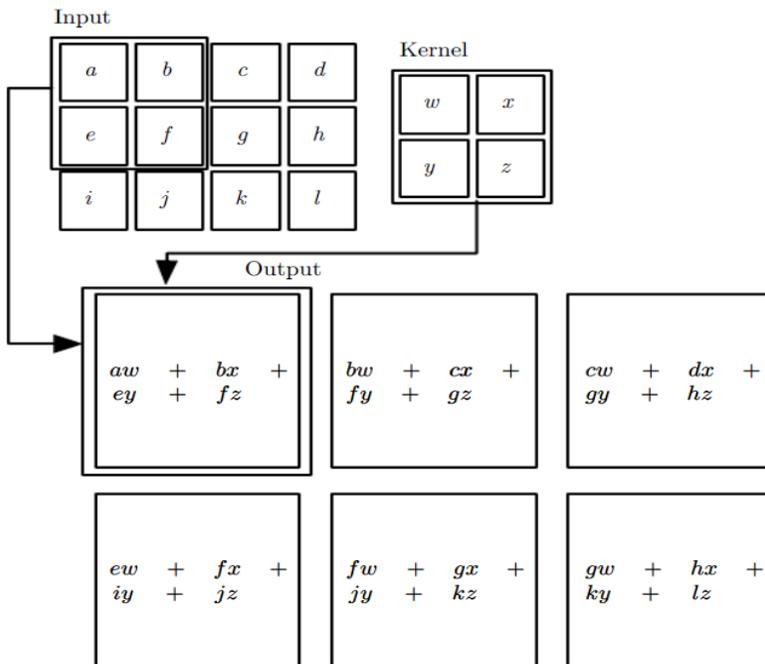


Figure 2.13: Convolution operation on an image of size 4×4 with a 2×2 kernel. Image obtained from: [79]

to produce the output, typically sigmoid. During training a loss layer will be present at the end of the CNN architecture, which is used by a backpropagation algorithm to update the weight of the convolution and FC layers. first described in (1), leading to the possible usage of higher learning rates. In principle, the method adds an additional step between the layers, in which the output of the layer before is normalized. Batch Normalization (BN) is an additional step that can be added between Convolution layer and ReLU, in which the output of the layer before is normalized. Batch Normalization will reduce the internal covariate shift in neural networks allowing higher learning rates [90].

A typical CNN architecture has a few main building blocks of [conv-ReLU-MP] with convolutional filter of size 5×5 or 7×7 , followed by a few FC layers, and a final activation layer, usually sigmoid. Simonyan and Zisserman [207] have shown the benefits of modifying the building block by adding extra convolutional and ReLU layers, [conv-ReLU-conv-ReLU-MP], but using only a small convolutional kernels with receptive field 3×3 . We use the same building block in the CNN architecture that we built in Chapter 6.

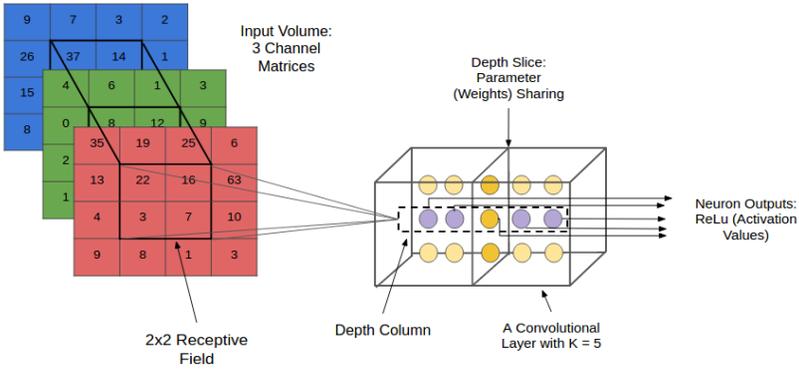


Figure 2.14: 3-D model of CNN convolution layer. Image obtained from: [153]

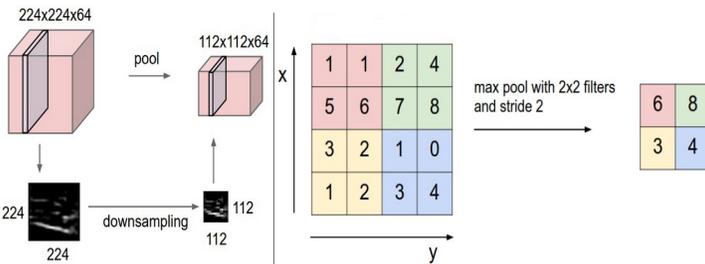


Figure 2.15: Demonstration of max-pooling operation. Image adapted from: [95]

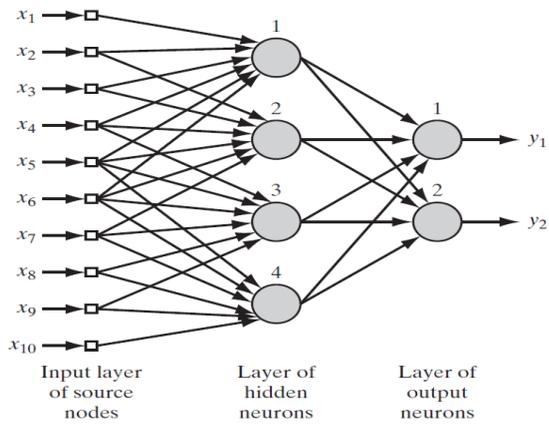


Figure 2.16: Fully connected layer with one hidden layer. Image adapted from: [83]

Chapter 3

Multi-parametric magnetic resonance imaging

3.1 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is a medical imaging technique that can visualize the anatomy and the physiological processes of the body. MRI has a wide range of applications in the medical field and is used on both human and non-human subjects. MRI is the investigative tool of choice in neuroimaging, especially for diagnosis and treatment of cancer, but it is also used in cardiovascular, musculoskeletal, liver and gastrointestinal imaging. MRI scanning is safer compared to other scanning techniques like Computed Tomography (CT), Positron Emission Tomography (PET), or Single-Photon Emission Computed Tomography (SPECT) as it is based on magnetic fields.

3.1.1 Principles of MRI

Magnetic resonance imaging is based on the principle of nuclear magnetic resonance, which uses the property that certain atomic nuclei can absorb and emit radio frequency energy under the influence of an external magnetic field. Atoms such as hydrogen (^1H), carbon (^{13}C), fluorine (^{19}F) and phosphorus (^{31}P) exhibit quantum mechanical property called spin, which enables them to have a magnetic dipole moment. Since hydrogen is the most abundant atom present in the human body, (^1H) is the most routinely used technique. In the

remainder of the thesis only ^1H MRI is considered. Under normal conditions magnetic fields from various hydrogen nuclei cancel each other, resulting in zero net magnetization. When a strong external magnetic field B_0 is applied, the hydrogen nuclei spins align themselves along the magnetic field. The direction of the applied magnetic field B_0 is also referred to as longitudinal axis. The individual spins align either in low energy state with direction parallel to B_0 or in the high energy state with direction anti-parallel to B_0 . The number of nuclei that occupy a low energy state is larger than that of a high energy state, resulting in a net magnetization, M , in the direction of B_0 as shown in Figure 3.1. The spins of hydrogen nuclei precess along the direction of the net magnetization M , and cannot be detected at this stage. This is called the resting or equilibrium state. The rate of precession is proportional to the applied magnetic field B_0 [70]. Larmor equation defines the relationship between the frequency of precession and the applied magnetic field, given by:

$$f_0 = \gamma B_0$$

where, f_0 is the frequency of precession, called Larmor frequency, and γ is the gyromagnetic ratio unique to each type of atom.

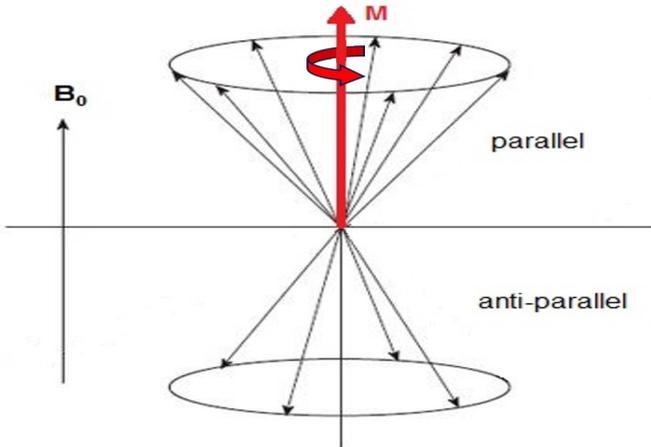


Figure 3.1: Graphical representation of precessing spins under an external magnetic field B_0 and the resulting net magnetization vector M . Source: [152]

In order to be able to measure the precession, an oscillating magnetic field B_1 called radio frequency (RF) pulse is applied in a plane perpendicular to B_0 , referred to as transverse direction. If the frequency of oscillating magnetic field B_1 is equal to the Larmor frequency of the hydrogen nucleus, some of the hydrogen nuclei in the low energy state will absorb the RF energy and jump to

a high energy state. This will result in a net magnetization in the transverse direction from the longitudinal direction of equilibrium state as shown in Figure 3.2. Once the RF pulse is removed, the hydrogen nuclei will return to the equilibrium state, this process is called relaxation. During relaxation, the nuclei lose energy by emitting an oscillating magnetic flux, which is recorded as an induction of a voltage in the RF coil by the rotating macroscopic magnetization according to Faraday's law and used to reconstruct a signal.

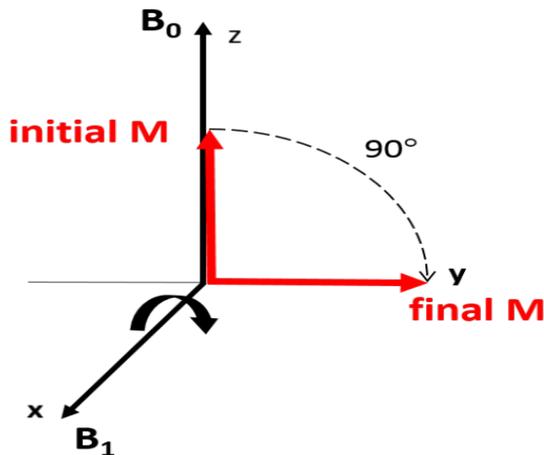


Figure 3.2: Illustration of change in the direction of magnetization vector M after applying the RF pulse. Source: [152].

The relaxation stage mainly consists of two events:

1) T_1 - or spin-lattice relaxation: here the magnetic component along the longitudinal direction is restored. The rate of restoration is described by time constant T_1 and different tissues have different values.

2) T_2 - or spin-spin relaxation: here the magnetization in the transverse direction gradually decays to zero and is described by a time constant T_2 .

T_2^* -relaxation: Local magnetic field inhomogeneities will result in transverse magnetization decay at a rate greater than T_2 . The combined effect of T_2 relaxation and additional factors will describes the actual decay of the transverse magnetization called as T_2^* -relaxation.

Different tissues have different hydrogen abundances and they also exhibit different relaxation times, these properties can be used to obtain contrast between tissues. Pulse sequence parameters such as echo time (TE) and repetition time (TR) can be used to exploit these properties and to obtain MRI images with different contrast between tissues. The echo time (TE) is the time between the RF excitation pulse and the peak in the MR signal induced in

the receiver coil, at which data acquisition starts. The repetition time (TR) is the time between successive pulse sequences and it determines recovery of longitudinal magnetization after each pulse. The next sections will briefly discuss some of the commonly used MRI techniques.

3.1.2 Conventional MRI

In neuroimaging Conventional MRI (cMRI) is the imaging modality of choice, which provides better spatial resolution and good soft tissue contrast. Using different pulse sequences and varying acquisition parameters, the echo time (TE) and repetition time (TR) will generate images with tissue contrast. In this thesis we have used T_1 -weighted, T_1 +contrast, T_2 -weighted and FLAIR (Figure 3.6) images of the brain.

T_1 -weighted imaging

T_1 -weighted imaging uses the differences in T_1 relaxation time among different tissue types and uses short TE and short TR for acquisition. Tissues having short T_1 relaxation time appear brighter in T_1 -weighted imaging. In tissues like fat, magnetic component along the longitudinal direction is restored quickly and they appear brighter (hyper-intense) on T_1 weighted images. The spins of hydrogen atoms in tissues with large amount of water realign slowly along the longitudinal direction and appear darker. In T_1 -weighted imaging of the brain, white matter appears bright, gray matter appears gray and cerebro-spinal fluid (CSF) appears dark.

T_1 -weighted imaging can be performed after the administration of contrast agent, generally gadolinium based compound [105]. This technique is known as T_1 +contrast imaging. Tissues like tumours where the blood-brain barrier is disrupted gets enhanced by gadolinium and appear extremely bright.

T_2 -weighted imaging

In T_2 -weighted imaging, the contrast between tissues is based on the T_2 (transverse) relaxation time. They require a long TE and TR and tissue with long T_2 appears brighter. Tissues like fat are dark and fluid is hyper-intense. In T_2 -weighted imaging of brain, gray matter appears gray, white matter appears dark and cerebral spinal fluid (CSF) appears hyper-intense.

FLAIR imaging

Fluid-attenuated inversion-recovery (FLAIR) is based on an inversion recovery technique, which is designed such that a signal from a particular tissue can be nulled by selecting the inversion time (TI). CSF and other fluids such as blood appear dark on a FLAIR image. Since CSF is suppressed in FLAIR imaging, it is good at detecting subtle changes in regions close to CSF and also small hyperintense lesions.

3.1.3 Magnetic resonance spectroscopic imaging

Magnetic resonance spectroscopy (MRS) is a non-invasive imaging technique that provides spectral profiles from which the metabolite concentrations of the tissue under investigation can be estimated. MRI mostly measures the water signal from the tissue, whereas MRS is capable of measuring magnetic resonance signals from nuclei present in other molecules. MRS is confined to a specific location (single voxel), while its extended technique, magnetic resonance spectroscopic imaging (MRSI), provides metabolite information in a two- or three- dimensional voxel grid.

MRS(I) is based on the principle of shielding effect or chemical shift effect. Due to a shielding effect of the atom, the local magnetic field is effectively reduced from the external magnetic field B_0 . Now the rate of precession is given by

$$f_{eff} = \frac{\gamma(1 - \sigma)B_0}{2\pi}$$

where, σ is the shielding constant, which depends on the chemical compound and the position of the nucleus in that compound. This results in a slightly different resonance frequency for different metabolites, which is called chemical shift from the reference frequency f_0 . The chemical shift is usually expressed in parts-per-million (ppm) since it is independent from the spectrometer frequency.

MRS can be used to measure signals from various nuclei such as hydrogen (^1H), carbon (^{13}C), fluorine (^{19}F) and phosphorus (^{31}P). ^1H -MRS is the most commonly used technique. Pulse sequences frequently used in MRS(I) are Point RESolved Spectroscopy (PRESS) [23] and STimulated Echo Acquisition Mode (STEAM) [73], which are shown in Figure 3.3. The concentration of water is significantly larger compared to other metabolites, typically in the range of 10,000 to 1. Therefore, in ^1H -MRS, a water suppression technique is employed to measure the metabolite signal of interest. Methods such as chemical shift selective suppression (CHESS) [82] and multiply optimized insensitive suppression train (MOIST) [134] are used for water suppression.

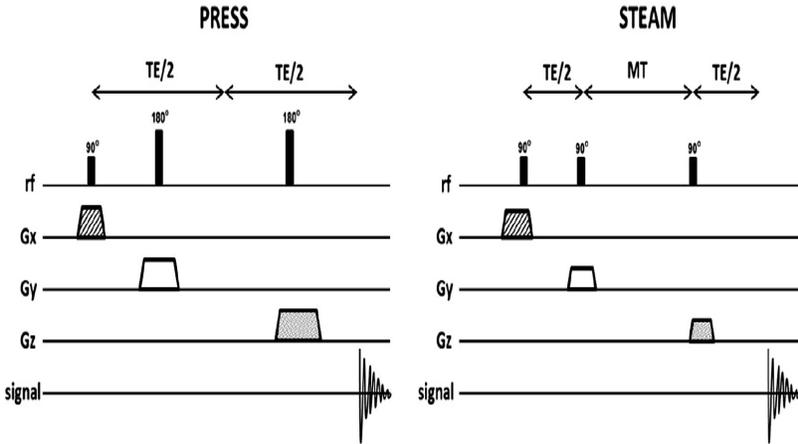


Figure 3.3: PRESS and STEAM acquisition sequences for MRS. MT stands for mixing time. Source: [15].

In MRS, acquisition is typically of short echo time ($TE \leq 50\text{ms}$) or long echo time ($TE \geq 130\text{ms}$). Long echo time signals show only a limited number of metabolites, whereas it is also easier to extract slowly decaying metabolites since it is less prone to baseline effects due to fast decaying macro-molecules. Short TE can detect large number of resonance frequencies, but also fast decaying macromolecular signals, overlapping with the metabolite signals of interest. A good quality single voxel MRS can be obtained from short acquisition times, whereas MRSI requires longer acquisition and processing times. Because of the low signal-to-noise ratio (SNR) in MRSI, the spatial resolution is restricted compared to other MRI modalities. Typically, a voxel size of $1 \times 1 \times 1\text{cm}^3$ is used in MRSI acquisition. Also, MRSI are measured from a limited region of interest instead of the whole brain in order to avoid long acquisition times. Figure 3.4 shows the MRSI signal measured from a region in the brain.

Important metabolites that are present in the human brain include N-acetyl aspartate (NAA), creatine (Cre), choline (Cho), glutamate (Glu), glutamine (Gln), myo-inositol (mI), glycine (Gly), lactate (Lac) and lipid (Lip). A short echo time MRS spectrum showing important metabolite peaks is shown in Figure 3.5. A complete list of metabolites along with their chemical shift can be found in [80]. The metabolite concentration present in the tissue is proportional to the area under the spectral peak of the corresponding metabolite. Peak area measurements are unreliable due to overlapping resonance peaks, baseline distortions and non-ideal lineshapes. Therefore, quantification methods such as Accurate Quantitation of Short Echo time domain Signals (AQSES) [144]

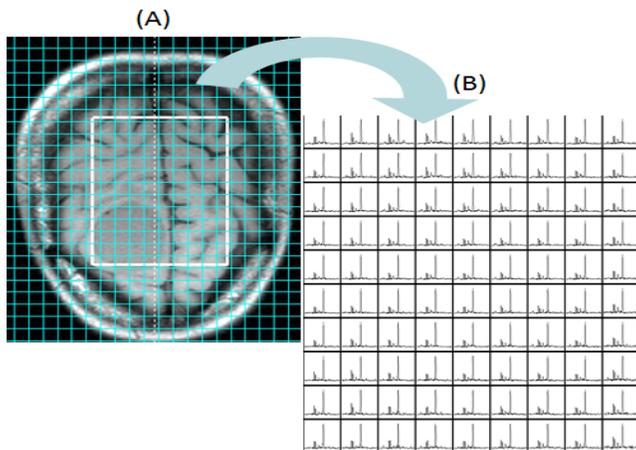


Figure 3.4: (A) MRSI grid overlaid on T_1 -weighted image. (B) Spectra corresponding to voxel grid within the white box.

and QUantitation based on quantum ESTimation (QUEST) [148], Totally Automatic Robust Quantitation in NMR (TARQUIN) [191] in the time-domain and LCModel [146] in the frequency domain are available. Pre-processing steps such as zero-filling, frequency alignment, phase correction, residual water suppression, signal normalization and baseline correction are usually performed before using the spectra for further analysis like quantification or classification.

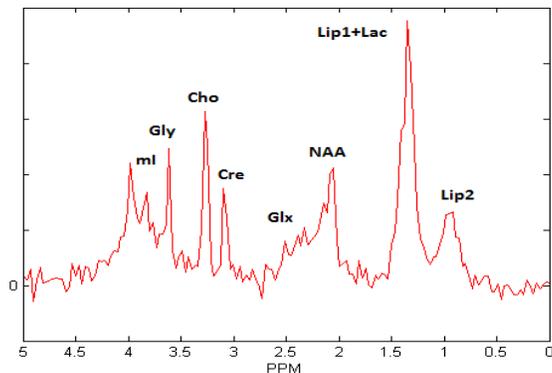


Figure 3.5: Absorption spectrum (real part) of a MRS signal obtained from a tumor region in brain.

3.1.4 Diffusion-weighted imaging

Diffusion-weighted imaging (DWI) is an advanced technique from which information about microscopic motion of water molecules within tissue can be obtained. Three types of diffusion can be found in the brain: free isotropic, restricted isotropic and anisotropic diffusion encountered mostly in CSF, gray matter and white matter regions, respectively. DWI uses the property where dephasing of the MR signal occurs due to random motion of water molecules in the presence of a varying magnetic field. Diffusion is made the dominant image contrast by applying large magnetic field gradients in particular directions. Apparent diffusion coefficient (ADC) maps can be obtained by acquiring DWI images with different weights. Diffusion tensor imaging (DTI) is a special kind of DWI, where the three-dimensional diffusion process is modelled [14]. From DTI, diffusion measures such as fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD) can be obtained. FA quantifies the degree of diffusion anisotropy, scaling from 0 (isotropic diffusion) to 1 (diffusion in only one direction). MD represents the magnitude of diffusion averaged over the applied gradient directions. AD and RD reflect the magnitude of diffusion along the principal axis of the ellipsoid and in a plane perpendicular to the principal axis, respectively.

Diffusion kurtosis imaging (DKI) is an emerging technique developed as an extension of DTI and is based on the non-gaussian diffusion of water in tissues, providing information on diffusional heterogeneity and micro-structural complexity [93]. Parameter maps such axial kurtosis (AK), radial kurtosis (RK), mean kurtosis (MK), and kurtosis anisotropy (KA) are obtained from DKI. AK represents the kurtosis coefficient in the direction of main diffusion, RK is the mean kurtosis value from all directions perpendicular to the direction of main diffusion and the MK is defined as the averaged kurtosis over all measured diffusion directions. In this thesis we have used FA, MD maps from DTI and MK map from DKI.

3.1.5 Perfusion-weighted imaging

Perfusion is the steady-state flow of blood to the tissue, measured as rate of blood delivery or blood flow per unit mass of tissue. Perfusion-weighted imaging (PWI) is an MRI imaging technique that provides insights into perfusion of blood in tissues, where several tissue hemodynamic parameters are estimated. Several PWI techniques have been proposed, which can be classified into two major approaches. In the first approach, an exogenous tracer such as gadolinium chelate is induced and dynamic scanning upon the passage of a tracer through the cerebro-vascular system is performed. Dynamic susceptibility-weighted MRI

(DSC-MRI) [157] and dynamic contrast-enhanced MRI (DCE-MRI) [173] belong to this class. In the second approach, techniques such as arterial spin labelling (ASL) use an endogenous contrast agent, where blood spins are magnetically labelled upstream from the imaging section with inverting or saturating RF-pulses [56]. DSC-MRI and DCE-MRI use a series of MR images to detect the passage of contrast agent through a given tissue. The injected contrast agent alters the T_2^* , T_2 and T_1 relaxation times locally, these changes picked up T_2^* , T_2 and T_1 -weighted images are used to quantify haemodynamic parameters. In DCE-MRI T_1 -weighted images are used, where permeability parameters like transfer constant (k^{trans}) and rate constant (k_{ep}) are estimated. DSC-MRI is the most relevant clinically used technique, which is also used in this thesis. In DSC-MRI, T_2^* - or T_2 - weighted images are used to quantify values such as cerebral blood volume (CBV), cerebral blood flow (CBF) a mean transit time (MTT). CBF represents instantaneous capillary flow in the tissue, CBV describes the blood volume of the cerebral capillaries and venules per cerebral tissue volume and MTT measures the length of time a certain volume of blood spends in the cerebral capillary circulation.

3.2 Glioma

Brain tumors emerge due to the formation of abnormal cells in the brain. They are classified into two groups: malignant tumors and benign tumors. Benign tumors usually do not grow back and lack the ability to invade neighboring tissue, whereas malignant tumors are aggressive with the potential to spread to other neighboring tissue. Brain tumors are one of the most fatal cancers [52], with a 5-year survival rate of 34.4% for primary malignant brain tumors [60].

Gliomas are the most common type of primary brain tumors originating from glial cells, which constitute 30% of primary brain tumors and 80% of malignant primary brain tumors [60]. The World Health Organization (WHO) categorises the gliomas into four grades [115]. Grade I and grade II are called low-grade glioma (LGGs), they are semi-malignant and have better prognosis. Similarly, grade III and grade IV are referred as high-grade glioma (HGG), they are malignant tumors and have very high fatality rate. The tumor sub-compartments of glioma contains active or enhancing tumor, necrosis and edema. Combination of all the three regions is referred to as "total tumor" and the combination of enhancing tumor and necrosis is referred to as "tumor core". Glioblastoma (GBM) is the most malignant type of tumor belonging to grade IV glioma, often characterized by the presence of necrosis (dead cell tissue). GBM is the most common type of glioma that is diagnosed in clinic with the lowest 5-year survival rate of 5%.

Common symptoms of glioma patients include headache, nausea, vomiting, fatigue, sleep disturbance and cognitive impairment [71, 162]. Medical history, age and neuro-images of the patient along with neurological examination are used in initial diagnosis of brain tumor. Currently, the gold standard for diagnosis and classification of gliomas consists of histological analysis of tissue samples obtained from biopsy. Treatment of gliomas consists of surgery, radiation therapy and/or chemotherapy and depends on the individual patient. After the treatment, the size of tumor resection plays an important role in prognosis and survival time of patients in GBMs.

3.2.1 Neuroimaging in glioma

Conventional MRI is the most common choice of imaging modality for diagnosis and treatment planning of glioma patients. Different MRI sequences are often combined in clinical practice for analysing and differentiating tumor sub compartments. cMRI is good at locating the tumor region. FLAIR imaging is better at capturing the whole tumor region including the edema, which appears brighter (hyper-intense). On T2-weighted images, the edema region adjacent to the tumor appears bright. In HGG, active tumor regions are enhanced in T1+contrast (T1c) images and the necrotic region appears dark as the contrast agent does not enter dead cells. Figure 3.6 shows one slice of T1+contrast, T2 and FLAIR image for a GBM patient.

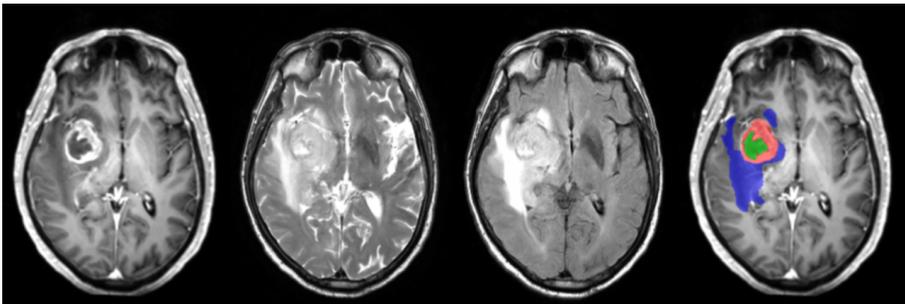


Figure 3.6: T1+contrast, T2, FLAIR and expert delineation overlaid on T1+contrast of a grade IV GBM patient left to right, respectively. Active tumor, necrosis and edema are shown in red, green and blue, respectively.

In clinical applications, PWI is used to assess regional variations in cerebral micro-vasculature in both normal and diseased brain. Conventional MRI is not a reliable indicator of malignancy. Combining CBV measurement from PWI with cMRI will help in evaluating tumor grading and malignancy. It also enables to

identify lesions such as cerebral abscesses or radiation necrosis, which resemble tumor. Pathologic differences in vascularization can be visualized using CBV maps, where radiation necrosis has lower CBV values and progressing tumor has higher CBV values. Figure 3.7 shows one slice of CBV map from a GBM patient.

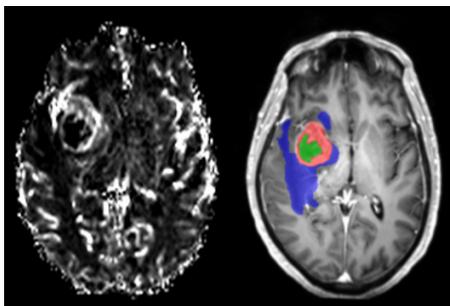


Figure 3.7: CBV map and expert delineation overlaid on T1+contrast of a grade IV GBM patient left to right, respectively. Active tumor, necrosis and edema are shown in red, green and blue, respectively.

The diffusion coefficient is usually higher in brain tumors compared to normal brain tissue [124]. Studies have shown that MK values increase with tumor grade and are better in differentiating between glioma grades than ADC and FA [147, 178]. MD and FA values measured from tumor region and peri-tumoral edema region have been successfully applied to differentiate between GBMs and metastases [30]. Figure 3.8 shows one slice of MD, FA, MK maps from a GBM patient.

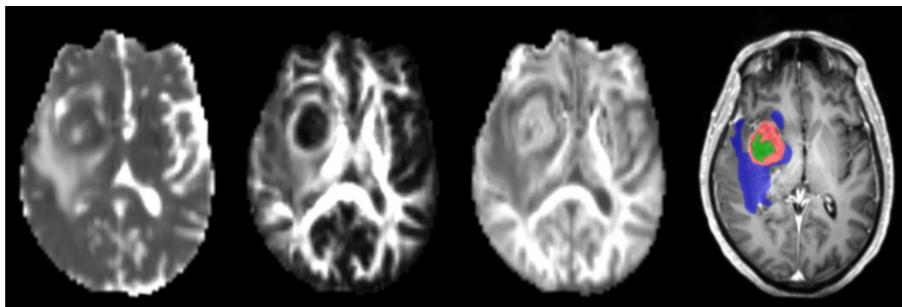


Figure 3.8: MD, FA, MK maps and expert delineation overlaid on T1+contrast of a grade IV GBM patient left to right, respectively. Active tumor, necrosis and edema are shown in red, green and blue, respectively.

MRSI helps in assessing the spatial variability of the tumor characteristics and is quite useful in determining glioma grade non-invasively. MRS(I) has the potential to monitor the glioma grade. The levels of NAA and mI are decreased and Cho is increased as glioma progresses upto grade III [29]. Presence of Lac indicates the malignant transformation of the tumor. Lipids are found in the MRS(I) of GBM patients, which indicate the presence of necrosis. MRSI is also useful in discriminating tumor recurrence from radiation necrosis, where studies have shown that Cho/NAA and Cho/Cre ratios are significantly higher in recurrent tumor than in radiation necrosis [164, 189]. Figure 3.9 shows the quantified NAA and Lip maps from a 2-D MRSI signal in a GBM patient.

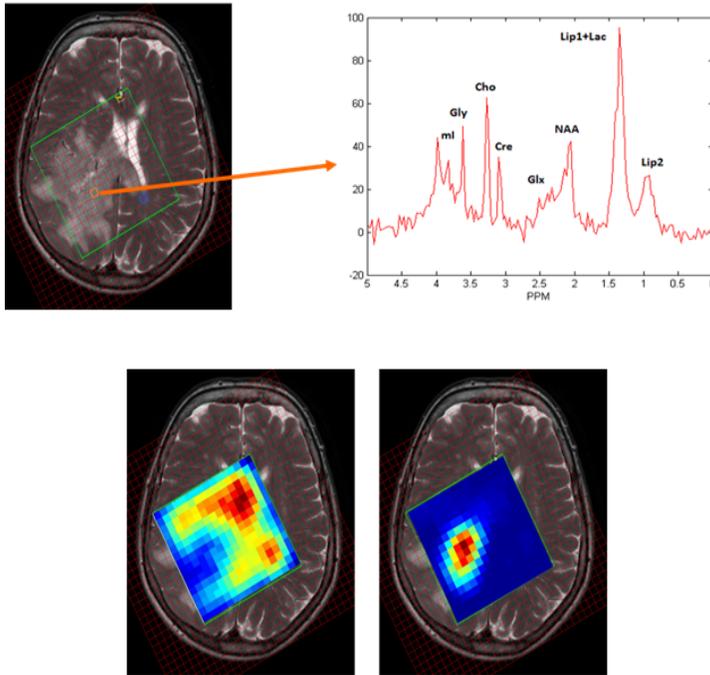


Figure 3.9: Top row: MRSI grid overlaid on T2 image and spectrum from a voxel in necrotic region (from left to right). Bottom row: Quantified NAA and Lip maps overlaid on T2 image (from left to right).

3.3 Data acquisition and pre-processing

3.3.1 UZ Leuven data

A total of 28 2-D- ^1H MRSI data was acquired on a 3T MR scanner (Achieva, Philips, Best, The Netherlands) at the University Hospital of Leuven from brain tumor patients. The study and the experimental procedures involving human subjects have been approved by the ethical committee of the institute.

All the MP-MRI acquisitions were performed on a 3T MR system (Philips Achieva, The Netherlands) at the University Hospital of Leuven, using a body coil for transmission and an 8-channel head coil for signal reception. The total time for all MRI modalities was around 45 min. 14 high-grade glioma patients (11 grade IV, 2 grade III and one grade II astrocytoma with focal progression to a grade III glioma) were scanned for this study. The lesions were classified according to grade using the 2007 WHO classification [115], with histopathological confirmation in all cases. The study involving human subjects have been approved by the human ethics review board of the institute. Written informed consent was obtained from every patient before participation.

Conventional magnetic resonance imaging (cMRI)

An axial spin echo T2-weighted MRI was acquired with the following parameters: repetition time (TR)/ echo time (TE): 3000/80 ms; slice/gap: 4/1 mm; turbo factor: 10; field of view [FOV]: $230 \times 184 \text{ mm}^2$; acquisition matrix: 400×300 .

A T1-weighted 3-D spoiled gradient echo MRI scan with contrast administration was performed with the following parameters: fast field echo, TR/TE/inversion time (TI): 9.7/4.6/900 ms; flip angle: 8° ; turbo field echo factor: 180; acquisition voxel size: $0.98 \times 0.98 \times 1 \text{ mm}^3$; 118 contiguous partitions.

An axial FLAIR MRI scan was acquired with the following parameters: TR/TE/TI: 11000/120/2800 ms, slice/gap: 4/1 mm, FOV: $230 \times 184 \text{ mm}^2$, acquisition matrix: 240×134 .

Perfusion weighted imaging (PWI)

Perfusion images were obtained using Dynamic Susceptibility-weighted Contrast-enhanced MRI with a gradient-echo EPI sequence: TR/TE: 1350/30 ms; slice/gap: 3/0 mm; dynamic scans: 60; FOV: $200 \times 200 \text{ mm}^2$; matrix: 112×109 ; EPI data were acquired during the first pass following a rapid injection of a

0.1mmol/kg body weight bolus of meglumine gadoterate via a mechanical pump at a rate of 4mL/s, followed by a 20-mL bolus of saline.

DSC data were analyzed using the DSCoMAN plugin [1] in ImageJ [154]. Relative cerebral blood volume (rCBV) maps were derived from the dynamic signal intensity curves using the method proposed by Boxerman et al. [25].

Magnetic resonance spectroscopic imaging (MRSI)

A 2D-¹H MRSI protocol was used as previously described in [177]. A point-resolved spectroscopy sequence was used as the volume selection technique with a bandwidth of 1.3kHz for the conventional slice-selective pulses; TR/TE: 2000/35ms; FOV: 160×160mm²; maximal volume of interest (VOI): 80×80mm²; slice thickness: 10mm; acquisition voxel size: 10×10mm²; reconstruction voxel size: 5×5mm²; receiver bandwidth: 2000Hz; samples: 2048; number of signal averages: 1; water suppression method: multiple optimizations insensitive suppression train [134]; first- and second-order pencil beam shimming; parallel imaging: sensitivity encoding with reduction factors of 2 (left-right) and 1.8 (anterior-posterior). Automated pre-scanning optimized the shim in order to yield a peak width consistently under 20Hz full-width half-maximum (FWHM). The slice was positioned in the center of the tumor. Voxels outside the MRSI PRESS excitation volume are excluded from the analysis.

The raw MRSI data were exported from the Philips platform after performing the following post-processing steps: zero filling in k space, transformation from k space to normal space, automatic phase correction and eddy current correction. Further processing was done with MATLAB based SPID software [143]. Two different pre-processing steps were used: 1) When only MRSI data was used for tumor characterization (Chapter 5 and 6): the residual water component was removed from the MRSI data using Hankel Lanczos singular value decomposition with partial reorthogonalization (HLSVD-PRO) [106]. A model order of 30 and a passband of 0.25 to 4.2 parts per million (ppm) were used in HLSVD-PRO algorithm. After removing the water component, baseline correction and baseline offset correction were performed. All the pre-processing were done using the Matlab (The MathWorks, Inc., Natick, Massachusetts, United States) based software, SPID [143]. The spectra were aligned in frequency using a simulated reference spectrum, which was generated using the parameters given in [80]. The complex-valued pre-processed spectra were truncated to the region 0.25-4.2 ppm and the truncated spectra were normalized to unit norm (l_2). Voxels outside the MRSI PRESS excitation volume are excluded from the analysis.

2) When MRSI data was used as part of MP-MRI signal: The residual water was removed using MPFIR (Maximum-Phase Finite Impulse Response) filtering

[145], a model order 25 and the spectral range from 0.25 to 4.2ppm were used. A band of voxels at the outer edges of the VOI were omitted to avoid chemical shift displacement artifacts and lipid contamination (band proportional to the grid size, maximally 3 voxels wide for a 16×16 grid). AQSES-MRSI [41] was used to quantify the following metabolites using an experimentally acquired metabolite basis set: Lip, Lac, NAA, Glx, Cre, Cho, mI and Gly. The basis set incorporates prior knowledge of the individual metabolites into the quantification procedure, i.e. their frequency, phase, and relative amplitudes within the multiplets. The quantified values were used in MP-MRI signal.

Diffusion kurtosis imaging (DKI)

An EPI DWI sequence with a spin-echo readout was used to acquire the DKI data, according to an optimized DKI protocol [142]. Implemented b-values were 0, 700, 1000, and 2800 s/mm^2 , applied respectively in 10, 25, 40, and 75 uniformly distributed directions. The following parameters were used in the acquisition: TR/TE: 3200/90 ms; gradient duration/diffusion time interval: 20/48.3 ms; FOV: $240 \times 240 \text{ mm}^2$; matrix: 96×96 ; number of signal averages: 1; slice/gap: 2.5/0 mm; parallel imaging: SENSE with factor 2 in the antero-posterior direction.

After motion and eddy current correction [8], isotropic smoothing was applied using a Gaussian kernel (FWHM=3mm) to reduce Gibbs ringing. However, smoothing was applied only to b=0 and b=700 images in order to avoid a noise bias in the tensor estimation [184]. Diffusion and kurtosis tensors were estimated in each voxel using a constrained weighted linear least-squares algorithm [184]. Mean diffusivity (MD), fractional anisotropy (FA) and mean kurtosis (MK) maps were derived from the tensors according to [142, 107].

Data co-registration

All the MP-MRI modalities were co-registered and brought to the same spatial resolution to perform voxel-wise analysis. Cubic spline interpolation was used to create an interpolated set of T1+C images, with one slice coinciding with the central plane of the MRSI volume of interest. Only the MRI data that are within the MRSI region of interest (ROI) were considered for analysis. Skull-stripping was applied to all images prior to co-registration. The cMRI dataset and the PWI dataset were rigidly co-registered to the interpolated T1+C reference set using SPM [2]. The normalized mutual information criterion was used for co-registration [122], with cubic spline interpolation for reslicing. Diffusion-weighted parameter maps were non-linearly co-registered to the T2-weighted

MRI using ExploreDTI [109] to minimize the local misalignment between the EPI distorted DKI data and the cMRI data. The MRSI data were spatially aligned with the reference set and resampled using cubic spline interpolation. All MRI parameters were brought to the spatial resolution of the original T1+C dataset, i.e. $0.98 \times 0.98 \times 1 \text{ mm}^3$. Only the voxels that are within the MRSI ROI are considered for analysis.

3.3.2 BRATS 2017 challenge dataset

The MP-MRI BRATS 2017 challenge dataset consists of four modalities T1, T1+contrast (T1+C), T2 and FLAIR [127, 4, 11, 12]. Datasets were acquired from 19 different institutions with various scanners and different clinical protocols. Provided MP-MRI images were co-registered to the same anatomical template, skull-stripped and interpolated to the same resolution (1mm^3). All datasets were manually segmented by one to four raters. All the raters followed the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. The expert labelling consisted of three classes: 1) enhancing tumor (ET), 2) peritumoral edema (ED) and 3) the necrotic and non-enhancing tumor (NCR/NET). The dataset was divided into three groups: 1) Training set, consisting of 210 high-grade glioma and 75 low-grade glioma patients. 2) Validation dataset containing 46 MP-MRI images. 3) Test dataset containing 146 MP-MRI images from both HGG and LGG patients. The user has access to expert labelling from the training set only.

Chapter 4

Residual water suppression in MRSI

Magnetic resonance spectroscopic imaging (MRSI) signals are often corrupted by residual water and artifacts. The aim of this chapter is to develop a tensor based method where the residual water is suppressed simultaneously in all voxels of the MRSI grid. A 3-D tensor is constructed by stacking the Löwner/ Hankel matrix from all the MRSI voxels in the third mode. Canonical polyadic decomposition (CPD)/ multi-linear singular value decomposition (MLSVD) is applied on the tensor to extract the water component, and to subsequently remove it from the original MRSI signal. Performance of tensor based methods is analyzed using simulations and in-vivo data and compared with the widely-used subspace based Hankel singular value decomposition (HSVD) method. The work presented in this chapter is based on [19].

4.1 Introduction

Magnetic resonance spectroscopic imaging (MRSI) is a non-invasive imaging technique that provides spectral profiles in a 2-D or 3-D voxel grid, from which the spatial distribution of metabolite concentrations or metabolite ratios can be estimated. Each voxel in the MRSI grid has a spectrum composed of several peaks corresponding to the metabolites present at that location. MRSI has many clinical applications and is used, among others, to investigate psychiatric disorders [42], for diagnosis and prognosis of brain tumors [45, 176], breast cancer [22] and autism [74]. Most of the clinical applications use metabolite

concentrations or metabolite ratios obtained from MRSI. Hence, an accurate and efficient quantification of metabolites is important. The metabolite levels in the human tissue are small compared to water, therefore ^1H MRSI signals typically contain a large water peak which is usually 10^3 to 10^4 larger than the metabolites of interest. This will affect the quantification of metabolites and has to be suppressed before applying any quantification algorithm. Typically, water suppression techniques are used during the acquisition of MRSI signals to get rid of large water peaks [134]. However, it is difficult to remove the water completely with these methods and some residual water will still be present in the spectra. It is important to suppress the water signal as much as possible for accurate and robust quantification of metabolites.

In general, residual water is suppressed before metabolite quantification, in a pre-processing step. Algorithms such as subspace-based Hankel singular value decomposition (HSVD) [13, 32, 180], multi-phase finite impulse response filtering [170], wavelet-based [191] and low rank methods based on union-of-subspaces [120] are available. Variants of these and other different methods are described in the review paper [66]. In the HSVD method, the water signal is first estimated using a subspace-based decomposition into a sum of complex damped exponentials and subsequently removed from the measured signal to suppresses the water component. HSVD is the most popular residual water suppression technique and is available in many software packages such as jMRUI [169], SPID [143], VeSPA [165] and TARQUIN [191] as a preprocessing step before quantification.

HSVD was originally developed to suppress residual water from single voxel magnetic resonance signal (MRS) signal. It can be applied to MRSI signal with 2D or 3D voxel grid, but they suppress water one voxel at a time. In MRSI, water signal from neighbouring voxels are closely related and they contain shared information. Since HSVD method works voxel-by-voxel in MRSI signals, they are not capable of exploiting the shared information present among the voxels in the MRSI grid. As such, the HSVD method might result in poor residual water suppression for some of the voxels in the MRSI grid. Processing one voxel at a time generated a matrix, where matrix based blind source separation technique was used to suppress water. On the other hand, processing all the voxels simultaneously generates a tensor, where tensor based BSS approach can be employed. This motivated us to develop tensor based approaches which enable us to exploit the shared information present among neighbouring voxels and result in better water suppression.

The water removal from the MRSI signals can be formulated as a blind source separation (BSS) problem. Recently, a Löwner-based BSS method has been developed, which can be used if the source signals can be approximated by rational functions. In this chapter, we propose a tensor based algorithm to

suppress the residual water simultaneously from all the voxels in the MRSI signal using the Löwner-based BSS method, under the assumption that the different MRSI components can be well approximated by low-degree rational functions. We have also explored a Hankel-tensor based exponential data fitting method for water suppression. These tensor based methods are compared against the matrix based HSVD method.

4.2 Tensor based methods for water removal

Section 4.2.1 explains the MRSI and the water signal model. Section 4.2.2 discusses residual water suppression using Löwner-tensor method. In Section 4.2.3, residual water suppression using Hankel-tensor method is explained.

4.2.1 MRSI and residual water

The MRSI time-domain signal is represented by a free-induction (FID). The complex time-domain FID signal in each voxel can be modeled by a sum of complex damped exponentials:

$$F(t) = \sum_{r=1}^R a_r e^{j\phi_r} e^{(-d_r + j2\pi f_r)t},$$

in which R is the number of resonance peaks in the signal, and f_r , a_r , ϕ_r and d_r are the frequency, amplitude, phase and damping of the r^{th} resonance peak, respectively. Similarly, in the frequency domain, the Fourier transform of the FID signal can be modeled as a sum of rational functions.

$$\begin{aligned} S(f) &= \sum_{r=1}^R \frac{a_r e^{j\phi_r} / 2\pi}{d_r + j2\pi(f - f_r)} \\ &= \sum_{r=1}^R \frac{c_r}{j\omega + p_r} \end{aligned} \tag{4.1}$$

where $c_r = a_r e^{j\phi_r} / 2\pi$ is the complex amplitude, $p_r = d_r - j2\pi f_r$ is the complex pole and $\omega = 2\pi f$ is the angular frequency.

The MRSI data matrix \mathbf{S} is constructed by stacking the spectra from each voxel in the columns. Similarly, the data matrix \mathbf{F} is defined by stacking the FID's from each voxel. The residual water present in the MRSI signals is

sometimes large and can strongly affect the metabolite peaks of interest, which belong to the region of interest in 0.25-4.2 ppm as shown in Figure 4.1. In theory, the water signal can be represented by only one exponential/rational function in time/frequency domain, but this model is not sufficient for an in-vivo signal. In practice, it is possible to model the residual water signal with a linear combination of several exponentials [32], which can then be extracted from \mathbf{S} (respectively, \mathbf{F}).

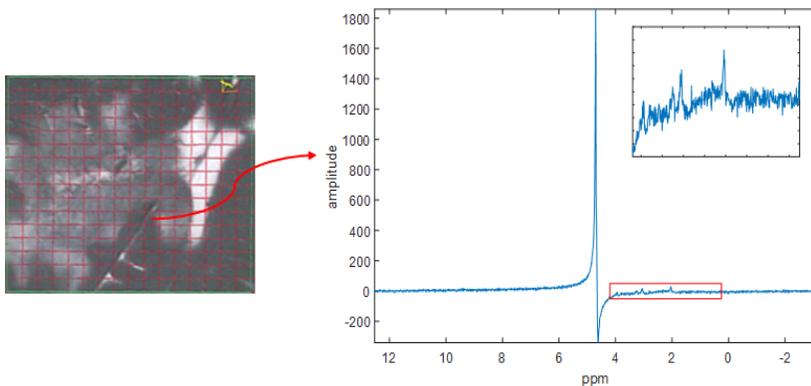


Figure 4.1: Absorption spectrum from one of the voxels in the MRSI grid with a large residual water signal. The region of interest for metabolites is shown within the red box.

4.2.2 Löwner-based water suppression

Löwner-based blind source separation in MRSI

It has been shown in [32] that it is possible to model the in-vivo residual water signal with a linear combination of several exponentials. Here, we assume that neighboring voxels in the MRSI signal share sources (rank-1 rational functions) that are used to model the residual water signal. Hence, the estimation of sources that model residual water and their corresponding abundances from the K measured MRSI signals can be formulated as a BSS problem:

$$\mathbf{S} = \mathbf{W}\mathbf{H}^T,$$

where columns of $\mathbf{S} \in \mathbb{C}^{N \times K}$ contain the measured spectra from all the voxels, the columns of $\mathbf{W} \in \mathbb{C}^{N \times R}$ represent the individual metabolite components and the columns of $\mathbf{H} \in \mathbb{C}^{K \times R}$ (the mixing matrix) represent their corresponding

abundances (weights) in each voxel. We use the Löwner-based BSS technique explained in Section 2.2.1 to estimate the individual metabolite sources. For each voxel, a Löwner matrix $\mathbf{L}_{\mathbf{s}_k}$ is constructed from the corresponding spectrum in the MRSI grid:

$$\mathbf{L}_{\mathbf{s}_k} = \begin{bmatrix} \frac{\mathbf{s}_k(x_1) - \mathbf{s}_k(y_1)}{x_1 - y_1} & \frac{\mathbf{s}_k(x_1) - \mathbf{s}_k(y_2)}{x_1 - y_2} & \cdots & \frac{\mathbf{s}_k(x_1) - \mathbf{s}_k(y_J)}{x_1 - y_J} \\ \frac{\mathbf{s}_k(x_2) - \mathbf{s}_k(y_1)}{x_2 - y_1} & \frac{\mathbf{s}_k(x_2) - \mathbf{s}_k(y_2)}{x_2 - y_2} & \cdots & \frac{\mathbf{s}_k(x_2) - \mathbf{s}_k(y_J)}{x_2 - y_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{s}_k(x_I) - \mathbf{s}_k(y_1)}{x_I - y_1} & \frac{\mathbf{s}_k(x_I) - \mathbf{s}_k(y_2)}{x_I - y_2} & \cdots & \frac{\mathbf{s}_k(x_I) - \mathbf{s}_k(y_J)}{x_I - y_J} \end{bmatrix},$$

in which \mathbf{s}_k is the spectrum of the k th voxel, and with $\{x_1, \dots, x_I\}$ and $\{y_1, \dots, y_J\}$ two partitions of the point set $\Omega = \{\omega_1, \dots, \omega_N\}$ with $N = I + J$. Two typical partitioning types for Ω are interleaved and block partitioning. Since the water signal is concentrated around 4.7 ppm, if we use the block partitioning it may result in one set containing large water peaks and the other set containing noise or small metabolite peaks. Interleaved partition guarantees that both sets contain water peaks. Therefore, we have used interleaved partition for constructing the Löwner matrix from the spectrum. The Löwner matrix is constructed using the spectrum from 0.25-6.5 ppm, which contains the region of interest in 0.25-4.2 ppm and the water region.

A third-order tensor $\mathcal{L}_{\mathbf{S}}$ is obtained by stacking the Löwner matrices along the third mode as shown in Figure 4.2. As it is assumed that each individual component \mathbf{w}_r can be well approximated by a degree-1 rational function with a single resonance peak as described in (4.2.1), each corresponding r th Löwner matrix has approximately rank 1 and can be described as $\mathbf{a}_r \mathbf{b}_r^T$. Hence, a CPD can be applied on $\mathcal{L}_{\mathbf{S}}$:

$$\mathcal{L}_{\mathbf{S}} \approx \sum_{r=1}^R \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{h}_r = \llbracket \mathbf{A}, \mathbf{B}, \mathbf{H} \rrbracket, \quad (4.2)$$

with $\mathbf{A} \in \mathbb{C}^{I \times R}$, $\mathbf{B} \in \mathbb{C}^{J \times R}$ and $\mathbf{H} \in \mathbb{C}^{K \times R}$. Each rank-1 tensor corresponds to the contribution of a particular component to the observed spectral data.

Estimation of source parameters

The abundance vectors \mathbf{h}_r can be directly identified from (4.2.2). A second goal is to identify the source components and their corresponding parameters as described in (4.2.1). The r th source is modeled by $\mathbf{w}_r(\omega) = \frac{c_r}{j\omega + p_r}$, and its

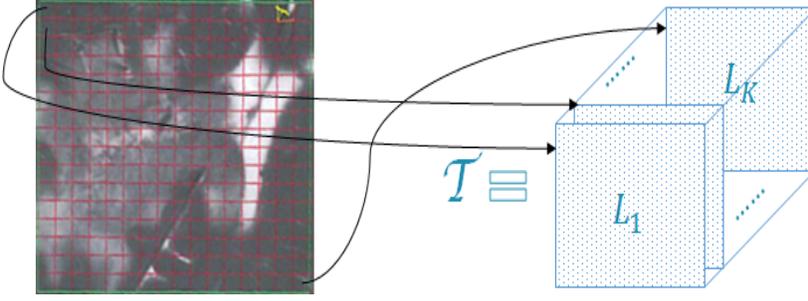


Figure 4.2: Construction of the Löwner tensor \mathcal{T} from the spectra of the different MRSI voxels

corresponding Löwner matrix $\mathbf{L}_{\mathbf{w}_r}$ can be written as:

$$\begin{aligned}
 \mathbf{L}_{\mathbf{w}_r} &= \begin{bmatrix} \frac{-jc_r}{(jx_1+p_r)(jy_1+p_r)} & \frac{-jc_r}{(jx_1+p_r)(jy_2+p_r)} & \cdots & \frac{-jc_r}{(jx_1+p_r)(jy_J+p_r)} \\ \frac{-jc_r}{(jx_2+p_r)(jy_1+p_r)} & \frac{-jc_r}{(jx_2+p_r)(jy_2+p_r)} & \cdots & \frac{-jc_r}{(jx_2+p_r)(jy_J+p_r)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-jc_r}{(jx_I+p_r)(jy_1+p_r)} & \frac{-jc_r}{(jx_I+p_r)(jy_2+p_r)} & \cdots & \frac{-jc_r}{(jx_I+p_r)(jy_J+p_r)} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{c_r^{(1)}}{jx_1+p_r} \\ \frac{c_r^{(1)}}{jx_2+p_r} \\ \vdots \\ \frac{c_r^{(1)}}{jx_I+p_r} \end{bmatrix} \begin{bmatrix} \frac{c_r^{(2)}}{jy_1+p_r} & \frac{c_r^{(2)}}{jy_2+p_r} & \cdots & \frac{c_r^{(2)}}{jy_J+p_r} \end{bmatrix} \\
 &= \mathbf{a}_r \mathbf{b}_r^T
 \end{aligned}$$

with $-jc_r = c_r^{(1)} c_r^{(2)}$. The parameters $c_r^{(1)}$ and p_r can be obtained from \mathbf{a}_r using least squares,

$$\begin{bmatrix} p_r \\ c_r^{(1)} \end{bmatrix} = [\mathbf{a}_r \quad -\mathbf{1}]^\dagger (j\mathbf{x} \circ \mathbf{a}_r),$$

where $\mathbf{1} \in \mathbb{R}^I$ is the vector with all ones. We can estimate $c_r^{(2)}$ and p_r from \mathbf{b}_r in a similar way. The final estimate of p_r can be obtained by averaging the estimates from \mathbf{a}_r and \mathbf{b}_r .

Water signal suppression

Once the source parameters are estimated, the model is extrapolated to the entire length of the frequency region. The abundance matrix \mathbf{H} is calculated using least squares from the source signals and measured spectra, $\mathbf{H} = (\mathbf{W}^\dagger \mathbf{S})^T$. The real part of each estimated pole gives the damping factor of the corresponding source (d_r) while the imaginary part returns the resonance frequency. The components whose resonance frequencies are outside the region of interest (0.25 - 4.2 ppm) belong to the water component or provide other nuisance peaks. Therefore, the influence of the water component on the observed spectral data is constructed using only those components and their corresponding abundance vectors. Let Φ denote the set of P indices corresponding to the P water sources. Then $\mathbf{W}_{water} = [\mathbf{w}_{\Phi_1} \ \cdots \ \mathbf{w}_{\Phi_P}]$ and $\mathbf{H}_{water} = [\mathbf{h}_{\Phi_1} \ \cdots \ \mathbf{h}_{\Phi_P}]$, and the contribution of the water component can be expressed as $\mathbf{S}_{water} = \mathbf{W}_{water} \mathbf{H}_{water}^T$. The water component can then be removed from the measured MRSI spectra as $\mathbf{S}_{suppressed} = \mathbf{S} - \mathbf{S}_{water}$.

After removing the water component from the MRSI signal, a small baseline will be present at the outer edges of the spectrum. This arises mainly because the Löwner method is not able to model the complex water signal properly at the outer edges of the spectrum when the water signal has some baseline. Also damped complex exponential function translates to rational function only when signal is continuous and infinitely long. The HSVD method can estimate a source with a broad peak (large damping) to model the edges of the water spectrum, whereas Löwner method fails to extract such broad peaks and hence fails to model the water spectrum at the outer edges and results in a baseline. This problem can be corrected by modeling the baseline using a polynomial function of degree D . Therefore, the polynomial functions are added to the estimated source matrix \mathbf{W} to obtain the matrix $\mathbf{W}_{poly} \in \mathbb{R}^{N \times (K+d+1)}$:

$$\mathbf{W}_{poly} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1R} & 1 & f_1 & f_1^2 & \cdots & f_1^d \\ w_{21} & w_{22} & \cdots & w_{2R} & 1 & f_2 & f_2^2 & \cdots & f_2^d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NR} & 1 & f_N & f_N^2 & \cdots & f_N^d \end{bmatrix}$$

The abundance matrix \mathbf{H} is recalculated in a least-square sense using the estimate of \mathbf{W}_{poly} and the measured spectra in \mathbf{S} . The residual water component is suppressed using the subtraction method as explained above. Each polynomial source is also considered as a water component.

In this method the water signal in each voxel is modeled as a linear combination of many (resonance peaks/rank-1 rational functions) (typically 20-30). For each voxel, the rows of the abundance matrix \mathbf{H} specify the subset of sources that

are used to model the water signal by means of their corresponding weights. Since these weight combinations are voxel-wise different, they can model voxel-wise variations in the water component. This will allow to handle the B0 inhomogeneity and spectrum distortions present in the MRSI signals.

Computation of CPD

The approach in this chapter makes use of a compression step and a CPD step [28]. The compression step applies an MLSVD with truncation to compute a smaller core tensor \mathcal{S} and corresponding factor matrices $\mathbf{U}^{(n)}$. Provided that the dimensions of \mathcal{S} exceed R , it can be shown that \mathcal{S} still has (approximately) rank R if \mathcal{T} has (approximately) rank R . In the second step, a CPD is performed on the smaller tensor \mathcal{S} (rather than on \mathcal{T}), which returns the factor matrices $\mathbf{B}^{(n)}$. CPD is computed using nonlinear least squares by minimizing the Frobenius norm of the error between the compressed core and its rank-1 approximation [166]. The n th factor matrix of \mathcal{T} is then equal to the $\mathbf{A}^{(n)} = \mathbf{U}^{(n)}\mathbf{B}^{(n)}$. This two-step procedure is especially beneficial if \mathcal{T} is large, as the computation complexity of the CPD algorithm is higher compared to the MLSVD algorithm. If \mathcal{T} only approximately has rank R , the two-step procedure still provides good estimates in various occasions which differ only minimally compared to the estimates from computing a CPD on \mathcal{T} directly.

The CPD algorithm requires an initial value for the factor matrices. Random initializations can sometimes result in poor water suppression. In order to overcome this problem, different initializations are used if the water suppression is not sufficient. To verify the quality of the water suppression, the variance in the water region and noise region are compared. If the variance in the water segment is larger than the variance in the noise segment by a given threshold, the water suppression is considered to be poor and a different initial value is used until a good suppression is obtained.

4.2.3 Hankel-tensor based water suppression in MRSI

In the time domain the BSS problem of separating individual resonance peaks can be formulated as:

$$\mathbf{F} = \mathbf{V}\mathbf{H}^T, \quad (4.3)$$

where columns of $\mathbf{F} \in \mathbb{C}^{N \times K}$ contain the measured FID from all the voxels, $\mathbf{V} \in \mathbb{C}^{N \times R}$ is the Vandermonde matrix of the R source poles (z_r) and the columns of $\mathbf{H} \in \mathbb{C}^{K \times R}$ (the mixing matrix) represent their corresponding abundances (weights) in each voxel.

Similarly to the Löwner method, for each voxel a Hankel matrix $\mathbf{H}_{\mathbf{f}_k}$ is constructed from the corresponding time-domain FID signal. A third-order tensor $\mathcal{H}_{\mathbf{f}}$ is obtained by stacking the Hankel matrices along the third mode as shown in Figure 4.2. We can estimate the poles $z_r = e^{-d_r + 2\pi f_r}$ of the MRSI signals by applying MLSVD to the Hankel tensor $\mathcal{H}_{\mathbf{f}}$ as explained in Section 2.2.1. The abundance matrix \mathbf{H} is then calculated using the least squares solution of the equation (4.2.3) in which the Vandermonde matrix V is derived from the estimated poles, $\mathbf{H} = (\mathbf{V}^\dagger \mathbf{F})^\top$. The real part of each $\log(z_r)$ gives the damping factor of the corresponding source (d_r) while the imaginary part returns the resonance frequency. The components with resonance frequencies outside the region of interest (0.25 - 4.2 ppm) are considered to reconstruct the water component and possibly other nuisance peaks. Finally, the residual water is suppressed from the MRSI signal by subtracting the estimated water component similarly to the Löwner method.

4.3 Results

To test the performance of the Löwner and Hankel-tensor based water suppression methods, they are applied on both simulated and in-vivo MRSI data. Section 4.3.2 discusses the performance of the Löwner, the Hankel-tensor and the HSVD methods on simulated datasets. In Section 4.3.3 the performance of the Löwner, Hankel-tensor and HSVD methods is assessed using in-vivo data. Tensorlab is used for the Löwner and Hankel matrix constructions and tensor computations [186]. The signal-to-noise ratio (SNR) is defined as the power of the signal to the power of the noise. Unless stated otherwise, a rank $R = 50$ is used for the CPD of the Löwner tensor $\mathcal{L}_{\mathbf{S}}$ and $d = 4$ degree polynomial sources are used for the entire MRSI in both the simulation and in-vivo cases. For the HSVD, an order of 50 was used for each voxel and for Hankel-tensor we have used an order of 100 in the MLSVD of the entire MRSI signal.

4.3.1 Spectral variations in MRSI voxels

To demonstrate that our proposed method can handle B0 inhomogeneity and spectrum distortions, we have applied the Löwner method to one in-vivo dataset. Figure 4.3 (a) & (c) show absorption spectra of an in-vivo MRSI signal from two of the voxels having different spectral shape (red, dashed-line). The estimated water signal (black, solid-line) is overlapped on the measured spectrum in the figure, and we clearly see that the estimated water signal (black, solid line) models both voxels that are having distinct spectral shapes. Figure 4.3 (b) & (d) show the individual resonance peaks used to model the water signal. From the

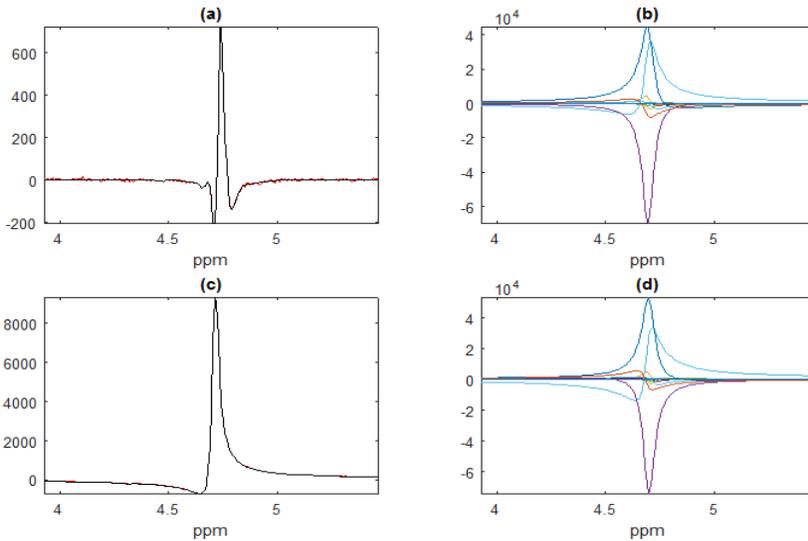


Figure 4.3: (a) & (c) Absorption spectra of an in-vivo signal without water suppression (red, dashed-line) overlapped with the estimated water spectra (black, solid-line) from two different voxels in the MRSI grid. (b) & (d) Individual resonance peaks used in the modelling of the water signal from (a) and (c), respectively.

figure we observe that the individual resonance peaks used to model the water component have different complex amplitude for both voxels, which enables to handle B_0 inhomogeneity and spectral distortions.

4.3.2 Simulations

The simulated signals containing residual water are generated using an in-vitro basis set which was obtained as described in [136]. The basis set consisting of in-vitro signals from Alanine (Ala), Aspartate (Asp), Choline (Cho), Creatine (Cre), γ -aminobutyric acid (GABA), Glutamate (Glu), Lactate (Lac), two lipids (Lip1 and Lip2), myo-Inositol (MI), N-Acetyl-Aspartate (NAA) and Taurine (Tau) metabolites was used to generate the spectra. A grid of MRSI signals of size 16×16 was generated for simulation. First, signals consisting of 12 metabolites are constructed using the basis set without any residual water. The amplitude of each metabolite in the grid is varied using a 2-D Gaussian window,

$$g(x, y) = e^{-(x^2+y^2)/2\sigma^2} + U(x, y)$$

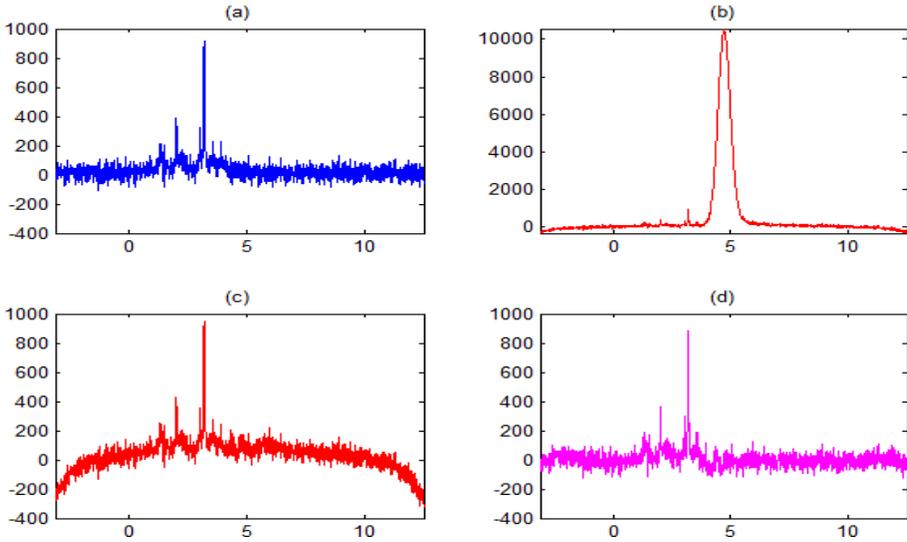


Figure 4.4: (a) Absorption spectrum of a simulated in-vitro signal without water from one of the MRSI voxels. (b) Absorption spectrum of a simulated in-vitro signal with large water peak from one of the MRSI voxels. (c) Absorption spectrum of the water-suppressed signal using the Löwner method without polynomial sources. (d) Absorption spectrum of the water-suppressed signal using the Löwner method with polynomial sources.

where $U(x, y)$ is a uniformly distributed random number, and $x = 0, y = 0$ is the central voxel. Circular Gaussian noise is added to the metabolite signals. Residual water signal was generated by scaling the in-vivo measured MRSI water reference signal. In each voxel the residual water signal is distorted by multiplying it with a Gaussian decaying signal $e^{-d_k t^2}$, where d_k is modelled as a uniformly distributed random variable between 0 and 0.005. Finally, residual water was added to the noisy metabolite signals to generate the MRSI data.

The Löwner method with and without polynomial sources was applied on a simulated MRSI signal to suppress the residual water. The result of water removal in one of the voxels is shown in Figure 4.4. From Figure 4.4(c)&(d) we clearly observe that the Löwner method will introduce a baseline and it can be addressed by including the polynomial sources in the least squares stage. In the remainder of the chapter, we have only considered the Löwner method with polynomial sources, unless explicitly mentioned.

The Löwner, Hankel-tensor and HSVD water-suppression methods are applied on the simulated MRSI signals for 100 different noise and metabolite amplitude

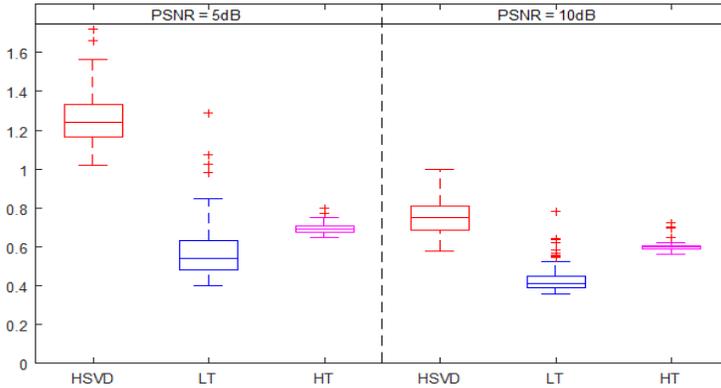


Figure 4.5: Boxplot of the residual errors after water suppression using the HSVD, the Löwner (LT) and the Hankel-tensor (HT) methods on 100 simulated in-vitro MRSI signals. The error is calculated as the l_2 -norm of the difference between the water-suppressed signal and the original water-free signal (with noise).

realizations. Figure 4.5 shows the boxplot of errors between the water-suppressed signal and the metabolite signal (with noise) for two different noise levels. The boxplot indicates that both the Löwner-based method and the Hankel-tensor method have a lower average error compared to the HSVD method and suppresses the residual water better without distorting the metabolite spectra. The Löwner-based method has the best performance in suppressing residual water compared to the other two methods.

4.3.3 In-vivo results

To test the performance of the algorithms we have applied the HSVD, Hankel tensor and Löwner methods on 28 in-vivo datasets. Figure 4.6 shows the real part of the spectra with residual water and after water suppression for two different voxels in a MRSI grid. In many voxels, all three methods give good water suppression as shown in Figure 4.6c. However, in some voxels the HSVD method does not perform well as shown in Figure 4.6f. In Figure 4.6b, we observe that an artefact is present at the right side of the water signal. Therefore the HSVD method fails to suppress water completely and results in a significant residue. However, both tensor methods are able to suppress water even in the presence of an artefact. Fig. 4.7 shows the water-suppressed spectra obtained with HSVD, Löwner and Hankel-tensor methods from different locations in the voxel grid.

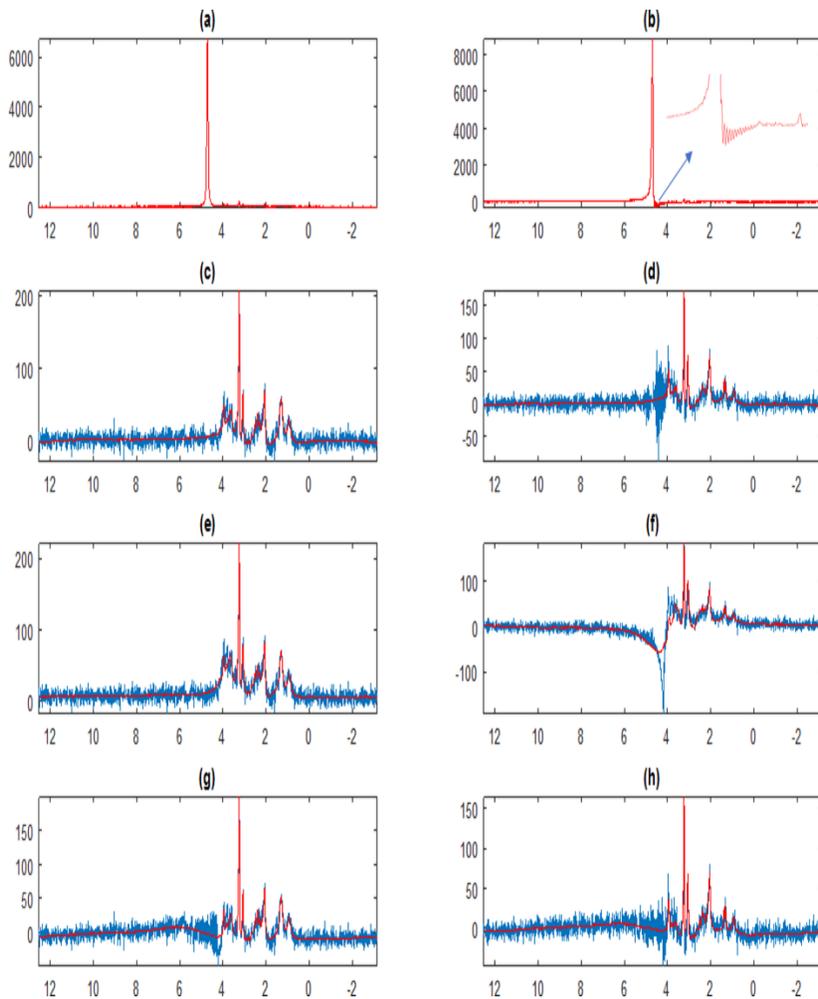


Figure 4.6: Residual water suppression in in-vivo MRSI signals. (a-b) Absorption spectra of measured MRSI signal with large residual water peak for two voxels. (c-d) Absorption spectrum of the water-suppressed signal (blue) using the Löwner method and the quantified signal (red) in the two corresponding voxels. (e-f) Absorption spectrum of the water-suppressed signal using HSVD method and the quantified signal (red) in the two corresponding voxels. (g-h) Absorption spectrum of the water-suppressed signal using the Hankel-tensor method and the quantified signal (red) in the two corresponding voxels.

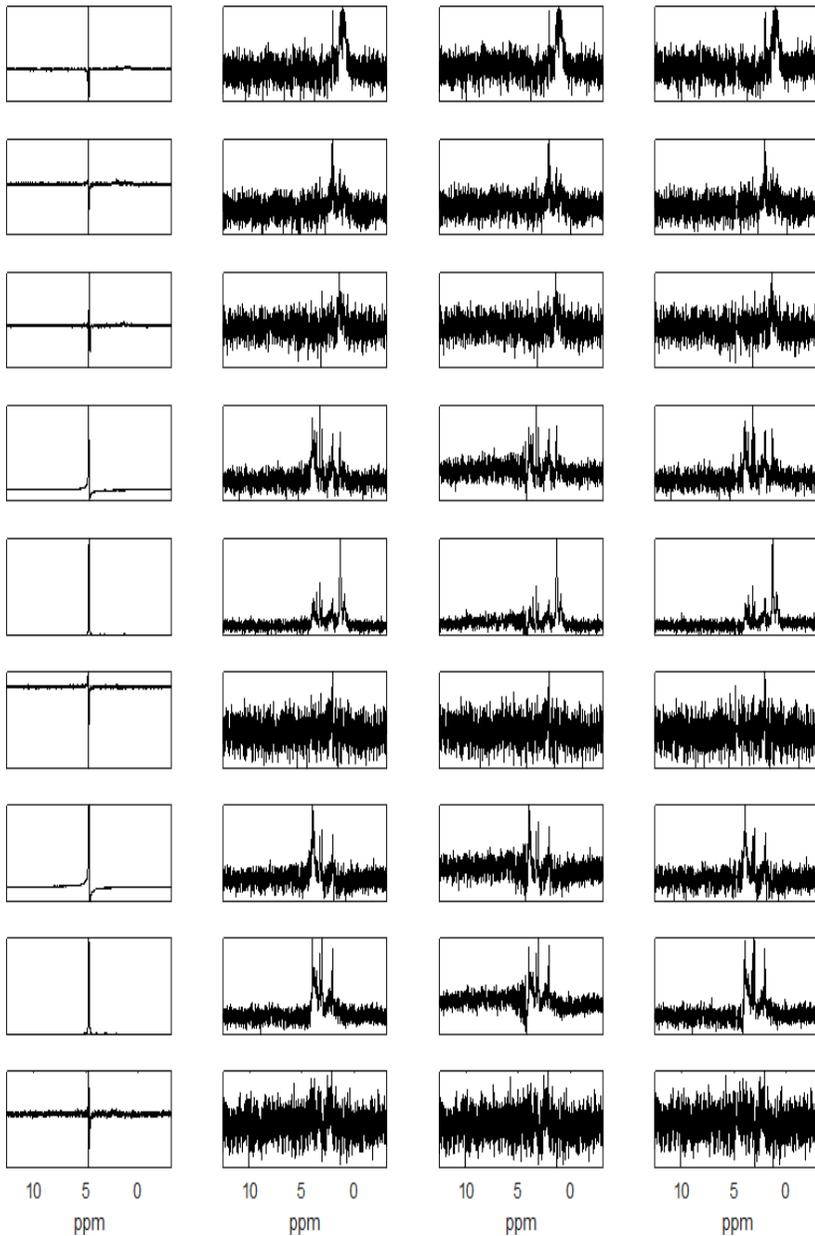


Figure 4.7: Original and residual water suppressed spectra from different locations in 16×16 voxel grid. First column: Un-suppressed original spectra, second column: water suppressed spectra using HSVD, third column: water suppressed spectra using the Hankel-tensor method, fourth column: water suppressed spectra using the Löwner tensor method. Voxel position in the 16×16 grid: Row-1: top left corner, Row-2: middle left, Row-3: bottom left corner, Row-4: middle top, Row-5: centre, Row-6: bottom middle, Row-7: top right corner, Row-8: middle right, Row-9: top right corner.

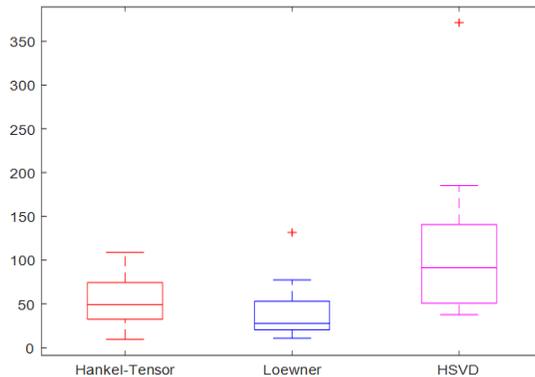


Figure 4.8: Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals. The error is calculated as the difference between the variance in water region segment and the variance from a segment in the noise region.

There is no ground truth available for in-vivo data to measure the quality of water suppression. To measure the performance, we calculate the difference in sample variance between the water region segment and the noise segment in each voxel. The spectrum in the region of 4.2-5.2 ppm is considered as water segment and the spectrum at the outer edges is considered as noise segment. For each MRSI signal, the average difference in variance is used as the performance measure. Figure 4.8 shows the boxplot of sample variance difference of 28 MRSI in-vivo data signals for the HSVD, Hankel tensor and Löwner methods.

Several fast acquisition schemes generates MRSI signals with small sample length to reduce the acquisition time. Therefore, we also applied the proposed methods on in-vivo MRSI signal with smaller sample length, which are obtained by truncating the original signal to 1024, 512 and 256 samples. Fig. 4.9- 4.11 shows the boxplot of sample variance difference using HSVD, Hankel tensor and Löwner methods for 28 MRSI in-vivo data signals with 1024, 512 and 256 samples per voxel, respectively. Table 4.1 shows the average error (difference in variance) over 28 in-vivo MRSI signals for Löwner-tensor, HSVD and Hankel-tensor methods. From the table we can observe that the Löwner-tensor method has better performance compared to HSVD for both original (2048) and truncated (1024, 512 and 256) FID signals. The Hankel-tensor method performs better on the original un-truncated signals and has similar performance in truncated (1024, 512 and 256) FID signals compared to HSVD.

Next, we tried to analyse the quality of water suppression by quantifying MRSI signals and examining the Cramér-Rao bounds of the metabolite amplitudes

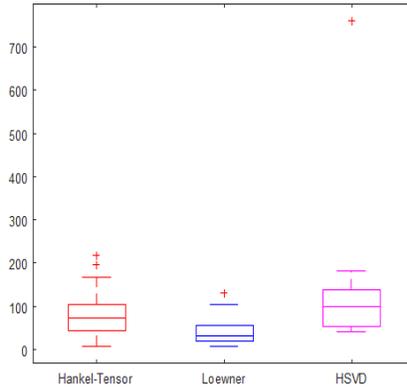


Figure 4.9: Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals truncated to 1024 samples per voxel. The error is calculated as the difference between the variance in the water region segment and the variance from a segment in the noise region.

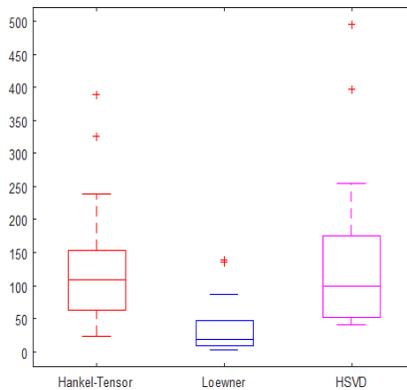


Figure 4.10: Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals truncated to 512 samples per voxel. The error is calculated as the difference between the variance in the water region segment and the variance from a segment in the noise region.

[35]. Two in-vivo MRSI signals measured from brain tumor patients with a grid size 16×16 were used in this analysis. A band of three voxels at the outer edges of the MRSI grid were omitted to avoid chemical shift displacement artefacts and bad quality spectra. This resulted in a reduced MRSI grid size of 10×10 . AQSES [144] was used to quantify the MRSI signal in Matlab based SPID software [143]. An in-vitro basis set consisting of two lipids (Lip1 and Lip2),

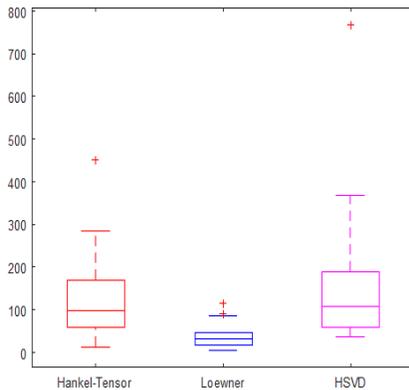


Figure 4.11: Boxplots of error after residual water suppression using the methods HSVD, Löwner and Hankel-tensor in 28 in-vivo MRSI signals truncated to 256 samples per voxel. The error is calculated as the difference between the variance in the water region segment and the variance from a segment in the noise region.

Table 4.1: Average residual water suppression error (difference in variance) obtained from HSVD, Löwner and Hankel-tensor methods over 28 in-vivo MRSI signals. The results are shown for original MRSI signal with 2048 samples for each voxel and truncated MRSI signals with 1024, 512 and 256 samples for each voxel.

FID length	HSVD	Hankel-tensor	Löwner-tensor
2048 (original)	100.26	53.56	37.33
1024	121.86	80.27	40.82
512	133.00	127.36	33.57
256	148.81	122.56	37.49

phosphocholine (PCh), Cre, Glu, glutamine (Gln), MI, Lac, N-Acetyl-Aspartate (NAA) and glycine (Gly) metabolites was used in the AQSES algorithm. Table 4.2 shows the average Cramér-Rao bounds in % of quantified amplitude of five metabolites along with the standard deviation for two of the in-vivo MRSI signals. In the first patient all three methods perform well and the average Cramér-Rao bounds are similar. However, for the second patient the metabolites estimated from the MRSI signals after water suppression using HSVD clearly show higher Cramér-Rao bounds for Glutamate and lipid metabolites because the HSVD method fails to suppress the water signal properly in many of the voxels as shown in Figure 4.12. From Figure 4.12, we can clearly see that

the improper suppression of the water signal results in bad quantification of metabolites such as Glutamate. Even though the suppression of the water signal in two of those voxels using Hankel tensor is better than using HSVD, the quantification is not good compared to Löwner water suppression.

Table 4.2: Mean and standard deviation of Cramér-Rao bounds in % of quantified amplitude for Lipid (Lip1), Glutamate, N-acetylaspartate (NAA) and phosphocholine (PCh) metabolites over 10×10 MRSI voxel grid.

Patient-1						
	Mean			Standard deviation		
	HSVD	Hankel-tensor	Löwner-tensor	HSVD	Hankel-tensor	Löwner-tensor
Lipid	19.7016	21.1315	19.9462	11.7981	12.5221	12.2064
Glutamate	22.5448	25.9515	21.2197	9.7971	8.9112	9.3843
NAA	9.5400	10.9208	9.3868	4.3972	5.5443	4.0182
PCh	9.5269	9.7131	9.3248	3.4515	2.7612	3.1800
Patient-2						
	Mean			Standard deviation		
	HSVD	Hankel-tensor	Löwner-tensor	HSVD	Hankel-tensor	Löwner-tensor
Lipid	511.8554	51.7515	46.8847	4693.97	91.6301	86.9021
Glutamate	37.1549	38.1211	27.5632	29.6739	21.7350	12.5496
NAA	13.1994	14.7912	13.3217	4.5156	5.4552	4.4907
PCh	14.9876	14.1948	15.0810	8.9005	9.0866	12.4655

4.4 Discussion

Residual water suppression is one of the common preprocessing steps used in the quantification of MRSI signals. T. Sundin et. al. [170] propose a maximum-phase finite impulse response filter for residual water suppression. This method will alter the amplitude and phase of the filtered signal, which may create problems in quantification methods where different phase variations are not allowed and where spectra instead of quantified metabolites are used for the analysis [21, 111]. HSVD is the most widely used method and is available as a preprocessing step in many MRSI software packages. As it is applied on a voxel-by-voxel basis and as it computes the water source components separately for each voxel, it does not exploit the information shared among the voxels in the MRSI grid. Therefore, the HSVD method can fail to suppress water completely in a particular voxel due to noise or artifacts present in that voxel. This problem can be seen in an in-vivo example shown in Figure 4.6 where the HSVD method fails to suppress the water in a particular voxel (Figure 4.6f) and performs better in another voxel (Figure 4.6e). This has motivated us

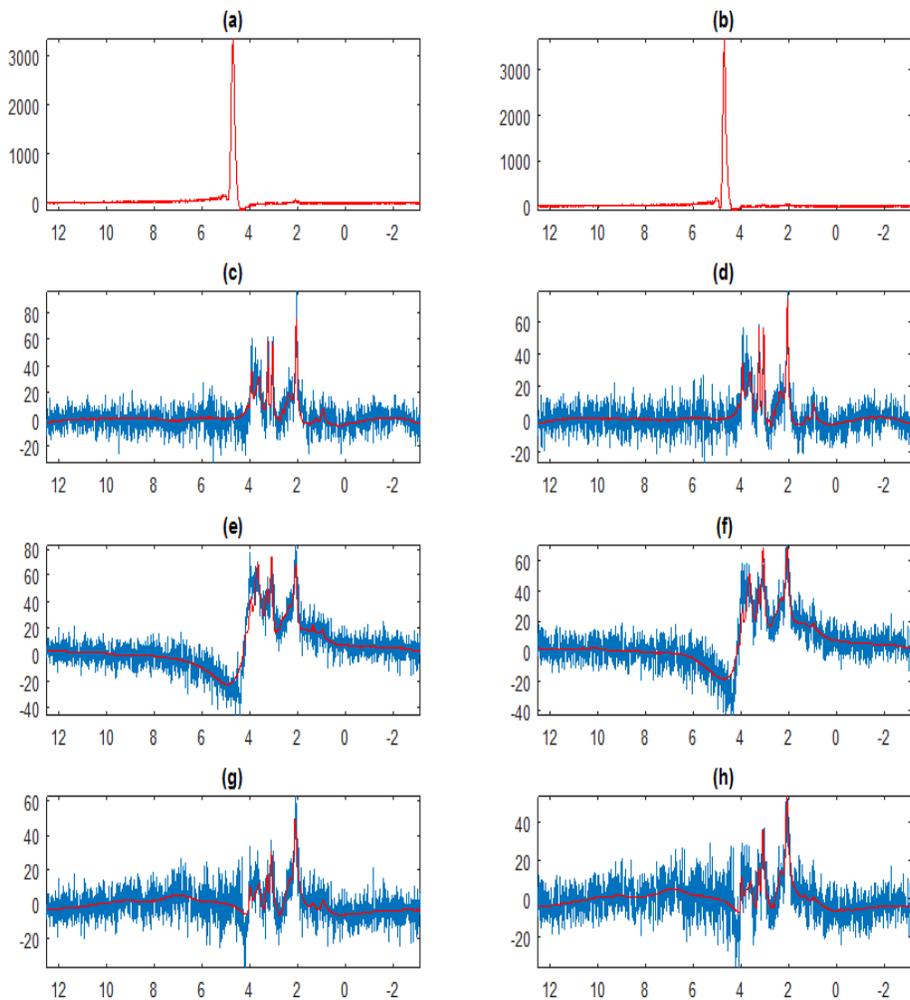


Figure 4.12: Improper water suppression using the HSVD method in two voxels of the 10×10 MRSI grid. (a-b) Absorption spectra of measured MRSI signal with large residual water peak for two voxels. (c-d) Absorption spectrum of the water-suppressed signal (blue) using the Löwner method and the quantified signal (red) in the two corresponding voxels. (e-f) Absorption spectrum of the water-suppressed signal using HSVD method and the quantified signal (red) in the two corresponding voxels. (g-h) Absorption spectrum of the water-suppressed signal using the Hankel-tensor method and the quantified signal (red) in the two corresponding voxels.

to develop a new algorithm, which can exploit the similarity in water sources present among all voxels.

In this chapter, we have represented the second-order MRSI data using a third-order tensor by means of a Löwner transform. A novel residual water-suppression method based on CPD, where sources are shared among many voxels, was developed. This work explored the feasibility and efficiency of the proposed algorithm in suppressing the residual water from MRSI data using both simulation and in-vivo signals. The Löwner-based method is applied simultaneously on the entire MRSI grid to estimate a large number of sources which can be used, in various combinations, to model the water component in each voxel. The water signal in each voxel is estimated as a linear combination of the sources with different voxel specific weights. This helps in preventing the failure of the water suppression in single voxels. The Löwner tensor is constructed using truncated spectra. Only the parts where the metabolite and water peaks are present are retained in the spectra. This helps in reducing the size of the tensor and the computational complexity of the algorithm without any significant impact on water suppression quality. The Löwner-based method requires fewer parameters to model the water signal in the MRSI grid. For example, in the HSVD method with a rank-50, for each voxel we estimate 50 complex amplitude + 50 complex poles. For an MRSI grid of 8×8 the total number of free parameters includes 50×64 complex amplitudes and 50×64 complex poles. In case of the Löwner method with a rank-50 and polynomial degree 4, the total number of free parameters for an MRSI grid of 8×8 includes 50 complex poles + 54×64 complex amplitudes.

When using the Löwner method to estimate the water sources we have applied the CPD only on the compressed core tensor S and the factor matrices are obtained as explained in the Section 2.1.1. It speeds up the algorithm without any significant negative consequences in the estimated water signal. In general, the factor matrices obtained from the compressed CPD step are typically used as the initial values in the computation of the CPD of the full tensor to further improve the decomposition, also known as the refinement step. Here, we have not used the refinement step as it is computationally intensive and did not provide any significant improvement in the estimated water signal. In the Löwner-based method we have added a quality verification on the water suppression to overcome the problems with random initialization. If the algorithm is used with different initializations to obtain a better water suppression, the compression step is only performed once as it is deterministic. Since the compression step takes most of the computation time and the CPD on the core tensor is relatively fast, running the algorithm again with a different initialization will not increase the computation significantly. A rank $R = 50$ was used for the CPD of the Löwner tensor \mathcal{L}_S based on the assumption that the water signal from all

the voxels can be modelled using 20-25 first-order rational functions and the remaining ones are sufficient to model the metabolites. The chosen rank was not sensitive to the grid size in the sense that similar performance was obtained on the larger voxel grid (16×16) as well as on the smaller voxel grid (8×8). Also, the selection of the rank itself is not so sensitive since the results did not change significantly when increasing (e.g, $R = 60$) or decreasing the rank (e.g, $R = 40$).

The Hankel-tensor based method has been used to estimate the parameters of exponentially damped sinusoids of multichannel signals [138]. In this chapter we have used the Hankel-tensor based method as a natural extension of the HSVD method to estimate the water signals from MRSI data. The Hankel-tensor method has better performance compared to the HSVD method in both simulations and in-vivo data, however its performance is worse than the Löwner method. Both Hankel-tensor and HSVD methods extract sources with broad peaks, which helps in tackling the baseline without the need for additional polynomial sources. A higher model-order (rank) of 100 was used, as the performance deteriorated when a model order similar to the HSVD-rank (50) was used. The reason is that a rank of 50 is typically sufficient for modelling a signal from an individual voxel, but is not sufficient to capture all the variations in damping and frequency shifts across signals in the MRSI grid. The Hankel tensor requires a higher model order (rank) compared to the Löwner method because the frequency domain decouples the noise, artefacts and trends present outside the region of interest and also a higher rank was needed to model the variations in trend across different voxels. The Hankel-tensor was constructed using the entire FID signal of length 2048, which resulted in a large tensor and higher computational complexity. A truncated FID (<2048 samples) can also be used to construct the Hankel-tensor to reduce the size and complexity, but its performance is worse than HSVD in in-vivo MRSI data. It is not clear why the performance of the Hankel-tensor based method does not perform as good as the Löwner based method. One possible reason is that in the frequency domain, the water signal and artefacts/nuisance parameters present at different frequency range can be decoupled, which is exploited by Löwner based method for extracting the water signal.

All three methods model the total MRS signal sufficiently well, however in some voxels HSVD fails to suppress the water completely. This happens mainly because the water signal is modeled from the sources whose frequencies are outside the region of interest (0.25 - 4.2 ppm) and in some cases (e.g., Figure 4.6) part of the water signal is modelled by a few sources whose frequencies lie in the region of interest (0.25 - 4.2 ppm). Therefore, a small residual water is present in the water suppressed signal, since part of the water signal modelled by sources in the region of interest (0.25 - 4.2 ppm) is not subtracted.

As a proof of concept, we have analyzed the proposed methods in terms of residual error in case of simulation and difference in variance for in-vivo examples. Also, we have assessed the water suppression quality of three methods using the average Cramér-Rao bounds of the metabolite amplitude in two in-vivo patients. Although Cramér-Rao bounds depend on more factors, they are often used in in-vivo studies to assess the reliability of metabolite quantification. In the future, these water suppression methods could also be compared in terms of reliability of metabolite quantification in a test-retest experiment.

4.5 Conclusion

A tensor-based method which suppresses residual water simultaneously in all MRSI voxels using a Löwner-based blind source separation technique and Hankel-tensor based exponential data-fitting technique are proposed. These methods were tested on both simulated and in-vivo ^1H MRSI signals. In both cases the tensor based methods perform better than the widely-used subspace-based HSVD method, which uses a single Hankel matrix from one spectrum at a time. Comparing the two tensor based methods, the Löwner-tensor based method was shown to better suppress residual water in MRSI. The main advantage of Löwner-based method is that it can handle the presence of artifacts in some voxels without significantly affecting the water suppression quality. In contrast, the HSVD method completely fails to suppress water in some voxels when artifacts are present, thus making the further analysis of those spectra difficult.

Chapter 5

Un-supervised tissue type differentiation in glioma

Magnetic resonance spectroscopic imaging (MRSI) reveals chemical information that characterizes different tissue types in brain tumors. Blind source separation techniques such as non-negative matrix factorization and hierarchical non-negative matrix factorization are used to extract the tissue-specific profiles and their corresponding distribution from the MRSI data. In this chapter we focus on automatic detection of the tumor, necrotic and normal brain tissue types using tensor based methods. A 3-dimensional MRSI tensor is constructed from in-vivo 2D-MRSI data of individual glioma patients. Non-negative canonical polyadic decomposition (NCPD) is applied to the MRSI tensor to differentiate various tissue types and to extract the tissue-specific profiles and their corresponding distribution. NCPD is applied to 28 MRSI datasets of UZ Leuven and compared with previous matrix-based decompositions. The work presented in this chapter is based on [18].

5.1 Introduction

Accurate characterisation and localization of pathologic tissue types play a key role in diagnosis and treatment planning of brain tumors. The tumor region of glioblastoma multiforme (GBM) could consist of several tissue types, which represent actively growing tumor, necrosis or normal brain tissue [111]. In recent years, many advanced MR modalities such as Magnetic Resonance

Spectroscopic Imaging (MRSI), perfusion-weighted MRI (PWI) and diffusion weighted MRI (DWI) are being used to characterize brain tumors and detect full tumor extent [58]. MRSI is a non-invasive imaging technique that provides spectral profiles in a two- or three- dimensional voxel grid, from which the spatial distribution of metabolite concentrations can be estimated. Each voxel in the MRSI grid has a spectrum composed of several peaks corresponding to the metabolites present in that grid. The area under the peak is proportional to the metabolite concentration. MRSI has been successfully applied to diagnosis and prognosis of brain tumors. There are many algorithms for MRSI data analysis available in the literature that aim at tissue characterisation, tumor localization and classification. In particular, nonnegative matrix factorization (NMF) and hierarchical NMF (hNMF) have shown potential to differentiate different tissue patterns in MRSI of GBM patients [111]. However, the performance of hNMF deteriorates in the presence of artifacts because it can handle only three tissue types (tumor, necrotic and normal).

Higher order tensors have certain properties that are not present in a matrix [38]. Tensor decompositions are now being used in various biomedical applications like genomics [131], EEG and fMRI data analysis [129] and smart patient monitoring [179]. The 2D MRSI signal can naturally be represented as a third-tensor \mathcal{P} as shown in Fig. 5.1. The mode-1 and mode-2 of the tensor \mathcal{P} represent the spatial dimension of the 2-D MRSI signal and mode-3 represents the spectra from all voxels. This motivated us to explore tensor based approached to obtain more refined tissue characterisation and localization.

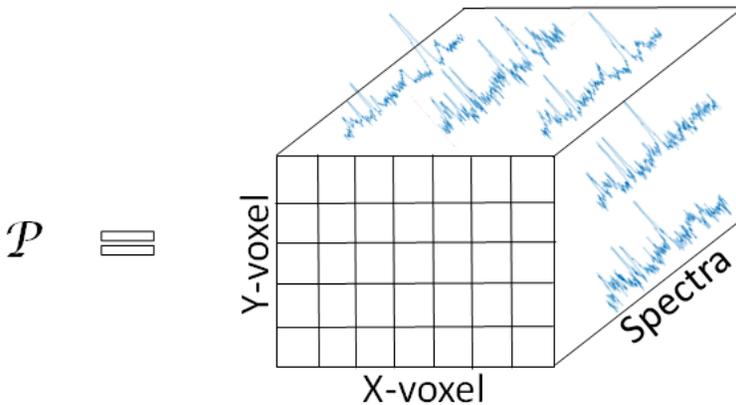


Figure 5.1: 3-way spatial tensor representation of 2-D MRSI data.

To extract different tissue types from the spatial tensor \mathcal{P} , we can use block term

decomposition (BTD) in R rank- $(L_r, L_r, 1)$ block terms [51]. The $(L_r, L_r, 1)$ BTD for a third-order tensor $\mathcal{P} \in \mathbb{R}^{I \times J \times K}$ can be written as:

$$\mathcal{P} \approx \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^T) \circ \mathbf{s}_r$$

where \mathbf{s}_r represents the tissue-specific spectral pattern, $\mathbf{A}_r \mathbf{B}_r^T$ having rank L_r represents the corresponding spatial distribution with $\mathbf{A}_r \in \mathbb{R}^{I \times L_r}$, $\mathbf{B}_r \in \mathbb{R}^{J \times L_r}$ and ‘ \circ ’ represents outer product. The rank L_r for each tissue type plays an important role in the decomposition, which depends on factors like size and shape of the tissue type. The rank L_r is not known a priori and it is difficult to estimate it from the MRSI data. Also there is no guarantee that the spatial distribution follows a low-rank structure so that it can be represented by a low rank matrix $\mathbf{A}_r \mathbf{B}_r^T$. Because of these problems, we developed a new tensorization of the 2-D MRSI signal which allows to exploit the low-rank structure.

In this chapter, we propose a novel algorithm for the detection of tumor, necrotic and normal tissue types from MRSI signals. The algorithm first applies a window method to enhance the peaks and reduce the length of the spectra, and then constructs a 3-D MRSI tensor. Decomposing this tensor using NCPD with common factor in mode-1 and mode-2, allows to retrieve the tissue-specific spectral profiles from the NCPD factor matrices. Preliminary studies have been previously presented in [20]. The chapter is organized as follows: In Section 5.2, the construction of the MRSI tensor and the NCPD algorithm for tissue type differentiation is explained. The performance evaluation of the proposed NCPD algorithm in comparison with one-level NMF and hNMF using short-echo time (TE) ^1H 2D-MRSI datasets from glioma patients is done in Section 5.3. Discussions are presented in Section 5.4 and finally the chapter is concluded in Section 5.5.

5.2 Method

5.2.1 MRSI tensor construction

For each voxel in the MRSI grid, a reduced real-valued spectrum \mathbf{x} is constructed from the corresponding complex-valued pre-processed spectrum. Elements of the vector \mathbf{x} are obtained by moving an overlapping window over the spectrum, where the i^{th} element of \mathbf{x} is the sum of squares of absolute value of all the

elements in the i^{th} window segment,

$$\mathbf{x}(i) = \sum_{j=1}^L s_i(j)s_i^*(j)$$

s_i is the spectrum at the i^{th} segment, ss_i^* is its complex conjugate and L is the length of the window segment. Fig. 5.2a shows the construction of a vector \mathbf{x} from the spectrum. The resulting vector X can be considered as a denoised and reduced-length version of the original spectrum. The window length L is chosen such that it covers the widest peak in the spectra and step value is chosen as $\frac{L}{4}$. Use of vector \mathbf{x} has the following advantages:

1. It reduces the length of spectra without losing vital information required for tumor tissue type differentiation.
2. It gives more weight to the peaks and makes the signal smoother and non-negative.

Instead of sum of squares we can use mean value over the window to reduce the length of the spectra. However, it does not enhance the peaks which contain most of the information.

The 2D-MRSI signal is naturally a third-order tensor. However, it does not exhibit low-rank structure. This prevents us from extracting tissue specific patterns by applying tensor decomposition methods directly on the 2D-MRSI. Other tensors obtained by stacking Löwner/Hankel matrix from each voxel can have low-rank structure. Using tensor decomposition such as CPD may extract individual peaks as sources. Such kind of problems will also arise when using NMF methods. These individual peaks do not correspond to a tumor or a normal tissue signature, which are composed of many peaks. Rank- $(L_r, L_r, 1)$ BTD can be used to extract sources containing many peaks corresponding to different tissues. Since signatures of tumor and normal tissue vary from patient to patient, it is not feasible to estimate the rank L_r for extracting good tumor and normal specific tissue signatures. Therefore, we developed a new way to construct tensor from reduced spectra to overcome such problems. A third-order MRSI tensor \mathcal{T} is constructed by stacking \mathbf{xx}^T from all the voxels in the MRSI grid as shown in Fig. 5.2(b).

Under the assumption that each voxel correspond to a specific tissue, the MRSI tensor \mathcal{T} will have a low-rank structure. The \mathbf{xx}^T structure in MRSI tensor couples the peaks in the reduced spectra and prevents extracting individual peaks as sources.

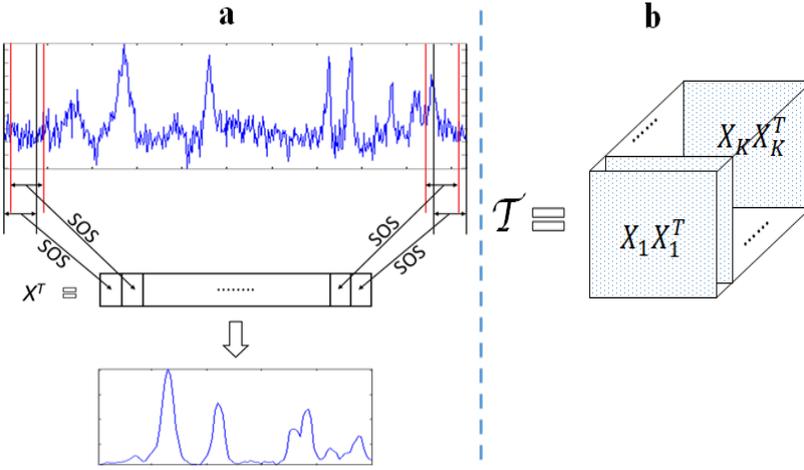


Figure 5.2: (a) Construction of reduced spectrum, \mathbf{x} from the pre-processed spectra. SOS: sum of squares. (b) Construction of the MRSI tensor \mathcal{T} from the reduced spectra, \mathbf{x} . K is the total number of voxels in the 2D MRSI excitation volume.

5.2.2 Non-negative CPD

Non-negative canonical polyadic decomposition (NCPD) is a tensor decomposition method, where the tensor is decomposed into a sum of rank-one tensors with non-negativity constraints on the factor matrices [37].

$$\mathcal{X} \approx [\mathbf{A}\mathbf{B}\mathbf{C}] \equiv \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad \mathbf{A}, \mathbf{B}, \mathbf{C} \geq 0$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R] \in \mathbb{R}_+^{I \times R}$, $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_R] \in \mathbb{R}_+^{J \times R}$ and $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_R] \in \mathbb{R}_+^{K \times R}$ are non-negative factor matrices. R is the rank, defined as the number of rank-one terms.

In the MRSI tensor \mathcal{T} , the frontal slices are symmetrical, therefore we constrain the frontal slices of each NCPD rank-one term to be symmetric. To maintain symmetry, a common factor matrix is used for mode-1 and mode-2 in the NCPD as shown in Fig. 5.3. After performing the NCPD on the MRSI tensor \mathcal{T} we obtain two factor matrices \mathbf{S} and \mathbf{H} , where \mathbf{S} represents the tissue-specific patterns of the reduced spectra and \mathbf{H} represents the spatial distribution of each tissue type.

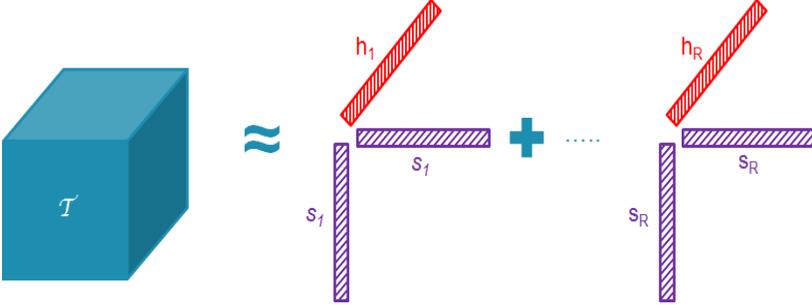


Figure 5.3: Non-negative CPD of MRSI tensor \mathcal{T} : MRSI tensor \mathcal{T} is decomposed into R rank-1 tensors. Common factor \mathbf{S} is used in mode-1 and mode-2 to maintain symmetry of frontal slices. Each \mathbf{s}_i gives a tissue-specific reduced spectral pattern and the corresponding \mathbf{h}_i gives the spatial distribution of the respective tissue type, upon reshaping. Non-negativity of \mathbf{S} and \mathbf{H} is imposed in the decomposition.

Each rank-one term obtained from the NCPD of the MRSI tensor \mathcal{T} is expected to correspond to a particular tissue type, although it is not always guaranteed with the current formulation. Here, we assume that in most of the voxels the spectra belong to a particular tissue type, with the exception of a few voxels whose spectra may contain a mixture of at most two tissue types. Therefore, the factor matrix H will be sparse, meaning that each row will mostly have only one high value. A further refinement of the NCPD method exploits the sparsity assumption in the factor matrix \mathbf{H} by imposing a l_1 regularization on it. The NCPD with l_1 regularization (NCPD- l_1) can be written as

$$[\mathbf{S}^*, \mathbf{H}^*] = \arg \min_{\mathbf{S} \geq 0, \mathbf{H} \geq 0} \|\mathcal{T} - \sum_{i=1}^R \mathbf{S}(:, i) \circ \mathbf{S}(:, i) \circ \mathbf{H}(:, i)\|^2 + \lambda \|\text{Vec}(\mathbf{H})\|_1 \quad (5.1)$$

where \mathbf{S} and \mathbf{H} are the aforementioned factor matrices and λ is the parameter which controls the sparsity.

In this work, the tensor decomposition was performed using the Tensorlab Matlab package [185]. Non-negativity constraints, common factors to maintain symmetry and l_1 regularization are imposed using the structured data fusion framework [167] available in Tensorlab. In structured data fusion framework,

transformation of variables is used to convert the constrained optimization problem in Equation (5.2.2) into an unconstrained problem, where squaring was used to get non-negative variables. Gauss-Newton algorithm with dogleg trust region (GNDL) is used to solve the resulting non-linear least-squares problem [167]. At each iteration of the GNDL algorithm, a step is calculated by iteratively solving a linearized version of Equation (5.2.2) using conjugate gradients. The SDF framework allows straightforward incorporation of regularization terms into the objective function.

5.2.3 Spectral recovery and non-negative least squares

The NCPD of the MRSI tensor \mathcal{T} gives the factor matrix \mathbf{S} , which contains the reduced spectral patterns specific to different tissue types. However, it is desirable to recover tissue-specific spectral sources as vectors of the same length as the pre-processed MRSI spectra, which are more interpretable since they can be directly compared to the original spectra. To this end, a least squares problem is solved, and the matrix of spectral sources representing the tissue-specific spectral patterns \mathbf{W} is:

$$\mathbf{W} = (\mathbf{H}^\dagger \mathbf{Y}^T)^T \quad (5.2)$$

where \mathbf{H}^\dagger is the Moore-Penrose pseudoinverse of the NCPD factor matrix \mathbf{H} and \mathbf{Y} is the matrix containing the unit normalized spectra from all voxels as its columns. The source spectra \mathbf{W} can be estimated as complex-valued or magnitude vectors by using complex-valued or magnitude spectra \mathbf{Y} in equation (5.2.3), respectively.

In the NCPD of MRSI tensor \mathcal{T} , the factor matrix \mathbf{H} corresponds to the weights of the linear combination of $\mathbf{S}(:, i)\mathbf{S}(:, i)^T$ and not the linear combination of source spectra \mathbf{W} . Also, the tensor is constructed using the normalised spectra, therefore voxels having relatively small values compared to other spectra will also get enhanced and will have higher abundances. The abundances of the sources in the original unnormalised spectra are more meaningful and represent the true distribution. Also we want the abundances to represent the weights in the linear combination of source spectra \mathbf{W} . To address these problems, spatial distributions, \mathbf{H}_D of the different tissue types is calculated using non-negative least squares with l_1 regularization:

$$\mathbf{H}_D(:, i) = \arg \min_{x \geq 0} \|\mathbf{W}x - \mathbf{Y}_{un}(:, i)\|_2^2 + \lambda_1 \|x\|_1 \quad (5.3)$$

where, $H_D(:, i)$ is the distribution of source spectra in each voxel, \mathbf{W} contains the estimated source spectra and $\mathbf{Y}_{un}(:, i)$ is the original unnormalised spectrum

of each voxel. The problem in equation (5.2.3) is solved for all the voxels in the MRSI grid using a Matlab based large-scale l_1 -regularized least squares problem solver [98]. When the estimated source spectra \mathbf{W} and the MRSI voxel spectra are complex-valued, the real and imaginary part are concatenated to form a single real-valued matrix, which is then used as input to the non-negative least squares problem with l_1 regularization. Tissue distribution maps are obtained by reshaping the rows of \mathbf{H}_D to the 2D-MRSI grid.

5.2.4 Initialization

The NCPD algorithm needs initial values for \mathbf{S} and \mathbf{H} and the non-negative least squares [98] algorithm needs initial values for \mathbf{H}_D . Initializing \mathbf{S} , \mathbf{H} and \mathbf{H}_D with uniformly distributed pseudorandom numbers between 0 and 1 gives good solution, but the results are not exactly the same between different runs. Although the solutions are similar between different runs, poor initial values may result in sub-optimal solutions. To find good initialization values, first we take the singular value decomposition (SVD) of the matrix \mathbf{Y} , $\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^H$, where the columns of the matrix \mathbf{Y} are the complex spectra from each voxel. Reduced spectra \mathbf{S}_{init} are constructed from R dominant left-singular vectors as explained in Section 5.2.1 and are used as initial value for \mathbf{S} in the NCPD algorithm. Initial values for \mathbf{H} are obtained through least squares as $\mathbf{H}_{init} = (\mathbf{S}_{init}^\dagger \mathbf{M})^T$, where \mathbf{M} is a matrix whose columns are the reduced spectra \mathbf{x} from each voxel. Least squares may introduce negative values in H_{init} . However, these negative values are typically rare and small in amplitude. Moreover, the NCPD algorithm in Tensorlab can handle negative initial values. A vector of all ones was used to initialize \mathbf{H}_D in the final non-negative least squares step.

5.2.5 Source number estimation

The NCPD algorithm needs the number of sources (i.e. decomposition rank) as input. Estimating the rank from the input spectra/ tensor \mathcal{T} is a difficult problem. The literature on estimation of decomposition rank from the tensor is limited. Tensorlab package [185] has a method *rankest*, which estimates the rank based on the L-curve of the number of rank-one terms in a CPD. However, this method gives good results when the noise is low or when the decomposition is exact, which is not the case for our MRSI tensor \mathcal{T} . The estimated rank R from this method is much higher than the required number of sources for a good tissue type differentiation. In [172] a Bayesian model based on automatic relevance determination is proposed for NMF, which also estimates the model order R along with non-negative factor matrices. In this method, the estimated

model order is dependent on the choice of the dispersion parameter, which represents the tradeoff between the data fidelity and the regularization terms. Selecting the optimal dispersion parameter for each MRSI dataset is difficult and it is as hard as selecting the rank itself. In this chapter we will use the covariance matrix based approach to estimate the number of sources.

Let \mathbf{A} be the data matrix of size $K \times N$ where the rows represent K spectra of length N from all the voxels in the MRSI grid. Then the $K \times K$ sample covariance matrix is estimated as $\mathbf{C} = \frac{1}{N-1}[(\mathbf{A} - \bar{\mathbf{A}}\mathbf{1}_N^T)(\mathbf{A} - \bar{\mathbf{A}}\mathbf{1}_N^T)^T]$, where $\bar{\mathbf{A}} = \frac{1}{N-1} \sum_{i=1}^N \mathbf{a}_i$ and $\mathbf{1}_N$ is a vector of all ones with length N . The eigenvalues of the covariance matrix \mathbf{C} are denoted by $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_K$. The number of sources is estimated as the minimum number R such that the cumulative sum of eigenvalues is greater than 99% of the sum of all eigen-values.

$$R^* = \min R, \quad \text{such that} \quad \left[\frac{\sum_{i=1}^{i=R} \lambda_i}{\sum_{i=1}^{i=K} \lambda_i} \geq 0.99 \right] \quad (5.4)$$

where R^* is the estimated number of sources. When the data matrix \mathbf{A} is constructed from the original complex valued spectra, the estimated number of sources is high. Therefore, we use the reduced spectra in \mathbf{A} to calculate the covariance matrix \mathbf{C} . Reduced spectra suppress noise and small variations present in the original spectra resulting in fewer eigenvalues significantly larger than zero in the covariance matrix \mathbf{C} . Therefore, using reduced spectra to estimate R provides a good estimate for the number of sources in many MRSI datasets. However, when the MRSI data contain spectra of less quality or having more artifacts the estimated number of sources is still too high. To overcome this problem we incorporate prior knowledge about the maximum number of sources (includes tissue types + artifacts) for estimating the number of sources. Let the maximum number of sources be P . If R^* obtained from (5.2.5) is less than or equal to P ($R^* \leq P$), then the number of sources is set to R^* and is used in the NCPD algorithm. When R^* obtained from (5.2.5) is greater than P ($R^* > P$), only the largest $P + 1$ eigenvalues of C are retained and the remaining ones are set to zero. Then the number of sources is estimated as in equation (5.2.5) with K set to $P + 1$, i.e the set of eigenvalues is reduced to the largest $P + 1$ values only.

5.2.6 Source and distribution correlation

The performance evaluation of the algorithms in the in-vivo study was analyzed using two measures:

1. Source correlation: In this chapter we have defined two types of source correlation:

- (a) Source correlation Type I ($SC1$): The source correlation is calculated as Pearson's linear correlation coefficient between the estimated source spectrum and the tissue-specific spectrum based on expert labeling of the in-vivo MRSI voxels [111]. The tissue-specific spectrum based on expert labeling (src_expt) for a particular tissue type is computed as the average of all the spectra from the voxels labelled by the expert as belonging to that tissue type. The construction of expert spectra, src_expt for the tumor tissue is shown in Fig. 5.4.

$$src_expt = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$$

$$SC1 = r(\mathbf{W}(:, T), src_expt)$$

where $\mathbf{W}(:, T)$ is the estimated source spectrum corresponding to a particular tissue type, src_expt is the expert labeled spectrum corresponding to that particular tissue type, Y_1, Y_2, \dots, Y_n are the spectra from the voxel that are marked by the expert as belonging to that particular tissue type and r is Pearson's linear correlation coefficient.

- (b) Source correlation Type II ($SC2$): First Pearson's linear correlation coefficients are calculated between the estimated source spectrum and all the spectra in the voxels that are marked by the expert as belonging to a particular tissue type. Then the source correlation $SC2$ is calculated by taking the average of Pearson's linear correlation coefficients:

$$SC2 = \frac{r(W(:, T), Y_1) + \dots + r(W(:, T), Y_n)}{n}$$

where $SC2$ is the Type II source correlation, $W(:, T)$ is the estimated source spectrum corresponding to a particular tissue type.

2. Distribution correlation (DC): Distribution correlation is calculated as Pearson's linear correlation coefficient between the estimated distribution map corresponding to a particular tissue type and the distribution map based on expert labeling ($dist_expt$). For each tissue type, a distribution map based on expert labeling is obtained by using values equal to the l_2 norm of the corresponding spectra for all voxels labeled as a certain tissue class, and values of 0 for the other voxels as shown in Fig. 5.4.

$$DC = r(H_D(:, T), dist_expt)$$

where DC is the distribution correlation, $H_D(:,T)$ is the estimated distribution vector corresponding to a particular tissue type, $dist_expt$ is the expert labeled distribution vector corresponding to that particular tissue type.

Because of heterogeneity, the tumor tissue is modelled by more than one source spectrum in many patients. In this case, the average of source spectra and sum of the corresponding distribution maps are used in the calculation of $SC1$ and DC , respectively. Whereas, for $SC2$, only the maximum correlation among the source spectra is retained for averaging. In NCPD algorithm the estimated source spectra and the MRSI voxel spectra are complex signals. Therefore, the real and imaginary part of the complex spectra are concatenated to form a real signal, which is then used in the calculation of correlation. Since absolute spectra were used in NMF and hNMF, source correlation was calculated on the spectra directly.

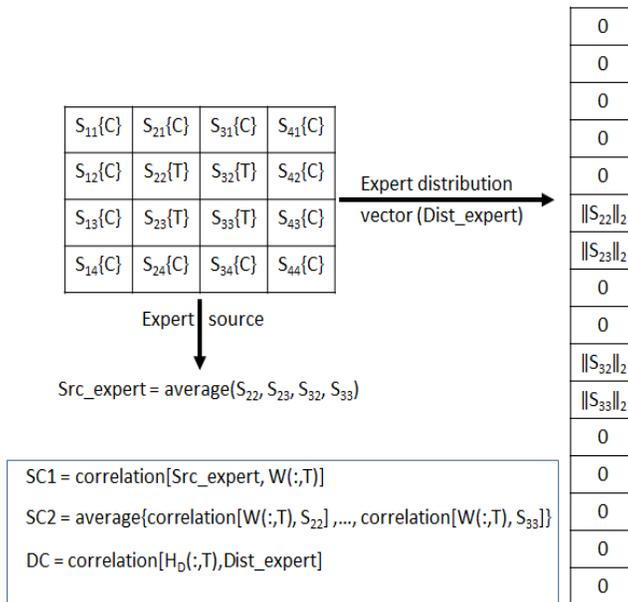


Figure 5.4: Generation of expert labeled tissue-specific (Tumor) spectrum and distribution vector. Calculation of source and distribution correlation is shown in the box. S_{ij} is the spectra at i^{th} column and j^{th} row. The tissue type T-tumor or C-control is shown between braces.

5.3 Results on brain tumor dataset

To test the feasibility of spatial tensor representation as shown in Fig.5.1 in tissue type differentiation, we constructed a spatial tensor \mathcal{P} using magnitude spectra and applied $(L_r, L_r, 1)$ BTD on the spatial tensor for one high grade (grade IV) MRSI dataset. The spatial tensor \mathcal{P} is decomposed into 6 rank $(L_r, L_r, 1)$ terms with rank $L_r = 10, 10, 9, 8, 5, 5$. The corresponding distribution maps of the different sources are shown in Fig. 5.5b. Non-negative constraints are applied to the source mode (mode-3) only. The rank L_r is chosen manually by trying different combinations and selecting the one which gives the best results. For comparison, we have shown the distribution maps obtained from the MRSI tensor \mathcal{T} (shown in Fig.5.2) of the same dataset using NCPD- l_1 algorithm with 6 sources in Fig. 5.5c. The MRSI grid superimposed on anatomical image and the expert labelling is shown in Fig. 5.5a. Comparing the distribution maps in Fig. 5.5b and 5.5c with the expert labelling in Fig. 5.5a, we can observe that the tissue type differentiation is not good using a spatial tensor representation with $(L_r, L_r, 1)$ BTD compared to an \mathbf{xx}^T based tensor representation with NCPD. The output of spatial tensor with $(L_r, L_r, 1)$ BTD is sensitive to choice of L_r and choosing a suitable L_r for all the R sources is difficult. Therefore the results are demonstrated for only one high grade glioma patient.

In order to evaluate the performance and validate the tissue differentiation ability, three algorithms, NCPD (with and without l_1 regularization), single stage NMF and hNMF were applied on 28 in-vivo ^1H MRSI datasets (22 grade IV, 3 grade II and 3 grade II astrocytoma with focal progression to a grade III glioma) from 17 patients with gliomas. The Type I, Type II source correlation and the distribution correlation for tumor and necrotic tissue obtained from NCPD without regularization, NCPD with l_1 regularization, single stage NMF and hNMF are shown as box-plots in Fig. 5.6. From Fig. 5.6 it is clearly evident that source and distribution correlation values are higher and less scattered when using NCPD- l_1 compared to single stage NMF and hNMF. The NCPD algorithm is unable to extract tumor tissue in 2 out of 28 datasets, whereas the single stage NMF and hNMF algorithms do not estimate tumor tissue in 5 datasets. The correlation values of the tissue types which are not recovered are set to zero (Fig. 5.6). A summary of the results i.e, mean, standard deviation (std dev), median, median absolute deviation (MAD) and range is shown in Table 5.1. In case of tumor tissue, NCPD- l_1 has the highest mean and median values for source and distribution correlation. To check whether there is a significant increase in the median, we have performed a one-sided Wilcoxon rank sum test with 1% significance level ($\alpha = 0.01$) [75]. The Wilcoxon rank sum test was performed between the correlations obtained from NCPD- l_1 and other algorithms and the corresponding p-values are shown in Table 5.1. There

(a) MRSI voxel grid and expert label

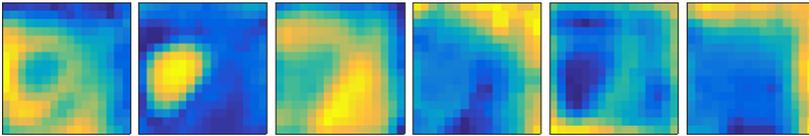
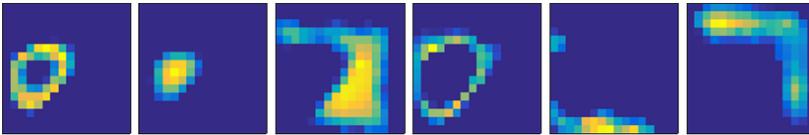
(b) Distribution maps: $(L_r, L_r, 1)$ BTD of spatial tensor \mathcal{P} (c) Distribution maps: NCPD of MRSI tensor \mathcal{T} 

Figure 5.5: (a) Left to right, First image: T2-weighted anatomical MR image of a brain tumor with areas of necrosis. Second image: voxels within the MRSI excitation volume superimposed on anatomical image. Third image: expert labeling, where yellow (horizontal + vertical line) indicates necrotic, red (horizontal line) indicates tumor, dark blue (slanted + horizontal line) indicates normal, light blue (vertical line) indicates normal/tumor and green (slanted line) indicates spectra of poor quality. (b) Tissue distribution maps obtained from an $(L_r, L_r, 1)$ block term decomposition of spatial tensor \mathcal{P} . First three images from the left correspond to tumor, necrotic and normal tissue distribution respectively. Remaining three images correspond to bad spectra/artefact. (c) Tissue distribution maps obtained using the NCPD- l_1 algorithm on the MRSI tensor \mathcal{T} . First three images from the left correspond to tumor, necrotic and normal tissue distribution respectively. Remaining three images correspond to tumor, normal tissue and bad spectra/artefact distribution respectively. In this dataset the tumor tissue is modelled by two sources.

was a significant increase in the median of $SC1$, $SC2$ and DC from NCPD- l_1 compared to single stage NMF and hNMF, which is evident from the p-values ($p < 0.01$). However, the increase in the median was not significant compared

to NCPD without regularization ($p > 0.05$). In case of necrotic tissue, the performance of all the algorithms is good. NCPD- l_1 has significantly higher mean and median values compared to single stage NMF and hNMF for DC ($p < 0.01$). On the other hand, single stage NMF and hNMF has slightly better median values for $SC2$ but the increase is not significant ($p > 0.01$). For the necrotic sources NCPD- l_1 has slightly better median values for correlation than NCPD (no regularization), but the differences are not significant ($p > 0.01$).

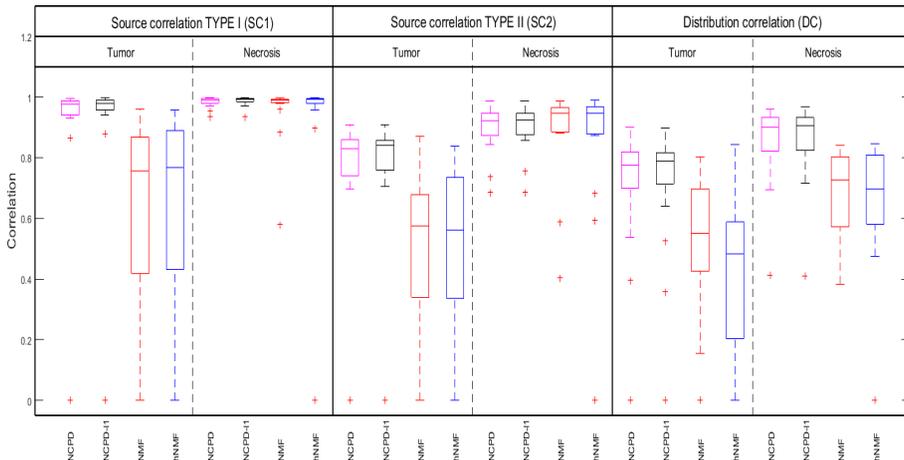


Figure 5.6: Box-plot of Type I source correlation, Type II source correlation and distribution correlation values obtained from NCPD without regularization, NCPD with l_1 regularization, single stage NMF and hNMF algorithms. Zero correlation values indicate that the specific tissue type was not recovered.

For all the MRSI datasets, the rank was automatically estimated using the method explained in section 5.2.5. Fig. 5.7 shows the estimated ranks for 28 MRSI datasets with and without using prior knowledge. From Fig. 5.7 we can observe that the estimated rank was higher in many MRSI datasets and after applying the maximum possible tissue types as prior knowledge in the second stage the estimated ranks are reduced. The tissue types are assigned to the sources manually by visualizing the estimated distribution maps, expert labelling and the estimated source spectra.

The result of an in-vivo example is shown in Fig. 5.8. NCPD- l_1 , single stage NMF and hNMF methods are applied to a 16×16 voxel grid shown in the second row of Fig. 5.8a. The spectra which are truncated to the region 0.25-4.2 ppm are of length 517. The reduced spectra are constructed using a window length $L = 20$ and the window is moved with a step-size of 5 samples. Therefore the size of the MRSI tensor \mathcal{T} with 16×16 ($K = 256$) voxels is $100 \times 100 \times 256$. For this

Table 5.1: Mean, standard deviation, median, MAD and range of source correlation (Type I and Type II) and distribution correlation for 28 glioma datasets. The results are shown on tumor and necrotic tissue type for NCPD without regularization, NCPD with l_1 regularization, single stage NMF and hNMF algorithms. P-values indicate the statistical significance of higher median for the NCPD- l_1 algorithm (* indicates statistical significance for $p < 0.01$)

	Spectral correlation Type I				Spectral correlation Type II				Distribution correlation				
	NCPD	NCPD- l_1	NMF	hNMF	NCPD	NCPD- l_1	NMF	hNMF	NCPD	NCPD- l_1	NMF	hNMF	
Tumor	Mean	0.891	0.898	0.619	0.617	0.752	0.759	0.496	0.496	0.699	0.702	0.477	0.403
	std dev	0.264	0.265	0.339	0.346	0.229	0.231	0.281	0.287	0.234	0.236	0.273	0.269
	median	0.975	0.978	0.757	0.767	0.829	0.841	0.574	0.560	0.775	0.789	0.552	0.482
	MAD	0.015	0.015	0.134	0.146	0.041	0.041	0.111	0.179	0.057	0.046	0.135	0.178
	Range	0.994-0.866	0.997-0.878	0.960-0.406	0.956-0.417	0.908-0.696	0.908-0.704	0.871-0.286	0.837-0.321	0.900-0.396	0.896-0.357	0.803-0.155	0.844-0.024
p-value	0.1753	-	<0.0001*	<0.0001*	0.3269	-	<0.0001*	<0.0001*	0.4745	-	<0.0001*	<0.0001*	
Necrosis	Mean	0.984	0.988	0.960	0.929	0.896	0.899	0.891	0.855	0.852	0.855	0.687	0.661
	std	0.017	0.015	0.098	0.233	0.079	0.076	0.151	0.238	0.132	0.132	0.134	0.203
	median	0.991	0.993	0.991	0.992	0.922	0.926	0.946	0.945	0.900	0.905	0.726	0.698
	MAD	0.005	0.004	0.003	0.004	0.037	0.037	0.025	0.032	0.044	0.045	0.089	0.114
	Range	0.998-0.935	0.998-0.936	0.996-0.581	0.998-0.897	0.987-0.686	0.987-0.686	0.987-0.405	0.989-0.593	0.961-0.411	0.967-0.409	0.840-0.384	0.847-0.473
p-value	0.2238	-	0.075	0.3879	0.3639	-	0.0844	0.1965	0.4622	-	0.0001*	0.0001*	

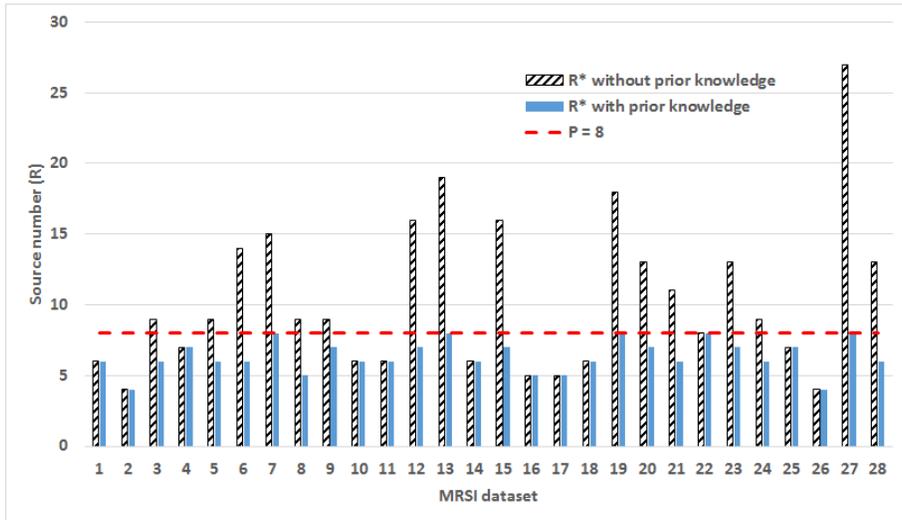


Figure 5.7: Estimated number of sources from the covariance matrix for 28 MRSI datasets with and without using the maximum number of tissue types as prior knowledge. The horizontal line indicates the maximum number of tissue types used as prior knowledge in the analysis ($P = 8$).

dataset the number of sources was estimated as $R = 7$, the same rank was used for the single stage NMF algorithm. The seven sources and their corresponding distribution maps obtained from NCPD- l_1 and NMF methods are shown in Fig. 5.8b, 5.8c, 5.8d and 5.8e. Using the hNMF method, only three sources are obtained as shown in Fig. 5.8f and 5.8g. Fig. 5.8g shows that the hNMF method identifies the normal and necrotic ($SC1 = 0.9971$) tissue properly but fails to recover the tumor tissue. Single-stage NMF identifies normal and necrotic tissue, but only the necrotic source is good ($SC1 = 0.9941$) and the normal source deviates a lot from the expert as shown in 5.8d (first row). In the single stage NMF method the recovered tumor tissue ($SC1 = 0.6041$) and its corresponding distribution ($DC = 0.4262$) are bad and it is difficult to identify it as tumor tissue from the source spectrum. The NCPD method identifies all three tumor ($SC1 = 0.9875$), necrotic ($SC1 = 0.9854$) and normal tissue types. Fig. 5.8b and 5.8c show that the estimated tissue sources and their corresponding spatial distribution are accurate when compared to expert labeling. In this example we have estimated seven sources and their corresponding distributions from rank-7 NCPD- l_1 . Three sources correspond to tumor, necrosis and normal tissue type, the other four sources correspond to artifacts (Fig. 5.8b and 5.8c: 7th row) and spectra from the outer edges of the voxel grid (Fig. 5.8b and 5.8c: 4th, 5th and 6th row) which are contaminated by the chemical shift displacement artifact.

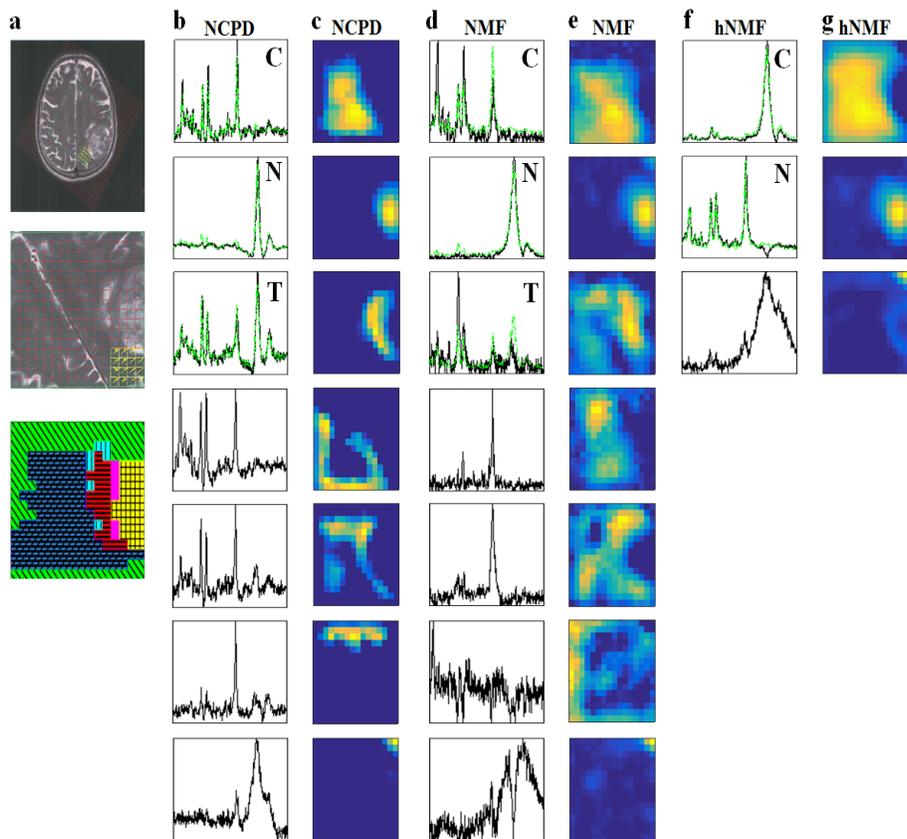


Figure 5.8: Tissue pattern differentiation using ^1H MRSI: C, T and N represent normal, tumor and necrosis, respectively. (a) First row: T2-weighted anatomical MR image of a brain tumor with areas of necrosis. Second row: voxels within the MRSI excitation volume superimposed on anatomical image. Third row: expert labeling, where yellow (horizontal + vertical line) indicates N, red (horizontal line) indicates T, magenta (no pattern) indicates T/N, dark blue (slanted + horizontal line) indicates C, light blue (vertical line) indicates C/T and green (slanted line) indicates spectra of poor quality. (b, c) results of NCPD. (b) The recovered sources from the NCPD- l_1 method are shown in black (solid line). First three rows represent C, N and T spectral sources in black (solid line), with tissue-specific spectra based on expert labeling overlaid in green (dash-dot line). The remaining four rows represent artifacts and spectra from outer edges. (c) Distribution maps corresponding to spectral profiles in (b). (d, e) Results of single-stage NMF. (d) The recovered sources are shown in black, overlaid with the expert-based tissue-specific spectra in red. First three rows show normal, necrotic and tumor spectra and the remaining rows show other spectra obtained using rank-7 NMF. (e) Distribution maps corresponding to (d). (f, g) Results of hNMF. (f) Recovered sources shown in black and expert-based tissue-specific spectra in red. First two rows show control and necrotic spectra. (g) Distribution maps corresponding to (f).

5.4 Discussion

The 2D-MRSI data can be directly represented as a third-order tensor. A $(L_r, L_r, 1)$ BTD based approach can be used to extract the different tissue types from this spatial tensor. The problem with this approach is that it is difficult to find the rank L_r of the factor matrices. The rank L_r is patient specific and it is different for different tissue types. Even when the ranks L_r are approximately known, this method does not perform better than single stage NMF or hNMF algorithms. This has motivated us to find a new way to represent 2-D MRSI data in a tensor.

In this chapter, we have proposed a method to represent the 2D-MRSI data in a tensor using a reduced format of the spectra. A novel tissue type differentiation algorithm based on non-negative canonical polyadic decomposition with l_1 regularization was developed. This study explored the feasibility and efficiency of the proposed algorithm (NCPD- l_1) in recovering the normal, tumor and necrotic tissue patterns for patients with glioma using short-TE MRSI data. The previous matrix-based algorithms, NMF and hNMF failed consistently in extracting tissue-specific spectral patterns. NMF failed because the tumor spectral profile is not sufficiently uncorrelated from a linear combination of other tissue patterns: normal and necrosis [111]. Also, sometimes the NMF algorithm extracts individual peaks as sources and these sources do not represent the tissue-specific spectral patterns. The problem with the hNMF [111] is that the algorithm is designed for a maximum of three sources and cannot handle artifacts. Therefore, in [111] the voxels at the outer edge of the PRESS excitation volume are removed to minimize the effect of artifacts. By doing this we can lose the voxels belonging to clinically relevant tissue types. In the MRSI grid shown in Fig. 5.8, if we remove 2 or 3 outer rows or columns of voxels the necrotic tissue is almost lost. Also, due to heterogeneity of the tissue some datasets require more than one source to model that tissue type. In this case, hNMF fails to model all the tissue types with only 3 sources.

The hNMF algorithm is mainly designed to handle GBMs. When the hNMF algorithm is used on low grade gliomas, the second stage of hNMF is not applied and hNMF reduces to single stage NMF with two sources. NMF with more than two sources performs better compared to hNMF in MRSI data, which does not contain necrotic tissue type (low grade gliomas). In high grade gliomas containing necrotic tissue type, hNMF performs better than single stage NMF because the second stage in the hNMF separates tumor and necrotic tissue type. In NCPD, MRSI data with more tissue types as in high grade glioma with necrotic tissue can be modelled using a higher number of sources and low grade gliomas with less artifacts can be modelled using a lower number of sources. The proposed NCPD algorithm can better separate tumor and necrotic tissue

type than hNMF in high grade gliomas and better separate tumor tissue from other ones in low grade gliomas than NMF.

We have proposed an initialization scheme for the NCPD algorithm based on SVD of the complex-valued spectra and spectral reduction of singular vectors. Random initialization is used in most of the non-negative tensor factorization applications. In non-negative RESCAL tensor factorization [102], the factor matrices are initialized using an NMF initialization, method Non-negative Double Singular Value Decomposition method (NNDSVD) [24]. When NCPD is initialised using NNDSVD on the reduced spectra, the initial factor matrix S_{init} contains many zero values, which are also retained in the decomposed factor matrix S . Therefore, the estimated reduced source spectra are unrealistic and deviate from the actual tissue-specific reduced spectra. Other initialization methods based on clustering will have more realistic source spectra compared to SVD-based methods. The problem with these methods is that they require some initialization and are computationally intensive [190, 34]. Although our initialization method is based on SVD, it does not suffer from too many zero values in the source initialization because the SVD is applied on the complex-valued spectra and the singular vectors are made positive by constructing the reduced spectra from them. The initialized sources are more realistic and close to the reduced spectra found in the MRSI voxels.

The number of sources for the NCPD algorithm is estimated using the eigenvalues of reduced spectra covariance matrix. The number of sources is overestimated when no prior knowledge is used as shown by striped bars in Fig. 5.7. Use of prior knowledge about the maximum number of tissue types prevents this overestimation and results in better number of sources estimation as shown by solid bars in Fig. 5.7. In this method we have used a cut-off of 99% on the cumulative sum of eigenvalues of C . Whereas, information theoretic criteria based methods such as in [188, 6] can determine the number of sources adaptively without the need for a cut-off value. However, these methods do not perform well in the presence of artifacts or when a linear model is not strictly satisfied. These methods highly overestimate the number of sources in our MRSI datasets.

The advantage of this tensor method is that the construction of the MRSI tensor couples the peaks in the spectra because of the XX^T in the frontal slices. Therefore, in the spectral sources obtained from the NCPD algorithm the peaks will be coupled, i.e. we will not get individual peaks as sources. The difference in the results between NCPD without regularization and NCPD- l_1 is negligible because the construction of the tensor and the extra sources already introduce sparsity in H . But NCPD- l_1 algorithm gives more stable results and sometimes models the tissue types with less sources. Also, the computational time is much less in NCPD- l_1 compared to NCPD (without regularization) as it converges in fewer iterations.

5.5 Conclusion

The NCPD- l_1 algorithm outperforms the existing tissue type differentiation methods based on NMF and hNMF. The worse performance of the hNMF is due to the fact that the voxels in the outer edge of MRSI excitation volume are included in the assessment. By contrast, NCPD can account for artifacts and bad voxels present in the outer edges because more sources ($R \geq 3$) are used in the decomposition. NCPD is also able to separate artifacts from tissue sources, but NMF fails to separate these properly even after using more sources in the decomposition. The NCPD algorithm has the potential to replace the hNMF method in unsupervised nosologic imaging for brain tumors [112], which can be used as a tool to assist brain tumor diagnosis. Recently, instead of using only the MRSI signal, a multiparametric (MRSI, cMRI, DWI, PWI) approach based on a modified hierarchical non-negative matrix factorization (hNMF) has been used to characterize brain tumor heterogeneity [152].

Chapter 6

Supervised tumor voxel classification in MRSI

The non-negative canonical polyadic decomposition (NCPD) algorithm developed in the previous chapter extracts the tissue-specific profiles and their corresponding distribution from the MRSI data. However, it cannot assign each voxel to a particular tissue. In this chapter we extend the NCPD algorithm and develop a CNN classifier to classify the voxels in the MRSI grid into three classes: tumor, normal and bad quality. We also analyse the effect of low-rank regularization on convolution layers in the CNN classifier. NCPD, CNN with and without low-rank regularization are tested on MRSI datasets of UZ Leuven and compared with a random forest classifier.

6.1 Introduction

Magnetic resonance spectroscopic imaging has been used in the diagnosis and prognosis of brain tumors. There are many algorithms for MRSI data analysis available in the literature that aim at tissue characterisation, tumor localization and classification. Machine learning algorithms are extensively used in the medical field, including cancer prognosis and prediction [100]. In the previous chapter we have developed a un-supervised non-negative canonical polyadic decomposition (NCPD) algorithm for tumor tissue characterisation and localization, which can be used to generate nosologic images for brain tumors. In clinical settings, automatic labelling of voxels in the MRSI grid as belonging to tumor or normal tissue is useful. Generally, un-supervised methods

are difficult to automate. The NCPD algorithm generates tissue specific sources and their corresponding abundances. However, it does not specify which source belongs to which tissue and it requires the user to manually specify the class of each source. Once we have the class for each source, voxels can be labelled using the corresponding abundances. It is possible to automatically estimate the class of each source using thresholds based on prior-knowledge. However, these thresholds are very sensitive and often result in poor decision. Supervised algorithms are widely used in classification task. With sufficient training data, supervised classification algorithms often outperform un-supervised methods. Also, supervised methods are easy to automate and require minimal to even no user interaction. This motivated us to explore supervised algorithms for classifying voxels in the MRSI signals as tumor or normal.

In MRS(I), supervised algorithms have been used for various classification tasks. Linear discriminant analysis, logistic regression, functional trees, support vector machines and decision stump logit boost classification algorithms are used in single voxel high resolution magic angle spinning spectroscopy to distinguish between recurrent low grade gliomas that transformed to a higher grade and those that remained grade II [40]. In [68] support vector machine (SVM) and logistic regression have been used to classify voxels as tumor or normal using ^{31}P -MRSI. These methods extract features by quantifying metabolites from MRS(I) signals. Luts et al. [118] studied the effect of feature extraction in classifying brain tumors into different grades using ^1H -MRS and showed that using full spectra and pattern recognition based feature extraction methods perform better than metabolite quantification.

In recent years deep learning algorithms that do not require any feature engineering have shown a lot of potential in medical image analysis [113]. Most state-of-art techniques for automatic tumor segmentation from multi-parametric MRI employ a particular class of deep learning, namely convolutional neural network (CNN) [92]. CNNs are good in identifying local input patterns by enforcing a local connectivity. Since local features in the spectra are useful in discriminating different tissue-specific patterns, CNNs will be more suited for classifying voxels in the MRSI dataset. Therefore, in this chapter we propose a CNN architecture to classify the voxels in the MRSI grid into three classes: tumor, normal and bad. We also apply multi-linear singular value decomposition (MLSVD) based low-rank regularization to convolution layers in CNN. We analyse the performance of the CNN classifiers with and without low-rank regularization and compare their performance to that of a random forest classifier.

6.2 Method

6.2.1 Feature extraction

Similarly to tissue type differentiation in the previous chapter, for each voxel in the MRSI grid a reduced real-valued spectrum X is constructed from the corresponding complex-valued pre-processed spectrum. Elements of the vector X are obtained by moving an overlapping window over the spectrum, where the i^{th} element of X is the sum of squares of the absolute values of all the elements in the i^{th} window segment as shown in Figure 5.2a,

$$X(i) = \sum_{j=1}^L s_i(j) s_i^*(j)$$

s_i is the spectrum at the i^{th} segment, s_i^* is its complex conjugate and L is the length of the window segment. This reduced spectrum is used as input feature for non-negative CPD and random forest classifier. For a convolution neural network classifier, the reduced spectrum is used to generate the input image.

6.2.2 Non-negative CPD for tumor classification

As described in the previous chapter, the 2D-MRSI data from the patient is represented in a 3-D tensor \mathcal{T} using the reduced spectra (Figure 5.2). Non-negative canonical polyadic decomposition (NCPD) with common factor in mode-1 and mode-2 and l_1 regularization on mode-3 is applied to the MRSI tensor \mathcal{T}

$$[\mathbf{S}^*, \mathbf{H}^*] = \arg \min_{\mathbf{S} \geq 0, \mathbf{H} \geq 0} \left\| \mathcal{T} - \sum_{i=1}^R \mathbf{S}(:, i) \circ \mathbf{S}(:, i) \circ \mathbf{H}(:, i) \right\|_2^2 + \lambda \|\text{Vec}(\mathbf{H})\|_1$$

where \mathbf{S} represents the tissue-specific patterns of the reduced spectra, \mathbf{H} represents the spatial distribution of each tissue type. R is the rank, defined as the number of rank-one terms and λ is a parameter which controls sparsity. Ideally, to classify a voxel into three classes (tumor, normal and bad) a rank of $R = 3$ should be used. Due to the heterogeneity of tumor and normal tissues it is difficult to model them using only one source for each type. Therefore each class is usually modelled by more than one source. The total number of sources (rank R) is estimated similarly to the NCPD tissue type differentiation method from the previous chapter. Once the source (\mathbf{S}) and abundance matrix (\mathbf{H}) are estimated from the NCPD, each source is then manually assigned to a particular class. The abundance values in \mathbf{H} corresponding to tumor, normal

and bad sources are summed respectively and the voxels are assigned to the corresponding class which has the largest value in the summed abundances. Instead of manual assignment of sources, it is also possible to assign the sources to a particular class automatically using the full length spectral sources instead of the reduced spectra \mathbf{S} . The full length spectral sources are estimated from the abundance matrix and the pre-processed MRSI spectra as described in Section 5.2.3 of the previous chapter. Tissue types are assigned to the source spectra using the maximum values of Cho, NAA, Cr and Lip peaks. Detailed description of this algorithm is given in the APPENDIX A.

6.2.3 Random forest

Random forest is one of the widely used supervised classification algorithms. They belong to the class of ensemble methods for classification, where a collection of decision trees are used [26, 27]. Each decision tree learns a rule and gives a classification output, called a vote for that class. The random forest then chooses the class voted by the majority of decision trees. We used the default MATLAB R2016a (Statistical Toolbox) implementation "TreeBagger" with 200 trees to classify each voxel into one of the three (tumor or normal or bad) classes. The reduced spectrum from each voxel is used as input feature for the random forest classifier. Since the number of voxels belonging to tumor is less than normal or bad, class imbalance is present in the training data. The class imbalance is handled by weighting the training data corresponding to the class it belongs to. The weights were chosen such that they are inversely proportional to the percentage of voxels belonging to that corresponding class in the dataset. A weight of $W_T = 1$, $W_N = 0.69$ and $W_B = 0.58$ was used for data belonging to tumor, normal and bad class, respectively.

6.2.4 Convolutional neural network

Löwner matrices are used in many applications such as system identification [9, 10] and blind source separation [54]. Consider a function $f(t) \in \mathbb{C}$, evaluated in the point set $T = \{t_1, t_2, \dots, t_N\}$. The point set T is partitioned into two distinct point sets, $X = \{x_1, x_2, \dots, x_I\}$ and $Y = \{y_1, y_2, \dots, y_J\}$ with $I + J = N$. The elements of the Löwner matrix $\mathbf{L} \in \mathbb{C}^{I \times J}$ are then defined as

$$\forall i, j : l_{ij} = \frac{f(x_i) - f(y_j)}{x_i - y_j}.$$

We thus obtain the following matrix:

$$\mathbf{L} = \begin{bmatrix} \frac{f(x_1)-f(y_1)}{x_1-y_1} & \frac{f(x_1)-f(y_2)}{x_1-y_2} & \dots & \frac{f(x_1)-f(y_J)}{x_1-y_J} \\ \frac{f(x_2)-f(y_1)}{x_2-y_1} & \frac{f(x_2)-f(y_2)}{x_2-y_2} & \dots & \frac{f(x_2)-f(y_J)}{x_2-y_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{f(x_I)-f(y_1)}{x_I-y_1} & \frac{f(x_I)-f(y_2)}{x_I-y_2} & \dots & \frac{f(x_I)-f(y_J)}{x_I-y_J} \end{bmatrix}.$$

Since CNN are more suited for analyzing images, we have used a Löwner matrix approach to convert the 1-D reduced spectra to a 2-D image. The construction of the Löwner matrix also couples different peaks in the reduced spectra, which can help CNNs to extract more relevant features for classification. The Löwner matrices from each voxel are used as input feature for the CNN model. Figure 6.1 shows the reduced spectrum and the corresponding Löwner matrix images from three voxels belonging to tumor, normal and bad classes.

A typical convolutional neural network architecture is made of several layers consisting of three basis blocks: 1) Convolution layer, 2) Rectified linear unit (ReLU) and 3) Max-pooling. A convolution layer is the most important layer in the CNN architecture: in this layer 2-D convolution is performed on the input using a kernel to generate a feature map. A single convolution layer can have multiple kernels, generating a set of feature maps. ReLU applies a non-saturating activation function, replaces negative values in the feature maps by zeroes and preserves the positive values. In the max-pooling layer, the dimensionality of an input feature map is reduced, by retaining only the maximum value over a defined neighborhood region of the feature map. The CNN architecture used for MRSI voxel classification consists of six convolution layers, six ReLU layers, three max pool layers, one fully connected layer and a final softmax layer for prediction, as shown in Figure 6.2. Batch normalization is used in between convolution layer and ReLU.

6.2.5 Low-rank regularization

Multilinear singular value decomposition (MLSVD) can be considered as a generalization of matrix singular value decomposition to higher order tensors. For a fourth-order tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$, MLSVD is defined as

$$\mathcal{T} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \times_4 \mathbf{U}^{(4)}.$$

where \mathcal{S} is an all-orthogonal, ordered, $(I_1 \times I_2 \times I_2 \times I_4)$ -tensor, $\mathbf{U}^{(1)} \in \mathbb{R}^{I_1 \times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{I_2 \times I_2}$, $\mathbf{U}^{(3)} \in \mathbb{R}^{I_3 \times I_3}$ and $\mathbf{U}^{(4)} \in \mathbb{R}^{I_4 \times I_4}$ are orthonormal matrices and \times_n is the n -mode product of a tensor by a matrix [48].

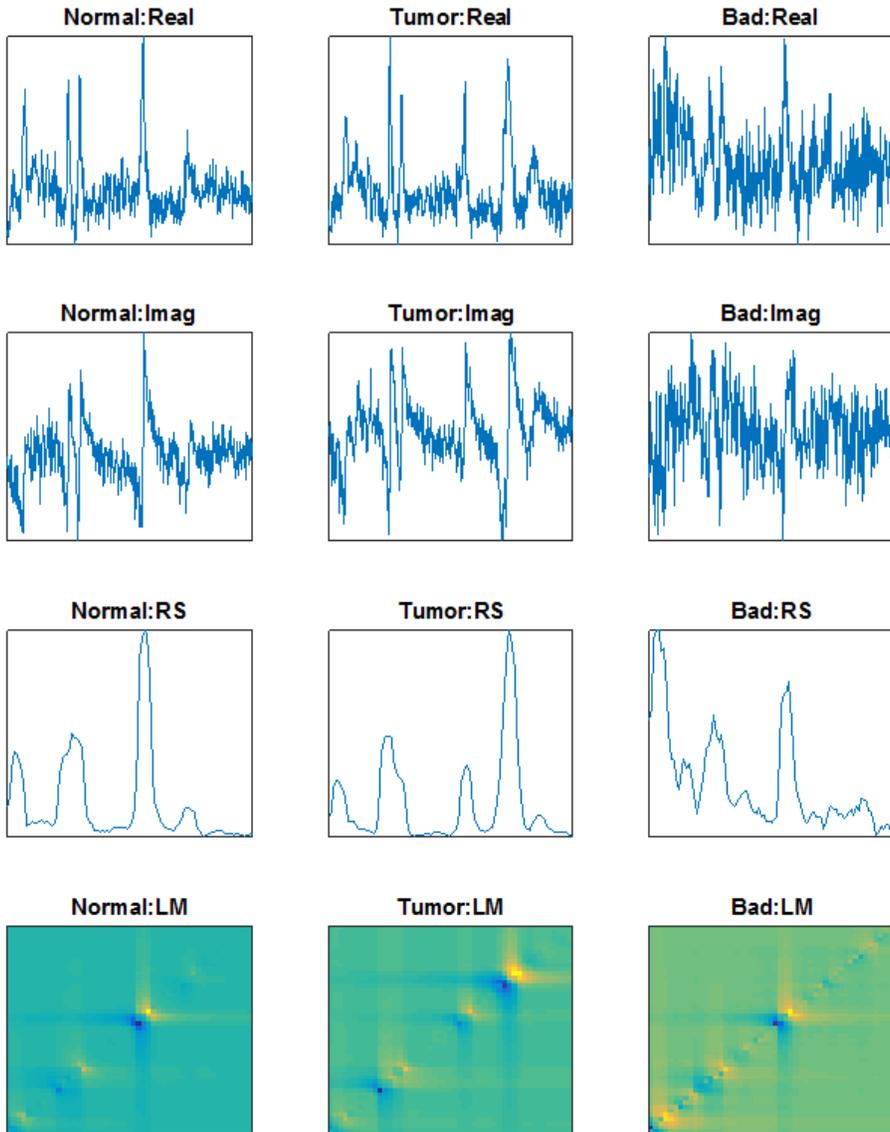


Figure 6.1: Original spectra, reduced spectra and Löwner matrices belonging to three different classes. First row: real part of the spectra, second row: imaginary part of the spectra, third row: reduced spectra and fourth row: Löwner matrices. First column: tumor, second column: normal and third column: bad.

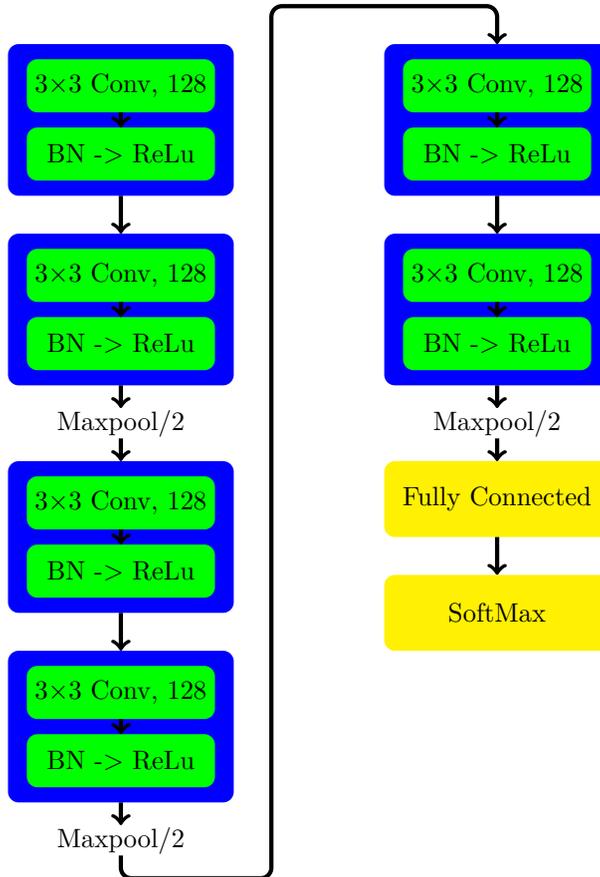


Figure 6.2: Convolutional neural network architecture used for voxel classification in MRSI. Each convolution layer shows the filter size (e.g. 3×3) and the number of output channels (e.g. 128). Maxpool/2 represents a maxpool layer with size 2×2 and stride 2, which reduces the dimension of the input image by half.

Filters of a convolution layer can be represented by a 4-D tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$, where I_1 and I_2 represents the filter size/receptive field width and height, respectively, I_3 represents the number of input channels (input depth) and I_4 represents the number of output filters (output depth). In the low-rank regularization model we replace the convolution tensor \mathcal{T} by a low rank tensor \mathcal{T}^* using truncated MLSVD.

$$\begin{aligned}\mathcal{T}^* &= \mathcal{S}^* \times_3 \mathbf{U}^{*(3)} \times_4 \mathbf{U}^{*(4)} \\ \mathcal{S}^* &= \mathcal{S}^{**} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)}\end{aligned}\tag{6.1}$$

where, $\mathcal{S}^{**} \in \mathbb{R}^{I_1 \times I_2 \times I_3^* \times I_4^*}$ is the truncated core tensor, $\mathbf{U}^{*(3)} \in \mathbb{R}^{I_3 \times I_3^*}$ and $\mathbf{U}^{*(4)} \in \mathbb{R}^{I_4 \times I_4^*}$ are truncated mode-3 and mode-4 factor matrices, respectively, with $I_3 > I_3^*$ and $I_4 > I_4^*$. The compression is not performed on the receptive field dimension (mode-1 and mode-2) since the filter size was small (3×3). Chen Yunpeng et al. [194] have formulated the relationship between the convolution layer and its tensor decomposed representation. For an input tensor $\mathcal{X} \in \mathbb{R}^{p \times q \times I_3}$ the convolution layer with kernel weights $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ will generate an output tensor of size $\mathcal{Y} \in \mathbb{R}^{p \times q \times I_4}$. The output of the modified convolution layer whose weights are represented by the truncated MLSVD (equation (6.2.5)) can be written as:

$$\begin{aligned}\mathcal{Y}(a, b, c) &= \sum_{l=1}^{I_4^*} \left[\sum_{k=1}^{I_3^*} \sum_{i=a-\delta_1}^{a+\delta_1} \sum_{j=b-\delta_2}^{b+\delta_2} \mathcal{S}^*(i-a+\delta_1, j-b+\delta_2, k, l) \right. \\ &\quad \left. \sum_{m=1}^{I_3} \mathcal{X}(i, j, m) \mathbf{U}^{*(3)}(m, l) \right] \mathbf{U}^{*(4)}(c, l)\end{aligned}\tag{6.2}$$

where, δ_1 and δ_2 represent the half-width of the filter size on each dimension. Equation (6.2.5) can be simplified by denoting the intermediate results with $\mathcal{T}^{(1)} \in \mathbb{R}^{p \times q \times I_3^*}$ and $\mathcal{T}^{(2)} \in \mathbb{R}^{p \times q \times I_4^*}$ as [194]:

$$\mathcal{T}^{(1)}(i, j, l) \triangleq \sum_{m=1}^{I_3} \mathcal{X}(i, j, m),\tag{6.3}$$

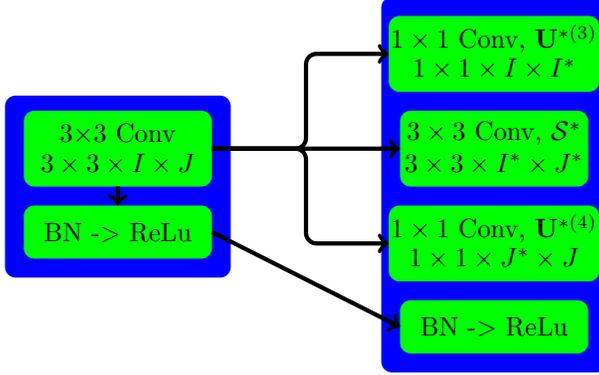


Figure 6.3: Low-rank regularized representation of convolution layer. The original convolution layer is shown in the left and the low-rank regularized representation is shown in the right. Each convolution layer shows the filter size (e.g. 3×3) and the total size of the kernel (e.g. $3 \times 3 \times I \times J$). Convolution layers in the low-rank regularized representation also show the factor matrix/core tensor corresponding to that layer

$$\mathcal{T}^{(2)}(a, b, l) \triangleq \sum_{k=1}^{I_3^*} \sum_{i=a-\delta_1}^{a+\delta_1} \sum_{j=b-\delta_2}^{b+\delta_2} \mathcal{T}^{(1)}(i, j, k) \quad (6.4)$$

$$\mathcal{S}^*(i - a + \delta_1, j - b + \delta_2, k, l),$$

$$\mathcal{Y}(a, b, c) = \sum_{l=1}^{I_4^*} \mathcal{T}^{(2)}(a, b, l) \mathbf{U}^{*(4)}(c, l). \quad (6.5)$$

The full convolution layers can be represented by a low-rank model using three convolution layers as shown in equations (6.2.5), (6.2.5) and (6.2.5). First the input tensor is convolved using a 1×1 kernel corresponding to the mode-3 (input) factor matrix, then a full convolution with truncated kernel in input and output dimension corresponding to the core tensor is performed, finally a 1×1 convolution corresponding to the mode-4 (output) factor matrix is performed. The low-rank convolution architecture is shown in Figure 6.3.

Initially, the CNN with full convolution architecture is trained for 10 epochs. Next, MLSVD is applied to all the convolution kernel weights except the first layer where the input is fed. The truncated input (I_3^*) and output (I_4^*)

dimensions of the low-rank architecture are estimated from the MLSVD singular values of the corresponding mode, respectively. In both modes it is calculated as the minimum number such that the cumulative sum of singular values is greater than the threshold $\tau = 0.8$ of the sum of all singular values. Finally, each convolution layer in the CNN architecture is replaced by a low-rank model as shown in Figure 6.3 with the estimated truncated dimensions (I_3^* and I_4^*).

Batch normalization was used after each convolution layer as shown in figure 6.2. Multiclass hinge loss (mhinge) was used as loss function for training the CNN. Similarly to random forest classifier, training data was weighted in order to handle the class imbalance. Both CNN's, with and without low-rank regularization, are implemented using Matlab based MatConvNet package [183].

6.3 Results and Discussion

The performance of the algorithms are evaluated using 28 MRSI signals from 17 glioma patients. Voxels are labelled by the expert as belonging to active tumor, necrosis, normal, bad or mixture of these (only a few cases). All the voxels belonging to active tumor, necrosis and their mixture are considered as tumor class. Some voxels are also labelled by the expert as a mixture of tumor and normal, in this case the preference is given to tumor and is considered as belonging to the tumor class. Numbers of MRSI scans and grid size vary over patients as shown in Table 6.1.

Leave-one-patient-out-cross-validation method was used for training and validating the supervised classifiers. Classifiers are trained from the labelled voxels of all the patients except one, which was used as a test patient. This is repeated until all the patients belong to the test group are used once. For each test patient all the MRSI voxels are classified using the trained model. To measure performance we compute the sensitivity (TPR, true positive rate), specificity (SPC, true negative rate), precision (PPV, positive predictive value) and F1 Score for tumor vs others (normal + bad) and normal vs others (tumor + bad). These performance measures are defined as:

Table 6.1: Number of scans along with the grid size of the MRSI signal for 17 different patients.

Patient number	number of scans	Grid size
1	1	16 × 16
2	2	16 × 16 and 8 × 16
3	2	16 × 16 and 16 × 16
4	1	8 × 16
5	1	8 × 16
6	2	16 × 16 and 16 × 8
7	2	16 × 16 and 12 × 16
8	2	16 × 16 and 16 × 16
9	1	8 × 16
10	1	16 × 16
11	3	16 × 16, 8 × 16 and 8 × 16
12	3	16 × 16, 12 × 14 and 12 × 14
13	2	16 × 16 and 16 × 16
14	2	16 × 16 and 16 × 16
15	1	8 × 8
16	1	8 × 16
17	1	12 × 16

$$TPR = \frac{TP}{TP + FN}$$

$$SPC = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP}$$

$$F1 = \frac{2TP}{2TP + FP + FN}$$

where, TP, TN, FP and FN are the number of true positive, true negative, false positive and false negative voxels in the test patient, respectively. Leave-one-patient-out-cross-validation method is used for random forest and CNN classifiers, whereas NCPD is applied on the MRSI dataset individually and the performance measures are calculated for each patient. Boxplots of all performance measures for tumor vs others (normal + bad) and normal vs others

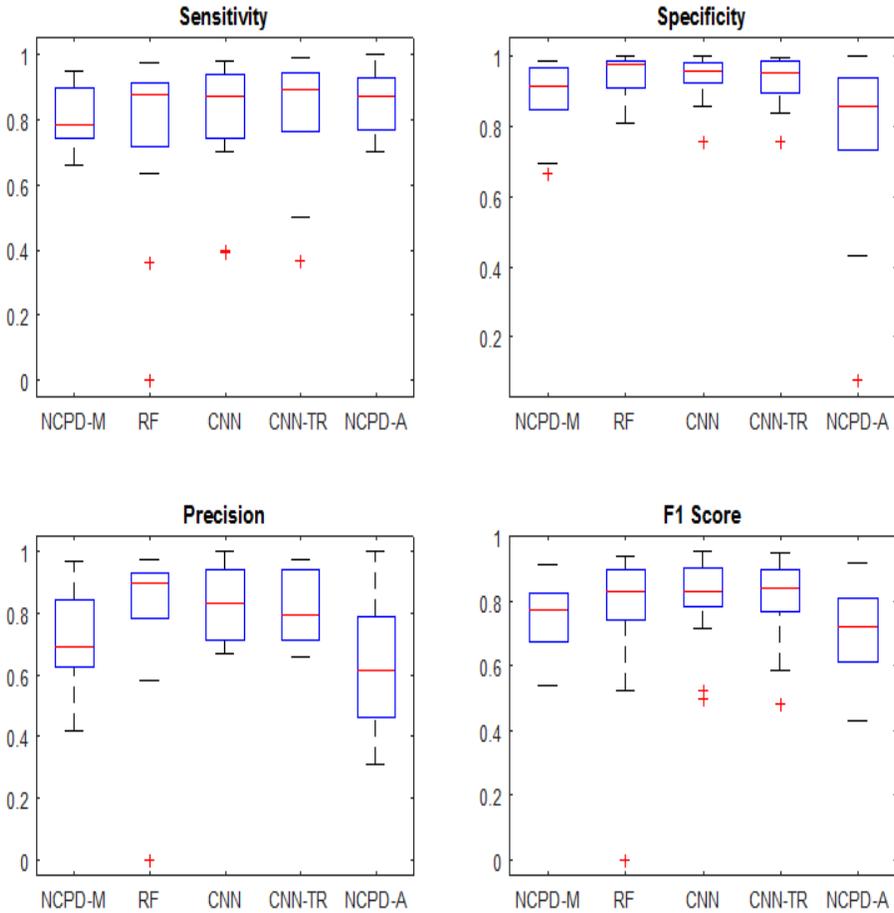


Figure 6.4: Boxplot of sensitivity, specificity, precision and F1 Score over 17 patients from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization (CNN-TR) and NCPD with automatic source assignment (NCPD-A). These values are calculated for Tumor vs others (normal + bad).

(tumor + bad) are shown in Figure 6.4 and Figure 6.5. Table 6.2 shows the average values of the performance measure from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization (CNN-TR) and NCPD with automatic source assignment (NCPD-M) algorithms.

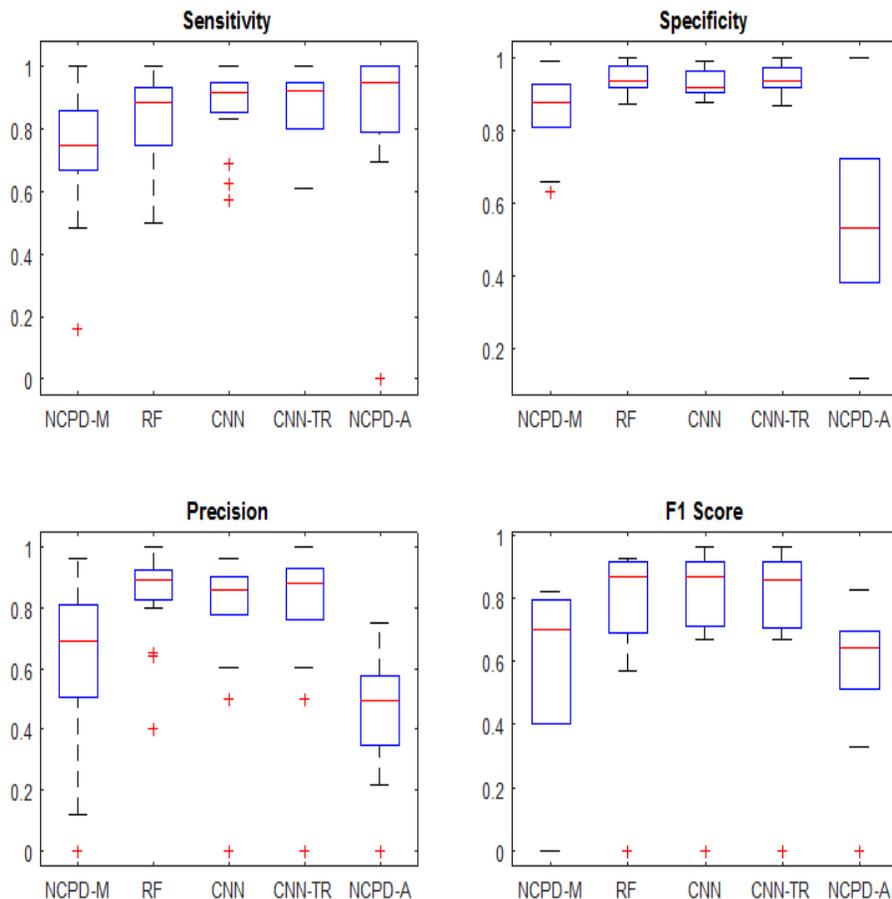


Figure 6.5: Boxplot of sensitivity, specificity, precision and F1 Score over 17 patients from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization (CNN-TR) and NCPD with automatic source assignment (NCPD-A). These values are calculated for Normal vs others (tumor + bad).

Table 6.2: Average sensitivity, specificity, precision and F1 Score over 17 patients from NCPD with manual source assignment (NCPD-M), Random forest (RF), convolutional neural network (CNN), convolutional neural network with low-rank regularization and NCPD with automatic source assignment (NCPD-A). These values are calculated for Tumor vs others (normal + bad) and Normal vs others (tumor + bad).

		Sensitivity		Specificity		precision		F1 Score	
		Mean	std dev	Mean	std dev	Mean	std dev	Mean	std dev
Tumor	NCPD-M	0.8101	0.0907	0.8796	0.1038	0.7036	0.1627	0.7437	0.1118
	RF	0.7689	0.2499	0.9446	0.0613	0.8076	0.2310	0.7739	0.2237
	CNN	0.8158	0.1798	0.9422	0.0604	0.8270	0.1156	0.8089	0.1288
	CNN-TR	0.8290	0.1694	0.9318	0.0647	0.8106	0.1180	0.8087	0.1237
	NCPD-A	0.8507	0.0906	0.7912	0.2357	0.6361	0.2178	0.7035	0.1391
Normal	NCPD-M	0.7169	0.2284	0.8606	0.1032	0.6302	0.2829	0.5851	0.2502
	RF	0.8415	0.1387	0.9427	0.0351	0.8408	0.1487	0.7702	0.2311
	CNN	0.8739	0.1273	0.9296	0.0335	0.7776	0.2343	0.7862	0.2254
	CNN-TR	0.8585	0.1289	0.9391	0.0364	0.7886	0.2413	0.7835	0.2245
	NCPD-M	0.8934	0.2383	0.5519	0.2532	0.4493	0.2294	0.5664	0.2465

Both NCPD-M and NCPD-A methods have a high value of sensitivity for tumor vs others because the tumor voxels are overestimated and a significant number of normal and bad voxels are classified as tumor. This is evident from the low values of specificity and precision. Also NCPD is not good in identifying bad voxels because bad spectra will not have any specific signature and it is difficult to represent these spectra as a linear combination of sources. NCPD with manual source assignment performs better than automatic source assignment. NCPD-automatic performance deteriorates because it fails to identify sources corresponding to bad class properly. The main advantage of NCPD is that it can be applied to any MRSI dataset without the need for training.

The random forest classifier has the lowest average sensitivity for tumor vs others compared to other methods. In one of the patients, it completely fails to identify tumor voxels resulting in a lower sensitivity. In other patients more tumor voxels are classified as bad or normal as compared to the CNN method. Random forest performs better than the un-supervised NCPD method in identifying tumor, which is reflected in a higher F1 score. However, its performance is still worse compared to the CNN method, which is evident from the lower value of sensitivity and F1 score. Random forest is good in identifying normal voxels rather than tumor voxels and has a similar performance to the CNN classifier for normal vs others (tumor + bad).

Both CNN models with and without low-rank regularization have the best

performance in identifying tumor voxels compared to NCPD and random forest. In identifying normal voxels CNN had lower specificity and precision than random forest. However, the overall performance is still similar to random forest, which can be seen from the higher sensitivity and F1 score. There is not much difference in performance between the two CNN methods with and without the low-rank regularization. The low-rank regularized CNN has less parameters compared to the full CNN architecture. In our model we have applied the low-rank model to 5 convolution layers with a kernel size of $3 \times 3 \times 128 \times 128$. Each of the original convolution layers will have $3 * 3 * 128 * 128 = 147456$ parameters and a total of $147456 * 5 = 737280$ parameters from 5 layers. After applying the low-rank regularization with a threshold $\tau = 0.8$, the truncated input and output dimension has an average value of $I_3^* = 95$ and $I_4^* = 95$, respectively. Each of the low-rank convolution layers will have $(1 * 1 * 128 * 95 + 3 * 3 * 95 * 95 + 1 * 1 * 95 * 128) = 105417$ parameters and a total of $105417 * 5 = 527085$ parameters from 5 such low-rank layers. A smaller number of learning parameters (convolution kernel weights) usually helps in preventing over-fitting of training data. However for the current problem we did not find any notable difference and there was a slight increase in the sensitivity for tumor vs others at the cost of specificity with the low-rank regularized model. The Löwner matrices generated from the MRSI signal are of size 50×49 . For an input image of this size, the original convolution layer will perform $3 * 3 * 128 * 128 * 50 * 49 = 361267200$ multiplications, where as a low-rank convolution model performs $50 * 49 * 128 * 95 + 50 * 49 * 3 * 3 * 95 * 95 + 50 * 59 * 94 * 128 = 264287650$ multiplications to generate the output. Since low-rank regularised CNN requires fewer multiplications than the un-regularised original CNN, it is faster in predicting the class for a given input voxel.

6.4 Conclusion

A convolutional neural network method with low-rank regularization is proposed for classifying voxels in the MRSI grid into tumor, normal and bad class. We have extended the NCPD tissue differentiation algorithm to perform voxel classification in MRSI signals. Both the CNN and NCPD algorithms are tested on in-vivo MRSI signals from brain tumor patients and are compared with a random forest classifier. CNN has the best performance when identifying the tumor voxels and for the normal voxel identification both the CNN and random forest perform similarly. In future research we would like to apply these algorithms to a much larger dataset and further classify the tumor voxels into active tumor and necrosis.

Chapter 7

Un-supervised tumor Segmentation in Multi-parametric MRI

In diagnosis and treatment planning of brain tumors, characterisation and localization of tissue plays an important role. In this chapter the non-negative CPD for tissue differentiation is adapted to deal multi-parametric MRI (MP-MRI) data, in particular: conventional MRI (T2, T1+C and FLAIR), PWI, DKI and MRSI modalities. A 3-dimensional tensor is constructed from in-vivo multi-parametric MRI of high grade glioma patients. Constrained canonical polyadic decomposition (CPD) with common factor in mode-1 and mode-2 and l_1 regularization on mode-3 is applied on the 3-dimensional multi-parametric tensor to characterize various tissue types. The work presented in this chapter is based on [17].

7.1 Introduction

Accurate characterisation and localization of tissue types play a key role in brain tumor diagnosis and treatment planning. Neuro-imaging methods provide anatomical and pathophysiological information about brain tumors and aid in diagnosis, treatment planning and follow-up of patients. Currently, conventional magnetic resonance imaging (cMRI) is mainly used for detection and analysis of brain tumor upon suspicion. In recent years, many advanced magnetic resonance

(MR) modalities, such as perfusion-weighted imaging (PWI), diffusion-weighted imaging (DWI) and MR spectroscopic imaging (MRSI) are being used for brain tumor diagnosis [111, 96, 58]. The tumor region of glioblastoma multiforme (GBM) could consist of several tissue types, which represent actively growing tumor, necrosis or normal brain tissue [111]. Many studies have investigated the potential of individual MRI modalities to characterize tumor grade, detect full tumor extent and assess early success of therapy. No specific modality is clearly capable of answering clinical questions unambiguously. Various MR modalities provide complementary information on the structural, hemodynamic and/or biochemical characteristics of the tissue. Instead of using the MR modalities independently, it has been shown that combining several MRI modalities (multi-parametric MRI) improves diagnostic accuracy when it comes to grading gliomas [176, 152].

Hierarchical non-negative matrix factorization (hNMF) has been applied to brain MRSI data to differentiate various tissue patterns in GBM patients [111]. In Chapter 5, we have developed a non-negative canonical polyadic decomposition (NCPD) based algorithm to differentiate various tissue patterns from MRSI data and showed that it works better than the NMF based algorithm developed for MRSI [111]. Recently, the same hierarchical non-negative matrix factorization (hNMF) algorithm developed for MRSI [111] has been modified to deal with MP-MRI data and to provide a more refined tissue characterization and segmentation [152]. These findings motivated us to explore tensor based approaches similar to NCPD on MP-MRI datasets.

In this chapter, a tensor based algorithm is proposed for tissue type differentiation in high-grade glioma patients from MP-MRI. A modified version of the tensor approach in [20] has been applied to MP-MRI imaging. The proposed method consists of representing the MP-MRI data in a 3-dimensional tensor and applying constrained canonical polyadic decomposition (CPD) with l_1 regularization to the MP-MRI tensor. The performance of the CPD algorithm is evaluated using expert segmentation and compared with hierarchical non-negative matrix factorization (hNMF) [152].

7.2 Method

7.2.1 Tensor construction

For each voxel in the ROI, a vector \mathbf{x} is constructed consisting of MP-MRI features as shown in Fig. 7.1a. The vector \mathbf{x} consists of:

1. eight metabolite concentrations mI, Gly, Cho, Cre, Glx, NAA, Lac and Lip from the MRSI modality.
2. T2, T1+C and FLAIR from the cMRI modality.
3. rCBV values from the PWI modality.
4. MD, MK and FA values from the DKI modality.
5. Smoothed version of T2, T1+C, FLAIR, rCBV, MD, MK and FA features using a moving average window with kernel size 3×3 .
6. Smoothed version T2, T1+C, FLAIR, rCBV, MD, MK and FA features using a moving average window with kernel size 5×5 .

A 3-way MP-MRI tensor \mathcal{T} is constructed by stacking $\mathbf{x}\mathbf{x}^T$ from all the voxels in the MP-MRI grid as shown in Fig. 7.1b.

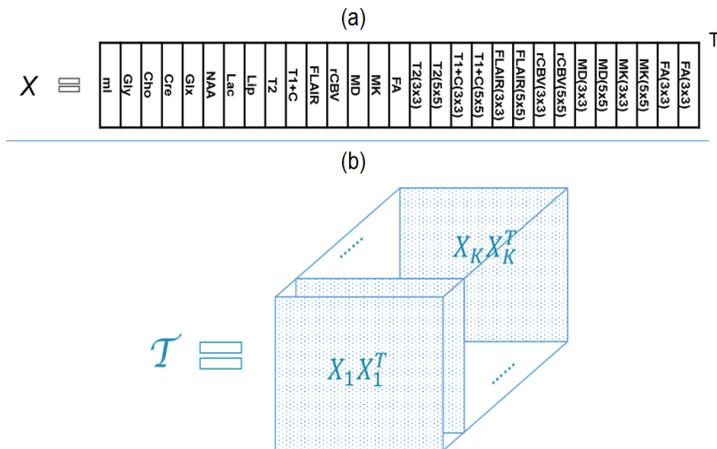


Figure 7.1: (a) Construction of feature vector \mathbf{x} from MP-MRI. (b) Construction of the MP-MRI tensor \mathcal{T} from the feature vector \mathbf{x} . K is the total number of voxels in the region of interest (ROI).

7.2.2 Constrained canonical polyadic decomposition

Canonical polyadic decomposition (CPD) is a tensor decomposition method, where the tensor is decomposed into a sum of rank-one tensors [38]. CPD of a

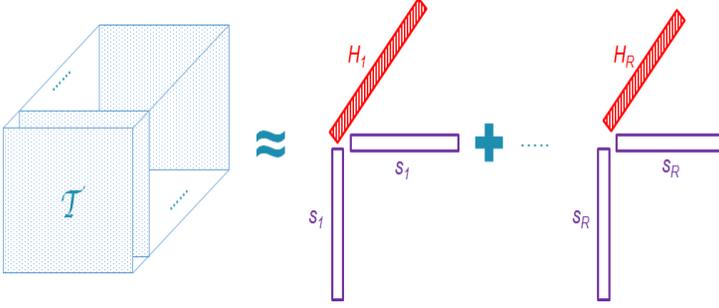


Figure 7.2: Partial non-negative CPD of MP-MRI tensor \mathcal{T} : MP-MRI tensor \mathbf{T} is decomposed into R rank-1 tensors. Common factor \mathbf{S} is used in mode-1 and mode-2 to maintain symmetry of frontal slices. Each \mathbf{s}_i gives a tissue-specific feature vector and the corresponding \mathbf{h}_i gives the spatial distribution of the respective tissue type, upon reshaping. A non-negativity constraint is imposed only on \mathbf{H} in the decomposition.

third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ can be written as

$$\mathcal{X} \approx \llbracket \mathbf{ABC} \rrbracket \equiv \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r,$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R] \in \mathbb{R}^{I \times R}$, $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_R] \in \mathbb{R}^{J \times R}$ and $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_R] \in \mathbb{R}^{K \times R}$ are factor matrices. R is the rank, defined as the number of rank-one terms. 'o' represents outer product.

In the MP-MRI tensor \mathcal{T} , the frontal slices are symmetrical, therefore we constrain the frontal slices of each CPD rank-one term to be symmetric. To maintain symmetry, a common factor matrix \mathbf{S} is used for mode-1 and mode-2 in the CPD as shown in Fig. 7.2. Tissue abundances are kept positive by imposing a non-negative constraint on the mode-3 factor matrix \mathbf{H} in the CPD. After performing the CPD on the MRSI tensor \mathcal{T} we obtain two factor matrices \mathbf{S} and \mathbf{H} , where \mathbf{S} represents the tissue-specific feature vector and \mathbf{H} represents the spatial distribution of each tissue type.

Each rank-one term obtained from the constrained CPD of the MP-MRI tensor \mathcal{T} is expected to correspond to a particular tissue type. Therefore, the mode-3 factor matrix, \mathbf{H} will be sparse, meaning that each row will mostly have only one high value. l_1 regularization is applied on the mode-3 factor matrix \mathbf{H} in the partial non-negative CPD to exploit the sparsity assumption of the tissue type distribution.

$$[\mathbf{S}^*, \mathbf{H}^*] = \arg \min_{\mathbf{S}, \mathbf{H} \geq 0} \left\| \mathcal{T} - \sum_{i=1}^R \mathbf{S}(:, i) \circ \mathbf{S}(:, i) \circ \mathbf{H}(:, i) \right\|_2^2 + \lambda \|\text{Vec}(\mathbf{H})\|_1,$$

where \mathbf{S} represents a matrix with tissue-specific feature vectors as columns, \mathbf{H} is a matrix containing their corresponding distribution vectors and λ is the parameter which controls the sparsity. The tensor decomposition was performed using Tensorlab Matlab package [185]. Common mode-1 and mode-2 factor, non-negativity of \mathbf{H} (mode-3) and l_1 regularization are applied using structured data fusion method [167] in Tensorlab.

7.2.3 Validation

Dice score: For each patient, the tissue segmentation were obtained from their corresponding abundance values, \mathbf{H} . The tissue segmentation was performed by assigning each voxel to a particular source using k-means clustering on the abundance values, \mathbf{H} . Number of classes is set equal to R , the number of sources obtained from CPD. The segmentation obtained from CPD was compared to the manual segmentation by a radiologist with 6 years of experience in brain tumor research using the dice score:

$$Dice_{tissue} = 2 \times \frac{A_{tissue,CPD} \cap A_{tissue,expert}}{A_{tissue,CPD} + A_{tissue,expert}}$$

where, $Dice_{tissue}$ is the dice score, $A_{tissue,CPD}$ is the area of tissue segmentation obtained from the CPD algorithm and $A_{tissue,expert}$ is the tissue area labelled by a radiologist.

Source correlation: In this method Pearson's linear correlation coefficient is calculated between the estimated feature vectors and the tissue-specific feature vectors based on expert labelling of the in-vivo MP-MRI voxels [111]. The expert label based feature vector for a particular tissue type is computed as the average of all the feature vectors from the voxels which are labelled by the expert as belonging to that tissue type.

7.3 Results and discussion

The feasibility of constrained CPD algorithm in differentiating tissue types is tested by applying it on fourteen MP-MRI datasets from patients with

high-grade glioma. The performance of the constrained CPD algorithm is compared with the MP-MRI hNMF algorithm [152]. Table 7.1 shows the dice score for active tumor and the tumor core region and source correlation for the active tumor tissue type as computed by the constrained CPD and the hNMF algorithm. Table 7.1 also shows the mean and standard deviation of the dice score and source correlation over 14 MP-MRI datasets. The constrained CPD algorithm has slightly better mean values for the dice score, however the increase was not significant. The mean source correlation value is clearly better using CPD compared to hNMF. For all the patients, the rank R is selected manually for both constrained CPD and hNMF. The l_1 regularization parameter of $\lambda = 0.2$ was used in the CPD algorithm.

Table 7.1: Dice scores and source correlation of fourteen HGG patients using CPD and hNMF algorithms. Dice scores are shown for active tumor and the tumor core (tumor + necrosis) region. Source correlation is shown for active tumor tissue type. Mean and standard deviation are shown in the last two rows, respectively.

	CPD- l_1			hNMF		
	Dice tumor	Dice core	Tumor source correlation	Dice tumor	Dice core	Tumor source correlation
PT01	0.83	0.96	0.94	0.81	0.98	0.50
PT02	0.77	0.85	0.95	0.78	0.86	0.92
PT03	0.90	0.95	0.97	0.81	0.92	0.51
PT04	0.81	0.88	0.95	0.74	0.82	0.88
PT05	0.89	0.89	0.98	0.71	0.71	0.78
PT06	0.88	0.95	0.99	0.82	0.94	0.62
PT07	0.76	0.88	0.80	0.74	0.90	0.54
PT08	0.81	0.94	0.91	0.75	0.86	0.93
PT09	0.69	0.83	0.93	0.76	0.90	0.98
PT10	0.79	0.59	0.95	0.54	0.50	0.74
PT11	0.87	0.87	0.98	0.88	0.88	0.96
PT12	0.76	0.76	0.98	0.75	0.75	0.98
PT13	0.92	0.92	0.98	0.93	0.93	1.00
PT14	0.88	0.88	0.99	0.89	0.89	0.98
Mean	0.83	0.87	0.95	0.78	0.85	0.81
std	0.07	0.10	0.05	0.09	0.13	0.19

The result of applying the constrained CPD algorithm on one of the MP-MRI datasets from a grade IV patient is shown in Fig 7.3 and 7.4. Fig 7.3 shows some of the relevant input MRI features (first two rows) and the abundance

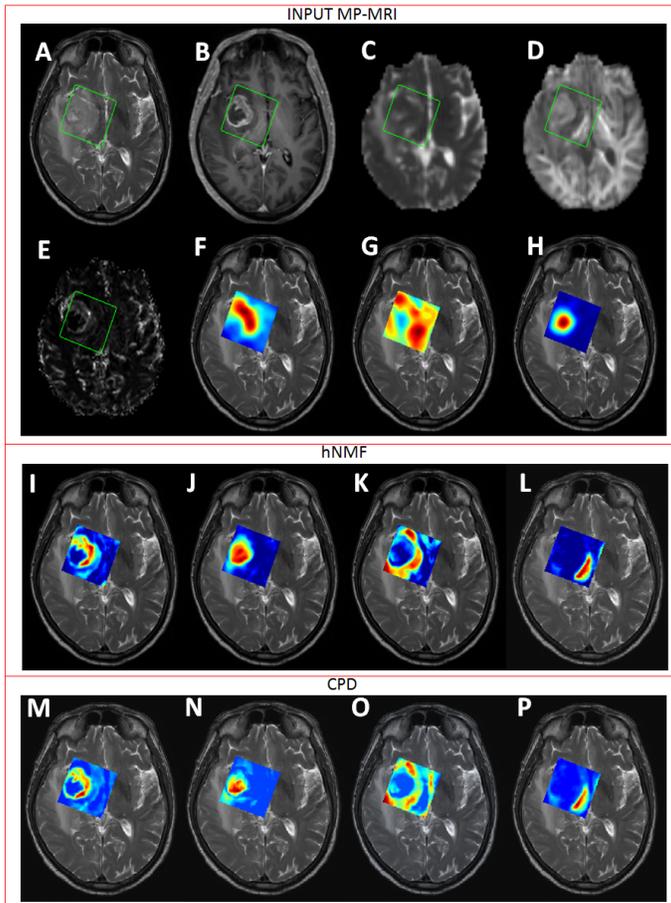


Figure 7.3: Coregistered MRI maps of several modalities and tissue abundance maps obtained from hNMF and CPD algorithms. The input maps: (A) T2, (B) T1+C, (C) MD, (D) MK, (E) rCBV, (F) Cho, (G) Cre and (H) Lip. The green box indicates the ROI. (I-L) tissue abundance maps obtained from hNMF ((I) Active tumor, (J) Necrosis, (K) Edema, (L) Cerebrospinal fluid). (M-P) tissue abundance maps obtained from constrained CPD ((M) Active tumor, (N) Necrosis, (O) Edema, (P) Cerebrospinal fluid).

maps obtained from hNMF and CPD algorithms. For hNMF only four out of six relevant abundance maps are shown, whereas in case of CPD a decomposition of rank $R = 4$ was performed and all abundance maps are shown. The sources which represent tissue-specific MP-MRI features obtained from CPD and hNMF

algorithms are shown in Fig. 7.4a. The segmented region of active tumor and necrosis obtained from the CPD and hNMF algorithms is shown in Fig. 7.4b along with the radiologist segmentation. From Fig 7.3 and 7.4 we can observe that for this patient (PT03 in Table 7.1) the CPD based algorithm provides a better active tumor segmentation than hNMF based algorithm ($\text{Dice tumor}_{CPD} = 0.90$ and $\text{Dice tumor}_{hNMF} = 0.81$).

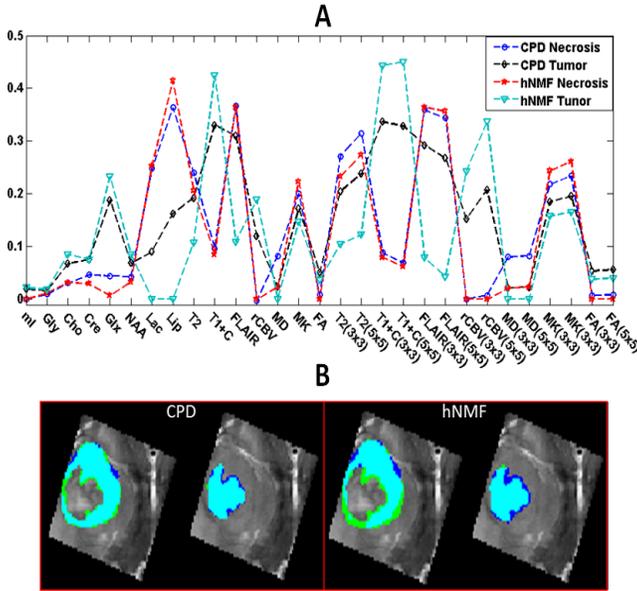


Figure 7.4: (A) Sources consisting of MP-MRI features obtained from CPD and hNMF algorithms. The estimated sources are shown for tumor and necrosis tissue type. (B) Comparison of the segmentation by radiologist (blue) with the segmentations obtained from CPD and hNMF (green). First two images correspond to tumor and necrosis segmentation from CPD, respectively. Last two images correspond segmentations from hNMF. Cyan indicates segmentation overlap.

The constrained CPD algorithm performs slightly better than the existing tissue type differentiation method based on hNMF, but the improvement in dice score is not significant. However, the CPD algorithm has clearly better source correlation than hNMF. Higher source correlation implies that the estimated source features for a particular tissue type better resembles the actual MP-MRI features observed in the region corresponding to that tissue type. This helps in identifying the tissue type from the sources. The main advantage of the tensor

approach is that we need to select only one rank R , whereas in the hNMF algorithm we need to select two ranks, one for each group in the second step of hNMF [152]. Therefore, automation of the CPD based algorithm will be much easier compared to hNMF. Also in the tensor approach, the tensor construction couples different MP-MRI features because of the \mathbf{xx}^T in the frontal slices. Therefore, in the sources obtained from the CPD algorithm the features will be coupled, meaning that an individual feature will not appear as a separate source. This is the reason for having higher source correlation values in the CPD algorithm. Moreover, for each voxel, the \mathbf{xx}^T in the frontal slices will give more weight to dominant features and least significant features will be suppressed.

7.4 Conclusion

In this chapter we have described a method to represent the MP-MRI data in a 3-D tensor. An algorithm using constrained canonical polyadic decomposition with l_1 regularization was proposed for tissue type differentiation in high-grade glioma patients from multi-parametric MRI. To explore feasibility of the proposed algorithm, it was tested on 14 MP-MRI datasets from patients having high-grade glioma. The constrained CPD algorithm has a better performance in terms source correlation and similar performance in terms of dice score compared to hNMF [152]. In this in-vivo study we have shown that a tensor formulation can be used for tumor characterization in multi-parametric MRI and further research in this direction is promising.

Chapter 8

Supervised tumor Segmentation in Multi-parametric MRI

In this chapter, we propose a fully automated superpixel-wise two-stage tumor tissue segmentation algorithm using random forest from MP-MRI data with only conventional MRI (T2, T1+C, T1 and FLAIR). The first stage is used to identify total tumor and the second stage to segment sub-regions. The features for the random forest classifier are extracted by constructing a tensor from the multimodal MRI data and applying multi-linear singular value decomposition. This method is trained using the BRATS 2017 training dataset and tested on validation and test datasets. The work presented in this chapter is based on [16].

8.1 Introduction

Accurate characterisation and localization of tissue types play a key role in brain tumor diagnosis and treatment planning. Neuro-imaging methods in particular magnetic resonance imaging (MRI) provide anatomical and pathophysiological information about brain tumors and aid in diagnosis, treatment planning and follow-up of patients. Manual segmentation of tumor tissue is a tedious and time consuming job, it also suffers from inter and intra-rater variability. An automated brain tumor segmentation algorithm will help to overcome those problems. However, automation of brain tumor tissue segmentation is

a difficult problem and often fails when applied on MRI images from different centres/scanners.

In the previous chapter we have developed a constrained CPD based un-supervised MP-MRI tumor segmentation method. This algorithm used MP-MRI data consisting of cMRI (T1c, T2 and FLAIR), DWI, PWI and MRSI and was analysed on a 2D region of interest specified by the MRSI modality. In practice, advanced MRI modalities like DWI, PWI and MRSI are often not available. Although, using advanced MRI modalities provides more refined segmentations, tumor segmentation algorithms should be capable providing good delineations without using advanced modalities. The performance of the constrained CPD algorithm deteriorated significantly when applied on the full slice without the MRSI modality. Since, in practice tumor segmentations are performed on 3D images, the un-supervised constrained CPD based algorithms are not suitable for tumour segmentation without including advanced modalities. Nowadays, supervised classification methods are receiving the most attention [127]. This is mainly due to availability of large public database, BRATS [127, 4, 11, 12] containing labelled MP-MRI images consisting of T2, T1, T1+contrast and FLAIR modality. All of the top performing algorithms in recent BRATS challenges use supervised algorithms. Therefore, in this chapter we use supervised methods to develop tumor segmentation algorithm using BRATS 2017 challenge dataset.

Supervoxels are gaining popularity in image segmentation algorithms, and have also been used in the context of brain tumor segmentation from MRI [192]. Performing supervoxel-level image segmentation offers certain advantages over pixel-level segmentation like spatial smoothness, capturing image redundancy and reducing computational complexity [192, 94]. These advantages motivated us to develop supervoxel level brain tumor segmentation. Supervoxels are mostly used for segmenting whole tumor and in this chapter we extend it to perform tumor sub-region segmentation.

Recently, tensor decompositions such as the canonical polyadic decomposition and the multilinear singular value decomposition (MLSVD) [160] have been used to extract features from high-dimensional data to use in classification algorithms [69]. Multimodal MRI consisting of T2, T1, T1+contrast and FLAIR imaging after co-registration and re-sampling to the same resolution, can be naturally represented as a third-order tensor. Our objective also includes exploring tensor decomposition method to extract feature that can exploit higher-order structure of MP-MRI data.

In this chapter we develop a fully automatic tumor supervoxel-level tissue segmentation algorithm using a random forest classifier. Multilinear singular value decomposition is used to extract features for the classifier.

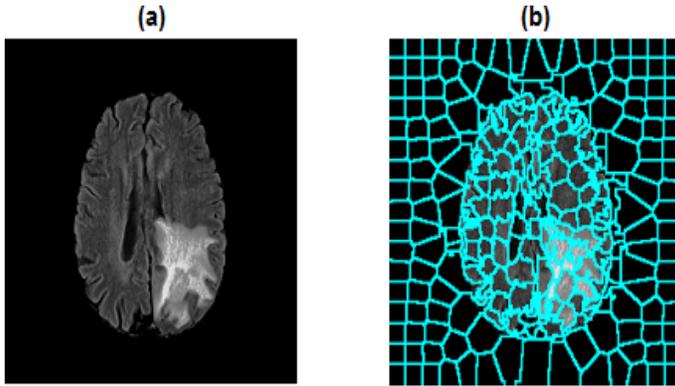


Figure 8.1: (a) One slice of the FLAIR image. (b) Generated superpixels for the slice in (a).

8.2 Method

8.2.1 Preprocessing

First, each individual 3D image is scaled to the range: 0-1. Next, intensities are normalized by applying histogram equalization. A reference histogram is generated by selecting 10 random images from the training set and extracting a histogram from the combined image. Histogram equalization is applied separately to different modalities. Background is removed from each slice using Ostu's image threshold method [137].

8.2.2 Feature extraction

The MR images are divided into smaller patches which are better aligned with intensity edges, called superpixels [5]. The superpixels are generated from each slice from one of the modalities as shown in Fig.8.1. The tissue assignment is done on superpixel-level instead of individual pixel, which helps to reduce computational cost and improve spatial smoothness [192].

Different features extracted from each superpixel are explained below

- Feature1: Intensity values are the most basic features that are used by almost all segmentation algorithms. Therefore for each superpixel, mean intensity values of all the four modalities and six difference images (e.g.: $\text{abs}(T1-T2)$) are extracted.

- Feature2: Entropy and standard deviation over each superpixel. Here we assume that these features are able to capture variations in each superpixel and these will be different for different tissues.
- Feature3: A third-order tensor is constructed for each superpixel, where the frontal slices are the covariance matrix of pixel-level features and the third mode is the modality and the difference images of the modalities. The covariance matrix is constructed over the pixels present in the superpixel. Pixel-level features consist of mean, median, standard deviation and entropy over a 5×5 window. Features are extracted by applying a rank-2 truncated multilinear singular value decomposition (MLSVD) on the third-order Tensor as shown in Fig. 8.2. Here we assume that the covariance between different pixel-level features will be different for different tissues.
- Feature4: A fourth-order tensor is constructed for each superpixel where the first two modes are 5×5 image patches with the main voxel at the centre, third mode consists of image patches from all four modalities (T1, T2, T1C and FLAIR) and patches from six difference images of the modalities ($\text{abs}(T1-T2)$, $\text{abs}(T1-T1C)$, $\text{abs}(T1-FLAIR)$, $\text{abs}(T1C-T2)$, $\text{abs}(T1C-FLAIR)$, $\text{abs}(T2-FLAIR)$) and the fourth mode consists of voxels within the superpixel. Again MLSVD is used for feature extraction. Only the mode-1, mode-2, mode-3 and the core tensor are used as feature. Here the features represent the dominant value of different modalities from a small 5×5 region measured across the superpixel, which was obtained by MLSVD.
- Feature5: A Third-order tensor is constructed for each superpixel, where the first mode is the pixels from 5×5 image patches with the main voxel at the centre, the second mode is the modality and the difference images (e.g.: $\text{abs}(T1-T2)$) of the modalities and the third mode consists of the voxels within the superpixel. Again, features are extracted by applying rank-2 MLSVD. Only the mode-1, mode-2 and core tensor are used as feature. This feature is similar feature4, where the tensor order is reduced by one by using a vectorised version of 5×5 image patches. Either feature4 or feature5 is used but not both.
- Feature6: For each superpixel a covariance matrix is estimated from the intensity values of all the modalities and the difference images. Covariance matrix plus two dominant eigenvectors and eigenvalues are used as features. This is also somewhat similar to feature3, where covariance is calculated between image modalities instead of pixel level feature.
- Feature7: Texture features are commonly used in segmentation. Local spectral histograms is one such feature, which represents texture

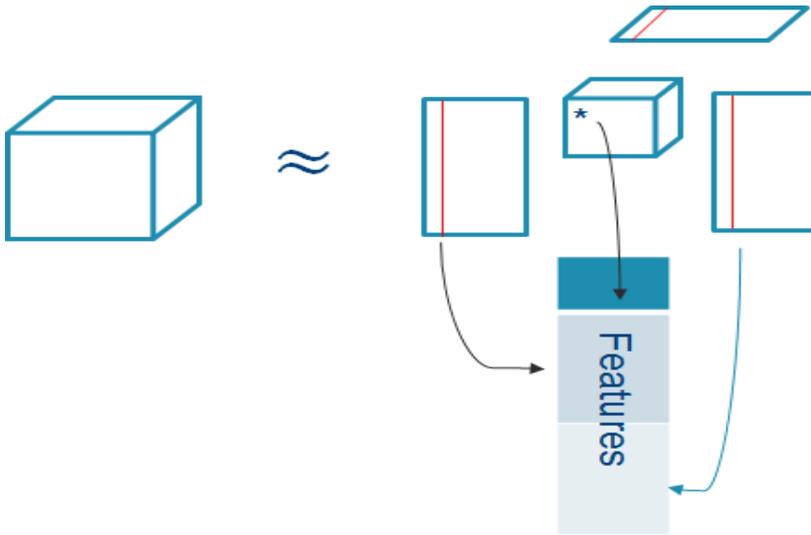


Figure 8.2: Truncated multilinear singular value decomposition and feature extraction.

descriptors based on a local distribution of filter responses [114]. Local spectral histograms with fifteen filters are estimated for all four modalities and six difference images. A third-order tensor is constructed for each superpixel, where the first mode includes local spectral histograms, second mode is the modality and the difference images of the modalities and the third mode consists of the voxels within the superpixel. Features are extracted by applying rank-2 MLSVD. Only the mode-1, mode-2 and core tensor are used as feature. The mean of the local spectral histograms over the superpixel are also used along with the MLSVD features.

8.2.3 Training and tissue segmentation

Tumor tissue segmentation was performed using a two-stage classifier. In the first stage a binary classification was performed on the superpixels to segment tumor and non-tumor regions. In the second-stage a multi-class classification was performed on the superpixels which are inside the estimated tumor region to segment enhancing tumor (ET), edema (ED), necrotic and non-enhancing tumor (NCR/NET) and healthy tissue. The two-stage operation is demonstrated in Fig. 8.3. For both stages a random forest classifier was used.

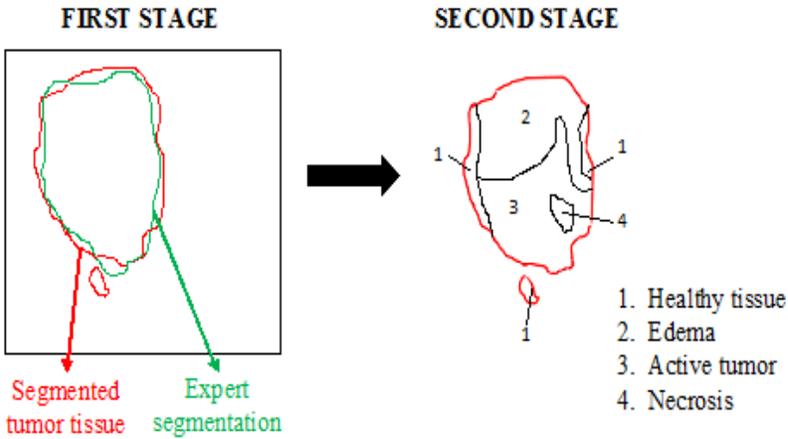


Figure 8.3: Demonstration of whole tumor segmentation in first stage and sub-tissue segmentation in second stage.

First stage:

In the first stage, superpixels are obtained from FLAIR because the total tumor is brighter in this modality. The feature set for this stage consists of Feature1, Feature2, Feature3 and Feature4. The dataset is divided in three groups, and random forest classifiers with 100 trees are trained from each group. The prediction is the result of majority voting of the classifiers learned from three data groups. For each model training is done iteratively, where a class balanced subset from the respective group is used for initial training. Next the trained model is tested on the remaining data from the respective group, the data that are classified wrongly are added to the initial subset and trained again with 100 tree random forest binary classifier. After the first stage classification at superpixel level, image filling and continuity-based denoising developed by [192] is performed on the whole tumor segmentation before going to the second stage.

Second stage:

In the second stage, superpixels are obtained from T1+contrast imaging modality because the enhancing tumor is brighter in this modality. Feature1, Feature5, Feature6 and Feature7 are used as features in this stage. Random forest classifiers with 250 trees are trained using an iterative method. Initially, 60 patients are randomly selected from the dataset for training and the trained model is tested on the remaining subset of the database. Next, the patients

which resulted in low Dice scores are included in the training set and a new model is trained. This is continued until all the patients in the dataset are used. Initially, the training was started with a balanced data. The generated superpixel in the second stage will not coincide exactly with the segmentation of the first stage. Therefore, all the superpixels which have less than 20% overlap with the segmentation of the first stage are discarded.

Initially, the first stage classifier was trained using Feature1, Feature2, Feature3 and Feature4. Since the performance on the validation set was satisfactory, we trained the second stage using the same features. However, the performance of the second-stage classifier on the validation set was bad. Therefore, additional features Feature5, Feature6 and Feature7 were included in the second-stage classifier to improve the performance. The list of all features with their corresponding dimension and the stage where they are used is shown Table 8.1.

Table 8.1: Features used in stage one and stage two along with their corresponding dimension.

	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Feature7
Dimension	10	20	36	48	78	122	196
Stage One	✓	✓	✓	✓	X	X	X
Stage Two	✓	X	X	X	✓	✓	✓

8.3 Results and Discussion

8.3.1 First Stage: whole tumor segmentation

A first stage model with three classifiers was trained using the BRATS 2017 training database [127, 4, 11, 12] containing 210 HGG and 75 LGG patients. The performance of the trained model in segmenting the whole tumor is tested on the validation dataset of BRATS 2017 challenge. Average Dice score and sensitivity obtained from the trained first stage model over 46 HGG patients are shown in Table 8.2. The boxplots of the Dice scores and sensitivity are shown in Fig. 8.4.

Table 8.2: Mean, standard deviation, median, 25 quantile and 75 quantile of Dice score and sensitivity for whole tumor (WT) over forty six patients from validation set using only first stage model.

	Dice WT	Sensitivity WT
Mean	0.8330	0.8574
Std	0.1186	0.1318
Median	0.8673	0.9024
25 quantile	0.8298	0.8114
75 quantile	0.9084	0.9415

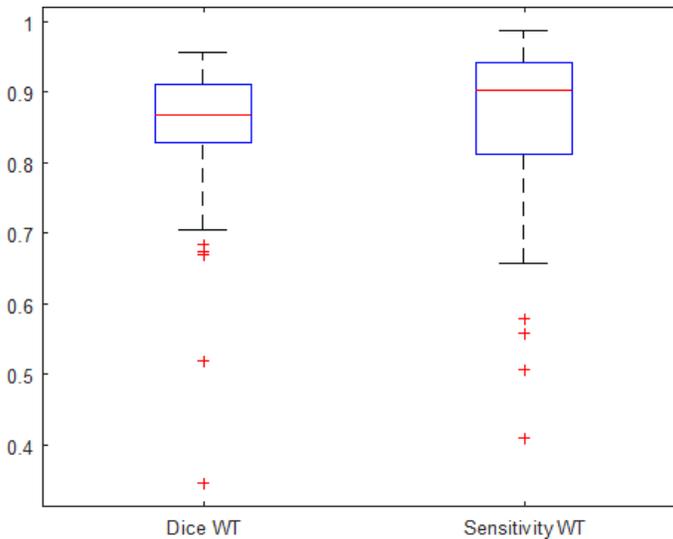


Figure 8.4: Boxplots of Dice scores and sensitivity for whole tumor (WT) obtained from first stage model on BRATS 2017 validation dataset of 46 patients.

8.3.2 High grade glioma

The BRATS 2017 high grade glioma database [127, 4, 11, 12] containing $N = 210$ patients is split into training set (70%) and test set (30%). A first stage model with a single classifier plus the second stage model is trained using only HGG and the trained model is tested on 63 HGG patients. Fig.8.5 shows the segmentation of tumor tissue for two different slices. We can observe from the figure that

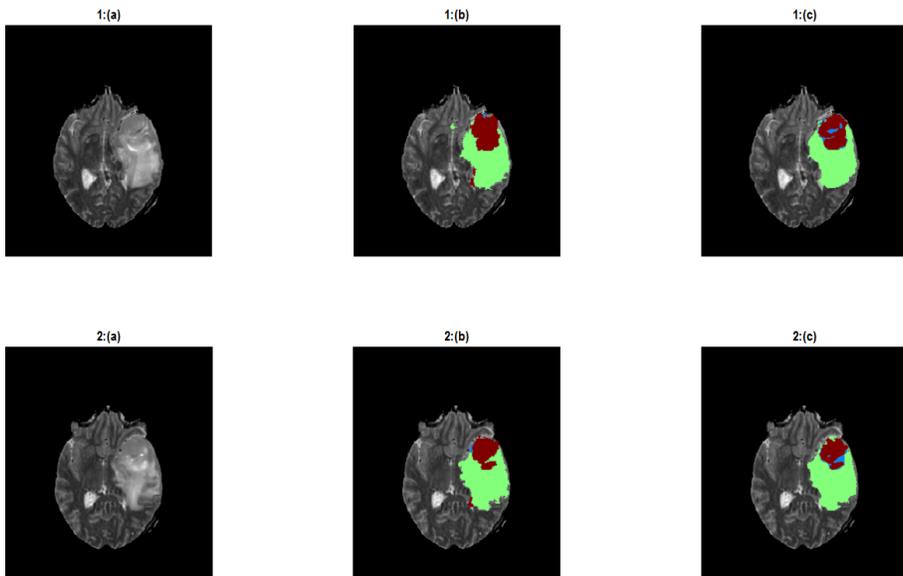


Figure 8.5: Segmentation results on two slices 1-2. (a) T2 image of one slice, (b) Estimated segmentation (c) Expert segmentation. Green-Edema, Brown-enhancing tumor and Blue- Necrosis.

the enhancing tumor and edema region are segmented properly. However the NCR/NET region is not identified properly.

The boxplot of the Dice scores is shown in Fig. 8.6 and the average Dice score and sensitivity obtained from the trained model for sixty three HGG patients are shown in Table 8.3. From the boxplot, we can observe that the algorithm performs well on most of the patients. however there are still some patients where the algorithm fails to segment properly.

Table 8.3: Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and sensitivity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over 63 HGG patients.

	Dice ET	Dice WT	Dice TC	Sensitivity ET	Sensitivity WT	Sensitivity TC
Mean	0.761	0.833	0.783	0.855	0.815	0.777
Std	0.106	0.096	0.147	0.126	0.090	0.191
Median	0.783	0.867	0.824	0.886	0.837	0.826
25 quantile	0.708	0.795	0.723	0.820	0.769	0.721
75 quantile	0.833	0.895	0.898	0.941	0.884	0.908

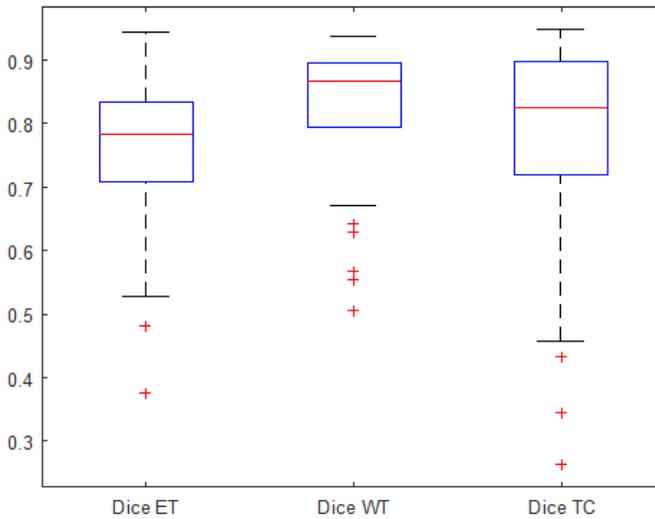


Figure 8.6: Boxplots of Dice scores for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) on BRATS 2017 training dataset of 63 patients.

8.3.3 Validation and test dataset results

The trained model consisting of both first and second stage is also tested on the BRATS 2017 validation dataset [127, 4, 11, 12]. The results are shown in Fig. 8.7 and Table 8.4. The performance is worse when compared to only HGG case.

Table 8.4: Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and sensitivity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over 46 validation dataset.

	Dice ET	Dice WT	Dice TC	Sensitivity ET	Sensitivity WT	Sensitivity TC
Mean	0.6125	0.7928	0.6734	0.6971	0.8320	0.6763
Std	0.3013	0.1217	0.2215	0.2347	0.1221	0.2297
Median	0.7419	0.8410	0.7291	0.7683	0.8614	0.6921
25 quantile	0.5240	0.7441	0.6607	0.6365	0.8053	0.5961
75 quantile	0.8383	0.8788	0.8213	0.8468	0.9169	0.8482

This algorithm does not identify necrotic and non-enhancing tumor (NCR/NET) tissue properly, it results in bad performance for LGG patients where NCR/NET tissue is larger than enhancing tumor in most cases. Also, when there is no enhancing tumor in patients, the algorithm identifies it falsely in some

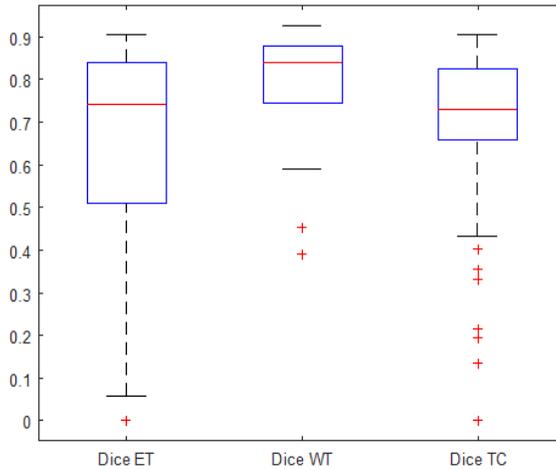


Figure 8.7: Boxplots of Dice scores for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) on BRATS 2017 validation dataset of 46 patients.

superpixels. This results in Dice score of zero for enhancing tumor, which can be seen from the boxplot in Fig. 8.7. Post-processing methods to remove such false positives may improve the average performance.

Finally the complete algorithm is applied on BRATS 2017 test dataset consisting of 146 patients. The average results are shown in Table 8.5.

Table 8.5: Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and Hausdorff95 for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over test dataset of 146 patients.

	Dice ET	Dice WT	Dice TC	Hausdorff95 ET	Hausdorff95 WT	Hausdorff95 TC
Mean	0.5032	0.7701	0.6105	71.27	31.33	38.90
Std	0.3054	0.1871	0.2954	132.94	30.61	84.13
Median	0.6376	0.8400	0.7416	6.56	14.56	9.56
25 quantile	0.2137	0.7202	0.4597	3.70	4.69	6.48
75 quantile	0.7447	0.8902	0.8358	39.46	54.40	29.73

8.4 Conclusion

In this chapter, we have developed a fully automated algorithm for brain tumor segmentation from multimodal MRI data. Superpixels and a tensor based feature extraction algorithm is proposed to be used with a two-stage random forest classifier for segmenting tumor tissue. The superpixels are restricted to 2-D slices because of the different resolution in the third dimension. In future work, the 2-D superpixels can be directly extended to 3D provided the difference in resolution is considered when constructing superpixels. The performance of the algorithm is comparable to the state-of-the-art methods when applied only to HGG patients. However, its performance deteriorates when tested on the BRATS 2017 validation and test database, which contains both low grade glioma (LGG) and HGG patients. The proposed superpixel method has good performance in segmenting the whole tumor using only the first stage model on the the BRATS 2017 validation set. However, this method does not perform well in segmenting the sub-regions, specifically the NCR/NET region. We assume that the features may not be discriminant enough to separate the NCR/NET region from others (mainly normal). In future work, identification of the NCR/NET region can be improved by using texture based features like Gabor and applying feature selection for selecting dominant and removing redundant features. Therefore, the proposed method is more suited to segment the whole tumor and not good in identifying sub-regions, especially in LGG patients.

Chapter 9

Conclusion and future perspectives

9.1 Conclusion

In this thesis we have explored the application of tensor based blind source separation techniques to magnetic resonance spectroscopic imaging and multi-parametric MRI. Tensor based blind source separation methods have been developed for simultaneous suppression of residual water from 2-D MRSI data. Both un-supervised algorithms using tensor based blind source separation techniques and supervised algorithms using random forest/CNN were developed for tumor characterization from 2-D MRSI and MP-MRI signals.

The mathematical and medical-physical background that is used in this thesis are briefly discussed in Chapter 1 and 2. Chapter 1 deals with the mathematical background, where basic concepts of tensor decompositions, various unsupervised blind source separation techniques and supervised classifiers are discussed. Medical-physical background is addressed in Chapter 2. Physical concepts of MRI and MRSI are explained along with a brief overview of advanced imaging modalities. Afterwards, the most common brain tumor, glioma is discussed along the applicability of neuroimaging techniques for diagnosis and treatment planning. In the last part, MRI acquisition parameters and data preprocessing pipelines were described for the UZ Leuven dataset.

Chapter 4: focuses on pre-processing of MRSI signal. In this chapter we have developed two tensor based methods to suppress residual water from MRSI signals. The first method used Löwner tensors constructed from spectra and CPD to extract water components. Similarly, in the second method a Hankel tensor constructed from FID and MLSVD was used to extract water components. The water components are then subtracted to remove residual water from the MRSI signal.

Tensor based methods were able to suppress the residual water simultaneously from all the voxels in the MRSI signal. In simulations, the Löwner-tensor based method resulted in the lowest average mean square error, followed by the Hankel-tensor based method. In some of the in-vivo MRSI signals, HSVD fails to suppress residual water in a few voxels. Whereas, tensor based methods overcome this problem by exploiting the shared information among the neighbouring voxels and thus resulting in much better residual water suppression. In in-vivo signals, the Löwner-tensor method has significantly lower difference in variance compared to the Hankel-tensor and the HSVD methods. The main conclusion is that in both simulated and in-vivo MRSI data, the tensor based methods outperform the widely-used subspace-based HSVD method.

One of the issues with the Löwner-tensor based water suppression method is the initialization of factor matrices. Some initializations might result in poor water suppression. In some cases, especially for FID's with small sample size a lot of initializations are required to obtain a proper residual water suppression. Another minor issue with the tensor based methods is the selection of rank R required for good quality water suppression. It is difficult to optimise R as a function of grid size and using a general R may not result in optimal water suppression.

In MRSI, B_0 field inhomogeneity can cause frequency shifts of complete spectra between voxels. The proposed tensor methods can handle only minor frequency shifts, and will result in improper water suppression in some of the voxels with bigger chemical shift. A pre-processing step where the spectra from all the voxels are aligned to a reference spectrum can overcome this problem.

HSVD has been widely used for residual water suppression in the clinical setting as a pre-processing step in metabolite quantification. A Löwner-tensor based method has the potential to replace the HSVD method in the clinical setting. Additional experiments such as analysing the effect of the rank on the voxel grid with various sizes, assessing the reliability of metabolite quantification in a test-retest experiment and testing on MRSI signals acquired using a different protocol are required to fine tune the algorithm before using it in MRSI processing software's for clinical applications.

Chapter 5: here we focused on applying tensor based BSS techniques for tumor tissue type differentiation in MRSI. Applying rank- $(L_r, L_r, 1)$ -BTD directly on the MRSI signal failed to extract any meaningful tissue patterns. To exploit the low-rank structure, a third-order tensor with \mathbf{xx}^T structure is constructed from the smoothed and reduced length spectra. Non-negative CPD based technique was applied to extract tissue specific spectra and their corresponding abundances. An automatic rank detection scheme was also proposed to estimate the number of tissue specific sources present in each of MRSI signals.

Tensor based methods can handle artefacts, extract tissue specific sources that characterize the actual tissue and generate more refined abundance maps. In extracting and localizing tumor tissue, the NCPD method has the best performance with a significantly higher mean and median values for source and distribution correlation compared to the NMF and the hNMF methods. For necrotic tissue, the performance of all the algorithms were good with the NCPD having slightly higher mean and median values for the source and the distribution correlation. The tensor based NCPD method was able to extract tumor tissue from patients where the matrix based NMF and hNMF algorithms failed. From the results, we can conclude that the NCPD algorithm significantly outperforms the existing tissue type differentiation methods based on NMF and hNMF. The use of l_1 regularization in the NCPD algorithm has minimal effect on the performance.

In the construction of the reduced spectra we used a window length of size $L = 20$ such that it covers the widest peak in the spectra. The choice of window length was flexible and resulted in similar performance when its size was increased or decreased. However, varying the window length by large amounts or using original spectrum has negative effect on the performance. The automatic rank estimation method used in this chapter provided a good estimate of rank for the MRSI dataset used in this thesis. However, it was not robust and was also sensitive to the cut-off value. The NCPD algorithm without automatic rank estimation can be easily adapted to handle different types of tumor, various MRSI signals acquired using different protocol and MRSI signals measured from various regions with a minimal adjustment to the window length.

Tissue specific sources can be used to identify the tissue type and nosologic images can be generated from abundances. These can be used by clinicians as an initial guess to identify the tumor region and various tissue types to further improve the diagnosis. To use this algorithm for identifying the voxels belonging to a particular tissue further modification and testing on large datasets are required.

Chapter 6: the NCPD algorithm proposed in Chapter 5 was extended to perform voxel classification. This chapter focuses on developing supervised algorithms, where a CNN based classifier is proposed to classify the voxels in the MRSI signal. Low-rank regularization was applied to the CNN classifier using MLSVD. The CNN methods have better performance with higher average sensitivity and F1 score when classifying tumors compared to that of the RF and the NCPD algorithms. The Low-rank regularized CNN has around 25% lower computational complexity than that of un-regularized CNN without any performance degradation. Although the CNN classifier has the best performance measure, other methods also have certain advantages. NCPD is an un-supervised method and does not require any training. It can be applied directly to a new MRSI signal which is obtained using different acquisition parameters (e.g. different length, sampling frequency). A random forest algorithm is faster compared to CNN and requires significantly less time for training.

In the random forest algorithm, using denoised original length (resolution) spectra resulted in lower performance compared to that of reduced spectra. This shows that construction of reduced spectra results in a better feature extraction. In CNN using Löwner matrix from denoised original length (resolution) spectra has similar performance with significant increase in computation time compared to that of Löwner matrix from reduced spectra.

To apply low-rank regularization, a cut-off value of 0.8 (80%) was used for truncating the MLSVD. The cut-off value was not sensitive for small changes and resulted in similar performance. The low-rank regularized CNN has fewer number of multiplications compared to that of un-regularized CNN, which enabled it to perform faster on a CPU. However, the low-rank regularized CNN was not faster on a GPU as these methods are not fully yet developed to exploit the parallel processing capabilities of GPU. These methods are still in the conceptual stage and requires extensive analysis using a large dataset for any potential clinical application.

Chapter 7: here more advanced MRI modalities are included to perform tumor segmentation from MP-MRI data. A method based on constrained canonical polyadic decomposition with l_1 regularization was developed to segment tumors from MP-MRI data consisting of cMRI (T1+C, T2 and FLAIR), MRSI, DWI and PWI. A constrained CPD algorithm provides tissue-specific signatures of the pathological tissue classes, as well as abundance maps of their spatial distribution. The abundance maps were converted into absolute tissue segmentation and good spatial alignment was found with manual segmentation by a neuro-radiologist. Results indicate that the constrained CPD algorithm has significantly better performance in terms of source correlation and similar performance in terms of

dice score compared to MP-MRI hNMF [152].

The tensor based blind source separation method, NCPD developed for tissue type differentiation in Chapter 5 can be easily adapted to handle MP-MRI data with very few modifications. However, such kind of adaptability is not possible for supervised algorithms. The automatic rank detection scheme developed for NCPD did not result in reliable rank estimation for constrained CPD. Therefore, the rank was selected manually for each patient. This method performed well when applied on a small region of interest which was defined by the scan area of the MRSI modality. However, when applied on the whole image by excluding the MRSI modality its performance was reduced significantly. Therefore, in the current state this algorithm has very limited clinical applicability.

Chapter 8: focuses on applicability of features generated from a tensor decomposition for supervised tumor segmentation algorithms from MP-MRI datasets. We have proposed a two stage random forest classifier using features generated from a superpixel by applying MLSVD. Its performance is good in segmenting the whole tumor using only the first stage model on the BRATS 2017 validation set. However, its performance deteriorates when segmenting the sub-regions, specifically the Necrotic(NCR)/non-enhancing tumor (NET) region. This study shows that tensor decompositions can be successfully applied to extract features for supervised tumor segmentation algorithms. However, the significance of such features has not been analysed. Further experiments in this direction are required.

We have made the assumption that the boundaries of the generated superpixels will coincide with the tumor delineation. This assumption was approximately satisfied in many patients when segmenting the whole tumor, which resulted in good dice scores for total tumor segmentation. However, the superpixel generated for sub-regions did not align well especially when the pathologic regions was small. This resulted in very low dice scores for some of the patients. The performance of the proposed method was not good compared to the state-of-the-art supervised algorithms. Therefore, further modifications such as working on pixel level instead of superpixel in the second stage and post-processing with conditional random field algorithms are required for this method to become clinically relevant.

9.2 Future perspectives

We have applied residual water suppression on 2D MRSI signals with a maximum grid size of 16×16 . In recent years MRSI signals are being acquired in 3D

[132] and in high resolution [121]. Applying tensor based water suppression algorithms to larger grids with many voxels is very demanding, both in terms of computation and memory. In the future, analysis of the proposed methods in terms of computation, memory, stability and performance can be performed when applied to very large voxel grids. In order to reduce the computation and memory load the tensor methods can be applied on smaller grids and then the results can be combined.

MRSI also suffers from other nuisance parameters like artefacts and baseline. A simple rank-1 model was sufficient to model water, but artefacts have much more variation and require higher rank models, which can be captured by block term decomposition. In future, block term decomposition based BSS algorithms have the potential to provide promising results in detecting and removing nuisance parameters from the MRSI signals.

In case of matrix based blind source separation, the hierarchical model was proved to be better for tissue characterization in both MRSI and MP-MRI [111, 152]. A fine tuned hierarchical model for the proposed tensor based algorithm will result in better tissue localization, especially for MP-MRI datasets. The proposed tensor based algorithms do not exploit the spatial information like tissue homogeneity in the local neighbourhood. Incorporating these in the proposed algorithms using a new tensor construction scheme or adding spatial regularization will result in a more refined tissue characterization.

The tensor based algorithm was proposed for brain tumor differentiation using short echo time MRSI signals. In future these methods can be applied to MRSI signals obtained with: 1) different acquisition parameters like long echo time, 2) from other parts of the body, 3) from other diseases such as prostate cancer or multiple sclerosis. We can also test this algorithm for other biomedical signals by slight modification as shown for MP-MRI case.

Construction of \mathbf{xx}^T structured tensor from a data matrices and applying rank- R canonical polyadic decomposition will implicitly perform clustering of data matrix into R groups. Therefore, exploring the usability of this tensor based method in various clustering applications is an interesting research direction. We can also compare its clustering performance with well known clustering methods such as k-means and spectral clustering.

In today's era of big-data, the MRSI dataset used to train CNN is small. In the future work the proposed CNN methods can be trained and tested on a much larger dataset. The voxels belonging to the tumor class can be further classified into active tumor and necrosis. This was not considered in this thesis as only a small number of necrotic samples were available for training. In future, when working on a large dataset, a four class (active tumor, necrosis, normal tissue

and bad quality) CNN can be considered. In this thesis, the Löwner matrix is used as input to CNN, but other methods of converting the spectra into a 2-D array such as the short time Fourier transform, wavelets and empirical mode decomposition can be considered as potential input feature for CNN. The superpixel and MLSVD based algorithm was trained and tested on BRATS 2017 challenge dataset. The post-operative GBM dataset of UZ Leuven also contains T1, T1+C, T2 Flair images, the same modalities as BRATS 2017 challenge dataset but with different resolution in the third direction. The proposed superpixel and MLSVD based algorithm can be applied with little modification to accommodate the post-operative MP-MRI scans. In this thesis we have used 2-D superpixels; this can be directly extended to 3-D by considering the difference in resolution in the other direction when constructing superpixels.

Most state-of-the-art tumor segmentations algorithms in MP-MRI are based on CNN. These CNN architecture are very big, containing millions of parameters. This poses a huge burden on the computation and memory during training and can also result in overfitting. Exploring tensor based low-rank regularization methods as a potential tool for reducing computational complexity and preventing overfitting is promising research.

Advanced methods proposed in this thesis need further fine tuning and more extensive validation before becoming clinical tools for the benefit of the patient. In this deep learning and big-data era, there are promising opportunities to further exploit such methods.

Appendix A

Tissue type assignment

Assigning a tissue type to the estimated source spectra will give valuable information. The spectra from the tumor tissue are characterised by an increase in choline (Cho) and decrease in N-acetylaspartate (NAA) and creatine (Cr) [72, 175, 88], whereas necrotic spectra are characterised by large values of lipid (Lip) and absence of other metabolites [55, 88]. To assign a tissue type to these source spectra, we have used the maximum values of Cho (3.13-3.30 ppm), NAA (1.95-2.10 ppm), Cr (2.92-3.08 ppm) and Lip (0.75-1.45 ppm) peaks (Figure A.1). The complete algorithm is described in the flow chart shown in Figure A.2. This method identifies seven tissue types: T (tumor), C (normal(control)), N (necrosis), B (bad/artifact), T/N (tumor and necrotic mixture), T/C (tumor and control mixture) and C/N (control and necrotic mixture). The flowchart outlining the tissue typing method is shown in Fig. A.2. The method starts by calculating the maximum, minimum and peak values from the real part of the source spectrum (absorption spectrum). The following steps are carried out for tissue typing:

- (I) Large negative peak?:- In this step we check for the presence of any large negative values in the real part of the spectrum using threshold τ_1 . If true the spectra is assigned as artifact (bad).
- (II) Large unexpected peak?:- This step checks for the presence of a any large peaks other than Naa, Cr, Cho or Lip using τ_2 . If true the spectrum is assigned as bad.
- (III) Max peak- In this step we check for the maximum peak among NAA, Cr, Lip and Cho. Here the algorithm splits into three branches corresponding

to maximum lipid peak, maximum NAA or Cr peak and maximum Cho peak as shown in the flowchart (Fig. A.2) .

- (IV) Too broad Lip?:- checks whether the Lip peak is too broad ($> 10\text{Hz}$). When the peak is too broad the spectral source is assigned to artifact (B).
- (V) One Large Peak?:- checks if any one peak far outweighs the other peaks using threshold τ_3 . When Lip outweigh others peaks, the source is assigned as N [111] and for other cases it is assigned as B.
- (VI) Lip present?:- checks for presence of significant amount of Lip using threshold τ_4 . This step is performed when NAA, Ch or Cho is the biggest peak. When a significant amount of Lip is present along with other peaks, the source spectrum will be assigned as a mixture of Necrotic and tumor (T/N) or normal (C/N).
- (VII) T or C?:- In this step we have used the ratio $\frac{\text{Cho}}{\text{NAA or Cr}}$ to differentiate between control (C) and tumor (T) tissue sources. NAA or Cr is chosen based on which is larger.

We have used threshold values of $\tau_1 = 0.6$, $\tau_2 = 0.6$, $\tau_3 = \frac{2}{3}$, $\tau_4 = 0.5$ and $\tau_5 = 0.8$, which can be adjusted to accommodate MRSI signals that are measured with a different acquisition protocol.

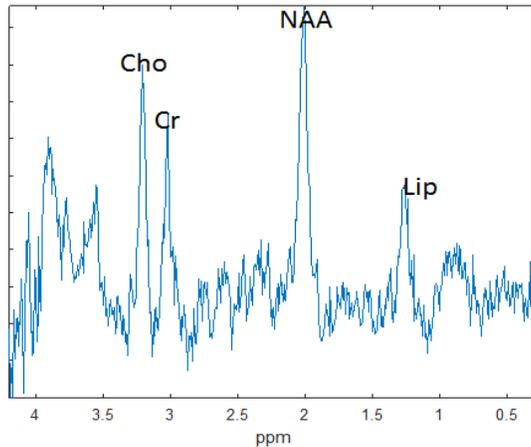


Figure A.1: Shows real part of the tissue specific spectrum ($real(W_i)$) obtained from NCPD. The peaks that are used for tissue type assignment are labelled.

For tumor voxel classification T, N, T/N, T/C, C/N are grouped into the tumor class.

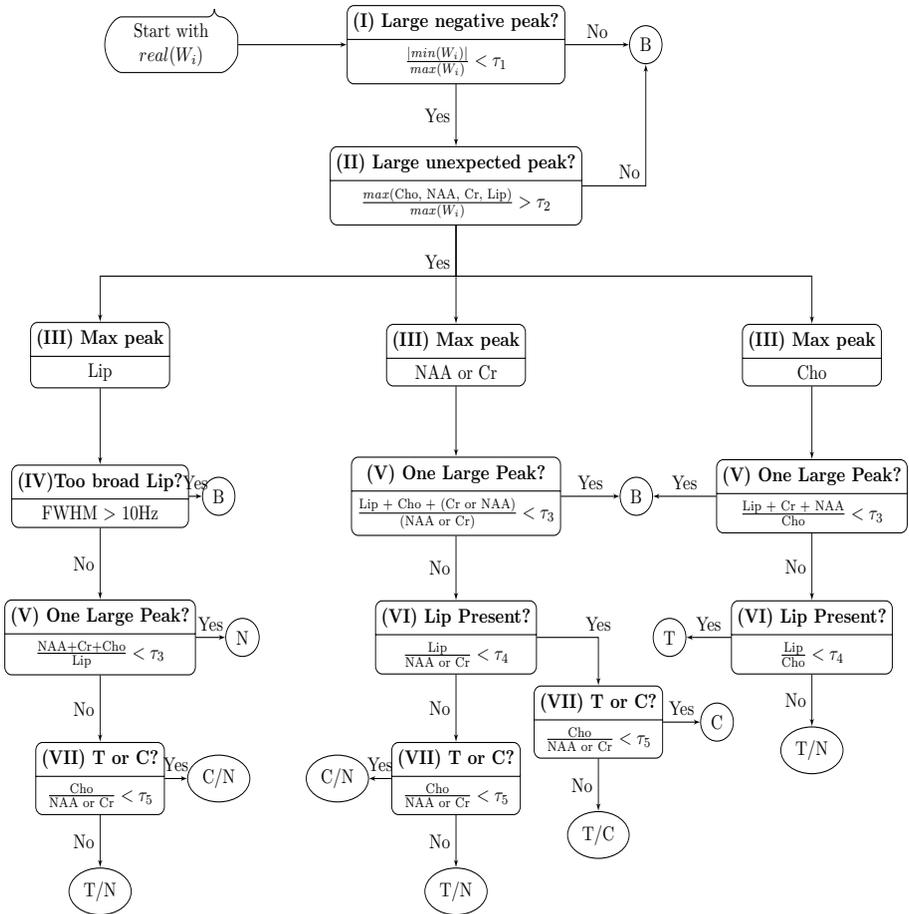


Figure A.2: Flowchart showing the procedure of assigning a tissue type to the source spectrum. B-bad; N-necrosis; C-control; T-tumor. FWHM stands for full width at half maximum.

Bibliography

- [1] Dynamic Susceptibility Contrast MR ANalysis (DSCoMAN). Available at: <https://sites.duke.edu/dblab/dscoman/>. Accessed: 2016-02-01.
- [2] Statistical Parametric Mapping (SPM). Available at: <http://www.fil.ion.ucl.ac.uk/spm/>. Accessed: 2016-02-01.
- [3] Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis, author=Harshman, Richard A.
- [4] Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, author=Bakas, Spyridon and Akbari, Hamed and Sotiras, Aristeidis and Bilello, Michel and Rozycki, Martin and Kirby, Justin S and Freymann, John B and Farahani, Keyvan and Davatzikos, Christos. *Scientific data* 4 (2017), 170117.
- [5] ACHANTA, R., SHAJI, A., SMITH, K., LUCCHI, A., FUA, P., AND SÜSTRUNK, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* 34, 11 (2012), 2274–2282.
- [6] AKAIKE, H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19, 6 (1974), 716–723.
- [7] ALPAYDIN, E. *Introduction to machine learning*. MIT press, 2014.
- [8] ANDERSSON, J., XU, J., YACOB, E., AUERBACH, E., MOELLER, S., AND UGURBIL, K. A comprehensive Gaussian process framework for correcting distortions and movements in diffusion images. In *Proceedings of the 20th Annual Meeting of ISMRM* (2012), p. 2426.
- [9] ANTOULAS, A., AND ANDERSON, B. On the scalar rational interpolation problem. *IMA Journal of Mathematical Control and Information* 3, 2-3 (1986), 61–88.

- [10] ANTOULAS, T. A tutorial introduction to the Loewner framework for model reduction. 9th Elgersburg Workshop Mathematische Systemtheorie, 2014.
- [11] BAKAS, S., AKBARI, H., SOTIRAS, A., BILELLO, M., ROZYCKI, M., KIRBY, J., FREYMAN, J., FARAHANI, K., AND DAVATZIKOS, C. Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. *The Cancer Imaging Archive* (2017).
- [12] BAKAS, S., AKBARI, H., SOTIRAS, A., BILELLO, M., ROZYCKI, M., KIRBY, J., FREYMAN, J., FARAHANI, K., AND DAVATZIKOS, C. Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. *The Cancer Imaging Archive* (2017).
- [13] BARKHUIJSEN, H., DE BEER, R., AND VAN ORMONDT, D. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *Journal of Magnetic Resonance (1969)* 73, 3 (1987), 553–557.
- [14] BASSER, P. J., MATTIELLO, J., AND LEBIHAN, D. MR diffusion tensor spectroscopy and imaging. *Biophysical journal* 66, 1 (1994), 259–267.
- [15] BERTHOLDO, D., WATCHARAKORN, A., AND CASTILLO, M. Brain proton magnetic resonance spectroscopy: introduction and overview. *Neuroimaging Clinics* 23, 3 (2013), 359–380.
- [16] BHARATH, H., COLLEMAN, S., SIMA, D., AND VAN HUFFEL, S. Tumor Segmentation from Multimodal MRI Using Random Forest with Superpixel and Tensor Based Feature Extraction. In: *Crimi A., Bakas S., Kuijff H., Menze B., Reyes M. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017. Lecture Notes in Computer Science 10670* (2018), 463–473.
- [17] BHARATH, H., SAUWEN, N., SIMA, D. M., HIMMELREICH, U., DE LATHAUWER, L., AND VAN HUFFEL, S. Canonical polyadic decomposition for tissue type differentiation using multi-parametric MRI in high-grade gliomas. In *Signal Processing Conference (EUSIPCO), 2016 24th European* (2016), IEEE, pp. 547–551.
- [18] BHARATH, H., SIMA, D., SAUWEN, N., HIMMELREICH, U., DE LATHAUWER, L., AND VAN HUFFEL, S. Nonnegative Canonical Polyadic Decomposition for Tissue-Type Differentiation in Gliomas. *IEEE journal of biomedical and health informatics* 21, 4 (2017), 1124–1132.
- [19] BHARATH, H. N., DEBALS, O., SIMA, D. M., HIMMELREICH, U., DE LATHAUWER, L., AND VAN HUFFEL, S. Tensor Based Method for

- Residual Water Suppression in ^1H Magnetic Resonance Spectroscopic Imaging. Accepted for publication in *IEEE Transactions on Biomedical Engineering*, 2018.
- [20] BHARATH, H. N., SIMA, D. M., SAUWEN, N., HIMMELREICH, U., DE LATHAUWER, L., AND VAN HUFFEL, S. Tensor based tumor tissue type differentiation using magnetic resonance spectroscopic imaging. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (Aug 2015), pp. 7003–7006.
- [21] BHARATH, H. N., SIMA, D. M., SAUWEN, N., HIMMELREICH, U., LATHAUWER, L. D., AND HUFFEL, S. V. Nonnegative Canonical Polyadic Decomposition for Tissue-Type Differentiation in Gliomas. *IEEE Journal of Biomedical and Health Informatics* 21, 4 (July 2017), 1124–1132.
- [22] BOLAN, P. J., NELSON, M. T., YEE, D., AND GARWOOD, M. Imaging in breast cancer: magnetic resonance spectroscopy. *Breast Cancer Research* 7, 4 (2005), 1.
- [23] BOTTOMLEY, P. A. Spatial localization in NMR spectroscopy in vivo. *Annals of the New York Academy of Sciences* 508, 1 (1987), 333–348.
- [24] BOUTSIDIS, C., AND GALLOPOULOS, E. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition* 41, 4 (2008), 1350–1362.
- [25] BOXERMAN, J., SCHMAINDA, K., AND WEISSKOFF, R. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *American Journal of Neuroradiology* 27, 4 (2006), 859–867.
- [26] BREIMAN, L. Bagging predictors. *Machine learning* 24, 2 (1996), 123–140.
- [27] BREIMAN, L. Random forests. *Machine learning* 45, 1 (2001), 5–32.
- [28] BRO, R., AND ANDERSSON, C. A. Improving the speed of multiway algorithms: Part II: Compression. *Chemometrics and intelligent laboratory systems* 42, 1 (1998), 105–113.
- [29] BULIK, M., JANCALEK, R., VANICEK, J., SKOCH, A., AND MECHL, M. Potential of MR spectroscopy for assessment of glioma grading. *Clinical neurology and neurosurgery* 115, 2 (2013), 146–153.
- [30] BYRNES, T. J., BARRICK, T. R., BELL, B. A., AND CLARK, C. A. Diffusion tensor imaging discriminates between glioblastoma and cerebral metastases in vivo. *NMR in biomedicine* 24, 1 (2011), 54–60.

- [31] C PEET, A., ARVANITIS, T., O LEACH, M., AND WALDMAN, A. Functional imaging in adult and paediatric brain tumours.
- [32] CABANES, E., CONFORT-GOUNY, S., LE FUR, Y., SIMOND, G., AND COZZONE, P. J. Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain. *Journal of Magnetic Resonance* 150, 2 (2001), 116–125.
- [33] CARROLL, J. D., AND CHANG, J.-J. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika* 35, 3 (1970), 283–319.
- [34] CASALINO, G., DEL BUONO, N., AND MENCAR, C. Subtractive clustering for seeding non-negative matrix factorizations. *Information Sciences* 257 (2014), 369–387.
- [35] CAVASSILA, S., DEVAL, S., HUEGEN, C., VAN ORMONDT, D., AND GRAVERON-DEMILLY, D. Cramér–Rao bounds: an evaluation tool for quantitation. *NMR in Biomedicine* 14, 4 (2001), 278–283.
- [36] CICHOCKI, A., ZDUNEK, R., AND AMARI, S.-I. Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In *International Conference on Independent Component Analysis and Signal Separation* (2007), Springer, pp. 169–176.
- [37] CICHOCKI, A., ZDUNEK, R., PHAN, A. H., AND AMARI, S. I. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [38] CICHOCKI, A. AND MANDIC, D. AND DE LATHAUWER, L. AND GUOXU ZHOU AND QIBIN ZHAO AND CAIAFA, C. AND PHAN, H.A. Tensor Decompositions for Signal Processing Applications: From two-way to multiway component analysis. *Signal Processing Magazine, IEEE* 32, 2 (March 2015), 145–163.
- [39] COMON, P., AND JUTTEN, C. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [40] CONSTANTIN, A., ELKHALED, A., JALBERT, L., SRINIVASAN, R., CHA, S., CHANG, S. M., BAJCSY, R., AND NELSON, S. J. Identifying malignant transformations in recurrent low grade gliomas using high resolution magic angle spinning spectroscopy. *Artificial intelligence in medicine* 55, 1 (2012), 61–70.

- [41] CROITOR SAVA A. R., SIMA, D. M., POULLET, J.-B., WRIGHT, A. J., HEERSCHAP, A., AND VAN HUFFEL, S. Exploiting spatial information to estimate metabolite levels in two-dimensional MRSI of heterogeneous brain lesions. *NMR in Biomedicine* 24, 7 (2011), 824–835.
- [42] DAGER, S. R., OSKIN, N., RICHARDS, T. L., AND POSSE, S. Research applications of magnetic resonance spectroscopy (MRS) to investigate psychiatric disorders. *Topics in magnetic resonance imaging: TMRI* 19, 2 (2008), 81.
- [43] DE ALMEIDA, A. L., FAVIER, G., AND MOTA, J. C. Space-time multiplexing codes: A tensor modeling approach. In *Signal Processing Advances in Wireless Communications, 2006. SPAWC'06. IEEE 7th Workshop on* (2006), IEEE, pp. 1–5.
- [44] DE ALMEIDA, A. L., FAVIER, G., AND MOTA, J. C. M. PARAFAC-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization. *Signal Processing* 87, 2 (2007), 337–351.
- [45] DE EDELENYI, F. S., SIMONETTI, A., POSTMA, G., HUO, R., AND BUYDENS, L. Application of independent component analysis to ^1H MR spectroscopic imaging exams of brain tumours. *Analytica Chimica Acta* 544, 1 (2005), 36–46.
- [46] DE LATHAUWER, L. Decompositions of a higher-order tensor in block terms—part II: definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications* 30, 3 (2008), 1033–1066.
- [47] DE LATHAUWER, L. Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices. *SIAM Journal on Matrix Analysis and Applications* 30, 3 (2008), 1022–1032.
- [48] DE LATHAUWER, L., DE MOOR, B., AND VANDEWALLE, J. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications* 21, 4 (2000), 1253–1278.
- [49] DE LATHAUWER, L., DE MOOR, B., AND VANDEWALLE, J. Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM journal on Matrix Analysis and Applications* 26, 2 (2004), 295–327.
- [50] DE LATHAUWER, L., AND NION, D. Decompositions of a higher-order tensor in block terms—Part III: Alternating least squares algorithms. *SIAM journal on Matrix Analysis and Applications* 30, 3 (2008), 1067–1083.

- [51] DE LATHAUWER, LIEVEN. Blind Separation of Exponential Polynomials and the Decomposition of a Tensor in Rank- $(L_R, L_R, 1)$ Terms. *SIAM J. Matrix Anal. Appl.* 32, 4 (Dec. 2011), 1451–1474.
- [52] DEANGELIS, L. M. Brain tumors. *New England Journal of Medicine* 344, 2 (2001), 114–123.
- [53] DEBALS, O. *Tensorization and applications in blind source separation*. PhD thesis, Department of Electrical Engineering, KU Leuven, Leuven, Belgium, 2017.
- [54] DEBALS, O., VAN BAREL, M., AND DE LATHAUWER, L. Löwner-Based Blind Signal Separation of Rational Functions With Applications. *Signal Processing, IEEE Transactions on* 64, 8 (April 2016), 1909–1918.
- [55] DELIKATNY, E. J., CHAWLA, S., LEUNG, D.-J., AND POPTANI, H. MR-visible lipids and the tumor microenvironment. *NMR in Biomedicine* 24, 6 (2011), 592–611.
- [56] DETRE, J. A., RAO, H., WANG, D. J., CHEN, Y. F., AND WANG, Z. Applications of arterial spin labeled MRI in the brain. *Journal of Magnetic Resonance Imaging* 35, 5 (2012), 1026–1037.
- [57] DEVARAJAN, K. Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS computational biology* 4, 7 (2008), e1000029.
- [58] DIMOU, S., BATTISTI, R., HERMENS, D., AND LAGOPOULOS, J. A systematic review of functional magnetic resonance imaging and diffusion tensor imaging modalities used in presurgical planning of brain tumour resection. *Neurosurgical review* 36, 2 (2013), 205–214.
- [59] DING, C. H., LI, T., AND JORDAN, M. I. Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence* 32, 1 (2010), 45–55.
- [60] DOLECEK, T. A., PROPP, J. M., STROUP, N. E., AND KRUCHKO, C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009. *Neuro-oncology* 14, suppl_5 (2012), v1–v49.
- [61] DOMANOV, I., AND DE LATHAUWER, L. On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors — Part I: Basic Results and Uniqueness of One Factor Matrix. *SIAM Journal on Matrix Analysis and Applications* 34, 3 (2013), 855–875.

- [62] DOMANOV, I., AND DE LATHAUWER, L. On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors — Part II: Uniqueness of the Overall Decomposition. *SIAM Journal on Matrix Analysis and Applications* 34, 3 (2013), 876–903.
- [63] DOMANOV, I., AND DE LATHAUWER, L. On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Basic results and uniqueness of one factor matrix. *SIAM Journal on Matrix Analysis and Applications* 34, 3 (2013), 876–903.
- [64] DOMANOV, I., AND DE LATHAUWER, L. Generic uniqueness of a structured matrix factorization and applications in blind source separation. *IEEE Journal of Selected Topics in Signal Processing* 10, 4 (June 2016), 701–711.
- [65] DOMANOV, I., AND LATHAUWER, L. D. Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition. *SIAM Journal on Matrix Analysis and Applications* 35, 2 (2014), 636–660.
- [66] DONG, Z. Proton MRS and MRSI of the brain without water suppression. *Progress in nuclear magnetic resonance spectroscopy* 86 (2015), 65–79.
- [67] DU, B., ZHANG, M., ZHANG, L., HU, R., AND TAO, D. PLTD: Patch-based low-rank tensor decomposition for hyperspectral images. *IEEE Transactions on Multimedia* 19, 1 (2017), 67–79.
- [68] ER, F. C., HATAY, G. H., OKEER, E., YILDIRIM, M., HAKYEMEZ, B., AND OZTURK-ISIK, E. Classification of phosphorus magnetic resonance spectroscopic imaging of brain tumors using support vector machine and logistic regression at 3T. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (2014), IEEE, pp. 2392–2395.
- [69] FARGEAS, A., ALBERA, L., KACHENOURA, A., DRÉAN, G., OSPINA, J.-D., COLOIGNER, J., LAFOND, C., DELOBEL, J.-B., DE CREVOISIER, R., AND ACOSTA, O. On feature extraction and classification in prostate cancer radiotherapy using tensor decompositions. *Medical Engineering and Physics* 37, 1 (2015), 126–131.
- [70] FÖRSTERLING, F. H. Spin dynamics: basics of nuclear magnetic resonance. *Medical Physics* 37, 1 (2010), 406–407.
- [71] FORSYTH, P. A., AND POSNER, J. B. Headaches in patients with brain tumors A study of 111 patients. *Neurology* 43, 9 (1993), 1678–1678.

- [72] FOUNTAS, K. N., KAPSALAKI, E. Z., GOTSIS, S. D., KAPSALAKIS, J. Z., SMISSON III, H. F., JOHNSTON, K. W., ROBINSON JR, J. S., AND PAPADAKIS, N. In vivo proton magnetic resonance spectroscopy of brain tumors. *Stereotactic and functional neurosurgery* 74, 2 (2000), 83–94.
- [73] FRAHM, J., MERBOLDT, K.-D., AND HÄNICKE, W. Localized proton spectroscopy using stimulated echoes. *Journal of Magnetic Resonance (1969)* 72, 3 (1987), 502–508.
- [74] FRIEDMAN, S., SHAW, D., ARTRU, A., RICHARDS, T., GARDNER, J., DAWSON, G., POSSE, S., AND DAGER, S. Regional brain chemical alterations in young children with autism spectrum disorder. *Neurology* 60, 1 (2003), 100–107.
- [75] GIBBONS, J. D., AND CHAKRABORTI, S. *Nonparametric statistical inference*. Springer, 2011.
- [76] GILLIS, N., ET AL. Nonnegative matrix factorization: Complexity, algorithms and applications. *Unpublished doctoral dissertation, Université catholique de Louvain. Louvain-La-Neuve: CORE* (2011).
- [77] GILLIS, N., AND GLINEUR, F. Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization. *Neural computation* 24, 4 (2012), 1085–1105.
- [78] GILLIS, N., KUANG, D., AND PARK, H. Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing* 53, 4 (2015), 2066–2078.
- [79] GOODFELLOW, I., BENGIO, Y., AND COURVILLE, A. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [80] GOVINDARAJU, V., YOUNG, K., AND MAUDSLEY, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine* 13, 3 (2000), 129–153.
- [81] HAARDT, M., ROEMER, F., AND DEL GALDO, G. Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. *IEEE Transactions on Signal Processing* 56, 7 (2008), 3198–3213.
- [82] HAASE, A., FRAHM, J., HANICKE, W., AND MATTHAEI, D. ^1H NMR chemical shift selective (CHESS) imaging. *Physics in Medicine & Biology* 30, 4 (1985), 341.

- [83] HAYKIN, S. S., HAYKIN, S. S., HAYKIN, S. S., AND HAYKIN, S. S. *Neural networks and learning machines*, vol. 3. Pearson Upper Saddle River, NJ, USA:, 2009.
- [84] HE, K., ZHANG, X., REN, S., AND SUN, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision (2015)*, pp. 1026–1034.
- [85] HINTON, G., DENG, L., YU, D., DAHL, G. E., MOHAMED, A.-R., JAITLY, N., SENIOR, A., VANHOUCHE, V., NGUYEN, P., SAINATH, T. N., ET AL. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine* 29, 6 (2012), 82–97.
- [86] HO, T. K. Random decision forests. In *Document analysis and recognition, 1995., proceedings of the third international conference on (1995)*, vol. 1, IEEE, pp. 278–282.
- [87] HORE, V., VIÑUELA, A., BUIL, A., KNIGHT, J., MCCARTHY, M. I., SMALL, K., AND MARCHINI, J. Tensor decomposition for multiple-tissue gene expression experiments. *Nature genetics* 48, 9 (2016), 1094.
- [88] HOWE, F., BARTON, S., CUDLIP, S., STUBBS, M., SAUNDERS, D., MURPHY, M., WILKINS, P., OPSTAD, K., DOYLE, V., MCLEAN, M., ET AL. Metabolic profiles of human brain tumors using quantitative in vivo ^1H magnetic resonance spectroscopy. *Magnetic Resonance in Medicine* 49, 2 (2003), 223–232.
- [89] HUNYADI, B., CAMPS, D., SORBER, L., VAN PAESSCHEN, W., DE VOS, M., VAN HUFFEL, S., AND DE LATHAUWER, L. Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing* 2014, 1 (2014), 139.
- [90] IOFFE, S., AND SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).
- [91] ISHTEVA, M. *Numerical methods for the best low multilinear rank approximation of higher-order tensors*. PhD thesis, Department of Electrical Engineering, KU Leuven, Leuven, Belgium, 2009.
- [92] IŞIN, A., DIREKOĞLU, C., AND ŞAH, M. Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science* 102 (2016), 317–324.

- [93] JENSEN, J. H., HELPERN, J. A., RAMANI, A., LU, H., AND KACZYNSKI, K. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic resonance in medicine* 53, 6 (2005), 1432–1440.
- [94] JIA, S., AND ZHANG, C. Fast and robust image segmentation using an superpixel based fcm algorithm. In *Image Processing (ICIP), 2014 IEEE International Conference on* (2014), IEEE, pp. 947–951.
- [95] KARPATY, A. CS231n convolutional neural networks for visual recognition. <http://cs231n.github.io/convolutional-networks/>, 2018. [Online; accessed 20-April-2018].
- [96] KEOGH, B. P., AND HENSON, J. W. Clinical manifestations and diagnostic imaging of brain tumors. *Hematology/oncology clinics of North America* 26, 4 (2012), 733–755.
- [97] KERSTING, K., WAHABZADA, M., THURAU, C., AND BAUCKHAGE, C. Hierarchical convex NMF for clustering massive data. In *Proceedings of 2nd Asian Conference on Machine Learning* (2010), pp. 253–268.
- [98] KOH, K., KIM, S., AND BOYD, S. l1_ls: A Matlab Solver for Large-Scale L1-Regularized Least Squares Problems. http://web.stanford.edu/~boyd/l1_ls/, 2008.
- [99] KOLDA, T. G., AND BADER, B. W. Tensor decompositions and applications. *SIAM review* 51, 3 (2009), 455–500.
- [100] KOUROU, K., EXARCHOS, T. P., EXARCHOS, K. P., KARAMOUZIS, M. V., AND FOTIADIS, D. I. Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal* 13 (2015), 8–17.
- [101] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.
- [102] KROMPASS, D., NICKEL, M., JIANG, X., AND TRESP, V. Non-negative tensor factorization with rescal. In *Tensor Methods for Machine Learning, ECML workshop* (2013).
- [103] KROONENBERG, P. M. *Applied multiway data analysis*, vol. 702. John Wiley & Sons, 2008.
- [104] KUNG, S.-Y., ARUN, K. S., AND RAO, D. B. State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem. *JOSA* 73, 12 (1983), 1799–1811.

- [105] LANIADO, M., WEINMANN, H., SCHÖRNER, W., FELIX, R., AND SPECK, U. First use of GdDTPA/dimeglumine in man. *Physiological chemistry and physics and medical NMR* 16, 2 (1984), 157–165.
- [106] LAUDADIO, T., MASTRONARDI, N., VANHAMME, L., VAN HECKE, P., AND VAN HUFFEL, S. Improved Lanczos algorithms for blackbox MRS data quantitation. *Journal of Magnetic Resonance* 157, 2 (2002), 292–297.
- [107] LE BIHAN, D., MANGIN, J.-F., POUPON, C., CLARK, C. A., PAPPATA, S., MOLKO, N., AND CHABRIAT, H. Diffusion tensor imaging: concepts and applications. *Journal of magnetic resonance imaging* 13, 4 (2001), 534–546.
- [108] LEE, D. D., AND SEUNG, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (1999), 788.
- [109] LEMANS, A., JEURISSEN, B., SIJBERS, J., AND JONES, D. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In *17th Annual Meeting of Intl Soc Mag Reson Med* (2009), vol. 209, p. 3537.
- [110] LI, S., WANG, W., QI, H., AYHAN, B., KWAN, C., AND VANCE, S. Low-rank tensor decomposition based anomaly detection for hyperspectral imagery. In *Image Processing (ICIP), 2015 IEEE International Conference on* (2015), IEEE, pp. 4525–4529.
- [111] LI, Y., SIMA, D. M., CAUTER, S. V., CROITOR SAVA A. R., HIMMELREICH, U., PI, Y., AND VAN HUFFEL, S. Hierarchical non-negative matrix factorization (hNMF): a tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI. *NMR in Biomedicine* 26, 3 (2013), 307–319.
- [112] LI, Y., SIMA, D. M., VAN CAUTER, S., HIMMELREICH, U., SAVA, A. R. C., PI, Y., LIU, Y., AND VAN HUFFEL, S. Unsupervised nosologic imaging for glioma diagnosis. *Biomedical Engineering, IEEE Transactions on* 60, 6 (2013), 1760–1763.
- [113] LITJENS, G., KOOI, T., BEJNORDI, B. E., SETIO, A. A. A., CIOMPI, F., GHAFORIAN, M., VAN DER LAAK, J. A., VAN GINNEKEN, B., AND SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. *Medical image analysis* 42 (2017), 60–88.
- [114] LIU, X., AND WANG, D. A spectral histogram model for texton modeling and texture discrimination. *Vision Research* 42, 23 (2002), 2617–2634.

- [115] LOUIS, D. N., OHGAKI, H., WIESTLER, O. D., CAVENEE, W. K., BURGER, P. C., JOUVET, A., SCHEITHAUER, B. W., AND KLEIHUES, P. The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica* 114, 2 (2007), 97–109.
- [116] LÖWNER, K. Über monotone matrixfunktionen. *Mathematische Zeitschrift* 38, 1 (1934), 177–216.
- [117] LU, H., PLATANIOTIS, K. N., AND VENETSANOPOULOS, A. N. A survey of multilinear subspace learning for tensor data. *Pattern Recognition* 44, 7 (2011), 1540–1551.
- [118] LUTS, J., POULLET, J.-B., GARCIA-GOMEZ, J. M., HEERSCHAP, A., ROBLES, M., SUYKENS, J. A., AND HUFFEL, S. V. Effect of feature extraction for brain tumor classification based on short echo time ^1H MR spectra. *Magnetic resonance in medicine* 60, 2 (2008), 288–298.
- [119] M., B. A., M., B. M., AND MICHAEL, Z. *Blind Source Separation*. American Cancer Society, 2006.
- [120] MA, C., LAM, F., JOHNSON, C. L., AND LIANG, Z.-P. Removal of nuisance signals from limited and sparse ^1H MRSI data using a union-of-subspaces model. *Magnetic resonance in medicine* 75, 2 (2016), 488–497.
- [121] MA, C., LAM, F., NING, Q., JOHNSON, C. L., AND LIANG, Z.-P. High-resolution ^1H -MRSI of the brain using short-TE SPICE. *Magnetic resonance in medicine* 77, 2 (2017), 467–479.
- [122] MAES, F., COLLIGNON, A., VANDERMEULEN, D., MARCHAL, G., AND SUETENS, P. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on* 16, 2 (1997), 187–198.
- [123] MAHYARI, A. G., ZOLTOWSKI, D. M., BERNAT, E. M., AND AVIYENTE, S. A Tensor Decomposition-Based Approach for Detecting Dynamic Network States From EEG. *IEEE Transactions on Biomedical Engineering* 64, 1 (2017), 225–237.
- [124] MAIER, S. E., SUN, Y., AND MULKERN, R. V. Diffusion imaging of brain tumors. *NMR in biomedicine* 23, 7 (2010), 849–864.
- [125] MARTINEZ-MONTES, E., VALDÉS-SOSA, P. A., MIWAKEICHI, F., GOLDMAN, R. I., AND COHEN, M. S. Concurrent EEG/fMRI analysis by multiway partial least squares. *NeuroImage* 22, 3 (2004), 1023–1034.

- [126] MCLEOD, K., SERMESANT, M., BEERBAUM, P., AND PENNEC, X. Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. *IEEE transactions on medical imaging* 34, 7 (2015), 1562–1575.
- [127] MENZE, B., JAKAB, A., BAUER, S., KALPATHY-CRAMER, J., FARAHANI, K., KIRBY, J., BURREN, Y., PORZ, N., SLOTBOOM, J., WIEST, R., LANCZI, L., GERSTNER, E., WEBER, M., ARBEL, T., AVANTS, B., AYACHE, N., BUENDIA, P., COLLINS, D., CORDIER, N., CORSO, J., CRIMINISI, A., DAS, T., DELINGETTE, H., DEMIRALP, C., DURST, C., DOJAT, M., DOYLE, S., FESTA, J., FORBES, F., GEREMIA, E., GLOCKER, B., GOLLAND, P., GUO, X., HAMAMCI, A., IFTEKHARUDDIN, K., JENA, R., JOHN, N., KONUKOGLU, E., LASHKARI, D., MARIZ, J., MEIER, R., PEREIRA, S., PRECUP, D., PRICE, S., RAVIV, T., REZA, S., RYAN, M., SARIKAYA, D., SCHWARTZ, L., SHIN, H., SHOTTON, J., SILVA, C., SOUSA, N., SUBBANNA, N., SZEKELY, G., TAYLOR, T., THOMAS, O., TUSTISON, N., UNAL, G., VASSEUR, F., WINTERMARK, M., YE, D., ZHAO, L., ZHAO, B., ZIKIC, D., PRASTAWA, M., REYES, M., AND VAN LEEMPUT, K. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging* 34, 10 (2015), 1993–2024.
- [128] MØRUP, M. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1, 1 (2011), 24–40.
- [129] MØRUP, M., HANSEN, L. K., ARNFRED, S. M., LIM, L.-H., AND MADSEN, K. H. Shift-invariant multilinear decomposition of neuroimaging data. *NeuroImage* 42, 4 (2008), 1439 – 1450.
- [130] MOUNTFORD, C. E., STANWELL, P., LIN, A., RAMADAN, S., AND ROSS, B. Neurospectroscopy: the past, present and future. *Chemical reviews* 110, 5 (2010), 3060–3086.
- [131] MURALIDHARA, C., GROSS, A. M., GUTELL, R. R., AND ALTER, O. Tensor decomposition reveals concurrent evolutionary convergences and divergences and correlations with structural motifs in ribosomal RNA. *PloS one* 6, 4 (2011), e18768.
- [132] MURUGANANDHAM, M., CLERKIN, P. P., SMITH, B. J., ANDERSON, C. M., MORRIS, A., CAPIZZANO, A. A., MAGNOTTA, V., MCGUIRE, S. M., SMITH, M. C., BAYOUTH, J. E., ET AL. 3-Dimensional magnetic resonance spectroscopic imaging at 3 Tesla for early response assessment of glioblastoma patients during external beam radiation

- therapy. *International Journal of Radiation Oncology • Biology • Physics* 90, 1 (2014), 181–189.
- [133] NGUYEN, C., WANG, Y., AND NGUYEN, H. N. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. *Journal of Biomedical Science and Engineering* 6, 05 (2013), 551.
- [134] OGG, R. J., KINGSLEY, R., AND TAYLOR, J. S. WET, a T1-and B1-insensitive water-suppression method for in vivo localized ^1H NMR spectroscopy. *Journal of Magnetic Resonance, Series B* 104, 1 (1994), 1–10.
- [135] ORTEGA-MARTORELL, S., LISBOA, P. J., VELLIDO, A., SIMÕES, R. V., PUMAROLA, M., JULIÀ-SAPÉ, M., AND ARÚS, C. Convex non-negative matrix factorization for brain tumor delimitation from MRSI data. *PloS one* 7, 10 (2012), e47824.
- [136] OSORIO GARCIA, M. I. *Advanced signal processing for magnetic resonance spectroscopy*. PhD thesis, Department of Electrical Engineering, KU Leuven, Leuven, Belgium, 2011.
- [137] OTSU, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* 9, 1 (1979), 62–66.
- [138] PAPY, J.-M., DE LATHAUWER, L., AND VAN HUFFEL, S. Exponential data fitting using multilinear algebra: the single-channel and multi-channel case. *Numerical linear algebra with applications* 12, 8 (2005), 809–826.
- [139] PHAM, T. D., AND YAN, H. Tensor Decomposition of Gait Dynamics in Parkinson’s Disease. *IEEE Transactions on Biomedical Engineering* (2017).
- [140] PHAN, A. H., AND CICHOCKI, A. Tensor decompositions for feature extraction and classification of high dimensional datasets. *Nonlinear theory and its applications, IEICE* 1, 1 (2010), 37–68.
- [141] PHAN, A.-H., TICHAVSKY, P., AND CICHOCKI, A. Low complexity damped Gauss–Newton algorithms for CANDECOMP/PARAFAC. *SIAM Journal on Matrix Analysis and Applications* 34, 1 (2013), 126–147.
- [142] POOT, D. H., DEN DEKKER, A. J., ACHTEN, E., VERHOYE, M., AND SIJBERS, J. Optimal experimental design for diffusion kurtosis imaging. *Medical Imaging, IEEE Transactions on* 29, 3 (2010), 819–829.
- [143] POULLET, J. B. *Quantification and classification of magnetic resonance spectroscopic data for brain tumor diagnosis*. PhD thesis, Department of Electrical Engineering, KU Leuven, Leuven, Belgium, 2008.

- [144] POULLET, J.-B., SIMA, D. M., SIMONETTI, A. W., DE NEUTER, B., VANHAMME, L., LEMMERLING, P., AND VAN HUFFEL, S. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. *NMR in Biomedicine* 20, 5 (2007), 493–504.
- [145] POULLET, J.-B., SIMA, D. M., VAN HUFFEL, S., AND VAN HECKE, P. Frequency-selective quantitation of short-echo time ^1H magnetic resonance spectra. *Journal of Magnetic Resonance* 186, 2 (2007), 293–304.
- [146] PROVENCHER, S. W. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic resonance in medicine* 30, 6 (1993), 672–679.
- [147] RAAB, P., HATTINGEN, E., FRANZ, K., ZANELLA, F. E., AND LANFERMANN, H. Cerebral gliomas: diffusional kurtosis imaging analysis of microstructural differences. *Radiology* 254, 3 (2010), 876–881.
- [148] RATINEY, H., SDIKA, M., COENRADIE, Y., CAVASSILA, S., ORMONDT, D. V., AND GRAVERON-DEMILLY, D. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR in Biomedicine* 18, 1 (2005), 1–13.
- [149] RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATY, A., KHOSLA, A., BERNSTEIN, M., BERG, A. C., AND FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115, 3 (2015), 211–252.
- [150] SAJDA, P., DU, S., BROWN, T. R., STOYANOVA, R., SHUNGU, D. C., MAO, X., AND PARRA, L. C. Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *IEEE transactions on medical imaging* 23, 12 (2004), 1453–1465.
- [151] SANEI, S., AND CHAMBERS, J. A. *EEG signal processing*. John Wiley & Sons, 2013.
- [152] SAUWEN, NICOLAS AND SIMA, DIANA M. AND VAN CAUTER, SOFIE AND VERAART, JELLE AND LEEMANS, ALEXANDER AND MAES, FREDERIK AND HIMMELREICH, UWE AND VAN HUFFEL, SABINE. Hierarchical non-negative matrix factorization to characterize brain tumor heterogeneity using multi-parametric MRI. *NMR in Biomedicine* 28, 12 (2015), 1599–1624.
- [153] SAXENA, A. Convolutional Neural Networks (CNNs): An Illustrated Explanation. <https://xrds.acm.org/blog/2016/06/convolutional->

- neural-networks-cnns-illustrated-explanation/, 2016. [Online; accessed 20-April-2018].
- [154] SCHNEIDER, C. A., RASBAND, W. S., ELICEIRI, K. W., ET AL. NIH Image to ImageJ: 25 years of image analysis. *Nat methods* 9, 7 (2012), 671–675.
- [155] SEN, B., AND PARHI, K. K. Extraction of common task signals and spatial maps from group fmri using a parafac-based tensor decomposition technique. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on* (2017), IEEE, pp. 1113–1117.
- [156] SHAHNAZ, F., BERRY, M. W., PAUCA, V. P., AND PLEMMONS, R. J. Document clustering using nonnegative matrix factorization. *Information Processing & Management* 42, 2 (2006), 373–386.
- [157] SHIROISHI, M. S., CASTELLAZZI, G., BOXERMAN, J. L., D’AMORE, F., ESSIG, M., NGUYEN, T. B., PROVENZALE, J. M., ENTERLINE, D. S., ANZALONE, N., DÖRFLER, A., ET AL. Principles of T2*-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging. *Journal of Magnetic Resonance Imaging* 41, 2 (2015), 296–313.
- [158] SIDIROPOULOS, N., DE LATHAUWER, L., FU, X., HUANG, K., PAPALEXAKIS, E., AND FALOUTSOS, C. Tensor Decomposition for Signal Processing and Machine Learning. *Signal Processing, IEEE Transactions on*. to be published.
- [159] SIDIROPOULOS, N. D., BRO, R., AND GIANNAKIS, G. B. Parallel factor analysis in sensor array processing. *IEEE transactions on Signal Processing* 48, 8 (2000), 2377–2388.
- [160] SIDIROPOULOS, N. D., DE LATHAUWER, L., FU, X., HUANG, K., PAPALEXAKIS, E. E., AND FALOUTSOS, C. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing* 65, 13 (2017), 3551–3582.
- [161] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLU, I., PANNEERSHELVAM, V., LANCTOT, M., ET AL. Mastering the game of Go with deep neural networks and tree search. *nature* 529, 7587 (2016), 484–489.
- [162] SIZOO, E. M., BRAAM, L., POSTMA, T. J., PASMAN, H. R. W., HEIMANS, J. J., KLEIN, M., REIJNEVELD, J. C., AND TAPHOORN, M. J. Symptoms and problems in the end-of-life phase of high-grade glioma patients. *Neuro-oncology* 12, 11 (2010), 1162–1166.

- [163] SMILDE, A., BRO, R., AND GELADI, P. *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons, 2005.
- [164] SMITH, E. A., CARLOS, R. C., JUNCK, L. R., TSIEN, C. I., ELIAS, A., AND SUNDGREN, P. C. Developing a clinical decision model: MR spectroscopy to differentiate between recurrent tumor and radiation change in patients with new contrast-enhancing lesions. *American Journal of Roentgenology* 192, 2 (2009), W45–W52.
- [165] SOHER, B., SEMANCHUK, P., TODD, D., STEINBERG, J., AND YOUNG, K. VeSPA: integrated applications for RF pulse design, spectral simulation and MRS data analysis. In *Proceedings of 19th Annual Meeting ISMRM* (2011), p. 1410.
- [166] SORBER, L., VAN BAREL, M., AND DE LATHAUWER, L. Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-($L_r, L_r, 1$) terms, and a new generalization. *SIAM Journal on Optimization* 23, 2 (2013), 695–720.
- [167] SORBER, L. AND VAN BAREL, M. AND DE LATHAUWER, L. Structured Data Fusion. *Selected Topics in Signal Processing, IEEE Journal of* 9, 4 (June 2015), 586–600.
- [168] SØRENSEN, M., AND DE LATHAUWER, L. Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition. *IEEE Transactions on Signal Processing* 61, 22 (2013), 5507–5519.
- [169] STEFAN, D., DI CESARE, F., ANDRASESCU, A., POPA, E., LAZARIEV, A., VESCOVO, E., STRBAK, O., WILLIAMS, S., STARCUK, Z., CABANAS, M., ET AL. Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package. *Measurement Science and Technology* 20, 10 (2009), 104035.
- [170] SUNDIN, T., VANHAMME, L., VAN HECKE, P., DOLOGLOU, I., AND VAN HUFFEL, S. Accurate quantification of ^1H spectra: From finite impulse response filter design for solvent suppression to parameter estimation. *Journal of Magnetic Resonance* 139, 2 (1999), 189–204.
- [171] SYMEONIDIS, P., NANOPOULOS, A., AND MANOLOPOULOS, Y. A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis. *IEEE Transactions on Knowledge and Data Engineering* 22, 2 (2010), 179–192.
- [172] TAN, V. Y., AND FÉVOTTE, C. Automatic Relevance Determination in Nonnegative Matrix Factorization with the β -Divergence. *Pattern*

- Analysis and Machine Intelligence, IEEE Transactions on* 35, 7 (2013), 1592–1605.
- [173] TOFTS, P. S., BRIX, G., BUCKLEY, D. L., EVELHOCH, J. L., HENDERSON, E., KNOPP, M. V., LARSSON, H. B., LEE, T.-Y., MAYR, N. A., PARKER, G. J., ET AL. Estimating kinetic parameters from dynamic contrast-enhanced T 1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *Journal of magnetic resonance imaging* 10, 3 (1999), 223–232.
- [174] TU, D., CHEN, L., LV, M., SHI, H., AND CHEN, G. Hierarchical online NMF for detecting and tracking topic hierarchies in a text stream. *Pattern Recognition* 76 (2018), 203–214.
- [175] TZIKA, A. A., ASTRAKAS, L. G., ZARIFI, M. K., ZURAKOWSKI, D., POUSSAINT, T. Y., GOUNNEROVA, L., TARBELL, N. J., AND BLACK, P. M. Spectroscopic and perfusion magnetic resonance imaging predictors of progression in pediatric brain tumors. *Cancer* 100, 6 (2004), 1246–1256.
- [176] VAN CAUTER, S., DE KEYZER, F., SIMA, D. M., CROITOR SAVA A. R., D’ARCO, F., VERAART, J., PEETERS, R. R., LEEMANS, A., VAN GOOL, S., WILMS, G., DEMAEREL, P., VAN HUFFEL, S., SUNAERT, S., AND HIMMELREICH, U. Integrating diffusion kurtosis imaging, dynamic susceptibility-weighted contrast-enhanced MRI, and short echo time chemical shift imaging for grading gliomas. *Neuro-oncology* 16, 7 (2014), 1010–1021.
- [177] VAN CAUTER, S., SIMA, D. M., LUTS, J., TER BEEK, L., RIBBENS, A., PEETERS, R. R., OSORIO GARCIA, M. I., LI, Y., SUNAERT, S., VAN GOOL, S. W., ET AL. Reproducibility of rapid short echo time CSI at 3 Tesla for clinical applications. *Journal of Magnetic Resonance Imaging* 37, 2 (2013), 445–456.
- [178] VAN CAUTER, S., VERAART, J., SIJBERS, J., PEETERS, R. R., HIMMELREICH, U., DE KEYZER, F., VAN GOOL, S. W., VAN CALENBERGH, F., DE VLEESCHOUWER, S., VAN HECKE, W., ET AL. Gliomas: diffusion kurtosis MR imaging in grading. *Radiology* 263, 2 (2012), 492–501.
- [179] VAN HUFFEL, S. Tensor Decompositions in Smart Patient Monitoring. *SIAM News* 42, 7 (Sep. 2015), 1.
- [180] VANHAMME, L., FIERRO, R. D., VAN HUFFEL, S., AND DE BEER, R. Fast removal of residual water in proton spectra. *Journal of Magnetic Resonance* 132, 2 (1998), 197–203.

- [181] VASILESCU, M. A. O., AND TERZOPOULOS, D. Multilinear analysis of image ensembles: Tensorfaces. In *European Conference on Computer Vision* (2002), Springer, pp. 447–460.
- [182] VAVASIS, S. A. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization* 20, 3 (2009), 1364–1377.
- [183] VEDALDI, A., AND LENC, K. MatConvNet – Convolutional Neural Networks for MATLAB. In *Proceeding of the ACM Int. Conf. on Multimedia* (2015).
- [184] VERAART, J., SIJBERS, J., SUNAERT, S., LEEMANS, A., AND JEURISSEN, B. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. *NeuroImage* 81 (2013), 335–346.
- [185] VERVLIET, N., DEBALS, O., SORBER, L., VAN BAREL, M., AND DE LATHAUWER, L. Tensorlab 3.0. <https://www.tensorlab.net>, 2016.
- [186] VERVLIET, N., DEBALS, O., SORBER, L., VAN BAREL, M., AND DE LATHAUWER, L. Tensorlab 3.0, March 2016.
- [187] WANG, Y., PENG, J., ZHAO, Q., LEUNG, Y., ZHAO, X.-L., AND MENG, D. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2017).
- [188] WAX, M., AND KAILATH, T. Detection of signals by information theoretic criteria. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 33, 2 (1985), 387–392.
- [189] WEYBRIGHT, P., SUNDGREN, P. C., MALY, P., HASSAN, D. G., NAN, B., ROHRER, S., AND JUNCK, L. Differentiation between brain tumor recurrence and radiation injury using MR spectroscopy. *American Journal of Roentgenology* 185, 6 (2005), 1471–1476.
- [190] WILD, S., CURRY, J., AND DOUGHERTY, A. Improving non-negative matrix factorizations through structured initialization. *Pattern Recognition* 37, 11 (2004), 2217–2232.
- [191] WILSON, M., REYNOLDS, G., KAUPPINEN, R. A., ARVANITIS, T. N., AND PEET, A. C. A constrained least-squares approach to the automated quantitation of in vivo ^1H magnetic resonance spectroscopy data. *Magnetic resonance in medicine* 65, 1 (2011), 1–12.
- [192] WU, W., CHEN, A. Y., ZHAO, L., AND CORSO, J. J. Brain tumor detection and segmentation in a CRF (conditional random fields)

- framework with pixel-pairwise affinity and superpixel-level features. *International journal of computer assisted radiology and surgery* 9, 2 (2014), 241–253.
- [193] XU, F., BAI, X., AND ZHOU, J. Non-local similarity based tensor decomposition for hyperspectral image denoising. In *Image Processing (ICIP), 2017 IEEE International Conference on* (2017), IEEE, pp. 1890–1894.
- [194] YUNPENG, C., XIAOJIE, J., BINGYI, K., JIASHI, F., AND SHUICHENG, Y. Sharing Residual Units Through Collective Tensor Factorization in Deep Neural Networks. *arXiv preprint arXiv:1703.02180* (2017).

Curriculum vitae

Bharath Halandur Nagaraja was born in Kundapura, Karnataka, India on January 1, 1984. He received his Bachelor of Engineering degree in the field of electronics and communication from The National Institute of Engineering, Mysore, India in 2005. He received his M.Sc (by research) degree in the field of electrical engineering from the Indian Institute of Technology Madras, India in 2012. His master thesis was on signal processing techniques for biomedical applications.

Between February 2013 and April 2014 he was employed as a Senior R&D Engineer at National Brain Research Centre, where he worked on the processing and analysis of the magnetic resonance spectroscopic imaging (MRSI) data. Prior to M.Sc, he has worked in WeP Peripherals Ltd, Mysore as a Design Engineer. He was involved in the design and development of firmware for retail billing and dot matrix printers

In May 2014, he joined KU Leuven Electrical Engineering Department (ESAT-STADIUS) as a PhD student under the supervision of Prof. Sabine Van Huffel. His research has been a part of work package six (WP6) in the BIOTENSORS project, funded by ERC Advanced Grant: BIOTENSORS (n° 339804), with the aim of applying tensor based blind source separation techniques to magnetic resonance spectroscopic imaging.

List of publications

Papers in international journals

1. **Bharath, H. N.**, Debals, O., Sima, D. M., Himmelreich, U., De Lathauwer, L., and Van Huffel, S. Tensor based method for residual water suppression in ^1H magnetic resonance spectroscopic imaging. Accepted for publication in IEEE Transactions on Biomedical Engineering, 2018.
2. Sauwen, N., Acou, M., **Bharath, H. N.**, Sima, D.M., Veraart, J., Maes, F., Himmelreich, U., Achten, E. and Van Huffel, S. The successive projection algorithm as an initialization method for brain tumor segmentation using non-negative matrix factorization. PloS one. 2017 Aug 28;12(8):e0180268.. <https://doi.org/10.1371/journal.pone.0180268>.
3. **Bharath, H. N.**, Sima, D. M., Sauwen, N., Himmelreich, U., Lathauwer, L. D., and Huffel, S. V. Nonnegative canonical polyadic decomposition for tissue-type differentiation in gliomas. IEEE Journal of Biomedical and Health Informatics 21, 4 (July 2017), 1124–1132.

Papers in proceedings of international conferences

1. **Bharath, H. N.**, Coleman, S., Sima, D.M., Van Huffel, S. (2018) Tumor Segmentation from Multimodal MRI Using Random Forest with Superpixel and Tensor Based Feature Extraction. In: Crimi A., Bakas S., Kuijff H., Menze B., Reyes M. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017. Lecture Notes in Computer Science, vol 10670. Springer, Cham.
2. **Bharath, H. N.**, Sauwen, N., Sima, D.M., Himmelreich, U., De Lathauwer, L., and Van Huffel, S. Canonical polyadic decomposition for tissue type

differentiation using multi-parametric MRI in high-grade gliomas. European Signal Processing Conference (EUSIPCO) 2016, Budapest, Aug. 2016 (pp. 547-551).

3. Sauwen, N., Acou, M., **Bharath, H. N.**, Sima, D.M., Veraart, J., Maes, F., Himmelreich, U., Achten, E., and Van Huffel, S. Initializing nonnegative matrix factorization using the successive projection algorithm for multi-parametric medical image segmentation. 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, May 2016 (pp.265-270).

4. **Bharath, H. N.**, Sima, D.M., Sauwen, N., Himmelreich, U., De Lathauwer, L., and Van Huffel, S. Tensor based tumor tissue type differentiation using magnetic resonance spectroscopic imaging. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Aug. 2015 (pp. 7003-7006).

Abstracts in proceedings of international conferences

1. **Bharath, H.N.**, Sima, D.M. and Van Huffel, S. Tensor Based Approaches in Magnetic Resonance Spectroscopic Signal Analysis. SIAM Conference on Computational Science and Engineering (SIAM CSE 17), Atlanta, USA, Feb. 2017.

2. **Bharath, H.N.**, Sima, D.M. and Van Huffel, S. Löwner Based Residual Water Suppression in Magnetic Resonance Spectroscopic Imaging. 6th Dutch Bio-Medical Engineering Conference (BME 2017), , Egmond aan Zee, The Netherlands, Jan. 2017.

3. **Bharath, H.N.**, Sauwen, N., Sima, D.M., Himmelreich, U., De Lathauwer, L., and Van Huffel, S. Canonical polyadic decomposition for tissue type differentiation using multi-parametric MRI in high-grade gliomas. 33rd Annual Conference of the European Society of Magnetic Resonance in Medicine and Biology (ESMRMB 2016), Vienna, Austria, Sept. 2016.

4. **Bharath, H.N.**, Sauwen, N., Sima, D.M., Himmelreich, U., and Van Huffel, S. Non-Negative Canonical Polyadic Decomposition Based Brain Tumor Characterization From Multi-Parametric MR Imaging. TDA 2016 workshop on Tensor Decompositions and Applications, Leuven, Belgium, Jan. 2016.

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
STADIUS CENTER FOR DYNAMICAL SYSTEMS, SIGNAL PROCESSING AND DATA ANALYTICS
Kasteelpark Arenberg 10, bus 2446
B-3001 Leuven
bhalandu@esat.kuleuven.be

