

© 2011 Hien Minh Nguyen

TOWARDS HIGH-RESOLUTION MAGNETIC RESONANCE SPECTROSCOPIC IMAGING:
SPATIOTEMPORAL DENOISING AND ECHO-TIME SELECTION

BY

HIEN MINH NGUYEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Associate Professor Minh N. Do, Chair
Professor Yoram Bresler
Professor Zhi-Pei Liang
Assistant Professor Bradley P. Sutton

Abstract

Magnetic resonance spectroscopic imaging (MRSI) enables acquisition of spatial-spectral nuclear spin distributions. Compared to conventional MRI, the additional spectral information provides a powerful tool for *in vivo* study of biological tissues. However, considerable practical challenges, which cannot be overcome using the traditional Fourier method, remain in obtaining spatial-spectral data with both high resolution and high signal-to-noise ratio (SNR). In addition, the choice of echo time (TE) during data acquisition affects the SNR and complexity of the model used to describe measured data. TE selection remains a complicated and controversial issue in proton MR spectroscopy. This dissertation addresses two issues, that is the problem of low SNR and the problem of TE selection.

To address SNR limitations of MRSI, we first investigate a new scheme for denoising MRSI data, incorporating both an anatomically adapted spatial-smoothness constraint and an autoregressive spectral constraint within the penalized maximum-likelihood framework. Theoretical analysis is provided to characterize the denoising performance of this approach. Results demonstrate the ability of the spectral constraint to suppress noise in non-metabolite regions. However, this non-uniform spectral denoising effect may not be visually accepted, limiting the practical use of this denoising approach.

We then further propose to denoise MRSI data by exploiting low-rank properties. These are two low-rank structures of MRSI data, one due to partial separability and the other due to linear predictability of MRSI data. Denoising is performed by arranging the measured data in appropriate matrix forms (i.e., Casorati and Hankel) and applying low-rank approximations by the singular value decomposition. Experimental results obtained from *in vivo* MRSI data demonstrate that

the combination and a particular order of the two proposed low-rank approximations provide an effective way to denoise MRSI data in the case of typical severe noise contamination.

To analyze the problem of TE selection for MRS data acquisition, we reconsider this problem from an estimation theoretic perspective. Specifically, we analyze the Cramér-Rao lower bound (CRB) on estimated spectral parameters as a function of TE, which serves as a metric to quantify the reliability of the estimation procedure. This new approach provides a quantitative method for identifying potentially useful TEs in contrast to a common heuristic choice of TE. In addition, unlike empirical studies which face practical limitations on acquisition time, the CRB analysis enables easy evaluation of an arbitrarily large range of TEs.

To my parents and brother

Acknowledgments

First, I would like to thank Professor Minh Do for giving me the opportunity to do my graduate study here at the University of Illinois at Urbana-Champaign (UIUC), and for patiently supporting my research focus in the field of magnetic resonance imaging through the years. My sincerest appreciation goes to Professor Zhi-Pei Liang for providing not only the very much needed research environment and computing facilities, but also for his constant support, encouragement, and guidance in various aspects of my work. My curiosity about the field started with his lectures, and his enthusiasm always motivated me to continue. I am deeply indebted to him. I would also like to thank Professor Brad Sutton for his detailed support in the earlier stage of my graduate study, during my master thesis. His great experience with pulse sequences was of tremendous help to me. I am grateful to Professor Yoram Bresler for his valuable comments on my dissertation. His work helped me to refine and enhance an important part of my dissertation.

I am thankful to my labmates at the Beckman Institute and Coordinated Science Laboratory — especially Justin Haldar, Rob Morrison, Diego Hernando, Chao Ma, Xi Peng, Cornelius Brinegar, Zhubin Gahvari, Bo Zhao, Fan Lam, Anthony Christodoulou, Xiaobo Qu, Maojing Fu, Professor Du Huiquan, Yue Lu, Dan Kubacki, Quang Nguyen, and Cheng Ouyang — for their friendship and collaborative environment, in which I am happy to be a member. My sincerest appreciation to Justin Haldar for his constant critical but encouraging comments during several joint projects that enhanced my learning curve.

My special appreciation to Duc Ngo for his patient moral support through all the years of my rough PhD journey. I am also grateful to my friends here at UIUC — in particular Lyudmila Kushnir, Tomoki Ozawa, Galina Yakubova, Nam Nguyen, Arely Cortez, and Dong-Ha Oh — for

memorable days and fun activities during my graduate study.

I would also like to acknowledge financial support from the Coordinated Science Laboratory, Beckman Institute, Department of Electrical and Computer Engineering, and Vietnam Education Foundation.

Finally, but most of all, I am grateful to my parents and my brother for their love, understanding, and support, especially when I am far from home. This dissertation is dedicated to them.

Table of Contents

List of Tables	x
List of Figures	xi
List of Abbreviations	xix
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Problem statement	3
1.3 Main results	4
1.4 Organization of the dissertation	5
Chapter 2 Background	7
2.1 Magnetic resonance spectroscopy	7
2.1.1 NMR experiment	8
2.1.2 Chemical shift	11
2.1.3 Metabolite content of <i>in vivo</i> brain ^1H MR spectroscopy	14
2.1.4 Water suppression	17
2.1.5 Baseline suppression	18
2.1.6 MRS quantitation	20
2.1.7 Single-voxel acquisition	23
2.2 Magnetic resonance spectroscopic imaging	25
2.2.1 Principle of MRSI	26
2.2.2 Practical limitations	27
2.2.3 MRSI quantitation	30
2.2.4 Data acquisition	31
2.3 Denoising in MRSI: A review	33
2.3.1 Denoising algorithms for general signal processing applications	34
2.3.2 Existing approaches for denoising of MRSI data	40
Chapter 3 Denoising of MRSI data with spatial-spectral regularization	42
3.1 Introduction	42
3.2 Proposed method	42
3.2.1 Proposed algorithm	44
3.2.2 Characteristics of the proposed algorithm	46

3.3	Discussion	51
3.3.1	Algorithm considerations	51
3.3.2	Estimation of linear prediction coefficients	54
3.3.3	Choice of regularization parameters	56
3.3.4	Single-voxel quantitation evaluation	60
3.4	Experimental results	66
3.4.1	Simulations	66
3.4.2	<i>In vivo</i> experiment	68
3.5	Summary	72
Chapter 4	Denoising of MRSI data with low-rank approximations	74
4.1	Introduction	74
4.2	Proposed method	75
4.2.1	Low-rankness due to spatiotemporal partial separability	75
4.2.2	Low-rankness due to linear predictability	77
4.2.3	Proposed algorithm	78
4.3	Discussion	80
4.3.1	Algorithm considerations	80
4.3.2	Rank selection	84
4.3.3	Denoising performance	87
4.4	Simulation and experimental results	92
4.4.1	Simulations	92
4.4.2	<i>In vivo</i> experiments	95
4.5	Summary	97
Chapter 5	Cramér-Rao bound analysis of echo-time selection	98
5.1	Introduction	98
5.2	Proposed formulation	100
5.2.1	Spectral model	100
5.2.2	Cramér-Rao bound expressions	102
5.3	Analysis of TE selection	103
5.3.1	Coefficient of variation bound	103
5.3.2	Simple case of one metabolite	105
5.3.3	Case of two metabolites	108
5.3.4	Effect of the baseline component	109
5.3.5	Effect of multi-echo data	114
5.3.6	General case of many metabolites	119
5.3.7	Case study	121
5.4	Summary	126
Chapter 6	Conclusions	128
Appendix A	Proof of (4.19)	132
Appendix B	Proof of (4.25)	134

Appendix C	Derivations of the Cramér-Rao bound	135
C.1	CRB expressions for the single-echo case	135
C.2	Asymptotic CRB expressions for the single-echo case	137
C.3	CRB expressions for the multi-echo case	138
References		141

List of Tables

3.1	Performance of the algorithm at each iteration: convergence at the i -th iteration was computed as $\frac{\ \mathbf{s}^{(i)} - \mathbf{s}^{(i-1)}\ _2}{\ \mathbf{s}^{(2)} - \mathbf{s}^{(1)}\ _2}$ and cost function at the i -th iteration was normalized by cost function at the first iteration.	47
3.2	Comparison of estimating linear prediction coefficients using different harmonic retrieval techniques at $\text{SNR}(\boldsymbol{\rho}^{\text{meas}})=17$ dB: mean relative errors (in dBs), computed according to (3.20) and averaged over 1024 noise realizations.	57
4.1	Monte Carlo experiment on choosing effective rank $\bar{L}_{\text{ps}}$ at 9 different SNR levels. Voxel 1 is in the region of low-SNR, while voxel 2 is in the region of high-SNR. All SNR values are reported in decibels.	85
5.1	Parameters for the simulation of lipid components.	114
5.2	Metabolite amplitudes a_n and lineshape parameters d_n used in the case study. Cases when the metabolite is not present are denoted as “NA”.	125
5.3	Case study - resulting CVBs. Cases when the metabolite is not present in the spectrum are denoted as “NA”.	125

List of Figures

2.1	Illustration of the time evolution of the transverse magnetization.	11
2.2	A typical peak with Lorentzian lineshape.	12
2.3	Illustration of the chemical shift: shielding of nuclei from the external magnetic field by electrons surrounding them results in a reduced effective magnetic field. . .	12
2.4	Spectral profiles of common metabolites in the human ^1H brain spectrum.	15
2.5	A schematic diagram of the PRESS sequence.	24
2.6	A schematic diagram of the STEAM sequence.	24
2.7	Comparison of conventional MRI versus MRSI data of the same subject: (a) Anatomical image acquired using spin-echo sequence, TE=40 ms, TR=1500 ms; (b)-(c) NAA and Lac spatial distributions obtained from MRSI data by integrating the complex spectrum at 2.02 and 1.33 ppm, respectively. Total acquisition time for the MRSI data was about 3.5 hours (8 signal averages). Notice that this measured MRSI data have low SNR and low spatial resolution.	29
2.8	A schematic diagram of the conventional CSI pulse sequence.	32
2.9	A schematic diagram of the EPSI pulse sequence.	33
2.10	Typical denoising of the <i>in vivo</i> MRSI data of the mouse brain: NAA map obtained (a) from the measured data, (b) from 32 signal averaging, (c) using a Gaussian smoothing filter, and (d) using 3-D wavelet shrinkage. Notice that signal averaging yielded the best denoising result, however undesirably prolonging acquisition time, while Gaussian filtering has the worst trade-off between the SNR and resolution. In this severe noise contamination case, wavelet shrinkage was not able to preserve spatial features well.	34
2.11	Illustration of nonlinear diffusion using Lorentzian edge-preserving function: (a) edge-preserving function $g(x)$; (b) influence function $\psi(x)$; (c) minimization function $\rho(x)$	37
3.1	Performance of the algorithm at each iteration: top row shows (a) noiseless and (b) noisy NAA spatial distributions; denoised NAA spatial distribution at (c) iteration 1 and (d) iteration 30. Notice that an additional noise reduction is achieved as the algorithm is iterated. Bottom row shows (e) noiseless and (f) noisy spectral distributions; denoised spectral distribution at (g) iteration 1 and (h) iteration 30. Notice that the non-uniform denoising effect in the spectral domain is not alleviated with iterations, as expected.	46

3.2	Illustration of the denoising property of the spectral constraint: (a) noisy spectrum; (b) reconstruction at an insignificant value of $\beta = 0.01$; (c) reconstruction at a moderate value of $\beta = 10$; (d) NRMSEs of the actual and weighted in frequency domain solutions as a function of β ; (e) weighted (in frequency domain) solution at $\beta = 0.01$; (f) weighted (in frequency domain) solution at $\beta = 10$. One can clearly see that the weighting is performed in the transformed domain other than the frequency domain. Noiseless spectrum is displayed in red.	49
3.3	Spatial-spectral response functions $h_n(\mathbf{r}, f)$: top row shows $h_n(\mathbf{r}, f_0)$ of the (a) Fourier reconstruction, (b) spatial denoising, and (c) spatial-spectral denoising; bottom row shows $h_n(\mathbf{r}_0, f)$ of the (d) Fourier reconstruction (which has the same $h_n(\mathbf{r}_0, f)$ as the spatial denoising), (e) spatial-spectral denoising at $f_n = 2$ ppm, and (f) spatial-spectral denoising at $f_n = 7.23$ ppm. Notice that spatial-spectral reconstruction does not introduce a significant additional spatial leakage compared to the spatial reconstruction, but instead introduces some spectral leakage.	51
3.4	Denoising in $\mathbf{k}$ - t domain versus denoising in $\mathbf{r}$ - t domain: top row shows (a) noiseless and (b) noisy spatial distributions $\rho(\mathbf{r}, f_0)$ at $f_0 = 7.8$ ppm; corresponding denoised $\rho(\mathbf{r}, f_0)$ (c) in $\mathbf{k}$ - t and (d) in $\mathbf{r}$ - t ; bottom row shows (e) noiseless and (f) noisy spectral distributions $\rho(\mathbf{r}_0, f)$; corresponding denoised $\rho(\mathbf{r}_0, f)$ in (g) $\mathbf{k}$ - t and (h) $\mathbf{r}$ - t domain.	53
3.5	SSRFs for denoising in $\mathbf{k}$ - t domain versus denoising in $\mathbf{r}$ - t domain: top row shows $h_n(\mathbf{r}, f_0)$ for the (a) Fourier reconstruction, (b) denoising in $\mathbf{k}$ - t , (c) denoising in $\mathbf{r}$ - t ; bottom row shows $h_n(\mathbf{r}_0, f)$ for the (d) Fourier reconstruction, (e) denoising in $\mathbf{k}$ - t , (f) denoising in $\mathbf{r}$ - t . Notice that the main lobe of the SSRF for the denoising in $\mathbf{k}$ - t has a larger spatial extent, resulting in an additional spatial blurring during the filtering process compared to the denoising in $\mathbf{r}$ - t . Here, $\mathbf{r}_0$ was chosen to be the voxel corresponding to the highest intensity in (a) and $f_0 = 2$ ppm. The SSRF $h_n(\mathbf{r}, f)$ was computed such that the corresponding $\mathbf{r}_n$ and f_n match $\mathbf{r}_0$ and f_0 . Figures in each row are shown on the same scale.	54
3.6	Simulated data used in the study on the estimation of linear prediction coefficients: top row shows simulated noiseless spectra in the case of (a) Lorentzian lineshape, (b) Gaussian lineshape, and (c) Gaussian lineshape with a baseline; bottom row shows the corresponding representative noisy spectra.	55
3.7	Weights w_k derived from the linear prediction coefficients, estimated using various techniques: (a) conventional LS, (b) TLS, (c) TLS Cadzow, and (d) IQML methods. Ground truth weights in the noiseless case are shown in red. Notice that the conventional LS estimation did not yield an acceptable performance, while the state-of-the-art techniques improved the estimation.	57
3.8	Single-voxel quantitation from the noisy vs. the denoised data in the case of Lorentzian spectral lineshapes: (a) original spectrum; (b) noisy spectrum; NRMSEs for the (c) amplitudes and (d) lineshape parameters obtained by quantitation from noisy and denoised data; variance for the (e) amplitudes and (f) lineshape parameters obtained by quantitation from noisy and denoised data.	61

3.9	Single-voxel quantitation from the noisy vs. the denoised data as a function of β in the case of Lorentzian spectral lineshapes: (a) NRMSE and (b) variance for the estimated area of peak 1; (c) NRMSE and (d) variance for the estimated area of peak 4. As β increases, the proposed denoising achieves smaller NRMSEs and variances of estimated metabolite areas.	62
3.10	Single-voxel quantitation from the noisy vs. the denoised data in the case of non-Lorentzian spectral lineshapes: (a) original spectrum; (b) noisy spectrum; NRMSEs for the (c) amplitudes and (d) lineshape parameters obtained by quantitation from the noisy and denoised data; variance for the (e) amplitudes and (f) lineshape parameters obtained by quantitation from the noisy and denoised data.	63
3.11	Single-voxel quantitation from the noisy vs. denoised data as a function of β in the case of non-Lorentzian spectral lineshapes: (a) NRMSE and (b) variance for the estimated area of peak 1; (c) NRMSE and (d) variance for the estimated area of peak 4. For a range of medium values of β , the proposed denoising reduces NRMSEs of metabolite areas compared to the quantitation directly from the noisy data. This improvement is more significant for a lower-SNR peak (peak 4). Notice that variance of the estimated area for each peak decreases as β increases.	64
3.12	Performance of a single-voxel quantitation from the noisy vs. denoised data in the case of overlapping Gaussian spectral lineshapes: (a) original spectrum; (b) noisy spectrum; NRMSE for the estimated (c) amplitudes and (d) lineshape parameters obtained by quantitation from noisy (red) and denoised (black) data; variance for the estimated (e) amplitudes and (f) lineshape parameters obtained by quantitation from noisy (red) and denoised (black) data.	65
3.13	Denoising results - spatial distributions $\rho(\mathbf{r}, f_0)$ at a low-SNR frequency in the metabolite-baseline region (4.86 ppm): $\rho(\mathbf{r}, f_0)$ obtained from (a) noisy data; (b) proposed denoising; (c) wavelet shrinkage; (d) noiseless data; (e) corresponding error of (b); (f) corresponding error of (c). Error maps in (e) and (f) were scaled by a factor of 0.4 with respect to the figures in (a)-(d). Wavelet denoising and the proposed method have approximately the same data consistency level for the reconstructed 3-D distribution.	67
3.14	Denoising results - spatial distributions $\rho(\mathbf{r}, f_0)$ at a low-SNR frequency at 7.8 ppm. $\rho(\mathbf{r}, f_0)$ obtained from (a) noisy data; (b) proposed denoising; (c) wavelet shrinkage; (d) noiseless data; (e) corresponding error of (b); (f) corresponding error of (c). Error maps in (e) and (f) were scaled by a factor of 0.4 with respect to the figures in (a)-(d). Wavelet denoising and the proposed method have approximately the same data consistency level for the reconstructed 3-D distribution.	68
3.15	Denoising results - spectra from a particular voxel, marked by an arrow in Fig. 3.13(d): (a) noisy spectrum; (b) spectrum obtained from the proposed denoising; (c) spectrum obtained from wavelet shrinkage; (d) denoising weights w_k for (b); (e) error of (b); (f) error of (c). Original spectrum is displayed in red.	69

3.16	<i>In vivo</i> experiment: (a) reference image; denoising weights w_k for (b) voxel 1 in the region of electrocoagulation and (c) voxel 2, outside of the electrocoagulation region. Denoising weights show that the AR constraint can accurately identify the presence or absence of the dominant metabolite peaks (NAA, Cr, Lac, and residual water peak). As expected, Lac is present at voxel 1, while it is absent at voxel 2.	70
3.17	Denoising results from <i>in vivo</i> experiment: top row shows NAA spatial distributions obtained from (a) noisy data; (b) spatial denoising; (c) proposed denoising; (d) wavelet denoising; bottom row shows Lac spatial distributions obtained from (e) noisy data; (f) spatial denoising; (g) proposed denoising; (h) wavelet denoising. Wavelet denoising and proposed method have approximately the same data consistency level.	71
3.18	Denoising results from <i>in vivo</i> experiment: top row shows spectra at voxel 2 obtained from (a) noisy data, (b) spatial denoising, (c) proposed denoising, and (d) wavelet denoising; bottom row shows spectra at voxel 1 obtained from (e) noisy data, (f) spatial denoising, (g) proposed denoising, and (h) wavelet denoising. Voxels 1 and 2 are indicated by arrows in Fig. 3.16(a). All spectra are shown in absolute mode and on the same scale.	72
4.1	Convergent solution of the Cadzow denoising approach: (a) singular values of $\mathbf{H}$ (first 100 values are shown only); (b) variance of elements along the main anti-diagonal of $\bar{\mathbf{H}}$; (c) error of the denoised signal and its Lorentzian fit. This experiment confirms that the convergent solution has a corresponding (a) low-rank matrix $\mathbf{H}$, (b) Hankel matrix $\bar{\mathbf{H}}$, and (c) Lorentzian lineshape.	81
4.2	Simulation study of low-rank filtering based on linear predictability: (a) original (red) and noisy (black) spectra at low and high noise levels; (b)-(e) denoised spectra from Cadzow and LORA, each with high ($L_{lp} = 20$) or low ($L_{lp} = 8$) model order.	82
4.3	Monte Carlo study of the bias introduced at each iteration by the Cadzow method. The corresponding simulated spectra are shown in Fig. 4.2(a), bottom row. The model order L_{lp} was 8. Notice that in this study the bias increases with iterations.	84
4.4	Monte Carlo study of rank L_{lp} selection based on AIC: mean AIC value (averaged over 256 noise realizations) as a function of $\hat{L}_{lp}$ at (a) low noise level with $\text{SNR}_e=26.61$ dB, $\text{SNR}_p=94.4$; (b) medium noise level with $\text{SNR}_e=17.03$ dB, $\text{SNR}_p=30.4$; (c) high noise level with $\text{SNR}_e=12.76$ dB, $\text{SNR}_p=17.9$. The true rank L_{lp} in this experiment was 20.	86
4.5	Noise reduction factor g for low-rank filtering based on partial separability as a function of the estimated rank $\hat{L}_{ps}$. Empirical values of g (black) and corresponding theoretical approximations (red) were computed according to (4.22) and (4.25), respectively. The true rank $L_{ps} = 8$. Notice that in the region $\hat{L}_{ps} > L_{ps}$ the theoretically predicted values of g tend to its empirical values and both approach 1 in the full-rank case.	88

4.6	Monte Carlo study of low-rank filtering based on partial separability: (a) representative noisy and denoised spectra at three SNR levels; (b) error distributions calculated as the histograms of e and $\bar{e}$ according to (4.27) at a particular location $(\mathbf{r}_0, f_0)$. For this illustrative example, $\mathbf{r}_0$ is a representative voxel in the white matter region marked as “2” in Fig. 4.9(e) and f_0 is the frequency point marked in Fig. 4.6(a). All spectra are shown in absolute mode. Notice the reduced noise variances achieved by denoising.	89
4.7	Monte Carlo study of low-rank filtering based on linear predictability: (a) representative noisy and denoised spectra at three noise levels; (b) error distributions calculated as the histograms of e and $\bar{e}$ according to (4.27) at a particular location $(\mathbf{r}_0, f_1)$. For this illustrative example, $\mathbf{r}_0$ is a representative voxel in the white matter region marked as “2” in Fig. 4.9(e) and f_1 is the frequency point marked in Fig. 4.7(a).	90
4.8	Monte Carlo study of low-rank filtering based on linear predictability (continuation of Fig. 4.7): (a),(b) error distributions calculated as the histograms of e and $\bar{e}$ according to (4.27) at a particular location $(\mathbf{r}_0, f_1)$, $(\mathbf{r}_0, f_2)$, and $(\mathbf{r}_0, f_3)$, respectively. For this illustrative example, $\mathbf{r}_0$ is a representative voxel in the white matter region marked as “2” in Fig. 4.9(e) and f_2, f_3 are the frequency points marked in Fig. 4.7(a). Representative noisy and denoised spectra at three noise levels are shown in Fig. 4.7(a).	91
4.9	Denoising results - spatial distributions $\rho(\mathbf{r}, f_0)$ at a low-SNR frequency (6.66 ppm, see Fig. 4.10(e)), obtained from (a) noisy data; (b) LORA denoising; (c) wavelet denoising; (d) Gaussian apodization; and (e) noiseless data. Corresponding errors of (b)-(d) are shown in (f)-(h), respectively. Parameters for both wavelet denoising and Gaussian apodization (wavelet shrinkage threshold and apodization constant, respectively) were chosen to yield approximately the same residual noise variance level as that of the proposed method.	93
4.10	Denoising results - spectra at a particular voxel, marked as “1” in Fig. 4.9(e), obtained from (a) noisy data; (b) LORA denoising; (c) wavelet shrinkage; (d) Gaussian apodization; (e) noiseless data. Corresponding errors of (b)-(d) are shown in (f)-(h), respectively. All spectra are shown in the real-valued mode. Parameters for both wavelet denoising and Gaussian apodization (wavelet shrinkage threshold and apodization constant, respectively) were chosen to yield approximately the same residual noise variance level as that of the proposed method.	94
4.11	Metabolite quantitation study from noisy and denoised data: histograms of the normalized root mean-squared error of the estimated metabolite amplitude (NRMSE_c) for (a) NAA, (b) Cr, and (c) Glu in GM region and (d) NAA, (e) Cr, and (f) Glu in WM region. NRMSE_c was computed according to (4.30). Notice that the proposed denoising helped to improve the accuracy of metabolite quantitation.	95
4.12	Denoising results from <i>in vivo</i> experiment: NAA (top row) and Lac (bottom row) spatial distributions obtained from (a) noisy data; (b) proposed method.	96

4.13	Denoising results from <i>in vivo</i> experiment: spectra at voxels in the region of electrocoagulation (voxel 1, top row) and outside of electrocoagulation area (voxel 2, bottom row), obtained from (a) noisy data; (b) proposed method. All spectra are shown in absolute mode and on the same scale.	97
5.1	A schematic diagram of the typical spin-echo CSI data acquisition: at a longer TE the signal intensity is decreased exponentially, resulting in a lower SNR.	99
5.2	Illustration of the performance of metabolite quantitation: (a)-(c) spectra containing resonances from NAA, Cr, Cho, Glu, Glx, mIns, and a simulated baseline at three SNR levels which correspond to the smallest, medium, and largest SNR levels in (d)-(f); (d)-(f) CVs, obtained from metabolite quantitation, versus CVBs as a function of the noise standard deviation. In all cases of different metabolites, the CVBs were lower than the CVs approximately by a factor of 2.	106
5.3	CRB analysis of TE selection in the reduced case of one metabolite: (a) spectra of NAA at TE=20 ms (red) and TE=140 ms (black); (b) representative noisy NAA spectrum; (c) CVB (black) and quantified CV (red) on the amplitude of NAA; (d) spectra of Lac at TE=20 ms (red) and TE=140 ms (black), shown in the same scale as (a); (e) representative noisy Lac spectrum; (f) CVB (black) and quantified CV (red) on the amplitude of Lac. Metabolite quantitation was performed with 1024 noise realizations. The non-monotonic behavior of the CVB curve of Lac compared to NAA can be explained by the fact that the spectrum of Lac experiences phase inversion significantly in the region of 1.3 ppm, while the spectrum of NAA does not, except for a decrease in amplitude.	107
5.4	Effect of the spectral overlap on the CVB: spectra of (a) Glu and (b) Gln at TE=5 ms, both on the same scale. Spectral overlap is observed in the regions from 1.8 ppm to 2.6 ppm and from 3.5 ppm to 4 ppm; representative noisy spectra of (c) Glu and (d) Gln; (e) CVB on the amplitude of Glu in the presence (in black) and absence (in red) of Gln; (f) CVB on the amplitude of Gln in the presence (in black) and absence (in red) of Glu. (g) CV for Glu in the presence (in black) and absence (in red) of Gln; (h) CV for Gln in the presence (in black) and absence (in red) of Glu. In both Glu and Gln cases, the CVB and CV of one metabolite is not significantly affected by the presence of the other metabolite. This is not expected if judging intuitively from the observed spectral overlap.	109
5.5	Monte Carlo study of the effect of the baseline amplitude on the CRB: top row shows representative noisy spectra containing NAA, Cr, Cho, Glx, m-Ins, and a simulated baseline at different scaling factors for the baseline amplitude; bottom row shows the CVBs on the amplitudes of (d) NAA, (e) Cr, and (f) Cho and the corresponding CVs after performing metabolite quantitation with 1024 noise realizations. Notice that the CVBs did not change as the baseline amplitude was scaled. Hence, we suggest that baseline amplitude should not affect the choice of TE from the CRB perspective.	112

5.6	Monte Carlo study of the effect of the baseline smoothness on the choice of TE from the CRB perspective: (a)-(e) representative noisy spectra containing a simulated baseline and a metabolite peak, which is gradually displaced across the whole frequency range; (f) CVB and quantified CV value for each case of the peak displacement. Quantitation was performed with 1024 noise realizations. Notice that as the peak moves closer to the dominant resonance range of the baseline (2 ppm to 6 ppm), CVB increases and changes according to the change in the baseline smoothness.	113
5.7	Baseline results in an increase in the CVB by a small factor: spectrum of the simulated baseline component originating from (a) macromolecules (MM) and (b) lipids (Lip); (c) spectrum containing simulated resonances of metabolites, MM, and Lip; (d)-(f) ratio of the CVB on the amplitudes of NAA, Cr, Cho, Glu, Gln, and mIns, respectively, in the case when the baseline is present versus the case when the baseline is absent.	115
5.8	Baseline leads to an increase in the CVB by a small factor (continuation of Fig. 5.7): (a)-(c) ratio of the CVB on the amplitudes of Glu, Gln, and mIns, respectively, in the case when the baseline is present versus the case when the baseline is absent.	116
5.9	Multi-echo acquisition at echo times $\{TE_i\}_{i=1}^D = 30$ ms: (a) Representative noisy spectrum, containing resonances from NAA, Cr, Cho, Glu, Gln, mIns, Lac, and a baseline; (b)-(f) CVBs and quantified CVs on the amplitudes of NAA, Cr, Cho, Glu, and Gln, respectively, as a function of D . As expected, the CVB curve decreases at the rate $1/\sqrt{D}$ for all metabolites.	119
5.10	Effect of multi-echo data on the CVB: metabolite spectral profiles in the case of (a) a single TE; (b) double TEs with $TE_1 = TE_2$; (c) double TEs with $TE_1 \neq TE_2$ and the spectral overlap (inner product) between two metabolites at TE_2 is reduced. CVB plots of the amplitudes of (d) metabolite 1 and (e) metabolite 2 show that CVB is reduced at least by a factor of $\sqrt{2}$ for the double-echo acquisition. An additional (although small) reduction factor is obtained if one chooses TE_2 so that the spectral overlap between two metabolites is reduced, as is the case in (c). . . .	120
5.11	CVB on the amplitude of (a) NAA, (b) Cho, and (c) Lac in the case of regular (17 metabolites and baseline) and reduced (7 and 4 metabolites) models. The CVB curve decreases as the model order decreases. For clarity, this figure should be viewed in color.	121
5.12	Illustration of water suppression using the HSVD method: (a) Acquired spectrum with water suppression frequency-selective pulse resulting in a large residual water peak; (b) water suppressed spectrum after performing HSVD. Notice that metabolite resonances can be clearly observed in (b).	123
5.13	<i>In vivo</i> MRS spectra at (a) TE=30 ms; (b) TE=144 ms; (c) TE=288 ms. Baseline extraction was performed on the spectrum acquired at TE=30 ms: (d) spectrum after removing 30 initial FID points; (e) back-extrapolated spectrum, containing metabolite components; (f) extracted baseline component after subtracting (e) from (a).	124

5.14	Glu and mIns resonances overlap with the baseline completely, while the spectral overlap of Gln with the baseline is reduced due to the presence of two resonance peaks of Gln at 6.9 ppm and 7.62 ppm.	126
------	---	-----

List of Abbreviations

AIC	Akaike Information Criteria
AMARES	Advanced Method for Accurate, Robust and Efficient Spectral Fitting
AQSES	Automated Quantitation of Short Echo Time Spectra
AQSES-MRSI	Automated Quantitation of Short Echo Time Spectra - Magnetic Resonance Spectroscopic Imaging
AR	AutoRegressive
CHESS	Chemical Shift Selective
Cho	Choline
Cr	Creatine
CRB	Cramér-Rao Bound
CSF	CerebroSpinal Fluid
CSI	Chemical Shift Imaging
CV	Coefficient of Variation
CVB	Coefficient of Variation Bound
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EPSI	Echo-Planar Spectroscopic Imaging
Eth	Ethanolamine
FID	Free Induction Decay
FOV	Field of View
GAVA	Gamma Visual Analysis

Glc	Glucose
Gln	Glutamine
Glu	Glutamate
Glx	Glutamine/Glutamate
GM	Gray Matter
His	Histidine
HLSVD	Hankel Lanczos Singular Value Decomposition
HR	Harmonic Retrieval
HSVD	Hankel Singular Value Decomposition
HTLS	Hankel Total Least Squares
HTLS-PK	Hankel Total Least Squares using Prior Knowledge
ICA	Independent Component Analysis
IQML	Iterative Quadratic Maximum Likelihood
KNOB-SVD	Knowledge-Based Singular Value Decomposition
KNOB-TLS	Knowledge-Based Total Least Squares
Lac	Lactate
LCModel	Linear Combination of Model
Lip	Lipid
LORA	Low-Rank Approximation
LPSVD	Linear Prediction and Singular Value Decomposition
LPTLS	Linear Prediction and Total Least Squares
mIns	Myo-Inositol
ML	Maximum Likelihood
mm	Millimeter
mM	Millimole
MM	Macromolecule
MR	Magnetic Resonance

MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
ms	Millisecond
MSE	Mean Squared Error
NAA	N-Acetylaspartate
NMF	Nonnegative Matrix Factorization
NMR	Nuclear Magnetic Resonance
NRMSE	Normalized Root Mean Squared Error
OVS	Outer Volume Suppression
PCA	Principal Component Analysis
PDE	Partial Differential Equation
PEEP	Phase-Encoded Echo-Planar
ppm	Parts Per Million
PREP	Projection-Reconstruction Echo-Planar
PRESS	Point Resolved Spectroscopy
PSF	Partially Separable Function
RF	RadioFrequency
Scy	Scyllo-inositol
SISSI	Single-Shot Spectroscopic Imaging
SNAP	Snapshot Fast Low-Angle Shot
SNR	Signal-to-Noise Ratio
SPLASH	Spectroscopic Fast Low-Angle Shot
SSRF	Spatial-Spectral Response Function
STEAM	Stimulated Echo Acquisition Mode
SVD	Singular Value Decomposition
Tau	Taurine

TE	Echo Time
Thr	Threonine
TLS	Total Least Squares
TMS	Tetramethylsilane
TR	Repetition Time
TV	Total Variation
VARPRO	Variable Projection
WM	White Matter

Chapter 1

Introduction

1.1 Motivation

Magnetic resonance spectroscopic imaging (MRSI) powerfully extends conventional MRI by providing an additional dimension of spectral information. This information, which has been shown to be useful in various applications such as chemical analysis, breast cancer imaging, and imaging of neuro-degenerative disorders, is extracted by treating nuclear spins in a more special and careful way than conventional MRI. This exciting field originated perhaps in 1939, when I. Rabi discovered the nuclear magnetic resonance (NMR) phenomenon in molecular rays [1]. In 1946 the first observation of NMR signals in bulk matter was made independently by two research groups, that of F. Bloch, W. W. Hansen and M. E. Packard and that of E. M. Purcell, H. C. Torrey and R. V. Pound [2]. Since then, the potential use of NMR for chemical analysis of *in vivo* biological tissue and later on for non-invasive metabolite imaging was soon recognized. Unlike other spectroscopy modalities such as ultra-violet spectroscopy or infra-red spectroscopy, NMR spectroscopy does not require long time sample preparation, is not limited by the sample depth penetration, operates in the radio-frequency range (therefore, there are no associated potential harmful effects), and can provide both qualitative and quantitative information. As a result, numerous studies of *in vivo* NMR spectroscopy (usually referred to as MRS) appeared, and its combination with MRI spatial discrimination methods resulted in the medical imaging technique of MRSI. MRSI has rapidly evolved into a complex technique that can be used to monitor metabolic processes [3] or to investigate many neurological disorders such as brain tumor, degenerative diseases such as Alzheimer's and Parkinson's diseases, cerebrovascular diseases, and metabolic disorders such as epilepsy and

multiple sclerosis [4].

The main limitations of MRSI are the low signal sensitivity and long acquisition times, which lead to typical low SNR and low resolution (usually on the order of 1 cm in contrast to 1 mm for anatomical imaging) MRSI data. Low sensitivity results from two factors: (i) the NMR phenomenon itself is relatively insensitive due to the very small magnetization difference between energy levels of a nucleus possessing angular momentum; (ii) metabolite concentrations are low (typically thousands-fold below that of tissue water, which provides the strong signal used for conventional MRI measurements [5, 6]). Long acquisition time (typically on the order of tens of minutes to hours, as compared to milliseconds to seconds for conventional anatomical imaging) has been a typical problem in MRSI due to the need to encode an additional spectral dimension together with conventional spatial dimensions.

In addition, unlike conventional MRI, MRSI faces other issues such as strong nuisance signals (e.g., water, macromolecule, and lipid resonances). Water resonance is usually eliminated from the spectrum using specialized pulse sequences or during a post-processing stage, while macromolecule and lipid resonances can be substantially eliminated by acquiring data at a long echo time (TE), due to short T_2 decay of these components. However, this in turn decreases SNR and leads to the loss of the observable signals from metabolites with short T_2 parameters such as glutamine, glutamate, and myo-inositol, which are important in metabolite studies [7]. Thus, the choice of TE has been a topic of numerous studies and remains a complicated issue in ^1H MRS.

As a result, the ability to improve the conventional Fourier reconstruction scheme to achieve high-resolution and high-SNR spatial-spectral information, combined with robust data acquisition and metabolite quantitation procedure is tremendously desired to enhance MRSI practical utility for clinical applications.

1.2 Problem statement

The acquired MRSI signal is conventionally modeled as

$$\mathbf{s}(\mathbf{k}, t) = \int \int \rho(\mathbf{r}, f) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}} e^{-i2\pi ft} d\mathbf{r} df + \xi(\mathbf{k}, t), \quad (1.1)$$

where $\rho(\mathbf{r}, f)$ denotes the desired spatial-spectral function and $\xi(\mathbf{k}, t)$ is the measurement noise often modeled as a complex white Gaussian process. The function $\rho(\mathbf{r}, f)$ provides valuable information on the spatial-spectral distribution of metabolites, and is useful for noninvasive metabolite imaging of living systems. For example, ^{13}C MRSI can be used to study glucose metabolism [2]; ^{31}P MRSI is capable of detecting metabolites participating in tissue energy metabolism [2]; ^1H MRSI can map out the spatial distributions of N-acetylaspartate, creatine, choline, and lactate metabolites that are useful for the investigation of neurological disorders [4]. However, considerable practical challenges remain in obtaining $\rho(\mathbf{r}, f)$ in both high spatial-spectral resolution and high SNR. In addition, the observed function $\rho(\mathbf{r}, f)$ is also dependent on the chosen TE. In particular, shorter TE spectra generally have higher SNR and contain significant contributions from a larger number of metabolites than long TE spectra; however, this advantage can be offset by the more complicated signal model that must be used to describe short TE data.

The goal of this research, thus, is to address the following issues:

1. Address the problem of low SNR associated with MRSI data. In particular, the denoising approach should exploit the spatial-spectral characteristics inherent in the MRS signal in a “soft” way. Since the ultimate goal of MRSI is metabolite quantitation, the question whether performing quantitation from the denoised data improves metabolite quantitation accuracy compared to that of the noisy data should be also addressed.
2. Provide a quantitative analysis of the effect of different TE values on the accuracy of metabolite quantitation in order to guide the design of clinical protocols.

1.3 Main results

- We have proposed and characterized a new approach to address the problem of low SNR in MRSI. The method formulates the denoising problem using the penalized maximum-likelihood framework, in which spatial constraints are imposed using a weighted spatial-smoothness regularization and spectral assumptions are incorporated with an autoregressive constraint. The filtering effects of the method were characterized. The method can improve quantitation accuracy if the spectral regularization term is properly designed for particular applications. Simulation results demonstrate the ability of the spectral constraint to suppress noise in non-metabolite regions, which gives rise to a non-uniform denoising effect spectrally. Related publication is Ref. [8].
- We have proposed and validated another denoising method that is more effective than the first approach in terms of not resulting in a non-uniform spectral denoising effect. The method exploits two low-rank properties inherent in MRSI data, one due to partial separability of the spatial-temporal data, and the other due to linear predictability of MRS signal. A singular value decomposition (SVD)-based algorithm is proposed to utilize these low-rank properties. Experimental validations demonstrate that the combination of these two low-rank properties of MRSI data provides an effective and flexible method which could prove useful for denoising MRSI data and other spatial-spectral and spatial-temporal imaging data as well. Related publications include Refs. [9–11].
- We provided a quantitative analysis of the effect of different TE values on the accuracy of metabolite quantitation. Specifically, based on the estimation theory, we computed the Cramér-Rao lower bound (CRB) on the variance of the estimated metabolite amplitudes as a function of TE. While retrospective use of the CRB is common in MR spectroscopy as a metric of quality for an acquired experimental dataset [12–14], we used the CRB prospectively to guide the design of the protocol. The CRB analysis showed that the coupling among spectral components and the existence of baseline signals at short TE that are usually con-

sidered as significant obstacles for metabolite quantitation, in fact are not problematic if the metabolite spectral profiles are known and the baseline can be accurately represented using some *a priori* chosen basis functions. We analyzed the effect of baseline signals and multi-echo data on the performance of metabolite quantitation in several specific scenarios. All observations were made based on the CRBs and were also supported by the results from an actual metabolite quantitation. A case study was also provided with parameters simulated based on the experimental data to demonstrate the use of the proposed CRB framework as a tool for choosing optimal TE. This CRB analysis provides a quantitative method for identifying potentially useful TEs from an arbitrarily large range of candidate values, in contrast to a common heuristic choice of TE. Related publication is Ref. [15].

1.4 Organization of the dissertation

The outline of the rest of the dissertation is as follows:

Chapter 2 presents background material that will be helpful for understanding the subsequent chapters of the dissertation. It contains a high-level overview of NMR physics, including important NMR chemical shift phenomena and basic signal generation. This chapter also reviews conventional MRS data acquisition and processing schemes (e.g., water and baseline suppression) and briefly describes information content of ^1H MR spectrum, which is the focus of this research. In addition, practical limitations of MRSI and established MRSI quantitation algorithms are discussed. In the last part of this chapter, we review denoising approaches for general signal processing applications as well as existing denoising methods proposed specifically for MRSI.

Chapters 3 and 4 present two new denoising approaches. The first approach combines a spectral autoregressive constraint with a weighted spatial-smoothness regularization, while the second approach exploits two low-rank properties of MRSI data, one due to partial separability and the other due to linear predictability of MRSI data. Both formulations are described and experimental results are provided to illustrate the denoising performance of these approaches.

Chapter 5 presents a quantitative analysis of the choice of TE for MRS acquisitions. The chapter analyzes the CRB on the variances of the estimated spectral parameters as a function of TE, which serves as a metric to quantify the reliability of the estimation procedure. The general case of ^1H MR-observable metabolites and their coupling interactions is considered, although a similar analysis can be easily expanded to the other forms of MRS.

Finally, Chapter 6 is devoted to conclusions and the Appendix contains proofs of several remarks presented in this dissertation.

Chapter 2

Background

Compared with conventional MRI, MR spectroscopy (MRS) and MR spectroscopic imaging (MRSI) aim to utilize the full potential of the MR signal by exploiting an additional dimension of spectral information. Although benefits of MRS and MRSI are tremendous, they face many obstacles that typically either do not exist or are much less prominent in conventional MRI (undesirable dominant water resonance, resonance from macromolecules and lipids, low spatial resolution, low SNR, long acquisition time, etc.). In order to overcome these obstacles, a careful treatment of nuclear spins ranging from the data acquisition design to the signal reconstruction that exploits signal properties inherent in MRS data must be applied. In this chapter, we present a high-level review of the basic NMR signal generation, conventional data processing, and data acquisition methods used in MRS and MRSI. Practical limitations of MRSI are also described. The last section of this chapter is devoted to reviewing existing methods addressing the problem of low SNR in MRSI - a topic of the two subsequent chapters in this dissertation.

2.1 Magnetic resonance spectroscopy

Two fundamental NMR phenomena that form a basis for the signal generation in MRS are *chemical shift* and *spin coupling*. Specifically, chemical shift is a phenomenon that results in the observable range of different molecule resonances, while spin coupling is an observable feature in high-resolution NMR spectra that leads to multiple resonances of an individual metabolite. A general description of chemical shift is presented in this section, while spin coupling is only briefly mentioned. We then review the metabolite information content of a typical ^1H MR spectrum of the

brain, which is of interest in this dissertation. Conventional signal localization techniques and routine MRS data pre-processing such as water suppression and baseline removal are also described.

2.1.1 NMR experiment

Elementary particles having periodical and orbital motion (such as protons, neutrons, electrons) can be considered as rotating objects carrying an electrical charge, thus creating a current loop. This current generates a magnetic field, which is characterized by the magnetic dipole moment $\boldsymbol{\mu}$, proportional to the angular momentum of the particle:

$$\boldsymbol{\mu} = \gamma \mathbf{L}, \quad (2.1)$$

where γ is the gyromagnetic ratio. The fundamental postulate in quantum mechanics states that the angular momentum of elementary particles is limited to discrete values. That is, its amplitude in the z direction is given by

$$L_z = \left(\frac{h}{2\pi} \right) m \quad (2.2)$$

with

$$m = I, I - 1, I - 2, \dots, -I, \quad (2.3)$$

where h is Planck's constant and I is the spin quantum number, being equal to $1/2$ for ^1H nuclei. As a consequence, the magnetic moment is also quantized.

When placed in an external magnetic field $\mathbf{B}_0$, the current loop will experience a torque which results in the precession of $\boldsymbol{\mu}$ about $\mathbf{B}_0$, described as

$$\frac{d\boldsymbol{\mu}}{dt} = \gamma \boldsymbol{\mu} \times \mathbf{B}_0. \quad (2.4)$$

In addition, the particle also acquires a magnetic energy

$$\begin{aligned} E &= -\boldsymbol{\mu} \cdot \mathbf{B}_0, \\ &= -\gamma \left(\frac{h}{2\pi} \right) m B_0. \end{aligned} \quad (2.5)$$

Since m can receive only two discrete values ($m = \pm \frac{1}{2}$) for spin-half particles (such as ^1H nuclei), magnetic energy of the particle is split into two energy levels, with the energy difference

$$\Delta E = \gamma \left(\frac{h}{2\pi} \right) B_0. \quad (2.6)$$

A coherent transition of spins from one energy level to another results in the resonance phenomenon in NMR. One achieves this by applying an oscillating magnetic field in the transverse xy plane with a frequency ω_{RF} so that the radiation energy $h\omega_{\text{RF}}/2\pi$ equals to the magnetic energy given by (2.6). It then follows that

$$\omega_{\text{RF}} = \gamma B_0, \quad (2.7)$$

also referred to as the *on-resonance condition*.

The population difference between the two spin energy states (α and β states) is given by

$$\frac{N_\alpha}{N_\beta} = 1 + \frac{\gamma h B_0}{kT} \quad (2.8)$$

and is usually very small, demonstrating that NMR is a low sensitive technique. As an example, a macroscopic sample with 10^6 nuclear spins at 37°C , 9.4 T magnetic field corresponding to 400 MHz has the population difference between α and β states being only 31 spins [2].

Total net magnetic moment or magnetization $\mathbf{M}$ of a macroscopic sample is the sum of all individual magnetic moments $\boldsymbol{\mu}_i$. Since $\boldsymbol{\mu}_i$ are randomly distributed, at the thermal equilibrium there is no net component of $\mathbf{M}$ in the xy plane (or no phase coherence), while the component M_z is non-zero, but it is static. In order to observe nuclear magnetization, the longitudinal magnetiza-

tion M_z must be rotated onto transverse plane. This is achieved by applying a B_1 field, oscillating in transverse plane in RF range. With the application of this field, the resulting transverse magnetization M_{xy} is non-zero and time-decaying due to T_2 relaxation and field inhomogeneity:

$$\begin{aligned} M_{xy}(t) &= M_{xy}(0_+)e^{-t/T_2^*}e^{-i\omega_0 t}, \\ M_z(t) &= M_z^0(1 - e^{-t/T_1}) + M_z(0_+)e^{-t/T_1} \end{aligned} \quad (2.9)$$

where M_z^0 denotes the initial longitudinal magnetization while $M_{xy}(0_+)$ and $M_z(0_+)$ are the transverse and longitudinal magnetization after applying B_1 field.

Time evolution of $\mathbf{M}$ described in (2.9) is illustrated in Fig. 2.1. The transverse magnetization M_{xy} induces an electromotive force (emf) in a receiving coil positioned in the transverse plane. The time-dependence of emf is called the *free induction decay* (FID) [2]. In the simple case of a spin system with one spectral component resonating at frequency ω_0 and ignoring field inhomogeneity, the observed FID can be expressed as

$$s(t) = M_0 e^{-t/T_2^*} e^{-i\omega_0 t}, \quad (2.10)$$

assuming a complete excitation with 90° RF pulse. The corresponding Fourier transform of $s(t)$ is

$$S(\omega) = \int_{-\infty}^{+\infty} s(t) e^{i\omega t} dt \quad (2.11)$$

$$= \frac{M_0 T_2^*}{1 + (\omega + \omega_0)^2 T_2^{*2}} + i \frac{M_0 (\omega + \omega_0) T_2^{*2}}{1 + (\omega + \omega_0)^2 T_2^{*2}}, \quad (2.12)$$

which is a Lorentzian function with real component also referred to as the *absorption* component and imaginary component referred to as the *dispersion* component, both illustrated in Fig. 2.2. Notice that the absorption component has width at half height equal to $\frac{1}{\pi T_2^*}$, while the dispersion component is broader. Thus, for the best separation of the peaks in an NMR spectrum, absorption mode spectra are typically desired.

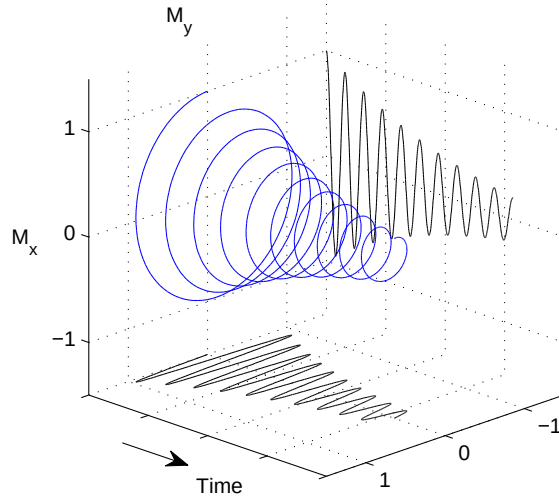


Figure 2.1: Illustration of the time evolution of the transverse magnetization.

2.1.2 Chemical shift

According to the on-resonance condition in (2.7), nuclei of the same isotope even in different molecules resonate at the same frequency because of their identical gyromagnetic ratio. This is undesirable for NMR spectroscopy, since we would not be able to observe and decouple resonances in different molecular environments. Fortunately, resonance frequency ω not only depends on the gyromagnetic ratio γ and the external magnetic field B_0 , but is also sensitive to the chemical environment of the nucleus of interest. This is commonly referred to as the *chemical shift*. This phenomenon is caused by the shielding of nuclei from the external magnetic field by electrons surrounding them, also referred to as electronic shielding. As shown in Fig. 2.3, the electrons surrounding a nucleus can be regarded as small currents, giving rise to a magnetic moment μ_e at the nucleus. Since in diamagnetic molecules the magnetic moment opposes the external magnetic field, the effective magnetic field at the nucleus is reduced as

$$B = B_0(1 - \sigma), \quad (2.13)$$

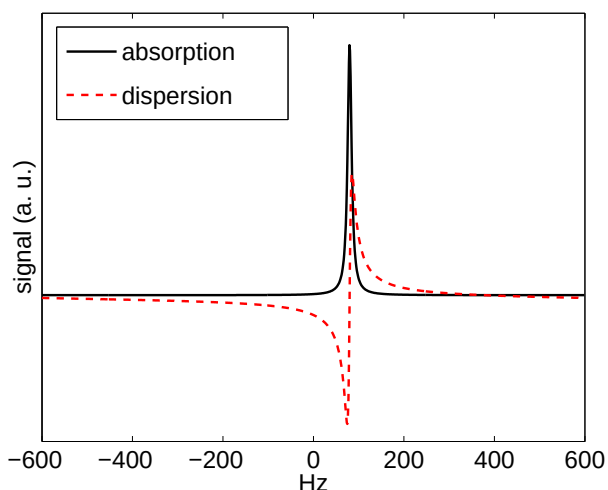


Figure 2.2: A typical peak with Lorentzian lineshape.

where σ denotes the *shielding constant*, which is a dimensionless number, normally expressed in parts per million (ppm). Shielding constant is on the order of 10^{-5} for protons and depends on the chemical environment of the nucleus, having larger values for heavy atoms, since the shielding increases with the number of electrons. Notice that σ is determined solely by the electronic and magnetic environment of the nuclei being observed, and hence, it does not depend on the magnetic field.

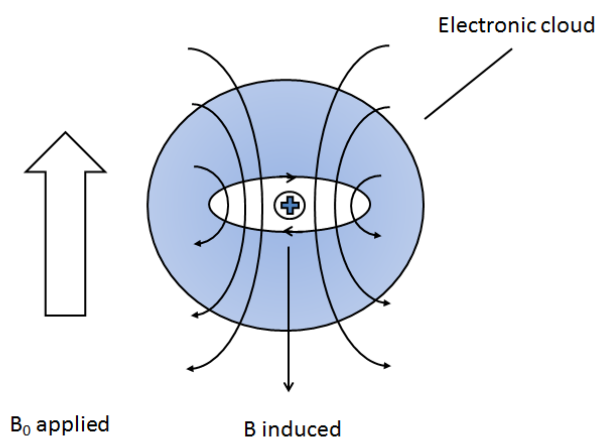


Figure 2.3: Illustration of the chemical shift: shielding of nuclei from the external magnetic field by electrons surrounding them results in a reduced effective magnetic field.

Using (2.13), the on-resonance condition in (2.7) becomes

$$f = \left(\frac{\gamma}{2\pi} \right) B_0 (1 - \sigma). \quad (2.14)$$

From (2.14) one can conclude that nuclei that are chemically non-equivalent are shielded to different extents and thus give separate resonance signals in the spectrum.

Notice that the measured frequency f depends on a particular value of B_0 . Thus, in NMR spectroscopy, there is no absolute spectral scale. A relative scale is used instead, whereby one measures the frequency difference Δf between the resonance signals of the sample and that of a reference compound. To remove the dependency of Δf on B_0 , a quantity, also referred to as the *chemical shift*, is defined as

$$\delta[\text{ppm}] = \frac{\Delta f}{\gamma B_0 / 2\pi} \times 10^6. \quad (2.15)$$

Notice that in (2.15) the numerator is typically no more than a few hundred Hz, whereas the denominator is usually several hundred MHz, both resulting in a very small number. Thus, δ_{sample} is usually given in ppm (by multiplying with 10^6). By convention, the δ -values are positive to the left and negative to the right of the origin at 0 ppm.

Although chemical shift is a basic phenomenon in NMR that results in the observable range of different molecule resonances, it also has an undesirable effect on slice-selection. Specifically, precession frequency of an isochromat in the presence of a slice selection gradient G_{ss} is

$$f = \frac{\gamma}{2\pi} (B_0 + G_{\text{ss}} z), \quad (2.16)$$

where z is the displacement from the gradient isocenter to the spatial position of the considered isochromat. When on-resonance condition in (2.7) is satisfied, it is clear that z must equal zero, i.e. the center of the slice passes through the gradient isocenter. However, since chemical shift causes off-resonance effect, the precession frequency of each isochromat is off by a certain value

as described in (2.15). It then follows that the center of the slice is effectively shifted by

$$\Delta z = \frac{B_0 \sigma}{G_{ss}} = \frac{2\pi \Delta f}{\gamma G_{ss}}. \quad (2.17)$$

Notice that this displacement is specific to each isochromat only, which means that different slices are selected for different isochromats. In this case, the RF carrier frequency needs to be adjusted by an amount Δf , which can be done only when there is one isochromat. Thus, the chemical shift artifact (also referred to as *chemical shift displacement*) cannot be completely eliminated, but rather only reduced by increasing the RF bandwidth at the expense of low SNR.

Notice that besides the chemical shift, an additional feature that can be observed in high-resolution NMR spectra is a splitting of resonances into several smaller lines, a phenomenon often referred to as *J-coupling or spin-spin coupling*. This NMR phenomenon originates from the fact that nuclei with magnetic moments can influence each other through electrons in chemical bonds. As a result, these interactions lead to multiple resonances of an individual molecule. For example, in a simple two-spin system, the resonance of each spin is typically split into two resonances, centered at the original frequency and with the interval between two spectral lines dictated by a scalar coupling constant. In practice, many more complicated types of coupling happen, resulting in a more complex resonance. A thorough explanation of the typical multiple resonances of metabolites is beyond the scope of this review chapter. An interested reader may refer to [16]. Figure 2.4 shows typical spectra with multiple resonances of some common ^1H brain metabolites such as N-acetylaspartate, creatine, choline, glutamate, and myo-Inositol.

2.1.3 Metabolite content of *in vivo* brain ^1H MR spectroscopy

Because of low sensitivity, the majority of *in vivo* MRS studies have been conducted using nuclei with high inherent NMR sensitivity such as protons (^1H). In this research, we focus on ^1H spectrum with a particular application to brain metabolism. Following we briefly review typical metabolites found in the brain ^1H MR spectrum that have numerous investigated clinical studies

to date. Spectral resonances of some metabolites are shown in Fig. 2.4.

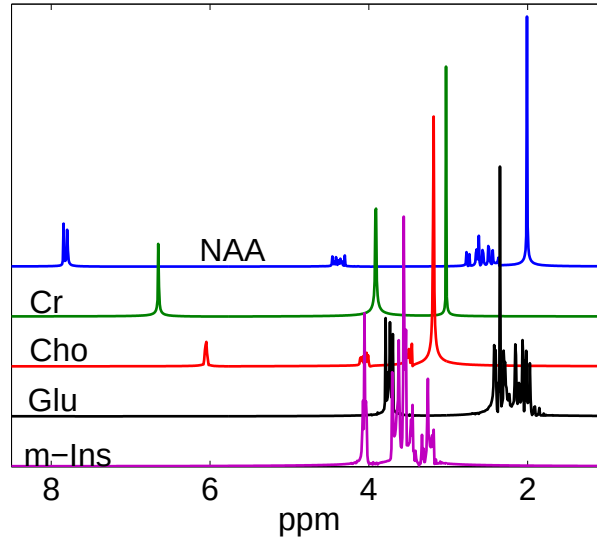


Figure 2.4: Spectral profiles of common metabolites in the human ^1H brain spectrum.

- *N-Acetyl aspartate (NAA)*: it is the largest signal in the normal adult brain spectrum, localized primarily in the central and peripheral nervous systems. A higher concentration of NAA is found in gray matter (GM) (8-11 mM) than in white matter (WM) (6-9 mM). NAA has resonances between 2.0 and 8.0 ppm, with the most prominent singlet at 2.01 ppm. In addition, there are other 3 doublets of doublets at 2.49, 2.67, 4.38 ppm, and a broad doublet at 7.28 ppm [17]. There is a strenuous debate on the function of NAA: on one hand, NAA is believed to serve as a marker of neuronal density; on the other hand, NAA concentrations differ among neuronal types and NAA has also been found in other non-neuronal cells, thus suggesting that NAA may not directly reflect neuron specifics. The first view is supported by various studies showing a decrease in NAA intensity in disorders accompanied with neuronal and/or axonal loss such as chronic stages of stroke, tumors, and multiple sclerosis. The second view is supported by a report of a child with mental retardation in whom NAA resonance was missing [7].
- *Creatine (Cr)*: Cr is involved in energy metabolism, generating adenosine triphosphate. It is synthesized in the liver and kidney and then transported to the brain. Thus, chronic liver

disease typically leads to lower cerebral Cr concentration. A higher concentration of Cr is found in GM (5.2-5.7 mM) than in WM (6.4-9.7 mM). In the normal adult brain spectrum, Cr has two singlet resonances at 3.03 ppm and 3.91 ppm [17].

- *Choline (Cho)*: Cho is the third most prominent resonance in ^1H NMR spectra. Contributions to the total Cho resonance originate from free Cho, glycerophosphorylcholine, and phosphorylcholine. Concentration of the total Cho in the human brain is 1-2 mM, non-uniformly distributed. Cho is involved in the membrane synthesis and degradation. An elevated level of Cho is observed in disease states where an increase in the membrane turnover is observed, such as cancer, Alzheimer's disease, and multiple sclerosis. A decreased level of Cho is associated with liver disease and stroke.
- *Lactate (Lac)*: Increased Lac typically indicates lack of oxygen. Therefore, increased levels of brain Lac have been observed using MRS in a variety of diseases, including ischemia, hypoxia, and brain tumor. Lac resonance contains a doublet at 1.31 ppm from 3 equivalent methyl protons and a quartet at 4.10 ppm from single methine proton [17]. In the healthy human brain, the Lac resonance at 1.31 ppm is normally below detectability in most studies, due to its low concentration (0.5 mM). In addition, Lac overlaps with macromolecular and lipid resonances, which makes it hard to distinguish its resonance. To observe Lac peak at 1.31 ppm, usually TE is chosen such that the peak inversion occurs, that is at $1/J_{\text{lac}} = 144$ ms, where J_{lac} is the J-coupling constant of Lac [5].

The information content of the ^1H MR brain spectrum depends on many factors such as the field strength, choice of timing parameter TE, and type of pulse sequence. At the common 3 T field strength and long TE (from around 140 to 280 ms), only signals from NAA, Cr, Cho, and Glx are clearly seen in the observed spectrum of the normal brain, while other metabolites are barely detectable. At short TEs (usually 35 ms or less) compounds with shorter T_2 relaxation times such as m-Ins become detectable.

2.1.4 Water suppression

Water is the most abundant component in living tissue. The concentration of tissue water is typically thousands of times higher than those of metabolites [5]. Thus, water suppression is a necessary step prior to metabolite quantitation. In ^1H spectrum the strong water resonance is observed around 4.7 ppm with broad lineshapes. This water resonance is rarely a pure exponentially decaying signal due to field inhomogeneities and/or partial water suppression pulse sequences. Water suppression is performed first during the data acquisition stage by employing frequency-selective excitation. However, a residual water peak typically remains in the acquired spectrum. Thus, spectral editing methods are used to further remove the residual water peak.

Frequency-selective excitation methods exploit the difference in the chemical shifts of the water and other resonances. These methods are widely used in the typical spectroscopic pulse sequences and they are very effective in suppressing water resonance. The first water suppression method employed the use of binomial pulses, specifically designed so that metabolite resonances are rotated to the transverse plane for detection, while the water magnetization is returned to the longitudinal axis at the end of the sequence, thus making it unobservable [18,19]. A more routinely used method is chemical shift selective (CHESS) water suppression [20]. This technique applies frequency-selective, shaped RF pulses to excite the water onto the transverse plane and dephase it by the gradient. A detailed explanation of the binomial method and CHESS can be found in [2].

In this dissertation, we focus on removing water using a spectral editing method, performed repeatedly in our data processing in the later chapters. A popular spectral editing method for water removal is the method proposed in [6]. The method exploits the assumption that water resonance is the most dominant component in the spectrum; thus its content can be represented well with a few Lorentzian basis functions. Specifically, a data fit of the time domain FID signal is performed using the HSVD method [21] with a small number of Lorentzian functions (typically up to three basis components were observed to be enough to describe the non-Lorentzian water peak, although this number can vary for a particular MRS dataset [6]). From the estimated parameters (amplitudes,

frequencies, and damping factors) we can further choose components that have frequencies corresponding to the water region (around 4.7 ppm), yielding the estimated time domain water signal. This signal is then subtracted from the original FID signal and Fourier transformed to yield a water suppressed spectrum. The described method was observed to be fast and quite robust in removing a dominant portion of the water peak. Choosing the model order (number of Lorentzian functions used to represent water resonance) is critical for the performance of this method. It is generally recommended to use a small model order to guarantee that the metabolite components are not removed as well, since a small resulting residual water peak will be not problematic for metabolite quantitation. The performance of this water suppression method is presented in Chapter 5, Section 5.3.7.

2.1.5 Baseline suppression

In the typical MR spectrum, besides the presence of metabolite signals of interest, resonances from nuisance signals are also present. These signals are mainly contributed due to macromolecules and lipids in the living systems and are often referred to as the baseline. The nuisance signals usually overlap with metabolite contributions and therefore are typically not of interest to metabolite quantitation. However, recent studies show that there is some correlation between altered macromolecule profile and several conditions such as stroke [22], brain tumor [23], and multiple sclerosis [24].

One straightforward way to remove baseline contribution is to acquire the MRS data at a longer TE, due to the fact that baseline signals have short T_2 decay. However, this is typically not desirable because in addition to the baseline suppression, metabolite signals with short T_2 decays are removed as well. In addition, exact *in vivo* determination of T_2 parameters for metabolites and baseline components is complicated. The same concern holds for the method which separates baseline and metabolite signals based on their differences in T_1 relaxation parameters (T_1 values of metabolites are well above 1000 ms, while T_1 values for lipids are in the range from 250 ms to 350 ms [2]). The presence of baseline becomes an even more serious problem when the spectral

resolution is low, since some metabolite peaks such as alanine or Lac in the range from 1.2 to 1.6 ppm are no longer resolved and are mixed with the baseline resonance [14].

The prior knowledge of *in vivo* MRS baseline signal is still not precise and complete, so there is no existing parametric model function associated to macromolecule components. Conventional metabolite quantitation methods assume the smoothness of the signal in the frequency domain either to remove the baseline in the preprocessing step [25, 26] or to model the signal in the quantitation step in a non-parametric way [14, 27]. Baseline suppression using frequency-selective pulses can also be used but may be less effective due to a strong overlap of baseline and metabolite components in the spectrum.

In the preprocessing step, a popular and rather effective semi-parametric approach is ‘Subtract’ [25]. It relies on the physical assumption that most of the baseline signal is contained within the first time-domain data points, due to the baseline smoothness. Thus, truncation of the initial data points should remove baseline contributions. Quantitation and back-extrapolation of this truncated signal is then performed, after which the resulting signal is subtracted from the original FID signal. The resulting signal is the estimated baseline and can be further parametrized with Lorentzian functions.

A popular method that models the baseline during the quantitation step is Linear Combination Model (LCModel) [14]. It represents the baseline signal in a non-parametric way as a linear combination of cubic B-splines with equally spaced knots. The quantitation procedure is performed to estimate both metabolite and baseline parameters. A different representation of the baseline signal was proposed in [26, 28]. The idea of this technique is that the baseline can be represented by a small number of wavelet coefficients, due to the baseline smoothness. Specifically, data fitting with *a priori* known metabolite basis set was performed to remove estimated metabolite contributions in the spectrum. Remaining narrow spectral features in the region close to known metabolite resonance locations were further removed. A wavelet transform of the resulting signal was performed to characterize the baseline using wavelet coefficients. The resulting parametrized baseline signal was subtracted from the original acquired signal to yield metabolite components. The whole

procedure is then repeated for several iterations. In [26, 28, 29] an empirical comparison of the baseline fitting models, using either wavelets or B-splines, was performed and the results did not show significant differences in the accuracy of the quantified metabolite values.

2.1.6 MRS quantitation

The objective of MRS is to yield quantitative determination of either relative metabolite amplitudes or absolute metabolite concentrations, also referred to as the *metabolite quantitation*. The oldest and still widely available (on the scanners) quantitation method is a simple integration of the area under the spectral peaks. However, due to substantial overlap of the peaks, this approach has low estimation accuracy. An enormous amount of research has been devoted to the development of accurate and robust MRS quantitation algorithms. A thorough review of the existing quantitation methods can be found in [30–32]. The purpose of this section is not to describe all these methods in detail, but rather to provide a summary of the major trends in the concepts of these quantitation methods.

MRS quantitation methods can be classified into two main categories:

1. The first generation of proposed methods (dating from the late 1980s to the late 1990s) considered the discrete MRS signal as a sum of exponentially decaying time-domain signals corresponding to the n -th individual resonance peak with amplitude c_n , phase ϕ_n , damping factor d_n , and resonance frequency f_n :

$$s[m] = \sum_{n=1}^N c_n e^{j(\phi_0 + \phi_n)} e^{(-d_n + j2\pi f_n)m\Delta t} + \xi[m], \quad (2.18)$$

where Δt is the sampling rate, ϕ_0 is the zero-order phase, and $\xi[m]$ denotes additive noise. The amplitude c_n is proportional to the number of nuclei contributing to the n -th spectral peak. The damping factor d_n relates to the peak line-width, reflecting T_2 decay and effect of field inhomogeneity. Equivalently, this model expresses spectral lineshapes in terms of Lorentzian functions. However, field inhomogeneity and limited spatial resolution can result

in non-Lorentzian lineshapes. In this case, alternative Gaussian or Voigt functions are used to represent non-ideal spectral lineshapes as

$$s[m] = \sum_{n=1}^N c_n e^{j(\phi_0 + \phi_n)} e^{(-g_n m \Delta t + j 2\pi f_n) m \Delta t} + \xi[m], \quad (2.19)$$

$$s[m] = \sum_{n=1}^N c_n e^{j(\phi_0 + \phi_n)} e^{(-d_n - g_n m \Delta t + j 2\pi f_n) m \Delta t} + \xi[m], \quad (2.20)$$

respectively. Estimation of metabolite amplitudes can be performed in two ways. The first approach, which works for all Lorentzian, Gaussian or Voigt lineshape models, is to perform the maximum likelihood (ML) estimation by minimizing the l_2 norm of the difference between the measured data and the model function. Two widely used quantitation methods in this category are Variable Projection (VARPRO) [33] and Advanced Method for Accurate, Robust and Efficient Spectral fitting (AMARES) [34]. VARPRO uses a Levenberg-Marquardt [35] optimization routine while AMARES uses NL2SOL [36] to minimize the cost function. The second approach is often referred to as a “black box” method. It was recognized that, in fact, (2.18) is related to the linear prediction or harmonic retrieval (HR) problems, widely investigated in the field of signal processing [37, 38]. Specifically, (2.18) can be rewritten as

$$s[m] = \sum_{n=1}^N a_n z_n^m + \xi[m], \quad (2.21)$$

where the absolute value of a_n is the amplitude of the n -th spectral peak, the absolute value of z_n is proportional to the damping factor, and the phase of z_n relates to the peak frequency. The HR technique considers estimating both a_n and z_n given the measured data $s[m]$. The estimation is usually performed indirectly by determining linear prediction coefficients p_n

in the following matrix equation:

$$\begin{bmatrix} s[N] \\ s[N+1] \\ \vdots \\ s[M-1] \end{bmatrix} = - \begin{bmatrix} s[0] & \dots & s[N-1] \\ s[1] & \dots & s[N] \\ \vdots & \ddots & \vdots \\ s[M-N-1] & \dots & s[M-2] \end{bmatrix} \begin{bmatrix} p_N \\ p_{N-1} \\ \vdots \\ p_1 \end{bmatrix}, \quad (2.22)$$

from which a_n and z_n are then computed. Different variants on how to estimate p_n accurately in the presence of noise were proposed [30], but among these methods, Linear Prediction and Singular Value Decomposition (LPSVD) (proposed in [39] and particularly suggested for MRS quantitation in [40]) and Linear Prediction and Total Least Squares (LPTLS) [41] methods are the most popular. Another approach to estimate a_n and z_n is to exploit SVD properties of the time-domain data matrix. Methods following this approach are often referred to as the *state-space* methods. A widely used state-space method is Hankel SVD (HSVD), originally proposed in [42] and particularly suggested for MRS quantitation in [21]. Improved variants of HSVD are Hankel Lanczos SVD (HLSVD) [43] and Hankel TLS (HTLS) [44]. The details and illustration of the performance of HSVD can be found in Chapter 5, Section 5.3.7.

2. The second generation of methods for MRS quantitation (dating from the late 1990s to the first decade of the twenty-first century) exploits prior knowledge of the spectral profiles of metabolites. Because of a substantial overlap of spectral peaks, incorporation of such prior information conceptually should improve the quantitation performance. The additional prior information can be incorporated in the frequency domain as in the LCModel method [14] or in the time domain as in HTLS-PK [45], KNOB-SVD [46], KNOB-TLS [47], and in the work by B. J. Soher et al. [48]. HTLS-PK assumes that coefficients z_n are known, which is equivalent to knowing damping factors and resonance frequencies. Although the later assumption is accurate, the fact that damping factors (lineshape parameters) can be known

precisely is not practical, since this information is dependent on a particular experiment such as the intrinsic T_2 decay of metabolites, field inhomogeneity, and magnetic susceptibility of the sample. A more practical approach to incorporate the prior spectral information was proposed in KNOB-SVD and KNOB-TLS, where only relative relations among resonance frequencies, amplitudes, and damping factors are assumed to be known. A similar idea is implemented with VARPRO and AMARES [34]. While the mentioned methods exploit “local” prior information, other methods assume that the whole spectral profile of each metabolite can be obtained *a priori*. For example, LCModel operates in the frequency domain and models the *in vivo* observed spectrum using an *in vitro* basis set but frequency-shifted and lineshape-broadened by unknown parameters. A similar concept was proposed in [48] but in the time-domain. The observed spectrum was modeled as a sum of known metabolite spectral profiles, generated by spectral quantum-mechanic simulation [49, 50], and additional frequency shifts and line-broadening parameters were incorporated.

2.1.7 Single-voxel acquisition

Single-voxel MRS localization techniques typically use three mutually orthogonal slice selective pulses and the echo signal is only collected from the volume in space where all three slices intersect. Two widely used localization techniques are PRESS (Point Resolved Spectroscopy) [51–54] and STEAM (Stimulated Echo Acquisition mode) [55–59].

In the PRESS sequence (also referred to as double spin echo localization technique), slice-selective excitation is combined with two slice-selective refocusing pulses (see Fig. 2.5). Specifically, a combination of $90^\circ - \tau_1 - 180^\circ - \tau_2 - 180^\circ$ spatially selective pulses is applied to localize a voxel of interest. These three pulses produce FID and two spin echoes at $2\tau_1$ and $2\tau_2$, respectively. Only the last echo is acquired for spectroscopic encoding.

In the STEAM sequence (also referred to as single-shot or single-scan localization technique), a combination of $90^\circ - \tau_1 - 90^\circ - \tau_2 - 90^\circ - \tau_3$ spatially selective pulses is applied, and the stimulated echo is acquired (see Fig. 2.6). This combination provides a maximum intensity of the

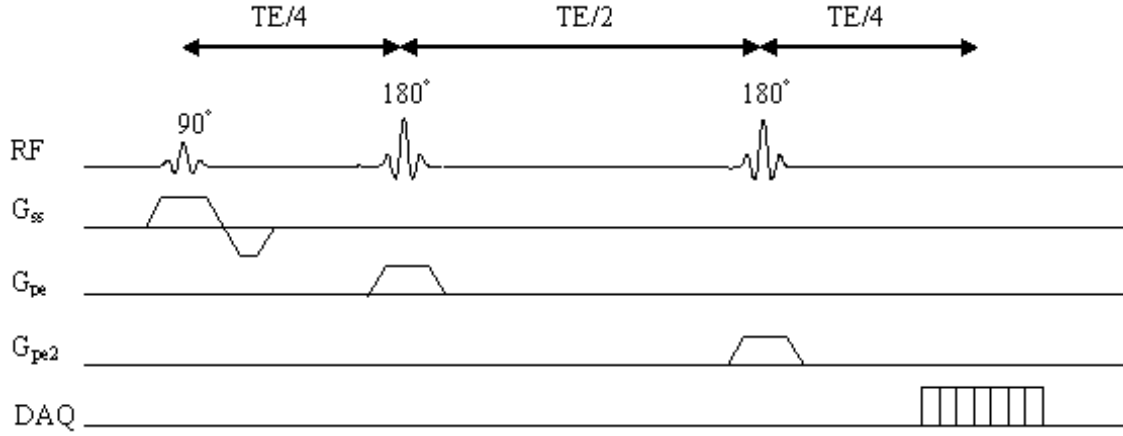


Figure 2.5: A schematic diagram of the PRESS sequence.

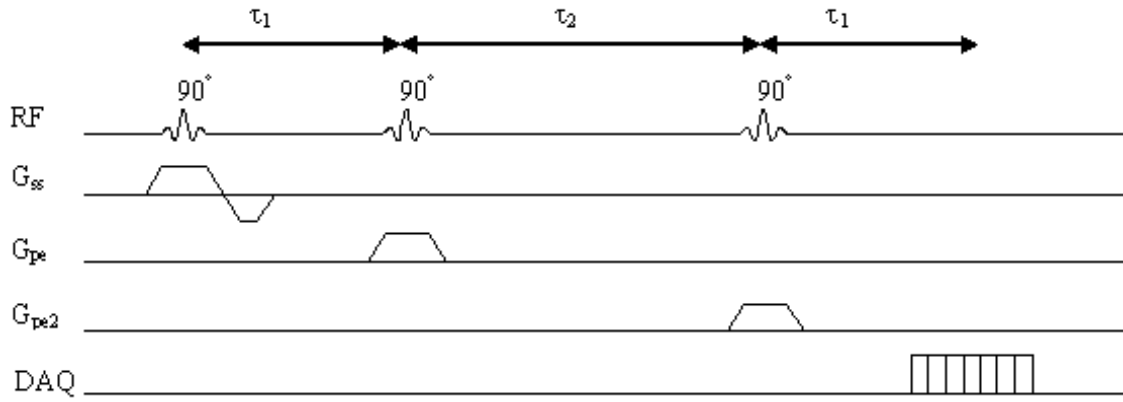


Figure 2.6: A schematic diagram of the STEAM sequence.

stimulated echo.

PRESS sequence has the advantage of twice the signal sensitivity of STEAM, and it is less susceptible to the effects of motion and/or diffusion. However, PRESS requires longer TE; thus, spectral components of interest must possess sufficiently long T_2 values to ensure a detectable NMR signal. In addition, STEAM can have slightly better water suppression performance because additional saturation pulse can be added during the τ_2 period. Overall, PRESS and STEAM pulse sequences are generally comparable in clinical brain spectroscopy and both provide highly effective localization [7, 60].

2.2 Magnetic resonance spectroscopic imaging

Single-voxel proton MR spectroscopy has been applied for characterizing the metabolic characteristics of brain tumors for some time, mainly in the 1990s [61]. Many studies investigated the reduction or increment of common brain metabolites in a variety of conditions. Most popular metabolites of brain studies were NAA, Cr, Cho, and Lac. Recent studies have also investigated on the effect of short and long TEs on the spectrum reproducibility [62].

However, although single-voxel MRS is a relatively rapid method for obtaining information about the metabolism, it does not address spatial information required for mapping of metabolite distributions in the volume of interest. The ability to acquire information from large parts of the subject is important for planning treatments such as radiation and surgical resection [61]. This issue can be addressed by considering multi-voxel MRSI, also known as *chemical shift imaging* (CSI). In CSI, phase-encoding in multiple directions is combined with a readout in the absence of any field gradients. MRSI was first demonstrated in phantoms in 1982 [63]. In 1985, Bottomley et al. reported the first spatially localized human brain spectrum, at 1.5 T using a slice-selective spin echo excitation technique, and frequency selective water suppression [64]. While in MRS the goal is to obtain quantitative chemical information typically expressed in numbers, MRSI provides metabolic information in an imaging format. In recent years, several works have investigated exploiting available spatial information to improve the metabolite quantitation procedure for MRSI, as opposed to performing conventional MRS quantitation voxel by voxel [65–71]. Although the current trend of the field is moving from MRS towards MRSI, nevertheless it is worth mentioning that single-voxel MRS still continues to play a role for many biomedical applications due to its advantages such as much shorter acquisition time (3-10 mins, depending on the chosen TR), good localization, smaller proneness to field inhomogeneity effects, and well-established acquisition protocols.

2.2.1 Principle of MRSI

Given the desired spatial-spectral function $\rho(\mathbf{r}, f)$ describing the sample, the acquired MRS signal in the absence of noise and phase-encoding gradients can be expressed as the sum of signals from the entire volume of interest

$$s_0(t) = \int_{-\infty}^{+\infty} s(\mathbf{r}, t) d\mathbf{r}, \quad (2.23)$$

where $s(\mathbf{r}, t)$ is the Fourier transform of $\rho(\mathbf{r}, f)$ from the frequency domain to the time domain. The observed spectrum is obtained as the Fourier transform of the acquired time-domain signal, as it was shown in (2.11):

$$\begin{aligned} S_0(f) &= \int_{-\infty}^{\infty} s_0(t) e^{i2\pi f t} dt \\ &= \int_{-\infty}^{+\infty} \rho(\mathbf{r}, f) d\mathbf{r}. \end{aligned} \quad (2.24)$$

According to (2.24), in the absence of phase-encoding gradients, we cannot obtain spatially resolved $S_0(f)$. By applying the gradients $\mathbf{G}_r$ as in the conventional MRI, a phase shift is induced at each spatial location and the observed spectrum is now expressed as

$$S_0(\mathbf{G}_r, f) = \int_{-\infty}^{+\infty} \rho(\mathbf{r}, f) e^{-i\gamma \mathbf{G}_r \cdot \mathbf{r} t} d\mathbf{r}, \quad (2.25)$$

where $\mathbf{G}_r = \begin{pmatrix} G_x & G_y & G_z \end{pmatrix}^T$.

Using the k -space interpretation that $k_x(\bar{t}) = \frac{\gamma}{2\pi} \int_0^{\bar{t}} G_x(\tau) d\tau$ and $k_y(\bar{t}) = \frac{\gamma}{2\pi} \int_0^{\bar{t}} G_y(\tau) d\tau$, (2.25) can be rewritten as

$$S_0(\mathbf{k}, f) = \int_{-\infty}^{+\infty} \rho(\mathbf{r}, f) e^{-i2\pi \mathbf{k} \cdot \mathbf{r}} d\mathbf{r}. \quad (2.26)$$

Thus, the acquired MRSI signal in the presence of phase-encoding gradients and absence of

noise can be expressed as

$$s_0(\mathbf{k}, t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho(\mathbf{r}, f) e^{-i2\pi\mathbf{k}\cdot\mathbf{r}} e^{-i2\pi ft} d\mathbf{r} df. \quad (2.27)$$

Equation (2.27) describes the basic relation between the acquired MRSI data and the desired spatial-spectral function. It follows from (2.27) that, in general, an infinite number of data points is required to reconstruct the spatial-spectral function. In the case of finite number of samples, one solution consistent with the measured data points is to take a multi-dimensional discrete Fourier transform (DFT) of the sampled data points $\{s_0[n_x\Delta k_x, n_y\Delta k_y, m\Delta t]\}_{n=0, m=0}^{N-1, M-1}$ under the Nyquist rate

$$\Delta k_x = \frac{1}{\text{FOV}_x}, \quad \Delta k_y = \frac{1}{\text{FOV}_y}, \quad \Delta t = \frac{1}{\text{BW}}, \quad (2.28)$$

where FOV_x , FOV_y , and BW are the field of view of interest in 2-D and the spectral bandwidth, respectively. Notice that because the data cannot be sampled infinitely and thus is expressed as the product with a rectangular sampling function, the reconstructed spatial-spectral function is equivalent to convolving the original spatial-spectral function with a kernel (sinc) function. This kernel function is also referred to as the *point spread function* and it is often used to characterize the effects of data truncation such as the resulting resolution and signal contamination/leakage. For MRSI in particular, the effects of data truncation are more pronounced than that of conventional MRI due to a limited number of phase-encodings. For example, according to [2], in a 1D MRSI experiment the actual voxel size can be 21% larger than its nominal voxel size, only 87.3% of the observed signal originates from the intended spatial location, and the remaining 12.7% of the signal is leaked to adjacent voxels.

2.2.2 Practical limitations

Several issues that limit practical utility of MRSI are as follows:

1. *Long acquisition time:* Long acquisition times of conventional MRSI sequences are at-

tributed inherently to the presence of additional spectral dimension. Typically, the total acquisition time required to acquire $N_x \times N_y \times N_z$ phase encodings is

$$T_{\text{acq}} = N_x N_y N_z \text{TR}, \quad (2.29)$$

where TR is the repetition time of the pulse sequence. Thus, to obtain information from $32 \times 32 \times 32$ phase encodings with $\text{TR} = 2$ s would require 18.2 hours, which is not acceptable. Recently developed fast MRSI sequences such as EPSI are able to reduce the acquisition time by a factor equivalent to the number of phase encodings along one dimension, thus for this example reducing 18.2 hours to about 34 mins. This time reduction typically comes at the expense of the reduced SNR, increasing sensitivity to field inhomogeneity effects, and in most cases, implementing such acquisition schemes requires delicate and precise calculations of acquisition parameters in order to obtain useful MRS signal. Because of long acquisition times, the numbers of signal averaging and phase encodings are usually limited, resulting in an undesirable trade-off between SNR and resolution. This is illustrated in Fig. 2.7, which shows a typical MR anatomical image of the mouse brain versus NAA and Lac spatial distributions resulting from the measured MRSI data of the same subject. Anatomical image was acquired using spin-echo sequence within 1500 ms, while the total acquisition time of the MRSI data was around 3.5 hours for all 8 signal averages [72]. Notice that even with this significantly longer acquisition time, MRSI data had much lower SNR and lower resolution compared to the anatomical image.

2. *Low signal sensitivity:* MRS can detect signals from many tissue metabolites, though for *in vivo* studies the performance is limited by sensitivity considerations: (a) the NMR phenomenon itself is relatively insensitive, since only very small number of spins can be “activated” for generation of NMR signals due to the very small population difference between spin energy levels; (b) metabolite concentrations (less than 10mM) are typically thousand of times lower than that of tissue water (40-45M), used for most conventional MRI acquisi-

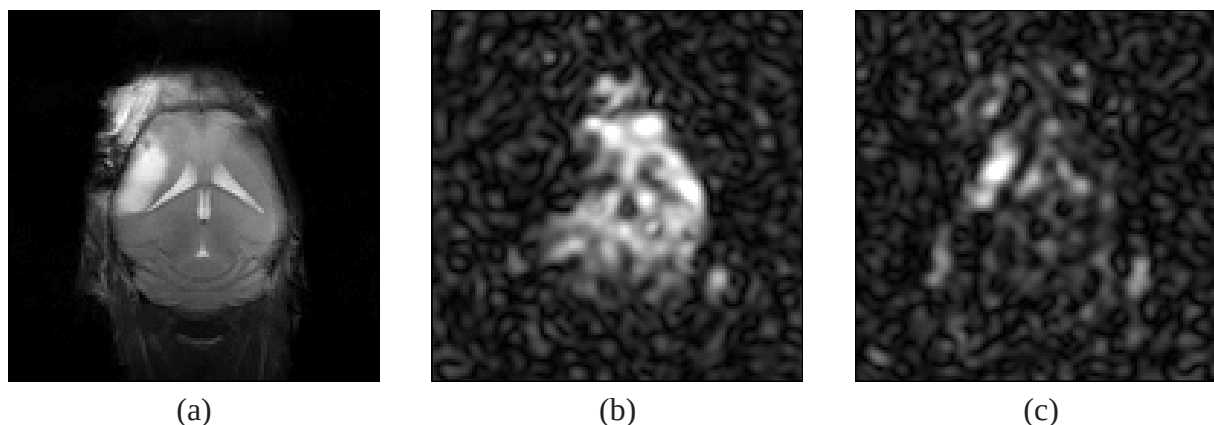


Figure 2.7: Comparison of conventional MRI versus MRSI data of the same subject: (a) Anatomical image acquired using spin-echo sequence, $TE=40$ ms, $TR=1500$ ms; (b)-(c) NAA and Lac spatial distributions obtained from MRSI data by integrating the complex spectrum at 2.02 and 1.33 ppm, respectively. Total acquisition time for the MRSI data was about 3.5 hours (8 signal averages). Notice that this measured MRSI data have low SNR and low spatial resolution.

tions [5].

3. *Field inhomogeneity:* Magnetic field inhomogeneity is a long-standing problem in MRSI.

The presence of field inhomogeneity is attributed to the spatial variation of the main magnetic field and the local differences in magnetic susceptibility of different tissues. With infinitely high spatial resolution, field inhomogeneity results in different frequency shifts for the same chemical species. These undesirable frequency shifts are easy to compensate given the known field map, acquired from conventional MRI scans. However, to maintain a reasonable scan time, the number of phase encodings that can be applied is limited, resulting in a low spatial resolution. In this case, field inhomogeneity leads not only to frequency shifts but also to line broadening and lineshape distortion of MRS resonance peaks in the spectrum, obtained using the traditional Fourier transform method. The later artifacts are hard to compensate and were the topics of intensive research in the recent years [73–79].

4. *Baseline signals:* As discussed in the previous section, baseline contribution from macro-molecule and lipid resonances is a problem for MRS. For MRSI, this problem is even more serious due to the inter-voxel contamination, caused by finite data sampling. For example,

extracranial lipids residing outside of the brain typically lead to a significant signal leakage into the region of interest. Conventional pulse sequence methods for lipid suppression are volume pre-localization and outer volume suppression (OVS) [2]. The first approach uses localization techniques such as Stimulated Echo Acquisition Mode (STEAM) and Point Resolved Spectroscopy (PRESS) (see Section 2.1.7) to excite a large rectangular volume inside the brain, not including the spatial origins of extracranial lipids. Thus, the lipids do not contribute to the observed signal. OVS technique can be considered as opposed to the volume pre-localization, where narrow slices in lipid regions are selectively excited and the resulting transverse magnetization is dephased by a gradient. Baseline can also be removed during the quantitation step, as we discussed earlier in Section 2.1.5. For MRSI in particular, additional regularization can be imposed to penalize a certain degree of smoothness of the baseline, as it was proposed in the method called Automated Quantitation of Short Echo time Spectra (AQSES) [27]. Specifically, a discrete derivative operator was introduced in the regularization term to characterize the baseline smoothness.

2.2.3 MRSI quantitation

A straightforward way to quantify MRSI data is to perform MRS quantitation voxel by voxel. Among various MRS quantitation methods, discussed in Section 2.1.6, LCModel is the most popular and seems to be a reliable quantitation method. However, since MRSI data contains additional spatial information, exploiting spatial correlation in the data should further improve the accuracy of metabolite quantitation. Several such methods include Refs. [65, 67–71]. A recently proposed method AQSES-MRSI [65] assumes that metabolite parameters do not vary much in some spatial neighborhood. Thus, the method adds a penalty term that promotes spatially smooth parameter maps.

Another class of MRSI quantitation methods has a conceptually different treatment of the problem than the quantitation methods described previously. Specifically, this statistical class of methods treats the quantitation as a feature extraction problem. Representative methods in this class

are peak quantitation using principal component analysis (PCA) [80–82], independent component analysis (ICA) [83, 84], and nonnegative matrix factorization (NMF) [85]. The basic idea of PCA is to examine the data-covariance structure of the MRSI dataset. Basic features that represent the largest variance of the dataset, also referred to as *principal components*, are then extracted. Since a major variability in the spectral dataset is due to variations of peak amplitudes, one can obtain the estimates of peak shapes and their amplitudes from the principal components. A disadvantage of PCA is that it is hard to quantify spectra with multiple, overlapping peaks. While PCA is a general statistical method that can be applied to any type of multi-dimensional data, ICA assumes further that metabolites are statistically independent signals and the spectrum is modeled as a linear combination of these signals. ICA extracts independent components from the dataset by maximizing non-Gaussianity of “distributions” of these components, resulting in the separation of metabolite signals from the observed multi-voxel dataset. NMF is another blind signal separation method applied to MRSI data. It formulates the separation of the observed spectral data into constituent metabolite spectra as an ML framework with a positivity constraint imposed on the spectral values. Specifically, a gradient-descent algorithm was proposed to solve the minimization problem to jointly estimate metabolite spectral profiles and their amplitudes at all voxels. Convergence of these update rules to a local optimum was proved in [85].

Although results on the performance of PCA, ICA, and NMF showed that these methods were able to automatically extract biochemical features out of a large dataset, a careful treatment of these approaches is preferred. A major concern with statistical approaches is that the extracted metabolite spectral profiles may not always be physically meaningful, especially in the case of low SNR. A thorough validation with *in vivo* experimental MRSI data would be desirable.

2.2.4 Data acquisition

Chemical Shift Imaging (CSI) technique is an established sequence used in MRSI and nowadays it is available on most scanners. In this conventional technique, a single excitation is followed by a brief spatial encoding step using phase encoding gradients. After that, a spectroscopic readout

with no gradients is applied to spectrally encode the FID signal. The process is repeated until all the necessary phase encoding steps have been performed (see Fig. 2.8). As such, one can see that a main limitation of CSI is long acquisition time. As discussed in Section 2.2.2, to acquire data in this manner with $32 \times 32 \times 32$ phase encodings would require 18.2 hours with TR of 2 s. Theoretical work on analyzing sensitivity CSI sequence was presented in [86].

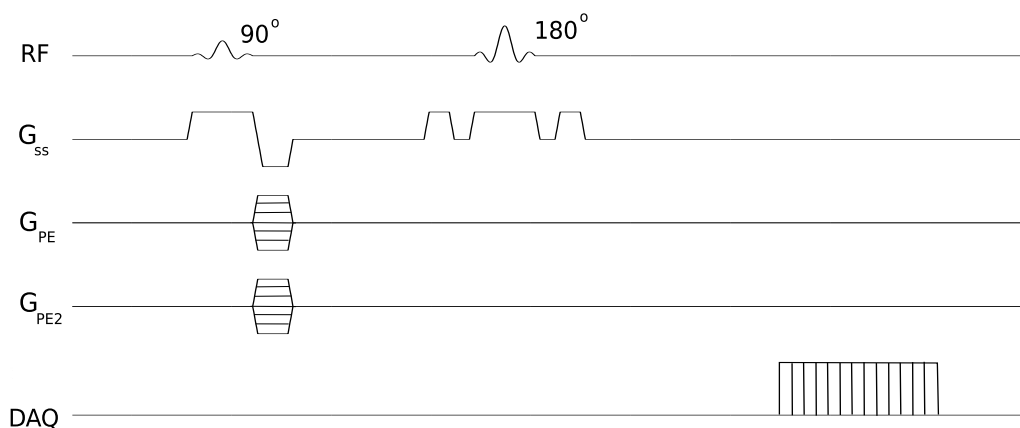


Figure 2.8: A schematic diagram of the conventional CSI pulse sequence.

Long experimental duration is the most serious limitation of CSI and a number of fast CSI techniques have been proposed in order to reduce the acquisition time. Most of these methods modify fast imaging sequences (EPI and FLASH) to include the spectroscopic information. Classical FLASH-based fast CSI techniques are spectroscopic FLASH (SPLASH) [87] and snapshot-FLASH (SNAP) [88]. Some representative EPI-based variations for spectroscopic applications are PREP [89, 90], PEEP [91], SISSI [92]. A thorough review and theoretical evaluation of the sensitivity of these fast CSI techniques can be found in [86], while the basic concept of EPI-based spectroscopic techniques is presented below.

Echo-planar spectroscopic imaging (EPSI): EPSI was first suggested by Mansfield [90] and was later adapted for use in the human brain by Posse et al. [93]. Differently from CSI, an oscillating gradient is applied during readout, with each data point in the time-domain corresponding to one lobe of the read-out gradient. One dimension of spatial-encoding is therefore applied in a single shot concurrently with collection of the spectral data (see Fig. 2.9). Thus, EPSI reduces

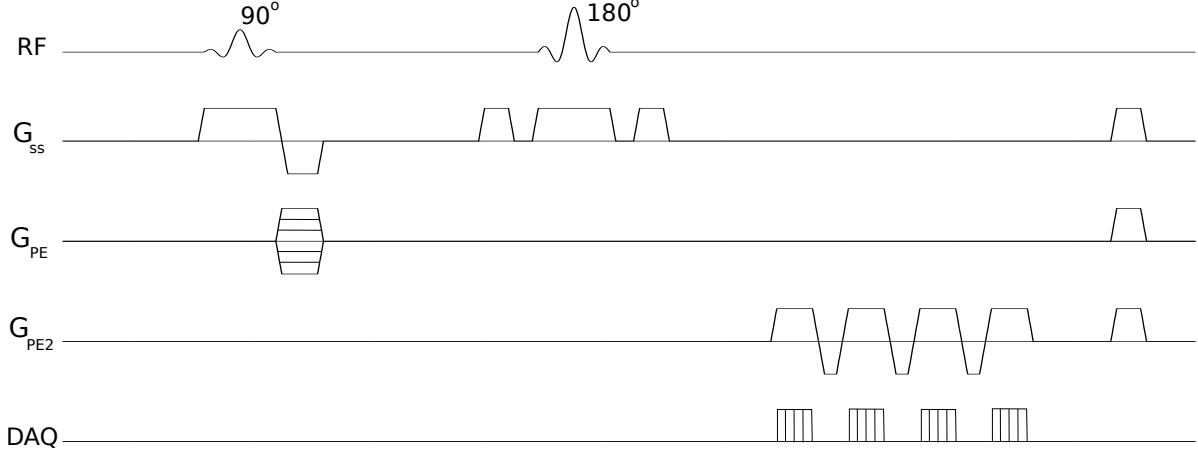


Figure 2.9: A schematic diagram of the EPI pulse sequence.

the order of phase-encoding by one dimension, achieving a significant scan time reduction (e.g. by a factor of 32, for a 32×32 2-D-MRSI). The major limitation of EPI is that the sequence is very sensitive to field inhomogeneity (as is also the case with conventional EPI), resulting in the spectral lineshape broadening and distortions. Methods to compensate for these spectral artifacts caused by field inhomogeneity are listed in Section 2.2.2.

2.3 Denoising in MRSI: A review

A straightforward way to improve the SNR of MRSI data is to acquire multiple sets of measurements for signal averaging, but at the expense of lengthening the already long data acquisition time. Another approach is to apply a linear shift-invariant filter such as a Gaussian smoothing filter. However, such an approach often has poor tradeoff between spatial-spectral resolution and SNR, as illustrated in Fig. 2.10(c). For a better trade-off between SNR and resolution, many advanced denoising methods proposed for general signal processing applications can be used. Most notable in this class are transform-based methods (e.g., wavelet shrinkage [94], SVD truncation [39,95,96], etc.) and PDE-based methods [97]. These methods utilize known signal properties such as piecewise smoothness for feature-preserving denoising.

They are effective when the SNR is beyond some threshold, but cannot well separate signal

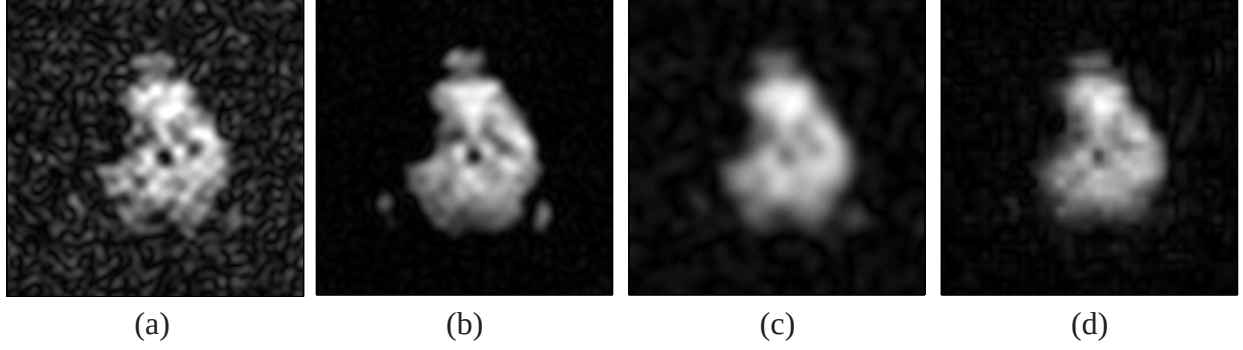


Figure 2.10: Typical denoising of the *in vivo* MRSI data of the mouse brain: NAA map obtained (a) from the measured data, (b) from 32 signal averaging, (c) using a Gaussian smoothing filter, and (d) using 3-D wavelet shrinkage. Notice that signal averaging yielded the best denoising result, however undesirably prolonging acquisition time, while Gaussian filtering has the worst trade-off between the SNR and resolution. In this severe noise contamination case, wavelet shrinkage was not able to preserve spatial features well.

from noise in the presence of severe noise contamination, as is often the case with practical MRSI data, illustrated in Fig. 2.10(d).

In the following subsections, we briefly review denoising approaches for general signal processing applications that can be applied to MRSI data, as well as the denoising methods proposed specifically for MRSI. Our work to address the problem of low SNR in MRSI by exploiting signal properties inherent in the spectroscopic data will be presented in the two subsequent chapters.

2.3.1 Denoising algorithms for general signal processing applications

There have been a tremendous number of proposed denoising methods for general signal processing applications, and this number keeps growing. We consider a few conventionally used methods to illustrate basic concepts of denoising as a filtering problem. Surely quite a number of other popular methods have to be uncovered in this chapter. For the sake of reviewing the key ideas, we ignore the spectral dimension. Specifically, we consider an original image u defined in a bounded spatial domain $\Omega \subset \mathbb{R}^2$ with $\mathbf{r} = (x, y)$, contaminated by additive white noise ξ :

$$v(\mathbf{r}) = u(\mathbf{r}) + \xi(\mathbf{r}). \quad (2.30)$$

Denoising of $\rho(\mathbf{r}, f)$ can be then obtained using various schemes such as by considering real and imaginary parts of $\rho(\mathbf{r}, f)$ frequency by frequency, or as an actual 3-D denoising along each spatial-spectral dimension consecutively, if the corresponding denoising filter is separable.

Linear filtering

A simple approach to denoise $v(\mathbf{r})$ is to apply a linear shift-invariant filter such as a Gaussian smoothing filter, centered at zero with an isotropic variance h^2 :

$$\hat{u}(\mathbf{r}, h) = v(\mathbf{r}) * G_h(\mathbf{r}), \quad (2.31)$$

where

$$G_h(\mathbf{r}) = \frac{1}{2\pi h^2} e^{\frac{-|\mathbf{r}|^2}{4h^2}}. \quad (2.32)$$

One can see from (2.32) that the Gaussian smoothing results in an isotropic filtering effect and different levels of smoothing are obtained for different values of h . Denoised image at any scale h_1 can be obtained from the scale h_0 , where $h_0 < h_1$. It was shown that $u(\mathbf{r})$ satisfies the following linear *heat flow* or *Laplace equation* [98]:

$$\frac{\partial u(\mathbf{r}, h)}{\partial h} = \text{div}(\nabla u), \quad (2.33)$$

where

$$\text{div}(\nabla u) = \frac{\partial \nabla u_x}{\partial x} + \frac{\partial \nabla u_y}{\partial y} \quad (2.34)$$

denotes the Laplace operator. According to (2.33), information (“heat”) diffuses equally in all directions and in the end, a uniform image, equal to the average of the initial image intensities (initial heat), is obtained. In [97], it was shown that Gaussian smoothing is optimal on harmonic functions and performs poorly on singular parts of u such as edges and textures, destroying the signal content in those regions. Thus, linear isotropic filtering in general poorly compromises spatial-spectral resolution for SNR improvement.

Anisotropic diffusion

Anisotropic diffusion is one class of edge-preserving spatial smoothing. It replaces the linear heat flow equation in (2.33) with a nonlinear partial differential equation (PDE) that does not diffuse the image in a uniform way. Instead, anisotropic diffusion smooths the original image while preserving brightness discontinuities. The first formulation of anisotropic diffusion was proposed by Perona and Malik by replacing the classical isotropic diffusion equation (2.33) with

$$\frac{\partial u(\mathbf{r}, h)}{\partial h} = \text{div} \left[g(||\nabla u||) \nabla u \right], \quad (2.35)$$

where $||\nabla u||$ is the gradient magnitude and $g(\cdot)$ is the edge-stopping function: $g(r) \rightarrow 0$ when $r \rightarrow \infty$ so that the diffusion is stopped across edges.

A discrete Perona-Malik formulation of anisotropic diffusion is shown in [98]:

$$\begin{aligned} u_l^{h+1} &= u_l^h + \frac{\lambda}{|\eta_l|} \sum_{k \in \eta_l} g(\nabla u_{l,k}) \nabla u_{l,k}, \\ \nabla u_{l,k} &= u_k^h - u_l^h, \quad k \in \eta_l, \end{aligned} \quad (2.36)$$

where u_l^h is the discretely sampled image, h denotes discrete iterations of the filtering process, l denotes the spatial location in a discrete 2-D grid, η_l is the spatial neighborhood of l with $|\eta_l|$ denoting the number of neighbors, and $\lambda \in \mathbb{R}^+$ represents the rate of diffusion. Interestingly, the formulation in (2.36) relates to the problem of minimizing the following criterion:

$$\min_u \sum_{l \in I} \sum_{k \in \eta_l} \rho(u_k - u_l, \epsilon), \quad (2.37)$$

where function ρ is defined in such a way that minimizes the effect of outliers at the boundaries between piecewise-constant image regions, thus reducing the smoothing effect across edges. Its derivative

$$\psi(\cdot) \triangleq \rho'(\cdot) \quad (2.38)$$

is usually referred to as the *influence function* [98]. As a connection to the anisotropic diffusion equation (2.35), the following relation holds:

$$g(x) = \frac{\psi(x)}{x}. \quad (2.39)$$

In the case of isotropic diffusion, $\rho(x)$ is quadratic, that is $\rho(x) = x^2/\epsilon^2$, or equivalently, the corresponding influence function $\psi(x)$ increases linearly without bound as x increases. Thus, the denoising is very sensitive to outliers, meaning that if the neighboring voxel k and the voxel of interest l lie in two different piecewise-constant regions, the denoised value at l will be biased.

To reject outliers, function $\rho(x)$ should increase less rapidly so that its gradient magnitude (the corresponding influence function) is reduced as x increases beyond a fixed point defined by ϵ , as illustrated in Fig. 2.11. In this figure, an edge-preserving function $g(x)$ is chosen to be a Lorentzian function

$$g(x) = \frac{1}{1 + 2\epsilon^2 x^2} \quad (2.40)$$

and equivalently,

$$\rho(x) = \epsilon^2 \log \left[1 + \frac{1}{2} \left(\frac{x^2}{\epsilon^2} \right) \right]. \quad (2.41)$$

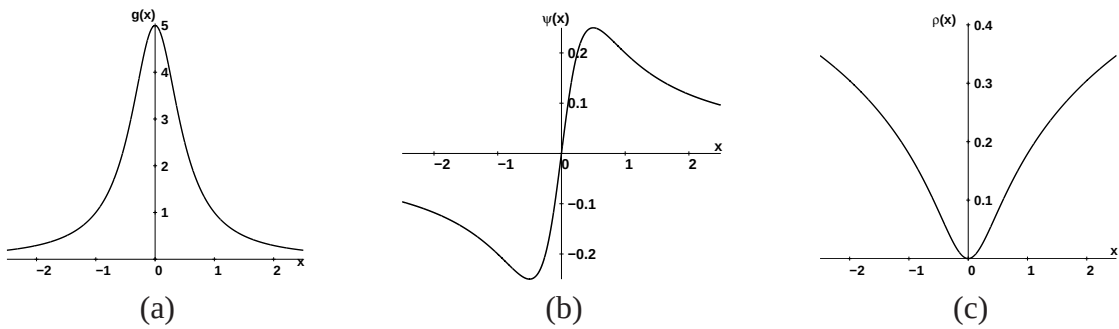


Figure 2.11: Illustration of nonlinear diffusion using Lorentzian edge-preserving function: (a) edge-preserving function $g(x)$; (b) influence function $\psi(x)$; (c) minimization function $\rho(x)$.

Other popular choices of functions $\rho(x)$, $\psi(x)$, and $g(x)$ leading to edge-preserving diffusion are *Tukey's biweight* and *Huber's minimax norm*. Details on the filtering effect of each of these edge-preserving functions as well as a comparison among the three types of anisotropic diffusion

can be found in [98]. Although anisotropic diffusion has the advantage of good edge preservation, it performs poorly on flat regions. In addition, this approach requires a very small time step h to ensure stability of the iterative filtering.

Total variation

Total variation approach was introduced by Rudin and Osher [99], and by Rudin, Osher, and Fatemi [100]. This approach, in some sense similarly to anisotropic diffusion, assumes that the original image is composed of a set of connected objects, each having its own smooth contours or edges. The image is smooth inside the objects but with finite number of jumps across the boundaries. Functional space modeling these assumptions is the space of integrable functions with finite *total variation*, defined as [97]

$$TV_{\Omega}(u) = \int_{\Omega} |\nabla u| d\Omega. \quad (2.42)$$

High total variation indicates signals with excessive and possibly spurious details. Thus, given a noisy image $v(\mathbf{r})$, the denoised image $\hat{u}(\mathbf{r})$ is obtained as

$$\hat{u}(\mathbf{r}) = \arg \min_u TV_{\Omega}(u) \quad (2.43)$$

subject to noise constraints

$$\begin{aligned} \int_{\Omega} (u(\mathbf{r}) - v(\mathbf{r})) d\mathbf{r} &= 0, \\ \int_{\Omega} |u(\mathbf{r}) - v(\mathbf{r})|^2 d\mathbf{r} &= \sigma^2. \end{aligned} \quad (2.44)$$

Minimization problem defined in (2.43) and (2.44) is equivalent to the following unconstrained problem:

$$\min_u TV_{\Omega}(u) + \lambda \int_{\Omega} |v(\mathbf{r}) - u(\mathbf{r})|^2 d\mathbf{r}, \quad (2.45)$$

with λ being a Lagrange multiplier. The above cost function is strictly convex; thus, the solution exists and is unique. A nonlinear optimization is required because the smoothing total variation term is non-differentiable. Since total variation penalizes the absolute magnitude of the gradient, i.e. brightness changes in the image, it is a reasonable denoising approach when the image is piecewise constant. However, in [97] it is shown analytically that the difference between the original and denoised images depends on the curvature of the image: straight edges are preserved because of small curvature, while details and texture having high curvature can be over-smoothed. This leads to a stair-casing effect that is typical for the total variation method.

Neighborhood filter

Unlike other approaches, neighborhood filter takes into account image intensities to define neighboring voxels. Thus, this approach is a fully non-local algorithm, since voxels from the whole image are used to estimate a particular voxel. A popular version of the neighborhood filter is bilateral filter [101]

$$\hat{u}_{\lambda,\beta}(\mathbf{r}) = \frac{1}{C(\mathbf{r})} \int_{\Omega} u(\boldsymbol{\nu}) e^{-\frac{|\boldsymbol{\nu}-\mathbf{r}|^2}{\lambda^2}} e^{-\frac{|u(\boldsymbol{\nu})-u(\mathbf{r})|^2}{\beta^2}} d\boldsymbol{\nu}, \quad (2.46)$$

where

$$C(\mathbf{r}) = \int_{\Omega} e^{-\frac{|\boldsymbol{\nu}-\mathbf{r}|^2}{\lambda^2}} e^{-\frac{|u(\boldsymbol{\nu})-u(\mathbf{r})|^2}{\beta^2}} d\boldsymbol{\nu} \quad (2.47)$$

is the normalization factor. A less known neighborhood filter is Yaroslavskiy filter

$$\hat{u}_{\lambda,\beta}(\mathbf{r}) = \frac{1}{C(\mathbf{r})} \int_{B_{\lambda,\mathbf{r}}} u(\boldsymbol{\nu}) e^{-\frac{|u(\boldsymbol{\nu})-u(\mathbf{r})|^2}{\beta^2}} d\boldsymbol{\nu}, \quad (2.48)$$

where $C(\mathbf{r})$ is the normalization factor:

$$C(\mathbf{r}) = \int_{B_{\lambda,\mathbf{r}}} e^{-\frac{|u(\boldsymbol{\nu})-u(\mathbf{r})|^2}{\beta^2}} d\boldsymbol{\nu} \quad (2.49)$$

and $B_{\lambda, \mathbf{r}}$ defines a spatial neighborhood that is a ball of center $\mathbf{r}$ and radius λ . It was found that there is no significant difference between bilateral and Yaroslavskiy filters and they both yield the same denoising behavior: inside a homogeneous region ($u(\boldsymbol{\nu}) - u(\mathbf{r}) \approx 0$), the Yaroslavskiy filter computes the arithmetic mean of the neighborhood, while the bilateral filter computes the Gaussian mean. At an edge boundary separating two regions, $u(\boldsymbol{\nu}) - u(\mathbf{r})$ is above some threshold, so the influence of voxel $\boldsymbol{\nu}$ is neglected [102]. As a result, neither filter blurs the edges of the image. Although possessing a good edge-preservation property, both neighborhood filters have a stair-casing effect, similarly to the total variation approach.

Wavelet shrinkage

This is one of the state-of-the-art denoising methods, proposed by Donoho and Johnstone [103], which transforms the signal to a de-correlated transform domain (wavelets) and performs thresholding of wavelet coefficients. The resulting signal is then transformed back to the original domain yielding a denoised signal. Wavelet shrinkage relies on the assumption that an image is smooth; thus, the image can be represented well with a few large-valued wavelet coefficients, whereas the noise is distributed across small-valued coefficients, which are removed using either hard thresholding [103] or soft thresholding [104]. Unfortunately, edges lead to a great number of non-zero wavelet coefficients lower than the selected threshold. Cancellation of these wavelet coefficients causes small oscillations near the edges (similarly to Gibbs phenomenon) and spurious wavelets (also referred to as the wavelet outliers). Soft thresholding is less prone to these artifacts than hard thresholding. Other methods to reduce the mentioned artifacts can be found in [105, 106]. In MRS, wavelet shrinkage has been applied to denoise single-voxel MRS data in [94].

2.3.2 Existing approaches for denoising of MRSI data

There are also several denoising methods specifically designed for MRSI signals [94, 107–110]. These methods can be classified into two classes, where the denoising is either performed separately before the MRS quantitation or implicitly during the quantitation stage. In the first case, a

popular method to denoise MRS signal is to apply the wavelet shrinkage [94]. In another work by Ahmed [107], MRS signals are denoised by consecutive projection onto different domains, represented by a set of linear time-frequency transforms. While some improvement was obtained compared with a single-voxel wavelet shrinkage, this method requires an increased computational time. Both wavelet and time-frequency transforms do not explore spectral properties of the MRS signal.

Explicit parametric models have also been used for denoising MRSI data either prior to or during metabolite quantitation [67, 68, 109, 111, 112]. These methods are very effective when the models are correct. However, generic parametric models that fully account for all spectral features are not yet available and incorrect models often create significant bias which can be very problematic in practical applications. For example, a popular denoising method described in [111] and originally proposed in [108] makes strict use of the Lorentzian lineshape of the spectral peaks and is very sensitive to noise. Less-constrained approaches were proposed in [8, 110]. An indirect way of denoising that is performed during the MRS quantitation stage typically incorporates a smoothing regularization to combat the noise. Such a method is AQSES-MRSI [65, 109] (see Section 2.2.3). While in this method the l_2 -type smoothing regularization reduces the noise variance, it is not efficient in preserving edge features. An edge-preserving regularization was used in [110], where a linear combination of the total variation and l_1 norms was used to reduce noise. However, this method models the spectral data as a linear combination of polynomial and spikes, while not utilizing available a priori knowledge of metabolite spectral profiles as in AQSES-MRSI.

Chapter 3

Denoising of MRSI data with spatial-spectral regularization

3.1 Introduction

As discussed in the previous chapter, the existing methods for denoising of MRSI data either have not explored the spectral constraint of the signal and rather treated the MRS signal as a general signal, or have used explicit parametric models, which can produce “noiseless” results, but overly bias the interpretation of the measured data towards the prior assumption. Thus, it is desirable to have a denoising method that incorporates the spectral information in a less-constrained way. We propose to achieve this by using general spatial-spectral constraints that capture typical MRSI signal characteristics in a non-parametric way. Specifically, we use an autoregressive (AR) spectral constraint to capture the general spectral structure. This spectral constraint is used in combination with a spatial constraint [113] to achieve spatial-spectral filtering.

3.2 Proposed method

We denote the spectral Fourier transform of the original spatial-spectral function $\rho(\mathbf{r}, f)$ as $s(\mathbf{r}, t)$ and parameterize $s(\mathbf{r}, t)$ as a linear combination of shifted basis functions $\phi(\mathbf{r} - \mathbf{r}_n, t - t_m)$:

$$s(\mathbf{r}, t) = \sum_{m=1}^M \sum_{n=1}^N s_{n,m} \phi(\mathbf{r} - \mathbf{r}_n, t - t_m). \quad (3.1)$$

We then formulate the denoising problem as solving the following penalized maximum-likelihood

optimization problem:

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}^{\text{meas}} - \mathbf{s}\|_2^2 + \lambda R^{\text{spat}}(\mathbf{s}) + \beta R^{\text{spec}}(\mathbf{s}, \boldsymbol{\alpha}), \quad (3.2)$$

where $\mathbf{s}^{\text{meas}}$ denotes the acquired data in $\mathbf{r}$ - t domain. In this expression, $R^{\text{spat}}(\mathbf{s})$ and $R^{\text{spec}}(\mathbf{s}, \boldsymbol{\alpha})$ are regularization functionals, and λ and β are regularization parameters.

The spatial regularization functional $R^{\text{spat}}(\mathbf{s})$ is used to impose a weighted spatial-smoothness constraint, as originally proposed in [113]:

$$\begin{aligned} R^{\text{spat}}(\mathbf{s}) &= \sum_{m=1}^M \sum_{n_1 \sim n_2} \mathbf{w}_{n_1 n_2} |s_{n_1, m} - s_{n_2, m}|^2, \\ &= \|\mathbf{B}\mathbf{s}\|_2^2, \end{aligned} \quad (3.3)$$

where $n_1 \sim n_2$ denotes the set of neighboring voxels and $\mathbf{w}_{n_1 n_2}$ are spatially varying weights derived from K anatomical reference images ref_k as

$$\mathbf{w}_{n_1 n_2} = \min \left(\frac{1}{\frac{1}{K} \sum_{k=1}^K a_k |\text{ref}_{k n_1} - \text{ref}_{k n_2}|}, w_{\max} \right). \quad (3.4)$$

Since w_{n_1, n_2} is inversely proportional to the variance for the difference between voxels n_1 and n_2 , these weights are used to encourage spatial smoothness and reduce noise in image regions that are expected to be smooth, while preserving edges where they are likely to exist. The threshold $w_{\max}$ is used to reduce the effect of the noise in the reference images and avoid imposing too strong smoothness constraints. For details, an interested reader is referred to [113].

The regularization functional $R^{\text{spect}}(\mathbf{s}, \boldsymbol{\alpha})$ is used to protect spectral features and is formulated based on an ‘‘autoregressive’’ (AR) constraint as

$$\begin{aligned} R^{\text{spect}}(\mathbf{s}, \boldsymbol{\alpha}) &= \sum_{m=L+1}^M \sum_{n=1}^N \left| s_{n, m} - \sum_{l=1}^L \alpha_{n, l} s_{n, m-l} \right|^2 \\ &= \|\mathbf{C}(\boldsymbol{\alpha})\mathbf{s}\|_2^2, \end{aligned} \quad (3.5)$$

where α_l are linear prediction (or also referred to as AR) coefficients. These coefficients are used to specify spectral content of the signal and are pre-estimated from the acquired data.

The spectral regularization term imposes a weaker spectral constraint than parametric modeling methods. It is motivated by the fact that in MR spectroscopy, the spectral lineshapes are commonly modeled as Lorentzian. If the spectral lineshape is Lorentzian, or, equivalently, the time-domain signal has an exponentially decaying form, then the signal is perfectly AR and hence, spectral regularization can be applied with a large value of β for appropriately chosen α_l . As a result, one can expect that in this case, quantitation from the denoised data would yield a more accurate estimate of the amplitude and lineshape parameters than quantitation from the noisy data. In addition, the proposed spectral constraint is also inspired by the fact that a signal that follows an AR process typically is smooth and its power spectrum is concentrated in some frequency bands, rather than spread over the whole spectrum uniformly, as is the case for white noise [114]. Thus, the AR process can characterize spectral correlation structures in the signal, which is used to distinguish signal from noise.

3.2.1 Proposed algorithm

The proposed denoising method requires minimizing the cost function in (3.2) jointly over s and α . However, this leads to a large nonlinear optimization problem (NM linear parameters s and NM nonlinear parameters α). We chose to use an iterative two-step procedure where we iterate between estimating either s or α while keeping the other fixed. In particular, we start from estimating α given the noisy data s^{meas} .

Accurate estimation of linear prediction coefficients α is essential for the proposed denoising approach since α relate to the locations and linewidths of the spectral peaks. A straightforward way to compute α at the first iteration is to solve (3.5) for each voxel independently. However, given that the acquired data have a low SNR, the least-squares solution would be severely contaminated with noise. Instead, we propose to estimate α at the first iteration of the algorithm from a low-rank approximation of the noisy k -t data. By doing so we gain SNR for the estimation of α significantly

and at the same time we can capture spatial variations of the spectral lineshapes, depending on the chosen rank. As an example, a simple case of rank-one approximation is equivalent to estimating α from the k -space center, i.e. from the spectrum, averaged over all voxels. In this case the resulting data from which we estimate α has high SNR, but the spatial variations of the lineshapes are ignored and in addition, the lineshapes may be distorted due to the averaging. Depending on the noise level, α can also be estimated from a first few data points and using state-of-the-art harmonic retrieval techniques that were proposed to solve (3.5) in the case of noise contaminated data (e.g., the total least-squares Prony (TLS Prony) method, iterated quadratic maximum likelihood (IQML) method [37, 115]).

Given the pre-estimated values of α at a current iteration, (3.2) has the closed-form solution

$$\hat{\mathbf{s}} = (\mathbf{I} + \lambda \mathbf{B}^H \mathbf{B} + \beta \mathbf{C}(\alpha)^H \mathbf{C}(\alpha))^{-1} \mathbf{s}^{\text{meas}}. \quad (3.6)$$

This solution can be found efficiently using the iterative conjugate gradient method. Note that $\mathbf{B}$ is a sparse matrix and $\mathbf{C}(\alpha)$ is a Toeplitz matrix, which enables efficient computation of the matrix multiplication $\mathbf{B}^H \mathbf{B}$ due to sparsity and $\mathbf{C}(\alpha)^H \mathbf{C}(\alpha)$ through the fast Fourier transform.

To illustrate convergence of the proposed iterative algorithm, we simulated low-resolution (32×32 spatial voxels and 1024 time points) spatial-spectral function $\rho(\mathbf{r}, f)$ with similar parameters as presented in Section 3.4.1. Noise was added yielding noisy spatial and spectral distributions shown in Figs. 3.1((b),(f) and the proposed denoising algorithm was performed with 30 iterations. Figures 3.1(c),(d) and (g),(h) show typical zero-padded reconstructions obtained from the denoised data at the first and last iterations. Notice that the noise variance was gradually reduced with iterations, as can be seen from Figs. 3.1(c),(d). However, as the algorithm is iterated, the non-uniform denoising effect in the spectral domain was not alleviated (see Figs. 3.1(g),(h)). This is expected, because the non-uniform denoising in the spectral domain is a direct result of the imposed spectral regularization. This denoising effect is discussed in detail in the next section. Table 3.1 shows the normalized cost function and convergence values at each iteration. One can

see that the cost function decreases at each iteration and the iterative estimation algorithm tends to converge within 20-30 iterations for the considered MRSI dataset.

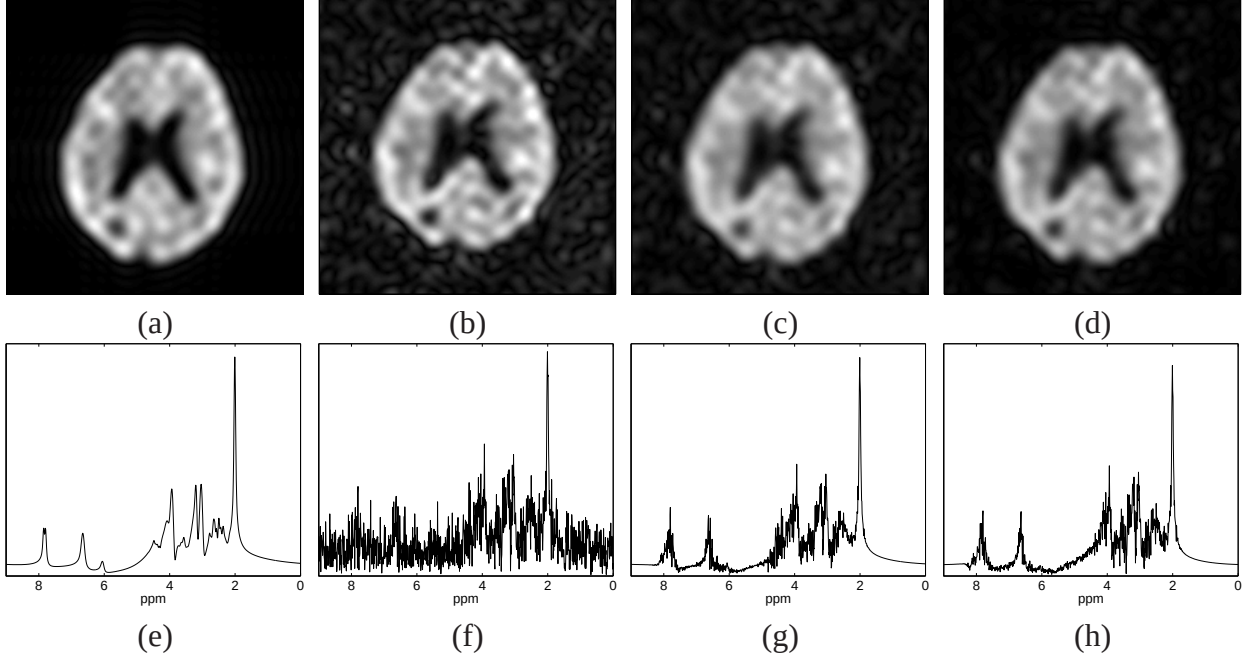


Figure 3.1: Performance of the algorithm at each iteration: top row shows (a) noiseless and (b) noisy NAA spatial distributions; denoised NAA spatial distribution at (c) iteration 1 and (d) iteration 30. Notice that an additional noise reduction is achieved as the algorithm is iterated. Bottom row shows (e) noiseless and (f) noisy spectral distributions; denoised spectral distribution at (g) iteration 1 and (h) iteration 30. Notice that the non-uniform denoising effect in the spectral domain is not alleviated with iterations, as expected.

3.2.2 Characteristics of the proposed algorithm

1) Denoising property

To illustrate the spatial-spectral filtering properties of the proposed approach, we note that the matrix $I + \lambda B^H B + \beta C(\alpha)^H C(\alpha)$ in (3.6) is symmetric. Thus, it can be decomposed as

$$I + \lambda B^H B + \beta C(\alpha)^H C(\alpha) = V^H \Sigma V, \quad (3.7)$$

Table 3.1: Performance of the algorithm at each iteration: convergence at the i -th iteration was computed as $\frac{\|\mathbf{s}^{(i)} - \mathbf{s}^{(i-1)}\|_2}{\|\mathbf{s}^{(2)} - \mathbf{s}^{(1)}\|_2}$ and cost function at the i -th iteration was normalized by cost function at the first iteration.

Iteration	1	2	3	4	5	6	7	8	9	10
Cost function	1	0.995	0.992	0.988	0.985	0.983	0.980	0.978	0.977	0.975
Convergence	NA	1	0.69	0.58	0.50	0.45	0.40	0.36	0.32	0.29
Iteration	11	12	13	14	15	16	17	18	19	20
Cost function	0.974	0.972	0.971	0.970	0.969	0.969	0.968	0.968	0.967	0.967
Convergence	0.26	0.24	0.22	0.21	0.20	0.19	0.18	0.17	0.16	0.15
Iteration	21	22	23	24	25	26	27	28	29	30
Cost function	0.967	0.966	0.966	0.965	0.965	0.965	0.964	0.964	0.964	0.964
Convergence	0.14	0.14	0.13	0.12	0.12	0.11	0.11	0.11	0.10	0.10

where the eigenvector matrix $\mathbf{V}$ is orthogonal. Substituting (3.7) into (3.6) and using the orthogonality of $\mathbf{V}$, one can show the following expression relating the solution and the measured data:

$$\bar{s}_k = w_k \bar{s}_k^{\text{meas}}, \quad (3.8)$$

where $\bar{s} = \mathbf{V}\mathbf{s}$ and w_k is the diagonal element of Σ^{-1} . Equation (3.8) characterizes the denoising property of the proposed method. Specifically, the denoised spatial-spectral function $\mathbf{s}$ is obtained by a point-wise weighting of the measured function $\mathbf{s}^{\text{meas}}$ in the transform domain defined by the eigenvectors of the matrix $\mathbf{I} + \lambda \mathbf{B}^H \mathbf{B} + \beta \mathbf{C}(\alpha)^H \mathbf{C}(\alpha)$.

A further understanding of this weighting property is achieved by approximating the Toeplitz matrix $\mathbf{C}(\alpha)$ with a circulant matrix and ignoring the spatial regularization term (by setting $\lambda = 0$). With this approximation, one can show that $\mathbf{V}$ is simply a unitary DFT matrix, and the weights in (3.8) are

$$w_k = \frac{1}{1 + \beta |h_k|^2}, \quad (3.9)$$

where

$$\{h_k\} = \text{DFT}\{-\alpha_L, -\alpha_{L-1}, \dots, -\alpha_1, 1, 0, \dots, 0\} \quad (3.10)$$

are the DFT samples of the frequency response of the L -th order FIR linear predictor/annihilating filter at each spatial location. Thus, in this case, the proposed spectral regularization performs weighting in the spectral domain: each frequency is scaled by a weighting coefficient dictated by the frequency response of the annihilating filter, resulting in the noise suppression in the frequency regions where there is no predicted (by the linear prediction coefficients) metabolite content, while the frequency regions with strong signal content are preserved.

The weighting effect discussed above resembles the one described in [116]. However, because the weights are generated from the AR constraint, they help to identify the frequency support of the spectrum, and using the AR constraint in combination with the spatial regularization helps to protect spatial-spectral features during the denoising process. In addition, the weighting property in the frequency domain, introduced by the circulant approximation, does not hold exactly. Figure 3.2 shows the noiseless spectrum, the reconstructed spectrum from the proposed denoising approach, and the explicitly weighted (in the frequency domain) spectrum using the pre-computed weights according to (3.9). As we can see from the figures, the reconstructed and weighted spectra do not coincide for all cases of insignificant and severe spectral constraint. As β increases, the difference between the weighted and reconstructed spectra becomes clearer. This study implies that the transform domain where the weighting is performed is not the frequency domain, as the circulant approximation suggests. A nice feature is that the denoising weights in (3.9), obtained using the circulant approximation, reflect noise variance reduction at each spectral point.

2) *Spatial-spectral response function*

The proposed reconstruction is linear given the fixed linear prediction coefficients $\alpha_{n,l}$ and hence the algorithm can be characterized in terms of the spatial-spectral resolution properties. In [113] the spatial response function has been analyzed. Here, we consider an additional spectral dimension and compute the spatial-spectral response function (SSRF). One can show that the estimated distribution $\hat{\rho}(\mathbf{r}, f)$ relates to the true distribution $\rho(\mathbf{r}, f)$ as

$$\hat{\rho}_{\mathbf{r}_n, f_n} = \int \int \rho(\mathbf{r}, f) h_n(\mathbf{r}, f) d\mathbf{r} df + \bar{\xi}_n(\mathbf{r}, f), \quad (3.11)$$

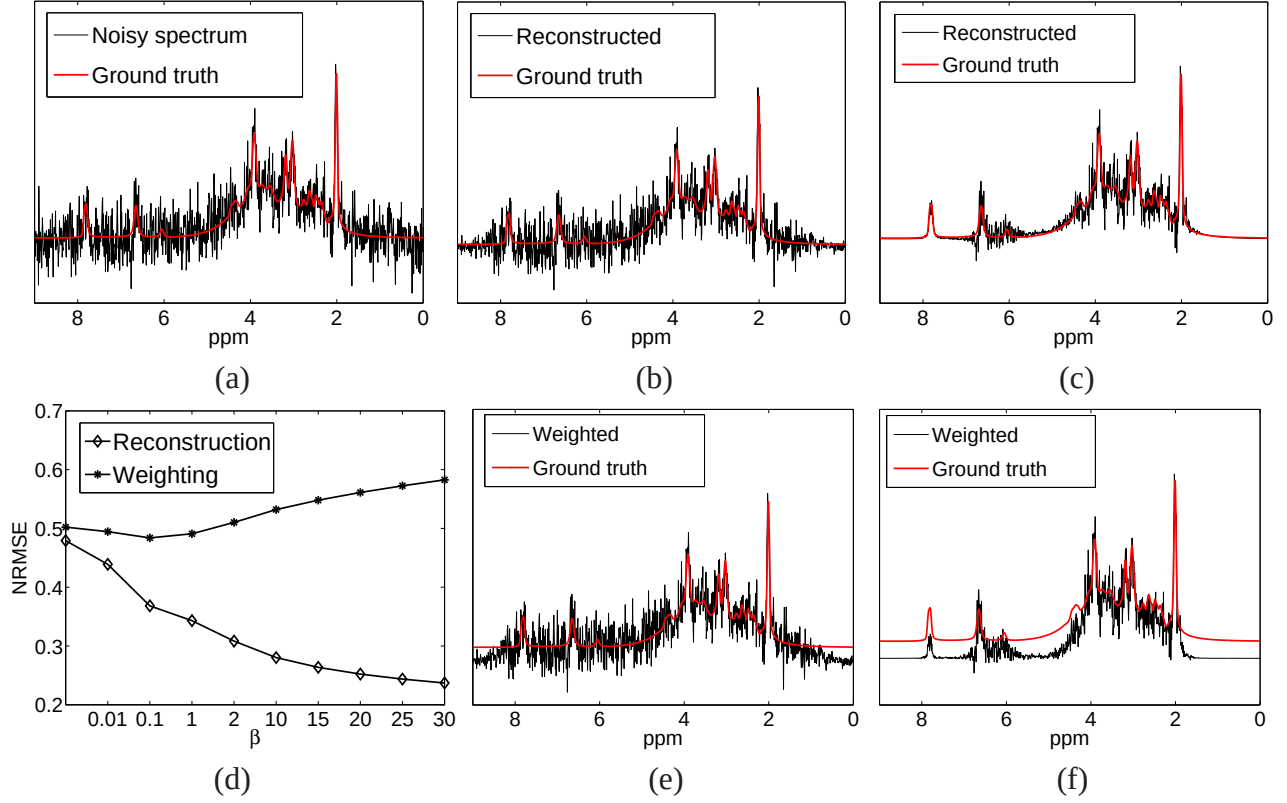


Figure 3.2: Illustration of the denoising property of the spectral constraint: (a) noisy spectrum; (b) reconstruction at an insignificant value of $\beta = 0.01$; (c) reconstruction at a moderate value of $\beta = 10$; (d) NRMSEs of the actual and weighted in frequency domain solutions as a function of β ; (e) weighted (in frequency domain) solution at $\beta = 0.01$; (f) weighted (in frequency domain) solution at $\beta = 10$. One can clearly see that the weighting is performed in the transformed domain other than the frequency domain. Noiseless spectrum is displayed in red.

where $h_n(\mathbf{r}, f)$ is the SSRF in the continuous spatial-spectral domain. From (3.6) and (3.11) it follows that the SSRF can be obtained as

$$h_n(\mathbf{r}, f) = \sum_p \sum_m e^{i2\pi \mathbf{k}_p \cdot \mathbf{r}_m} \left(\sum_q G_{qm} e^{i2\pi t_q f_n} \right) e^{-i2\pi \mathbf{k}_p \cdot \mathbf{r}} e^{-i2\pi t_m f}, \quad (3.12)$$

where G_{qm} are elements of the matrix

$$\mathbf{G} = \mathbf{I} + \lambda \mathbf{B}^H \mathbf{B} + \beta \mathbf{C}^H \mathbf{C} \quad (3.13)$$

and $\bar{\xi}_n(\mathbf{r}, f)$ is the zero-mean Gaussian noise. Notice that $h_n(\mathbf{r}, f)$ is a continuous function and can

be discretized arbitrarily. This function relates the reconstruction and the ground truth distribution and it is signal-dependent; i.e., the computation of $h_n(\mathbf{r}, f)$ involves the use of spatial weights w_{n_1, n_2} and linear prediction coefficients $\alpha_{n, l}$. The width of the main lobe of SSRF can be used to characterize the spatial-spectral resolution of the reconstruction. In addition, the SSRF provides valuable information about the spatial-spectral leakage introduced by the reconstruction.

Ideally, it is desirable that a reconstruction would have a corresponding SSRF in the form of a δ -function in both spatial and spectral domains. However, an SSRF of any non-perfect reconstruction would deviate from the δ -function and this deviation could be used as a characterization of the spatial-spectral leakage introduced by the reconstruction. Specifically, Fig. 3.3 shows a typical SSRF for the Fourier reconstruction, spatial denoising (by setting $\beta=0$), and the proposed spatial-spectral denoising. Spectral regularization parameter β was intentionally chosen to be overly large so that its effect on the SSRF could be clearly observed. Due to Fourier series, the Fourier reconstruction introduces a spatial leakage in the form of the ringings in $h_n(\mathbf{r}, f)$, as shown in Fig. 3.3(a). This leakage is attenuated by applying the spatial denoising at the expense of increasing the width of the main lobe of the SSRF, i.e. at the expense of degrading spatial resolution. The proposed spatial-spectral reconstruction does not notably worsen spatial resolution compared to the spatial denoising (compare Figs. 3.3(b) and (c)). However, it introduces some spectral leakage. This leakage amount depends on β and the frequency of interest f_n . If f_n is one of the frequencies determined by linear prediction coefficients α_l , which we refer to as *harmonic frequencies*, the spectral leakage is negligible. Figure 3.3(e) illustrates this by showing SSRF at a resonance frequency of NAA (2 ppm), which has the same amount of spectral leakage as the spatially denoised reconstruction shown in Fig. 3.3(d). However, if f_n is not one of the harmonic frequencies, then the amount of spectral leakage increases. As shown in Fig. 3.3(f), SSRF at a resonance frequency $f_n = 7.23$ ppm indicates significant spectral leakage.

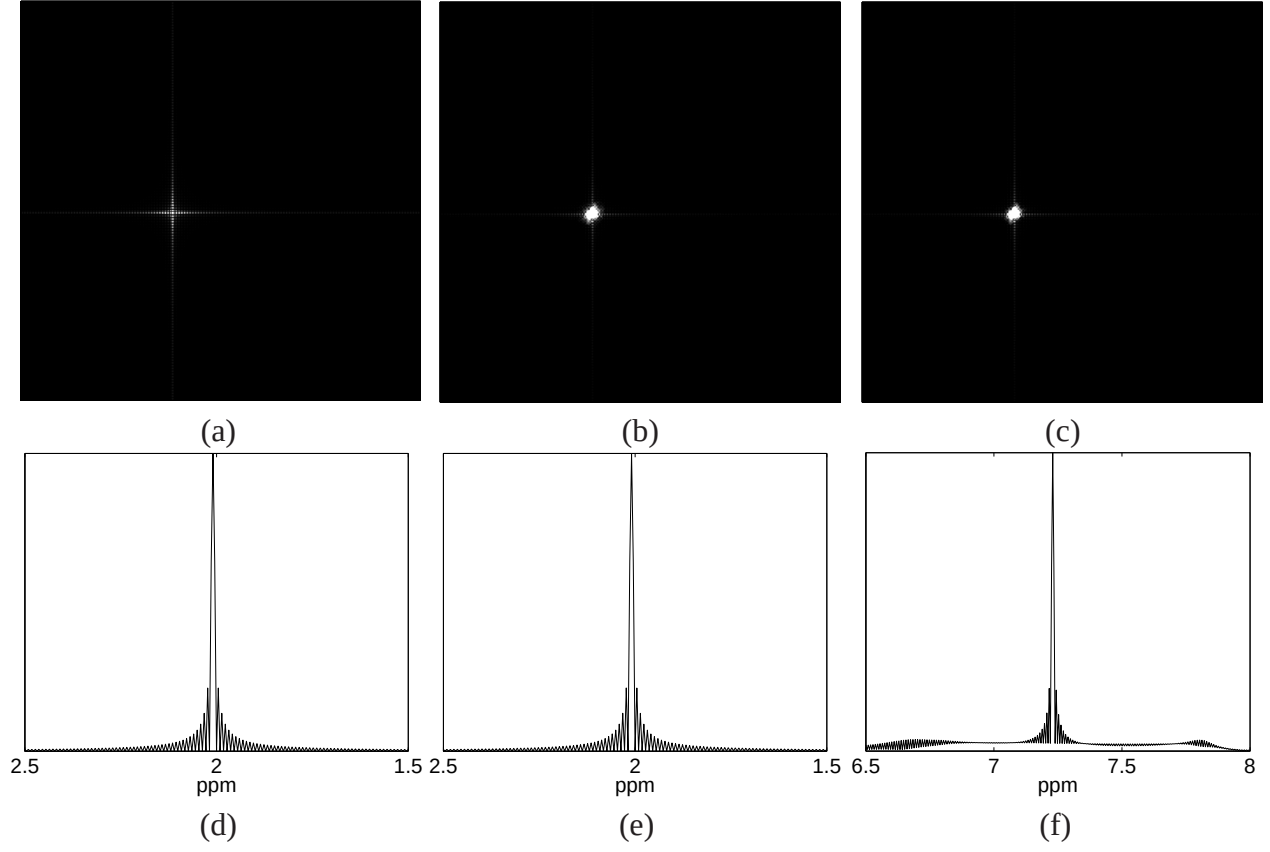


Figure 3.3: Spatial-spectral response functions $h_n(\mathbf{r}, f)$: top row shows $h_n(\mathbf{r}, f_0)$ of the (a) Fourier reconstruction, (b) spatial denoising, and (c) spatial-spectral denoising; bottom row shows $h_n(\mathbf{r}_0, f)$ of the (d) Fourier reconstruction (which has the same $h_n(\mathbf{r}_0, f)$ as the spatial denoising), (e) spatial-spectral denoising at $f_n = 2$ ppm, and (f) spatial-spectral denoising at $f_n = 7.23$ ppm. Notice that spatial-spectral reconstruction does not introduce a significant additional spatial leakage compared to the spatial reconstruction, but instead introduces some spectral leakage.

3.3 Discussion

3.3.1 Algorithm considerations

The proposed formulation for denoising of MRSI data described in (3.2) seeks a solution in $\mathbf{r}$ - t domain. The denoising problem can be similarly formulated in $\mathbf{k}$ - t domain. In this case, the corresponding penalized maximum-likelihood optimization problem becomes

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}^{\text{meas}} - \mathbf{s}\|_2^2 + \lambda R^{\text{spat}}(\mathbf{s}) + \beta R^{\text{spec}}(\mathbf{s}, \boldsymbol{\alpha}), \quad (3.14)$$

where $\mathbf{s}$ is now defined in $\mathbf{k}$ - t domain and

$$R^{\text{spat}}(\mathbf{s}) = \|\mathbf{B}\mathbf{F}_{\mathbf{k} \rightarrow \mathbf{r}}\mathbf{s}\|_2^2, \quad (3.15)$$

$$R^{\text{spect}}(\mathbf{s}, \boldsymbol{\alpha}) = \|\mathbf{C}(\boldsymbol{\alpha})\mathbf{s}\|_2^2 \quad (3.16)$$

with $\mathbf{F}_{\mathbf{k} \rightarrow \mathbf{r}}$ denoting the Fourier transform from $\mathbf{k}$ to $\mathbf{r}$ and $\boldsymbol{\alpha}$ defined in $\mathbf{k}$ - t domain. However, such a formulation may not be desirable because the “physical” space of signal properties inherent in MRSI data is actually $\mathbf{r}$ - t domain (spatial smoothing in $\mathbf{r}$ and linear predictability in t). Hence, it follows naturally to formulate the denoising problem in $\mathbf{r}$ - t domain. In addition, in the case of field inhomogeneity, formulating the denoising problem in $\mathbf{k}$ - t domain might not be straightforward. Particularly, it would be hard to choose the linear prediction order L because field inhomogeneity causes different frequency shifts of the spectral peaks at various locations $\mathbf{r}$, which in turn may lead to more peaks (due to averaging) in $\mathbf{k}$ domain. In such a case, L will not physically correspond to the number of the spectral peaks.

To illustrate the performance of the denoising approach in $\mathbf{k}$ - t domain in the absence of field inhomogeneity, Fig. 3.4 shows typical denoising results obtained by solving (3.14) for a low-resolution dataset used in Section 3.2.1. For comparison, Fig. 3.4 also shows the results obtained by formulating the denoising problem in $\mathbf{r}$ - t domain as described before in (3.2). The spatial regularization parameter λ was set to be the same in both cases, while the spectral regularization parameter β was chosen so that we obtain approximately the same spectral filtering level. Notice that with the same λ , the denoised solution obtained from filtering in $\mathbf{k}$ - t domain experiences more spatial blurring.

Spatial blurring of the denoising in $\mathbf{k}$ - t domain as compared to the denoising in $\mathbf{r}$ - t domain can be explained using SSRFs. Specifically, the SSRF for the denoising formulation in $\mathbf{k}$ - t domain can be defined similarly as in (3.11), Section 3.2.2. One can show that in this case the SSRF is

$$h_n(\mathbf{r}, f) = \sum_m \left(\sum_p \sum_q G_{l(p,q),m} e^{i2\pi \mathbf{k}_p \cdot \mathbf{r}_n} e^{i2\pi t_q f_n} \right) e^{-i2\pi \mathbf{k}_m \cdot \mathbf{r}} e^{-i2\pi t_m f}, \quad (3.17)$$

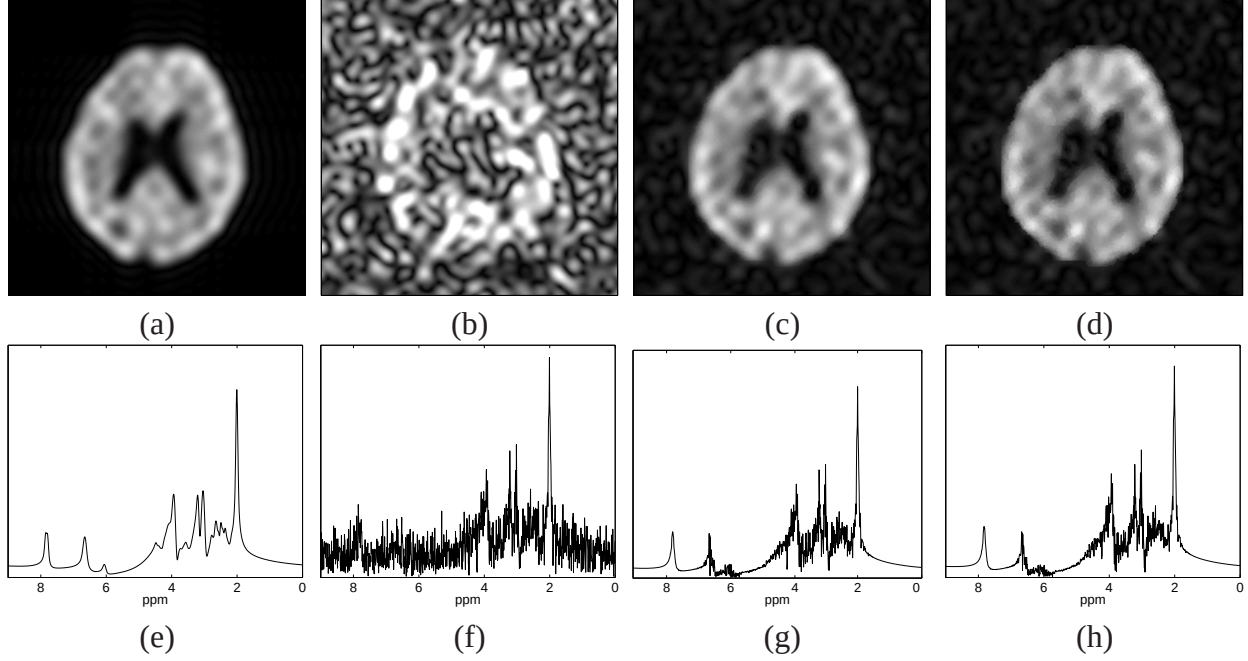


Figure 3.4: Denoising in k - t domain versus denoising in r - t domain: top row shows (a) noiseless and (b) noisy spatial distributions $\rho(\mathbf{r}, f_0)$ at $f_0 = 7.8$ ppm; corresponding denoised $\rho(\mathbf{r}, f_0)$ (c) in k - t and (d) in r - t ; bottom row shows (e) noiseless and (f) noisy spectral distributions $\rho(\mathbf{r}_0, f)$; corresponding denoised $\rho(\mathbf{r}_0, f)$ in (g) k - t and (h) r - t domain.

where $l(p, q)$ denotes the index $l = 1, \dots, MN$ that depends on both p and q . $G_{l(p,q),m}$ is the element of matrix

$$\mathbf{G} = \mathbf{I} + \lambda \mathbf{F}_{k \rightarrow r}^H \mathbf{B}^H \mathbf{B} \mathbf{F}_{k \rightarrow r} + \beta \mathbf{C}^H \mathbf{C}. \quad (3.18)$$

Notice the difference between (3.12) and (3.17). In (3.17) before multiplying with $\mathbf{G}$ that defines the filtering process of the algorithm, the DFT is performed in both spatial and spectral domains (terms $e^{i2\pi \mathbf{k}_p \cdot \mathbf{r}_n}$ and $e^{i2\pi t_q f_n}$, respectively), while in (3.12) the DFT is performed in the spectral domain (term $e^{i2\pi t_q f_n}$) only. This essentially results in an additional spatial blurring for the denoising in k - t . Figure 3.5 shows SSRFs computed for the denoising in k - t domain and in r - t . One can see that although both SSRFs reduce Gibbs ringing compared to the conventional Fourier reconstruction, the main lobe of the SSRF for the denoising in k - t has a larger spatial extent compared to that of the denoising in r - t , which explains the typical spatial blurring effect in Figs. 3.4(c),(g). In addition, the SSRF for the denoising in k - t has more spectral leakage as can

be seen from Figs. 3.4(e) and (f).

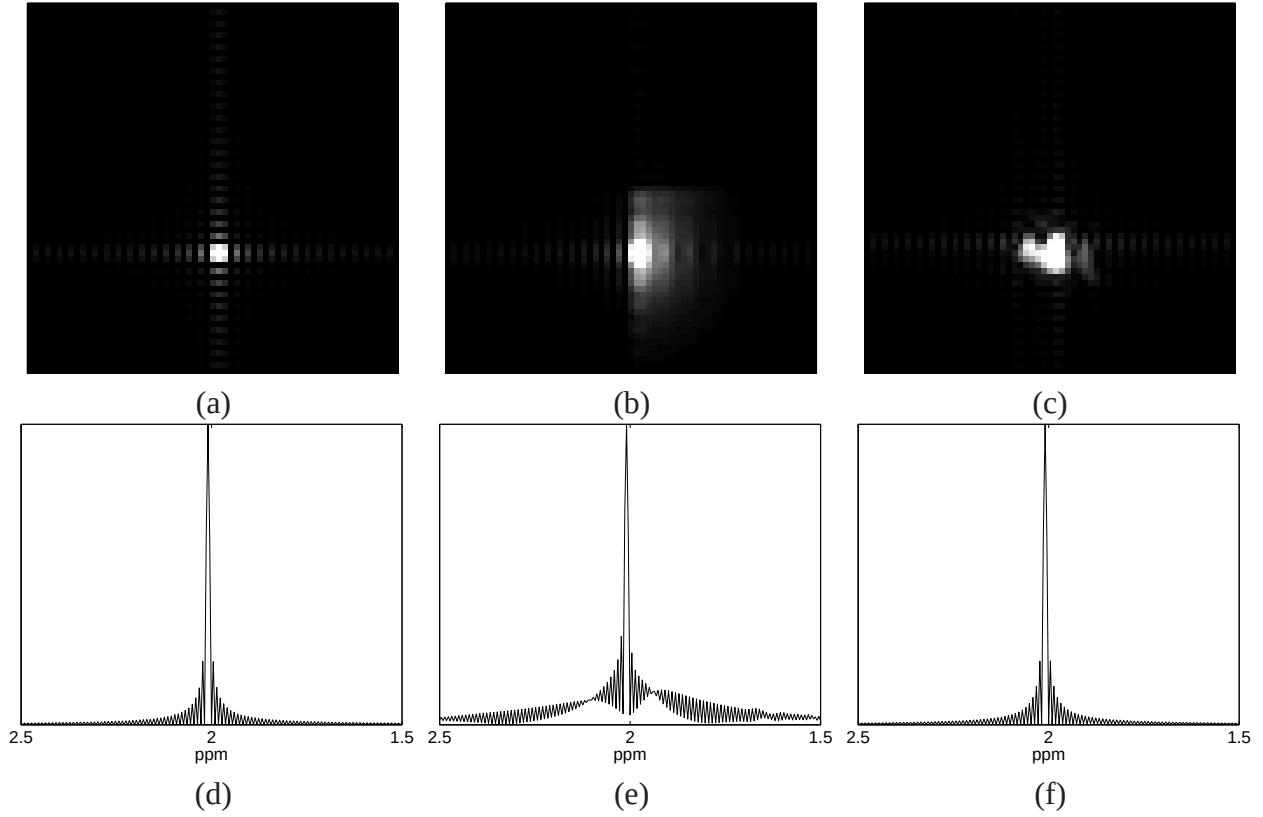


Figure 3.5: SSRFs for denoising in k - t domain versus denoising in r - t domain: top row shows $h_n(\mathbf{r}, f_0)$ for the (a) Fourier reconstruction, (b) denoising in k - t , (c) denoising in r - t ; bottom row shows $h_n(\mathbf{r}_0, f)$ for the (d) Fourier reconstruction, (e) denoising in k - t , (f) denoising in r - t . Notice that the main lobe of the SSRF for the denoising in k - t has a larger spatial extent, resulting in an additional spatial blurring during the filtering process compared to the denoising in r - t . Here, $\mathbf{r}_0$ was chosen to be the voxel corresponding to the highest intensity in (a) and $f_0 = 2$ ppm. The SSRF $h_n(\mathbf{r}, f)$ was computed such that the corresponding $\mathbf{r}_n$ and f_n match $\mathbf{r}_0$ and f_0 . Figures in each row are shown on the same scale.

3.3.2 Estimation of linear prediction coefficients

Since linear prediction coefficients α determine resonance frequencies and lineshape parameters of the spectral peaks, it is important to estimate these coefficients as accurately as possible in the presence of noise. Among numerous developed techniques, we consider the TLS, TLS combined with Cadzow formulation (TLS Cadzow), and IQML methods. To assess the performance of these methods, we simulate the spectrum with Lorentzian lineshape, Gaussian lineshape, and Gaussian

lineshape mixed with the baseline. For each case, we generate 1024 realizations of the complex Gaussian zero-mean noise. Noise variance was defined such that the SNR of the resulting noisy spectrum was 20 dB, where the SNR is defined as

$$\text{SNR}(\mathbf{s}^{\text{meas}}) = 20 \log_{10} \left(\frac{|A|}{\sigma_0} \right) \text{dB}, \quad (3.19)$$

with $|A|$ denoting the amplitude of the largest peak in the spectrum and σ_0 is the standard deviation of the noise. Figure 3.6 shows the original and noisy spectra for each considered scenario.

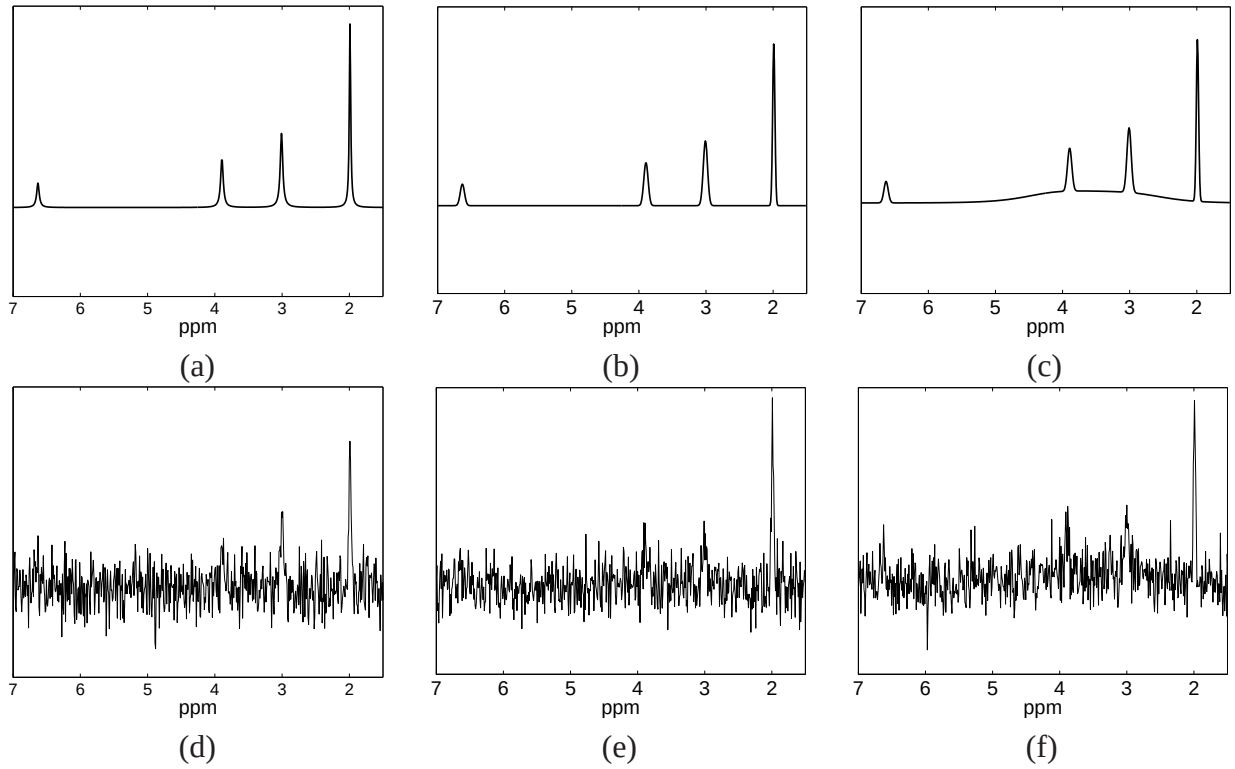


Figure 3.6: Simulated data used in the study on the estimation of linear prediction coefficients: top row shows simulated noiseless spectra in the case of (a) Lorentzian lineshape, (b) Gaussian lineshape, and (c) Gaussian lineshape with a baseline; bottom row shows the corresponding representative noisy spectra.

Estimation of α was performed based on the first 64 noisy time points and using the LS, TLS, TLS Cadzow, and IQML methods. Relative error of the estimated $\hat{\alpha}$ at each noise realization was then

computed as

$$\text{Err}(\boldsymbol{\alpha}) = 20 \log_{10} \left(\frac{\|\boldsymbol{\alpha}\|_2}{\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2} \right) \text{dB} \quad (3.20)$$

and the mean (over noise realizations) of these errors is shown in Table 3.2. For a visual assessment of how the estimation of $\boldsymbol{\alpha}$ affected the denoising performance, Fig. 3.7 shows the corresponding weights w_k computed according to (3.9) using the estimated $\boldsymbol{\alpha}$. We observed that in all different signal lineshape scenarios, all three later harmonic retrieval techniques (TLS, TLS Cadzow, and IQML) increase accuracy of estimating $\boldsymbol{\alpha}$ significantly, compared to the LS estimations. First iterations of the TLS Cadzow and IQML techniques coincide with the TLS method, while performing further iterations improves the accuracy of estimation. This improvement was observed to be significant in the case when the signal followed a Lorentzian lineshape model. In the case of the Gaussian lineshape with the baseline, the performance improvements of the TLS Cadzow and IQML techniques compared to the TLS method were the least significant.

From the above experiment, we see that a conventional LS estimation of linear prediction coefficients did not yield an acceptable accuracy performance in the presence of high noise level. The choice of a particular estimation method among the three considered methods (TLS, TLS Cadzow, and IQML) was observed to be dependent on the signal lineshape; i.e., if the spectral lineshapes are Lorentzian, then the TLS Cadzow and IQML techniques performed the best and IQML was observed to be the most stable. However, if the signal is not Lorentzian, then both TLS Cadzow and IQML techniques yielded on average less significant improvements than the TLS estimation.

3.3.3 Choice of regularization parameters

Choice of regularization parameter has been extensively analyzed in the literature for a general regularization-based reconstruction [117–121]. In our case, increasing the spatial regularization parameter λ helps to reduce noise variance non-uniformly in the spatial domain and uniformly in the spectral domain. Over-regularization introduces a spatial over-smoothing effect and can lead

Table 3.2: Comparison of estimating linear prediction coefficients using different harmonic retrieval techniques at $\text{SNR}(\rho^{\text{meas}})=17$ dB: mean relative errors (in dBs), computed according to (3.20) and averaged over 1024 noise realizations.

Case	LS estimation	TLS Prony	TLS Prony Cadzow	IQML
Lorentzian lineshape	0.69	4.16	11.54	11.69
Gaussian lineshape	1.00	7.62	26.93	24.73
Gaussian lineshape and baseline	1.49	10.30	15.56	11.21

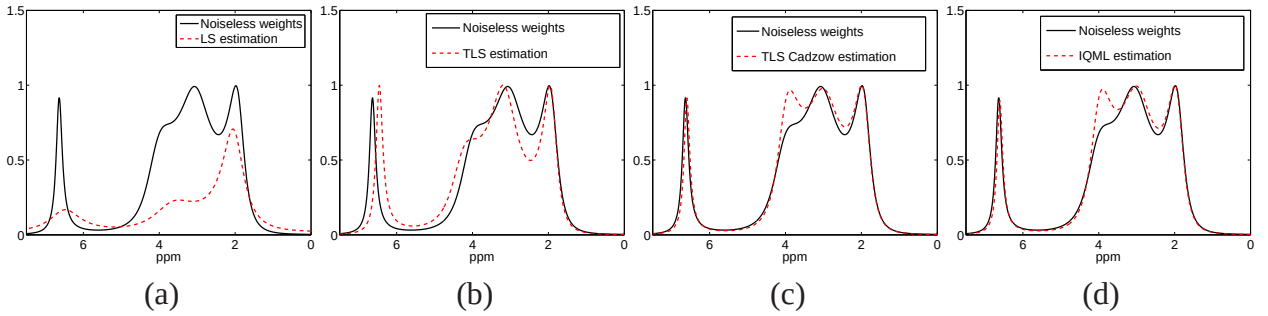


Figure 3.7: Weights w_k derived from the linear prediction coefficients, estimated using various techniques: (a) conventional LS, (b) TLS, (c) TLS Cadzow, and (d) IQML methods. Ground truth weights in the noiseless case are shown in red. Notice that the conventional LS estimation did not yield an acceptable performance, while the state-of-the-art techniques improved the estimation.

to a reduction of the spectral peak amplitudes. For the spectral regularization, increasing β too much asymptotically implies a strict assumption of the Lorentzian lineshape signal model. There are multiple criteria that can be used to choose regularization parameters. One way is to choose parameters based on a trade-off between the resulting SNR and spatial-spectral leakage of the reconstruction. However, due to the coupling of spatial and spectral dimensions, computing noise variance over both spatial and spectral dimensions is time-consuming. If the goal is to obtain a rough initial guess for regularization parameters, as it is often the case, a simplified approach can be used where spatial and spectral dimensions are considered separately. Selecting a regularization parameter λ for the spatial denoising was addressed in [72, 113], where λ was chosen based on the desired trade-off between the corresponding spatial response function [113] and SNR. Here, we focus on the choice of the spectral regularization parameter β .

Spectral regularization parameter can be chosen based on (i) the maximum allowed level of spectral leakage, (ii) the desired level of noise variance reduction, (iii) the trade-off between noise reduction and amount of lineshape distortions. The first choice follows straightforwardly from Section 3.2.2. The second choice can be realized by noting that in the case of a complex white Gaussian noise $\mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I})$ and ignoring spatial regularization, the noise covariance matrix of the reconstruction is given by

$$\Sigma = \sigma_0^2 \mathbf{F}_{t \rightarrow f}^H (\mathbf{I} + \beta \mathbf{C}^H \mathbf{C})^{-1} (\mathbf{I} + \beta \mathbf{C}^H \mathbf{C})^{-H} \mathbf{F}_{t \rightarrow f}, \quad (3.21)$$

where $\mathbf{F}_{t \rightarrow f}$ is the DFT matrix transforming from t to f . Approximating $\mathbf{C}$ with a circulant matrix, one can show that noise after reconstruction is uncorrelated and at each k -th spectral point the noise variance relates to β as:

$$\begin{aligned} \sigma_k^2 &= \sigma_0^2 w_k^2 \\ &= \sigma_0^2 \frac{1}{(1 + \beta |h_k|^2)^2}, \quad \text{for } k = 1, \dots, M, \end{aligned} \quad (3.22)$$

where h_k is defined as in (3.10). Thus, using (3.22) one can choose β to yield a desirable noise variance reduction.

The third choice of β based on the trade-off between noise reduction and amount of lineshape distortions can be realized using elliptical constraints, originally proposed in [122]. Specifically, ignoring again spatial regularization, (3.2) becomes

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}^{\text{meas}} - \mathbf{s}\|_2^2 + \beta \|\mathbf{C}(\boldsymbol{\alpha}) \mathbf{s}\|_2^2. \quad (3.23)$$

If $\beta = 0$, the reconstruction is the same as the noisy measured data, while in the other extreme as β approaches infinity, the reconstructed spatial-spectral distribution has strictly Lorentzian lineshapes, hence distorting the spectral lineshapes if the true MR signal is not Lorentzian. Thus, β acts as a balancing factor between the noise level and the lineshape distortion. As a criterion for

the noise level, we consider the linear prediction error. The noisier the signal, the less correlation structure it has, and hence the less predictive it is. Conversely, a linear predictive signal is typically smooth with infinite SNR. As a criterion for the lineshape distortion, we consider the data consistency, by noting that although the measured data is noisy, it contains the underlying true lineshape, while any denoised data may have distorted lineshapes, unless it is a perfect reconstruction. Thus, the problem of choosing β to balance between the SNR and lineshape distortions becomes the problem of balancing between the linear prediction error and data consistency. For a desired reconstruction, we want to satisfy the following conditions:

$$\|C\mathbf{s}\|_2^2 \leq \epsilon_{\text{lp}}^2, \quad (3.24)$$

$$\|\mathbf{s} - \mathbf{s}^{\text{meas}}\|_2^2 \leq \epsilon_{\text{d}}^2, \quad (3.25)$$

where ϵ_{lp} and ϵ_{d} are the thresholds on the level of SNR and lineshape distortions, respectively. In general, there is a family of solutions satisfying both (3.24) and (3.25). Following [122], we notice that a particular solution can be found as

$$\boldsymbol{\rho}_c = \left(\mathbf{I} + \frac{\epsilon_{\text{d}}^2}{\epsilon_{\text{lp}}^2} \mathbf{C}^H \mathbf{C} \right)^{-1} \boldsymbol{\rho}^{\text{meas}}, \quad (3.26)$$

which corresponds to

$$\beta = \epsilon_{\text{d}}^2 / \epsilon_{\text{lp}}^2. \quad (3.27)$$

For the estimates of ϵ_{d}^2 and ϵ_{lp}^2 we heuristically make use of a spatially denoised reconstruction at a chosen λ . Specifically, ϵ_{d} is the data consistency error of the Lorentzian approximation of the spatially denoised solution $\hat{s}_{\lambda \neq 0, \beta=0}(\mathbf{r}, t)$ and ϵ_{lp} is the corresponding linear prediction error. This approach to select β was applied to the simulations discussed in Section 3.4.

3.3.4 Single-voxel quantitation evaluation

Denoising of MRSI data is an initial step towards the ultimate goal of quantifying metabolite relative concentrations; thus, an important question is: Does the weighting effect of spectral denoising improve spectral quantitation beyond improving visual quality of the spectra? The answer depends on several factors such as the spectral characteristics of the signal, the parameters of the proposed denoising method (e.g., method used to estimate linear prediction coefficients, regularization parameters, linear prediction order), and the quantitation method used.

We focus on a simple single-voxel case to illustrate the effect of the spectral denoising in three cases: (i) spectral lineshapes are Lorentzian (ideal condition); (ii) spectral lineshapes are Gaussian and a baseline is included into the spectrum; (iii) spectral lineshapes are Gaussian with a strong overlap. For each case 64 noise realizations were considered. Linear prediction coefficients were estimated using the TLS Cadzow method on the first 64 data points. In the absence of a baseline, metabolite quantitation was formulated as solving the following optimization problem:

$$\{\hat{\mathbf{c}}, \hat{\mathbf{d}}\} = \arg \min_{\mathbf{c}, \mathbf{d}} \|\mathbf{s} - \mathbf{Z}(\mathbf{d})\mathbf{c}\|_2, \quad (3.28)$$

where $\mathbf{c}$ and $\mathbf{d}$ are the metabolite amplitude and lineshape parameters, respectively. Prior information on the peak resonance frequencies f_n was incorporated into $\mathbf{Z}$ as

$$Z_{m,n} = e^{-i2\pi f_n t_m} e^{-d_n t_m}, \quad (3.29)$$

if the lineshapes are Lorentzian, or

$$Z_{m,n} = e^{-i2\pi f_n t_m} e^{-d_n t_m^2}, \quad (3.30)$$

if the lineshapes are Gaussian.

In the presence of a baseline, metabolite quantitation was formulated as solving the following

optimization problem:

$$\{\hat{\mathbf{c}}, \hat{\mathbf{b}}, \hat{\mathbf{d}}\} = \arg \min_{\mathbf{c}, \mathbf{b}, \mathbf{d}} \|\mathbf{s} - \mathbf{Z}(\mathbf{d})\mathbf{c} - \mathbf{\Upsilon}\mathbf{b}\|_2, \quad (3.31)$$

where $\mathbf{b}$ denotes the baseline coefficients. The baseline component of the MRS signal was represented using B-spline basis functions in the frequency domain and then Fourier-transformed to the time domain, forming matrix $\mathbf{\Upsilon}$. In all cases, the corresponding minimization problem was solved jointly over the parameter space using Levenberg-Marquardt non-linear optimization routine in Matlab version 7.11.0.584 (R2010b).

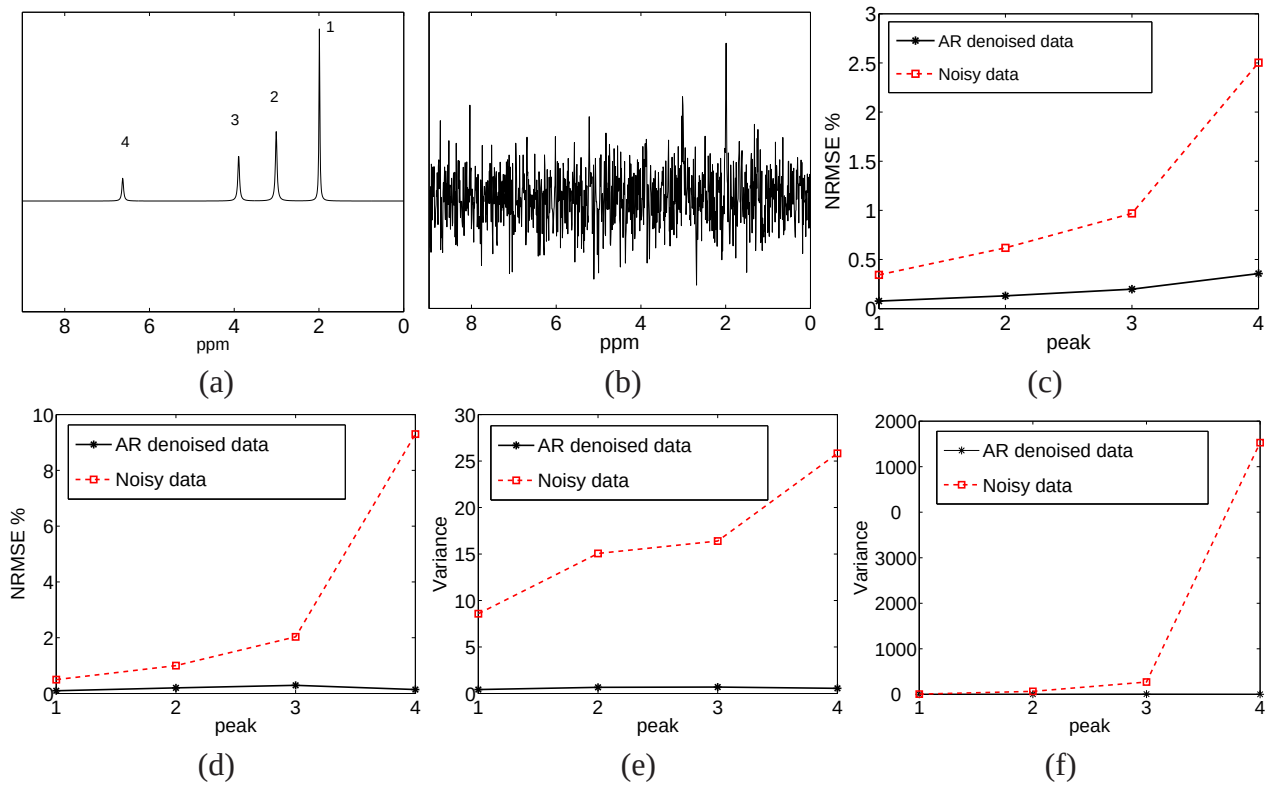


Figure 3.8: Single-voxel quantitation from the noisy vs. the denoised data in the case of Lorentzian spectral lineshapes: (a) original spectrum; (b) noisy spectrum; NRMSEs for the (c) amplitudes and (d) lineshape parameters obtained by quantitation from noisy and denoised data; variance for the (e) amplitudes and (f) lineshape parameters obtained by quantitation from noisy and denoised data.

We consider first a case of Lorentzian spectral lineshapes. The spectrum consists of 4 peaks, shown in Fig. 3.8(a), and a representative noisy spectrum is shown in Fig. 3.8(b). Figures 3.8(c)-(f) compare a single-voxel quantitation from the proposed denoised data with a quantitation obtained

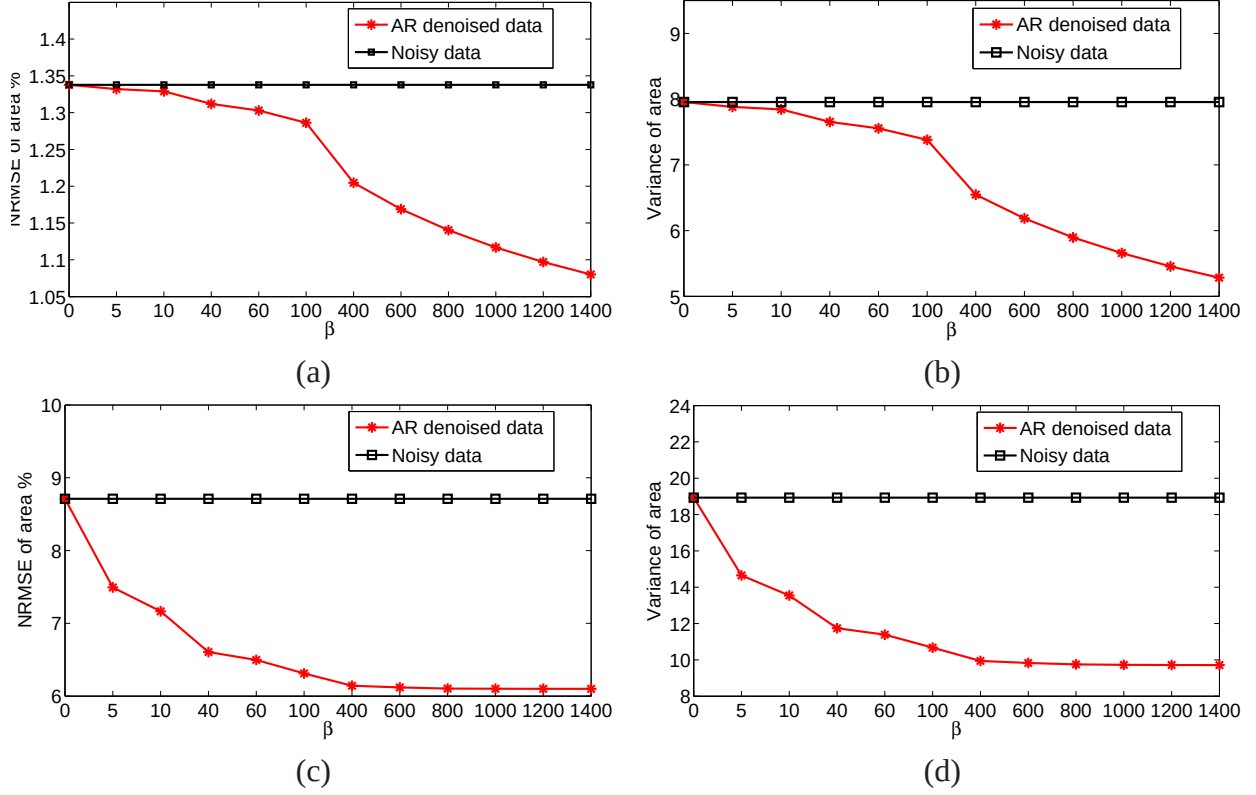


Figure 3.9: Single-voxel quantitation from the noisy vs. the denoised data as a function of β in the case of Lorentzian spectral lineshapes. (a) NRMSE and (b) variance for the estimated area of peak 1; (c) NRMSE and (d) variance for the estimated area of peak 4. As β increases, the proposed denoising achieves smaller NRMSEs and variances of estimated metabolite areas.

directly from the noisy data in terms of the normalized root mean square error (NRMSE) and variance of the estimated amplitude and lineshape parameters. We observed that the proposed denoising helped to achieve both smaller NRMSEs and variances of the estimated parameters. Notice that the improvement was most significant for peak 4, which has the lowest SNR. Because the spectral lineshape is Lorentzian, it is expected that as the spectral regularization parameter increases, the improvement in NRMSEs and variances of metabolite areas obtained from the proposed denoising would be more significant. This is illustrated in Fig. 3.9, which shows the estimated areas of two representative peaks: peak 1 with the highest SNR and peak 4 with the lowest SNR. Areas were computed based on the estimated amplitudes and lineshape parameters.

Next, the case of non-Lorentzian spectral lineshapes is considered. Specifically, the spectrum consists of 4 Gaussian peaks and a baseline, as shown in Fig. 3.10(a). Figures 3.10 and 3.11

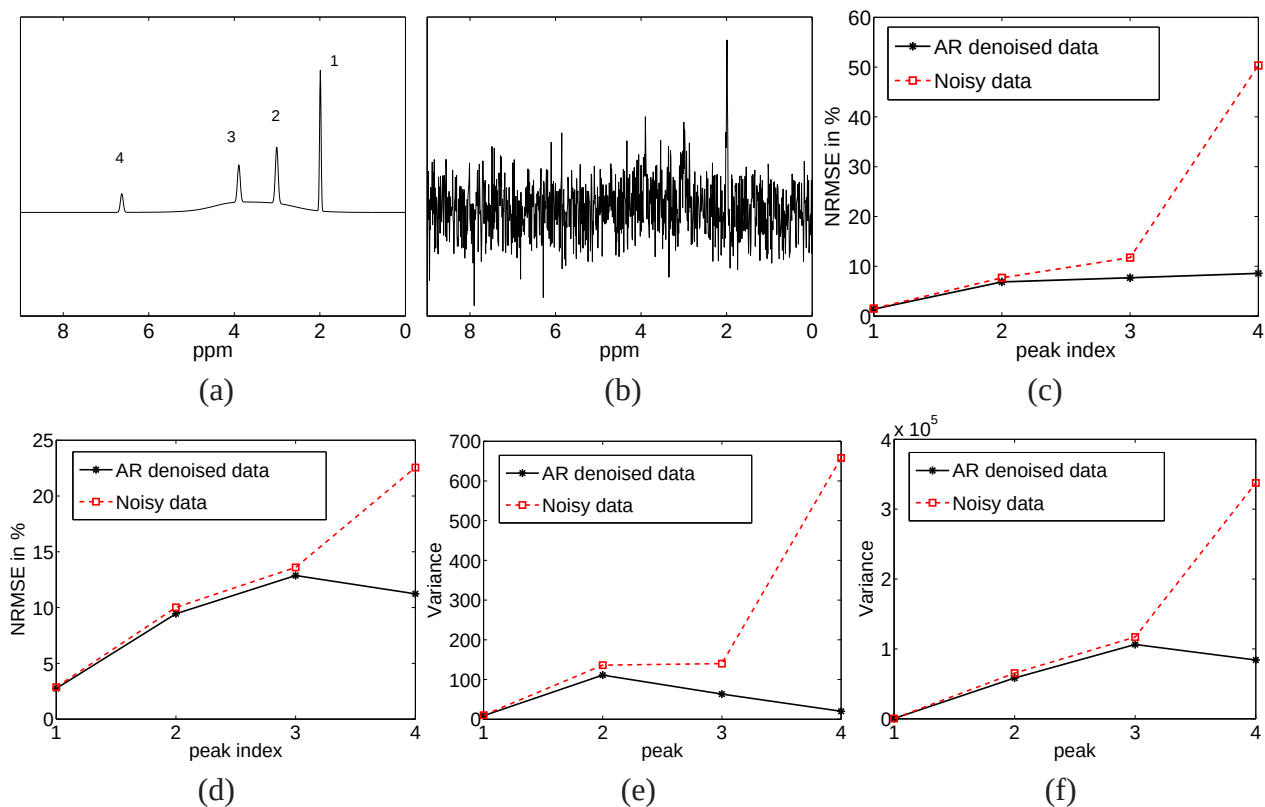


Figure 3.10: Single-voxel quantitation from the noisy vs. the denoised data in the case of non-Lorentzian spectral lineshapes: (a) original spectrum; (b) noisy spectrum; NRMSEs for the (c) amplitudes and (d) lineshape parameters obtained by quantitation from the noisy and denoised data; variance for the (e) amplitudes and (f) lineshape parameters obtained by quantitation from the noisy and denoised data.

illustrate the performance of metabolite quantitation from the denoised data. In this case, choosing overly large β does not yield decreased NRMSEs of metabolite parameters (although variances decrease) because we are biasing the signal lineshapes to a Lorentzian model. However, within a range of medium values of β the proposed denoising improves the quantitation accuracy. Notice that for peak 1, which has the highest SNR, quantitation accuracy of the estimated area is slightly improved, while for peak 4, which has the lowest SNR, the improvement in quantitation accuracy is remarkable.

Lastly, we consider the case of a strong overlap of Gaussian lineshape peaks. The original spectrum, shown in Fig. 3.12(a), has four peaks at the same resonance frequencies as in the previous studies, and two additional peaks simulating the spectral overlap. Specifically, peaks 2

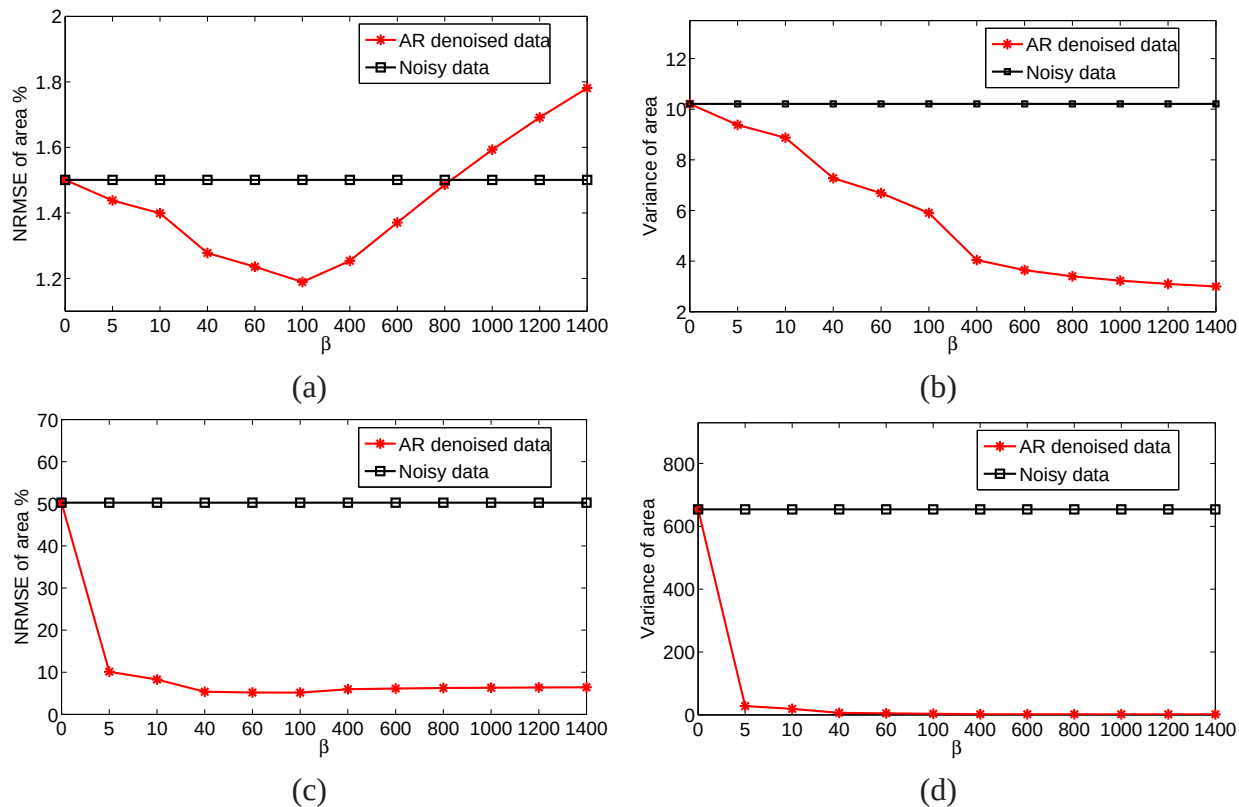


Figure 3.11: Single-voxel quantitation from the noisy vs. denoised data as a function of β in the case of non-Lorentzian spectral lineshapes: (a) NRMSE and (b) variance for the estimated area of peak 1; (c) NRMSE and (d) variance for the estimated area of peak 4. For a range of medium values of β , the proposed denoising reduces NRMSEs of metabolite areas compared to the quantitation directly from the noisy data. This improvement is more significant for a lower-SNR peak (peak 4). Notice that variance of the estimated area for each peak decreases as β increases.

and 5 strongly overlap with each other, peaks 1 and 6 have close resonance frequencies, and peaks 3 and 4 are relatively well separated from the other peaks. A representative noisy spectrum is shown in Fig. 3.12(b). Quantitation study shows that the proposed denoising improves quantitation accuracy of peak 4 the most. This is explained due to the fact that peak 4 is the most isolated peak from the other peaks in the spectrum. Improvement in the quantitation accuracy of peak 3 was less than that of peak 4, while improvement in the quantitation accuracy of peaks 1 and 2 was insignificant. In the case of peak 5, which strongly overlaps with peak 2, the proposed denoising slightly decreased the quantitation accuracy of both amplitude and lineshape parameters (see Fig. 3.12(c)). Notice that although peak 6 is closely spaced to peak 1, because peak 6 has an extremely

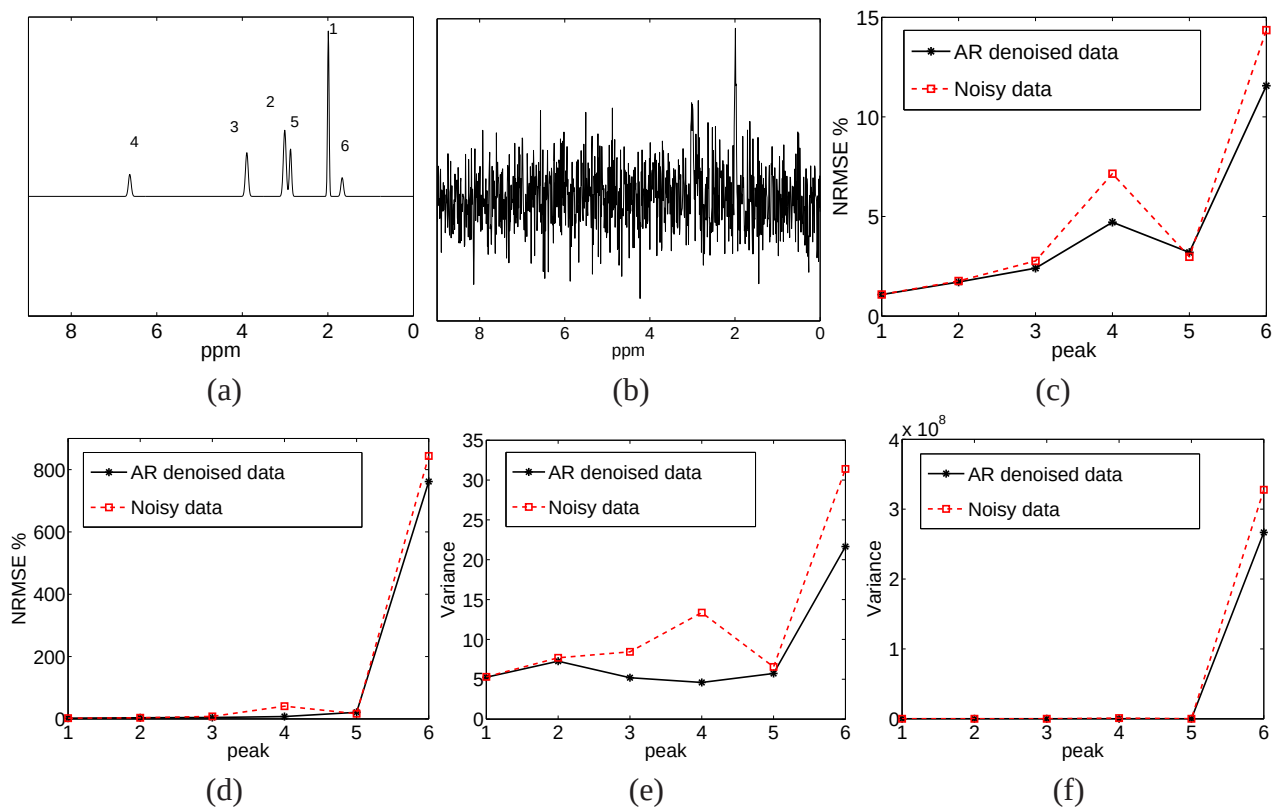


Figure 3.12: Performance of a single-voxel quantitation from the noisy vs. denoised data in the case of overlapping Gaussian spectral lineshapes: (a) original spectrum; (b) noisy spectrum; NRMSE for the estimated (c) amplitudes and (d) lineshape parameters obtained by quantitation from noisy (red) and denoised (black) data; variance for the estimated (e) amplitudes and (f) line-shape parameters obtained by quantitation from noisy (red) and denoised (black) data.

low SNR, the proposed denoising still improves the quantitation accuracy of this peak significantly.

The above single-voxel quantitation studies suggest that in the case of well-separated peaks, the proposed spectral denoising helps to improve the accuracy of metabolite quantitation. Specifically, if the lineshapes are Lorentzian, then the larger the spectral regularization parameter β , the more quantitation accuracy improvement can be achieved. Otherwise, a range of medium values of β can yield the accuracy improvement. If the peaks are closely located and have a significantly low SNR, then the AR denoising may still improve the quantitation accuracy. However, if the peaks strongly overlap, then the AR denoising does not achieve an improvement on the quantitation accuracy, and in some cases, it can degrade the quantitation performance.

3.4 Experimental results

3.4.1 Simulations

To illustrate the performance of the proposed method, we simulate a spatial-spectral distribution function based on literature values of ^1H metabolites and experimental data. We consider 5 commonly MR-observable metabolites in the human brain [7, 17]: *N-acetylaspartate (NAA)*, *creatine (Cr)*, *choline (Cho)*, *glutamate/glutamine (Glx)*, and *myo-inositol (m-Ins)*. For each metabolite, a spectral profile was obtained from quantum mechanical simulations of a spin-echo MR experiment [49]. The spatial distribution of each metabolite was created based on commonly reported literature values from 3 segmented regions of cerebrospinal fluid, grey matter, and white matter [17, 123, 124]. To evaluate the ability of the proposed method to handle non-ideal conditions, Gaussian (instead of pure Lorentzian) lineshape was used with an additional baseline signal. Lineshape parameters were chosen based on commonly reported values in the literature [125, 126]. The baseline signal was extracted from single-voxel CSI PRESS experimental data of the brain with $\text{TE}=30$ ms. MRSI data were simulated at the magnetic field strength of 3 T with a spectral bandwidth of 1,200 Hz and complex Gaussian noise was added.

The spatial regularization parameter λ was chosen so that a factor of around 3 was gained in the SNR, while β was first chosen based on the elliptical constraint in (3.27) and then slightly increased (by a factor of 1.4). Since the lineshape variations were not significant across the voxels, linear prediction coefficients were estimated from the k -space center using the TLS estimation technique. The linear prediction order L was chosen to be 8.

Figures 3.13 and 3.14 show the spatial distributions of the obtained denoised spatial-spectral function at a low-SNR frequency in the metabolite-baseline region at 4.86 ppm and a low-SNR resonance frequency of the metabolite peak at 7.8 ppm, respectively. We compared the proposed approach with the wavelet soft shrinkage. The threshold for the wavelet denoising method was chosen to yield the same level of data consistency as for the proposed approach. Notice that because of a high noise level, wavelet shrinkage improves SNR of the data at the expense of blurring

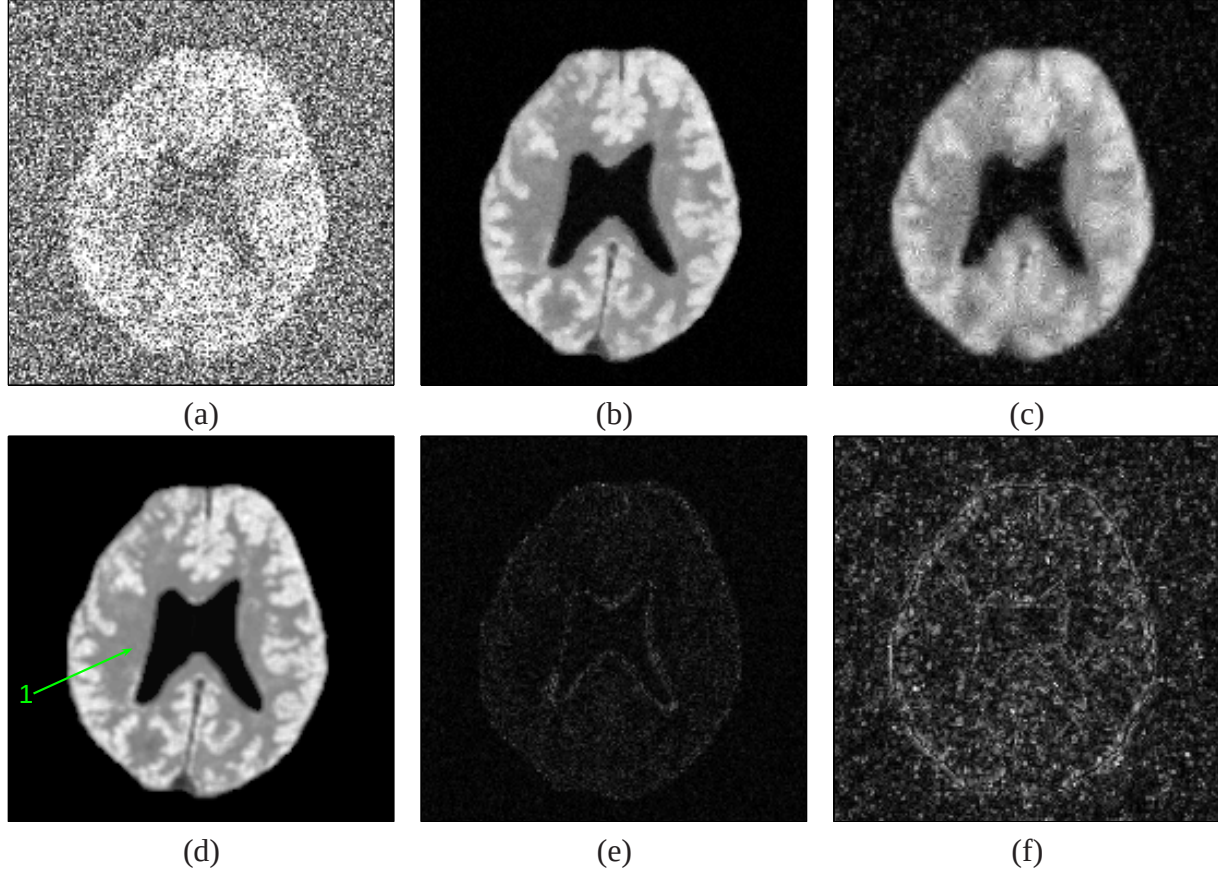


Figure 3.13: Denoising results - spatial distributions $\rho(\mathbf{r}, f_0)$ at a low-SNR frequency in the metabolite-baseline region (4.86 ppm): $\rho(\mathbf{r}, f_0)$ obtained from (a) noisy data; (b) proposed denoising; (c) wavelet shrinkage; (d) noiseless data; (e) corresponding error of (b); (f) corresponding error of (c). Error maps in (e) and (f) were scaled by a factor of 0.4 with respect to the figures in (a)-(d). Wavelet denoising and the proposed method have approximately the same data consistency level for the reconstructed 3-D distribution.

spatial features significantly, while spatial-spectral denoising preserves these features better. The proposed method achieves a better denoising performance for the peak at 4.86 ppm than for the peak at 7.8 ppm due to the smaller denoising weight at 4.86 ppm than that at 7.8 ppm (see Fig. 3.15 (d)).

Figure 3.15 shows representative denoised spectra from a particular voxel, marked by an arrow in Fig. 3.13(d). From the denoising weights computed as in (3.9) and shown in Fig. 3.15(d), we observe that these weights predict accurately the frequency bands which do not have major metabolite content. Compared to the wavelet denoising, the proposed denoising approach reduces

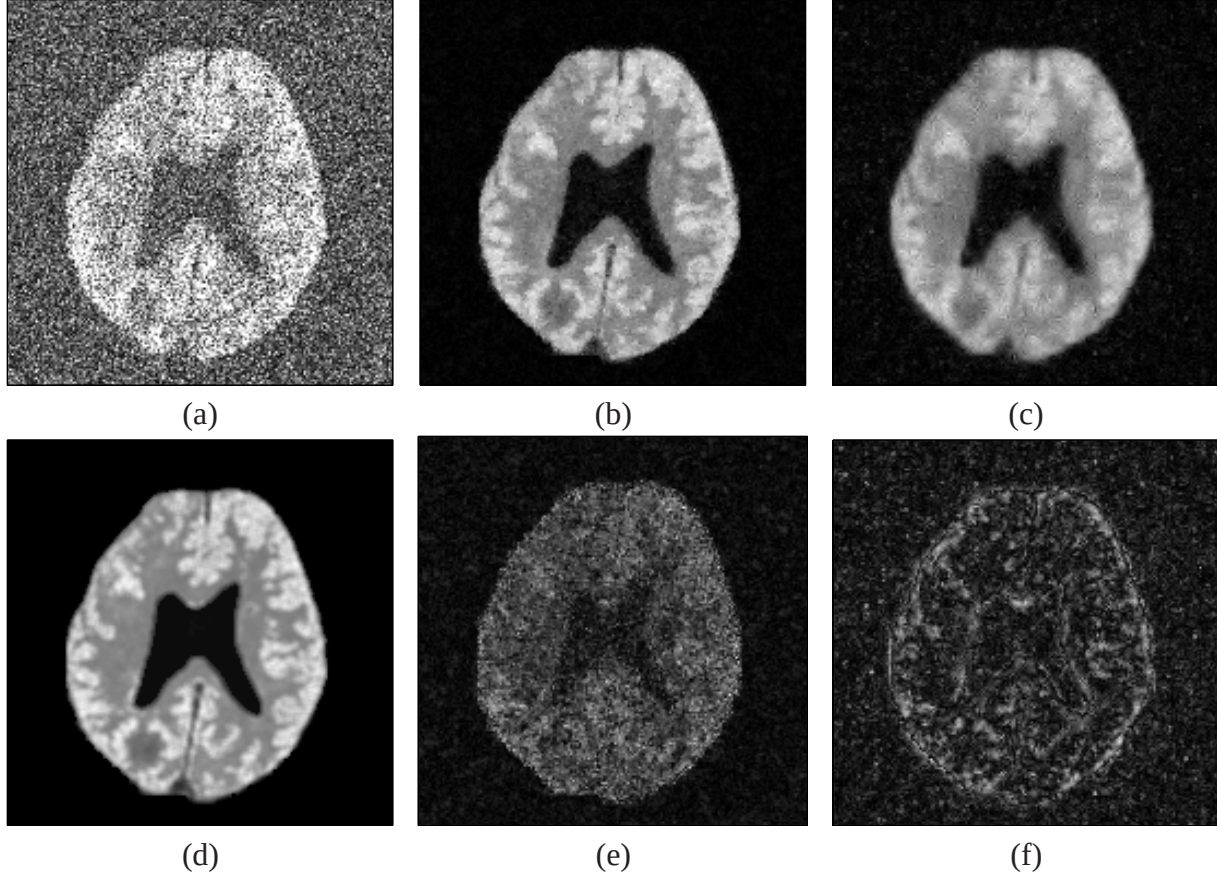


Figure 3.14: Denoising results - spatial distributions $\rho(\mathbf{r}, f_0)$ at a low-SNR frequency at 7.8 ppm. $\rho(\mathbf{r}, f_0)$ obtained from (a) noisy data; (b) proposed denoising; (c) wavelet shrinkage; (d) noiseless data; (e) corresponding error of (b); (f) corresponding error of (c). Error maps in (e) and (f) were scaled by a factor of 0.4 with respect to the figures in (a)-(d). Wavelet denoising and the proposed method have approximately the same data consistency level for the reconstructed 3-D distribution.

the noise variance in the frequency regions where the weights are small (see Figs. 3.15(e),(f)), but could not suppress noise in the other regions (as expected).

3.4.2 *In vivo* experiment

We have applied the proposed denoising technique on *in vivo* MRSI data of the mouse brain, acquired using a Varian INOVA 11.74 T (500 MHz) MRI scanner at Washington University in St. Louis. This dataset was previously described and used in [72]. A craniectomy was performed on the mouse and the middle cerebral artery (MCA) was electrocoagulated. *In vivo* MRSI data was acquired using a CSI sequence with CHES water suppression, TE=270 ms, TR=1500 ms,

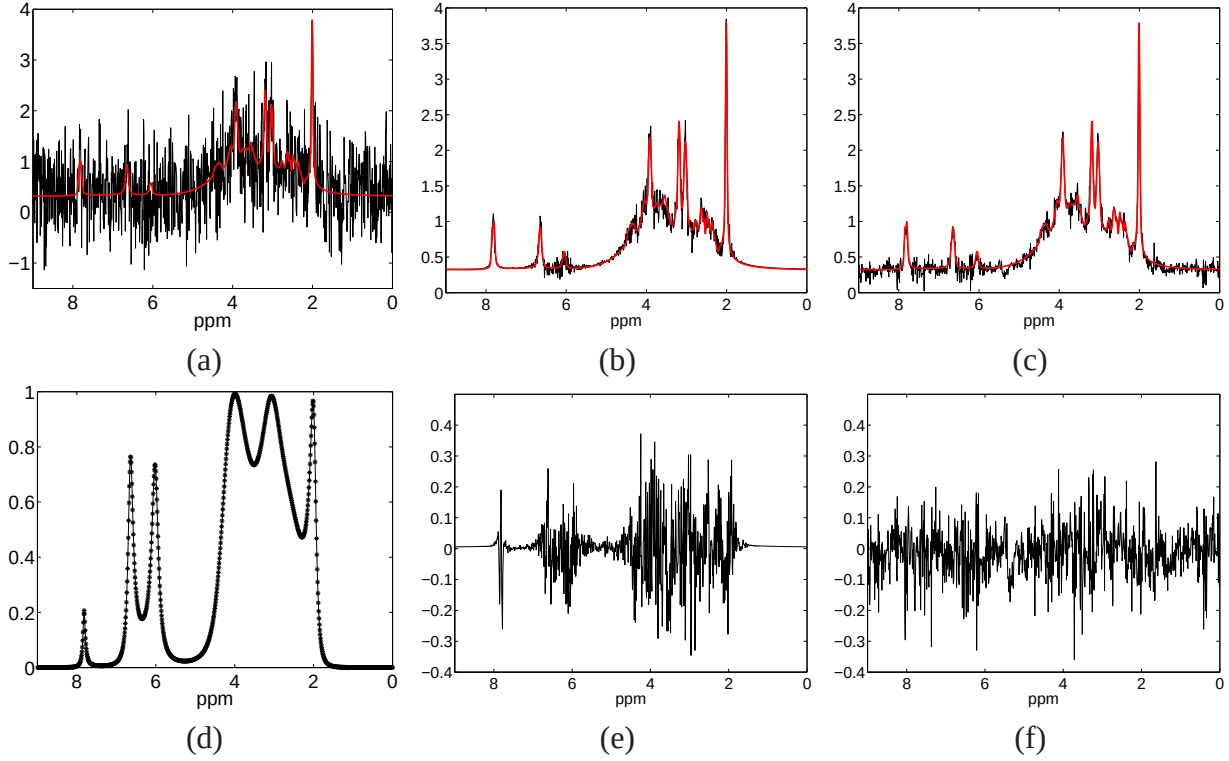


Figure 3.15: Denoising results - spectra from a particular voxel, marked by an arrow in Fig. 3.13(d): (a) noisy spectrum; (b) spectrum obtained from the proposed denoising; (c) spectrum obtained from wavelet shrinkage; (d) denoising weights w_k for (b); (e) error of (b); (f) error of (c). Original spectrum is displayed in red.

bandwidth=6000 Hz, 1024 FID data points, and 8 averages. A co-registered anatomical image was acquired using the spin-echo sequence with FOV=2.2 cm $\times$ 2.2 cm, slice thickness = 1.5 mm, matrix size=128 $\times$ 128, TE=40 ms, TR=1500 ms. The CSI dataset was pre-processed to compensate for field inhomogeneity and remove water resonance using HSVD [21]. Notice that the residual water was not completely removed. For a detailed description of the dataset, see [72].

Spatial-spectral denoising was performed with linear prediction order 4 to capture the dominant peaks present in the spectrum: NAA, Cr, Lac, and residual water. Linear prediction coefficients were estimated from a rank-9 approximation of the noisy k - t data using first 512 FID data points. Figure 3.16 shows pre-computed denoising weights at a representative voxel 1 where the electrocoagulation was performed and at a voxel 2 in the lower part of the brain far from the region of the electrocoagulation (see Fig. 3.16(a) for the indicated voxels). Electrocoagulation blocked the

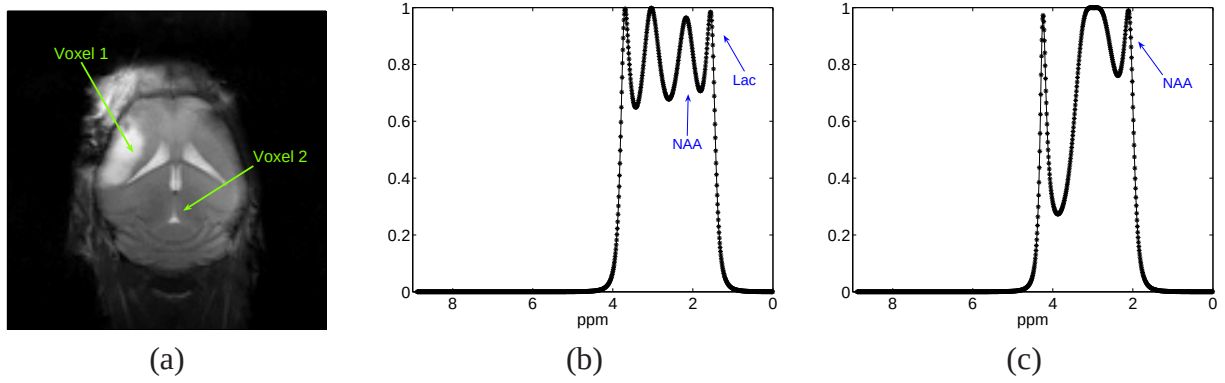


Figure 3.16: *In vivo* experiment: (a) reference image; denoising weights w_k for (b) voxel 1 in the region of electrocoagulation and (c) voxel 2, outside of the electrocoagulation region. Denoising weights show that the AR constraint can accurately identify the presence or absence of the dominant metabolite peaks (NAA, Cr, Lac, and residual water peak). As expected, Lac is present at voxel 1, while it is absent at voxel 2.

MCA, resulting in a stoppage of blood supply. This in turn caused the lack of oxygen and glucose. Thus, we expect elevated Lac concentration in the electrocoagulation and CSF regions. As we can see from Figs. 3.16(b),(c), the denoising weights indeed confirm these expectations: the weights clearly indicate the presence of a Lac peak at voxel 1, while a Lac peak at voxel 2 is absent.

NAA and Lac spatial maps were obtained by integrating the complex spectrum at 2.02 ppm and 1.33 ppm, respectively. Figure 3.17 shows these maps obtained from the measured, spatially denoised, spatially-spectrally denoised, and wavelet soft-shrinkage denoised data. Wavelet denoising was performed with a soft shrinkage threshold that yielded the same data consistency level as the AR denoised data. Compared to the spatial denoising alone, the proposed spatial-spectral reconstruction helped to further suppress noise, both spatially and spectrally. In addition, some features such as the electrocoagulation region, CSF region, and boundaries of the brain were enhanced (see Figs. 3.17(c), (g)), compared to the wavelet reconstruction (see Figs. 3.17(d), (h)). Notice that one should be careful with choosing parameters for the AR denoising such as the linear prediction model order and the number of the first FID data points used to estimate linear prediction coefficients, since a particular choice of these parameters can result in the false features in the reconstruction.

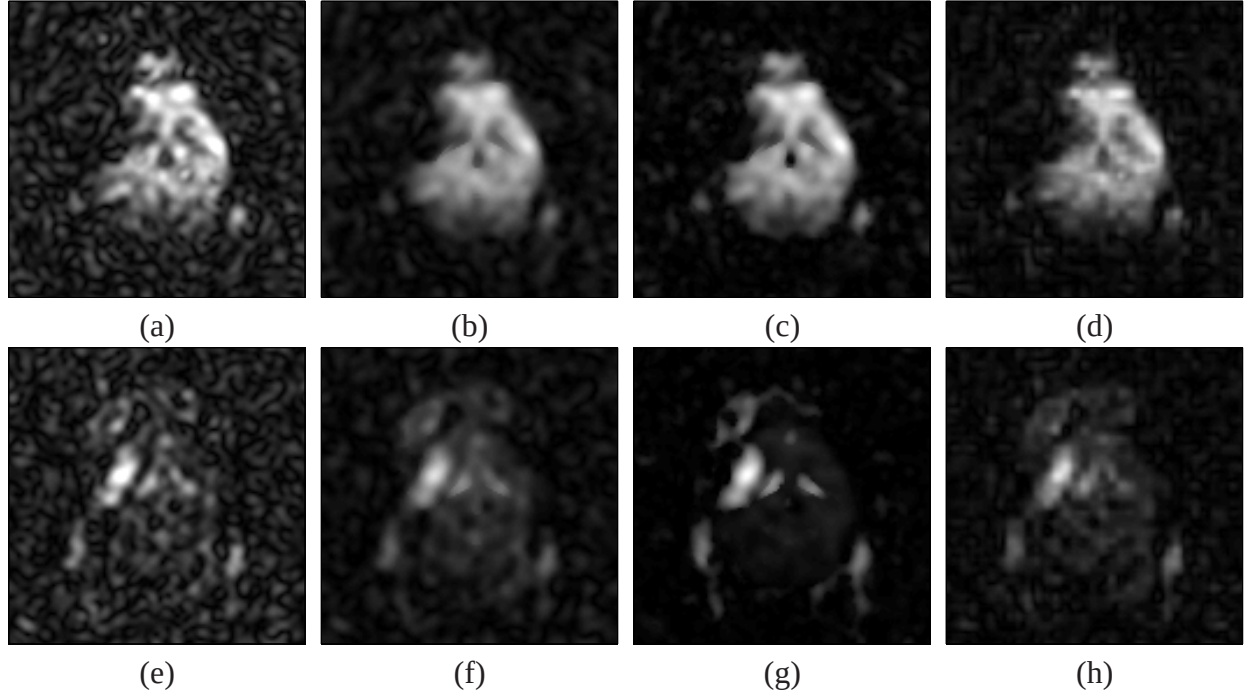


Figure 3.17: Denoising results from *in vivo* experiment: top row shows NAA spatial distributions obtained from (a) noisy data; (b) spatial denoising; (c) proposed denoising; (d) wavelet denoising; bottom row shows Lac spatial distributions obtained from (e) noisy data; (f) spatial denoising; (g) proposed denoising; (h) wavelet denoising. Wavelet denoising and proposed method have approximately the same data consistency level.

Figure 3.18 shows spectra obtained from the noisy, spatially denoised, spatially-spectrally denoised, and wavelet-denoised reconstructions at voxels 1 and 2. Among different reconstructions, the spatial-spectral denoising clearly suppresses the noise in non-metabolite regions, indicated by the weights in Fig. 3.16. However, this non-uniform spectral denoising behavior is not desirable visually. Due to this filtering property, the proposed algorithm is very sensitive to the AR model order. With the model order $L = 4$ used in this study, the method suppressed noise variance in the regions other than from 1.3 ppm to 4.7 ppm, as indicated by the weights in Fig. 3.16(a) and observed from the denoised spectra in Figs. 3.18(c),(g). This strong sensitivity to the model order is not desirable. Notice also that a residual water peak was alleviated significantly by both spatial and spatial-spectral denoising, shown in Figs. 3.18((b),(c),(f), and (g) respectively. This can be explained due to the fact that a spatial phase incoherence of the water peak was introduced by the pre-processing of the data for water removal. Because the anatomical constraint relies on a

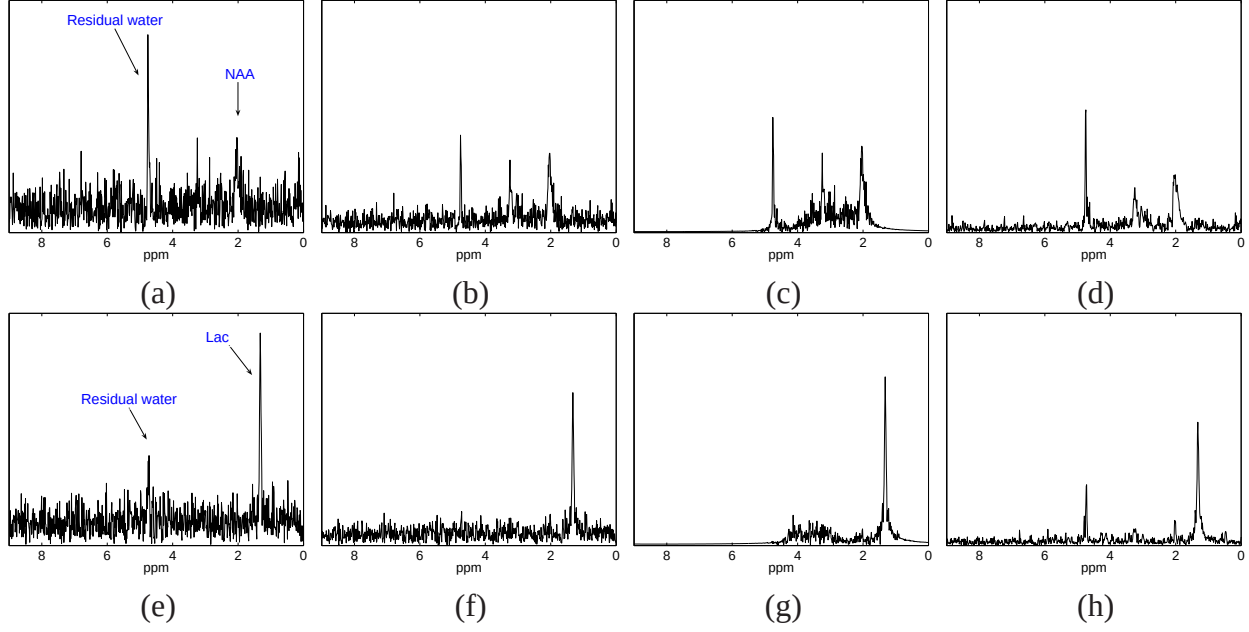


Figure 3.18: Denoising results from *in vivo* experiment: top row shows spectra at voxel 2 obtained from (a) noisy data, (b) spatial denoising, (c) proposed denoising, and (d) wavelet denoising; bottom row shows spectra at voxel 1 obtained from (e) noisy data, (f) spatial denoising, (g) proposed denoising, and (h) wavelet denoising. Voxels 1 and 2 are indicated by arrows in Fig. 3.16(a). All spectra are shown in absolute mode and on the same scale.

spatial smoothness assumption, applying spatial and/or spatial-spectral denoising resulted in the alleviated water peak amplitude.

3.5 Summary

This chapter investigated a new method for spatial-spectral denoising of MRSI data. The method formulates the denoising problem using the penalized maximum-likelihood framework, in which spatial constraints are imposed using a weighted spatial-smoothness regularization and spectral assumptions are incorporated with an AR constraint. The spectral constraint is motivated by a physical assumption on the lineshape characteristics in MR spectroscopy. The denoising effect of the method was characterized and quantitation evaluation studies were performed to assess the value of the denoised data. Parameters such as the linear prediction order and the method used to estimate linear prediction coefficients affect denoising performance of the algorithm significantly.

Experimental results demonstrate the ability of the spectral constraint to suppress noise in the non-metabolite regions. This non-uniform spectral denoising property of the method may not be desirable visually. Nevertheless, the method can improve quantitation accuracy if the spectral regularization term and linear prediction order are properly designed for particular applications.

Chapter 4

Denoising of MRSI data with low-rank approximations

4.1 Introduction

In the previous chapter we have considered denoising MRSI data using the penalized maximum-likelihood framework, in which spatial constraints are imposed using a weighted spatial-smoothness regularization and spectral assumptions are incorporated with an autoregressive constraint. The investigated approach showed the ability to improve metabolite quantitation accuracy if the parameters are chosen appropriately. However, the main disadvantage of the approach is a non-uniform spectral denoising. This behavior was observed experimentally and explained theoretically. In this chapter, another new scheme for MRSI denoising is introduced, which overcomes the non-uniform filtering effect and could prove useful for denoising MRSI data and other spatial-spectral and spatial-temporal imaging data. The method, coined LORA (Low Rank Approximations), utilizes low-rank structures of the spatial-spectral data. Specifically, we incorporate low-rank approximation in k - t domain by assuming that spatial variations are separable from temporal variations to some low order [127]. In addition, we exploit the low-rank structure of the temporal signal due to its linear predictability. The combination of these two low-rank properties provides an efficient spatial-spectral filtering.

The rest of this chapter is organized as follows. Section 4.2 presents the proposed method, including a description of the low-rank properties and the associated denoising algorithm. Section 4.3 discusses the denoising performance of the proposed method. Section 4.4 shows simulation and experimental results, followed by the summary in Section 4.5.

4.2 Proposed method

The proposed method is based on low-rank approximations of MRSI data. Specifically, it exploits two low-rank properties (one due to partial separability and the other due to linear predictability) for noise removal.

4.2.1 Low-rankness due to spatiotemporal partial separability

The spatial-spectral distribution function $\rho(\mathbf{r}, f)$ can be expressed as

$$\rho(\mathbf{r}, f) = \sum_{l=1}^{L_{\text{ps}}} a_l(\mathbf{r}) \psi_l(f). \quad (4.1)$$

This model is called L_{ps} -th-order partially separable (between space “ $\mathbf{r}$ ” and frequency “ f ”). Equivalently, the noiseless MRSI data $s_0(\mathbf{k}, t)$ can be expressed as partially separable in k -space “ $\mathbf{k}$ ” and time “ t ” to the L_{ps} -th order:

$$s_0(\mathbf{k}, t) = \sum_{l=1}^{L_{\text{ps}}} c_l(\mathbf{k}) \phi_l(t). \quad (4.2)$$

In (4.2) we assume that field inhomogeneity effects have been previously removed. As an example of a field-correction technique, the interested reader may refer to [73–75].

Validity of the PS model (4.1) can be justified as follows. Noting that $\rho(\mathbf{r}, f)$ is an $\mathcal{L}_2$ -function, model (4.1) is always valid for $L_{\text{ps}} = \infty$. In practice, the PS model is valid for a finite (small) L_{ps} because there is a finite number of resonances (spectral components) in any practical MRS experiment. In this case, function $\psi_l(f)$ can be viewed as the spectral function of the l -th resonance component and $a_l(\mathbf{r})$ is its corresponding spatial distribution. Model (4.1) also arises when $\rho(\mathbf{r}, f)$ can be decomposed into a summation of compartmental spectral functions. This simplified form of the PS model was previously used in [71] for spectroscopic imaging. In this case L_{ps} is the

number of compartments, $a_l(\mathbf{r}) = 1$, and

$$\psi_l(f) = \begin{cases} \frac{1}{V_l} \int \rho(\mathbf{r}, f) d\mathbf{r}, & \mathbf{r} \in \mathcal{D}_l \\ 0, & \text{otherwise,} \end{cases} \quad (4.3)$$

where $\mathcal{D}_l$ represents the l -th compartment and V_l is its corresponding volume. While compartmental spatial-spectral functions are partially separable, the PS model in (4.1) represents a much broader class of functions and does not impose compartmental homogeneity as in [71], which is often problematic for practical MRSI data [128].

An important property of model (4.2) is that the Casorati matrix formed from $s_0(\mathbf{k}, t)$ samples is low-rank. This property is summarized below.

Remark 1: *Let*

$$\mathbf{C}_0 = \begin{bmatrix} s_0(\mathbf{k}_1, t_1) & s_0(\mathbf{k}_1, t_2) & \dots & s_0(\mathbf{k}_1, t_M) \\ s_0(\mathbf{k}_2, t_1) & s_0(\mathbf{k}_2, t_2) & \dots & s_0(\mathbf{k}_2, t_M) \\ & & \dots & \\ s_0(\mathbf{k}_N, t_1) & s_0(\mathbf{k}_N, t_2) & \dots & s_0(\mathbf{k}_N, t_M) \end{bmatrix} \quad (4.4)$$

for any point set $\{s(\mathbf{k}_n, t_m)\}_{n=1, m=1}^{N, M}$. Then, model (4.2) implies that $\mathbf{C}_0$ has at most rank L_{ps} [127].

In practice, L_{ps} is much smaller than $\min(N, M)$ due to the small number of spectral components (about 5 ~ 14 commonly MR-observable metabolites, 4 resonance components from macromolecules, and 5 components from lipids in the human brain [7, 23]). This enables us to use low-rank property for effective denoising.

4.2.2 Low-rankness due to linear predictability

The time-domain signal of a spin system with L_{lp} spectral components resonating at frequency f_l with damping factor d_l can be expressed as [21]

$$s_0(t) = \sum_{l=1}^{L_{\text{lp}}} \alpha_l e^{-(d_l + i2\pi f_l)t}. \quad (4.5)$$

Equivalently in the Fourier domain, the spectrum consists of L_{lp} Lorentzian resonance lineshapes $\varphi_l(f)$:

$$\begin{aligned} \rho(f) &= \sum_{l=1}^{L_{\text{lp}}} \alpha_l \varphi_l(f), \\ \varphi_l(f) &= \frac{1/d_l}{1 + 4\pi^2(f + f_l)^2/d_l^2} - i2\pi \frac{(f + f_l)/d_l^2}{1 + 4\pi^2(f + f_l)^2/d_l^2}. \end{aligned} \quad (4.6)$$

In practice, resonance lineshapes $\varphi_l(f)$ may deviate from Lorentzian lineshapes because of magnetic field inhomogeneity and limited spatial resolution, etc. So, an associated practical question is whether the observed resonance lineshapes can be well represented using (4.6). This question was previously raised in [21, 129] and it was found that any observed lineshape can be fitted using a linear combination of Lorentzian basis functions. For example, the broad resonance from residual water was found to be well-represented using 3 Lorentzians [21]. Therefore, (4.6) is valid (in a mathematical sense), although in this case L_{lp} does not physically represent the number of spectral peaks. In practice, L_{lp} is larger than the number of spectral peaks but much smaller than the number of available samples M .

An important property of model (4.5) is that the discrete time-domain signal is linearly predictable, i.e.

$$s_0[m] = \sum_{l=1}^{L_{\text{lp}}} \beta_l s_0[m - l], \quad (4.7)$$

where $s_0[m] = s_0(m\Delta t)$ with Δt being the sampling interval. A Hankel matrix formed from $s_0[m]$ is low-rank, as stated below.

Remark 2: $\{s_0[m]\}_{m=1}^M$ is L_{lp} -th order linearly predictable as in (4.7) if and only if the Hankel matrix

$$\mathbf{H}_0 = \begin{bmatrix} s_0[1] & s_0[2] & \dots & s_0[K] \\ s_0[2] & s_0[3] & \dots & s_0[K+1] \\ \dots & \dots & \dots & \dots \\ s_0[M-K+1] & s_0[M-K+2] & \dots & s_0[M] \end{bmatrix} \quad (4.8)$$

has rank L_{lp} [130].

4.2.3 Proposed algorithm

1) Low-rank matrix approximation

Assume that an arbitrary rank- L matrix $\mathbf{A}_0$ is perturbed by noise (denoted by a random matrix $\mathbf{E}$) as

$$\mathbf{A} = \mathbf{A}_0 + \mathbf{E}. \quad (4.9)$$

In the case when rank L is known *a priori*, the low-rank approximation of $\mathbf{A}$ is given by

$$\bar{\mathbf{A}} = \arg \min_{\text{rank}(\hat{\mathbf{A}})=L} \|\mathbf{A} - \hat{\mathbf{A}}\|, \quad (4.10)$$

where the norm $\|\cdot\|$ denotes either spectral or Frobenius norm. It is well known that $\bar{\mathbf{A}}$ can be obtained using SVD as

$$\bar{\mathbf{A}} = \sum_{l=1}^L \sigma_l(\mathbf{A}) \mathbf{u}_l \mathbf{v}_l^H, \quad (4.11)$$

where σ_l , $\mathbf{u}_l$, and $\mathbf{v}_l$ are the singular values, left singular vectors, and right singular vectors of $\mathbf{A}$, respectively.

2) Rank determination

Case 1: $\mathbf{A}_0$ is a general $N \times M$ low-rank matrix and $\mathbf{E}$ is an $N \times M$ random Gaussian matrix.

In this case, L can be determined using the results from random matrix theory. Specifically, we first estimate noise standard deviation σ_0 from the measured data $\mathbf{s}(\mathbf{k}_n, t_m)$ and use σ_0 to estimate

$\|\mathbf{E}\|_2$. We then choose rank $\hat{L}$ so that

$$\sigma_{\hat{L}+1}(\mathbf{A}) \leq \|\mathbf{E}\|_2 \leq \sigma_{\hat{L}}(\mathbf{A}), \quad (4.12)$$

where $\|\mathbf{E}\|_2$ can be estimated using the Marchenko-Pastur distribution of the eigenvalues λ of $\mathbf{E}^H \mathbf{E}$ (under asymptotic assumption that $M, N \rightarrow \infty$ with $M/N = \text{const}$), which is given by [131]

$$\text{pdf}(\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\sigma_0^2\lambda\beta}, \quad (4.13)$$

where $\lambda_- \leq \lambda \leq \lambda_+$, $\lambda_{\pm} = \sigma_0^2(1 \pm \sqrt{\beta})^2$, and $\beta = M/N$. Consequently, $\|\mathbf{E}\|_2$ can be estimated as

$$\|\mathbf{E}\|_2 \approx \sigma_0(1 + \sqrt{\beta}). \quad (4.14)$$

Case 2: $\mathbf{A}_0$ is a low-rank Hankel matrix and $\mathbf{E}$ is a Gaussian Hankel matrix. The rank selection problem in this case has been well studied and is related to the model order estimation problem for autoregressive models. The rank $\hat{L}$ can be determined using various criteria such as the Akaike Information Criterion (AIC) [132]. Specifically, $\hat{L}$ is chosen to yield a minimum change in $AIC(\hat{L}) - AIC(\hat{L} + 1)$, where

$$AIC(\hat{L}) = M \log e(\hat{L}) + 2\hat{L} \quad (4.15)$$

and $e(\hat{L})$ is the error of a least squares fit of model (4.5) to the noisy data, given a particular candidate rank value $\hat{L}$.

3) Denoising by Low-Rank Approximations (LORA)

The proposed denoising algorithm is summarized below:

1. Given noisy data $\mathbf{s}(\mathbf{k}_n, t_m)$, construct matrix $\mathbf{C}$ according to (4.4) and solve the following optimization problem:

$$\bar{\mathbf{C}} = \arg \min_{\text{rank}(\hat{\mathbf{C}})=L_{\text{ps}}} \|\mathbf{C} - \hat{\mathbf{C}}\| \quad (4.16)$$

by performing rank- L_{ps} approximation using SVD.

2. Take the 2-D discrete Fourier transform (DFT) of each column of matrix $\bar{\mathbf{C}}$ to obtain $\tilde{s}(\mathbf{r}_n, t_m)$.
3. For each voxel $\mathbf{r}_n$, construct matrix $\mathbf{H}$ from $\tilde{s}(\mathbf{r}_n, t_m)$ according to (4.8) and solve the following optimization problem:

$$\bar{\mathbf{H}} = \arg \min_{\text{rank}(\hat{\mathbf{H}})=L_{\text{lp}}} ||\mathbf{H} - \hat{\mathbf{H}}|| \quad (4.17)$$

by performing rank- L_{lp} approximation using SVD.

4. Denoised data at each voxel $\mathbf{r}_n$ is obtained by extracting the elements from the first row and last column of $\bar{\mathbf{H}}$.

4.3 Discussion

4.3.1 Algorithm considerations

The performance of low-rank filtering based on linear predictability is rather sensitive to the noise level because the signal subspace in the Hankel matrix $\mathbf{H}$ is usually not well-conditioned (the rank- L signal component has a large condition number). For the low SNR values in typical *in vivo* MRSI data, this filtering step can result in spectral artifacts such as loss of signal components and introduction of spurious peaks. This is especially the case if both low-rank and Hankel structures are imposed on $\hat{\mathbf{H}}$ as is done in the Cadzow signal enhancement algorithm [108], which has been used for denoising [94, 111, 112]. Specifically, in [108] both low-rank and Hankel structures were imposed by averaging the elements along the anti-diagonals of matrix $\bar{\mathbf{H}}$ (instead of step 4 in the proposed algorithm) and the algorithm was iterated until convergence. The convergent solution of the Cadzow method has the following desired properties:

- (a) Hankel matrix $\mathbf{H}$ approaches a low-rank matrix;

(b) low-rank matrix $\bar{\mathbf{H}}$ approaches a Hankel matrix;

(c) spectral lineshape approaches Lorentzian;

(d) SNR of the denoised solution approaches infinity.

Figure 4.1 confirms the convergence properties mentioned above. The simulated noiseless and noisy spectra are shown in Fig. 4.2(a), bottom row. Specifically, Fig. 4.1(a) shows that even by approximating the noisy data matrix to a low rank matrix, the resulting Hankel matrix $\mathbf{H}$ after extracting the denoised data samples is close to the full rank (black curve). However, as we iterate, the rank approaches L (red curve), which confirms convergence property (a). Figure 4.1(b) shows variances of the elements along the main anti-diagonal of $\bar{\mathbf{H}}$ after SVD truncation. As we can see, these variances are large at the first iteration of the algorithm (black curve) and are negligible as the algorithm converges (red curve). A similar effect is observed for other anti-diagonals. This demonstrates that $\bar{\mathbf{H}}$ approaches a Hankel matrix, thus confirming convergence property (b). Lastly, by performing a Lorentzian fit of the denoised spectrum using the HSVD method [21], Fig. 4.1(c) illustrates that as the number of iterations increases, the denoised spectrum tends to have Lorentzian lineshapes, which confirms convergence properties (c) and (d).

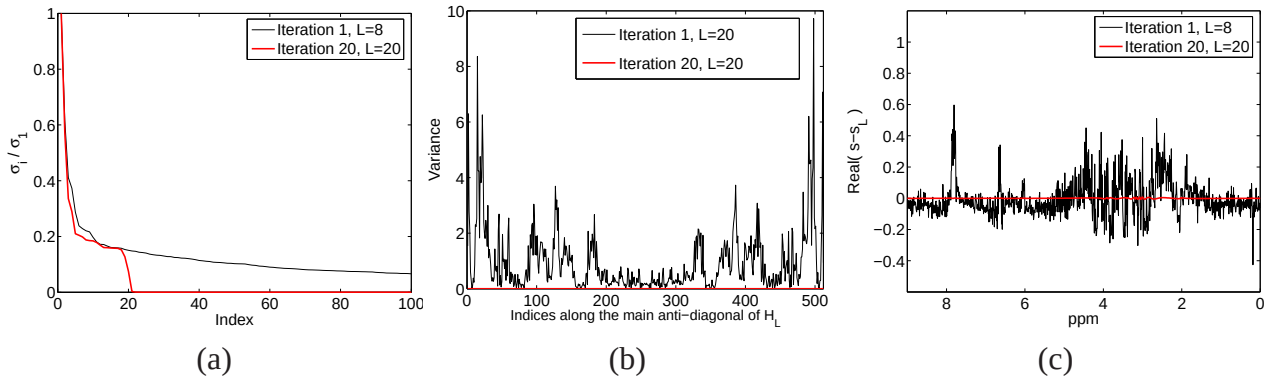


Figure 4.1: Convergent solution of the Cadzow denoising approach: (a) singular values of $\mathbf{H}$ (first 100 values are shown only); (b) variance of elements along the main anti-diagonal of $\bar{\mathbf{H}}$; (c) error of the denoised signal and its Lorentzian fit. This experiment confirms that the convergent solution has a corresponding (a) low-rank matrix $\mathbf{H}$, (b) Hankel matrix $\bar{\mathbf{H}}$, and (c) Lorentzian lineshape.

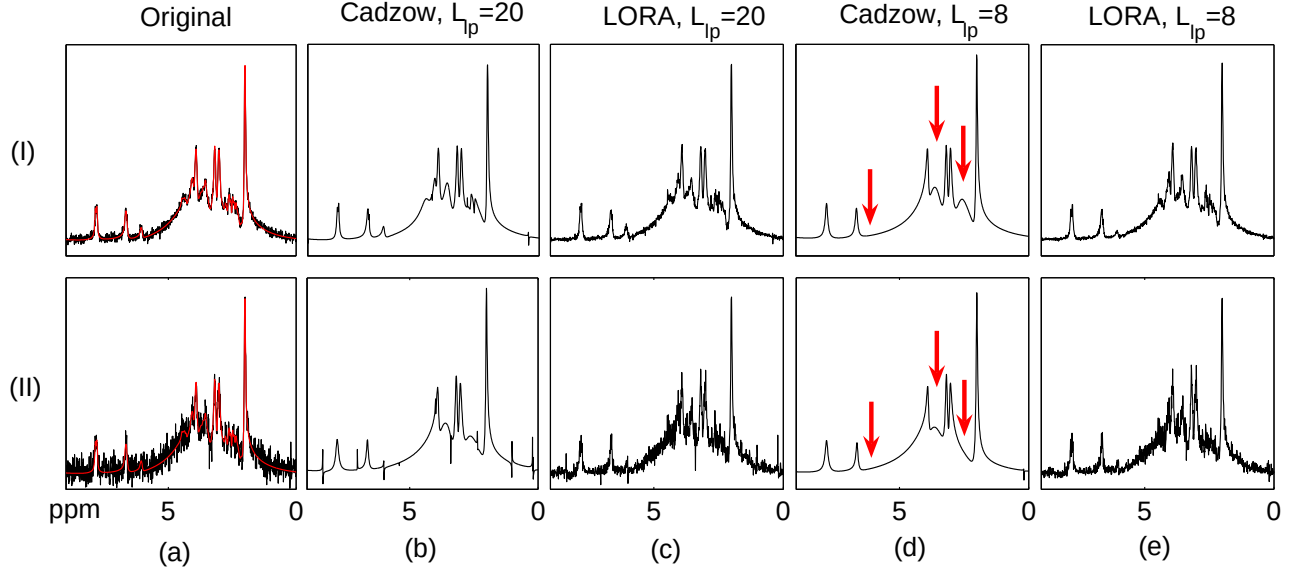


Figure 4.2: Simulation study of low-rank filtering based on linear predictability: (a) original (red) and noisy (black) spectra at low and high noise levels; (b)-(e) denoised spectra from Cadzow and LORA, each with high ($L_{lp} = 20$) or low ($L_{lp} = 8$) model order.

Proposition 1: Let $\mathbf{H}_0$ be rank- L_{lp} matrix formed from the noiseless data points according to (4.8), $\tilde{\mathbf{H}}$ is the rank- L_{lp} approximation as obtained from (4.17), and $\bar{\mathbf{H}}$ denotes a Hankel matrix obtained by averaging elements of $\tilde{\mathbf{H}}$ along anti-diagonals, i.e.

$$\tilde{s}[n] = \frac{1}{p(n)} \sum_{0 \leq i \leq I(n), 0 \leq j \leq J(n): i+j=n} \bar{s}_{i,j}, \quad (4.18)$$

where $\tilde{s}[n]$ and $\bar{s}_{i,j}$ are the elements of $\tilde{\mathbf{H}}$ and $\bar{\mathbf{H}}$ respectively, and $p(n)$ indicates how many times an element $\tilde{s}_n$ appears along the corresponding matrix anti-diagonal. Then (4.18) yields a smaller error in the Frobenius norm, that is

$$\|\tilde{\mathbf{H}} - \mathbf{H}_0\|_F \leq \|\bar{\mathbf{H}} - \mathbf{H}_0\|_F \quad (4.19)$$

with equality if and only if $\bar{\mathbf{H}}$ is Hankel.

For the proof of Proposition 1, see Appendix A. While according to this proposition, averaging elements of $\bar{\mathbf{H}}$ leads to a smaller error in the Frobenius norm with respect to the original matrix $\mathbf{H}_0$, averaging in general over-smooths spectral features and distorts spectral lineshapes, as shown in Figs. 4.2(b),(d). We suggest that averaging should be performed only when a low-rank matrix $\bar{\mathbf{H}}$ is close to Hankel (such as when the signal is generated by a low-order Lorentzian model and the noise level is relatively low). By not imposing the Hankel structure on $\hat{\mathbf{H}}$ in step 3 of the algorithm, the “rank” constraint is actually weakly imposed, which enables the algorithm to accommodate non-Lorentzian lineshapes (see Remark 2 on correspondence of low-rank and Hankel structures to exact Lorentzian spectral model) and avoid spectral artifacts in low-SNR cases. This is an advantage of LORA over the Cadzow algorithm. In addition, the low-rank filtering based on partial separability improves the SNR of the data before low-rank filtering based on linear predictability is applied, which further improves robustness of the LORA.

Figure 4.2 shows a set of simulation results to illustrate the sensitivity (to noise and model order) of low-rank filtering based on linear predictability. Figure 4.2(a) shows ground truth (red) and noisy (black) spectra at two SNR levels. The denoised spectra in Figs. 4.2(b)-(e) were obtained with Cadzow (imposing both low-rank and Hankel structures) and LORA (imposing low-rank structure only), respectively with two model orders ($L_{\text{lp}} = 20$ and 8). Note that in the high-SNR case the Cadzow method produced an almost “perfect” denoised result with a correct model order (top row, (b)) but lost some spectral components with an under-estimated model order (top row, (d)).

In the low SNR case, the Cadzow method introduced spurious spectral features (bottom row, (b)) even with the correct model order. The problem can be alleviated with a small model order but at the expense of missing spectral components (such as the loss of the smallest peak at 6.05 ppm and over-smoothing of the non-Lorentzian, closely spaced spectral peaks in the regions from 2.2 ppm to 2.8 ppm and from 3.3 ppm to 3.8 ppm; see bottom row in (d)). Both problems (spurious peaks and missing peaks) were alleviated in LORA (see (c) and (e)). Figure 4.3 shows the normalized bias of the Cadzow denoising procedure as a function of the number of iterations. This

bias was computed from a Monte Carlo study with 1024 noise realizations as

$$\text{Bias} = \frac{\|\mathbb{E}\{\hat{s}_i\} - s_0\|_2}{\|s_0\|_2}, \quad i = 1, \dots, 1024, \quad (4.20)$$

where $\hat{s}_i$ are the denoised data given the noisy signal realization s_i and $\mathbb{E}$ denotes expectation over noise realizations. It can be clearly seen that as the number of iterations increases, the denoised solution becomes more biased.

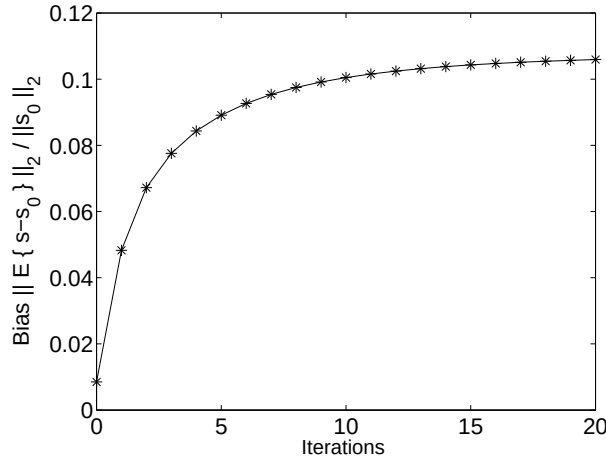


Figure 4.3: Monte Carlo study of the bias introduced at each iteration by the Cadzow method. The corresponding simulated spectra are shown in Fig. 4.2(a), bottom row. The model order L_{lp} was 8. Notice that in this study the bias increases with iterations.

4.3.2 Rank selection

For low-rank filtering based on partial separability, in the case of high/moderate SNR (ideally as long as $\|E\|_2 \leq \sigma_L(A_0)$ or in the range above 18 dB for the simulated ^1H MRSI dataset described in Section 4.4.1) we choose rank L_{ps} of C_0 according to (4.12). To evaluate the performance of the described rank selection method, we performed rank selection on the simulated dataset with true rank $L_{ps} = 8$. Specifically, we performed a Monte Carlo study with 8 different noise levels, 32 noise realizations per noise level and estimated $\|E\|_2$ using the Marchenko-Pastur distribution as in (4.14). The effective rank $\hat{L}_{ps}$ was selected according to (4.12). For this experiment and the

rest of the paper, we define SNR in terms of signal energy and signal amplitude as

$$\begin{aligned}\text{SNR}_e &= 20 \log_{10} \frac{\|\mathbf{s}_0\|_2}{\|\mathbf{s} - \mathbf{s}_0\|_2}, \\ \text{SNR}_p &= \frac{|A|}{\sigma_0},\end{aligned}\tag{4.21}$$

where $\mathbf{s}$ and $\mathbf{s}_0$ are the noisy and noiseless signals, respectively, $|A|$ is the amplitude of the largest metabolite peak and σ_0 is the standard deviation of the noise.

Table 4.1 shows the mean effective rank $\bar{L}_{\text{ps}}$, averaged over all the noise realizations. Notice that $\|\mathbf{E}\|_2$ is closely estimated using the Marchenko-Pastur distribution at every SNR level, as can be seen from the reported relative error $\frac{\|\mathbf{E}\|_2 - \|\hat{\mathbf{E}}\|_2}{\|\mathbf{E}\|_2}$. The results also show that as the SNR decreases, the discussed approach under-estimates L_{ps} . In practice, the choice of L_{ps} can also be guided by the known number of spectral components present in the spectrum [17].

Table 4.1: Monte Carlo experiment on choosing effective rank $\bar{L}_{\text{ps}}$ at 9 different SNR levels. Voxel 1 is in the region of low-SNR, while voxel 2 is in the region of high-SNR. All SNR values are reported in decibels.

$\text{SNR}_e(\mathbf{s}(\mathbf{k}, t))$	23.72	22.23	20.94	19.83	18.85	17.97	17.17	16.45
$\text{SNR}_e(\mathbf{s}(f))$ at voxel 1	6.42	5.17	4.45	3.83	3.26	2.83	2.27	1.99
$\text{SNR}_e(\mathbf{s}(f))$ at voxel 2	32.46	31.31	29.96	28.52	27.82	27.01	26.22	25.41
$\frac{\ \mathbf{E}\ _2 - \ \hat{\mathbf{E}}\ _2}{\ \mathbf{E}\ _2}$	0.0057	0.0051	0.0053	0.0058	0.0046	0.0050	0.0054	0.0050
L_{ps}	8	8	7.03	6.63	6.06	6.00	5.25	5.03

For low-rank filtering based on linear predictability, the choice of rank L_{lp} of matrix $\mathbf{H}_0$ is based on the AIC as specified in (4.15). We performed a Monte Carlo study on rank L_{lp} selection with the simulated dataset at 3 different noise levels, 256 noise realizations were considered. Figure 4.4 shows the computed mean AIC values (averaged over the noise realizations) at each noise level as a function of candidate rank values. The corresponding representative noisy spectra are shown in Fig. 4.7(a) on page 90. The effective rank $\hat{L}_{\text{lp}}$ is typically chosen to minimize AIC. However, this usually leads to an over-estimated rank [132]. To achieve a closer estimate, we suggest choosing

the rank to be the threshold point at which there is no significant reduction in the AIC value when the rank is further increased. In practice, the choice of L_{lp} can also be guided by the known number of spectral peaks present in the spectrum.

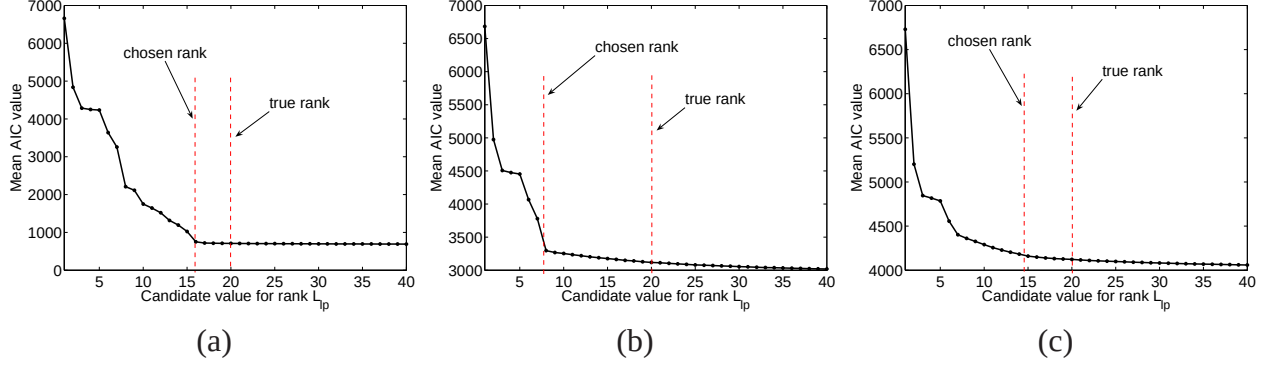


Figure 4.4: Monte Carlo study of rank L_{lp} selection based on AIC: mean AIC value (averaged over 256 noise realizations) as a function of $\hat{L}_{lp}$ at (a) low noise level with SNR_e=26.61 dB, SNR_p=94.4; (b) medium noise level with SNR_e=17.03 dB, SNR_p=30.4; (c) high noise level with SNR_e=12.76 dB, SNR_p=17.9. The true rank L_{lp} in this experiment was 20.

Inaccurate rank estimation for low-rank filtering of the linear prediction matrix is more problematic than that of the Casorati matrix. Specifically, an over-estimated rank L_{lp} ($\hat{L}_{lp} > L_{lp}$) can result in spurious peaks. Using an under-estimated rank (without imposing the Hankel structure) can alleviate the problem. However, a highly under-estimated rank $\hat{L}_{lp}$ can lead to a loss of useful spectral information, which is not desirable. Typically, peaks with small amplitudes are distorted or lost first [133]. Possible peak loss due to rank under-estimation can be alleviated by imposing low-rank structure only, as opposed to imposing both low-rank and Hankel structures iteratively as discussed in Section 4.2.3. For low-rank filtering of the Casorati matrix, severe rank under-estimation can lead to spatial blurring and spectral lineshape broadening. An over-estimated rank L_{ps} in this case does not directly introduce bias and spurious peaks, but would reduce the filtering effectiveness.

4.3.3 Denoising performance

To analyze the denoising performance of the proposed low-rank filtering, we rewrite $\mathbf{C}$ and $\bar{\mathbf{C}}$ as $\mathbf{C} = \mathbf{C}_0 + \mathbf{E}$ and $\bar{\mathbf{C}} = \mathbf{C}_0 + \mathbf{\Delta}$, where $\mathbf{C}_0$ is an $N \times M$ data matrix with rank L_{ps} as described in (4.4), $\mathbf{E}$ is the noise matrix, $\bar{\mathbf{C}}$ is formed from the denoised data as obtained from (4.16), and $\mathbf{\Delta}$ is the residual error, which can contain both residual noise and possible signal loss due to low-rank filtering. Without loss of generality, we assume $M < N$. The denoised matrix $\bar{\mathbf{C}}$ is obtained according to (4.16). We define the noise reduction factor as

$$g = \frac{\|\mathbf{E}\|_F}{\|\mathbf{\Delta}\|_F}. \quad (4.22)$$

By defining the SNR before and after denoising as $20 \log_{10} \frac{P_s}{\|\mathbf{E}\|_F^2}$ and $20 \log_{10} \frac{P_s}{\|\mathbf{\Delta}\|_F^2}$, where P_s denotes the signal power, the SNR gain can be calculated as $20 \log_{10} g$. Ideally, one would want $g = \infty$. In the full-rank case when $L_{\text{ps}} = M$ and no denoising is achieved, g equals 1.

Computing g as in (4.22) requires knowing the ground truth in order to compute $\|\mathbf{\Delta}\|_F$. However, notice that we only need to estimate the norm $\|\mathbf{\Delta}\|_F$ rather than knowing $\mathbf{\Delta}$ completely. Thus, one would hope that there is a signal-independent expression to predict $\|\mathbf{\Delta}\|_F$ closely. In general, it is hard to accurately characterize singular values of $\mathbf{C}_0$, $\mathbf{C}$, and $\bar{\mathbf{C}}$. However, under assumptions that (a) the signal and noise (before and after denoising) subspaces are uncorrelated in the sense that $\mathbf{C}_0^H \mathbf{E} = \mathbf{0}$ and $\mathbf{C}_0^H \mathbf{\Delta} = \mathbf{0}$, and (b) the noise is complex white, i.e. $\mathbf{E}^H \mathbf{E}$ and $\mathbf{\Delta}^H \mathbf{\Delta}$ are scalar multiples of identity matrix [134], which imply the following relation among the singular values:

$$\sigma_i^2(\mathbf{\Delta}) = \sigma_i^2(\bar{\mathbf{C}}) - \sigma_i^2(\mathbf{C}_0) \quad (4.23)$$

$$\sigma_i^2(\mathbf{E}) = \sigma_i^2(\mathbf{C}) - \sigma_i^2(\mathbf{C}_0), \quad (4.24)$$

we can obtain the following approximate formula based on the theoretical results of SVD of noisy

matrices in [134]:

$$g = \sqrt{\frac{\sum_{i=1}^M \sigma_i^2(\mathbf{E})}{\sum_{i=1}^{\hat{L}_{\text{ps}}} \sigma_i^2(\mathbf{E})}}, \quad \text{if } \hat{L}_{\text{ps}} \geq L_{\text{ps}}. \quad (4.25)$$

Proof of (4.25) is straightforward and can be found in Appendix B. Expression (4.25) provides a signal-independent prediction of g . It suggests that (i) in the case of accurate rank estimation ($\hat{L}_{\text{ps}} = L_{\text{ps}}$) and $\sigma_i(\mathbf{E})$ being all equal, the noise reduction factor is on the order of $\sqrt{\frac{M}{\hat{L}_{\text{ps}}}}$; (ii) in the case of over-estimated rank ($\hat{L}_{\text{ps}} > L_{\text{ps}}$), the residual noise level is larger than the residual noise level with correct rank determination. In practice, the conditions (a) and (b) which are required for (4.25) to hold are never satisfied exactly. However, we observed that a mild violation of these conditions may still give a rough guess of g , as shown in our simulation study discussed below.

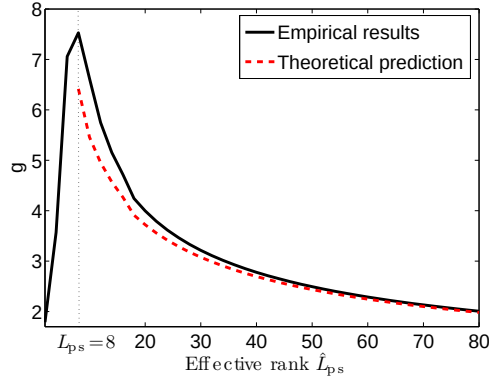


Figure 4.5: Noise reduction factor g for low-rank filtering based on partial separability as a function of the estimated rank $\hat{L}_{\text{ps}}$. Empirical values of g (black) and corresponding theoretical approximations (red) were computed according to (4.22) and (4.25), respectively. The true rank $L_{\text{ps}} = 8$. Notice that in the region $\hat{L}_{\text{ps}} > L_{\text{ps}}$ the theoretically predicted values of g tend to its empirical values and both approach 1 in the full-rank case.

Based on the simulated ^1H MRSI dataset described in Section 4.4.1, we performed a Monte Carlo study to compute g according to (4.22), averaged over 32 noise realizations. The noise reduction factor g is plotted as a function of the estimated rank $\hat{L}_{\text{ps}}$ in Fig. 4.5. We observe that when the rank is properly chosen, g reaches its maximum value (about 7.6 for the simulated dataset). As the rank is over-estimated ($\hat{L}_{\text{ps}} > L_{\text{ps}}$) and approaches the full rank, g decreases to 1, as expected. The reduction of g reflects the residual noise variance in the over-estimated rank case and the signal loss in the under-estimated rank case. A nice feature of this plot is that it shows how

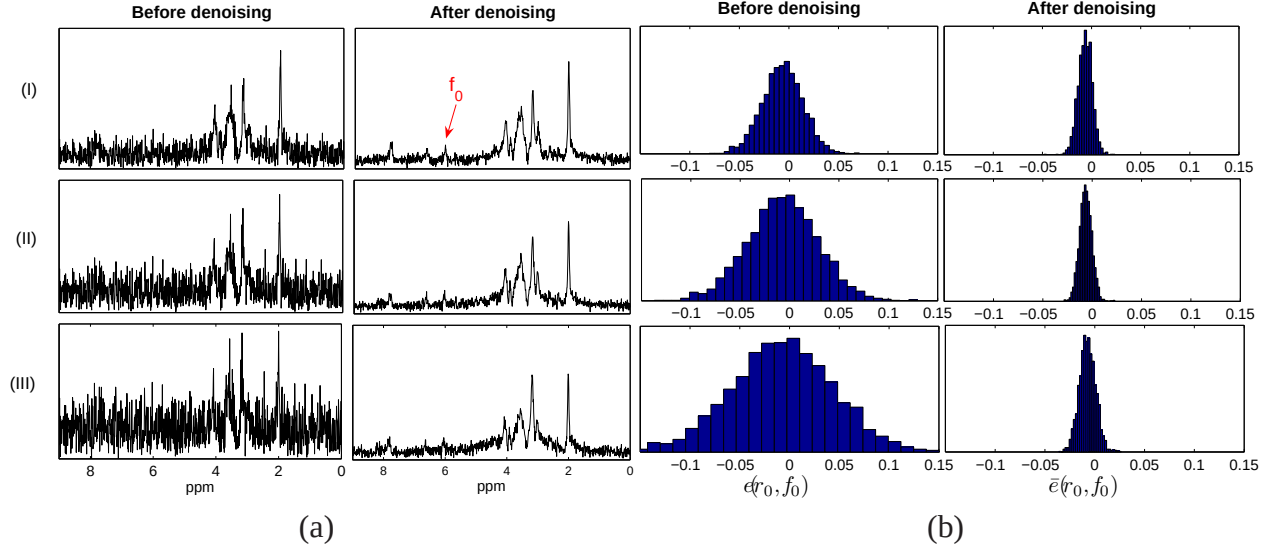


Figure 4.6: Monte Carlo study of low-rank filtering based on partial separability: (a) representative noisy and denoised spectra at three SNR levels; (b) error distributions calculated as the histograms of e and $\bar{e}$ according to (4.27) at a particular location (r_0, f_0) . For this illustrative example, r_0 is a representative voxel in the white matter region marked as “2” in Fig. 4.9(e) and f_0 is the frequency point marked in Fig. 4.6(a). All spectra are shown in absolute mode. Notice the reduced noise variances achieved by denoising.

much of the noise reduction we can achieve even in the case when the rank is incorrectly estimated. For example, if the rank was over-estimated to be 32, while the true rank is 8, then according to this plot, we still achieve a factor of about 3 (or 9.5 dB) of improvement in the noise reduction. Thus, for the sake of improving SNR, estimating rank precisely is not a critical point, as long as we do not under-estimate the rank causing the signal loss. Note that with the considered dataset in Fig. 4.5, the relative error $\frac{\|E^H E - M\sigma_0^2 I\|_F}{M\sigma_0} = 0.99$ for condition (a) and $\frac{\|C_0^H E\|_F}{\sqrt{\|C_0^H C_0\|_F \|E^H E\|_F}} = 0.15$ for condition (b). These numbers did not significantly change as the SNR was varied from low to high (see Fig. 4.6(a) for representative SNR levels). Conditions (a) and (b) for Δ are satisfied closer as $\hat{L}$ approaches M . The predicted value from (4.25) is plotted in the same figure as the empirical value of g in Fig. 4.5. Notice that in the region $\hat{L}_{ps} > L_{ps}$ the theoretically predicted values of g approach its empirical value as $\hat{L}_{ps}$ tends to the full-rank case.

We have performed a similar analysis on the noise reduction factor g , for the case of low-rank filtering of linear prediction matrix. However, to evaluate the level of SNR improvement for

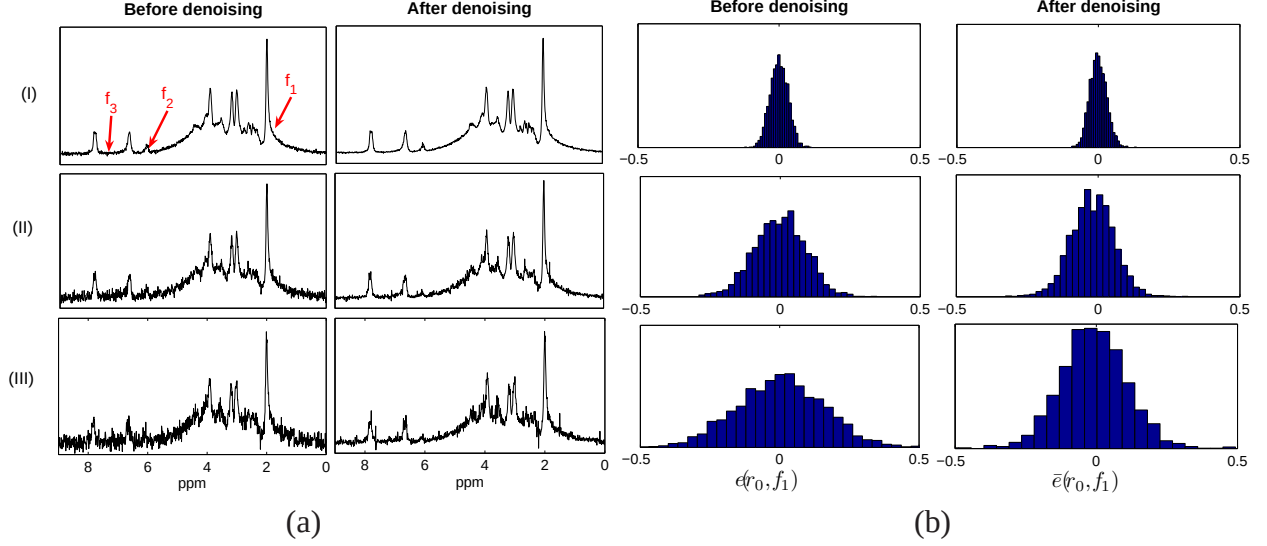


Figure 4.7: Monte Carlo study of low-rank filtering based on linear predictability: (a) representative noisy and denoised spectra at three noise levels; (b) error distributions calculated as the histograms of e and $\bar{e}$ according to (4.27) at a particular location (r_0, f_1) . For this illustrative example, r_0 is a representative voxel in the white matter region marked as “2” in Fig. 4.9(e) and f_1 is the frequency point marked in Fig. 4.7(a).

low-rank filtering of linear prediction matrix, we rather define the following factor:

$$h = \frac{\|\xi\|_2}{\|\eta\|_2}, \quad (4.26)$$

where ξ and η are vectors extracted from the first row and last column of E and Δ , respectively. We performed a Monte Carlo study with 256 noise realizations to compute h at different noise levels shown in Fig. 4.7(a). The reduction factor h depends on the conditioning of the Hankel matrix H_0 . For the case considered here, we observed that h achieves the value of around 1.7. However, for better-conditioned cases with well-separated spectral peaks, h can be more substantial (in the range of 5 for a spectrum with 3 peaks, separated by 0.57 ppm). This experiment suggests that low-rank filtering based on linear predictability can be less effective than low-rank filtering based on partial separability.

We performed a Monte Carlo study to further evaluate the denoising effectiveness of low-rank filtering with 2048 noise realizations to obtain histograms of the errors before and after denoising

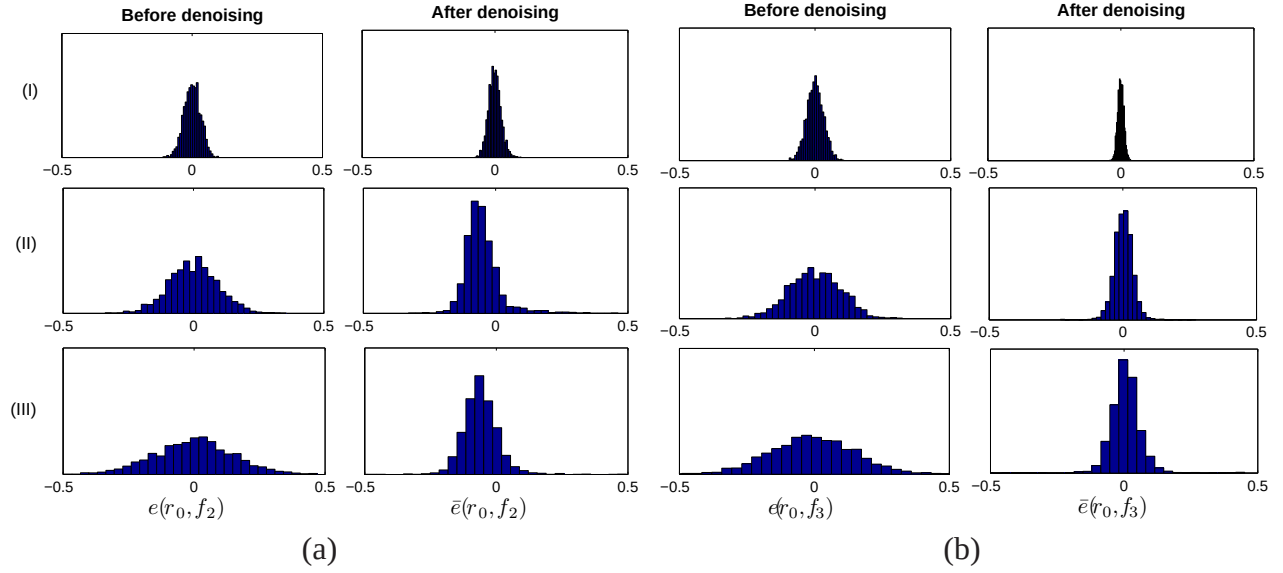


Figure 4.8: Monte Carlo study of low-rank filtering based on linear predictability (continuation of Fig. 4.7): (a),(b) error distributions calculated as the histograms of e and $\bar{e}$ according to (4.27) at a particular location $(\mathbf{r}_0, f_1)$, $(\mathbf{r}_0, f_2)$, and $(\mathbf{r}_0, f_3)$, respectively. For this illustrative example, $\mathbf{r}_0$ is a representative voxel in the white matter region marked as “2” in Fig. 4.9(e) and f_2, f_3 are the frequency points marked in Fig. 4.7(a). Representative noisy and denoised spectra at three noise levels are shown in Fig. 4.7(a).

as

$$\begin{aligned} e &= s - s_0 \\ \bar{e} &= \bar{s} - s_0, \end{aligned} \tag{4.27}$$

where $\bar{s}$ denotes the denoised data. Figure 4.6 shows some representative results for the low-rank filtering based on partial separability. It is easy to see that a significant reduction in the noise variance was achieved at every SNR level. Note that low-rank filtering is a “biased” estimator. However, the resulting bias is negligible once the rank is correctly chosen and SNR is not extremely low. In the case of extremely low SNR, using an over-estimated rank reduces bias but at the expense of reduced filtering effect.

Figures 4.7 and 4.8 show some representative results from a Monte Carlo study of low-rank filtering based on linear predictability. Notice that the denoising performance is frequency-dependent.

The histogram at a low-SNR frequency f_2 in Fig. 4.8(a) shows that a significant noise variance reduction is still achieved; however, a larger bias is introduced, compared to the case of a higher-SNR frequency f_1 in Fig. 4.7(b). For the frequency regions that contain mostly noise (such as the frequency f_3 in Fig. 4.8(b)), we observe significant noise variance reduction with no bias.

4.4 Simulation and experimental results

4.4.1 Simulations

To illustrate the performance of the proposed method, we reconsider the simulated MRSI data previously described and used in Section 3.4.1. Figure 4.9 shows the spatial distribution of the spatial-spectral function at a low-SNR frequency of 6.66 ppm, indicated in Fig. 4.10(e) as “ f_0 ”. We compare the proposed approach with 3-D wavelet soft shrinkage (Daubechies 4-tap kernel filter, 4 levels) and conventional Gaussian apodization. The wavelet shrinkage threshold and apodization constant were chosen to yield the same level of the residual noise variance as that for the proposed approach (threshold $T = 2.28\sigma_0$). It can be clearly seen that the corresponding wavelet-denoised spatial-spectral function in Fig. 4.9(c) has an improved SNR at the expense of blurring the spatial features significantly, while the proposed denoising preserves the spatial features better. Gaussian apodization shown in Fig. 4.9(d) has the poorest trade-off between SNR and resolution among all methods, as expected.

Figure 4.10 shows representative spectra from a particular voxel marked as “1” in Fig. 4.9(e). Both Gaussian apodization and wavelet denoising improve SNR at an expense of lineshape broadening and loss of metabolite amplitudes, with Gaussian apodization performing the worst. This amplitude loss can be seen from the corresponding residual spectra, which are not centered at zero, as shown in Figs. 4.10(g),(h).

To assess whether the proposed denoising helped to improve the accuracy of the spectral quantitation, we performed multi-voxel quantitation based on the noisy and denoised data. Quantitation

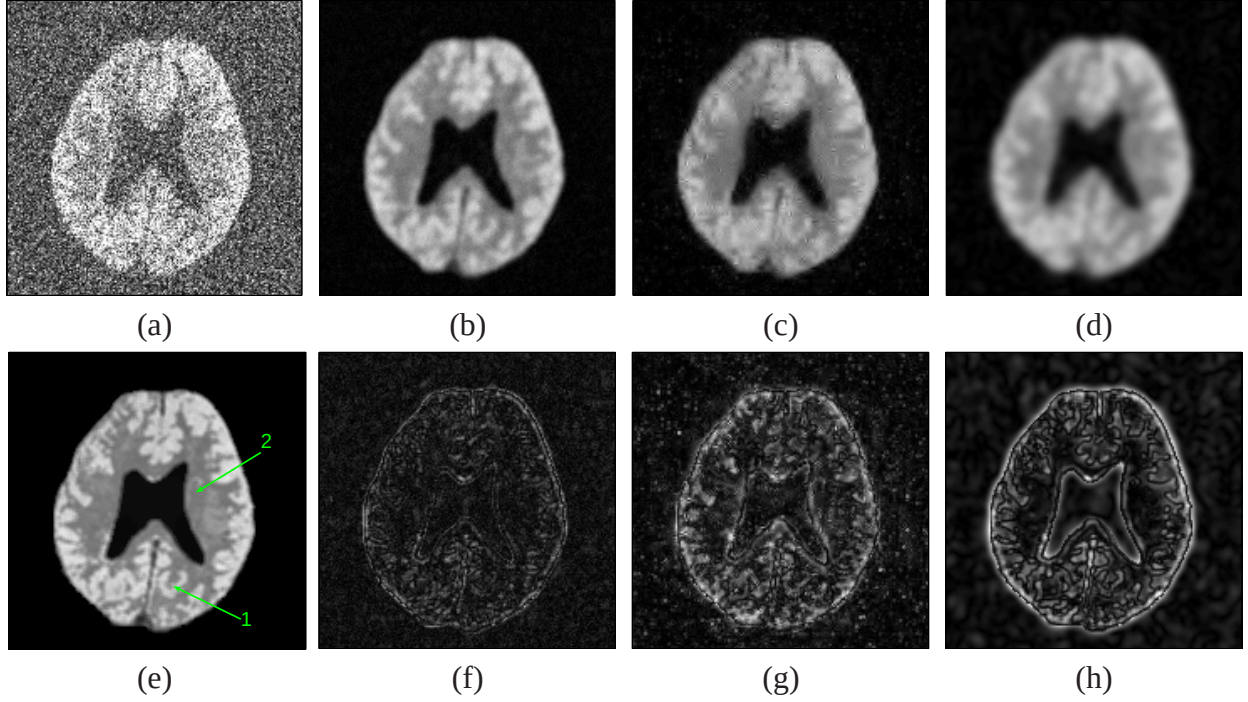


Figure 4.9: Denoising results - spatial distributions $\rho(\mathbf{r}, f_0)$ at a low-SNR frequency (6.66 ppm, see Fig. 4.10(e)), obtained from (a) noisy data; (b) LORA denoising; (c) wavelet denoising; (d) Gaussian apodization; and (e) noiseless data. Corresponding errors of (b)-(d) are shown in (f)-(h), respectively. Parameters for both wavelet denoising and Gaussian apodization (wavelet shrinkage threshold and apodization constant, respectively) were chosen to yield approximately the same residual noise variance level as that of the proposed method.

is formulated similarly to the single-voxel quantitation procedure described in Section 3.3.4, but in this case instead of assuming known frequencies of each individual spectral peak, we assume known spectral profiles φ_n of each metabolite. In addition, in order to avoid the local-minima issues during non-linear optimization, we assume that the ground truth lineshape parameters d_n are known. Specifically, we solve the following linear optimization problem:

$$\{\hat{\mathbf{c}}, \hat{\mathbf{b}}\} = \arg \min_{\mathbf{c}, \mathbf{b}} \|\mathbf{s} - \mathbf{Z}\mathbf{c} - \mathbf{\Upsilon}\mathbf{b}\|_2, \quad (4.28)$$

where

$$Z_{m,n} = \varphi_n[m\Delta t]e^{-d_n m^2 \Delta t^2} \quad (4.29)$$

and $\mathbf{s}$ is in turn chosen to be the noisy data and the LORA-denoised data. Parameters $\mathbf{c}$ and $\mathbf{b}$

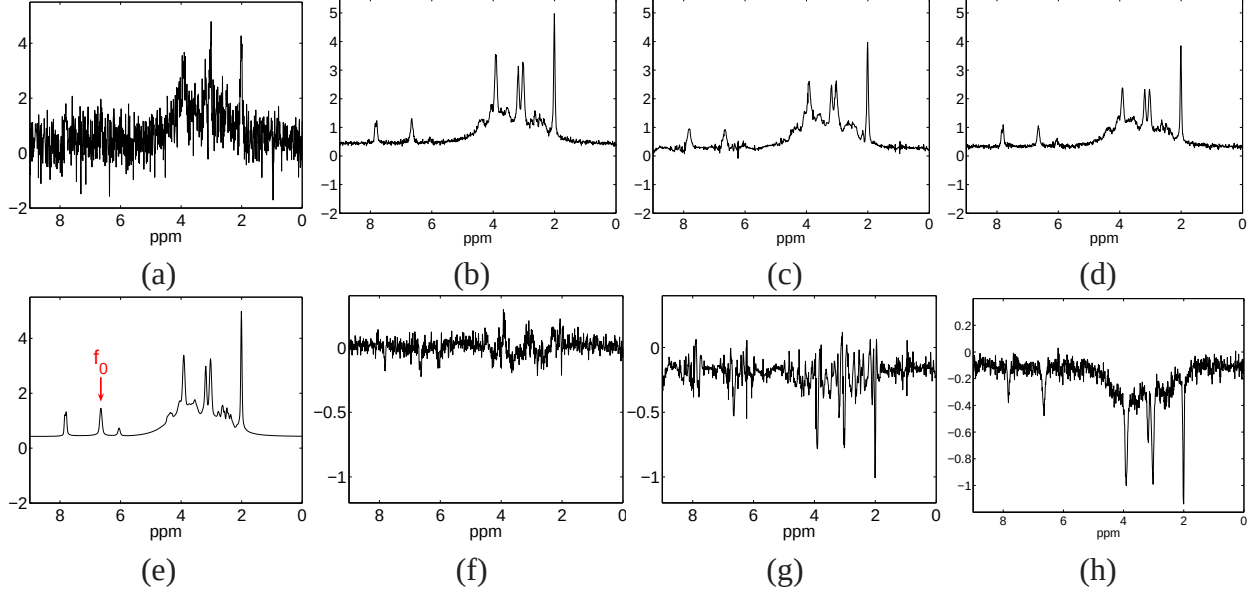


Figure 4.10: Denoising results - spectra at a particular voxel, marked as “1” in Fig. 4.9(e), obtained from (a) noisy data; (b) LORA denoising; (c) wavelet shrinkage; (d) Gaussian apodization; (e) noiseless data. Corresponding errors of (b)-(d) are shown in (f)-(h), respectively. All spectra are shown in the real-valued mode. Parameters for both wavelet denoising and Gaussian apodization (wavelet shrinkage threshold and apodization constant, respectively) were chosen to yield approximately the same residual noise variance level as that of the proposed method.

are metabolite amplitudes and baseline coefficients, respectively, while Υ represents the baseline basis functions in the frequency domain, Fourier-transformed to the time domain, as described in Section 3.3.4.

Figure 4.11 shows histograms of the normalized root mean-squared errors of metabolite amplitudes $\hat{c}_{\text{metab}}$ estimated from the noisy and LORA-denoised data over GM and WM regions. The error was computed as

$$\text{NRMSE}_c = \frac{|\hat{c}_{\text{metab}} - c_{\text{metab}}|}{c_{\text{metab}}} \cdot 100\%. \quad (4.30)$$

For each region, three such histograms are shown for the cases of quantifying NAA (high-SNR metabolite), Cr (medium-SNR metabolite), and Glu (low-SNR metabolite). Notice that histograms obtained from the LORA-denoised data are more concentrated at the origin and have much narrower tails than the corresponding histograms obtained from the noisy data. This is observed in all cases of different metabolites and over both GM and WM regions, which suggests that the

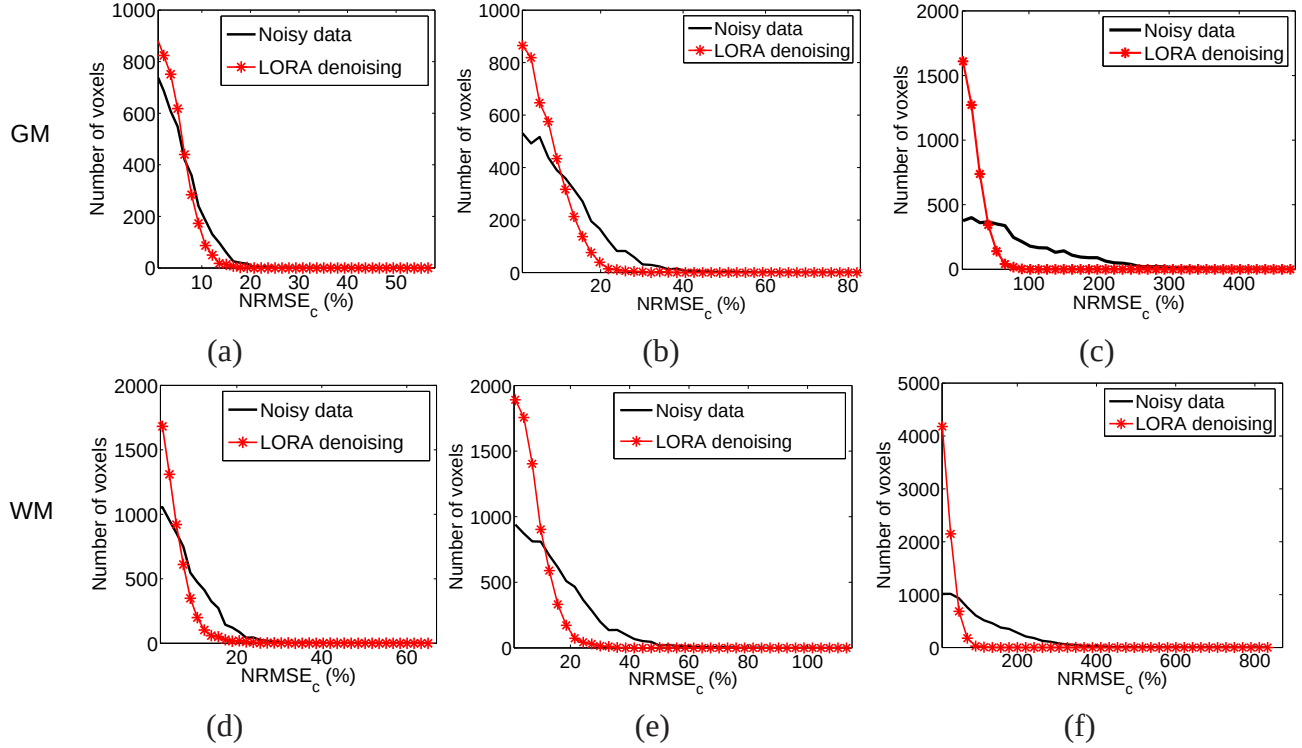


Figure 4.11: Metabolite quantitation study from noisy and denoised data: histograms of the normalized root mean-squared error of the estimated metabolite amplitude (NRMSE_c) for (a) NAA, (b) Cr, and (c) Glu in GM region and (d) NAA, (e) Cr, and (f) Glu in WM region. NRMSE_c was computed according to (4.30). Notice that the proposed denoising helped to improve the accuracy of metabolite quantitation.

proposed denoising helped to improve the accuracy of metabolite quantitation.

4.4.2 *In vivo* experiments

We have applied LORA to denoise *in vivo* MRSI data of the mouse brain, previously described and used in Section 3.4.2. Before performing LORA-denoising, data was pre-processed with spatial denoising using anatomical reconstruction [113] to further protect spatial features. The edge-enhancing effect of this spatial denoising is used to compensate for the possible spatial blurring introduced by low-rank approximation due to spatiotemporal partial separability.

NAA and Lac spatial maps were obtained by integrating the complex spectrum at 2.02 ppm and 1.33 ppm, respectively. Figure 4.12 shows NAA and Lac maps obtained from the measured and LORA-denoised data. Figure 4.13 shows spectra obtained from these reconstructions at voxels

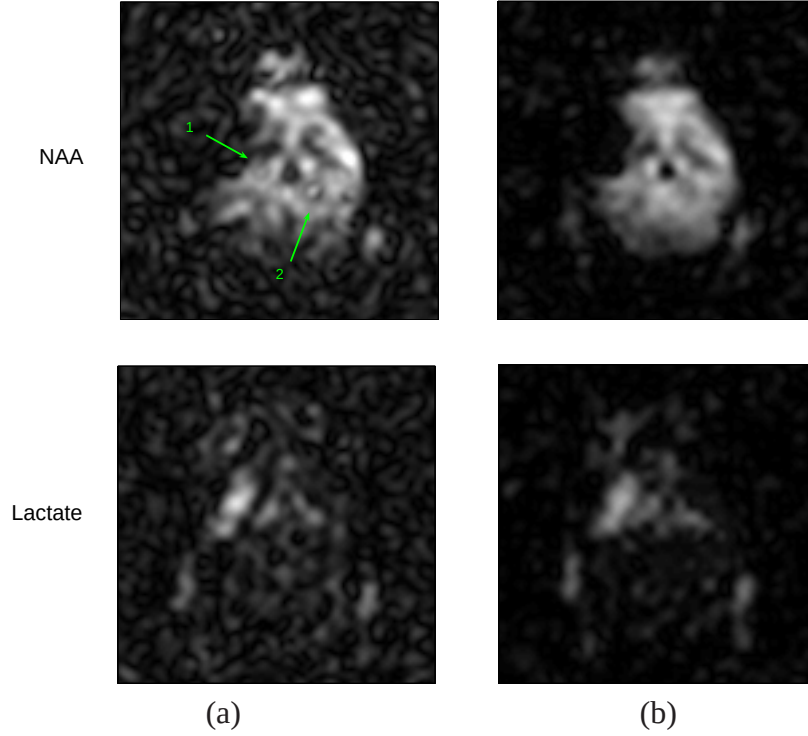


Figure 4.12: Denoising results from *in vivo* experiment: NAA (top row) and Lac (bottom row) spatial distributions obtained from (a) noisy data; (b) proposed method.

within and outside the region of electrocoagulation, marked as “1” and “2” in the top row of Fig. 4.12(a), respectively. Notice that the anatomical constraint resulted in the amplitude loss of the water peak due to the phase incoherence created during water removal processing. It can be clearly seen that the noise has been significantly suppressed while preserving spatial-spectral features. As mentioned in Section 3.4.2, electrocoagulation blocked MCA, resulting in a stoppage of blood supply and lack of oxygen. Thus, we expect elevated Lac and reduced NAA concentrations in the electrocoagulation region, which can be seen from the LORA-denoised reconstruction in Fig. 4.12(b) and Fig. 4.13(b). In addition, the LORA method does not experience a non-uniform spectral filtering property as the denoising method previously investigated in Chapter 3.

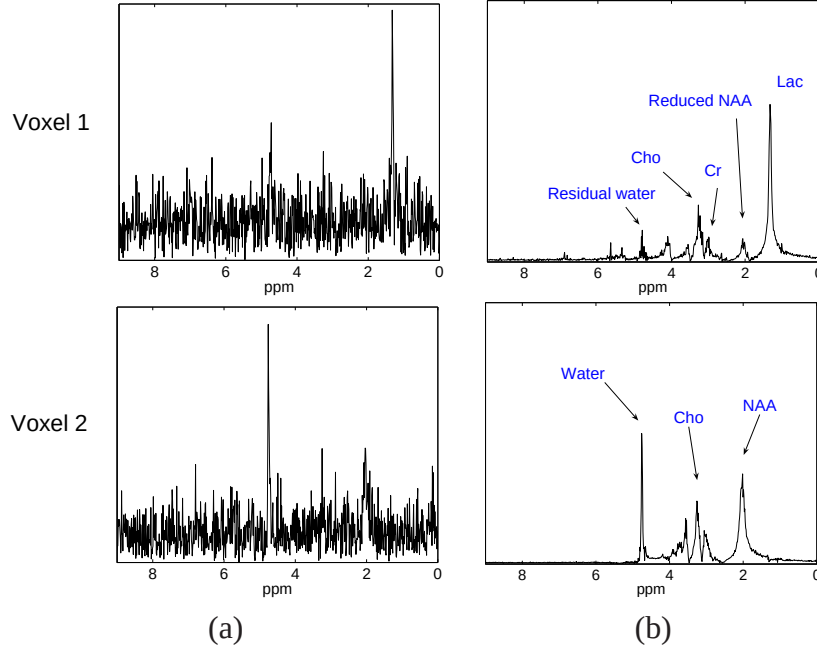


Figure 4.13: Denoising results from *in vivo* experiment: spectra at voxels in the region of electrocoagulation (voxel 1, top row) and outside of electrocoagulation area (voxel 2, bottom row), obtained from (a) noisy data; (b) proposed method. All spectra are shown in absolute mode and on the same scale.

4.5 Summary

In this chapter we have presented a new method for spatial-spectral denoising of MR spectroscopic imaging data. The method exploits the low-rank properties of MRSI data, one due to partial separability and the other due to the linear predictability of MRSI data. The combination of partial separability and linear predictability models provides a new principled way to improve SNR for spatiotemporal imaging. Unlike the denoising method investigated in Chapter 3, which incorporates the linear predictability assumption as a regularization within the penalized maximum-likelihood framework, this method does not have a non-uniform spectral denoising property. Both experimental results and the denoising performance study demonstrated that the combination of these two low-rank properties of MRSI data provides an effective method for denoising, which should prove useful for practical MRSI applications.

Chapter 5

Cramér-Rao bound analysis of echo-time selection

5.1 Introduction

In MRS experiments, the *echo time* (TE) is an important timing parameter that affects the signal-to-noise ratio (SNR), the number of metabolites that contribute significantly to the observed data, the complexity of the spectral baseline, and the spectral profile of each individual metabolite. All of these factors affect the interpretation of the observed spectrum to varying degrees, and the choice of TE remains controversial [135]. In particular, shorter TE spectra generally have higher SNR, as can be seen from a schematic diagram of the conventional spin-echo CSI sequence in Fig. 5.1, and contain significant contributions from a larger number of metabolites than long TE spectra; however, this advantage can be offset by the more complicated signal model that must be used to describe short TE data. In other words, for the sake of obtaining the highest SNR, one should choose the shortest TE possible; however, in that case the signal model becomes complex (model order increases and nuisance signal components must be considered), which can often result in reduced accuracy of the metabolite quantitation procedure.

In the existing literature, the TE parameter was often chosen based on qualitative metrics, such as the visual characteristics of the spectrum, or to be consistent with previous studies. Some quantitative empirical comparisons between short and long TE spectra have been reported in [62, 136, 137], though these studies only investigate up to three different TE s. The influence of TE on the performance of quantitation of metabolite amplitudes for single-voxel proton spectroscopy has been examined by existing empirical studies [135, 137–141]. There is variability in the conclusions regarding whether short or long TE values are preferred.

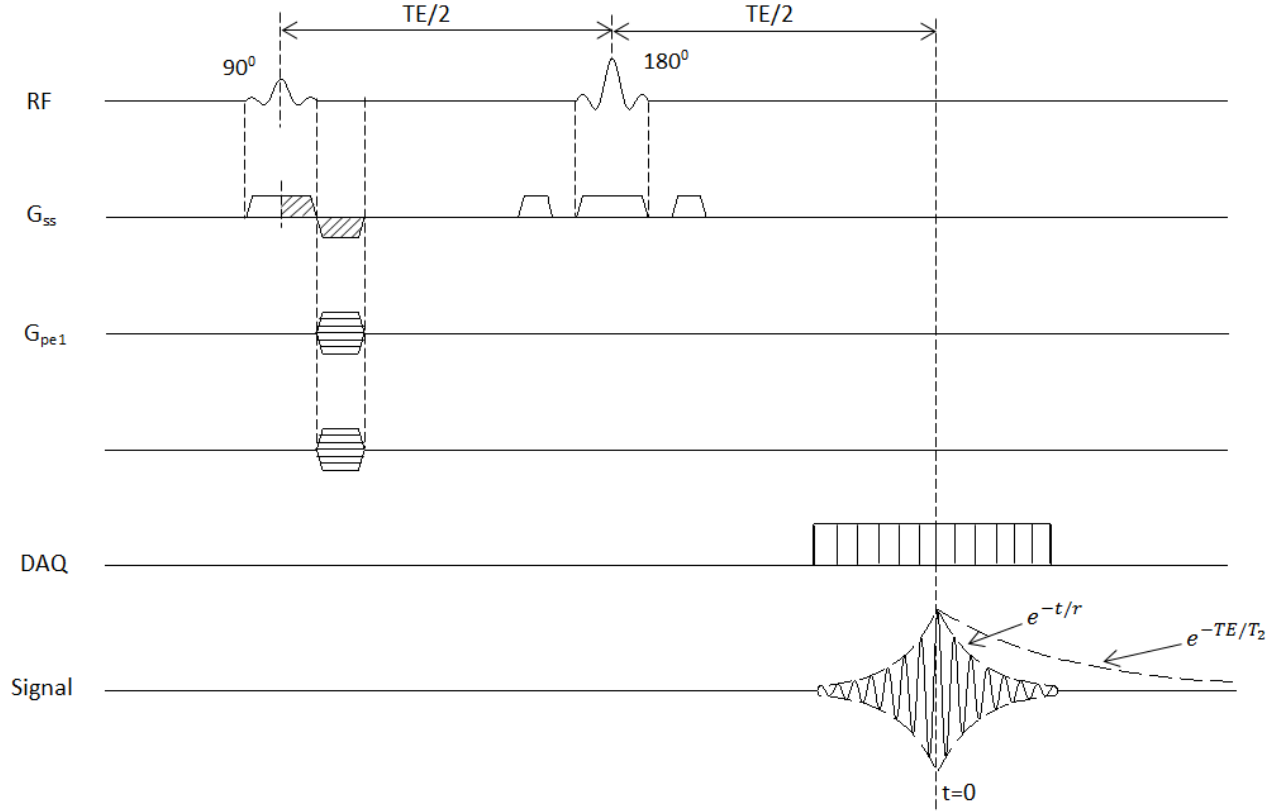


Figure 5.1: A schematic diagram of the typical spin-echo CSI data acquisition: at a longer TE the signal intensity is decreased exponentially, resulting in a lower SNR.

In this chapter, we consider the choice of TE from an estimation theoretic perspective. Specifically, we analyze the Cramér-Rao lower bound (CRB) on the variance of the metabolite amplitude estimates as a function of TE. While retrospective use of the CRB is common in MRS as a metric of quality for an acquired experimental dataset [12–14], we are interested in using the CRB prospectively to guide the design of an experiment. This new approach provides a quantitative method for identifying potentially useful TEs from an arbitrarily large range of candidate values, in contrast to a common heuristic choice of TE.

5.2 Proposed formulation

5.2.1 Spectral model

The observed discrete signal from a system with N spectral components originating from metabolite signals and K spectral components originating from baseline signals (such as macromolecules, lipids) can be modeled as the following:

$$\begin{aligned}
 s[m] &= s_{\text{metab}} + s_{\text{baseline}} + \xi[m], \\
 s_{\text{metab}}[m] &= e^{i\phi_0} \sum_{n=1}^N a_n(TE) \varphi_{n,TE}[m] \psi_{n,d_n}[m], \\
 s_{\text{baseline}}[m] &= \sum_{k=1}^K b_k v_{k,TE}[m], \\
 m &= 0, \dots, M-1.
 \end{aligned} \tag{5.1}$$

Each spectral metabolite component is characterized by the real, positive amplitude $a_n(TE)$, known metabolite basis function $\varphi_{n,TE}[m]$, and the signal decay $\psi_{n,d_n}[m]$, where d_n denotes a lineshape parameter. In addition, the model includes a zero-order phase term ϕ_0 for the whole spectrum, and $\xi[m]$ denotes additive noise.

As TE increases, the amplitudes $a_n(TE)$ decrease due to spin relaxation, leading to a reduction in SNR. In this work, we assume exponential decay such that

$$a_n(TE) = c_n e^{-TE/T_{2,n}}, \tag{5.2}$$

where $T_{2,n}$ is a metabolite-dependent relaxation constant.

The basis functions $\varphi_{n,TE}[m]$ can change with TE due to quantum mechanical effects, and, as we will see later, this plays an important role in the behavior of the CRB as a function of TE . We

assume that the basis function is known and has the form

$$\varphi_{n,TE}[m] = \sum_{l=1}^{L_n} \alpha_{l,n}(TE) e^{-i\beta_{l,n}(TE)} e^{-i2\pi f_{l,n}(TE)m\Delta t}, \quad (5.3)$$

where Δt denotes the sampling time and $\alpha_{l,n}(TE)$, $\beta_{l,n}(TE)$, and $f_{l,n}(TE)$ are the relative amplitude, phase, and frequency of the l -th resonance, belonging to the n -th metabolite [48, 142]. These parameters can be obtained from quantum mechanical simulations [49, 50, 143].

In addition, we model the signal decay of each metabolite component as Lorentzian

$$\psi_{n,d_n}[m] = e^{-m\Delta t/d_n}, \quad (5.4)$$

although this could be easily generalized to accommodate more complicated lineshapes.

For spectral components originating from baseline signals, since no model is yet available to describe accurately and completely these components, we model them in a non-parametric way as a linear combination of the baseline basis functions $v_{k,TE}[m]$ weighted by complex weighting coefficients b_k . Notice that we assumed b_k to be independent of TE, while the dependence on TE is embedded in the basis functions $v_{k,TE}[m]$. These baseline basis functions can be obtained by first extracting the baseline part of the MRS signal (acquired at a certain TE) using the Subtract method [25] and then performing data fitting of the resulting baseline component using harmonic retrieval or B-spline functions. The Subtract method is based on the fact that macromolecular signals decay more rapidly than metabolite signals. Thus, by truncating some initial FID points, we can extract metabolite component from the total signal. Subtraction of this signal from the original MRS signal yields the baseline component.

Notice that the number of metabolite components N and the number of baseline components K in general are unknown and need to be determined from the data. In practice, one knows roughly N and K based on the literature [17] (about 5 ~ 14 commonly MR-observable metabolites in the human brain [7], 5 baseline components for lipids and 4 baseline components for macromolecules

[23]). This dissertation focuses on the TE selection aspect of the problem, assuming N and K to be known.

Given the model (5.1), we are interested in analyzing CRBs on the metabolite amplitudes $a_n(TE)$ as a function of TE. The optimal TE from estimation theoretic perspective is the one that yields the minimum CRB. For notational convenience, in the following sections of the dissertation, we assume implicitly that metabolite amplitudes depend on TE and we drop TE dependence in the notation by denoting $a_n(TE)$ as a_n .

5.2.2 Cramér-Rao bound expressions

We assume the unknown parameters are the metabolite amplitudes, lineshape parameters, zero-order phase, magnitude and phase of baseline amplitudes, i.e.,

$$\boldsymbol{\theta} = \{a_1, \dots, a_N, d_1, \dots, d_N, \phi_0, |b_1|, \dots, |b_K|, \angle b_1, \dots, \angle b_K\}.$$

Cramér-Rao bound theory states that the variance of unbiased estimates is always lower bounded by

$$\text{var}(\hat{\theta}_k) \geq \text{CRB}_{\theta_k}, \quad (5.5)$$

where

$$\text{CRB}_{\theta_k} = [\mathbf{F}^{-1}]_{kk} \quad (5.6)$$

and $\mathbf{F}$ is the Fisher information matrix [144].

We consider the context of additive complex Gaussian white noise $\boldsymbol{\xi} \sim \mathcal{N}\{\mathbf{0}, \sigma_0^2 \mathbf{I}\}$, for which the log-likelihood function $\ln L(\mathbf{s})$ of the noisy data $\mathbf{s}$ can be easily expressed as

$$\ln L(\mathbf{s}) = \text{const} - \frac{1}{\sigma_0^2} \sum_{m=0}^{M-1} \left| s[m] - e^{i\phi_0} \sum_{n=1}^N a_n \varphi_{n,TE}[m] \psi_{n,d_n}[m] - \sum_{k=1}^K b_k \nu_{k,TE}[m] \right|^2. \quad (5.7)$$

Each element of the Fisher information matrix can be computed as

$$\mathbf{F}_{\theta_i \theta_j} = \mathbb{E} \left[\left(\frac{\partial \ln L(\mathbf{y})}{\partial \theta_i} \right) \left(\frac{\partial \ln L(\mathbf{y})}{\partial \theta_j} \right)^T \right], \quad (5.8)$$

where the symbol $\mathbb{E}$ denotes the mean over the noise. As an example, a representative element from the Fisher information matrix has the following expression:

$$\mathbf{F}_{a_i a_j} = \frac{2}{\sigma_0^2} \text{Re} \left\{ \sum_{m=0}^{M-1} \varphi_{i,TE}^*[m] \psi_{i,d_i}^*[m] \varphi_{j,TE}[m] \psi_{j,d_j}[m] \right\}. \quad (5.9)$$

Other elements in the Fisher information matrix have similar expressions, shown in Appendix C.1. Inverting $\mathbf{F}$ gives the desired CRB matrix, with the first N diagonal elements being the CRBs on the amplitudes a_n , which are of interest in this work. Notice that the matrix entries depend on an inner product involving the basis functions and signal decay functions, which reflects interactions among metabolites and metabolites or among metabolites and baselines. As will be seen later, if metabolite spectral profiles are known *a priori*, the effect of spectral overlap is not a significant obstacle for metabolite quantitation as it may seem to be visually.

5.3 Analysis of TE selection

5.3.1 Coefficient of variation bound

We consider the following commonly MR-observable metabolites in the human brain [7, 17] which contribute to either small or large signals in a wide range of TEs (typically from 25 ms to 288 ms): *N*-acetylaspartate (NAA), choline (Cho), creatine (Cr), glutamate (Glu), glutamine (Gln), myo-inositol (mIns), lactate (Lac), scyllo-inositol (Scy), aspartate (Asp), taurine (Tau), glucose (Glc), histidine (His), threonine (Thr), and ethanolamine (Eth).

For each metabolite, we simulated 1024 data points according to (5.1) at the magnetic field strength of 3T and spectral bandwidth of 1, 200 Hz. Spectral parameters $\alpha_{l,n}(TE)$, $f_{l,n}(TE)$, and

$\beta_{l,n}(TE)$ were obtained from quantum mechanical simulations of a spin-echo MR experiment [49]. Basis functions were then computed as in (5.3). Metabolite amplitudes, T_2 values, and the lineshape parameter d_n were chosen to match reported values in the existing literature [17, 125, 126, 145].

Baseline signal was extracted using the Subtract method [25] from the acquired MRS data using PRESS CSI sequence, TE=30 ms, TR=2s. In our experiments the number of the truncated initial FID points used for baseline separation was 30. The extracted baseline component was then fitted with harmonic functions using the HSVD method [21] to obtain parameters b_k with basis functions $v_k[m]$. For the details on the Subtract and HSVD methods, see Sections 2.1.5 and 2.1.6, respectively. We found that 6 baseline components were sufficient to represent the baseline signal accurately.

The *coefficient of variation* (CV), defined as the standard deviation of the estimated metabolite amplitude divided by its mean, is a common metric used in the literature to express quality of parameter estimation. Thus, we consider the square root of the CRB, normalized by the metabolite amplitude, as a bound on the coefficient of variation. We denote this bound as the *coefficient of variation bound* (CVB):

$$CVB_\theta = \frac{\sqrt{CRB_\theta}}{\theta}. \quad (5.10)$$

From the perspective of signal estimation, the desired TE is the one that has the smallest CVB. Since the CVB serves as a lower bound on the actual CV of the quantified metabolite amplitudes, obtaining the lowest CVB should guarantee the lowest theoretical estimation error of the quantitation procedure.

To verify observations from the CRB analysis, we formulate the metabolite quantitation procedure according to the signal model (5.1) as the following nonlinear optimization problem:

$$\{\hat{\mathbf{a}}, \hat{\mathbf{d}}, \hat{\mathbf{b}}\} = \arg \min_{\mathbf{a}, \mathbf{d}, \mathbf{b}} \|\mathbf{s} - \mathbf{Z}_{TE}(\mathbf{d})\mathbf{a} - \mathbf{\Upsilon}\mathbf{b}\|_2^2, \quad (5.11)$$

where the *a priori* known spectral profile $\varphi_{n,TE}[m]$ and signal decay $\psi_{n,d_n}[m]$ were incorporated

into $\mathbf{Z}_{TE}(\mathbf{d})$ as

$$Z_{TE}(\mathbf{d})[m, n] = \varphi_{n, TE}[m] \psi_{n, d_n}[m] \quad (5.12)$$

and the baseline basis functions $v_{k, TE}[m]$ were incorporated into Υ as

$$\Upsilon_{TE}[m, k] = v_{k, TE}[m]. \quad (5.13)$$

Initial guess for $\{\mathbf{a}, \mathbf{d}, \mathbf{b}\}$ was chosen to be the solution in the absence of the signal decay, that is $\mathbf{d}^{(0)} = \mathbf{0}$ and $\mathbf{a}, \mathbf{b}$ are computed as the least-squares inverse of $\mathbf{Z}_{TE}(\mathbf{d}^{(0)})$ given the noisy data $\mathbf{s}$. The minimization problem in (5.11) was solved using the Levenberg-Marquardt optimization routine in Matlab version 7.11.0.584 (R2010b). CVs of the estimated metabolite amplitudes were then computed as

$$CV_{a_n} = \frac{\mathbb{E}\{|\hat{a}_n - a_n|^2\}}{a_n} \quad (5.14)$$

over 1024 noise realizations. Figures 5.2(a)-(c) show the simulated spectra containing common metabolites such as NAA, Cr, Cho, Glu, Gln, and mIns and a baseline at three SNR levels. Figures 5.2(d)-(f) illustrate the performance of the implemented metabolite quantitation with respect to the computed CVBs. The plots show that the CVBs for all metabolites were less than CVs approximately by a factor of 2.

5.3.2 Simple case of one metabolite

To gain insight into the evolution of the CVB with TE, consider a reduced case where there exists only one spectral component, i.e., $N = 1$ and $K = 0$. Due to decreased SNR at long TE, one might expect that as TE increases, the corresponding CVB increases. However, this is not true in general and it depends on how the metabolite profile changes with TE. From the general expression of the elements of the Fisher information matrix shown in (5.9) and (C.4), one can show that the CRB on

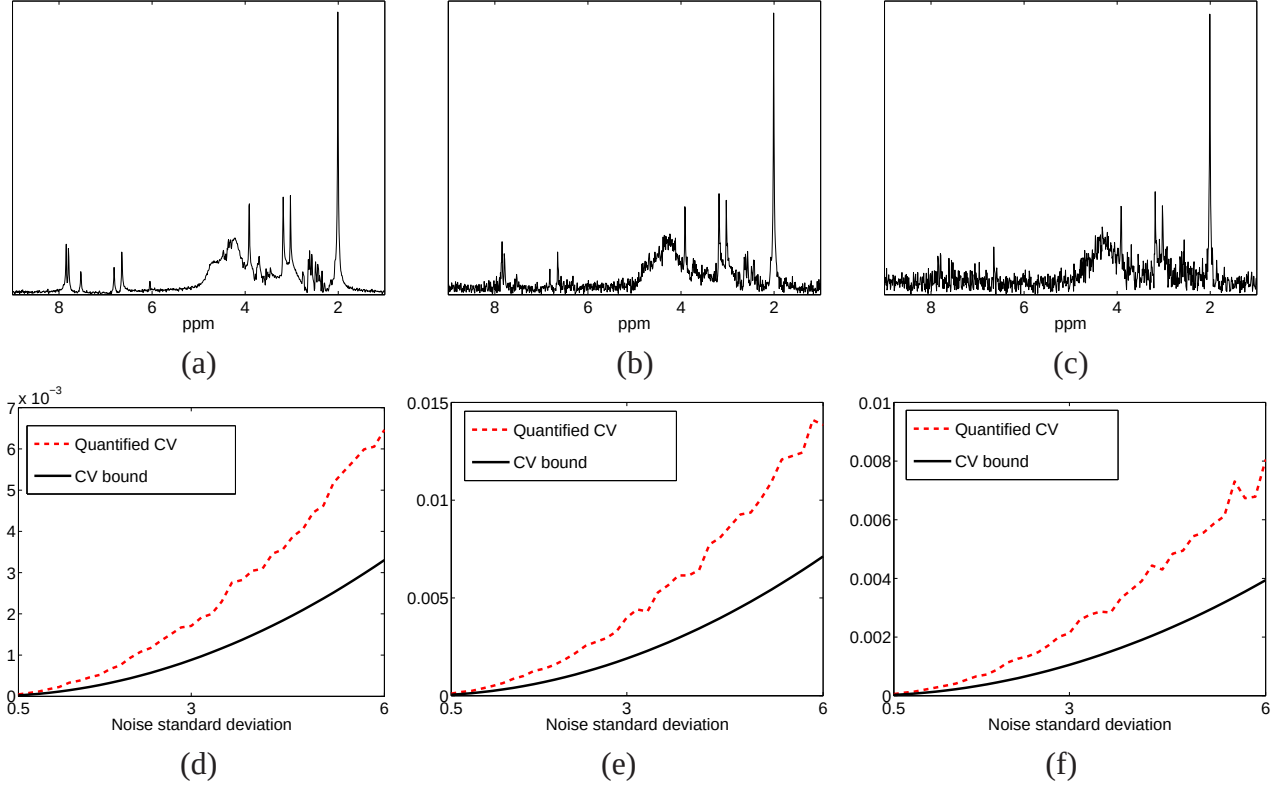


Figure 5.2: Illustration of the performance of metabolite quantitation: (a)-(c) spectra containing resonances from NAA, Cr, Cho, Glu, Glx, mIns, and a simulated baseline at three SNR levels which correspond to the smallest, medium, and largest SNR levels in (d)-(f); (d)-(f) CVs, obtained from metabolite quantitation, versus CVBs as a function of the noise standard deviation. In all cases of different metabolites, the CVBs were lower than the CVs approximately by a factor of 2.

the amplitude can be reduced to the following compact form:

$$CVB_{a_1} = \frac{\sigma_0^2}{2c_1 e^{-TE/T_{2,1}}} \left(\sum_{m=0}^{M-1} |\varphi_{1,TE}[m]|^2 |\psi_{1,d_1}[m]|^2 - \frac{(\sum_{m=1}^{M-1} m |\varphi_{1,TE}[m]|^2 |\psi_{1,d_1}[m]|^2)^2}{\sum_{m=1}^{M-1} m^2 |\varphi_{1,TE}[m]|^2 |\psi_{1,d_1}[m]|^2} \right)^{-1} \quad (5.15)$$

We observe that there are two factors affecting the change in the CVB as TE changes. The first factor is the term $e^{-TE/T_{2,1}}$, which shows that the CVB of the amplitude has an exponential dependence on TE. This factor comes from the inherent loss of signal at long TE. However, the second factor, which is the basis function $\varphi_{1,TE}[m]$, can offset the first factor by changing the three summation series involved in (5.15).

Figure 5.3 shows CVBs and actual quantified CVs of the amplitudes of NAA and Lac as a func-

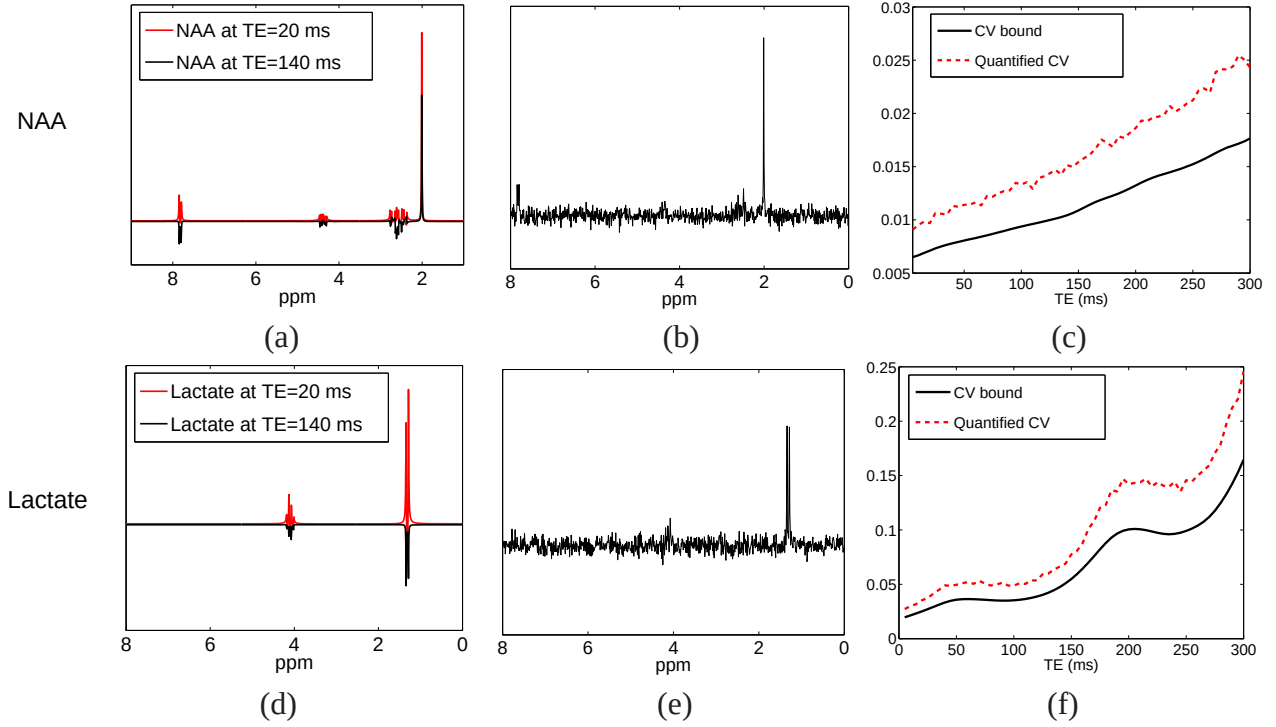


Figure 5.3: CRB analysis of TE selection in the reduced case of one metabolite: (a) spectra of NAA at TE=20 ms (red) and TE=140 ms (black); (b) representative noisy NAA spectrum; (c) CVB (black) and quantified CV (red) on the amplitude of NAA; (d) spectra of Lac at TE=20 ms (red) and TE=140 ms (black), shown in the same scale as (a); (e) representative noisy Lac spectrum; (f) CVB (black) and quantified CV (red) on the amplitude of Lac. Metabolite quantitation was performed with 1024 noise realizations. The non-monotonic behavior of the CVB curve of Lac compared to NAA can be explained by the fact that the spectrum of Lac experiences phase inversion significantly in the region of 1.3 ppm, while the spectrum of NAA does not, except for a decrease in amplitude.

tion of TE, together with their spectra at two representative TEs of 20 ms and 140 ms. Metabolite quantitation was performed with 1024 noise realizations at the noise level shown in Figs. 5.3(b) and (e). Notice that the computed CVB is fairly tight: it is lower than the actual CV roughly by the factor of $1.3 \sim 1.6$ for different TE values. The difference between the CVB curves of NAA and lactate is remarkable. In Figs. 5.3(c) and (f) we observe that the CVB curve of NAA increases monotonically with TE, while the CVB curve of Lac does not. One possible explanation for this difference is that NAA has only one dominant peak which does not change its phase except for a decrease in the amplitude, while two dominant Lac peaks at 1.3 ppm and 4.1 ppm experience phase inversion (see Figs. 5.3(a) and (d)). The experiment shows that even in the reduced case of

one metabolite, an increase in TE does not necessarily increase the CVB. This behavior may not have been expected intuitively. The considered CRB study leads to the remark below.

Remark 1: *In the case of one metabolite, the shortest TE should be preferred because it results in the lowest CVB. If the metabolite spectral profile does not change with TE, then as TE increases, the CVB increases exponentially with a rate $1/T_2$, where T_2 is the T_2 -relaxation decay of the metabolite. Otherwise, an increase in TE may not necessarily increase the CVB. For a particular metabolite, the CRB analysis indicates a range of TE values within which the CVB changes insignificantly (for example, $50 \sim 100$ ms and $215 \sim 250$ ms for Lac, as can be seen from CVB curve in Fig. 5.3(f) in black). Hence, for the purpose of metabolite quantitation, the choice of any TE value in the indicated range makes no significant difference. This observation is further confirmed by the CV curve obtained from metabolite quantitation (see Fig. 5.3(f) in red).*

5.3.3 Case of two metabolites

Spectral overlap is typically considered as a confounding problem in spectroscopy. To gain insight into the effect of the spectral overlap on the CVB, we consider the case of Glu and Gln. It is reported that a strong overlap of Glu and Gln resonances complicates their detection [17], as shown in Figs. 5.4(a),(b). We compute the CVB on the amplitude of Glu as a function of TE when Glu is present in the spectrum alone and when Gln is added to the spectrum. Metabolite quantitation with 1024 noise realizations was performed on the noisy data shown in Figs. 5.4(c),(d). We observe that even when there is overlap between these two metabolites, such as in the regions from 1.8 ppm to 2.6 ppm and from 3.5 ppm to 4 ppm, the CVB increases by a small factor ranging from 1.0 to 1.2 at different TEs for the case of Glu (see Fig. 5.4(e)). Similar CVB behavior is observed for the case when Gln is present in the spectrum alone and when Glu is added to the spectrum, shown in Fig. 5.4(f). Thus, the CVB plots show that overlaps in the spectra do not necessarily cause a large increment in the CVBs. Furthermore, the CVs estimated from metabolite quantitation confirm the same observations, as can be seen from Figs. 5.4(g,h). This leads to the remark below.

Remark 2: *From the CRB perspective, spectral overlap is not a significant obstacle for the*

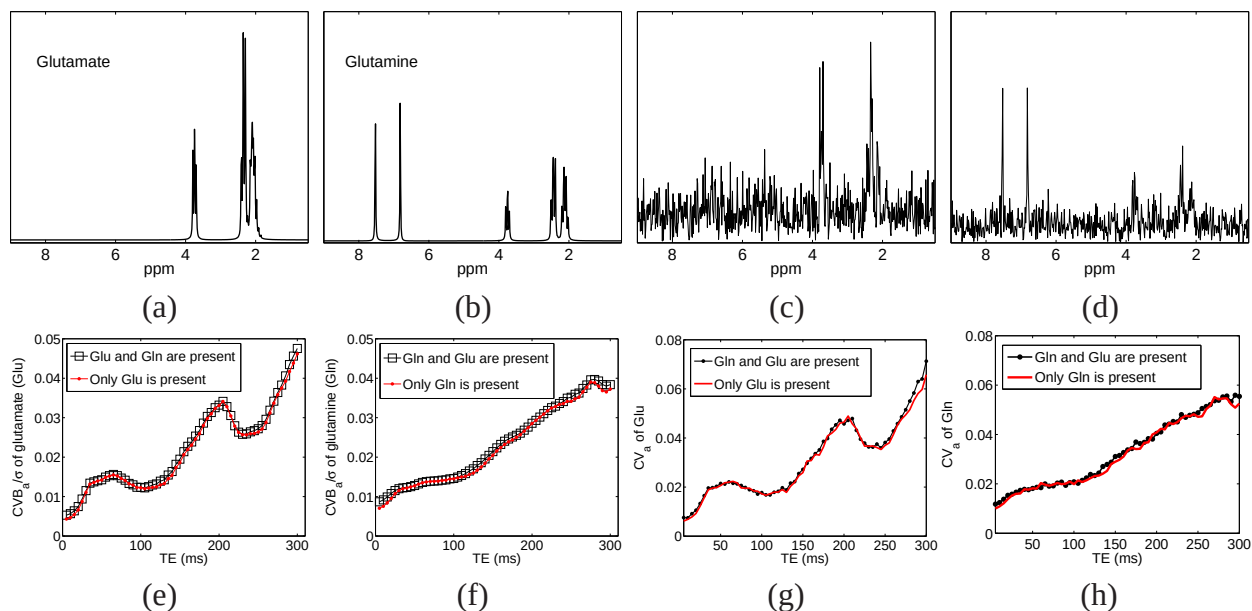


Figure 5.4: Effect of the spectral overlap on the CVB: spectra of (a) Glu and (b) Gln at TE=5 ms, both on the same scale. Spectral overlap is observed in the regions from 1.8 ppm to 2.6 ppm and from 3.5 ppm to 4 ppm; representative noisy spectra of (c) Glu and (d) Gln; (e) CVB on the amplitude of Glu in the presence (in black) and absence (in red) of Gln; (f) CVB on the amplitude of Gln in the presence (in black) and absence (in red) of Glu. (g) CV for Glu in the presence (in black) and absence (in red) of Gln; (h) CV for Gln in the presence (in black) and absence (in red) of Glu. In both Glu and Gln cases, the CVB and CV of one metabolite is not significantly affected by the presence of the other metabolite. This is not expected if judging intuitively from the observed spectral overlap.

quantitation of metabolite concentrations, unlike what visual intuition may suggest.

5.3.4 Effect of the baseline component

Baseline signals originating from the resonance of macromolecules and/or lipids are usually considered as nuisance components in the acquired MRS signal because they overlap with metabolite resonances in the spectrum. For example, Lac peak at 1.3 ppm is an important indicator of several neuro-degenerative diseases such as stroke, ischemia, and hypoxia. Unfortunately, Lac peak usually overlaps with broad lipid resonances that are present in the spectrum or introduced from subcutaneous lipid regions due to inaccurate localization [5]. Resonance of macromolecules typically covers a broad range from 1 ppm to 3 ppm, thus overlapping with a large number of metabolites

such as NAA, Cr, Glu, and Gln. Several typical baseline spectra from macromolecules and lipids in normal white matter, glial brain tumor, and multiple sclerosis are shown in [23].

In order to see how the existence of the baseline affects the choice of TE, we start by analyzing the elements of the Fisher information matrix $\mathbf{F}$ in (C.6). In [146, 147] a sum of pulses and band-limited functions was used to represent radio emissions of the sky. Specifically, pulses in the spatial domain corresponded to stars, while smooth functions represented extended objects such as galaxies and nebulae. The CRB was derived for this particular model to study the fundamental limitations on the precision with which the model parameters can be fitted to the data. Based on the results from [146, 147] and applying to our problem, one can show the following expressions for the elements of $\mathbf{F}$, assuming that the number of FID samples M asymptotically approaches infinity and the sampling rate Δt approaches zero:

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial a_p} \right) \left(\frac{\partial \ln L}{\partial |b|_q} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \left\{ e^{i(b_q - \phi_0)} \sum_{l=1}^{L_p} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \Upsilon_q(f) \right]_{f=f_l(p)} \right\} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial a_p} \right) \left(\frac{\partial \ln L}{\partial \angle b_q} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \left\{ b_q e^{-i\phi_0} \sum_{l=1}^{L_p} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \Upsilon_q(f) \right]_{f=f_l(p)} \right\} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial d_p} \right) \left(\frac{\partial \ln L}{\partial |b|_q} \right)^T \right] &= -\frac{4\pi}{\sigma_0^2} \text{Im} \left\{ \frac{a_p}{d_p^2} e^{i(b_q - \phi_0)} \sum_{l=1}^{L_p} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \dot{\Upsilon}_q(f) \right]_{f=f_l(p)} \right\} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial d_p} \right) \left(\frac{\partial \ln L}{\partial \angle b_q} \right)^T \right] &= \frac{4\pi}{\sigma_0^2} \text{Re} \left\{ \frac{a_p}{d_p^2} b_q e^{-i\phi_0} \sum_{l=1}^{L_p} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \dot{\Upsilon}_q(f) \right]_{f=f_l(p)} \right\} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial |b|_p} \right) \left(\frac{\partial \ln L}{\partial \phi_0} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \left\{ e^{i(\phi_0 - \angle b_p)} \sum_{n=1}^N a_n \sum_{l=1}^{L_n} a_l(n) e^{i\phi_l(n)} \left[\Psi_p(f) \otimes \dot{\Upsilon}_n(f) \right]_{f=f_l(n)} \right\} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \angle b_p} \right) \left(\frac{\partial \ln L}{\partial \phi_0} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \left\{ b_p e^{i\phi_0} \sum_{n=1}^N a_n \sum_{l=1}^{L_n} a_l(n) e^{i\phi_l(n)} \left[\Psi_p(f) \otimes \dot{\Upsilon}_n(f) \right]_{f=f_l(n)} \right\},
\end{aligned} \tag{5.16}$$

where $\Psi_p(f)$ and $\Upsilon_q(f)$ denote the Fourier transform of the functions $\psi_{p,d_p}(t)$ and $v_q(t)$, respectively, and $\dot{\Upsilon}(f)$ denotes the derivative of $\Upsilon(f)$. For derivations of (5.16), see Appendix C.2. From

the shown expressions we observe the following:

1. Coupling between metabolite amplitude a and baseline coefficient b is determined by the value of the baseline spectrum, convolved with metabolite relaxation function (resulting in the broadening of the baseline), evaluated at the frequencies of metabolite peaks.
2. Coupling between metabolite relaxation parameter d and baseline coefficient b or between metabolite global phase ϕ_0 and baseline coefficient b is determined by the value of the *derivative* of the baseline spectrum, convolved with metabolite relaxation function, evaluated at the frequencies of metabolite peaks.
3. In addition, because metabolite amplitudes and baseline coefficients are linear parameters, it can be shown that CRBs on the MSE of a_i are independent of a_j and b_k for all $j = 1, \dots, N$ and $k = 1, \dots, K$ [146].

The above observations lead to the remark below.

Remark 3: *In the presence of the baseline component, CRB (or equivalently, CVB) depends on the baseline smoothness rather than its amplitude.*

To confirm Remark 3, we performed two experiments, both in the presence of a simulated baseline and metabolites NAA, Cr, Cho, Glu, Gln, m-Ins. In the first experiment, the baseline was gradually scaled from small to large amplitude values and a Monte Carlo study was performed with 1024 noise realizations for each of the scaling case. Representative spectra at the smallest, medium, and largest baseline scaling factors are shown in Figs. 5.5(a)-(c). Interestingly, the CVB plots for each metabolite in Figs. 5.5(d)-(f) show that the CVB does not change as the baseline is scaled (black curves), thus confirming Remark 3 partly. In addition, the quantified CVs (red curves) do not have a clear increasing or decreasing trend as the baseline is scaled, but rather there are only small fluctuations in the values of the CVs in a random manner. From this experiment, we suggest that baseline amplitude should not affect the choice of TE from a CRB perspective.

In the second experiment, we consider the spectrum containing a single metabolite peak and the same baseline used in the first experiment. The metabolite peak was gradually displaced 20

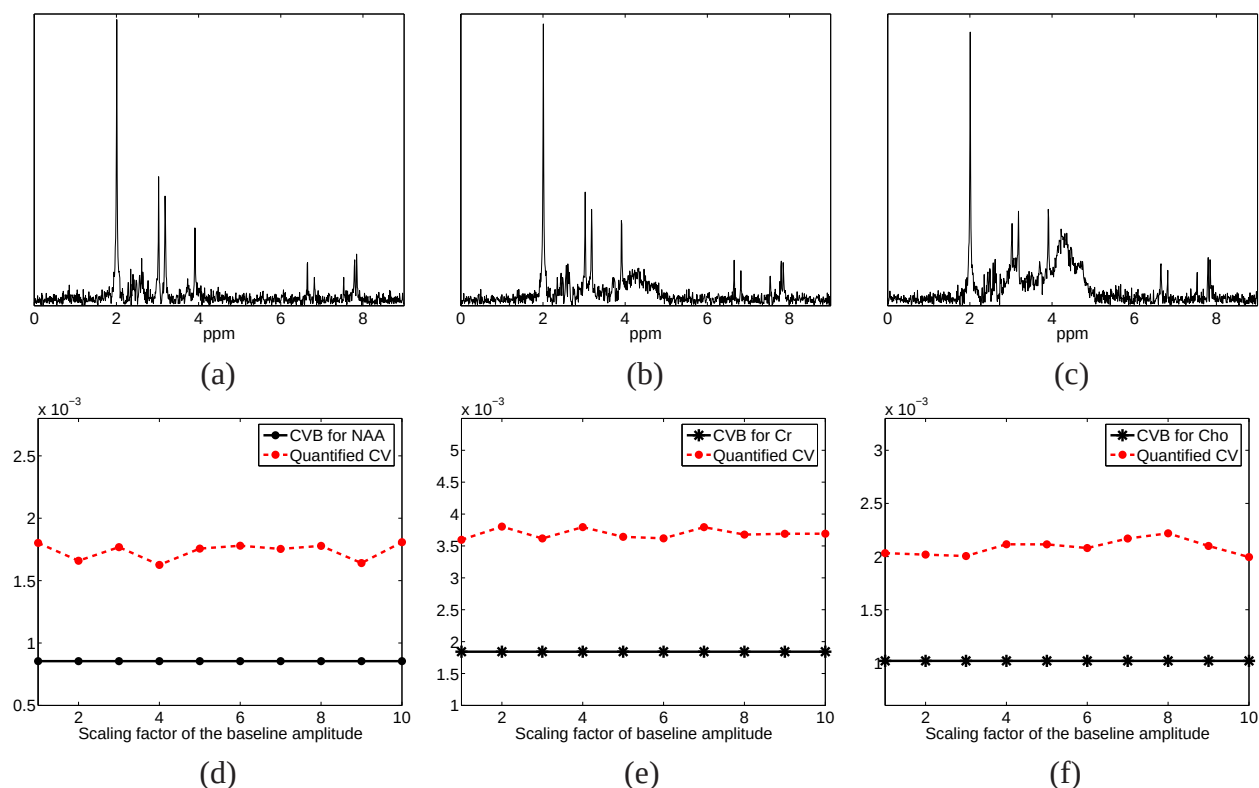


Figure 5.5: Monte Carlo study of the effect of the baseline amplitude on the CRB: top row shows representative noisy spectra containing NAA, Cr, Cho, Glx, m-Ins, and a simulated baseline at different scaling factors for the baseline amplitude; bottom row shows the CVBs on the amplitudes of (d) NAA, (e) Cr, and (f) Cho and the corresponding CVs after performing metabolite quantitation with 1024 noise realizations. Notice that the CVBs did not change as the baseline amplitude was scaled. Hence, we suggest that baseline amplitude should not affect the choice of TE from the CRB perspective.

times, each time by 50 Hz, to overlap with the baseline resonance, and the CVB was computed accordingly. Quantitation was then performed with 1024 noise realizations to obtain the CV for each displacement case. Representative spectra are shown in Figs. 5.6(a)-(e). Figure 5.6 shows that (i) as the peak is outside of the baseline region (2 ppm to 6 ppm), the CVB is small and almost does not change; (ii) as the peak is moved into the baseline region, the CVB increases and changes according to the change in the baseline smoothness; (iii) as the peak is moved out from the baseline region, the CVB decreases to the values that are typical for the case (i). Interestingly, a roughly similar trend is observed in the quantified CV curve. This experiment confirms Remark 3 completely.

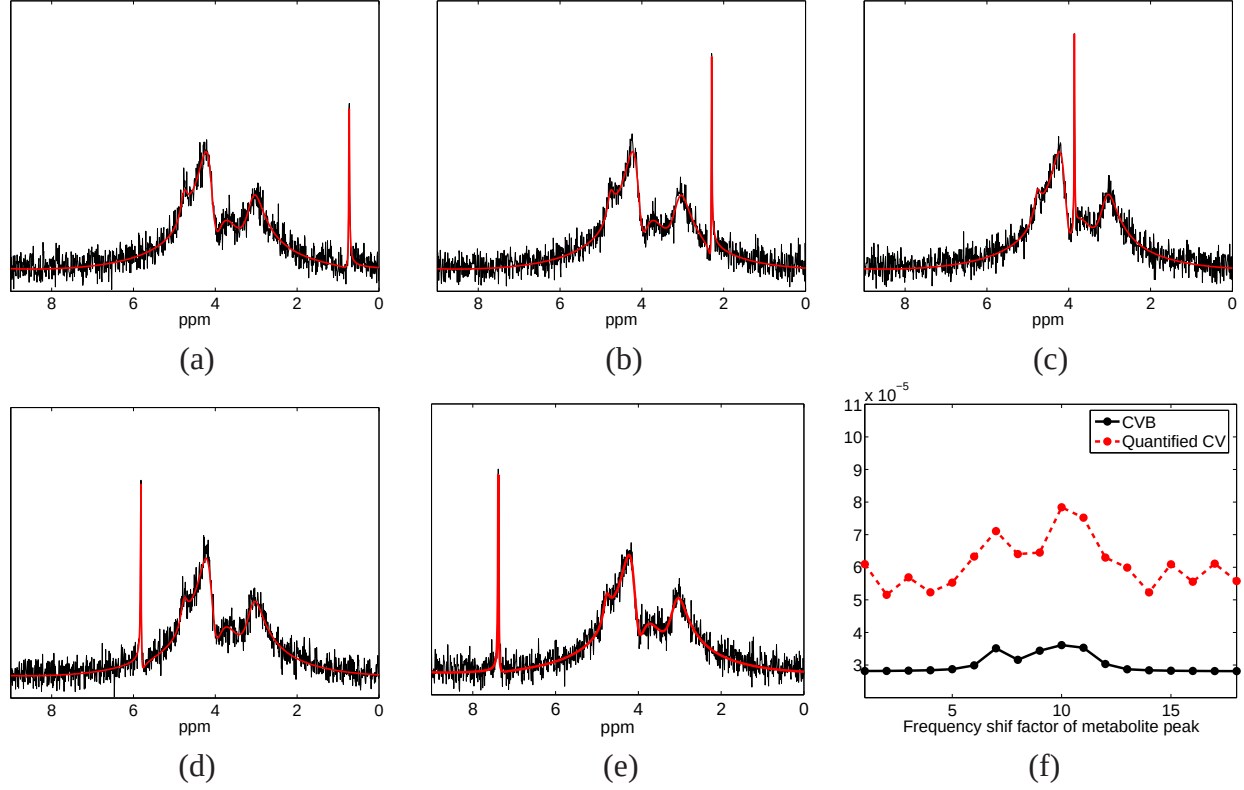


Figure 5.6: Monte Carlo study of the effect of the baseline smoothness on the choice of TE from the CRB perspective: (a)-(e) representative noisy spectra containing a simulated baseline and a metabolite peak, which is gradually displaced across the whole frequency range; (f) CVB and quantified CV value for each case of the peak displacement. Quantitation was performed with 1024 noise realizations. Notice that as the peak moves closer to the dominant resonance range of the baseline (2 ppm to 6 ppm), CVB increases and changes according to the change in the baseline smoothness.

To analyze the increment amount of the CVB caused by the baseline, we reconsidered the case of 6 metabolite resonances (NAA, Cr, Cho, Glu, Gln, mIns) and in addition to the baseline originating from macromolecules, simulated in the first experiment and shown in Fig. 5.7(a), we also considered baseline contributions from lipids. Lipid components were simulated based on the values reported in [23], where the parametrization of the metabolite-nulled spectra of high-grade gliomas with large dominating lipids and negligible physiological macromolecules was obtained. This lipid pattern was observed to be consistent with only mild variations across different spectra. Specifically, frequencies and linewidths of the lipid peaks in this experiment were simulated according to Table 5.1, resulting in the lipid spectrum shown in Fig. 5.7(b) and the total spectrum

shown in Fig. 5.7(c).

Table 5.1: Parameters for the simulation of lipid components.

Lipid components	Frequency (ppm)	Linewidth (Hz)	Lineshape	Relative amplitude
Lip1	0.89	79.6	Lorentzian	0.43
Lip2	1.30	49.6	Lorentzian	1
Lip3	2.05	26.2	Gaussian	0.06
Lip4	2.24	33.10	Gaussian	0.07
Lip5	2.81	32.2	Gaussian	0.04

Figures 5.7(d)-(f) and 5.8(a)-(c) show the ratio of the CVB on the variance of each metabolite amplitude in the presence of both metabolites and baseline versus the CVB in the presence of metabolites only. One can see that for each metabolite, the presence of the baseline increases the CVB by an insignificant factor ranging maximally to 1.21 for metabolites that strongly overlap with the baseline. This experiment suggests that as long as we can accurately extract the baseline and model it using *a priori* chosen basis functions, the existence of the baseline should not be a significant obstacle for metabolite quantitation, since the baseline is smooth and its smoothness does not change significantly with TE.

5.3.5 Effect of multi-echo data

In practice, with a typical TR=2 s one can consider acquiring single-voxel MRS data from several TEs. A practical question arises in this case: does the multi-echo MRS data help to improve metabolite quantitation? And if it does, then what is the maximally achievable improvement factor?

To analyze these questions, we follow (5.1) and model the observed MRS signal acquired at an echo time TE_i as

$$s_{TE_i}[m] = e^{i\phi_0} \sum_{n=1}^N c_n e^{-TE_i/T_{2,n}} \varphi_{n,TE_i}[m] \psi_{n,d_n}[m] + \sum_{k=1}^K b_k v_{k,TE_i}[m] + \xi_{TE_i}[m]. \quad (5.17)$$

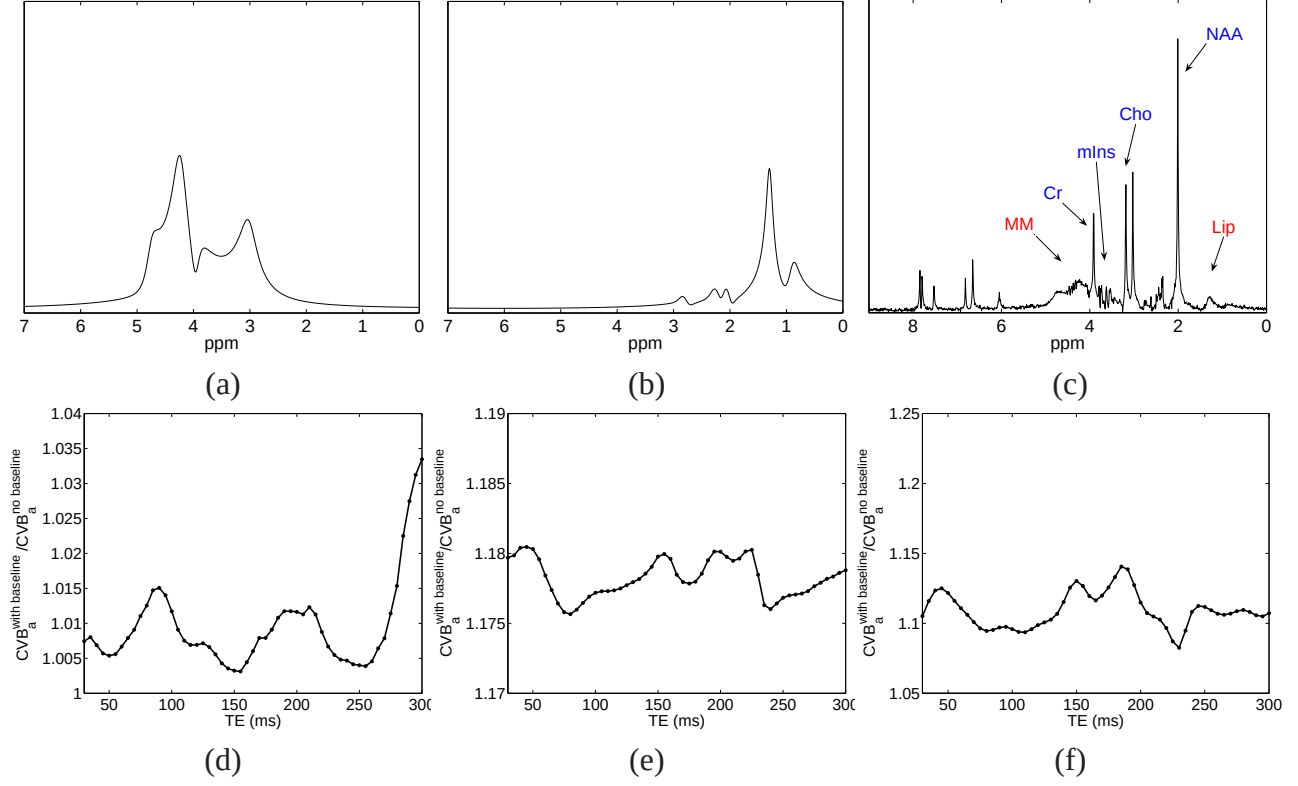


Figure 5.7: Baseline results in an increase in the CVB by a small factor: spectrum of the simulated baseline component originating from (a) macromolecules (MM) and (b) lipids (Lip); (c) spectrum containing simulated resonances of metabolites, MM, and Lip; (d)-(f) ratio of the CVB on the amplitudes of NAA, Cr, Cho, Glu, Gln, and mIns, respectively, in the case when the baseline is present versus the case when the baseline is absent.

In (5.17) we assumed that coefficients c_n and b_k are independent from TE_i , while the decay of metabolite and baseline signals as TE_i increases is embedded in the rest of the terms (metabolite spectral profile $\varphi_{n,TE_i}[m]$, metabolite decay term $e^{-TE_i/T_{2,n}}$, and baseline basis function $v_{k,TE_i}[m]$).

Rewriting (5.17) in matrix form, we have

$$s_{TE_i} = \left(\Phi_{TE_i} \odot \Psi \odot Q_{TE_i} \right) e^{i\phi_0} \mathbf{c} + \Upsilon_{TE_i} \mathbf{b} + \xi_{TE_i}, \quad (5.18)$$

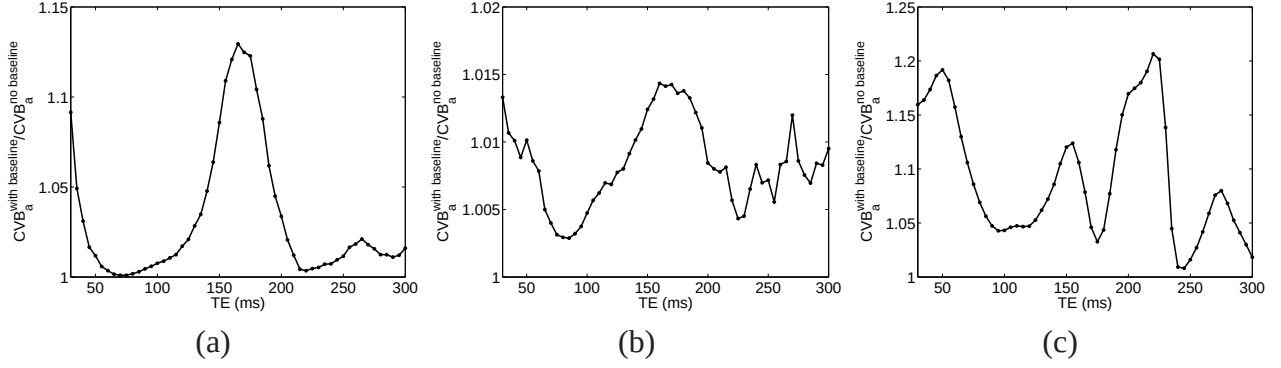


Figure 5.8: Baseline leads to an increase in the CVB by a small factor (continuation of Fig. 5.7): (a)-(c) ratio of the CVB on the amplitudes of Glu, Gln, and mIns, respectively, in the case when the baseline is present versus the case when the baseline is absent.

where Φ_{TE_i} , Ψ , Q_{TE_i} , and Υ_{TE_i} are $M \times N$ matrices defined as

$$\Phi_{TE_i} = \begin{bmatrix} \varphi_{1,TE_i}[0] & \varphi_{2,TE_i}[0] & \dots & \varphi_{N,TE_i}[0] \\ \varphi_{1,TE_i}[1] & \varphi_{2,TE_i}[1] & \dots & \varphi_{N,TE_i}[1] \\ \dots & \dots & \dots & \dots \\ \varphi_{1,TE_i}[M-1] & \varphi_{2,TE_i}[M-1] & \dots & \varphi_{N,TE_i}[M-1] \end{bmatrix}, \quad (5.19)$$

$$\Psi = \begin{bmatrix} \psi_{1,d_1}[0] & \psi_{2,d_2}[0] & \dots & \psi_{N,d_N}[0] \\ \psi_{1,d_1}[1] & \psi_{2,d_2}[1] & \dots & \psi_{N,d_N}[1] \\ \dots & \dots & \dots & \dots \\ \psi_{1,d_1}[M-1] & \psi_{2,d_2}[M-1] & \dots & \psi_{N,d_N}[M-1] \end{bmatrix}, \quad (5.20)$$

$$Q_{TE_i} = \begin{bmatrix} e^{-TE_i/T_{2,1}} & e^{-TE_i/T_{2,2}} & \dots & e^{-TE_i/T_{2,N}} \\ e^{-TE_i/T_{2,1}} & e^{-TE_i/T_{2,2}} & \dots & e^{-TE_i/T_{2,N}} \\ \dots & \dots & \dots & \dots \\ e^{-TE_i/T_{2,1}} & e^{-TE_i/T_{2,2}} & \dots & e^{-TE_i/T_{2,N}} \end{bmatrix}, \quad (5.21)$$

$$\mathbf{r}_{TE_i} = \begin{bmatrix} v_{1,TE_i}[0] & v_{2,TE_i}[0] & \dots & v_{K,TE_i}[0] \\ v_{1,TE_i}[1] & v_{2,TE_i}[1] & \dots & v_{K,TE_i}[1] \\ & & \dots & \\ v_{1,TE_i}[M-1] & v_{2,TE_i}[M-1] & \dots & v_{K,TE_i}[M-1] \end{bmatrix}. \quad (5.22)$$

Denoting

$$\begin{aligned} \tilde{\mathbf{s}} &= \begin{bmatrix} \mathbf{s}_{TE_1}^T & \mathbf{s}_{TE_2}^T & \dots & \mathbf{s}_{TE_D}^T \end{bmatrix}^T \\ \tilde{\boldsymbol{\xi}} &= \begin{bmatrix} \boldsymbol{\xi}_{TE_1}^T & \boldsymbol{\xi}_{TE_2}^T & \dots & \boldsymbol{\xi}_{TE_D}^T \end{bmatrix}^T \\ \tilde{\boldsymbol{\Phi}} &= \begin{bmatrix} \boldsymbol{\Phi}_{TE_1}^T & \boldsymbol{\Phi}_{TE_2}^T & \dots & \boldsymbol{\Phi}_{TE_D}^T \end{bmatrix}^T \\ \tilde{\boldsymbol{\Psi}} &= \begin{bmatrix} \boldsymbol{\Psi}^T & \boldsymbol{\Psi}^T & \dots & \boldsymbol{\Psi}^T \end{bmatrix}^T \\ \tilde{\mathbf{Q}} &= \begin{bmatrix} \mathbf{Q}_{TE_1}^T & \mathbf{Q}_{TE_2}^T & \dots & \mathbf{Q}_{TE_D}^T \end{bmatrix}^T \\ \tilde{\mathbf{Y}} &= \begin{bmatrix} \mathbf{r}_{TE_1}^T & \mathbf{r}_{TE_2}^T & \dots & \mathbf{r}_{TE_D}^T \end{bmatrix}^T, \end{aligned} \quad (5.23)$$

we obtain the following equation which describes the observed data at echo times $TE_1, TE_2, \dots, TE_D$:

$$\tilde{\mathbf{s}} = \left(\tilde{\boldsymbol{\Phi}} \odot \tilde{\boldsymbol{\Psi}} \odot \tilde{\mathbf{Q}} \right) e^{i\phi_0} \mathbf{c} + \tilde{\mathbf{Y}} \mathbf{b} + \tilde{\boldsymbol{\xi}}. \quad (5.24)$$

Given the signal model (5.24), the unknown parameters are

$$\boldsymbol{\theta} = \{c_1, \dots, c_N, d_1, \dots, d_N, \phi_0, |b_1|, \dots, |b_K|, \angle b_1, \dots, \angle b_K\}.$$

From (5.24), one can derive expressions for the elements of the Fisher information matrix in the multi-echo case. Following equation shows one particular element:

$$\{\mathbf{F}\}_{a_i a_j} = \frac{2}{\sigma_0^2} \text{Re} \left\{ \sum_{d=1}^D \sum_{m=0}^{M-1} \varphi_{i,TE_d}^*[m] \psi_{i,d_i}^*[m] \varphi_{j,TE_d}[m] \psi_{j,d_j}[m] \right\}. \quad (5.25)$$

Other elements of $\mathbf{F}$ have similar expressions, shown in Appendix C.3. Consider the case when $\{TE_d\}_{d=1}^D$ are all equal to TE_0 . Then we can rewrite (5.25) as

$$\{\mathbf{F}\}_{a_i a_j} = \frac{2}{\sigma_0^2/D} \text{Re} \left\{ \sum_{m=0}^{M-1} \varphi_{i,TE_0}^*[m] \psi_{i,d_i}^*[m] \varphi_{j,TE_0}[m] \psi_{j,d_j}[m] \right\}. \quad (5.26)$$

Comparing (5.26) with (5.9) and applying similar derivations to the other elements in $\mathbf{F}$, we obtain the remark below.

Remark 4: *The CRB is reduced by a factor of D if MRS data is acquired at an echo time TE_0 multiple (D) times, compared to the case when the data is acquired at TE_0 only once. Equivalently, the CVB is reduced by a factor of $\sqrt{D}$. In other words, the scenario when the MRS data is acquired at the same echo time D times is equivalent to the case when the data is acquired only once but with noise variance reduced by a factor of D . In this sense, acquiring multi-echo MRS data has similar effect as signal averaging.*

Figure 5.9 shows the computed CVB and the quantified CV as a function of D . The spectrum contained the resonances from NAA, Cr, Cho, Glu, Gln, mIns, Lac, and a baseline, as shown in Fig. 5.9(a). As expected, the CVB curve decreases at the rate $1/\sqrt{D}$ for all metabolites. This confirms Remark 4.

In addition to the reduction of the CRB due to an “implicit” averaging, a further reduction of the CRB is achieved if the echo times are chosen so that the spectral overlap (inner product) among spectral components is more resolved compared to the single-echo case. This fact is confirmed in Fig. 5.10, where we considered the resonance of two metabolites in three cases: (i) case of a single TE, (ii) case of double TEs with $TE_1 = TE_2$, and (iii) case of double TEs with $TE_1 \neq TE_2$ and the spectral overlap between two metabolites at TE_2 was reduced. As expected, case (i) yielded the largest CVB, while case (ii) had a $\sqrt{2}$ -fold reduction in the CVB for each metabolite and case (iii) provided an additional, slight reduction in the CVB compared to the case (ii) due to the fact that the TEs were chosen to resolve the overlap between two metabolite spectral profiles. Notice that this additional reduction factor for the experiment considered here was small (around 1.1). This

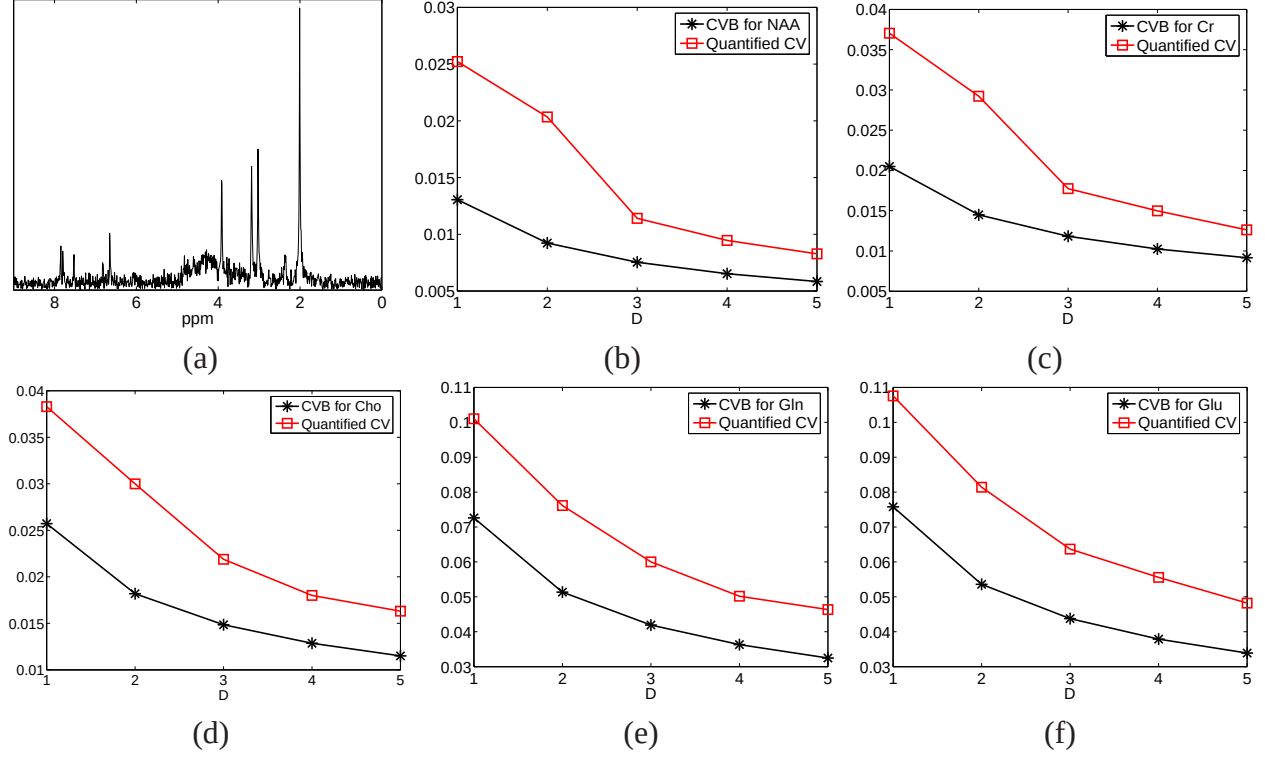


Figure 5.9: Multi-echo acquisition at echo times $\{\text{TE}_i\}_{i=1}^D = 30$ ms: (a) Representative noisy spectrum, containing resonances from NAA, Cr, Cho, Glu, Gln, mIns, Lac, and a baseline; (b)-(f) CVBs and quantified CVs on the amplitudes of NAA, Cr, Cho, Glu, and Gln, respectively, as a function of D . As expected, the CVB curve decreases at the rate $1/\sqrt{D}$ for all metabolites.

factor would be more significant if many other metabolites with resolved spectral overlap were present.

5.3.6 General case of many metabolites

To gain insight into the CVB evolution as a function of TE in a more complete model, we consider a spectrum containing all metabolites and the baseline, listed in Section 5.3.1. In addition, to examine whether a reduced model order can lead to a further decrease in the CVB, we also consider two additional models. The first reduced model has 7 commonly reported metabolites that give rise to large signals at short TE: *NAA*, *Cho*, *Cr*, *Glu*, *Gln*, *mIns*, and *Lac* [7]. A further reduced model consists of 4 commonly observed metabolites that have long T_2 values and contribute mainly to large signals at long TE: *NAA*, *Cho*, *Cr*, and *Lac* [7].

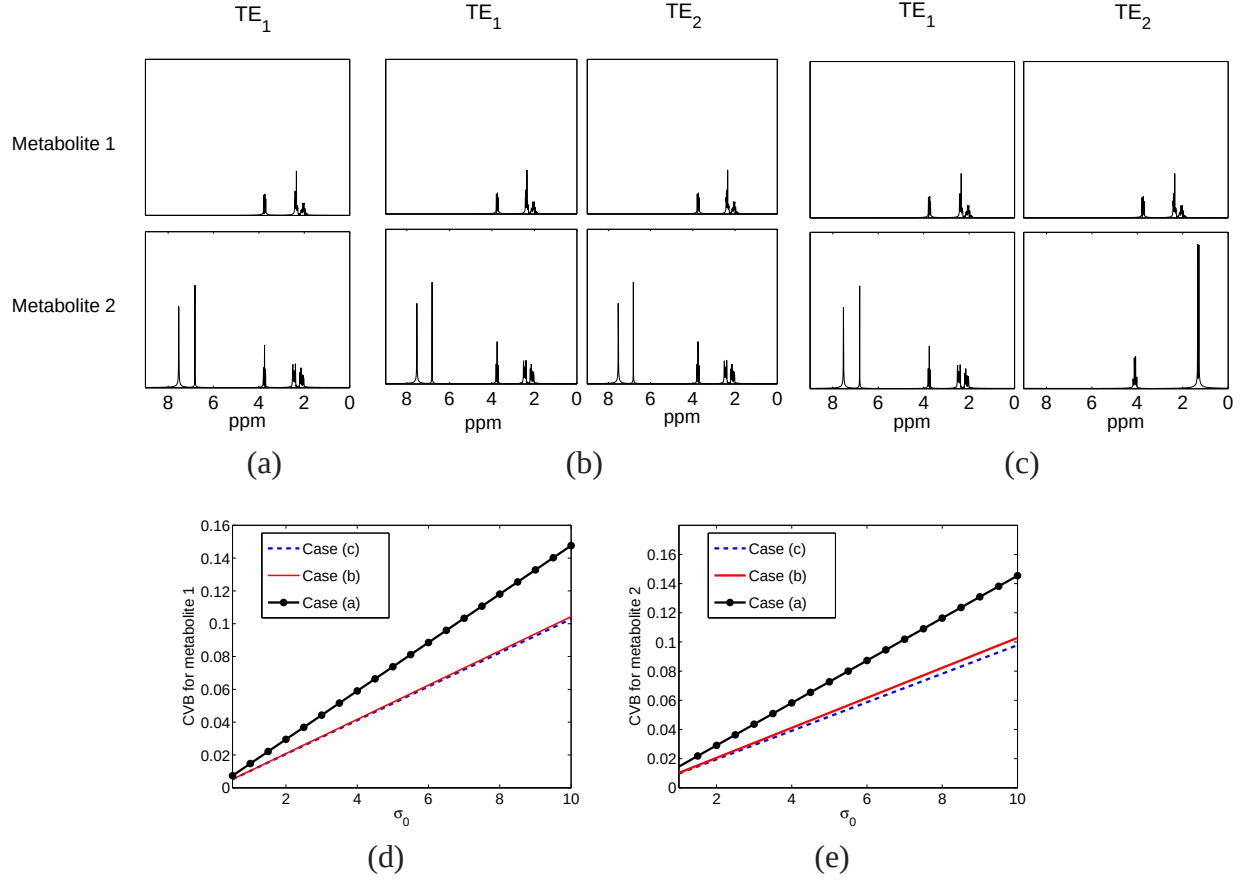


Figure 5.10: Effect of multi-echo data on the CVB: metabolite spectral profiles in the case of (a) a single TE; (b) double TEs with $TE_1 = TE_2$; (c) double TEs with $TE_1 \neq TE_2$ and the spectral overlap (inner product) between two metabolites at TE_2 is reduced. CVB plots of the amplitudes of (d) metabolite 1 and (e) metabolite 2 show that CVB is reduced at least by a factor of $\sqrt{2}$ for the double-echo acquisition. An additional (although small) reduction factor is obtained if one chooses TE_2 so that the spectral overlap between two metabolites is reduced, as is the case in (c).

Theoretically, it is clear that the reduced model describes the observed MR resonance process less accurately than the original model with more metabolites. In practice though, any loss of accuracy of the reduced model decreases at longer TE, because the signal contribution from metabolites with short T_2 becomes very small. However, the exact value of TE at which we can reliably assume a reduced model is still an open question and needs future investigation.

Figure 5.11 shows CVBs of the amplitude of NAA, Cho, and Lac for the original model with 14 metabolites and the baseline, the reduced model with 7 metabolites, and the reduced model with 4 metabolites. In the case of NAA we see that reducing the model order from 14 to 7 or

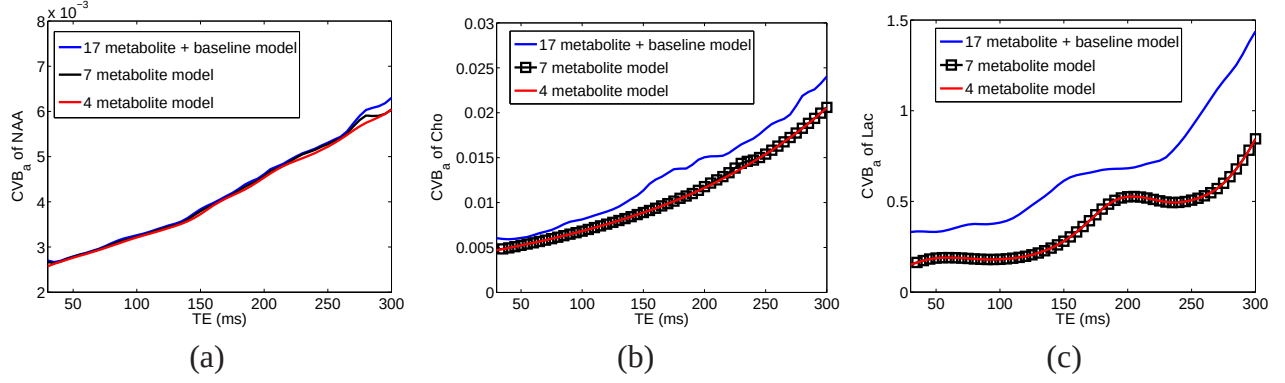


Figure 5.11: CVB on the amplitude of (a) NAA, (b) Cho, and (c) Lac in the case of regular (17 metabolites and baseline) and reduced (7 and 4 metabolites) models. The CVB curve decreases as the model order decreases. For clarity, this figure should be viewed in color.

to 4 metabolites did not yield a noticeable improvement in the CVB. This can be explained due to the unique spectral profile of NAA that does not strongly couple with other metabolite spectral profiles. In the cases of Cho and Lac we see that reducing the model order from 14 to 7 metabolites yielded a smaller CVB at each TE by a factor ranging from 1.2 to 1.7. This confirms that the decreased correlation between metabolites lowers the CVB. However, further reducing the model order from 7 to 4 metabolites did not yield a noticeable improvement, as can be seen from Fig. 5.11. Assuming we know the value of TE at which the reduced model is reasonable, we can tell from the corresponding CVB plots whether using the reduced model with prolonged TE could yield a smaller CVB compared to the original model at short TE. Notice that because the CVB curves tend to increase as TE increases, the CVB criterion suggests that excessively long TE values should not be chosen even when using the reduced model.

5.3.7 Case study

In previous sections we proposed the CRB framework as a quantitative tool to choose the optimal TE from an estimation theoretic perspective and we analyzed several factors affecting the CRB. However, a particular value for the optimal TE depends on a specific scenario such as the metabolites of interest, their spectral overlap in the observed spectrum, and the baseline characteristics. In this section, we consider a case study with a specific MRS signal of interest. Notice that the goal

of this section is to demonstrate the use of the proposed CRB methodology, thus the selected TE values provided in this section should be treated as optimal for the specific case study considered here only.

In vivo MRS data from a voxel in the region of GM of the healthy human brain was acquired using a Siemens Magnetom Allegra 3 T MRI scanner. A PRESS CSI sequence was employed with CHESS water suppression (water suppression bandwidth=35 Hz), FOV= $18 \times 18 \times 18 \text{ mm}^3$, TR=2 s, bandwidth=2,400 Hz, 2068 FID data points, and 128 averages. The CSI dataset was processed to remove residual water resonance using HSVD [6]. Specifically, the acquired time-domain signal was modeled using Lorentzian basis functions as in (2.21) with a_n and z_n defined according to (2.18). A data fit to this model was performed using a state-space HSVD method [21], where an SVD decomposition of the data matrix

$$\mathbf{H} = \begin{bmatrix} s[0] & s[1] & \dots & s[M/2] \\ s[1] & s[2] & \dots & s[M/2 + 1] \\ \vdots & \vdots & \vdots & \vdots \\ s[M/2 - 1] & s[M/2] & \dots & s[M - 1] \end{bmatrix} \quad (5.27)$$

with $M = 2068$ was performed as

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \quad (5.28)$$

from which the submatrices $\mathbf{U}^\uparrow$ and $\mathbf{U}_\downarrow$ were computed by removing the first and last rows of $\mathbf{U}$, respectively. Coefficients z_n were then estimated as eigenvalues of the matrix $\mathbf{Q}$ satisfying

$$\mathbf{U}^\uparrow = \mathbf{U}_\downarrow \mathbf{Q}. \quad (5.29)$$

Given z_n , coefficients a_n were estimated from (2.21) as the least-squares estimates. Each set $\{a_i, z_i\}$ represented a Lorentzian basis component of the signal. The largest such component corresponded to the water resonance and thus was removed. As we can see from Fig. 5.12, a large portion of the water peak at 4.7 ppm has been removed and metabolite resonances are clearly seen

in the spectrum shown in Fig. 5.12(b).

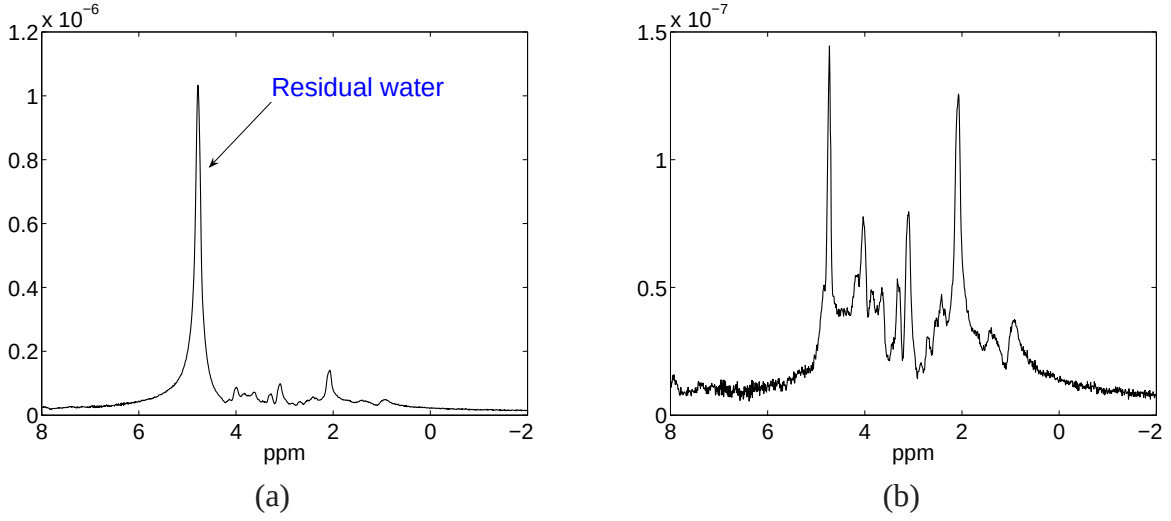


Figure 5.12: Illustration of water suppression using the HSVD method: (a) Acquired spectrum with water suppression frequency-selective pulse resulting in a large residual water peak; (b) water suppressed spectrum after performing HSVD. Notice that metabolite resonances can be clearly observed in (b).

Figures 5.13(a)-(c) show the resulting spectra after water removal at TE=30, 144, and 288 ms, respectively. Several metabolite peaks were identified such as NAA, Cr, Cho, Glu, Gln, and mIns. Notice that the spectrum at TE=30 ms also contains a significant amount of the baseline, while spectra at TE=144 ms and 288 ms can be considered as containing metabolite components only. In order to separate the baseline and metabolite components for the case of TE=30 ms, we followed the Subtract method described in Section 2.1.5. Below we summarize the main steps of this baseline suppression method:

- Truncate an appropriate number N_{trunc} of initial time-domain data points (in this study, $N_{\text{trunc}} = 30$) to obtain the signal s_{trunc} , shown in Fig. 5.13(d).
- Back-extrapolate s_{trunc} to time $t = 0$, obtaining an approximation of the metabolite component s_{metab} , shown in Fig. 5.13(e).
- Subtract s_{metab} from the original time-domain signal s to obtain the baseline component s_{baseline} .

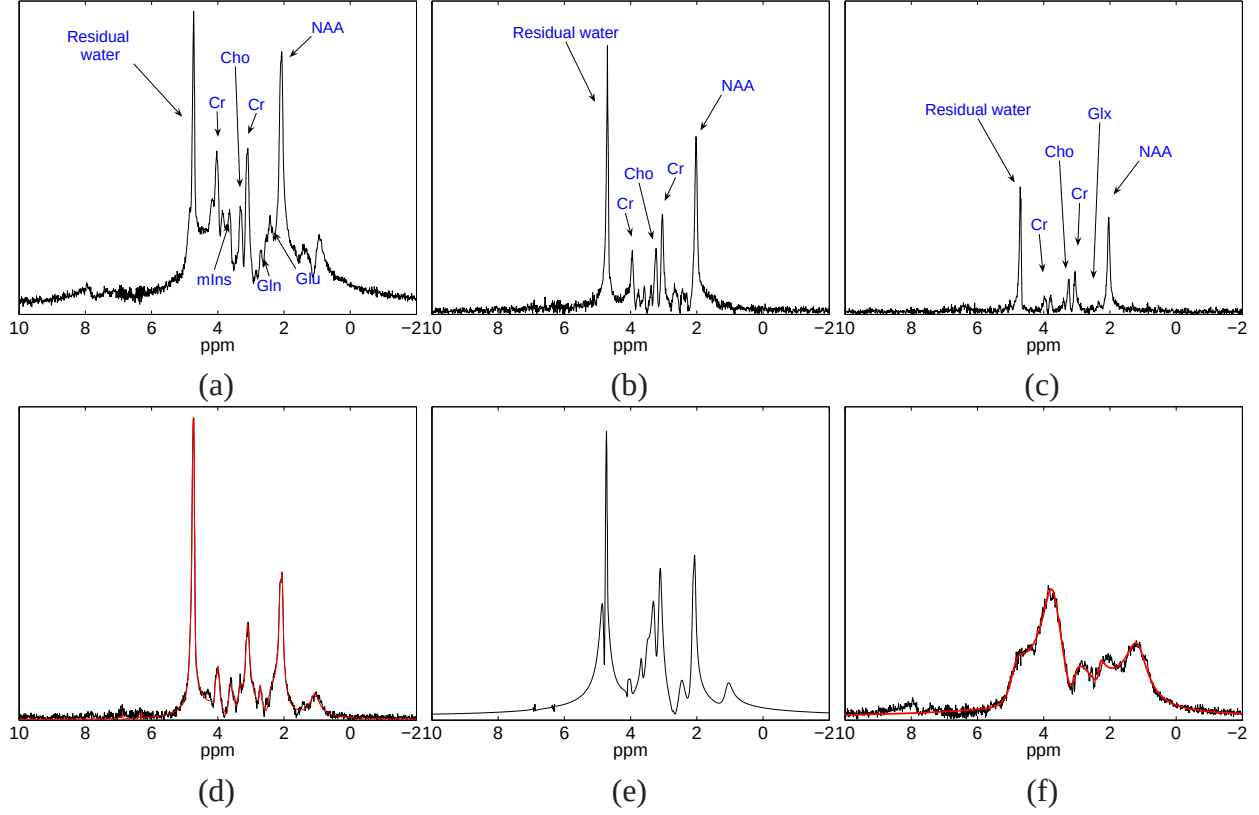


Figure 5.13: *In vivo* MRS spectra at (a) TE=30 ms; (b) TE=144 ms; (c) TE=288 ms. Baseline extraction was performed on the spectrum acquired at TE=30 ms: (d) spectrum after removing 30 initial FID points; (e) back-extrapolated spectrum, containing metabolite components; (f) extracted baseline component after subtracting (e) from (a).

- Parametrize s_{baseline} using 6 Lorentzian basis functions according to the HSVD method mentioned in Section 2.1.6. The resulting parametrized baseline is shown in Fig. 5.13(f).
- Subtract the parametrized baseline signal from s to yield the extracted metabolite signal s_{subt} .

Based on the extracted metabolite signals at TE=30, 144, and 288 ms, we assigned metabolite amplitudes a_n and lineshape parameters d_n to the simulated (using GAVA) metabolite spectral profiles $\varphi_{n,TE}[m]$. Table 5.2 shows the chosen values for a_n and d_n . In this case study, we assumed Lorentzian lineshapes with $d_n = 1/T_{2,n}$. In addition, a frequency shift was applied to the simulated metabolite spectral profiles to align them with the corresponding resonances of the metabolite signal that is extracted from the experimental data.

Table 5.2: Metabolite amplitudes a_n and lineshape parameters d_n used in the case study. Cases when the metabolite is not present are denoted as “NA”.

Metabolite	a_n at TE=30 ms	a_n at TE=144 ms	a_n at TE=288 ms	d_n (Hz)
NAA	50	40	24	16.7
Cr	27	22	9	16.7
Cho	10	10	6	33.3
Glu	10	2	NA	50
Gln	16	4	NA	55.6
mIns	14	3	NA	52.6
Scy	8	NA	NA	100
Asp	5	NA	NA	83.3
Tau	6	NA	NA	83.3
Glc	8	NA	NA	66.7
His	10	NA	NA	66.7
Thr	12	NA	NA	58.8
Eth	11	NA	NA	58.8

Based on the simulated parameters described above, we computed the CVBs for each metabolite in order to determine the optimal TE. Table 5.3 shows the resulting CVBs. One can see that the optimal TE for high-SNR metabolites such as NAA, Cr, Cho was 144 ms. This can be explained by the fact that there was a significant reduction of the spectral overlap among metabolites at TE=144 ms (signal from the resonance of 7 metabolites and the baseline had decayed at TE=144 ms). However, prolonging TE more to 288 ms did not yield a minimum CVB because the SNR was not sufficient at this TE. Notice also that the CVB for NAA improved by an insignificant factor of

Table 5.3: Case study - resulting CVBs. Cases when the metabolite is not present in the spectrum are denoted as “NA”.

Metabolite	CVB at TE=30 ms	CVB at TE=144 ms	CVB at TE=288 ms	Optimal TE
NAA	0.1967×10^{-5}	0.1786×10^{-5}	0.2418×10^{-5}	144 ms
Cr	0.6512×10^{-5}	0.2274×10^{-5}	0.4794×10^{-5}	144 ms
Cho	0.1854×10^{-5}	0.0173×10^{-5}	0.0203×10^{-5}	144 ms
Glu	0.0023×10^{-5}	0.0013×10^{-5}	NA	144 ms
Gln	0.1218×10^{-5}	0.1910×10^{-5}	NA	30 ms
mIns	0.0012×10^{-5}	0.0002×10^{-5}	NA	144 ms

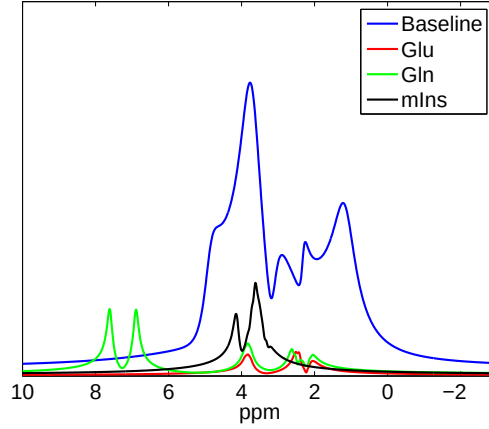


Figure 5.14: Glu and mIns resonances overlap with the baseline completely, while the spectral overlap of Gln with the baseline is reduced due to the presence of two resonance peaks of Gln at 6.9 ppm and 7.62 ppm.

1.1 at TE=144 ms as compared to the case of TE=30 ms. This behavior was observed before (see Fig. 5.11) and can be explained by the fact that NAA has a unique spectral profile that does not strongly couple with other spectral components. For low-SNR metabolite such as Gln, the optimal TE was 30 ms. In this case, because the SNR of Gln peak was already low, a longer TE did not help to reduce the CVB. One would expect the same conclusion with low-SNR metabolites such as Glu and mIns; however, in these two cases there was a significant spectral overlap between each of these two metabolites with the baseline, as shown in Fig. 5.14. One can see that both Glu and mIns resonances lacked the distinct spectral peaks outside of the baseline region that were observed for Gln at around 6.9 ppm and 7.62. These two peaks helped to reduce the spectral overlap between Gln and the baseline.

5.4 Summary

In proton MRS, echo time (TE) is a crucial parameter that influences the appearance of the observed spectrum. In this chapter, we provided a quantitative analysis of the effect of different TE values. Specifically, based on estimation theory, we used the CRB to compute a bound on the coefficient of variation of the unknown spectral amplitudes as a function of TE. We showed that the

form of the spectral basis functions and their interactions through inner products play important roles in the achievable estimation performance. In addition, we also showed that the existence of the baseline, which is usually considered a significant obstacle for metabolite quantitation at short TE, in fact is not problematic if the baseline can be accurately extracted and modeled using *a priori* chosen basis functions. This is due to the fact that the CRB depends on the baseline smoothness rather than its amplitude, and in practice the baseline is smooth and its smoothness does not change significantly with TE. We also observed that the use of multi-echo data improves the CRB, and thus should improve metabolite quantitation. Compared to the single-echo case, data acquisition at D number of TEs leads to a minimum theoretical factor of D improvement on the variance of the estimated metabolite amplitudes. This is equivalent to performing signal averaging D times. An additional improvement factor can be achieved if the TEs are chosen so that the spectral overlap (inner product) between spectral components is reduced compared to the single-echo case. All observations were supported from the CRB plots and the plots of the coefficient of variation, obtained from metabolite quantitation. A case study was also provided with parameters simulated based on the experimental data to demonstrate the use of the proposed CRB framework as a tool for choosing optimal TE. Unlike empirical studies, which face practical limitations on acquisition time, this CRB analysis enables easy evaluation of an arbitrarily large range of TEs.

Chapter 6

Conclusions

Magnetic resonance spectroscopic imaging (MRSI) extends conventional MRI by providing an additional dimension of spectral information. As a result, MRSI provides a powerful tool for *in vivo* metabolite studies and has important applications such as breast cancer imaging and neuroimaging. However, despite this potential, MRSI faces various issues limiting its practical utility.

This dissertation has addressed one of the issues, that is, the problem of low SNR. It has been recognized that exploiting signal properties inherent in MRSI data in a “soft” way is necessary to denoise MRSI data effectively in a severe noise contamination regime, which is typical for MRSI. In this dissertation, a new approach for denoising of MRSI data was introduced. The basic idea of the approach was to exploit the following signal properties of MRSI data: linear predictability, partial separability, and spatial correlation. A particular combination utilizing these properties led to the two novel denoising algorithms. The problem of low SNR in MRSI can also be alleviated prospectively (before acquiring data) by setting the echo time (TE) to be as small as possible. However, this approach aggravates another issue in MRSI - undesirable strong contribution from the baseline/nuisance signals. Thus, the choice of TE remains complicated and controversial. This issue has also been analyzed in the dissertation. Specifically, we proposed a novel, quantitative approach which considers the problem of TE selection from an estimation theoretic perspective.

The main contributions of this work include:

- The development and characterization of a novel denoising method that formulates the denoising problem using the penalized maximum-likelihood framework, in which spatial constraints are imposed using a weighted spatial-smoothness regularization and spectral assumptions are incorporated with an autoregressive constraint. The method has the capabil-

ity to preserve spatial-spectral features while suppressing noise in non-metabolite regions, which gives rise to a non-uniform denoising effect spectrally. This non-uniform filtering behavior was observed experimentally and explained theoretically. Quantitative study was provided to evaluate the usefulness of the denoised data.

- The development and characterization of a novel denoising method that exploits two low-rank properties inherent in MRSI data, one due to partial separability and the other due to linear predictability of MRSI data. The method has the capability to effectively reduce noise while preserving spatial-spectral features in the case of severe noise contamination. The combination of partial separability and linear predictability models provides a new principle for improving SNR for spatiotemporal imaging. Unlike the first proposed method, this approach does not have a non-uniform spectral filtering property.
- The quantitative analysis of the effect of different TE values using a Cramér-Rao bound (CRB) on the variance of metabolite amplitudes. The proposed approach provides a quantitative tool for choosing TE to guide the design of acquisition protocol instead of a general heuristic guess. We showed that (i) the form of the spectral basis functions and their interactions through spectral overlaps (inner products) play important roles in the achievable estimation performance; (ii) the existence of the baseline and the spectral overlap among spectral components, which are usually considered as significant obstacles for metabolite quantitation at short TE, in fact are not problematic if the baseline can be accurately represented using *a priori* chosen basis functions and the metabolite spectral profiles are known, respectively; (iii) the use of multi-echo data improves the CRB by a minimum improvement factor that is equivalent to performing signal averaging, while an additional improvement factor can be achieved if the spectral overlap between spectral components can be further reduced compared to the single-echo case. A metabolite quantitation study was provided to support the conclusions made from the CRB perspective. A case study was also considered to demonstrate the use of the proposed CRB framework.

Even though the results have shown the superior performance of the proposed denoising approach over conventional methods (Gaussian apodization and wavelet shrinkage) and the promising utility of the proposed approach to the TE selection for MRS acquisition protocols, further research is necessary to make the methods practically useful in clinics. One of the obstacles in practical applications of the proposed denoising methods is their computational intensity. This is particularly the case for the first denoising method. Specifically, in the first method the main computation time is devoted to computing the linear prediction coefficients voxel by voxel and then obtaining the denoised spatial-spectral solution based on these coefficients, while in the second proposed method the main computation time is spent on performing SVD of a large spatial-spectral data matrix (N spatial voxels $\times M$ time points) and then performing SVD of a smaller matrix ($M/2 \times M/2$) but voxel by voxel. Efforts have been made to speed up the computation time for both methods (in particular, the implementation of fast computation of the matrix multiplications through the fast Fourier transform and utilizing a parallel execution of the loop command in Matlab), reducing the computation time significantly. Currently, to denoise *in vivo* MRSI data with 32×32 spatial encodings and 1024 time points takes about 40 minutes using the first method but only 4.5 minutes using the second method on an Intel Xeon dual quad-core 3.16 Hz CPU, 13.8 GB RAM computer. Future use of highly computationally-efficient tools such as parallel processing should be useful in speeding up the computation of operations that are performed individually voxel by voxel, such as the estimation of linear prediction coefficients for the first method and the low-rank approximation of a Hankel matrix using SVD truncation for the second method. On the other hand, the CRB tool on TE selection for MRS acquisitions does not suffer from intensive computation time and is easy to use.

Although the proposed denoising methods were developed for denoising of MRSI data, it is possible that they can be applied to other spatial-spectral and spatial-temporal imaging data as well. This is more promising for the second proposed method (LORA), since the first denoising method has the main drawback in the non-uniform spectral filtering property, which is a direct result of imposing the autoregressive constraint as a regularization. Perhaps a different application that

requires spectral peak identification could exploit the full potential of the first approach in utilizing the autoregressive property of the spectral signal. An interesting application of the LORA method could be denoising of fluorine (^{19}F) MRSI data, which has a very large chemical shift range (over 300 ppm) with well-separated peaks, resulting in the extended ability of the method to work in a lower-SNR regime than the one considered in this dissertation. Further research may also be needed to improve the current trade-off between SNR and resolution of LORA by replacing a hard thresholding of singular values by a more advanced technique such as robust PCA. A clinical study on a large number of healthy and non-healthy subjects would be desirable to thoroughly determine the value of the denoised MRSI data in improving the accuracy of metabolite quantitation. As for the proposed tool for the TE selection using the CRB, additional clinical validations are needed and a further effort is necessary to make this technique available online on the scanners.

Appendix A

Proof of (4.19)

According to (4.18), we have

$$\begin{aligned}
 |\tilde{s}[n] - s_0[n]|^2 &= \left| \frac{1}{p(n)} \sum_{1 \leq i \leq I(n), 1 \leq j \leq J(n): i+j=n} \bar{s}_{i,j} - s_0[n] \right|^2 \\
 &= \left| \sum_{i,j: i+j=n} \frac{1}{p(n)} \left(\bar{s}_{i,j} - s_0[n] \right) \right|^2 \\
 &\leq \left(\frac{1}{p(n)} \right)^2 \sum_{i,j: i+j=n} |\bar{s}_{i,j} - s_0[n]|^2,
 \end{aligned} \tag{A.1}$$

where the last line follows from the triangle inequality. Note that

$$\|\tilde{\mathbf{H}} - \mathbf{H}_0\|_F^2 = \sum_n p(n) |\tilde{s}[n] - s_0[n]|^2 \tag{A.2}$$

and

$$\begin{aligned}
 \|\bar{\mathbf{H}} - \mathbf{H}_0\|_F^2 &= \sum_n \sum_{i,j: i+j=n} |\bar{s}_{i,j} - s_0[n]|^2 \\
 &\geq \sum_n p(n)^2 |\tilde{s}[n] - s_0[n]|^2
 \end{aligned} \tag{A.3}$$

$$\geq \sum_n p(n) |\tilde{s}[n] - s_0[n]|^2, \tag{A.4}$$

where (A.3) follows from (A.1), while (A.4) is true because $p(n) \geq 1$. Comparing (A.2) and (A.4) it follows that

$$\|\bar{\mathbf{H}} - \mathbf{H}_0\|_F \geq \|\tilde{\mathbf{H}} - \mathbf{H}_0\|_F \tag{A.5}$$

and the proof is finished.

Appendix B

Proof of (4.25)

The proof is given under assumptions (4.23) and (4.24) and $\hat{L}_{\text{ps}} \geq L_{\text{ps}}$. Due to SVD truncation, we have

$$\sigma_i(\bar{\mathbf{C}}) = \begin{cases} \sigma_i(\mathbf{C}), & i = 1, \dots, \hat{L}_{\text{ps}}, \\ 0, & i = \hat{L}_{\text{ps}} + 1, \dots, M. \end{cases} \quad (\text{B.1})$$

Using (4.23) and (B.1), we have

$$\sigma_i(\bar{\mathbf{C}}) = \begin{cases} \sigma_i^2(\mathbf{C}_0) + \sigma_i^2(\mathbf{E}), & \text{for } i = 1, \dots, \hat{L}_{\text{ps}}, \\ 0, & \text{for } i = \hat{L}_{\text{ps}} + 1, \dots, M. \end{cases} \quad (\text{B.2})$$

From (4.24) and (B.2) it follows that

$$\sigma_i^2(\Delta) = \begin{cases} \sigma_i^2(\mathbf{E}), & \text{for } i = 1, \dots, \hat{L}_{\text{ps}}, \\ 0, & \text{for } i = \hat{L}_{\text{ps}} + 1, \dots, M. \end{cases} \quad (\text{B.3})$$

The second line in (B.3) follows directly from that fact that $\mathbf{C}_0$ has rank $L_{\text{ps}} \leq \hat{L}_{\text{ps}}$.

From (B.3) it follows that

$$\|\Delta\|_F = \sqrt{\sum_{i=1}^{\hat{L}_{\text{ps}}} \sigma_i^2(\mathbf{E})} \quad (\text{B.4})$$

and hence, the proof is finished.

Appendix C

Derivations of the Cramér-Rao bound

C.1 CRB expressions for the single-echo case

For convenience, in this section we drop the TE dependence in the notations by denoting $\varphi_{n,TE}[m]$, $v_{n,TE}[m]$, Φ_{TE} , and Υ_{TE} as $\varphi_n[m]$, $v_n[m]$, Φ , and Υ , respectively. The signal model (5.1) can be expressed in matrix form as

$$\mathbf{s} = \mathbf{Z}e^{i\phi_0}\mathbf{a} + \Upsilon\mathbf{b} + \boldsymbol{\xi}, \quad (\text{C.1})$$

where $\mathbf{Z} = \Phi \odot \Psi$ with $\odot$ denoting the element-wise product. Thus, the log-likelihood function is given by

$$\ln L(\mathbf{s}) = \text{const} - \frac{1}{\sigma_0^2} \|\mathbf{s} - \mathbf{Z}e^{i\phi_0}\mathbf{a} - \Upsilon\mathbf{b}\|_2^2. \quad (\text{C.2})$$

Derivatives of the log-likelihood function with respect to the unknown parameters are

$$\begin{aligned} \frac{\partial \ln L}{\partial a_k} &= -\frac{2}{\sigma_0^2} e^{-i\phi_0} \sum_{m=0}^{M-1} \text{Re} \left\{ \xi[m] \varphi_k^*[m] \psi_{k,d_k}^*[m] \right\} \\ \frac{\partial \ln L}{\partial d_k} &= -\frac{2}{\sigma_0^2} e^{-i\phi_0} \sum_{m=0}^{M-1} \text{Re} \left\{ \xi[m] a_k \varphi_k^*[m] \frac{\partial \psi_{k,d_k}^*[m]}{\partial d_k} \right\} \\ \frac{\partial \ln L}{\partial \phi_0} &= -\frac{2}{\sigma_0^2} e^{-i\phi_0} \sum_{n=1}^N \sum_{m=0}^{M-1} \text{Im} \left\{ \xi[m] a_n \varphi_n^*[m] \psi_{n,d_n}^*[m] \right\} \\ \frac{\partial \ln L}{\partial |b_k|} &= -\frac{2}{\sigma_0^2} \sum_{m=0}^{M-1} \text{Re} \left\{ \xi[m] e^{-i\angle b_k} v_k^*[m] \right\} \\ \frac{\partial \ln L}{\partial \angle b_k} &= -\frac{2}{\sigma_0^2} \sum_{m=0}^{M-1} \text{Im} \left\{ \xi[m] b_k v_k^*[m] \right\}. \end{aligned} \quad (\text{C.3})$$

Thus, the elements of the Fisher information matrix are

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial a_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{\sum_{m=0}^{M-1} \varphi_p^*[m] \psi_{p,d_p}^*[m] \varphi_q[m] \psi_{q,d_q}[m]\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial \phi_0}\right)\left(\frac{\partial \ln L}{\partial \phi_0}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{a_p a_q \sum_{p=1}^N \sum_{q=1}^N \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} \varphi_p^*[k] \psi_{p,d_p}^*[k] \varphi_q[l] \psi_{q,d_q}[l]\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial d_p}\right)\left(\frac{\partial \ln L}{\partial d_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{a_p a_q \sum_{m=0}^{M-1} \varphi_p^*[m] \frac{\partial \psi_{p,d_p}^*[m]}{\partial d_p} \varphi_q[m] \frac{\partial \psi_{q,d_q}[m]}{\partial d_q}\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial |b|_p}\right)\left(\frac{\partial \ln L}{\partial |b|_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{e^{i(\angle b_q - \angle b_p)} \sum_{m=0}^{M-1} v_p^*[m] v_q[m]\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial \angle b_p}\right)\left(\frac{\partial \ln L}{\partial \angle b_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{b_p^* b_q \sum_{m=0}^{M-1} v_p^*[m] v_q[m]\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial \phi_0}\right)\right] &= -\frac{2}{\sigma_0^2} \text{Im}\left\{\sum_{n=1}^N a_n \sum_{m=0}^{M-1} \varphi_p^*[m] \psi_{p,d_p}^*[m] \varphi_n[m] \psi_{n,d_n}[m]\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial d_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{a_q \sum_{m=0}^{M-1} \varphi_p^*[m] \psi_{p,d_p}^*[m] \varphi_q[m] \frac{\partial \psi_{q,d_q}[m]}{\partial d_q}\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial |b|_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re}\left\{e^{i(\angle b_q - \phi_0)} \sum_{m=0}^{M-1} \varphi_p^*[m] \psi_{p,d_p}^*[m] v_q[m]\right\} \\
\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial \angle b_q}\right)\right] &= -\frac{2}{\sigma_0^2} \text{Im}\left\{b_q e^{-i\phi_0} \sum_{m=0}^{M-1} \varphi_p^*[m] \psi_{p,d_p}^*[m] v_q[m]\right\}. \tag{C.4}
\end{aligned}$$

By defining matrices

$$\begin{aligned}
\mathbf{A} &= \text{diag}\{\mathbf{a}\}, \\
\mathbf{B} &= \text{diag}\{\mathbf{b}\}, \\
\mathbf{G} &= \text{diag}\{e^{i\angle \mathbf{b}}\}, \\
\mathbf{D} &= \begin{bmatrix} \frac{\partial \mathbf{Z}_1}{\partial d_1} & \frac{\partial \mathbf{Z}_2}{\partial d_2} & \dots & \frac{\partial \mathbf{Z}_N}{\partial d_N} \end{bmatrix}, \tag{C.5}
\end{aligned}$$

where $\mathbf{Z}_i$ denotes the i -th column of $\mathbf{Z}$, we can express each block in $\mathbf{F}$ in a matrix form as

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{a}} \right) \left(\frac{\partial \ln L}{\partial \mathbf{a}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{Z}^H \mathbf{Z} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \phi_0} \right) \left(\frac{\partial \ln L}{\partial \phi_0} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{a}^H \mathbf{Z}^H \mathbf{Z} \mathbf{a} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{r}} \right) \left(\frac{\partial \ln L}{\partial \mathbf{r}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{A}^H \mathbf{D}^H \mathbf{D} \mathbf{A} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial |\mathbf{b}|} \right) \left(\frac{\partial \ln L}{\partial |\mathbf{b}|} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{G}^H \mathbf{\Upsilon}^H \mathbf{\Upsilon} \mathbf{G} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \angle \mathbf{b}} \right) \left(\frac{\partial \ln L}{\partial \angle \mathbf{b}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{B}^H \mathbf{\Upsilon}^H \mathbf{\Upsilon} \mathbf{B} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{a}} \right) \left(\frac{\partial \ln L}{\partial \phi_0} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \{ \mathbf{Z}^H \mathbf{Z} \mathbf{a} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{a}} \right) \left(\frac{\partial \ln L}{\partial \mathbf{r}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{Z}^H \mathbf{D} \mathbf{A} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{a}} \right) \left(\frac{\partial \ln L}{\partial |\mathbf{b}|} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{Z}^H \mathbf{\Upsilon} \mathbf{G} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{a}} \right) \left(\frac{\partial \ln L}{\partial \angle \mathbf{b}} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \{ \mathbf{Z}^H \mathbf{\Upsilon} \mathbf{B} \}.
\end{aligned} \tag{C.6}$$

The other blocks of $\mathbf{F}$ are the conjugate transpose of the corresponding blocks in (C.6).

C.2 Asymptotic CRB expressions for the single-echo case

Consider (C.4) and let the number of data points $M \rightarrow \infty$ and the sampling rate $\Delta t \rightarrow 0$. Then

$$\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial a_p} \right) \left(\frac{\partial \ln L}{\partial |b|_q} \right) \right] = \frac{2}{\sigma_0^2} \text{Re} \left\{ \int_0^\infty e^{i(\angle b_q - \phi_0)} \varphi_p^*(t) \psi_{p,d_p}^*(t) v_q(t) dt \right\}. \tag{C.7}$$

Substituting (5.3) into (C.7), we obtain

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial |b|_q}\right)\right] &= \frac{2}{\sigma_0^2} \text{Re} \left\{ e^{i(\angle b_q - \phi_0)} \sum_{l=1}^{L(p)} a_l(p) e^{i\phi_l(p)} \int_0^\infty \psi_{p,d_p}^*(t) v_q(t) e^{i2\pi f_l(p)t} dt \right\} \\ &= \frac{2}{\sigma_0^2} \text{Re} \left\{ e^{i(\angle b_q - \phi_0)} \sum_{l=1}^{L(p)} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \Upsilon_q(f) \right]_{f=f_l(p)} \right\},\end{aligned}\quad (\text{C.8})$$

where $\Psi_p(f)$ and $\Upsilon_q(f)$ are the Fourier transforms of $\psi_{p,d_p}(t)$ and $v_q(t)$, respectively. The last line in (C.8) follows from a property of the Fourier transform and taking into account that $\psi_{p,d_p}(t)$ is a real-valued function. Similarly, one can show the following asymptotical expressions:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial \ln L}{\partial a_p}\right)\left(\frac{\partial \ln L}{\partial \angle b_q}\right)\right] &= -\frac{2}{\sigma_0^2} \text{Im} \left\{ b_q e^{-i\phi_0} \sum_{l=1}^{L(p)} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \Upsilon_q(f) \right]_{f=f_l(p)} \right\} \\ \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial d_p}\right)\left(\frac{\partial \ln L}{\partial |b|_q}\right)\right] &= -\frac{4\pi}{\sigma_0^2} \text{Im} \left\{ \frac{a_p}{d_p^2} e^{i(\angle b_q - \phi_0)} \sum_{l=1}^{L(p)} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \dot{\Upsilon}_q(f) \right]_{f=f_l(p)} \right\} \\ \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial d_p}\right)\left(\frac{\partial \ln L}{\partial \angle b_q}\right)\right] &= \frac{4\pi}{\sigma_0^2} \text{Re} \left\{ \frac{a_p}{d_p^2} b_q e^{-i\phi_0} \sum_{l=1}^{L(p)} a_l(p) e^{i\phi_l(p)} \left[\Psi_p(f) \otimes \dot{\Upsilon}_q(f) \right]_{f=f_l(p)} \right\} \\ \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial |b|_p}\right)\left(\frac{\partial \ln L}{\partial \phi_0}\right)\right] &= -\frac{2}{\sigma_0^2} \text{Im} \left\{ e^{i(\angle b_q - \phi_0)} \sum_{n=1}^N a_n \sum_{l=1}^{L(n)} a_l(n) e^{i\phi_l(n)} \left[\Psi_n(f) \otimes \dot{\Upsilon}_p(f) \right]_{f=f_l(p)} \right\} \\ \mathbb{E}\left[\left(\frac{\partial \ln L}{\partial \angle b_p}\right)\left(\frac{\partial \ln L}{\partial \phi_0}\right)\right] &= -\frac{2}{\sigma_0^2} \text{Im} \left\{ b_p e^{-i\phi_0} \sum_{n=1}^N a_n \sum_{l=1}^{L(n)} a_l(n) e^{i\phi_l(n)} \left[\Psi_n(f) \otimes \dot{\Upsilon}_p(f) \right]_{f=f_l(p)} \right\},\end{aligned}\quad (\text{C.9})$$

where $\dot{\Upsilon}(f)$ denotes the derivative of $\Upsilon(f)$.

C.3 CRB expressions for the multi-echo case

Given the signal model (5.24), the log-likelihood function is expressed as

$$\ln L(\tilde{\mathbf{s}}) = \text{const} - \frac{1}{\sigma_0^2} \|\mathbf{s} - \tilde{\mathbf{Z}} e^{i\phi_0} \mathbf{c} - \tilde{\mathbf{\Upsilon}} \mathbf{b}\|_2^2, \quad (\text{C.10})$$

where $\tilde{\mathbf{Z}} = \tilde{\Phi} \odot \tilde{\Psi} \odot \tilde{\mathbf{Q}}$.

Derivatives of the log-likelihood function with respect to the unknown parameters are

$$\begin{aligned}
\frac{\partial \ln L}{\partial c_k} &= -\frac{2}{\sigma_0^2} e^{-i\phi_0} \sum_{m=0}^{DM-1} \text{Re} \left\{ \tilde{\xi}[m] \tilde{\varphi}_k^*[m] \tilde{q}_k^*[m] \tilde{\psi}_{k,d_k}^*[m] \right\} \\
\frac{\partial \ln L}{\partial d_k} &= -\frac{2}{\sigma_0^2} e^{-i\phi_0} \sum_{m=0}^{DM-1} \text{Re} \left\{ \tilde{\xi}[m] c_k \tilde{\varphi}_k^*[m] \tilde{q}_k^*[m] \frac{\partial \tilde{\psi}_{k,d_k}^*[m]}{\partial d_k} \right\} \\
\frac{\partial \ln L}{\partial \phi_0} &= -\frac{2}{\sigma_0^2} e^{-i\phi_0} \sum_{n=1}^N \sum_{m=0}^{DM-1} \text{Im} \left\{ \tilde{\xi}[m] c_n \tilde{\varphi}_n^*[m] \tilde{q}_n^*[m] \tilde{\psi}_{n,d_n}^*[m] \right\} \\
\frac{\partial \ln L}{\partial |b_k|} &= -\frac{2}{\sigma_0^2} \sum_{m=0}^{M-1} \text{Re} \left\{ \tilde{\xi}[m] e^{-i\angle b_k} \tilde{v}_k^*[m] \right\} \\
\frac{\partial \ln L}{\partial \angle b_k} &= -\frac{2}{\sigma_0^2} \sum_{m=0}^{M-1} \text{Im} \left\{ \tilde{\xi}[m] b_k \tilde{v}_k^*[m] \right\}.
\end{aligned} \tag{C.11}$$

Thus, the elements of the Fisher information matrix are

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{c}} \right) \left(\frac{\partial \ln L}{\partial \mathbf{c}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \tilde{\mathbf{Z}}^H \tilde{\mathbf{Z}} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \phi_0} \right) \left(\frac{\partial \ln L}{\partial \phi_0} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{c}^H \tilde{\mathbf{Z}}^H \tilde{\mathbf{Z}} \mathbf{c} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{r}} \right) \left(\frac{\partial \ln L}{\partial \mathbf{r}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{C}^H \tilde{\mathbf{D}}^H \tilde{\mathbf{D}} \mathbf{C} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial |\mathbf{b}|} \right) \left(\frac{\partial \ln L}{\partial |\mathbf{b}|} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{G}^H \tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}} \mathbf{G} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \angle \mathbf{b}} \right) \left(\frac{\partial \ln L}{\partial \angle \mathbf{b}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \mathbf{B}^H \tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}} \mathbf{B} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{c}} \right) \left(\frac{\partial \ln L}{\partial \phi_0} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \{ \tilde{\mathbf{Z}}^H \tilde{\mathbf{Z}} \mathbf{c} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{c}} \right) \left(\frac{\partial \ln L}{\partial \mathbf{r}} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \tilde{\mathbf{Z}}^H \tilde{\mathbf{D}} \mathbf{C} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{c}} \right) \left(\frac{\partial \ln L}{\partial |\mathbf{b}|} \right)^T \right] &= \frac{2}{\sigma_0^2} \text{Re} \{ \tilde{\mathbf{Z}}^H \tilde{\mathbf{Y}} \mathbf{G} \} \\
\mathbb{E} \left[\left(\frac{\partial \ln L}{\partial \mathbf{c}} \right) \left(\frac{\partial \ln L}{\partial \angle \mathbf{b}} \right)^T \right] &= -\frac{2}{\sigma_0^2} \text{Im} \{ \tilde{\mathbf{Z}}^H \tilde{\mathbf{Y}} \mathbf{B} \},
\end{aligned} \tag{C.12}$$

where $\tilde{\Phi}$, $\tilde{\Psi}$, $\tilde{\Upsilon}$ are defined as in (5.23), B , G are defined as in (C.5), and

$$\begin{aligned} \boldsymbol{C} &= \text{diag}\{\boldsymbol{c}\}, \\ \tilde{\boldsymbol{D}} &= \begin{bmatrix} \frac{\partial \tilde{\boldsymbol{Z}}_1}{\partial d_1} & \frac{\partial \tilde{\boldsymbol{Z}}_2}{\partial d_2} & \dots & \frac{\partial \tilde{\boldsymbol{Z}}_N}{\partial d_N} \end{bmatrix} \end{aligned} \tag{C.13}$$

with $\tilde{\boldsymbol{Z}}_i$ denoting the i -th column of $\tilde{\boldsymbol{Z}}$. The other blocks of $\boldsymbol{F}$ are the conjugate transpose of the corresponding blocks in (C.12).

References

- [1] H. C. Andrews and C. L. Patterson, “Singular value decompositions and digital image processing,” *IEEE Trans. Acoust. Speech Sign. Process.*, vol. 24, pp. 26–53, 1976.
- [2] R. A. de Graaf, *In vivo NMR spectroscopy: Principles and techniques*, 2nd ed. Chichester, UK: John Wiley & Sons, 2007.
- [3] A. I. Scott and R. L. Baxter, “Applications of ^{13}C NMR to metabolic studies,” *Annu. Rev. Biophys. Bioeng.*, vol. 10, pp. 151–174, 1981.
- [4] A. D. Costanzo et al., “Proton MR spectroscopy of the brain at 3 T: an update,” *Eur. Radiol.*, vol. 17, pp. 1651–1662, 2007.
- [5] A. W. Toga and J. C. Mazziotta, *Brain mapping: The methods*. San Diego, CA: Academic Press, 2002.
- [6] A. van den Boogaart, D. van Ormondt, W. W. F. Pijnappel, R. de Beer, and M. Ala-Korpela, “Removal of the water resonance from 1H magnetic resonance spectra,” in *Mathematics in Signal Processing III*, J. G. McWhirter, Ed. Oxford, UK: Clarendon Press, 1994, pp. 175–195.
- [7] P. B. Barker and D. D. M. Lin, “In vivo proton MR spectroscopy of the human brain,” *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 49, pp. 99–128, 2006.
- [8] H. M. Nguyen, J. P. Haldar, M. N. Do, and Z.-P. Liang, “Denoising of MR spectroscopic imaging data with spatial-spectral regularization,” in *Proc. IEEE Intl. Symposium Biomed. Imag.*, 2010, pp. 720–723.
- [9] H. M. Nguyen, X. Peng, M. N. Do, and Z.-P. Liang, “Spatiotemporal filtering of MR spectroscopic imaging data by low-rank approximations,” in *Proc. IEEE Intl. Symposium Biomed. Imag.*, 2011, pp. 857 – 860.
- [10] H. M. Nguyen, X. Peng, M. N. Do, and Z.-P. Liang, “Spatiotemporal filtering of MR spectroscopic imaging data by low-rank approximations,” in *Proc. Intl. Soc. Mag. Reson. Med.*, 2011, p. 1436.
- [11] H. M. Nguyen, X. Peng, M. N. Do, and Z.-P. Liang, “Spatiotemporal filtering of MR spectroscopic imaging data by low-rank approximations,” *IEEE Trans. Med. Imag.*, 2011, submitted on July 6.

- [12] S. Cavassila, S. Deval, C. Huegen, D. van Ormondt, and D. Graveron-Demilly, "Cramer-Rao bound expressions for parametric estimation of overlapping peaks: influence of prior knowledge," *J. Magn. Reson.*, vol. 143, pp. 311–320, 2000.
- [13] S. Cavassila, S. Deval, C. Huegen, D. van Ormondt, and D. Graveron-Demilly, "Cramer-Rao bounds: an evaluation tool for quantitation," *NMR Biomed.*, vol. 14, pp. 278–283, 2001.
- [14] S. W. Provencher, "Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra," *Magn. Reson. Med.*, vol. 30, pp. 672–679, 1993.
- [15] H. M. Nguyen, Z. J. Gahvari, J. P. Haldar, M. N. Do, and Z.-P. Liang, "Cramér-Rao bound analysis of echo time selection for ^1H -MR spectroscopy," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2009, pp. 2692–2695.
- [16] H. Friebolin, *Basic one- and two-dimensional NMR spectroscopy*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2005.
- [17] V. Govindaraju, K. Young, and A. A. Maudsley, "Proton NMR chemical shifts and coupling constants for brain metabolites," *NMR Biomed.*, vol. 13, pp. 129–153, 2000.
- [18] P. J. Hore, "Solvent suppression in Fourier transform nuclear magnetic resonance," *J. Magn. Reson.*, vol. 55, pp. 283–300, 1983.
- [19] G. A. Morris and R. Freeman, "Selective excitation in Fourier transform nuclear magnetic resonance," *J. Magn. Reson.*, vol. 29, pp. 433–462, 1978.
- [20] A. Haase, J. Frahm, W. Hanicke, and D. Matthaei, " ^1H NMR chemical shift selective (CHESS) imaging," *Phys. Med. Biol.*, vol. 30, pp. 341–344, 1985.
- [21] H. Barkhuysen, R. de Beer, and D. van Ormondt, "Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals," *J. Magn. Reson.*, vol. 73, pp. 553–557, 1987.
- [22] G. D. Graham, J.-H. Hwang, D. L. Rothman, and J. W. Prichard, "Spectroscopic assessment of alterations in macromolecule and small-molecule metabolites in human brain after stroke," *Stroke*, vol. 32, pp. 2797–2802, 2001.
- [23] U. Seeger and U. Klose, "Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases," *Magn. Reson. Med.*, vol. 49, pp. 19–28, 2003.
- [24] I. Mader, U. Seeger, J. Karitzky, M. Erb, F. Schick, and U. Klose, "Proton magnetic resonance spectroscopy with metabolite nulling reveals regional differences of macromolecules in normal human brain," *J. Magn. Reson. Imag.*, vol. 16, pp. 538–546, 2002.
- [25] H. Ratiney, M. Sidka, Y. Coenradie, S. Cavassila, D. V. Ormondt, and D. Graveron-Demilly, "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR Biomed.*, vol. 18, pp. 1–13, 2005.

- [26] B. J. Soher, K. Young, and A. A. Maudsley, "Representation of strong baseline contributions in ^1H MR spectra," *Magn. Reson. Med.*, vol. 45, pp. 966–972, 2001.
- [27] J.-B. Pouillet, D. M. Sima, A. W. Simonetti, B. de Neuter, L. Vanhamme, P. Lemmerling, and S. V. Huffel, "An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES," *NMR Biomed.*, vol. 20, pp. 493–504, 2007.
- [28] K. Young, B. J. Soher, and A. A. Maudsley, "Automated spectral analysis II: application of wavelet shrinkage for characterization of non-parameterized signals," *Magn. Reson. Med.*, vol. 40, pp. 816–821, 1998.
- [29] B. J. Soher, K. Young, and A. A. Maudsley, "Comparison of methods for baseline characterization of *in vivo* ^1H MR spectra," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2000, p. 1871.
- [30] S. Mierisová and M. Ala-Korpela, "MR spectroscopy quantitation: a review of frequency-domain methods," *NMR in Biomed.*, vol. 14, pp. 247–259, 2001.
- [31] L. Vanhamme, T. Sundin, P. V. Hecke, and S. V. Huffel, "MR spectroscopy quantitation: a review of time-domain methods," *NMR in Biomed.*, vol. 14, pp. 233–246, 2001.
- [32] J.-B. Pouillet, D. M. Sima, and S. V. Huffel, "MRS signal quantitation: a review of time- and frequency-domain methods," *J. Magn. Reson.*, vol. 195, pp. 134–144, 2008.
- [33] G. H. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Probl.*, vol. 19, pp. R1–R26, 2003.
- [34] L. Vanhamme, A. van den Boogaart, and S. V. Huffel, "Improved method for accurate and efficient quantification of MRS data with use of prior knowledge," *J. Magn. Reson.*, vol. 129, pp. 35–43, 1997.
- [35] J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Lecture Notes in Mathematics*, G. A. Watson, Ed. Berlin, Germany: Springer Verlag, 1978, pp. 104–116.
- [36] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimisation and nonlinear equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [37] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments an estimation theory perspective," *IEEE Trans. Sign. Process.*, vol. 52, pp. 1814–1829, 2004.
- [38] J. Makhoul, "Linear prediction: a tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, 1975.
- [39] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 30, pp. 833–840, 1982.
- [40] H. Barkhuysen, R. de Beer, W. M. M. J. Bovée, and D. van Ormondt, "Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure," *J. Magn. Reson.*, vol. 61, pp. 1465–481, 1985.

- [41] C. F. Tirendi and J. F. Martin, "Quantitative analysis of NMR spectra by linear prediction and total least squares," *J. Magn. Reson.*, vol. 85, pp. 162–169, 1989.
- [42] S. Y. Kung, K. S. Arun, and D. V. B. Rao, "State space and singular value decomposition based on approximation methods for harmonic retrieval," *J. Opt. Soc. Am.*, vol. 73, pp. 1799–1811, 1983.
- [43] W. W. F. Pijnappel, A. van den Boogaart, R. de Beer, and D. van Ormondt, "SVD-based quantification of magnetic resonance signals," *J. Magn. Reson.*, vol. 97, pp. 122–134, 1992.
- [44] S. V. Huffel, H. Chen, C. Decanniere, and P. V. Hecke, "Algorithm for time-domain NMR data fitting based on total least squares," *J. Magn. Reson. Ser. A*, vol. 110, pp. 228–237, 1994.
- [45] H. Chen, S. V. Huffel, and J. Vandewalle, "Improved methods for exponential parameter estimation in the presence of known poles and noise," *IEEE Trans. Sign. Process.*, vol. 45, pp. 1390–1393, 1997.
- [46] P. Stoica, Y. Selén, N. Sandgren, and S. V. Huffel, "Using prior knowledge in SVD-based parameter estimation for magnetic resonance spectroscopy - the ATP example," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1568–1578, 2004.
- [47] T. Laudadio, Y. Selén, L. Vanhamme, P. Stoica, P. V. Hecke, and S. V. Huffel, "Subspace-based MRS data quantitation of multiplets using prior knowledge," *J. Magn. Reson.*, vol. 168, pp. 53–65, 2004.
- [48] B. J. Soher, K. Young, V. Govindaraju, and A. A. Maudsley, "Automated spectral analysis III: Application to in vivo proton MR spectroscopy and spectroscopic imaging," *Magn. Reson. Med.*, vol. 40, pp. 822–831, 1998.
- [49] B. J. Soher, K. Young, A. Bernstein, Z. Aygula, and A. A. Maudsley, "GAVA: Spectral simulation for in vivo MRS applications," *J. Magn. Reson.*, vol. 185, pp. 291–299, 2007.
- [50] S. A. Smith, T. O. Levante, B. H. Meier, and R. R. Ernst, "Computer simulations in magnetic resonance. An object-oriented programming approach," *J. Magn. Reson. Ser. A*, vol. 106, pp. 75–105, 1994.
- [51] P. A. Bottomley, "Spatial localization in NMR spectroscopy in vivo," *Ann. N. Y. Acad. Sci.*, vol. 508, pp. 333–348, 1987.
- [52] P. A. Bottomley, "Selective volume method for performing localized NMR spectroscopy," US patent 4 480 228, 1984.
- [53] W. I. Jung, "Localized double spin echo proton spectroscopy. part I: Basic concepts," *Concepts Magn. Reson.*, vol. 8, pp. 1–15, 1996.
- [54] W. I. Jung, "Localized double spin echo proton spectroscopy. part II: Weakly coupled homonuclear spin systems," *Concepts Magn. Reson.*, vol. 8, pp. 105–118, 1996.

- [55] J. Frahm, K.-D. Merboldt, and W. Hänicke, “Localized proton spectroscopy using stimulated echoes,” *J. Magn. Reson.*, vol. 72, pp. 502–508, 1987.
- [56] J. Frahm, K.-D. Merboldt, W. Hänicke, and A. Haase, “Stimulated echo imaging,” *J. Magn. Reson.*, vol. 64, pp. 81–93, 1985.
- [57] J. Granot, “Selected volume excitation using stimulated echoes (VEST). Application to spatially localized spectroscopy and imaging,” *J. Magn. Reson.*, vol. 70, pp. 488–492, 1986.
- [58] R. Kimmich and D. Hoepfel, “Volume-selective multipulse spin-echo spectroscopy,” *J. Magn. Reson.*, vol. 72, pp. 379–384, 1987.
- [59] J. Frahm, H. Bruhn, M. L. Gyngell, K. D. Merboldt, W. Hänicke, and R. Sauter, “Localized high-resolution proton NMR spectroscopy using stimulated echoes: Initial applications to human brain *in vivo*,” *Magn. Reson. Med.*, vol. 9, pp. 79–93, 1989.
- [60] C. T. W. Moonen et al., “Comparison of single-shot localization methods (STEAM and PRESS) for *in vivo* proton NMR spectroscopy,” *NMR Biomed.*, vol. 2, pp. 201–208, 1989.
- [61] S. J. Nelson, “Multivoxel magnetic resonance spectroscopy of brain tumors,” *Mol. Cancer Ther.*, vol. 2, pp. 497–507, 2003.
- [62] I. Marshall, J. Wardlaw, C. Graham, L. Murray, and A. Blane, “Repeatability of long and short echo-time *in vivo* proton chemical-shift imaging,” *Neuroradiol.*, vol. 44, pp. 973–980, 2002.
- [63] T. R. Brown, B. M. Kincaid, and K. Ugurbil, “NMR chemical shift imaging in three dimensions,” *Proc. Natl. Acad. Sci.*, vol. 79, pp. 3523–3526, 1982.
- [64] P. A. Bottomley, W. A. Edelstein, T. H. Foster, and W. A. Adams, “*In vivo* solvent-suppressed localized hydrogen nuclear magnetic resonance spectroscopy: a window to metabolism?” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 82, pp. 2148–2152, 1985.
- [65] A. C. Sava, D. M. Sima, J.-P. Pouillet, A. J. Wright, A. Heerschap, and S. V. Huffel, “Exploiting spatial information to estimate metabolite levels in two-dimensional MRSI of heterogeneous brain lesions,” *NMR Biomed.*, 2010, preprint, available at <http://onlinelibrary.wiley.com/doi/10.1002/nbm.1628/pdf>.
- [66] B. M. Kelm, F. O. Kaster, A. Henning, M.-A. Weber, P. Bachert, P. Boesiger, F. A. Hamprecht, and B. H. Menze, “Using spatial prior knowledge in the spectral fitting of MRS images,” *NMR Biomed.*, 2011, preprint, available at <http://onlinelibrary.wiley.com/doi/10.1002/nbm.1704/pdf>.
- [67] J. Kornak, K. Young, N. Schuff, A. A. Maudsley, and M. W. Weiner, “Bayesian reconstruction of low resolution magnetic resonance imaging modalities,” in *Leeds Annual Statistical Research Workshop - Bioinformatics, Images, and Wavelets*, 2004, pp. 89–92.
- [68] Y. Bao and A. A. Maudsley, “Improved reconstruction for MR spectroscopic imaging,” *IEEE Trans. Med. Imag.*, vol. 26, pp. 686–695, 2007.

- [69] S. K. Plevritis and A. Macovski, "MRS imaging using anatomically based k-space sampling and extrapolation," *Magn. Reson. Med.*, vol. 34, pp. 686–693, 1995.
- [70] Z.-P. Liang and P. C. Lauterbur, "A generalized series approach to MR spectroscopic imaging," *IEEE Trans. Med. Imag.*, vol. 10, pp. 132–137, 1991.
- [71] X. Hu, D. N. Levin, P. C. Lauterbur, and T. Spraggins, "SLIM: Spectral localization by imaging," *Magn. Reson. Med.*, vol. 8, pp. 314 – 322, 1988.
- [72] J. P. Haldar, T.-H. Wu, Q. Wang, C.-I. Chen, S.-K. Song, and Z.-P. Liang, "Further development in anatomically constrained MR image reconstruction: Application to multimodal imaging of mouse stroke," in *Proc. Intl. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 422–425.
- [73] X. Peng, H. M. Nguyen, J. P. Haldar, D. Hernando, X.-P. Wang, and Z.-P. Liang, "Correction of field inhomogeneity effects in limited k-space coverage MRSI data with anatomical constraint," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2010, pp. 883–886.
- [74] A. Bashir and D. A. Yablonskiy, "Natural linewidth chemical shift imaging (NL-CSI)," *Magn. Reson. Med.*, vol. 56, pp. 7–18, 2006.
- [75] I. Khalidov, D. van de Ville, M. Jacob, F. Lazeyras, and M. Unser, "BSLIM: spectral localization by imaging with explicit B_0 field inhomogeneity compensation," *IEEE Trans. Med. Imag.*, vol. 26, pp. 990–1000, 2007.
- [76] R. Eslami and M. Jacob, "Correction of B_0 inhomogeneity distortion in magnetic resonance spectroscopic imaging," in *Proc. IEEE Intl. Conf. Acoust. Speech Sig. Process.*, 2009, pp. 405–408.
- [77] M. E. Halse and P. T. Callaghan, "Imaged deconvolution: a method for extracting high-resolution NMR spectra from inhomogeneous fields," *J. Magn. Reson.*, vol. 185, pp. 130–137, 2007.
- [78] Z. Dong and B. S. Peterson, "Spectral resolution amelioration by deconvolution (SPREAD) in MR spectroscopic imaging," *J. Magn. Reson. Imag.*, vol. 29, pp. 1395–1405, 2009.
- [79] A. Macovski, "Volumetric NMR imaging with time-varying gradients," *Magn. Reson. Imag.*, vol. 2, pp. 29–40, 1985.
- [80] R. Stoyanova, A. C. Kuesel, and T. R. Brown, "Application of principal-component analysis for NMR spectral quantitation," *J. Magn. Reson. Ser. A*, vol. 115, pp. 265–269, 1995.
- [81] T. R. Brown and R. Stoyanova, "NMR spectral quantitation by principal-component analysis. II. Determination of frequency and phase shifts," *J. Magn. Reson. Ser. B*, vol. 112, pp. 32–43, 1996.
- [82] R. Stoyanova and T. R. Brown, "NMR spectral quantitation by principal component analysis III. A generalized procedure for determination of lineshape variations," *J. Magn. Reson.*, vol. 154, pp. 163–175, 2002.

- [83] C. Ladrone, F. A. Howe, J. R. Griffiths, and A. R. Tate, "Independent component analysis for automated decomposition of in vivo magnetic resonance spectra," *Magn. Reson. Med.*, vol. 50, pp. 697–703, 2003.
- [84] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *IEEE Tran. Neural Netw.*, vol. 13, pp. 411–430, 2000.
- [85] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Adv. in Neural Info. Process.*, vol. 13, pp. 556–562, 2001.
- [86] R. Pohmann, M. von Kienlin, and A. Haase, "Theoretical evaluation and comparison of fast chemical shift imaging methods," *J. Magn. Reson.*, vol. 129, pp. 145–160, 1997.
- [87] A. Haase and D. Matthaei, "Spectroscopic FLASH NMR imaging (SPLASH) imaging," *J. Magn. Reson.*, vol. 71, pp. 550–553, 1987.
- [88] A. Haase, "Snapshot FLASH MRI. applications to T_1 , T_2 , and chemical-shift imaging," *Magn. Reson. Med.*, vol. 13, pp. 77–89, 1990.
- [89] M. Doyle and P. Mansfield, "Chemical-shift imaging: a hybrid approach," *Magn. Reson. Med.*, vol. 5, pp. 255–261, 1987.
- [90] P. Mansfield, "Spatial mapping of the chemical shift in NMR," *Magn. Reson. Med.*, vol. 1, pp. 370–386, 1984.
- [91] D. N. Guilfoyle, A. Blamire, B. Chapman, R. J. Ordidge, and P. Mansfield, "PEEP - a rapid chemical-shift imaging method," *Magn. Reson. Med.*, vol. 10, pp. 282–287, 1989.
- [92] P. M. Jakob, F. Kober, R. Pohmann, and A. Haase, "Single-shot spectroscopic imaging (SISSI) using a PEEP/BURST hybrid," *J. Magn. Reson. Ser. B*, vol. 110, pp. 278–283, 1996.
- [93] S. Posse, G. Tedeschi, R. Risinger, R. Ogg, and D. L. Bihan, "High speed ^1H spectroscopic imaging in human brain by echo planar spatial-spectral encoding," *Magn. Reson. Med.*, vol. 33, pp. 34–40, 1995.
- [94] H. F. Cancino-De-Greiff, R. Ramos-Garcia, and J. V. Lorenzo-Ginori, "Signal de-noising in magnetic resonance spectroscopy using wavelet transforms," *Concepts Magn. Reson.*, vol. 14, pp. 388–401, 2002.
- [95] H. C. Andrews and C. L. Patterson, "Singular value decompositions and digital image processing," *IEEE Trans. Acoust. Speech Sign. Process.*, vol. 24, pp. 26–53, 1976.
- [96] K. Konstantinides, B. Natarajan, and G. S. Yovanof, "Noise estimation and filtering using block-based singular value decomposition," *IEEE Trans. Image Process.*, vol. 6, pp. 479–483, 1997.
- [97] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, pp. 490–530, 2005.

- [98] G. Sapiro, *Geometric partial differential equations and image analysis*. Cambridge, UK: Cambridge University Press, 2001.
- [99] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, 1994, pp. 31–35.
- [100] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–265, 1992.
- [101] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Intl. Conf. Comp. Vis.*, 1998, pp. 839–846.
- [102] A. Buades, B. Coll, and J. M. Morel, "Neighborhood filters and PDEs," *Numer. Math.*, pp. 1–34, 2005.
- [103] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [104] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613–627, 1995.
- [105] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. Berlin, Germany: Springer, 1995, pp. 125–150.
- [106] S. Durand and M. Nikolova, "Restoration of wavelet coefficients by minimizing a specially designed objective function," in *IEEE Workshop on Var. and Level Set Meth. in Comp. Vision*, 2003, pp. 145–152.
- [107] O. A. Ahmed, "New denoising scheme for magnetic resonance spectroscopy signals," *IEEE Trans. Med. Imag.*, vol. 24, pp. 809–816, 2005.
- [108] J. A. Cadzow, "Signal enhancement - a composite property mapping algorithm," *IEEE Trans. Acoust. Speech Sign.*, vol. 36, pp. 49–62, 1988.
- [109] A. C. Sava, D. M. Sima, J.-P. Pouillet, and S. V. Huffel, "Exploiting spatial information for estimating metabolite concentration in MRSI," in *17th Proc. Intl. Soc. Mag. Reson. Med.*, 2009, p. 4280.
- [110] R. Eslami and M. Jacob, "Reduction of distortions in MRSI using a new signal model," in *Proc. IEEE Intl. Symposium Biomed. Imag.*, 2009, pp. 438–441.
- [111] A. Diop, Y. Zaim-Wadghiri, A. Briguët, and D. Graveron-Demilly, "Improvements of quantitation by using the Cadzow enhancement procedure prior to any linear-prediction methods," *J. Magn. Reson.*, vol. 105, pp. 17–24, 1994.
- [112] R. de Beer, "Quantitative *in vivo* NMR," 1994, lecture notes, University of Technology Delft, available at <http://www.intermnet.ua.es/interMNet/c59/index.html>.

- [113] J. P. Haldar, D. Hernando, S.-K. Song, and Z.-P. Liang, "Anatomically constrained reconstruction from noisy data," *Magn. Reson. Med.*, vol. 59, pp. 810–818, 2008.
- [114] S. V. Vaseghi, *Advanced Digital Signal Processing and noise reduction*, 2nd ed. New York, NY: John Wiley and Sons, Ltd, 2000.
- [115] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust. Speech Sign. Process.*, vol. 34, pp. 1081–1089, 1986.
- [116] D. L. Donoho, I. M. Johnstone, A. S. Stern, and J. C. Hoch, "Does the maximum entropy method improve sensitivity?" *Proc. Natl. Acad. Sci.*, vol. 87, pp. 5066–5068, 1990.
- [117] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Process.*, vol. 1, pp. 322–336, 1992.
- [118] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, pp. 561–580, 1992.
- [119] N. P. Galatsanos and A. K. Katsaggelos, "Cross validation and other criteria for estimating the regularization parameter and the noise variance," in *Proc. IEEE Intl. Conf. Acoust. Speech Sig. Process.*, 1991, pp. 3021–3024.
- [120] G. H. Golub, M. T. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometr.*, vol. 21, pp. 215–223, 1979.
- [121] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 326–339, 1991.
- [122] A. K. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "An iterative method for restoring noisy blurred images," *Circ. Sys. Sign. Process.*, vol. 3, pp. 139–160, 1984.
- [123] R. Srinivasan, N. Sailasuta, R. Hurd, S. Nelson, and D. Pelletier, "Evidence of elevated glutamate in multiple sclerosis using magnetic resonance spectroscopy at 3 T," *Brain*, vol. 128, pp. 1016–1025, 2005.
- [124] Y. Wang and S.-J. Li, "Differentiation of metabolic concentrations between gray matter and white matter of human brain by *in vivo* ^1H magnetic resonance spectroscopy," *Magn. Reson. Med.*, vol. 39, pp. 28–33, 1998.
- [125] F. Traber, W. Block, R. Lamerichs, J. Gieseke, and H. H. Schild, " ^1H metabolite relaxation times at 3.0 Tesla: measurements of T1 and T2 values in normal brain and determination of regional differences in transverse relaxation," *J. Magn. Reson. Imag.*, vol. 19, pp. 537–545, 2004.
- [126] C. Choi et al., "T₂ measurement and quantification of glutamate in human brain *in vivo*," *Magn. Reson. Med.*, vol. 56, pp. 971–977, 2006.

- [127] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE Intl. Symposium Biomed. Imag.*, 2007, pp. 988–991.
- [128] Z.-P. Liang and P. C. Lauterbur, "A theoretical analysis of the SLIM technique," *J. Magn. Reson. Ser. B*, vol. 102, pp. 54–60, 1993.
- [129] W. W. F. Pijnappel, A. van den Boogaart, R. de Beer, and D. van Ormondt, "SVD-based quantification of magnetic resonance signals," *J. Magn. Reson.*, vol. 97, pp. 122–134, 1992.
- [130] I. S. Iohvidov, *Hankel and Toeplitz matrices and forms*. Boston, MA: Birkhauser, 1982.
- [131] V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR - Sbornik*, vol. 1, pp. 457 – 483, 1967.
- [132] R. Shibata, "Selection of the order of an autoregressive model by Akaike's information criterion," *Biometrika*, vol. 63, pp. 117–126, 1976.
- [133] F. S. V. Bazan and P. L. Toint, "Conditioning of infinite Hankel matrices of finite rank," *Syst. Control Lett.*, vol. 41, pp. 347–359, 2000.
- [134] B. D. Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE Sign. Process.*, vol. 41, pp. 2826–2838, 1993.
- [135] R. Kreis, "Quantitative localized ^1H MR spectroscopy for clinical use," *J. Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 31, pp. 155–195, 1997.
- [136] D. J. O. McIntyre, R. A. Charlton, H. S. Markus, and F. A. Howe, "Long and short echo time proton magnetic resonance spectroscopic imaging of the healthy aging brain," *J. Magn. Reson. Imag.*, vol. 26, pp. 1596–1606, 2007.
- [137] M. Inglese, M. Spindler, J. S. Babb, P. Sunenshine, M. Law, and O. Gonen, "Field, coil, and echo-time influence on sensitivity and reproducibility of brain proton MR spectroscopy," *Am. J. Neuroradiol.*, vol. 27, pp. 684–688, 2006.
- [138] K. Zhong and T. Ernst, "Localized *in vivo* human ^1H MRS at very short echo times," *Magn. Reson. Med.*, vol. 52, pp. 898–901, 2004.
- [139] B. J. Soher et al., "Short TE *in vivo* ^1H MR spectroscopic imaging at 1.5 T: acquisition and automated spectral analysis," *Magn. Reson. Imag.*, vol. 18, pp. 1159–1165, 2000.
- [140] B. A. Holshouser, S. Ashwal, and D. B. Hinshaw, "Proton MR spectroscopy in children with acute brain injury: comparison of short and long echo time acquisitions," *J. Magn. Reson. Imag.*, vol. 11, pp. 9–19, 2000.
- [141] E. Hattingen, H. Lanfermann, J. Quick, K. Franz, F. E. Zanella, and U. Pilatus, " ^1H MR spectroscopic imaging with short and long echo time to discriminate glycine in glial tumors," *Magn. Reson. Mater Phy.*, vol. 22, pp. 33–41, 2009.

- [142] B. J. Soher and A. A. Maudsley, "Evaluation of variable line-shape models and prior information in automated ^1H spectroscopic imaging analysis," *Magn. Reson. Med.*, vol. 52, pp. 1246–1254, 2004.
- [143] K. Young, V. Govindaraju, B. J. Soher, and A. A. Maudsley, "Automated spectral analysis I: Formation of a priori information by spectral simulation," *Magn. Reson. Med.*, vol. 40, pp. 812–815, 1998.
- [144] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [145] P. B. Barker, D. O. Hearshen, and M. D. Boska, "Single-voxel proton MRS of the human brain at 1.5T and 3.0T," *Magn. Reson. Med.*, vol. 45, pp. 765–769, 2001.
- [146] S. P. Litke, "Maximum likelihood parameter estimation of a mixture of exponentials and known basis functions," M.S. thesis, University of Illinois at Urbana-Champaign, Department of Electrical and Computer Engineering, 1989.
- [147] Y. Bresler and S. P. Litke, "A parametric technique for superresolution image reconstruction," in *Proc. IEEE Intl. Conf. Acoust. Speech Sig. Process.*, vol. 3, 1990, pp. 1205–1208.