SPECTRAL ESTIMATION WITH SPATIO-SPECTRAL CONSTRAINTS FOR MAGNETIC RESONANCE SPECTROSCOPIC IMAGING

BY

QIANG NING

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Adviser:

Professor Zhi-Pei Liang

ABSTRACT

Magnetic resonance spectroscopic imaging (MRSI) is a promising tool to acquire in vivo biochemical information, and spectral estimation (quantification) of MRSI data is an important step towards quantitative studies. Although a large body of work has been done on spectral estimation over the past decades, it remains challenging due to model nonlinearity and extremely low signal-to-noise ratio (SNR). Building on the existing methods which effectively incorporate spectral prior knowledge in the form of basis functions, this work addresses the spectral estimation problem by incorporating both spectral and spatial prior information. Specifically, we jointly estimate the spectra over all the voxels of interest, incorporating prior spatial information in a regularization framework. The effectiveness of the proposed method has been evaluated using both simulated and experimental data. A theoretical analysis based on Cramér-Rao Bound is proposed to further assess the performance improvement of the proposed method over state-of-the-art methods. The proposed spectral estimation method should prove useful in various MRSI studies.

To my parents and brother, for their unconditional support.
To my wife, for her beautiful love.

ACKNOWLEDGMENTS

I would like to express my thanks to my adviser, Prof. Zhi-Pei Liang, without whose constant support this work would not be possible. It is his devotion to cutting-edge research that encourages me to move forward firmly on the road of science. His constructive criticism always motivates me to strengthen my capabilities which are necessary for an innovative researcher.

I also wish to thank Dr. Chao Ma, who provided much valuable advice in preparing this work. His penetrating insights helped me better understand research problems, and his patient guidance was indispensable for me to grow up as a researcher.

In addition, my sincere thanks also go to Dr. Fan Lam, Dr. Anthony Christodoulou and Dr. Bo Zhao, who shared their precious experience and insights with me. I would also like to express my gratitude to my other colleagues, Maojing Fu, Bryan Clifford, and Mohammed Sheikh, for the deadlines we worked for together, and for the generous help they provided to me in the past two years.

I would like to say many thanks to my parents and brother, for their selfless love to me and unwavering faith in me all the time. I also feel grateful to have my wife, who always supports me unconditionally, and who is the driving force behind me through hard times.

Finally, I wish to thank the Department of ECE for the second-to-none academic resources, and Dr. Yee and his family for their generous donation to the Yee memorial fellowship.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

MRSI is a unique, non-invasive tool to acquire in vivo physiological information without using molecular probes or radionuclide tracers. MRSI integrates the capability of conventional MRI and MR spectroscopy to obtain spatially resolved spectra, which have proved to be valuable biochemical information about the imaging subject. For example, MRSI can be used to study metabolism and to understand various physiological processes [1]; it can also be used for the detection, diagnosis, and treatment of diseases [2, 3]. One critical step for in vivo applications of MRSI is to extract quantitative information from the spatially resolved spectra through spectral estimation (also know as the quantification of MRSI). However, the inherent low sensitivity of MR techniques and the resulting low SNR make spectral estimation for MRSI a challenging problem.

A simple and still widely used spectral estimation method is to perform spectral integration, but its practical use is often limited to MRSI experiments with high SNR and the quantification of molecules without overlapping spectral components. For better quantification reliability, especially for low SNR MRSI data, model-based analysis is often preferred. Existing model-based methods usually exploit spectral prior knowledge by incorporating spectral bases into the forward model, and determine the model parameters by maximum likelihood estimation, which significantly improve spectral estimates over classical spectral integration methods [4, 5, 6, 7, 8, 9]. However, limited work has been done to exploit the spatial characteristics of metabolite distributions (e.g., smoothness or transform sparsity) during the estimation [10]. In particular, state-of-the-art methods (e.g., LCModel [6], and QUEST [8] method) perform spectral estimation of MRSI data voxel by

voxel and the estimated spectral parameters still have large variances. This work hence addresses the spectral estimation problem by incorporating both spectral and spatial prior information, which can significantly reduce the degree of freedom of the parameter space and achieve better tradeoff between bias and variance.

1.2 Main Results

This thesis proposes a novel method to address the spectral estimation problem for MRSI. The main contributions are summarized as following.

1. We reformulate the spectral estimation problem as a maximum likelihood estimation with regularization. The new formulation enables the incorporation of both spectral prior information and spatial prior information. Specifically, spectral prior information is imposed in the forward signal model in the form of spectral basis functions obtained from quantum simulation, and spatial prior information is absorbed into the regularization term. As a result, we quantify the spectra from all the voxels, instead of quantifying the spectra voxel by voxel independently as in existing methods.

2. Since all the voxels are quantified simultaneously, the resulting optimization problem from the proposed estimation formulation is computationally challenging. An efficient two-step algorithm to solve this large-scale nonlinear least-squares problem is provided, which makes use of the limited-memory BFGS (L-BFGS) and the alternating direction method of multipliers (ADMM).

3. A performance characterization based on Cramér-Rao bound (CRB) analysis for the proposed method is carried out, which shows the theoretical benefits of incorporating spatial constraints in terms of reducing estimation variance. Monte Carlo simulations are performed to validate the consistency between empirical performance and theoretical prediction.

4. The effectiveness of the proposed method is evaluated via both simulated and experimental data. Comparison with one of the state-of-

the-art methods, QUEST [8], is provided to demonstrate the improved performance provided by the proposed method.

1.3 Organization

The rest of the thesis is organized as follows. Chapter 2 firstly introduces some theoretical foundations for this work, including least-squares methods and estimation theory in the Bayesian framework, followed by a brief review of MRSI and existing spectral estimation methods. In chapter 3, we focus on describing the proposed method, from data preprocessing, to formulation that incorporates both spectral and spatial prior information, to the algorithm that solves the resulting optimization problem. Results (from both simulation data and experimental data) are shown in chapter 4 to evaluate the effectiveness of the proposed method, from which the significant improvement of spectral estimation will be presented. Finally, we discuss another possible method of imposing spatial prior information in chapter 5. We also carry out a theoretical performance analysis of the proposed method, which has been validated by Monte Carlo simulations.

CHAPTER 2

PRELIMINARIES

In this chapter, we firstly introduce some necessary concepts in least-squares (LS) methods and parameter estimation theory which are fundamental signal processing theories for analyzing MRSI data. Then a brief introduction of MRSI is provided, followed by a review of existing works related to the analysis of MRSI data.

2.1 Least-squares Methods

2.1.1 Linear Vector Spaces, and Hilbert Spaces

Linear least-squares methods are usually applied in general linear vector spaces, which are defined as following.

Definition 2.1 (Linear Vector Spaces (LVS)[11])**.** *A linear vector space is a set* $\mathcal{X}$ *defined over a field of scalars* $\mathbb{F}$ *with two operations:* $+: \ \mathcal{X} \times \mathcal{X} \mapsto \mathcal{X}$*, and* $\cdot: \ \mathbb{F} \times \mathcal{X} \mapsto \mathcal{X}$ *such that* $\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{X}$ *and* $\forall \alpha, \beta \in \mathbb{F}$

1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
2. $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$
3. $\exists \mathbf{0} \in \mathcal{X}$ *s.t.* $\mathbf{a} + \mathbf{0} = \mathbf{a}$
4. $\alpha \cdot (\beta \cdot \mathbf{a}) = (\alpha\beta) \cdot \mathbf{a}$
5. $\alpha \cdot (\mathbf{a} + \mathbf{b}) = \alpha \cdot \mathbf{a} + \alpha \cdot \mathbf{b}$
6. $(\alpha + \beta) \cdot \mathbf{a} = \alpha \cdot \mathbf{a} + \beta \cdot \mathbf{a}$
7. $0 \cdot \mathbf{a} = \mathbf{0}$*, and* $1 \cdot \mathbf{a} = \mathbf{a}$

The elements of $\mathcal{X}$ are called vectors. Note $\mathbb{C}^n$ is a special case of LVS, where all the vectors are tuples of complex numbers. Definition 2.1 is in fact a generalization of $\mathbb{C}^n$. This generalization allows us to represent much more real world relationships, which makes vector spaces an extremely useful tool [11].

Within an LVS $\mathcal{X}$, there are two important concepts: linear dependency, and basis.

Definition 2.2 (Linearly Dependent)**.** *A subset $\mathcal{S} \subset \mathcal{X}$ is linearly dependent if $\exists\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset S$ and $\exists \boldsymbol{\alpha} \in \mathbb{C}^n$, s.t. $\sum_{i=1}^n \alpha_i \mathbf{s}_i = \mathbf{0}$ and $\boldsymbol{\alpha} \neq \mathbf{0}$.*

Definition 2.3 (Basis of an LVS)**.** *A subset $\mathcal{S} \subset \mathcal{X}$ is said to be the basis of $\mathcal{X}$ if*

1. *The elements of $\mathcal{S}$ are linearly independent.*
2. *The Span$\{\mathcal{S}\} = \mathcal{X}$, where span is the set of all possible linear combinations of elements of $\mathcal{S}$.*

Note for any LVS $\mathcal{X}$, such a subset $\mathcal{S}$ always exists and has the same cardinality, which is called the dimension of $\mathcal{X}$ [11].

The convenience of using basis $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ for an LVS $\mathcal{X}$ is that $\forall \mathbf{x} \in \mathcal{X}$, there always exists a unique set of scalars, $\alpha_i \in \mathbb{F}$, $i = 1, \ldots, n$, such that $\sum_{i=1}^n \alpha_i \mathbf{s}_i = \mathbf{x}$. The scalar vector $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$ is called the coordinate of $\mathbf{x}$. Therefore, we can represent any LVS by a coordinate system, which makes it much easier to analyze general linear vector spaces.

Before we introduce the definition of Hilbert spaces, which are a special kind of LVS, we need to define vector norms and inner products:

Definition 2.4 (Vector Norms)**.** *A norm of a vector in $\mathcal{X}$ is such a mapping $\|\cdot\| : \mathcal{X} \mapsto \mathbb{R}$ that $\forall \mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\forall \alpha \in \mathbb{F}$*

1. *$\|\mathbf{x}\| \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{0}$.*
2. *$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$*
3. *$\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$*

An LVS associated with a definition of norm is called a normed vector space (NVS).

Definition 2.5 (Inner Products)**.** *An inner product is such a mapping* $< \cdot, \cdot >$: $\mathcal{X} \times \mathcal{X} \mapsto \mathbb{F}$ *that* $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}$ *and* $\forall \alpha \in \mathbb{F}$*,*

1. $< \mathbf{x} + \mathbf{y}, \mathbf{z} >=< \mathbf{x}, \mathbf{z} > + < \mathbf{y}, \mathbf{z} >$
2. $< \alpha\mathbf{x}, \mathbf{y} >= \alpha < \mathbf{x}, \mathbf{y} >$
3. $< \mathbf{x}, \mathbf{y} >= (< \mathbf{y}, \mathbf{x} >)^*$*, where "*$*$*" denotes complex conjugate*
4. $< \mathbf{x}, \mathbf{x} > \geq 0$ *with equality if and only if* $\mathbf{x} = \mathbf{0}$*.*

An LVS associated with a definition of inner products is called an inner product space (IPS). A fact is that $\sqrt{< \mathbf{x}, \mathbf{x} >}$ follows the definition of a norm. Therefore, it is usually called the induced norm.

Definition 2.6 (Hilbert Spaces)**.** *An LVS* $\mathcal{X}$ *is called a Hilbert space if it is associated with an inner product, and it is complete, i.e., any Cauchy sequence converges to a vector* $\mathbf{x} \in \mathcal{X}$*, in the sense of the induced norm.*

2.1.2 Linear Inverse Problems, and Orthogonality Principle

Linear inverse problems (i.e., solving $\mathbf{x}$ from $\mathbf{b}$ by the linear relationship $\mathcal{A}(\mathbf{x}) = \mathbf{b}$) are frequently encountered in reality, the solutions to which often do not exist due to the noise disturbance in $\mathbf{b}$ (and sometimes also in $\mathcal{A}$). Therefore, people sometimes solve linear inverse problems by solving the following least-squares problem instead.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|. \tag{2.1}$$

This problem has been readily solved in the conventional LVS (i.e., $\mathbb{R}^n$ or $\mathbb{C}^n$), but in a generalized LVS, the problem is generally difficult. Fortunately, if $\mathcal{A}$ is a bounded linear operator between two Hilbert spaces, and the range space of $\mathcal{A}$ is complete, we can solve Eq. (2.1) using the so-called orthogonality principle [11]. Before entering it, we need to bring up some basic concepts in Hilbert spaces, such as bounded linear operators, range spaces, adjoints, and the orthogonality principle.

Definition 2.7 (Bounded Linear Operators)**.** *A linear operator* $\mathcal{A}$ *is a mapping defined over two linear vector spaces and a scalar field* $\mathbb{F}$*,* $\mathcal{A}: \mathcal{X} \mapsto \mathcal{Y}$*, such that*

1. *If* $\mathcal{A}(\mathbf{x}) = \mathbf{y}$*, then* $\forall \alpha \in \mathbb{F}$*,* $\mathcal{A}(\alpha\mathbf{x}) = \alpha\mathbf{y}$*.*

2. *If* $\mathcal{A}(\mathbf{x}_1) = \mathbf{y}_1$ *and* $\mathcal{A}(\mathbf{x}_2) = \mathbf{y}_2$*, then* $\mathcal{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{y}_1 + \mathbf{y}_2$*.*

A linear operator is called bounded if both $\mathcal{X}$ *and* $\mathcal{Y}$ *are associated with norms (say,* $\|\cdot\|_{\mathcal{X}}$ *and* $\|\cdot\|_{\mathcal{Y}}$*), and* $\exists \alpha \in \mathbb{R}$ *s.t.* $\|\mathcal{A}(\mathbf{x})\|_{\mathcal{Y}} \leq \alpha\|\mathbf{x}\|_{\mathcal{X}}$ *holds true for all* $\mathbf{x} \in \mathcal{X}$*.*

Definition 2.8 (Adjoints of linear operators)**.** *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be two LVS each associated with an inner product* $<\cdot,\cdot>_{\mathcal{X}}$ *and* $<\cdot,\cdot>_{\mathcal{Y}}$*. The adjoints of a linear operator* $\mathcal{A}: \mathcal{X} \mapsto \mathcal{Y}$ *is a linear operator* $\mathcal{A}^*$ *such that* $\forall \mathbf{x} \in \mathcal{X}$ *and* $\forall \mathbf{y} \in \mathcal{Y}$*,* $<\mathcal{A}(\mathbf{x}), \mathbf{y}>_{\mathcal{Y}} = <\mathbf{x}, \mathcal{A}^*(\mathbf{y})>_{\mathcal{X}}$*.*

Definition 2.9 (Range Spaces and Null Spaces of Linear Operators)**.** *The range space of a linear operator* $\mathcal{A}: \mathcal{X} \mapsto \mathcal{Y}$ *is* $\mathcal{R}(\mathcal{A}) = \{\mathbf{y} \in \mathcal{Y} : \exists \mathbf{x} \in \mathcal{X}, s.t.\ \mathcal{A}(\mathbf{x}) = \mathbf{y}\}$*. The null space is* $\mathcal{N}(\mathcal{A}) = \{\mathbf{x} \in \mathcal{X} : \mathcal{A}(\mathbf{x}) = \mathbf{0}\}$*.*

Theorem 2.1 (Orthogonality Principle [11])**.** *Let* $\mathcal{S}$ *be a subspace of an IPS* $\mathcal{X}$*. Then* $\forall \mathbf{x} \in \mathcal{X}$*, finding* $\hat{\mathbf{x}} \in \mathcal{S}$ *such that*

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{x}\|, \ \forall \mathbf{y} \in \mathcal{S}$$

is equivalent to finding $\hat{\mathbf{x}} \in \mathcal{S}$ *such that*

$$\hat{\mathbf{x}} - \mathbf{x} \in \mathcal{S}^{\perp},$$

where $\mathcal{S}^{\perp} \equiv \{\mathbf{x} \in \mathcal{X} :< \mathbf{x}, \mathbf{s} >= 0, \forall \mathbf{s} \in \mathcal{S}\}$*. Furthermore, if* $\mathcal{S}$ *itself is a Hilbert space, then such a* $\hat{\mathbf{x}} \in \mathcal{S}$ *can always be found and is unique.*

Therefore, according to the orthogonality principle, we can handle the problem in Eq. (2.1) by seeking for a vector $\hat{\mathbf{x}} \in \mathcal{X}$ such that $\mathcal{A}(\hat{\mathbf{x}}) - \mathbf{b} \in \mathcal{R}(\mathcal{A})^{\perp} = \mathcal{N}(\mathcal{A}^*)$ [11], i.e., $\mathcal{A}^*(\mathcal{A}(\hat{\mathbf{x}})) = \mathcal{A}^*\mathbf{b}$. If $\mathcal{A}^*\mathcal{A}$ is a one-to-one mapping, then

$$\hat{\mathbf{x}} = (\mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^*\mathbf{b} \tag{2.2}$$

solves problem in Eq. (2.1). More generally,

$$\hat{\mathbf{x}} = \mathcal{A}^{\dagger}\mathbf{b},$$

where "†" denotes the Moore-Penrose pseudoinverse.

The evaluation of Eq. (2.2) usually requires the calculation of a matrix inverse, i.e., the inverse of A^*A, where A is the matrix representation of $\mathcal{A}$. In practice, we usually solve the linear equation

$$A^*A\mathbf{x} = A^*\mathbf{b},$$

instead of directly calculating the matrix inverse, which might be expensive. However, when the matrix size is very large, an exact solution may still be too difficult to resolve due to the ill-conditioning of the problem. A numerical approximation is desired in this case. Specifically, $\mathbf{x}$ can be obtained by solving the following convex optimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^*A^*A\mathbf{x} - b^*A\mathbf{x}.$$

Then various convex optimization algorithms can be applied, e.g., gradient descent, conjugate gradient (CG) descent, and Newton's method.

2.1.3 Nonlinear Inverse Problems, and the Variable Projection Method

General inverse problems are to solve $\mathbf{x}$ when measurement $\mathbf{b}$ is obtained through a general function $f(\cdot)$ (i.e., is not necessary to be linear), and can be expressed as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|f(\mathbf{x}) - \mathbf{b}\|. \tag{2.3}$$

It is generally difficult to solve such a problem, but if function $f(\cdot)$ is a mixture of linear functions and nonlinear functions, the variable projection method can then be applied to improve the estimation performance. Specifically, separate $\mathbf{x}$ into two parts, $\mathbf{x}_1$ and $\mathbf{x}_2$, such that the function $f(\cdot)$ is nonlinear in $\mathbf{x}_1$ and linear in $\mathbf{x}_2$, i.e.,

$$f(\mathbf{x}) = \mathcal{A}_{\mathbf{x}_1}(\mathbf{x}_2),$$

where $\mathcal{A}_{\mathbf{x}_1}(\cdot)$ is a linear operator dependent on parameter $\mathbf{x}_1$. This kind of problems in which the variables separate into two categories, linear and nonlinear, is ubiquitous in practice. In the scope of spectral estimation of MRSI data, it is a very important technique to reduce the degree of freedom

of variables and reduce the variances of estimates.

The motivation behind the variable projection method is to simplify the nonlinear optimization problem by ruling out the linear parts. If $\mathbf{x}_1$ is perfectly known, then by the solution to linear LS problems (Eq. (2.2)), we can project the measurement $\mathbf{b}$ onto the range space of $\mathcal{A}_{\mathbf{x}_1}$ and get

$$\hat{\mathbf{x}}_2 = \mathcal{A}_{\mathbf{x}_1}^{\dagger}\mathbf{b}.$$

If we can substitute the expression above into Eq. (2.3), then we get the following problem instead.

$$\hat{\mathbf{x}}_1 = \arg\min_{\mathbf{x}_1} \|\mathcal{A}_{\mathbf{x}_1}(\mathcal{A}_{\mathbf{x}_1}^{\dagger}\mathbf{b}) - \mathbf{b}\|, \tag{2.4}$$

$$\hat{\mathbf{x}}_2 = \mathcal{A}_{\hat{\mathbf{x}}_1}^{\dagger}\mathbf{b}. \tag{2.5}$$

In Eq. (2.4), although the problem is still nonlinear, there are fewer parameters to solve. In Eq. (2.5), it is simply a projection operation. Therefore, the original nonlinear optimization problem is turned into two relatively simpler subproblems by the variable projection method. Fortunately, the validity of such a substitution is supported by theorems proved by Golub and Pereyra in [12].

Theorem 2.2 ("Equivalence" of the Two Functionals [12])**.** *Assume that the range space of $\mathcal{A}_{\mathbf{x}_1}(\cdot)$ has a constant dimension for all $\mathbf{x}_1$ in an open set Ω.*

1. *If $\hat{\mathbf{x}}_1$ is a critical point (or a global minimizer in Ω) of Eq. (2.4), and*

$$\hat{\mathbf{x}}_2 = \mathcal{A}_{\hat{\mathbf{x}}_1}^{\dagger}\mathbf{b}, \tag{2.6}$$

   then $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ is a critical point of Eq. (2.3) (or a global minimizer for $\mathbf{x}_1 \in \Omega$).

2. *If $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ is a global minimizer of Eq. (2.3) for $\mathbf{x}_1 \in \Omega$, then $\hat{\mathbf{x}}_1$ is a global minimizer of Eq. (2.4) in Ω. Furthermore, if there is a unique $\hat{\mathbf{x}}_2$ among the minimizing pairs of Eq. (2.3), then $\hat{\mathbf{x}}_2$ must satisfy (2.6).*

Theorem 2.2 provides a theoretical guarantee that

1. Solving a local minimum for Eq. (2.4) and (2.5) also yields a local minimum for Eq. (2.3),

2. The global minimum(s) for Eq. (2.4) is consistent with the global minimum(s) for Eq. (2.3),

which makes the variable projection method extremely useful in practice.

2.2 Basics of Parameter Estimation

The spectral estimation of MRSI data is a specific topic for which we estimate the spectral parameters of the MRSI signal model. Therefore, some parameter estimation theories are necessary. In this section, we will provide some basics of statistical parameter estimation, including the biases and variances of estimators, performance evaluation of estimators, and several common estimation methods.

2.2.1 Bias, Variance and Efficiency of Estimators

Let $\mathbf{x} \in \mathcal{X}$ (e.g., $\mathbb{R}^n$, or $\mathbb{C}^n$) be the parameter vector we need to estimate, $\mathbf{Y}$ be the measurement, which is a random process, and $\hat{\mathbf{x}}(\mathbf{Y})$ be the designed estimator to estimate $\mathbf{x}$ from $\mathbf{Y}$, which is also a random process because of its dependence on $\mathbf{Y}$. The performance of $\hat{\mathbf{x}}(\mathbf{Y})$ is often evaluated by its bias and variance.

Definition 2.10 (Bias)**.** *The bias of an estimator* $\hat{\mathbf{x}}$ *is defined as*

$$b(\hat{\mathbf{x}}) = \mathbf{x} - \mathbb{E}\left[\hat{\mathbf{x}}(\mathbf{Y})\right].$$

If $b(\hat{\mathbf{x}}) = \mathbf{0}$*, then we call* $\hat{\mathbf{x}}$ *an unbiased estimator.*

Definition 2.11 (Variance)**.** *Assume* $\mathbf{x}$ *is an* n *dimensional vector, then the variance of* $\hat{\mathbf{x}}_i$*,* $i = 1, 2, \ldots, n$*, is defined as*

$$Var(\hat{\mathbf{x}}_i) = [Cov(\hat{\mathbf{x}}(\mathbf{Y}))]_{ii},$$

where

$$Cov(\hat{\mathbf{x}}(\mathbf{Y})) = \mathbb{E}\left[(\hat{\mathbf{x}}(\mathbf{Y}) - \mathbb{E}\left[\hat{\mathbf{x}}(\mathbf{Y})\right])(\hat{\mathbf{x}}(\mathbf{Y}) - \mathbb{E}\left[\hat{\mathbf{x}}(\mathbf{Y})\right])^T\right],$$

is the conventional covariance matrix of $\hat{\mathbf{x}}$*.*

The statistical dependence of $\mathbf{Y}$ on $\mathbf{x}$ can be described by a cumulative distribution function $F(\mathbf{y};\mathbf{x}) = \mathrm{P}\{\mathbf{Y} \leq \mathbf{y}\}$ (or, if $\mathbf{Y}$ is a continuous random variable, $p_{\mathbf{x}}(\mathbf{y})$). This underlying dependence between the measurement and the parameters to estimate determines how "difficult" the estimation problem will be, i.e., the variance of any estimator is bounded from below by the so-called Cramér-Rao bound (CRB) [13].

Theorem 2.3 (Cramér-Rao Bound)**.** *Given parameter* $\mathbf{x}$*, measurement* $\mathbf{Y}$*, and statistical dependence described by the probability distribution function* $p_{\mathbf{x}}(\mathbf{y})$*. For any unbiased estimator* $\hat{\mathbf{x}}(\mathbf{Y})$*, the following inequality holds.*

$$Var(\hat{\mathbf{x}}(\mathbf{Y})) \geq J^{-1}(\mathbf{x}),$$

where

$$Var(\hat{\mathbf{x}}(\mathbf{Y})) = E[(\mathbf{x} - \hat{\mathbf{x}}(\mathbf{Y}))(\mathbf{x} - \hat{\mathbf{x}}(\mathbf{Y}))^T] \tag{2.7}$$

is the error covariance matrix of $\hat{\mathbf{x}}$*, and*

$$J(\mathbf{x}) = \mathbb{E}\left[\bigtriangledown_{\mathbf{x}} ln f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{x})(\bigtriangledown_{\mathbf{x}} ln f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{x}))^T\right] \tag{2.8}$$

is the so-called Fisher Information in $\mathbf{Y}$ *about the parameters in* $\mathbf{x}$*.*

The Cramér-Rao bound is very useful in estimation theory, since it provides an approach to evaluating the performance of an unbiased estimator.

Definition 2.12 (Efficient Estimators)**.** *An unbiased estimator is efficient if it reaches its CRB, which means* $C_E(\hat{\mathbf{x}}) - J^{-1}(\mathbf{x}) = 0$*.*

Note an efficient estimator does not necessarily exist.

If we know beforehand that the parameter $\mathbf{x}$ lies in a lower dimensional subspace of the whole parameter space, and incorporate this prior knowledge into the estimation process, the estimator might actually achieve lower variance than the Cramér-Rao bound indicates. However, the conventional CRB does not take the constraints into account. As a result, the corresponding CRB is too "pessimistic", and the achievable estimator variance can even be lower (better) than the CRB. Given this situation, we can directly derive the so-called constrained CRB [14] using the Chapman-Robbins bound [15].

Theorem 2.4 (Chapman-Robbins Bound)**.** *Given a parameter* $\mathbf{x} \in \mathbb{R}^k$*, assume the measurement data* $\mathbf{Y} \in \mathbb{R}^n$ *have a probability distribution* $f_{\mathbf{x}}(\mathbf{y})$*.*

Then for any estimator $\hat{\mathbf{X}}(\mathbf{Y})$*, the following inequality holds true.*

$$\mathbf{\Sigma}(\hat{\mathbf{X}}) \geq [\delta \mathbf{m}_{\mathbf{x}}]^T \left(E\left[\left[\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}} \right] \left[\frac{\delta f_{\mathbf{x}}}{f_{\mathbf{x}}} \right]^T \right] \right)^{\dagger} [\delta \mathbf{m}_{\mathbf{x}}],$$

where $\mathbf{\Sigma}(\hat{\mathbf{X}})$ *is the covariance matrix of* $\hat{\mathbf{X}}(\mathbf{Y})$*,* $\mathbf{m}_{\mathbf{x}} \equiv E[\hat{\mathbf{X}}]$*, "*$\dagger$*" denotes the Moore-Penrose pseudoinverse,*

$$\delta \mathbf{m}_{\mathbf{x}} \equiv \left[\frac{\mathbf{m}_{\mathbf{x}+\mathbf{v_1}} - \mathbf{m}_{\mathbf{x}}}{\|\mathbf{v}_1\|}, \ldots, \frac{\mathbf{m}_{\mathbf{x}+\mathbf{v_p}} - \mathbf{m}_{\mathbf{x}}}{\|\mathbf{v}_p\|} \right]^T,$$

and

$$\delta f_{\mathbf{x}}(\mathbf{y}) \equiv \left[\frac{f_{\mathbf{x}+\mathbf{v}_1}(\mathbf{y}) - f_{\mathbf{x}}(\mathbf{y})}{\|\mathbf{v}_1\|}, \ldots, \frac{f_{\mathbf{x}+\mathbf{v_p}}(\mathbf{y}) - f_{\mathbf{x}}(\mathbf{y})}{\|\mathbf{v}_p\|} \right]^T,$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^k$ *are* p *arbitrary vectors. Again this bound can be improved by achieving the supremum of the right-hand side over all possible* $\mathbf{v}_1, \ldots, \mathbf{v}_p$*.*

It can be seen that the Chapman-Robbins bound (by taking the supremum over all possible deviations) is at least as sharp as the CRB (by taking the derivatives) [15], but it is also more difficult to compute in general. But it is a very important theoretical result from which we can derive the following constrained Cramér-Rao bound.

Theorem 2.5 (Constrained Cramér-Rao Bound)**.** *Let the parameter* $\mathbf{x}$ *to be estimated lie in a constrained space* $\mathcal{X} \subset \mathbb{R}^k$*. Let* $\mathbf{v}_1, \ldots, \mathbf{v}_p$ *be* p *linear independent vectors with sufficiently small lengths such that* $\mathbf{x} + \mathbf{v}_i \in \mathcal{X}, i = 1, \ldots, p$*. Assume the measurement data* $\mathbf{Y} \in \mathbb{R}^n$ *have a probability distribution* $f_{\mathbf{x}}(\mathbf{y})$*. Then for any estimator* $\hat{\mathbf{X}}(\mathbf{Y})$ *with mean* $\mathbf{m}_{\mathbf{x}}$*, the following inequality holds true:*

$$\mathbf{\Sigma}(\hat{\mathbf{X}}) \geq \limsup_{\substack{\mathbf{x}+\mathbf{v}_i \in \mathcal{X}, \|\mathbf{v}_i\| \to 0 \\ i=1,\ldots,p}} B_c,$$

where B_c *is the defined as the Chapman-Robbins bound over* $\mathbf{v}_1, \ldots, \mathbf{v}_p$*. If some additional conditions hold (see [14]), then*

$$\mathbf{\Sigma}(\hat{\mathbf{X}}) \geq [\nabla_x \mathbf{m}_{\mathbf{x}}]^T \mathbf{A} [\mathbf{A}^T \mathbf{J} \mathbf{A}]^{\dagger} \mathbf{A}^T [\nabla_x \mathbf{m}_{\mathbf{x}}],$$

where $\mathbf{A}$ *is any matrix whose column space is the same as span*$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$*, and* $\mathbf{J}$ *is the Fisher information matrix.*

Based on the constrained CRB, one can characterize the performance of an estimator given sparsity constraint [16] or low-rank constraint [17], etc.

2.2.2 Maximum Likelihood Estimation

A maximum likelihood (ML) estimator is to find an estimator which maximizes the likelihood function $p_{\mathbf{x}}(\mathbf{y})$:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p_{\mathbf{x}}(\mathbf{y}),$$

or

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \ln p_{\mathbf{x}}(\mathbf{y}), \tag{2.9}$$

which conceptually is to give the "most possible" estimate of $\mathbf{x}$ given $\mathbf{y}$. The maximum likelihood estimator is commonly used in practice when no prior distribution of $\mathbf{x}$ can be assumed. For example, the linear inverse problems, which we introduced in section 2.1, are usually formulated as least squares problems. The reason behind this formulation can be explained by the maximum likelihood estimation: if we assume $\mathcal{A}(\mathbf{x}) + \mathbf{N} = \mathbf{Y} \in \mathbb{R}^m$, where $\mathbf{N} \sim \mathcal{N}(0, \sigma^2 I_m)$, then

$$p_{\mathbf{x}}(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^m}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathcal{A}(\mathbf{x}))^T(\mathbf{y} - \mathcal{A}(\mathbf{x}))\right),$$

or

$$\ln p_{\mathbf{x}}(\mathbf{y}) = -\frac{1}{2}\|\mathcal{A}(\mathbf{x}) - \mathbf{y}\| - \frac{m}{2}\ln 2\pi.$$

Since the term $-\frac{m}{2}\ln 2\pi$ is a constant and does not change for different $\mathbf{x}$, the ML estimate for $\mathbf{x}$ can be reformulated as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathcal{A}(\mathbf{x}) - \mathbf{y}\|, \tag{2.10}$$

which is identical to Eq. (2.3). Therefore, the least squares solution to $\mathbf{x}$ is actually a maximum likelihood estimator given Gaussian noise perturbation.

A fact is that all ML estimators have these two properties [13]:

1. Asymptotically unbiased, i.e., $\lim_{N\to\infty} b(\hat{\mathbf{x}}) = 0$;

2. Asymptotically efficient, i.e., $\lim_{N\to\infty} \mathbb{E}\left[(\mathbf{x}_i - \hat{\mathbf{x}}_i(Y))^2\right] = [J^{-1}]_{ii}(\mathbf{x})$;

where N is the total number of observations. These two asymptotical properties provide us with a theoretical guarantee for the performance of ML estimators.

2.2.3 Bayesian Estimation

In the maximum likelihood estimation, parameter $\mathbf{x}$ is treated as an unknown constant vector. In practice, however, we sometimes can treat it as a realization of a certain random variable $\mathbf{X}$, the distribution of which, $p(\mathbf{x})$, is usually referred as a priori knowledge of $\mathbf{x}$. If we can obtain a priori knowledge beforehand, it can be built in our estimator, and achieve better performance, which is called the Bayesian estimation.

In Bayesian estimation, since the parameter is not a constant anymore, the performance against a specific $\mathbf{x}$ is not that insightful anymore. Instead, the expectation of a given cost function, $\ell(\mathbf{x}, \hat{\mathbf{x}})$, is used to evaluate the performance, i.e.,

$$\hat{\mathbf{x}}(\mathbf{Y}) = \arg\min_{\hat{\mathbf{x}}} \mathbb{E}\left[\ell(\mathbf{X}, \hat{\mathbf{x}}(\mathbf{Y}))\right].$$

Using the Bayes rule, i.e., $p_{\mathbf{x}}(\mathbf{y})p(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$, we have

$$\begin{aligned}
\mathbb{E}\left[\ell(\mathbf{X}, \hat{\mathbf{x}}(\mathbf{Y}))\right] &= \int \ell(\mathbf{x}, \hat{\mathbf{x}}) p_{\mathbf{x}}(\mathbf{y}) p(\mathbf{x}) d\mathbf{x} d\mathbf{y} \\
&= \int p(\mathbf{y}) d\mathbf{y} \int \ell(\mathbf{x}, \hat{\mathbf{x}}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \\
&\triangleq \int \ell(\hat{\mathbf{x}}|\mathbf{y}) p(\mathbf{y}) d\mathbf{y},
\end{aligned}$$

where $\ell(\hat{\mathbf{x}}|\mathbf{y})$ is the so-called posterior cost function. Therefore, the Bayes estimation is actually to minimize the posterior cost:

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg\min_{\mathbf{x}} \ell(\mathbf{x}|\mathbf{y}).$$

The cost function $\ell(\mathbf{x}, \hat{\mathbf{x}})$, for instance, can be $\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2$, or $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$, by minimizing the expectation of which, we get the so-called minimum mean squared error (MMSE) estimator, or minimum mean absolute error (MMAE)

estimator. Particularly, if

$$\ell(\mathbf{x}, \hat{\mathbf{x}}) = \mathbb{1}_{\{\|\mathbf{x}-\hat{\mathbf{x}}\| \leq \varepsilon\}} = \begin{cases} 1 & \text{if } \|\mathbf{x}-\hat{\mathbf{x}}\| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases},$$

then as $\varepsilon \to 0$, the corresponding estimator is

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}),$$

which is known as the maximum a posterior (MAP) estimator. By the Bayes rule, we can also write a MAP estimator as

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \ln p_{\mathbf{x}}(\mathbf{y}) + \ln p(\mathbf{x}). \tag{2.11}$$

In Eq. (2.11), if $\mathbf{X}$ is uniformly distributed over a specific subset of $\mathcal{X}$, then $p(\mathbf{x})$ would be a constant and can be safely dropped. In that case, it is identical to the ML estimator Eq. (2.9). Therefore, the ML estimator can also be treated as a MAP estimator when uniform distribution was assumed as a priori.

2.3 MRSI and Related Work

Spectroscopy in general is the study of the interaction between matter and electromagnetic radiation. MR spectroscopy (MRS), specifically, is a kind of spectroscopy study of the interaction between nuclei and magnetic fields. It has become one of the most versatile forms of spectroscopy (de Graaf [1]), since the discovery by Proctor and Yu [18], and Dickinson [19] that nuclei from the same kind of atom absorb energy at different resonance frequencies, by virtue of what later came to be called chemical shift effects. The chemical shift effects are related to the chemical environment of the nuclei. Therefore, multiple chemicals can be detected and differentiated using the MR spectroscopy method. Compared to MR spectroscopy in the early days, modern MRS is a combination of pulse excitation and Fourier transform (Ernst and Anderson [20]). The acquired spectra can then be represented as

$$\rho(f) = \int \rho(t) e^{-j2\pi ft} dt, \tag{2.12}$$

where different chemical shifts form different peaks in the frequency domain, as shown in Fig. 2.1.

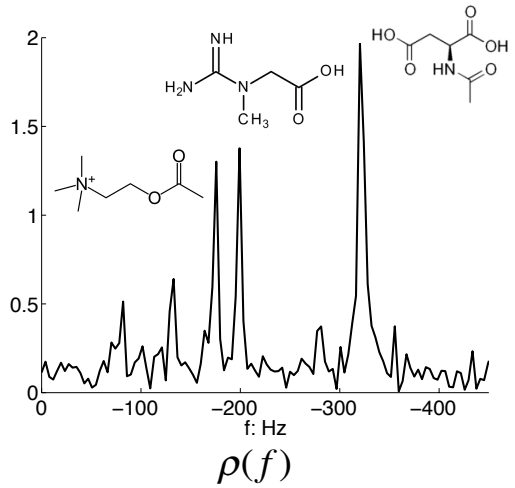


Figure 2.1: Example spectra and molecular structures of N-acetyl aspartate (NAA), creatine, and myo-inositol, three common metabolites in the human brain.

Conventional magnetic resonance imaging (MRI) focuses on obtaining "body" images through imaging the hydrogen protons in the human body. Resulting images with various kinds of contrast can provide very useful physiology information, including anatomical structures and tumors (Fig. 2.2). In MRI, the spatial localization was made possible by using magnetic gradients [21]. The acquired image is modeled as

$$\rho(\mathbf{x}) = \int \rho(\mathbf{k}) e^{j2\pi \mathbf{k}\cdot\mathbf{x}} d\mathbf{k}, \tag{2.13}$$

where $\rho(\mathbf{k})$ is the measurement in the so-called k-space (i.e., spatial frequency domain), and $\mathbf{k} = \int \gamma \mathbf{G}(t)dt$ can be controlled through gradient $\mathbf{G}(t)$.

MR spectroscopic imaging (MRSI) is an integration of conventional MRI and MRS, where both spatial information and spectral information are analyzed. These spectra provide physiological information of the local tissue in a heterogeneous sample, like the brain or the heart. To interpret MRSI data quantitatively, it is usually necessary to estimate those parameters with underlying physical or physiological meanings from the measurement.

Mathematically, MRSI can be viewed as a spectral estimation problem

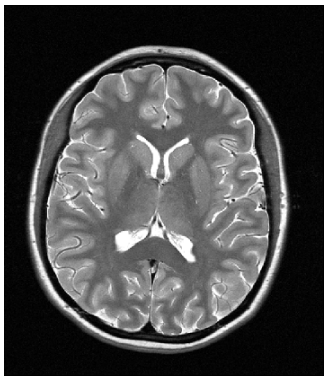



Figure 2.2: Example MRI images of the brain (left) and heart (right).

formulated as

$$d(\mathbf{x}, t) = \int \rho(\mathbf{x}, f) e^{j2\pi ft} df + n(\mathbf{x}, t), \tag{2.14}$$

where $d(\mathbf{x}, t)$ is the measured data in the spatiotemporal domain, $\rho(\mathbf{x}, f)$ is the desired localized spectrum, and $n(\mathbf{x}, t)$ represents additive white Gaussian noise. The main challenge for interpreting MRSI lies in the extremely low SNR.

Over the past few decades, researchers have made significant efforts to address this spectral estimation problem. There are two primary categories of existing spectral estimation methods for MRSI: nonparametric methods and parametric methods. An early nonparametric method was to perform the Fourier transform, that is,

$$\hat{\rho}(\mathbf{x}, f) = \int d(\mathbf{x}, t) e^{-j2\pi ft} dt.$$

The corresponding metabolite "intensity" at spatial location $\mathbf{x}$ would then be simply measured as the area beneath the corresponding peak in $\hat{\rho}(\mathbf{x}, f)$. In high-SNR scenarios, the Fourier transform method is useful and efficient; however, the SNR is often so low in practice that the peak locations cannot even be identified within a spectrum.

Common parametric approaches rely on a damped complex sinusoidal model for MRSI signals [22, 23]:

$$\rho(\mathbf{x}, f) = \int \left(\sum_{l=1}^{L} a_l(\mathbf{x}) z_l(\mathbf{x})^t \right) e^{-j2\pi ft} dt,$$

or in the time-domain,

$$\rho(\mathbf{x}, t) = \sum_{l=1}^{L} a_l(\mathbf{x}) z_l(\mathbf{x})^t, \tag{2.15}$$

where $a_l(\mathbf{x})$ is the concentration of the l-th peak at location $\mathbf{x}$, and $z_l(\mathbf{x})$ are usually called the poles of the linear predictable signal,

$$z_l(\mathbf{x}) = e^{-\theta_l(\mathbf{x}) - j2\pi f_l(\mathbf{x})},$$

where $\theta_l(\mathbf{x})$, and $f_l(\mathbf{x})$ are the damping factor, and resonance frequency, respectively. Plug Eq. (2.15) in Eq. (2.14), we get

$$d(\mathbf{x}, t) = \sum_{l=1}^{L} a_l(\mathbf{x}) z_l(\mathbf{x})^t + n(\mathbf{x}, t).$$

This model motivates the application of linear prediction–based methods for spectral estimation, such as LPSVD [22], HSVD [23], and HLSVD [24]. Let $d_{\mathbf{x}}[m]$ denote $d(\mathbf{x}, m\Delta t)$, i.e., the discretized time sequence at location $\mathbf{x}$. Then the LPSVD method, for instance, says that $\forall \mathbf{x}$, $d_{\mathbf{x}}[m], m = 0, 1, \ldots, M-1 (M > L)$, satisfies the following equation:

$$\begin{bmatrix} d_{\mathbf{x}}[L-2] & d_{\mathbf{x}}[L-3] & \cdots & d_{\mathbf{x}}[0] \\ d_{\mathbf{x}}[L-1] & d_{\mathbf{x}}[L-2] & \cdots & d_{\mathbf{x}}[1] \\ \vdots & \vdots & \ddots & \vdots \\ d_{\mathbf{x}}[M-2] & d_{\mathbf{x}}[M-3] & \cdots & d_{\mathbf{x}}[M-L] \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{bmatrix} = \begin{bmatrix} d_{\mathbf{x}}[L-1] \\ d_{\mathbf{x}}[L] \\ \vdots \\ d_{\mathbf{x}}[M-1] \end{bmatrix},$$

from which we can determine the linear prediction coefficients $\beta_1, \ldots, \beta_L$ in the least squares sense (Eq. (2.2)), or in the total least squares sense. Rooting the polynomial equation

$$z^L + \beta_1 z^{L-1} + \cdots + \beta_{L-1} z + \beta_L = 0$$

yields estimates of the poles in (2.15). The linear parameters (i.e., $a_l(\mathbf{x})$) can then be determined by projecting $d(\mathbf{x}, t)$ onto the subspace spanned by $\{z_l(\mathbf{x})^t\}_{l=1}^{L}$. By using a parametric model, linear prediction–based methods can perform better spectral estimation than nonparametric methods. However, an inherent drawback of those methods is their limitation in in-

corporating prior knowledge (despite the existence of some variants of linear prediction methods that can impose some special forms of prior knowledge, e.g., [25, 26]). This limitation strongly prohibits linear prediction methods from being applied to low-SNR scenarios, which is often the case in practice. Therefore, state-of-the-art methods usually take advantage of the spectral basis functions, and estimate the spectral parameters in the maximum likelihood sense.

CHAPTER 3

METHOD

3.1 Data Preprocessing

In ^{1}H-MRSI, the desired metabolite signals are usually overwhelmed by water and lipid signals due to the huge abundance of hydrogen protons and lipid protons in the human body. Besides, macromolecular signals also generate wavy baselines (i.e., spectral components with a broad line width), which makes it more difficult to separate out the metabolite signals. These contaminating signals are usually called the nuisance signals in spectral estimation of ^{1}H-MRSI. Effectively removing nuisance signals as well as protecting desired metabolite signals is very important for practical ^{1}H-MRSI.

Another practical issue for MRSI is the spatial inhomogeneity of magnetic fields. Since field inhomogeneity brings additional line shape distortions to the observed spectra, we will also need to correct it in the preprocessing step.

3.1.1 Water and Lipid Removal

A common procedure to remove water and lipid signals is to apply suppression pulses during the data acquisition process. For example, the chemical-shift selective saturation (CHESS) pulses suppress water by selectively saturating signals within a certain frequency band that is assumed to be where water signals reside [27, 28]; the outer-volume-suppression (OVS) [29, 30] can be applied to suppress lipid signals. However, due to some practical issues (e.g., field inhomogeneity), the effectiveness of these suppression techniques is strongly limited, and the residual water and lipid signals still overwhelm the desired metabolite signals so that postprocessing techniques are usually needed for further nuisance removal [22, 23, 31].

In this work, we propose to use the Union-of-Subspaces (UoSS) model [32]

as a postprocessing method. The UoSS modeling of an MRSI experiment is given by

$$\rho(\mathbf{x},t)=\sum_{p=1}^{P_M}u_{M,p}(\mathbf{x})v_{M,p}(t)+\sum_{p=1}^{P_W}u_{W,p}(\mathbf{x})v_{W,p}(t)+\sum_{p=1}^{P_L}u_{L,p}(\mathbf{x})v_{L,p}(t),$$

where $v(t)$ is the temporal basis, $u(t)$ is the spatial coefficients, and subscripts "M", "W", and "L" represent metabolite, water, and lipid, respectively. The UoSS model exploits the partial separability [33, 34] of medical images, which decouples the determination of temporal basis and spatial coefficients, and enables more efficient nuisance removal schemes.

The determination of temporal bases $v(t)$ is achieved through constructing the Casorati matrix, i.e.,

$$C(\rho)=\begin{bmatrix}\rho(\mathbf{x}_1,t_1) & \rho(\mathbf{x}_1,t_2) & \cdots & \rho(\mathbf{x}_1,t_M)\\ \rho(\mathbf{x}_2,t_1) & \rho(\mathbf{x}_2,t_2) & \cdots & \rho(\mathbf{x}_2,t_M)\\ \vdots & \vdots & \ddots & \vdots\\ \rho(\mathbf{x}_P,t_1) & \rho(\mathbf{x}_P,t_2) & \cdots & \rho(\mathbf{x}_P,t_M)\end{bmatrix},$$

and then performing a singular value decomposition (SVD) truncation to obtain its principal right eigenvectors as temporal bases. This process is performed for water, lipids, and metabolites individually. In practice, we retrieve the temporal bases from a calibration dataset with high temporal resolution, i.e., the chemical shift imaging (CSI) dataset.

Given the retrieved temporal bases $\mathbf{V}_M$, $\mathbf{V}_W$, and $\mathbf{V}_L$, we determine the spatial coefficients via the following optimization problem.

$$\hat{\mathbf{U}}=\arg\min_{\mathbf{U}}\|\mathbf{d}-\mathbf{F}\{\mathbf{U}\hat{\mathbf{V}}^{\mathrm{H}}\}\|_2^2+\lambda\mathrm{R}(\mathbf{U}),$$

where $\mathbf{d}$ is the vector form of the (k,t)-space measurement, $\mathbf{F}$ is the forward model transforming (x,t)-space data to (k,t)-space, which is usually the discrete Fourier transform (DFT), $\hat{\mathbf{V}}=[\mathbf{V}_M,\mathbf{V}_W,\mathbf{V}_L]$ is a stack of all the estimated temporal basis, and $\mathrm{R}(\mathbf{U})=\|\mathbf{U}\|_{\mathrm{F}}^2$ is a regularization term for spatial coefficient $\mathbf{U}$. Once we obtain $\hat{\mathbf{U}}$, the corresponding water and lipid signals can be removed from the data.

3.1.2 Baseline Removal

In ^{1}H-MRSI, there are many MR measurable metabolites containing useful information, as well as some macromolecules. These macromolecular signals usually have a broader spectral line width (baseline), so that they overlap with most of the metabolites, making the estimation of metabolite signals very ill-conditioned [35, 36, 37]. There are many methods to remove the baselines in practice [6, 7, 8, 9, 38, 39]. Here we use the so-called "Subtract" method [8], which uses smooth functions to model the baseline signals.

After water and lipid removal, the residual signal at each spatial location can be modeled as

$$\rho(t) = \rho_M(t) + \rho_B(t),$$

where "M" and "B" denote metabolite and baseline, respectively. Since baselines have broad line widths, they decay much faster than the metabolites in the time domain. The Subtract method takes advantages of this property by assuming

$$\rho(t) \approx \rho_M(t), t \geq t_0,$$

where t_0 is a selected time point after which the baselines are below the noise flaw. Then by estimating $\rho_M(t)$ from $\rho(t)$, $t \geq t_0$, using a parametric model, we can recover $\hat{\rho}_M(t)$, $t < t_0$, by extrapolation. Then "subtract" the recovered metabolite signal,

$$r(t) = \rho(t) - \hat{\rho}_M(t),$$

from which we can perform a curve fit to $r(t)$ using either smoothing splines or other smooth functions (e.g., wavelets). The fitted curve is the estimated baselines, $\hat{\rho}_B(t)$, which can be readily removed from $\rho(t)$.

3.1.3 Field Inhomogeneity Correction

The field inhomogeneity comes from

1. Instrumental imperfection;
2. Interference by the imaging object.

MRS signals are acquired in the so-called k-space, which is a Fourier transform of the x-space:

$$d(\mathbf{k},t) = \mathcal{F}_{\mathbf{x}}\{\sum_{n=1}^{N} a_n(\mathbf{x})e^{-\theta_n(\mathbf{x})t}e^{-j2\pi\Delta f(\mathbf{x})t}\varphi_n(t)\} + \xi(\mathbf{k},t),$$

where $\mathcal{F}_{\mathbf{x}}$ is the Fourier transform applied to $\mathbf{x}$, $\Delta f(\mathbf{x})$ is the inhomogeneous magnetic field map, and $\xi(\mathbf{k},t)$ is the thermal noise (assumed to be white Gaussian). In practice, only a finite number of k-space points can be covered. If the center rectangular of k-space is covered, a most commonly used reconstruction (i.e., reconstruction from k-space to x-space) method is to perform the inverse Fourier transform, which yields

$$d(\mathbf{x},t) = \sum_{n=1}^{N} a_n(\mathbf{x})e^{-\theta_n(\mathbf{x})t}e^{-j2\pi\Delta f(\mathbf{x})t}\varphi_n(t) *_{\mathbf{x}} w(\mathbf{x}) + \tilde{\xi}(\mathbf{x},t), \tag{3.1}$$

where "$*_{\mathbf{x}}$" represents the convolution operation in spatial domain, $w(\mathbf{x})$ is the so-called point spread function brought by finite coverage of the k-space (i.e., k-space truncation), and $\tilde{\xi}(\mathbf{x},t) = \xi(\mathbf{x},t)*_{\mathbf{x}}w(\mathbf{x})$. In the one-dimensional case, specifically, if the k-space is covered from $-k_m/2$ to $k_m/2$, then

$$w(x) = \mathrm{sinc}(k_m x),$$

which obviously has infinite spatial support. Because of this convolution operation, signals from the surrounding spatial locations will also make contributions to a location $\mathbf{x}$, and the original spectrum at location $\mathbf{x}$ is distorted.

There exist multiple approaches to correct this field inhomogeneity effect [40, 41, 42]. Here we resort to another experiment specifically designed to acquire the field inhomogeneity map, and treat that field map as a priori. Then the conjugate phase (CP) algorithm [41] is used to correct the field inhomogeneity.

3.2 Spectral Estimation for MRSI

The spectral estimation has been a longstanding problem for quantitatively interpreting MRSI data. Along with those practical issues, such as strong

nuisance contamination, wavy baselines, and B_0 field inhomogeneity, its main challenges are the large overlaps between peaks originating from different metabolites, and the extremely low SNR.

By taking quantum mechanics and molecular structures into consideration, we are able to acquire prior information of the spectra of different metabolites in the form of spectral constraints, so that overlapping peaks can be separated into different metabolites, which provides more robustness against low SNR. To further improve the estimation robustness, we propose to incorporate spatial constraints as well, since in real biological samples, the spectral parameters always have spatial correlations that can be exploited in estimation.

3.2.1 Spectral Constraints

State-of-the-art methods (e.g., VARPRO [4], LCModel [6], Young et al. [7], QUEST [8], and AQSES [9]) have proposed to impose spectral constraints in the form of spectral basis functions. This is motivated by the fact that within one spectrum $\rho(\mathbf{x}, f)$, different spectral components (i.e., peaks) originating from the same metabolite should be related to each other in terms of their concentration ratios, phases, damping factors, and frequency locations. For example, Fig. 3.1 shows the spectra of several typical metabolites, from which we can clearly see that the concentration ratios and frequency locations are fixed. Compared to Eq. (2.15), which treats every spectral component separately, $\rho(\mathbf{x}, t)$ can be modeled using basis functions instead:

$$\rho(\mathbf{x}, t) = \sum_{n=1}^{N} a_n(\mathbf{x}) e^{-\theta_n(\mathbf{x})t} \varphi_n(t), \tag{3.2}$$

where $\varphi_n(t)$ is the so-called basis function for the n-th metabolite, which can be accurately obtained beforehand using quantum simulation (for example, GAVA [43]) or in vitro experiments. The spectral basis functions can be expressed as

$$\varphi_n(t) = \sum_{l=1}^{L_n} \alpha_{l,n} e^{j\beta_{l,n}} e^{-j2\pi f_{l,n} t},$$

where L_n is the number of spectral components in the spectrum of the n-th metabolite, and $\alpha_{l,n}, \beta_{l,n}$ and $f_{l,n}$ are the spin density ratio, relative phase,

and resonance frequency of its l-th peak. Obviously, the introduction of basis functions $\varphi_n(t)$ greatly reduces the dimension of the parameter space, and helps improve spectral estimation. For instance, one major problem of treating every peak separately in a spectrum is the overlapping of peaks in the frequency domain, but by introducing full spectra, we might still distinguish two overlapping peaks if their corresponding metabolites have different peaks at other chemical shifts. From the perspective of optimization, incorporating $\varphi_n(\cdot)$ is equivalent to imposing equality constraints over the spectral parameters, and the parameter search space is constrained in a much lower dimensional subspace. From the perspective of estimation, the resulting estimation problem becomes less ill-conditioned since overlapping peaks can be differentiated by their corresponding peaks at other chemical shifts.

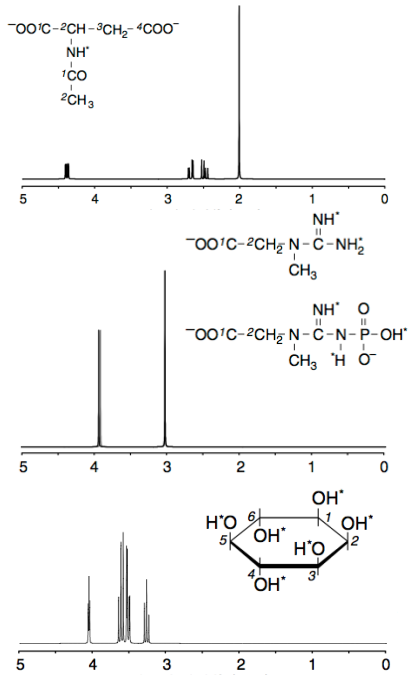


Figure 3.1: Example basis functions [1] and molecular structures of NAA, creatine, and myo-inositol, three common metabolites in the human brain.

The discretized MRSI measurement $d(\mathbf{x}, t)$ at spatial location $\mathbf{x}_p$ and time instant t_q for $p = 1, \cdots, P$, and $q = 1, \cdots, Q$, from a sample with N metabolites can be expressed as

$$d(\mathbf{x}_p, t_q) = \rho(\mathbf{x}_p, t_q) + \xi_{pq}, \tag{3.3}$$

where $\rho(\mathbf{x}, t)$ is defined in Eq. (3.2), and ξ_{pq} represents measurement noise (often assumed to be white Gaussian). Note that we have assumed perfect removal of nuisance signals and field correction here.

For notation convenience, Eq. (3.3) can be written in a vector-matrix form as

$$\mathbf{d}_p = \mathbf{K}(\boldsymbol{\theta}_p)\mathbf{a}_p + \boldsymbol{\xi}_p, \quad p = 1, \cdots, P, \tag{3.4}$$

where $\mathbf{d}_p$ contains all the measured data at location $\mathbf{x}_p$ put in a vector form, i.e.,

$$\mathbf{d}_p = \begin{bmatrix} d(\mathbf{x}_p, t_1) \\ d(\mathbf{x}_p, t_2) \\ \vdots \\ d(\mathbf{x}_p, t_Q) \end{bmatrix},$$

and similarly, $\mathbf{a}_p$ and $\boldsymbol{\xi}_p$ are the concentration vector and noise vector, respectively. $\mathbf{K}(\boldsymbol{\theta}_p)$ is the model matrix with $\boldsymbol{\theta}_p$ containing all the nonlinear unknown parameters (i.e., $\theta_n(\mathbf{x}_p)$) at location $\mathbf{x}_p$ as defined in (3.2), i.e.,

$$\mathbf{K}(\boldsymbol{\theta}_p) = \begin{bmatrix} e^{-\theta_1(\mathbf{x}_p)t_1}\varphi(t_1) & e^{-\theta_2(\mathbf{x}_p)t_1}\varphi(t_1) & \cdots & e^{-\theta_N(\mathbf{x}_p)t_N}\varphi(t_1) \\ e^{-\theta_1(\mathbf{x}_p)t_2}\varphi(t_1) & e^{-\theta_2(\mathbf{x}_p)t_1}\varphi(t_2) & \cdots & e^{-\theta_N(\mathbf{x}_p)t_N}\varphi(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\theta_1(\mathbf{x}_p)t_1}\varphi(t_Q) & e^{-\theta_2(\mathbf{x}_p)t_1}\varphi(t_Q) & \cdots & e^{-\theta_N(\mathbf{x}_p)t_N}\varphi(t_Q) \end{bmatrix}.$$

The goal of spectral estimation in MRSI is to determine $\mathbf{a}_p$ and $\boldsymbol{\theta}_p$ from $\mathbf{d}_p$. This is done in existing methods by solving the following optimization problems (or maximum likelihood estimation problems for Gaussian noise, as we introduced in Eq. (2.10)):

$$(\hat{\mathbf{a}}_p, \hat{\boldsymbol{\theta}}_p) = \arg\min_{\mathbf{a}_p, \boldsymbol{\theta}_p} \|\mathbf{d}_p - \mathbf{K}(\boldsymbol{\theta}_p)\mathbf{a}_p\|_2^2, \tag{3.5}$$

for $p = 1, 2, \cdots, P$. In the absence of any spatial constraints, the above estimation problems can be solved independently for each p, and various nonlinear LS algorithms can be applied.

3.2.2 Spatial Constraints

In contrast to the existing methods, we propose to solve these estimation problems described in Eq. (3.5) by imposing spatial constraints as well as spectral constraints on both $\mathbf{a}_p$ and $\boldsymbol{\theta}_p$. Let $\mathbf{d} = \left[\mathbf{d}_1^T,\ \mathbf{d}_2^T,\ \cdots,\ \mathbf{d}_P^T\right]^T$, $\mathbf{a} = \left[\mathbf{a}_1^T,\ \mathbf{a}_2^T,\ \cdots,\ \mathbf{a}_P^T\right]^T$, and

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{K}(\boldsymbol{\theta}_1) & & \\ & \ddots & \\ & & \mathbf{K}(\boldsymbol{\theta}_P) \end{bmatrix}.$$

Then we can reformulate the spectral quantification problem as a joint estimation problem in a regularization framework:

$$(\hat{\mathbf{a}}, \hat{\boldsymbol{\theta}}) = \arg\min_{\mathbf{a},\boldsymbol{\theta}} \|\mathbf{d} - \mathbf{K}(\boldsymbol{\theta})\mathbf{a}\|_2^2 + \mathrm{R}(\mathbf{a}, \boldsymbol{\theta}), \tag{3.6}$$

where $\mathrm{R}(\mathbf{a}, \boldsymbol{\theta})$ is a penalization functional to absorb spatial regularizations. Note we can also view Eq. (3.6) as a MAP estimator (see Eq. (2.11)), where

$$\mathrm{R}(\mathbf{a}, \boldsymbol{\theta}) = -\ln p(\mathbf{a}, \boldsymbol{\theta}),$$

and $p(\mathbf{a}, \boldsymbol{\theta})$ is the assumed prior distribution of parameter $\mathbf{a}$ and $\boldsymbol{\theta}$. Therefore, $\mathrm{R}(\mathbf{a}, \boldsymbol{\theta})$ is not only a penalizing term, but also a representation of prior knowledge of the probability distribution of parameters. In practice, we have various prior information, such as tissue edges, transform sparsity, and spatial smoothness, so $\mathrm{R}(\mathbf{a}, \boldsymbol{\theta})$ can be purposely designed to absorb these prior information, further improving the spectral estimation performance.

In the present study, we focus on imposing smoothness constraints. This is motivated by the fact that in most MRSI data obtained from biological samples, spatial distributions of spectral parameters (e.g., T_2^* relaxation time and metabolite concentration) within a tissue are rather smooth. There exist various ways to impose smoothness, for example, ℓ_2 regularization, wavelet sparsity, and total variation. Specifically, we propose to impose spatial constraints on $\boldsymbol{\theta}$ and $\mathbf{a}$ by

$$\mathrm{R}(\mathbf{a}, \boldsymbol{\theta}) = \sum_{n=1}^{N} \mathrm{R}^{(n)}(\mathbf{a}, \boldsymbol{\theta}),$$

where superscript "(n)" represents the n-th metabolite, and

$$\mathrm{R}^{(n)}(\mathbf{a},\boldsymbol{\theta}) = \lambda_n \|\mathbf{W}_{\boldsymbol{\theta}}\boldsymbol{\theta}^{(n)}\|_2^2 + \eta_n \|\mathcal{W}_{\mathbf{a}}\{\mathbf{a}^{(n)}\}\|_1,$$

where λ_n and η_n are both regularization parameters. Therefore, the proposed formulation is

$$(\hat{\mathbf{a}},\hat{\boldsymbol{\theta}}) = \arg\min_{\mathbf{a},\boldsymbol{\theta}} \|\mathbf{d} - \mathbf{K}(\boldsymbol{\theta})\mathbf{a}\|_2^2 + \sum_{n=1}^{N} \lambda_n \|\mathbf{W}_{\boldsymbol{\theta}}\boldsymbol{\theta}^{(n)}\|_2^2 + \sum_{n=1}^{N} \eta_n \|\mathcal{W}_{\mathbf{a}}\{\mathbf{a}^{(n)}\}\|_1. \tag{3.7}$$

The weighting matrix $\mathbf{W}_\theta$ (derived from one or multiple references such as anatomic images) is used to preserve edges. Explicitly, this smoothness-based regularization functional is defined in the same way as in [44]:

$$\|\mathbf{W}_{\boldsymbol{\theta}}\boldsymbol{\theta}^{(n)}\|_2^2 = \sum_{p_1=1}^{P} \sum_{\substack{p_2>p_1 \\ p_2\in\Omega_{p_1}}} w_{p_1,p_2} |\theta_{p_1}^{(n)} - \theta_{p_2}^{(n)}|^2, \tag{3.8}$$

where w_{p_1,p_2} is a positive weighting coefficient, Ω_{p_1} is the set of all the voxels that are in the adjacent neighbor to voxel p_1, and $\theta_p^{(n)}$ is the p-th nonlinear parameter for the n-th metabolite, $p = 1,\ldots,P, n = 1,\ldots,N$. The weighting coefficients w_{p_1,p_2} can be derived from one or multiple reference images (see [44] for details). Conceptually, the coefficient should be large in smooth areas (to promote smoothness), and be small around edge areas (to preserve edges). Finally, $\mathcal{W}_{\mathbf{a}}$ is a sparsifying operator, e.g., the wavelet transform, or total variation (TV) transform, which all promote spatial smoothness in $\mathbf{a}$. Here we have chosen $\mathcal{W}_{\mathbf{a}}$ to be the second order total generalized variation (TGV) transform [45]:

$$\|\mathcal{W}_{\mathbf{a}}\{\mathbf{u}\}\|_1 = \min_{\mathbf{p_1},\mathbf{p_2}} \alpha_0 \|\nabla\mathbf{u} - \mathbf{p}\|_1 + \alpha_1 \sum_{j=1}^{4} \|(\mathcal{E}(\mathbf{p}))_j\|_1, \tag{3.9}$$

where $\alpha_1, \alpha_2 > 0$ are two regularization parameters. The operators ∇ and $\mathcal{E}$ are given by

$$\nabla\mathbf{u} = \begin{bmatrix} \partial_x\mathbf{u} \\ \partial_y\mathbf{u} \end{bmatrix},$$

$$\mathcal{E}(\mathbf{p}) = \begin{bmatrix} \partial_x\mathbf{p}_1 & \frac{1}{2}(\partial_y\mathbf{p}_1 + \partial_x\mathbf{p}_2) \\ \frac{1}{2}(\partial_y\mathbf{p}_1 + \partial_x\mathbf{p}_2) & \partial_y\mathbf{p}_2 \end{bmatrix},$$

where ∂_x and ∂_y denote the finite difference along x and y direction, respectively. Next we will discuss how to solve Eq. (3.7) efficiently.

3.3 Algorithm

The problem in Eq. (3.7) is a large-scale, joint estimation problem with nonlinearity. To solve it in practice, we propose to solve the nonlinear parameter $\boldsymbol{\theta}$ firstly, then fix it and solve the linear parameter $\mathbf{a}$. Specifically, we sequentially solve the following two problems:

$$(\tilde{\mathbf{a}}, \hat{\boldsymbol{\theta}}) = \arg\min_{\mathbf{a},\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\boldsymbol{\theta})\mathbf{a}\|_2^2 + \sum_{n=1}^{N} \lambda_n \|\mathbf{W}_{\boldsymbol{\theta}} \boldsymbol{\theta}^{(n)}\|_2^2, \tag{3.10}$$

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2 + \sum_{n=1}^{N} \eta_n \|\mathcal{W}_{\mathbf{a}}\{\mathbf{a}^{(n)}\}\|_1, \tag{3.11}$$

where in Eq. (3.10), $\tilde{\mathbf{a}}$ is obtained along with $\hat{\boldsymbol{\theta}}$, but $\hat{\boldsymbol{\theta}}$ is fixed for Eq. (3.11), while $\tilde{\mathbf{a}}$ can be the initial value for $\hat{\mathbf{a}}$.

3.3.1 Proposed Algorithm-Step 1

To solve the problem in Eq. (3.10), we can use the variable projection strategy to reduce the number of parameters, which is to solve the following problem instead.

$$\begin{cases} \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\boldsymbol{\theta})\mathbf{K}(\boldsymbol{\theta})^{\dagger}\mathbf{d}\|_2^2 + \sum_{n=1}^{N} \lambda_n \|\mathbf{W}_{\boldsymbol{\theta}} \boldsymbol{\theta}^{(n)}\|_2^2, \\ \tilde{\mathbf{a}} = \mathbf{K}(\hat{\boldsymbol{\theta}})^{\dagger}\mathbf{d}. \end{cases} \tag{3.12}$$

The reason that we can reformulate Eq. (3.10) into Eq. (3.12) can be summarized as the following theorem.

Theorem 3.1 ("Equivalence" of Eq. (3.10) and Eq. (3.12))**.** *Assume that the range space of* $\mathbf{K}(\boldsymbol{\theta})$ *has a constant dimension for all* $\boldsymbol{\theta}$ *in an open set* $\boldsymbol{\Theta} \subset \mathbb{R}^{L \times N}$.

1. *If* $\hat{\boldsymbol{\theta}}$ *is a critical point (or a global minimizer in* $\boldsymbol{\Theta}$*) of Eq.* (3.12)*, and*

$$\tilde{\mathbf{a}} = \mathbf{K}(\hat{\boldsymbol{\theta}})^{\dagger}\mathbf{d}, \tag{3.13}$$

then $(\hat{\boldsymbol{\theta}}, \tilde{\mathbf{a}})$ *is a critical point of Eq.* (3.10) *(or a global minimizer for* $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$*).*

2. *If* $(\hat{\boldsymbol{\theta}}, \tilde{\mathbf{a}})$ *is a global minimizer of Eq.* (3.10) *for* $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$*, then* $\hat{\boldsymbol{\theta}}$ *is a global minimizer of Eq.* (3.12) *in* $\boldsymbol{\Theta}$*. Furthermore, if there is a unique* $\tilde{\mathbf{a}}$ *among the minimizing pairs of Eq.* (3.10)*, then* $\tilde{\mathbf{a}}$ *must satisfy (3.12).*

Conceptually, the reformulation of Eq. (3.10) into Eq. (3.12) follows the variable projection strategy: to represent the linear term by its closed-form solution as if the nonlinear term is fixed. Theorem 3.1 is an immediate extension of Thm 2.2 with regularization on the nonlinear term, and can be proved in a similar way (see Appendix A for details).

The joint quantitation problem in (3.12) is still much larger than the individual quantitation problems in (3.5), so it is desirable to solve them efficiently. The problem in (3.12) is a weighted ℓ_2-norm regularized nonlinear least-squares problem, which can be solved using a quasi-Newton method where only gradient evaluation is required. We used the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [46] to solve (3.12) because of its computational efficiency and relatively low memory usage. The gradient of (3.12) can be derived as follows. Let $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})^{\dagger}\mathbf{d}$, and

$$f_1(\boldsymbol{\theta}) = \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\boldsymbol{\theta})\mathbf{a}(\boldsymbol{\theta})\|_2^2, \quad f_2(\boldsymbol{\theta}) = \sum_{n=1}^{N} \lambda_n \|\mathbf{W}_{\boldsymbol{\theta}}\boldsymbol{\theta}^{(n)}\|_2^2$$

Then the gradient desired is simply the sum of ∇f_1 and ∇f_2 by the linearity of gradient. It is easier to obtain ∇f_2:

$$\nabla f_2 = 2\sum_{n=1}^{N} \lambda_n \mathbf{W}_{\boldsymbol{\theta}}^{*}\mathbf{W}_{\boldsymbol{\theta}}\boldsymbol{\theta}^{(n)}.$$

As for ∇f_1, we have

$$\nabla f_1 = \left(\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}}\mathbf{a} + \mathbf{K}\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^{*} \left(\mathbf{K}(\boldsymbol{\theta})\mathbf{a}(\boldsymbol{\theta}) - \mathbf{d}\right), \tag{3.14}$$

where $\partial \mathbf{a}/\partial \boldsymbol{\theta}$ is yet to be determined. Since $\mathbf{a}(\boldsymbol{\theta}) = (\mathbf{K}(\boldsymbol{\theta})^{*}\mathbf{K}(\boldsymbol{\theta}))^{-1}\mathbf{K}(\boldsymbol{\theta})^{*}\mathbf{d}$, we have

$$\mathbf{K}(\boldsymbol{\theta})^{*}\mathbf{K}(\boldsymbol{\theta})\mathbf{a}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})^{*}\mathbf{d}.$$

Taking derivatives to both sides, we have

$$(\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}})^*\mathbf{K}\mathbf{a} + \mathbf{K}^*\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}}\mathbf{a} + \mathbf{K}^*\mathbf{K}\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}} = (\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}})^*\mathbf{d},$$

i.e.,

$$\mathbf{K}^*\mathbf{K}\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}} = (\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}})^*\mathbf{d} - (\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}})^*\mathbf{K}\mathbf{a} - \mathbf{K}^*\frac{\partial \mathbf{K}}{\partial \boldsymbol{\theta}}\mathbf{a},$$

from which we can obtain $\partial \mathbf{a}/\partial \boldsymbol{\theta}$ by solving the equation.

Note in the gradients obtained above, ∇f_1 is constituted by complex numbers since $\mathbf{d}$ and $\mathbf{K}(\boldsymbol{\theta})$ are constituted by complex numbers. In gradient based optimization algorithms, the parameter updates will also be complex numbers. However, the parameter $\boldsymbol{\theta}$ should be real in practice (e.g., T_2 relaxation constants are real), and this constraint is not taken into consideration in the algorithm above. To address this issue, we treat the real part and imaginary part of complex numbers separately. Let

$$\mathbf{r}_1(\boldsymbol{\theta}) = \mathbf{d} - \mathbf{K}(\boldsymbol{\theta})\mathbf{a}(\boldsymbol{\theta}),\ \tilde{\mathbf{r}}_1 = \begin{bmatrix} \mathrm{Re}\{\mathbf{r}_1\} \\ \mathrm{Im}\{\mathbf{r}_1\} \end{bmatrix},\ \tilde{f}_1 = \frac{1}{2}\|\tilde{\mathbf{r}}_1\|_2^2.$$

Then $f_1 \equiv \tilde{f}_1$, and

$$\nabla \tilde{f}_1 = \begin{bmatrix} \mathrm{Re}\{\nabla f_1\} \\ \mathrm{Im}\{\nabla f_1\} \end{bmatrix}, \tag{3.15}$$

where ∇f_1 is defined in Eq. (3.14). Using (3.15) as gradient, the constraint that $\boldsymbol{\theta}$ is real can thus be incorporated.

3.3.2 Proposed Algorithm-Step 2

After $\hat{\boldsymbol{\theta}}$ is estimated, we then solve the ℓ_1-norm regularized least-squares problem in (3.11) using a so-called alternating direction method of multipliers (ADMM) [47]. For the following problem,

$$\begin{aligned} &\min_{x,z} f(x) + g(z), \\ &\text{s.t.}\ \ Ax + Bz = c, \end{aligned}$$

the ADMM is to construct

$$L(x, z; y) = f(x) + g(z) + \frac{\mu}{2}\|Ax + Bz - c - y\|_2^2,$$

and iteratively solve the following subproblems,

$$\begin{cases} x^{(k+1)} = \arg\min_x L(x, z^{(k)}; y^{(k)}), \\ z^{(k+1)} = \arg\min_z L(x^{(k+1)}, z; y^{(k)}), \\ y^{(k+1)} = y^{(k)} + \gamma(c - Ax^{(k+1)} - By^{(k+1)}). \end{cases}$$

It is shown that under certain convex assumptions, if $\mu > 0$ and $0 < \gamma < \frac{\sqrt{5}+1}{2}$, the algorithm above converges [47]. For simplicity, we firstly present the algorithm for the case that $\mathcal{W}_{\mathbf{a}}\{\cdot\}$ in (3.11) is a linear operator of $\mathbf{a}$ (e.g., the wavelet and TV transform) and can thus be represented by a matrix form, i.e., $\mathcal{W}_{\mathbf{a}}\{\mathbf{a}^{(n)}\} = \mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n)}$. As proposed by Guo et al. [48], we introduce an auxiliary variable $\mathbf{u}$, so that (3.11) is equivalent to

$$\begin{aligned} &\min_{\mathbf{a},\mathbf{u}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2 + \sum_{n=1}^{N} \eta_n \|\mathbf{u}^{(n)}\|_1, \\ &\text{s.t. } \mathbf{u}^{(n)} = \mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n)}, \text{ for } n = 1, \ldots, N, \end{aligned} \tag{3.16}$$

where again, the superscript "(n)" represents the n-th metabolite. Obviously, we can now use ADMM to solve Eq. (3.16). Let

$$L(\mathbf{a}, \mathbf{u}; \tilde{\mathbf{u}}) = \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2 + \sum_{n=1}^{N} \Big[\eta_n \|\mathbf{u}^{(n)}\|_1 + \frac{\mu_n}{2}\|\mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n)} - \mathbf{u}^{(n)} - \tilde{\mathbf{u}}^{(n)}\|_2^2\Big].$$

Then we can iteratively solve the following subproblems,

$$\begin{cases} \mathbf{u}^{(k+1)} = \arg\min_{\mathbf{u}} L(\mathbf{a}^{(k)}, \mathbf{u}; \tilde{\mathbf{u}}^{(k)}), \\ \mathbf{a}^{(k+1)} = \arg\min_{\mathbf{a}} L(\mathbf{a}^{(k)}, \mathbf{u}^{(k+1)}; \tilde{\mathbf{u}}^{(k)}), \\ \tilde{\mathbf{u}}^{(k+1)} = \tilde{\mathbf{u}}^{(k)} + \gamma(\mathbf{u}^{(k)} - \mathbf{W}_{\mathbf{a}}\mathbf{a}^{(k)}). \end{cases}$$

That is,

$$\mathbf{u}^{(n,k+1)} = \arg\min_{\mathbf{u}^{(n)}} \|\mathbf{u}^{(n)}\|_1 + \frac{\mu_n}{2\eta_n}\|\mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n,k)} - \mathbf{u}^{(n)} - \tilde{\mathbf{u}}^{(n,k)}\|_2^2, \ \forall n, \tag{3.17}$$

$$\mathbf{a}^{(k+1)} = \arg\min_{\mathbf{a}} \|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2 + \sum_{n=1}^{N} \mu_n \|\mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n)} - \mathbf{u}^{(n,k+1)} - \tilde{\mathbf{u}}^{(n,k)}\|_2^2, \tag{3.18}$$

$$\tilde{\mathbf{u}}^{(n,k+1)} = \tilde{\mathbf{u}}^{(n,k)} + \gamma(\mathbf{u}^{(n,k)} - \mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n,k)}), \forall n, \tag{3.19}$$

where (n) and (k) represent the metabolite index and iteration index respectively.

Subproblem (3.17) can be solved explicitly using shrinkage. In the scalar case, the shrinkage problem is to minimize

$$f(x) = \frac{1}{2}(x - x')^2 + \lambda|x|,$$

and the minimizing value of x is

$$\text{shrink}(x', \lambda) = x' \max\left(1 - \frac{\lambda}{|x'|}, 0\right).$$

In the vector case as in (3.17), $\mathbf{u}^{(n,k+1)}$ can be solved entry-by-entry:

$$\mathbf{u}^{(n,k+1)} = \text{shrink}(\mathbf{W_a}\mathbf{a}^{(n,k)} - \tilde{\mathbf{u}}^{(n,k)}, \eta_n/\mu_n), \forall n = 1, \ldots, N,$$

where

$$\text{shrink}(\mathbf{v}, \eta) = \mathbf{v}. * \max\left(1 - \eta./|\mathbf{v}|, 0\right).$$

Subproblem (3.18) is a large-scale quadratic optimization problem, and can be readily solved by standard convex optimization tools. Therefore, for linear $\mathcal{W}_\mathbf{a}$, we can summarize the algorithm as follows.

Algorithm 1 The ADMM algorithm for solving (3.11) for linear $\mathcal{W}_\mathbf{a}\{\cdot\}$

1: Input: measurement $\mathbf{d}$, nonlinear parameter estimate $\hat{\boldsymbol{\theta}}$, convergence criteria $\mathcal{H}$, spatial regularization parameters $\{\eta_n\}_{n=1}^N$, ADMM parameters γ and $\{\mu_n\}_{n=1}^N$.
2: Initialization: $\mathbf{a}^{(0)}$, $\mathbf{u}^{(0)}$, and $\tilde{\mathbf{u}}^{(0)}$
3: **while** $\mathcal{H}$ not satisfied **do**
4: **for** $j = 1 : N$ **do**
5: $\mathbf{u}^{(j,k+1)} \leftarrow \text{shrink}(\mathbf{W_a}\mathbf{a}^{(j,k)} - \tilde{\mathbf{u}}^{(j,k)}, \eta_j/\mu_j)$
6: **end for**
7: $\mathbf{a}^{(k+1)} \leftarrow$ solve (3.18) using the conjugate gradient descent algorithm
8: **for** $j = 1 : N$ **do**
9: $\tilde{\mathbf{u}}^{(j,k+1)} \leftarrow \tilde{\mathbf{u}}^{(j,k)} + \gamma(\mathbf{u}^{(j,k)} - \mathbf{W_a}\mathbf{a}^{(j,k)})$
10: **end for**
11: **end while**
12: Output: $\mathbf{a}^{k+1}$ as the solution to Eq. (3.11).

3.3.3 Extension to TGV

The algorithm above is for linear $\mathcal{W}_{\mathbf{a}}\{\cdot\}$. Next we show that it can also be extended to the case that $\mathcal{W}_{\mathbf{a}}\{\cdot\}$ represents the TGV transform. Plug TGV (3.9) into (3.11):

$$(\hat{\mathbf{a}}, \hat{\mathbf{p}}) = \arg\min_{\mathbf{a},\mathbf{p}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2 + \alpha_0 \sum_{n=1}^{N} \eta_n \|\nabla \mathbf{a}^{(n)} - \mathbf{p}^{(n)}\|_1 + \alpha_1 \sum_{n=1}^{N} \eta_n \sum_{j=1}^{4} \|(\mathcal{E}(\mathbf{p}^{(n)}))_j\|_1. \quad (3.20)$$

We can see that the optimization functional is not linear in $\mathbf{a}$, so Algorithm 1 does not apply immediately. However, let

$$\tilde{\mathbf{a}} = \begin{bmatrix} \mathbf{a} \\ \mathbf{p} \end{bmatrix}, \tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}, \tilde{\mathbf{K}}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \mathbf{K}(\hat{\boldsymbol{\theta}}) & \\ & \mathbf{0} \end{bmatrix},$$

and

$$\tilde{\mathcal{E}}_j(\tilde{\mathbf{a}}^{(n)}) = (\mathcal{E}(\mathbf{p}^{(n)}))_j, \ j = 1, 2, 3, 4.$$

Then we can reformulate (3.20) as

$$\hat{\tilde{\mathbf{a}}} = \arg\min_{\tilde{\mathbf{a}}} \frac{1}{2}\|\tilde{\mathbf{d}} - \tilde{\mathbf{K}}(\hat{\boldsymbol{\theta}})\tilde{\mathbf{a}}\|_2^2 + \sum_{n=1}^{N} \eta_n \|\tilde{\mathbf{W}}\tilde{\mathbf{a}}^{(n)}\|_1, \quad (3.21)$$

where

$$\tilde{\mathbf{W}}\tilde{\mathbf{a}}^{(n)} = \begin{bmatrix} \alpha_0\nabla, -\alpha_0 I \\ \alpha_1\tilde{\mathcal{E}}_1 \\ \vdots \\ \alpha_1\tilde{\mathcal{E}}_4 \end{bmatrix} \tilde{\mathbf{a}}^{(n)}.$$

Comparing Eq. (3.21) with Eq. (3.11), we can see that when $\mathcal{W}_{\mathbf{a}}\{\cdot\}$ is TGV, Algorithm 1 can still be used after the simple reformulation introduced in (3.21).

CHAPTER 4

RESULTS

The effectiveness of the proposed quantification method is demonstrated via both simulation and experimental results. In this chapter, we will describe the setup of both the simulation and in vivo experimental datasets, and show the spectral estimation results by the proposed method in comparison to one of the state-of-the-art methods.

4.1 Simulation

4.1.1 Setup

A realistic 2-dimensional MRSI simulation dataset for the human brain was generated for our analysis. The synthetic model used in simulation was

$$d(\mathbf{x}, t) = \sum_{n=1}^{6} a_n(\mathbf{x}) e^{-R_{2,n}(\mathbf{x})t} \varphi_n(t) + n(\mathbf{x}, t).$$

Six common metabolites, NAA, creatine (Cre), choline (Cho), glutamate (Glu), myo-inositol (mI), and glutamine (Gln), were used. True metabolite distributions (a_n) are shown in Fig. 4.1, which were designed such that they are smooth within each type of tissue, but different across different types of tissues (gray matter, white matter, and cerebrospinal fluid), so that spatial smoothness can be exploited. The ratios between metabolites was adjusted based on literature values from [1]. The true R_2 map is shown in Fig. 4.2, which is piece-wise constant. The corresponding spectral basis functions, $\varphi_n(t)$, were obtained by quantum mechanics simulation (GAVA [43]) and shown in Fig. 4.3. The spectral integration of the simulated dataset is shown in Fig. 4.4a. White Gaussian noise was also added to the dataset (Fig. 4.4b) to mimic practical thermal noise, with three typical spectra at different SNR

levels shown in Fig. 4.4c.

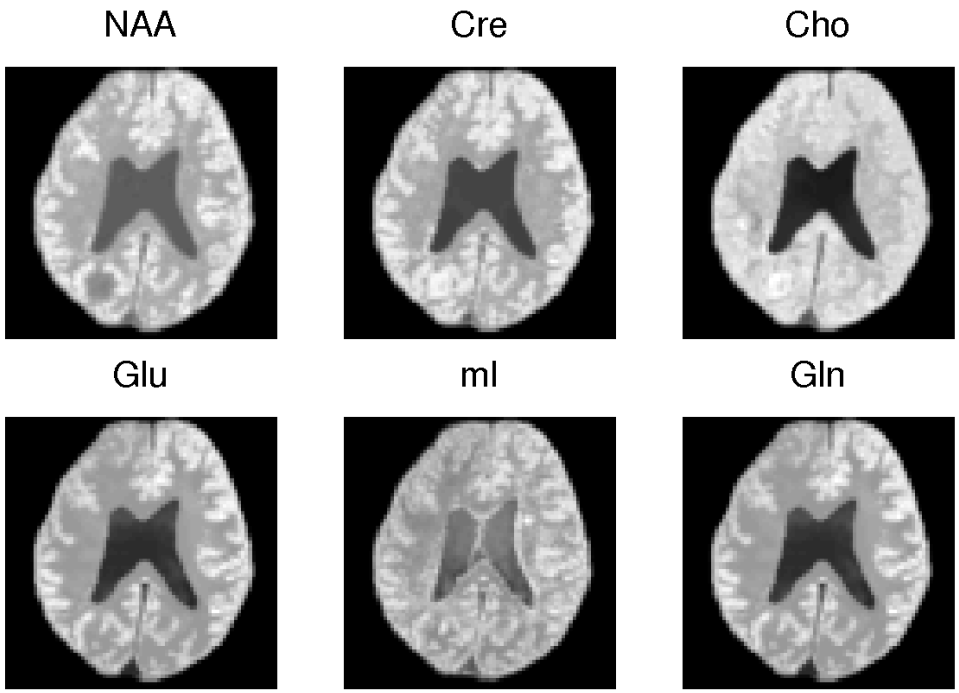


Figure 4.1: True spatial distributions of metabolites used in simulation.

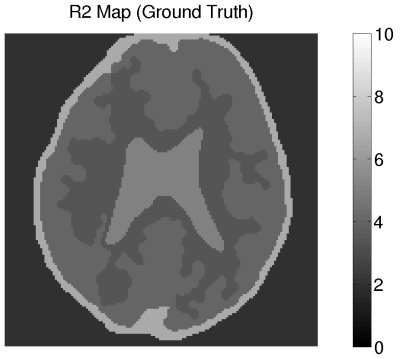


Figure 4.2: True spatial distributions of the R_2 values used in simulation.

4.1.2 Results

We have evaluated the performance of the proposed method and compared it with QUEST [8], a standard method used in practice. As is done in other

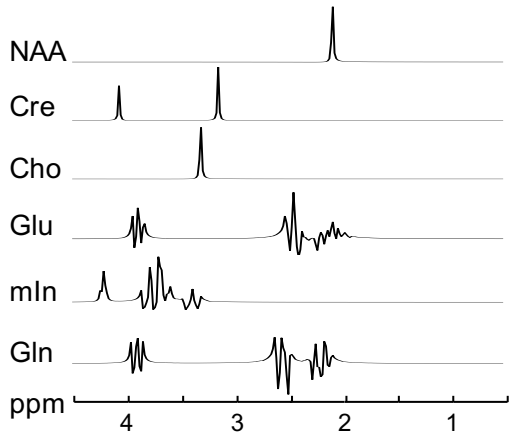


Figure 4.3: Spectral basis functions of the six metabolites used to synthesize the simulation dataset.

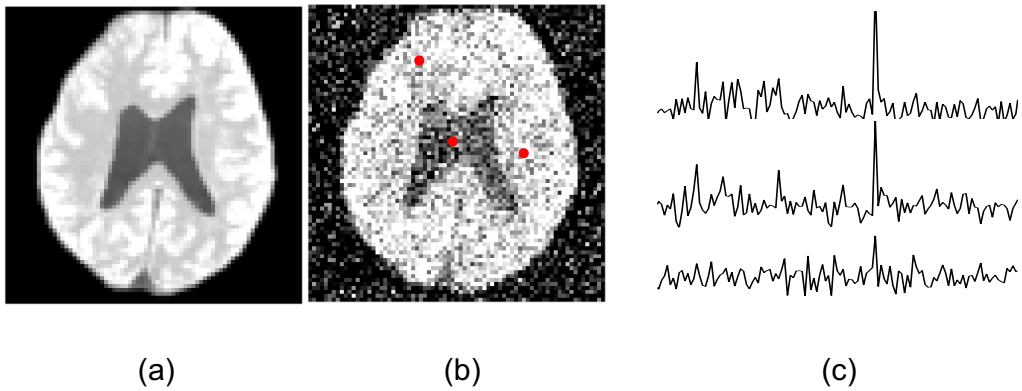


Figure 4.4: (a) Spectral integral of the synthesized dataset; (b) spectral integral after adding white Gaussian noise; and (c) representative spectra, with their locations indicated by the red dots in (b).

state-of-the-art frequency-domain methods (LCModel [6], etc.) and time-domain methods (AQSES [9], etc.), QUEST performs spectral quantification voxel by voxel using a set of pre-determined spectral bases without incorporating spatial constraints.

Figure 4.5 shows a comparison of the typical quantification results for NAA, Cre, Cho, Glx (Glutamate+Glutamine) and mI using QUEST and the proposed method, respectively, on the simulation MRSI dataset. The quantification problem is fundamentally a parameter estimation problem. To better illustrate the estimation performance, Fig. 4.6 shows the synthetic spectra from the true spectral parameters and the estimated parameters for one spatial location.

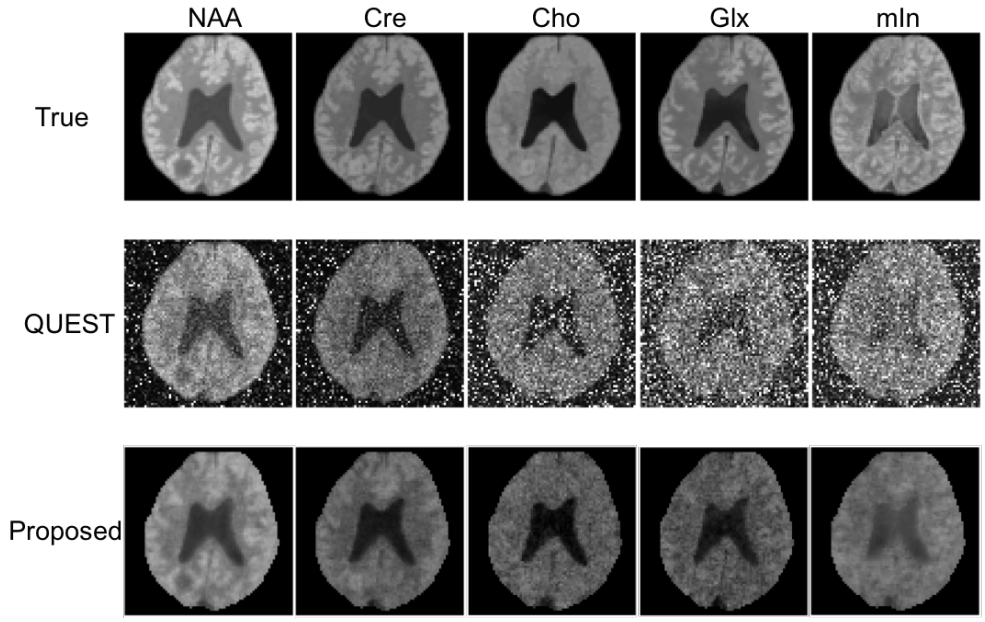


Figure 4.5: True metabolite distributions, metabolite distributions estimated by QUEST and by the proposed are shown in three rows, respectively. Because of the inherent difficulty in separating Glu from Gln, the conventional way is to show Glx (Glu+Gln) instead. Note the significant improvement of the proposed over QUEST.

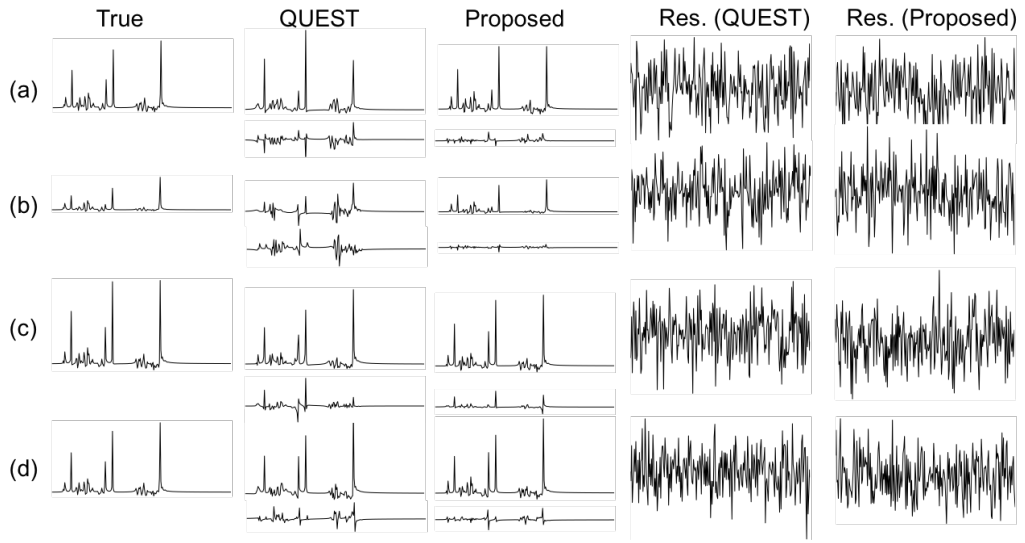


Figure 4.6: Spectral estimation results at some typical points are shown in (a)-(d). For both QUEST and the proposed, the difference between the true spectrum and the estimated spectrum is shown under the estimated spectrum. The right two columns represent the fitting residuals of two methods.

4.2 In Vivo Experiment

4.2.1 Setup

Some representative in vivo results are shown here to demonstrate the effectiveness of the proposed method in practice. The MRSI data were acquired from a healthy volunteer on a Siemens 3 Tesla MRI scanner with a 2-dimensional bipolar echo-planar spectroscopic imaging (EPSI) sequence. The readout bandwidth was 100 kHz, the echo time was 30 ms, the echo spacing was 1.42 ms, and the nominal resolution was approximately 5 mm (field-of-view 220 mm $\times$ 220 mm, matrix size 48 $\times$ 48). The position information (OVS bands, shimming volume, slice position, etc.) is summarized in Fig. 4.7.

4.2.2 Results

Figure 4.8 shows the magnetic field map obtained from a calibration scan, which determines the field inhomogeneity based on the phase evolution over time of water signals. A polynomial interpolation was also performed to

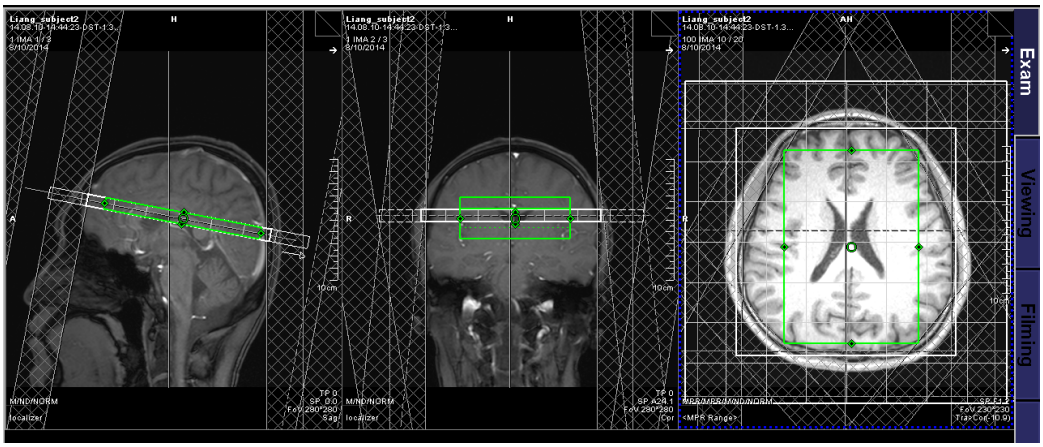


Figure 4.7: Setup for the in vivo experiment.

improve "bad" estimation regions where water signals were weak (e.g., bone regions). The estimated R_2 maps for NAA, Cre, and Cho are shown in Fig. 4.9, and the quantified concentration maps are also shown in Fig. 4.10. Note that the QUEST results showed significant spatial variations (indicating large estimation variance), which were reduced by the proposed method.

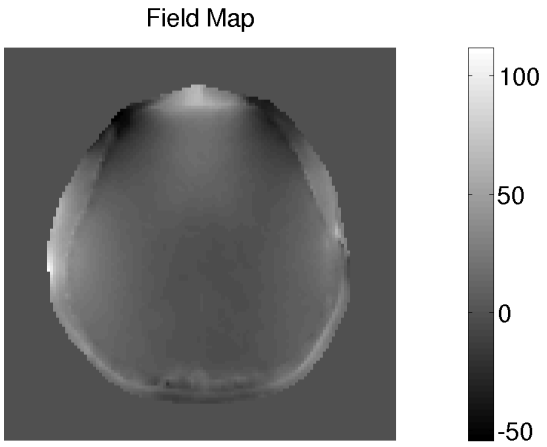


Figure 4.8: The magnetic field map (Hz) acquired from a calibration scan.

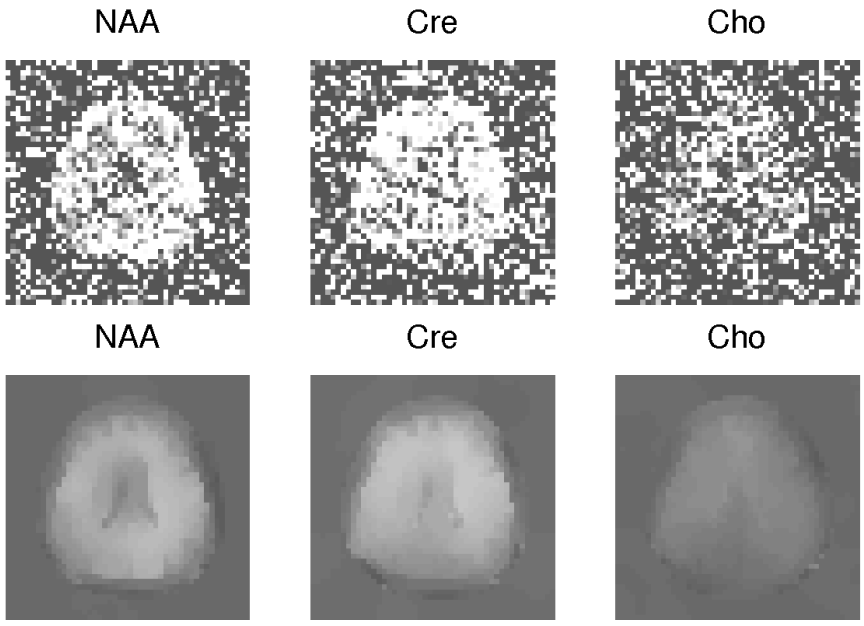


Figure 4.9: The estimated R_2 map by QUEST (top) and the proposed method (bottom).

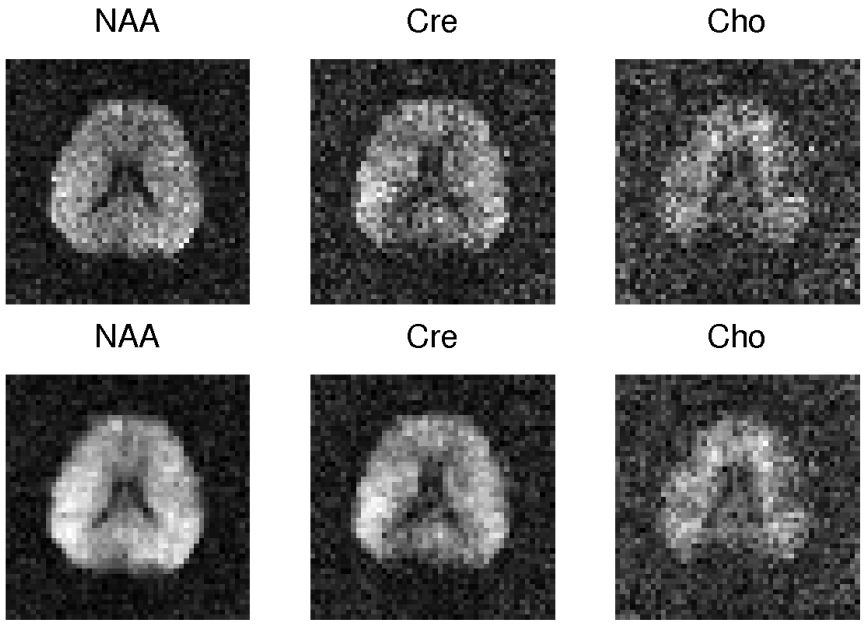


Figure 4.10: The estimated concentration map by QUEST (top) and the proposed method (bottom).

CHAPTER 5

DISCUSSION

From Fig. 4.5, we can see that the estimated concentration maps by QUEST were very noisy (large spatial variations); the proposed method significantly reduced the estimation variance. Obviously, the proposed method has smaller estimation error than QUEST when compared with the ground truth. In Fig. 4.6, both methods had residuals close to the noise floor (columns 4 and 5), indicating comparable performance in terms of data fitting. However, the spectrum from QUEST parameters showed noticeable errors, which were significantly reduced by the proposed method (due to the spatial constraint). The performance improvement by the proposed method is further validated by in vivo experimental data.

In this chapter, we will also discuss one straightforward method of imposing spatial constraints and its disadvantages. Then we provide a theoretical analysis of how much improvement the proposed method can achieve.

5.1 Direct Denoising

One simple idea to apply spatial constraints is to denoise the estimates of a state-of-the-art method by enforcing the spatial smoothness constraint:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{\beta}{2}\|\mathbf{a} - \tilde{\mathbf{a}}\|_2^2 + \mathrm{R}(\mathbf{a}), \tag{5.1}$$

where $\tilde{\mathbf{a}}$ denotes the current estimation of $\mathbf{a}$ by a state-of-the-art method (here we used QUEST as well). As indicated by Fig. 5.1, such a straightforward method can help reduce the noise, but it also introduces some blurring artifacts, and the estimates may also be biased due to non-Gaussianity of the noise in the QUEST estimates. The proposed method overcomes these problems, and perhaps more importantly, the performance of the proposed

method can be much more easily characterized, as we will discuss in the next section.

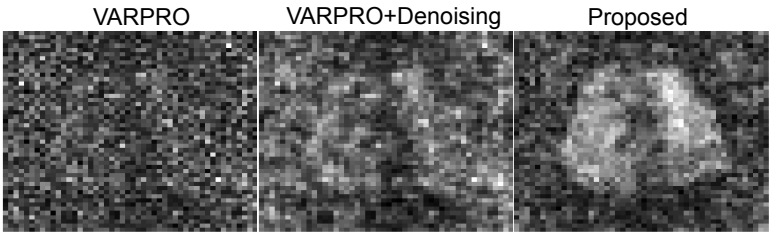


Figure 5.1: Estimated concentration maps of metabolite Glx (Glutamate+Glutamine) using QUEST, denoising of QUEST and the proposed method, respectively, based on the same in vivo MRSI dataset as shown in Fig. 4.10. Note the improved performance of the proposed method over the straightforward denoising method.

5.2 Performance Evaluation

Performance analysis of a spectral estimation method is strongly desired in practical applications. In this section, we propose to use the constrained Cramér-Rao bound theory to characterize the performance of the proposed method. Specifically, in the proposed method, spatial prior information is incorporated in the form of ℓ_1-norm regularization, which can be seen as a convex relaxation of the ℓ_0-norm. The estimation variance can thus be bounded by the constrained Cramér-Rao bound (CRB). Additionally, by comparing the constrained CRB (characterizing the proposed joint spectral estimation method) with the conventional CRB (characterizing point-by-point spectral estimation methods), we can predict how much improvement the proposed method can achieve by incorporating spatial prior information. In this section, the theoretical analysis is also validated by Monte Carlo simulations.

5.2.1 Theory

It is well known that if the concentration $\mathbf{a}$ in (3.16) satisfies a certain level of transform sparsity, then a solution to (3.16) is also a solution to the following

equality constrained optimization problem:

$$\begin{aligned}&\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2,\\ &\text{s.t. } \|\mathbf{W}_{\mathbf{a}}\mathbf{a}^{(n)}\|_0 = M^{(n)}, \text{ for } n = 1, \ldots, N.\end{aligned}$$

or more compactly,

$$\begin{aligned}&\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}(\hat{\boldsymbol{\theta}})\mathbf{a}\|_2^2,\\ &\text{s.t. } \|\mathbf{W}\mathbf{a}\|_0 = M,\end{aligned} \tag{5.2}$$

where $\mathbf{W}$ and M have obvious definitions. Assuming the estimation of $\boldsymbol{\theta}$ is accurate (i.e., $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$), the performance of the proposed method can be characterized by calculating the constrained CRB [16, 49] of the estimator in (5.2), which is a lower bound on the variance of the estimated parameters that reside in a constrained space.

For simplicity, assume that $\mathbf{a}$ is real and $\mathbf{W}$ is invertible. The total variance for $\hat{\mathbf{a}}$ in (5.2) is bounded as [49]

$$\sum_i \mathrm{Var}(\hat{\mathbf{a}}_i)_{\text{constrained}} \geq \mathrm{Tr}(\mathbf{A}[\mathbf{A}^T\mathbf{F}\mathbf{A}]^{-1}\mathbf{A}^T), \tag{5.3}$$

where $\mathbf{F}$ is the Fisher information matrix (see Appendix B for the derivation of $\mathbf{F}$), and $\mathbf{A}$ is selected to be those columns of $\mathbf{W}^{-1}$ corresponding to the support of $\mathbf{W}\mathbf{a}$. In comparison, the estimator in (3.5) has no sparsity constraint (as in state-of-the-art methods which perform quantification point-by-point); its performance can be characterized by the unconstrained CRB, which can be obtained by letting $\mathbf{A}$ be an identity matrix:

$$\sum_i \mathrm{Var}(\hat{\mathbf{a}}_i)_{\text{unconstrained}} \geq \mathrm{Tr}(\mathbf{F}^{-1}). \tag{5.4}$$

In the analysis above we have assumed $\boldsymbol{\theta}$ to be accurate. However, in practice, $\hat{\boldsymbol{\theta}}$ is a random variable dependent on noise realization, and cannot be exact the same with $\boldsymbol{\theta}$. In this case, we absorb the model discrepancy caused by $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ into an additional noise term. The resulting total noise thus has a larger variance. Therefore, we adjust corresponding CRB according to the new noise variance.

5.2.2 Validation

In this section, we perform corresponding Monte Carlo simulations to validate the constrained CRB and conventional CRB proposed above. We will firstly describe the simulation setup, and then show the results from Monte Carlo simulations.

Setup

A simulated dataset was designed for Monte Carlo simulations. We performed a wavelet transform to the concentration maps used in section 4.1, truncated out the principal non-zero entries, and then performed an inverse wavelet transform to create the true concentration maps used to generate the dataset, so that the ground truth of $\mathbf{a}$ is indeed transform sparse. The rest of the simulated dataset remains the same as the one used in section 4.1.

The Monte Carlo simulations were carried out as follows. White Gaussian noise was added to the simulated dataset, and then spectral estimation was performed (either using the proposed method or using a state-of-the-art method as baseline). The same procedure was performed for 100 times, and the additive noise was guaranteed to be independent from time to time. We can then obtain an estimate of the variance of $\hat{\mathbf{a}}_i$ for all i from the 100 $\hat{\mathbf{a}}$'s. To further analyze the characteristics of the proposed method at different sparsity levels, several sets of Monte Carlo simulations were carried out, each of which corresponded to a simulated dataset with concentration maps truncated using different thresholds in the transformed domain. Note sparsity level is defined as the proportion of non-zero terms of $\mathbf{Wa}$ in (5.2) in the following discussion. A higher sparsity level here indicates more non-zero entries.

Results and Discussion

To begin with, we assume that $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$. Figure 5.2 shows the consistency between CRB and Monte Carlo simulations for the case where QUEST was used as the spectral estimation method. It is as expected since given true $\boldsymbol{\theta}$, the estimation of $\mathbf{a}$ is simply a linear LS problem. Moreover, since QUEST is a point-by-point quantification method, the estimation variance of $\hat{\mathbf{a}}$ does

not depend on the transform sparsity level of $\mathbf{a}$, which is also reflected in Fig. 5.2. As for the proposed method, results are shown in Fig. 5.3, from which we can observe that if $\mathbf{Wa}$ is sparser, the constraint becomes more effective, and yields lower estimation variance.

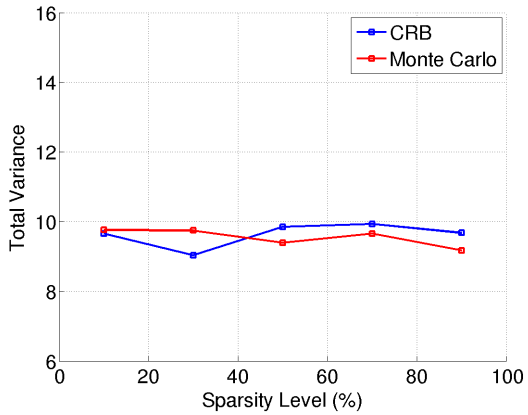


Figure 5.2: The total variance of $\hat{\mathbf{a}}$ using QUEST given true $\boldsymbol{\theta}$. We can see that the unconstrained CRB (blue) is rather consistent with Monte Carlo simulation (red).

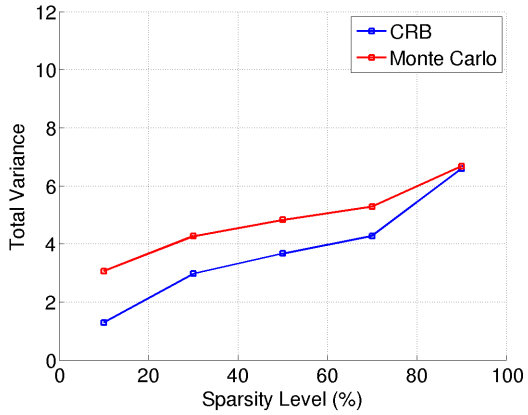


Figure 5.3: The total variance of $\hat{\mathbf{a}}$ using the proposed method given true $\boldsymbol{\theta}$. We can see that the estimation variance increases as $\mathbf{Wa}$ becomes more dense (higher sparsity level).

We define the ratio between the constrained CRB in (5.3) and the unconstrained CRB in (5.4) to describe the theoretical improvement when using spatial sparsity constraints for quantitation:

$$\mathrm{R} = \frac{\mathrm{Tr}(\mathbf{F}^{-1})}{\mathrm{Tr}(\mathbf{A}[\mathbf{A}^T\mathbf{F}\mathbf{A}]^{-1}\mathbf{A}^T)}. \tag{5.5}$$

Therefore, when true $\boldsymbol{\theta}$ is used, the improvement of incorporating spatial constraints is shown in Fig. 5.4. We can see that the improvement is larger for smaller sparsity level, and the improvement is roughly 1/sparsity level. Note for very small sparsity levels, the mismatch between CRB analysis and Monte Carlo simulation becomes increasingly large, because the basis pursuit by any practical algorithm might not be perfect. In practice, we need to

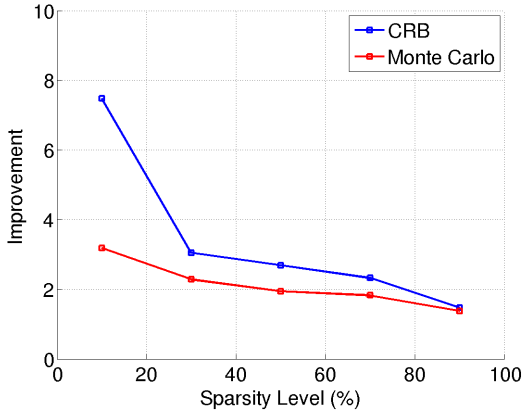
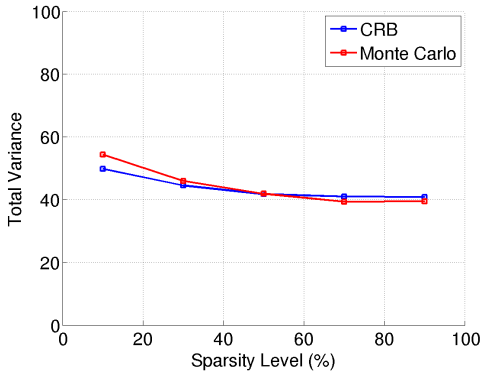
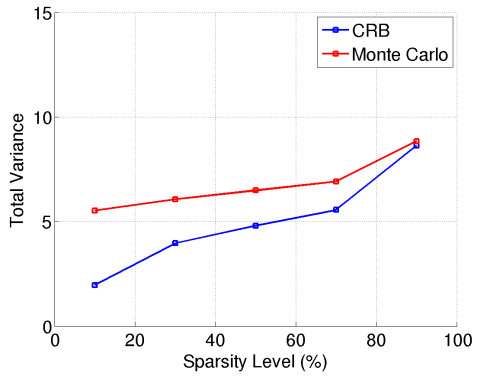


Figure 5.4: Blue line: the improvement R of the proposed method over QUEST given true $\boldsymbol{\theta}$. Red line: real improvement obtained from Monte Carlo simulation.

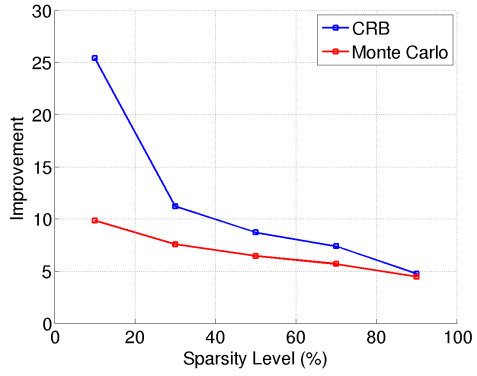
consider the inaccuracy of using $\hat{\boldsymbol{\theta}}$ instead of $\boldsymbol{\theta}$. Using the method described in the previous section, we carried out same comparisons as we did for using true $\boldsymbol{\theta}$, as shown in Fig. 5.5. As expected, the improvement obtained by incorporating sparsity constraint increases as the sparsity level decreases.



(a) The total variance of $\hat{\mathbf{a}}$ using QUEST given estimated $\hat{\boldsymbol{\theta}}$. Note the total variance is much larger than that using the true $\boldsymbol{\theta}$ (Fig. 5.2), because of the nonlinearity of the problem.



(b) The total variance of $\hat{\mathbf{a}}$ using the proposed method given estimated $\hat{\boldsymbol{\theta}}$.



(c) Blue line: the improvement R of the proposed method over QUEST given estimated $\hat{\boldsymbol{\theta}}$. Red line: real improvement obtained from Monte Carlo simulation.

Figure 5.5: Performance analysis when $\hat{\boldsymbol{\theta}}$ is not perfect.

CHAPTER 6

CONCLUSION

The practical use of spectral estimation for MRSI has long been limited by the inherently low sensitivity of NMR techniques and model nonlinearity. Although state-of-the-art methods have effectively imposed spectral prior information, various readily available spatial prior information is not incorporated in estimation. In this thesis, we present a novel method to address this problem. The proposed method jointly estimates the spectra from all voxels, and enforces spatial smoothness and transform sparsity of spectral parameters. Both simulated and experimental results showed the improved performance of the proposed method. We can summarize the conclusion as follows.

1. Spectral estimation for MRSI can be improved by incorporating spatial prior information, e.g., smoothness or transform sparsity.

2. The performance of the proposed method can be theoretically characterized by the constrained Cramér-Rao bound. The sparser the spectral parameters (in the transformed domain), the lower the estimation variances that can be expected.

Therefore, the proposed method should prove useful for spectral estimation in various MRSI studies.

APPENDIX A

THE PROOF OF THEOREM 3.1

In this chapter, we reiterate Theorem 3.1 as follows. Let

$$\mathbf{r}(\mathbf{c}, \boldsymbol{\alpha}) = [\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\alpha})\mathbf{c}, \mathbf{W}\boldsymbol{\alpha}],$$

and

$$\mathbf{r}_2(\boldsymbol{\alpha}) = \left[\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\alpha})\boldsymbol{\Phi}(\boldsymbol{\alpha})^{\dagger}\mathbf{y}, \mathbf{W}\boldsymbol{\alpha}\right],$$

where $\mathbf{c} \in \mathbb{R}^m$, $\boldsymbol{\alpha} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^k$, $\boldsymbol{\Phi} : \mathbb{R}^p \mapsto \mathbb{R}^{m \times k}$, and $\mathbf{W} \in \mathbb{R}^{m \times n}$. It is equivalent to show the "equivalence" between the following two optimization problems.

$$(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}}) = \arg\min_{\mathbf{c},\boldsymbol{\alpha}} \|\mathbf{r}(\mathbf{c}, \boldsymbol{\alpha})\|_2^2 \tag{P1}$$

$$\begin{cases} \hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{r}_2(\boldsymbol{\alpha})\|_2^2 \\ \hat{\mathbf{c}} = \boldsymbol{\Phi}(\hat{\boldsymbol{\alpha}})^{\dagger}\mathbf{y} \end{cases} \tag{P2}$$

Next we follow the original proof of variable projection (where $\mathbf{W} = 0$) provided by Golub and Pereyra (1973) [12], and make necessary modifications. Before proving Theorem 3.1, we need to firstly introduce the following lemma, which was also provided by [12].

Lemma A.1. *Let A^- be an $n \times m$ matrix function such that $AA^-A = A$ and $(AA^-)^T = AA^-$. Then*

$$DP_A = P_{A^\perp}(DA)A^- + (P_{A^\perp}(DA)A^-)^T, \tag{A.1}$$

where D represents a more general derivative in Banach spaces (the so called Fréchet derivative).

Proof of Lemma A.1. Since P_A is a projector onto the subspace of A, we

have

$$P_A A = A.$$

Therefore,

$$\begin{aligned} DA &= D(P_A A) = (DP_A)A + P_A(DA), \\ (DP_A)A &= DA - P_A(DA) = P_{A^\perp}(DA). \end{aligned} \tag{A.2}$$

We can see that the A^- here is a generalization or relaxation of the Moore-Penrose pseudo-inverse of A. It can be verified that AA^- is also a valid orthogonal projector onto the subspace of A. Moreover, because of the uniqueness of a projector, we have $P_A = AA^-$. Because projectors are idempotent, we have $P_A^2 = P_A$, and then

$$\begin{aligned} DP_A &= D(P_A^2) \\ &= (DP_A)P_A + ((DP_A)P_A)^T \\ &= (DP_A)AA^- + ((DP_A)AA^-)^T. \end{aligned} \tag{A.3}$$

Then by using (A.2), we have the result of (A.1). □

As $P_{A^\perp} = I - P_A$, we easily have $DP_{A^\perp} = -DP$, which is useful in the proof following. And we should also notice that if A^- is replaced by $A^\dagger$ in (A.1), the result remains unchanged.

Proof of Theorem 3.1.

1. As we only consider an open set Ω for $\boldsymbol{\alpha}$, the critical points are only those with zero gradient. Hence, to show that the critical points of (P1) and (P2) coincide, we need to find their corresponding gradients to begin with.

   Since $\|\mathbf{r}_2(\boldsymbol{\alpha})\|_2^2 = \|P_{\boldsymbol{\Phi}^\perp(\alpha)}\mathbf{y}\|_2^2 + \|\mathbf{W}\boldsymbol{\alpha}\|_2^2$, by the chain rule of gradient, we have

$$\begin{aligned} \frac{1}{2}\nabla\|\mathbf{r}_2\|_2^2 &= \mathbf{y}^T P_{\boldsymbol{\Phi}^\perp}(DP_{\boldsymbol{\Phi}^\perp})\mathbf{y} + \mathbf{W}^T\mathbf{W}\boldsymbol{\alpha} \\ &= -\mathbf{y}^T P_{\boldsymbol{\Phi}^\perp}[P_{\boldsymbol{\Phi}^\perp}(D\boldsymbol{\Phi})\boldsymbol{\Phi}^\dagger + (P_{\boldsymbol{\Phi}^\perp}(D\boldsymbol{\Phi})\boldsymbol{\Phi}^\dagger)^T]\mathbf{y} \\ &\quad + \mathbf{W}^T\mathbf{W}\boldsymbol{\alpha}. \end{aligned} \tag{A.4}$$

Because $P_{\mathbf{\Phi}^{\perp}(\alpha)}(\mathbf{\Phi}^{\dagger})^T = (\mathbf{\Phi}^{\dagger})^T - \mathbf{\Phi}\mathbf{\Phi}^{\dagger}(\mathbf{\Phi}^{\dagger})^T = 0$, the expression above can be simplified to

$$\frac{1}{2}\nabla\|\mathbf{r}_2(\boldsymbol{\alpha})\|_2^2 = -\mathbf{y}^T P_{\mathbf{\Phi}^{\perp}}(D\mathbf{\Phi})\mathbf{\Phi}^{\dagger}\mathbf{y} + \mathbf{W}^T\mathbf{W}\boldsymbol{\alpha}. \tag{A.5}$$

Similarly, for $\|\mathbf{r}(\mathbf{c}, \boldsymbol{\alpha})\|_2^2$, we have

$$\frac{1}{2}\nabla\|\mathbf{r}(\mathbf{c}, \boldsymbol{\alpha})\|_2^2 = \begin{bmatrix} -(\mathbf{y} - \mathbf{\Phi}\mathbf{c})^T(D\mathbf{\Phi})\mathbf{c} + \mathbf{W}^T\mathbf{W}\boldsymbol{\alpha} \\ -(\mathbf{y} - \mathbf{\Phi}\mathbf{c})^T\mathbf{\Phi} \end{bmatrix},$$

which comes from taking gradient with respect to $\boldsymbol{\alpha}$ and $\mathbf{c}$, respectively. If we let $\mathbf{c} = \mathbf{\Phi}^{\dagger}\mathbf{y}$ as in (P2), then

$$\frac{1}{2}\nabla\|\mathbf{r}(\mathbf{\Phi}^{\dagger}\mathbf{y}, \boldsymbol{\alpha})\|_2^2 = \begin{bmatrix} -(\mathbf{y} - \mathbf{\Phi}\mathbf{\Phi}^{\dagger}\mathbf{y})^T(D\mathbf{\Phi})\mathbf{c} + \mathbf{W}^T\mathbf{W}\boldsymbol{\alpha} \\ -(\mathbf{y} - \mathbf{\Phi}\mathbf{\Phi}^{\dagger}\mathbf{y})^T\mathbf{\Phi} \end{bmatrix}.$$

Given the fact that $\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^{\dagger} = P_{\mathbf{\Phi}^{\perp}}$, orthogonal projectors are symmetric (i.e., $P_{\mathbf{\Phi}^{\perp}} = P_{\mathbf{\Phi}^{\perp}}^T$), and $P_{\mathbf{\Phi}^{\perp}}\mathbf{\Phi} = 0$, the gradient above can be further simplified as

$$\begin{aligned} \frac{1}{2}\nabla\|\mathbf{r}(\mathbf{\Phi}^{\dagger}\mathbf{y}, \boldsymbol{\alpha})\|_2^2 &= \begin{bmatrix} -\mathbf{y}^T P_{\mathbf{\Phi}^{\perp}}(D\mathbf{\Phi})\mathbf{\Phi}^{\dagger}\mathbf{y} + \mathbf{W}^T\mathbf{W}\boldsymbol{\alpha} \\ \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2}\nabla\|\mathbf{r}_2(\boldsymbol{\alpha})\|_2^2 \\ \mathbf{0} \end{bmatrix}. \end{aligned}$$

Thus, if $\hat{\boldsymbol{\alpha}}$ is a critical point of (P2), and $\hat{\mathbf{c}} = \mathbf{\Phi}^{\dagger}(\hat{\boldsymbol{\alpha}})\mathbf{y}$, then $(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})$ is a critical point of (P1) as well.

2. Next we need to show that the global minimizer of these two problems coincide. Assume $\hat{\boldsymbol{\alpha}}$ is a global minimizer of (P2), and $\hat{\mathbf{c}}$ is calculated as $\mathbf{\Phi}(\boldsymbol{\alpha})^{\dagger}\mathbf{y}$. Then obviously, $\|\mathbf{r}_2(\hat{\boldsymbol{\alpha}})\|_2^2 = \|\mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})\|_2^2$.

   Suppose that there exists $(\mathbf{c}^*, \boldsymbol{\alpha}^*)$, $\boldsymbol{\alpha}^* \in \Omega$, s.t. $\|\mathbf{r}(\mathbf{c}^*, \boldsymbol{\alpha}^*)\|_2^2 < \|\mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})\|_2^2$. Because (P2) takes an extra step to minimize over $\mathbf{c}$, we have $\|\mathbf{r}_2(\boldsymbol{\alpha})\|_2^2 \leq \|\mathbf{r}(\mathbf{c}, \boldsymbol{\alpha})\|_2^2$, $\forall \boldsymbol{\alpha}, \mathbf{c}$. Then it follows that $\|\mathbf{r}_2(\boldsymbol{\alpha}^*)\|_2^2 \leq \|\mathbf{r}(\mathbf{c}^*, \boldsymbol{\alpha}^*)\|_2^2 < \|\mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})\|_2^2 = \|\mathbf{r}_2(\hat{\boldsymbol{\alpha}})\|_2^2$. But this is contradictory to the fact that $\hat{\boldsymbol{\alpha}}$ was a global minimizer of (P2). So the hypothesis is not valid, and $(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})$ is also a global minimizer of (P1).

Conversely, if $(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})$ is a global minimizer of (P1) and letting $\mathbf{c}^* = \boldsymbol{\Phi}^\dagger(\hat{\boldsymbol{\alpha}})\mathbf{y}$, then $\|\mathbf{r}_2(\hat{\boldsymbol{\alpha}})\|_2^2 = \|\mathbf{r}(\mathbf{c}^*, \hat{\boldsymbol{\alpha}})\|_2^2 \leq \|\mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})\|_2^2$. Since $\|\mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})\|_2^2$ is a global minimizer, it has to be equal sign in the inequality. Assume $\exists \boldsymbol{\alpha}' \neq \hat{\boldsymbol{\alpha}}$ such that $\|\mathbf{r}_2(\boldsymbol{\alpha}')\|_2^2 < \|\mathbf{r}_2(\hat{\boldsymbol{\alpha}})\|_2^2$. Let $\mathbf{c}' = \boldsymbol{\Phi}^\dagger(\boldsymbol{\alpha}')\mathbf{y}$. Then $\|\mathbf{r}_2(\boldsymbol{\alpha}')\|_2^2 = \|\mathbf{r}(\mathbf{c}', \boldsymbol{\alpha}')\|_2^2 < \|\mathbf{r}_2(\hat{\boldsymbol{\alpha}})\|_2^2 = \|\mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})\|_2^2$, which is contradictory to the fact that $(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})$ is a global minimizer of (P1). Therefore, $\hat{\boldsymbol{\alpha}}$ is a global minimizer of (P2) as well. Furthermore, if $\hat{\mathbf{c}}$ is unique, then it must be the same as $\mathbf{c}^* = \boldsymbol{\Phi}^\dagger(\hat{\boldsymbol{\alpha}})\mathbf{y}$, because $\mathbf{r}(\mathbf{c}^*, \hat{\boldsymbol{\alpha}}) = \mathbf{r}(\hat{\mathbf{c}}, \hat{\boldsymbol{\alpha}})$.

□

APPENDIX B

DERIVATION OF THE FISHER INFORMATION MATRIX

Since the parameters at different spatial locations are disjoint, the derivation of the Fisher information matrix (FIM) $\mathbf{F}$ in Eq. (5.3) can be decomposed into deriving the FIM for each spatial point. Therefore, we only discuss how to derive the FIM for a single voxel MRSI signal model in this chapter. The discretized signal model for a single voxel can be expressed as follows.

$$\begin{aligned} s[m] &= e^{i\phi_0} \sum_{n=1}^{N} a_n(T_{\mathrm{E}}) \varphi_{n,T_{\mathrm{E}}}[m] \psi_{n,d_n}[m] + \xi[m], \qquad \text{(B.1)} \\ m &= 0, 1, \ldots, M-1, \end{aligned}$$

where m is the time index, $\xi[m] \sim N(0, \sigma^2)$ is a complex Gaussian noise, ϕ_0 is a zero-order term phase, $a_n(T_{\mathrm{E}})$ is a real positive amplitude assumed to exponentially decay with respect to T_{E}:

$$a_n(T_{\mathrm{E}}) = c_n e^{-T_{\mathrm{E}}/T_{2,n}}, \qquad \text{(B.2)}$$

and $\varphi_{n,T_{\mathrm{E}}}[m]$ and $\psi_{n,d_n}[m]$ are metabolite basis functions and signal decay functions, respectively, defined as

$$\begin{aligned} \varphi_{n,T_{\mathrm{E}}}[m] &= \sum_{l=1}^{L_n} \alpha_{l,n}(T_{\mathrm{E}}) e^{-i\beta_{l,n}(T_{\mathrm{E}})} e^{-i2\pi f_{l,n}(T_{\mathrm{E}}) m\Delta t}, \qquad \text{(B.3)} \\ \psi_{n,d_n}[m] &= e^{-m\Delta t/d_n}. \qquad \text{(B.4)} \end{aligned}$$

In the above formulations, $T_{2,n}$ is a metabolite-dependent relaxation constant, d_n is a real lineshape parameter and Δt is the sampling time. $\alpha_{l,n}(T_{\mathrm{E}})$, $\beta_{l,n}(T_{\mathrm{E}})$ and $f_{l,n}(T_{\mathrm{E}})$ are relative amplitude, phase and frequency of the l-th resonance of the n-th metabolite which can all be determined from quantum mechanical simulations. The parameter vector that we are going to estimate

is

$$\boldsymbol{\theta} = [a_1, \ldots, a_N, d_1, \ldots, d_N, \phi_0]^T. \tag{B.5}$$

B.1 Entrywise Derivation of CRB

Since $\xi(t)$ is a complex white Gaussian noise, the likelihood function of $s[m], m = 0, \ldots, M-1$ is

$$\begin{aligned} L(s[m]) &= \frac{1}{(\pi\sigma^2)^M} e^{-\frac{|\xi[m]|^2}{\sigma^2}}, \\ \Rightarrow \ln L(s[m]) &= \text{const} - \frac{1}{\sigma^2} \sum_{m=0}^{M-1} |\xi[m]|^2, \end{aligned} \tag{B.6}$$

where $\xi[m] \sim N(0, \sigma^2)$. Using the result that $\frac{\partial \|z\|^2}{\partial \alpha} = z(\frac{\partial z^*}{\partial \alpha}) + z^*(\frac{\partial z}{\partial \alpha}) = 2Re\{z(\frac{\partial z^*}{\partial \alpha})\}$ where $z \in \mathbb{C}$, $\alpha \in \mathbb{R}$, we have

$$\frac{\partial \ln L}{\partial \theta_k} = -\frac{1}{\sigma^2} \sum_{m=0}^{M-1} \left\{ \xi[m] \frac{\partial \xi^*[m]}{\partial \theta_k} + \xi^*[m] \frac{\partial \xi[m]}{\partial \theta_k} \right\}, \tag{B.7}$$

where

$$\begin{aligned} \xi^*[m] &= s^*[m] - e^{-i\phi_0} \sum_{n=1}^{N} a_n(T_{\mathrm{E}}) \varphi^*_{n,T_{\mathrm{E}}}[m] \psi^*_{n,d_n}[m], \\ (\frac{\partial \xi[m]}{\partial a_k})^* &= \frac{\partial \xi^*[m]}{\partial a_k} = -e^{-i\phi_0} \varphi^*_{k,T_{\mathrm{E}}}[m] \psi^*_{k,d_k}[m], \qquad \text{(B.8)} \\ (\frac{\partial \xi[m]}{\partial d_k})^* &= \frac{\partial \xi^*[m]}{\partial d_k} = -e^{-i\phi_0} a_k(T_{\mathrm{E}}) \varphi^*_{k,T_{\mathrm{E}}}[m] \frac{\partial \psi^*_{k,d_k}[m]}{\partial d_k}, \qquad \text{(B.9)} \\ (\frac{\partial \xi[m]}{\partial \phi_0})^* &= \frac{\partial \xi^*[m]}{\partial \phi_0} = ie^{-i\phi_0} \sum_{n=1}^{N} a_n(T_{\mathrm{E}}) \varphi^*_{n,T_{\mathrm{E}}}[m] \psi^*_{n,d_n}[m], \qquad \text{(B.10)} \\ k &= 1, 2, \ldots, N. \end{aligned}$$

By Eq. (2.8), we have

$$F_{p,q}(\theta) = E_\xi \left[\left(\frac{\partial \ln L}{\partial \theta_p} \right) \left(\frac{\partial \ln L}{\partial \theta_q} \right) \right]. \tag{B.11}$$

In calculating that, we will need the expectation of $\xi^*[m_1]\xi[m_2]$ and $\xi[m_1]\xi[m_2]$, $\forall m_1, m_2 = 1, \ldots, M$. For white Gaussian noise, it is easy to get

$$\begin{aligned} E_\xi\{\xi^*[m_1]\xi[m_2]\} &= \begin{cases} \sigma^2, & \text{for } m_1 = m_2 \\ 0, & \text{for } m_1 \neq m_2 \end{cases} \\ E_\xi\{\xi[m_1]\xi[m_2]\} &= 0, \text{for } m_1 \neq m_2. \end{aligned} \tag{B.12}$$

To handle $\xi[m]\xi[m]$, we firstly decompose it into real variables:

$$\xi[m] = \xi_r[m] + i\xi_i[m], \tag{B.13}$$

where $\xi_r[m], \xi_i[m] \sim N(0, \frac{\sigma^2}{2})$ and are independent with each other. Then

$$E_\xi\{\xi[m]\xi[m]\} = E_\xi\{\xi_r^2 + 2i\xi_r\xi_i - \xi_i^2\} = E_\xi\{\xi_r^2 - \xi_i^2\} = 0.$$

Therefore,

$$E_\xi\{\xi[m_1]\xi[m_2]\} = 0, \forall m_1, m_2 = 1, \ldots, M. \tag{B.14}$$

Taking (B.12) and (B.14) into consideration, the entries of the Fisher Information matrix is

$$\begin{aligned} F_{p,q}(\theta) &= \frac{1}{\sigma^4} E_\xi \left\{ \sum_{m_1=0}^{M-1} \left[\xi[m_1] \frac{\partial \xi^*[m_1]}{\partial \theta_p} + \xi^*[m_1] \frac{\partial \xi[m_1]}{\partial \theta_p} \right] \right. \\ &\quad \left. \sum_{m_2=0}^{M-1} \left[\xi[m_2] \frac{\partial \xi^*[m_2]}{\partial \theta_q} + \xi^*[m_2] \frac{\partial \xi[m_2]}{\partial \theta_q} \right] \right\} \\ &= \frac{1}{\sigma^4} E_\xi \left\{ \sum_{m=0}^{M-1} \left[\xi[m]\xi^*[m] \frac{\partial \xi^*[m]}{\partial \theta_p} \frac{\partial \xi[m]}{\partial \theta_q} + \xi^*[m]\xi[m] \frac{\partial \xi[m]}{\partial \theta_p} \frac{\partial \xi^*[m]}{\partial \theta_q} \right] \right\} \\ &= \frac{1}{\sigma^2} \sum_{m=0}^{M-1} \left[\frac{\partial \xi^*[m]}{\partial \theta_p} \frac{\partial \xi[m]}{\partial \theta_q} + \frac{\partial \xi[m]}{\partial \theta_p} \frac{\partial \xi^*[m]}{\partial \theta_q} \right] \\ &= \frac{2}{\sigma^2} \sum_{m=0}^{M-1} \mathrm{Re} \left\{ \frac{\partial \xi^*[m]}{\partial \theta_p} \frac{\partial \xi[m]}{\partial \theta_q} \right\}. \end{aligned} \tag{B.15}$$

(B.16)

Substituting (B.8), (B.9) and (B.10) into (B.15), we get for $\forall p, q = 1, \ldots, N$

$$F_{a_p,a_q}(\theta) = \frac{2}{\sigma^2}\sum_{m=0}^{M-1}\mathrm{Re}\left\{\varphi^*_{p,T_\mathrm{E}}[m]\psi^*_{p,d_p}[m]\varphi_{q,T_\mathrm{E}}[m]\psi_{q,d_q}[m]\right\}, \tag{B.17}$$

$$F_{d_p,d_q}(\theta) = \frac{2}{\sigma^2}\sum_{m=0}^{M-1}\mathrm{Re}\left\{a_p a_q\varphi^*_{p,T_\mathrm{E}}[m]\frac{\partial\psi^*_{p,d_p}[m]}{\partial d_p}\varphi_{q,T_\mathrm{E}}[m]\frac{\partial\psi_{q,d_q}[m]}{\partial d_q}\right\}, \tag{B.18}$$

$$F_{\phi_0,\phi_0}(\theta) = \frac{2}{\sigma^2}\sum_{m=0}^{M-1}\mathrm{Re}\left\{\sum_{n_1=1}^{N}\sum_{n_2=1}^{N}a_{n_1}a_{n_2}\varphi^*_{n_1,T_\mathrm{E}}[m]\psi^*_{n_1,d_{n_1}}[m]\varphi_{n_2,T_\mathrm{E}}[m]\psi_{n_2,d_{n_2}}[m]\right\}, \tag{B.19}$$

$$F_{a_p,d_q}(\theta) = \frac{2}{\sigma^2}\sum_{m=0}^{M-1}\mathrm{Re}\left\{a_q\varphi^*_{p,T_\mathrm{E}}[m]\psi^*_{p,d_p}[m]\varphi_{q,T_\mathrm{E}}[m]\frac{\partial\psi_{q,d_q}[m]}{\partial d_q}\right\}, \tag{B.20}$$

$$F_{a_p,\phi_0}(\theta) = -\frac{2}{\sigma^2}\sum_{m=0}^{M-1}\mathrm{Im}\left\{\varphi^*_{p,T_\mathrm{E}}[m]\psi^*_{p,d_p}[m]\sum_{n=1}^{N}a_n\varphi_{n,T_\mathrm{E}}[m]\psi_{n,d_n}[m]\right\}, \tag{B.21}$$

$$F_{d_p,\phi_0}(\theta) = -\frac{2}{\sigma^2}\sum_{m=0}^{M-1}\mathrm{Im}\left\{a_p\varphi^*_{p,T_\mathrm{E}}[m]\frac{\partial\psi^*_{p,d_p}[m]}{\partial d_p}\sum_{n=1}^{N}a_n\varphi_{n,T_\mathrm{E}}[m]\psi_{n,d_n}[m]\right\}. \tag{B.22}$$

Here we have got the entries of $F_{p,q}(\theta)$ above its diagonal, and the rest of the entries are the conjugate transpose of the upper triangular part.

$$\mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} F_{a,a} & F_{a,d} & F_{a,\phi_0} \\ F^H_{a,d} & F_{d,d} & F_{d,\phi_0} \\ F^H_{a,\phi_0} & F^H_{d,\phi_0} & F_{\phi_0,\phi_0} \end{pmatrix}. \tag{B.23}$$

B.2 Matrix Derivation of CRB

In deriving the CRB entrywisely, we see the formulations are much complicated due to a large number of summations. Here we show an alternative way of calculating $\mathbf{F}$ which takes advantage of the matrix derivatives. Rewrite the signal model in matrix form.

$$\mathbf{s} = e^{i\phi_0}\mathbf{Z}\mathbf{a} + \boldsymbol{\xi}, \tag{B.24}$$

where $\mathbf{Z}_{m,n} = \varphi_{n,T_{\mathrm{E}}}[m]\psi_{d,d_n}[m],\ m = 0, \ldots, M-1,\ n = 1, \ldots, N$ and

$$\mathbf{s} = \begin{pmatrix} s[0] \\ s[1] \\ \vdots \\ s[M-1] \end{pmatrix}, \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{pmatrix}, \boldsymbol{\xi} = \begin{pmatrix} \xi[0] \\ \xi[1] \\ \vdots \\ \xi[M-1] \end{pmatrix}. \tag{B.25}$$

Then the likelihood function is

$$\ln L(\mathbf{s}) = \text{const} - \frac{1}{\sigma^2}\|\boldsymbol{\xi}\|^2. \tag{B.26}$$

$$\Rightarrow \nabla_{\boldsymbol{\theta}} \ln L = -\frac{1}{\sigma^2}(\mathbf{J}^H\boldsymbol{\xi} + (\mathbf{J}^H\boldsymbol{\xi})^*), \tag{B.27}$$

where $\mathbf{J}$ is the Jacobian matrix of $\boldsymbol{\xi}$ over $\boldsymbol{\theta}$.

$$\begin{aligned} \mathbf{J} &= \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\theta}} \\ &= \left[\begin{array}{ccc} \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{a}} & \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{d}} & \frac{\partial \boldsymbol{\xi}}{\partial \phi_0} \end{array} \right] \\ &= \left[\begin{array}{ccc} -e^{i\phi_0}\mathbf{Z} & -e^{i\phi_0}\mathbf{D}\mathbf{A} & -ie^{i\phi_0}\mathbf{Z}\mathbf{a} \end{array} \right], \end{aligned} \tag{B.28}$$

where

$$\mathbf{D} = \left[\begin{array}{ccc} \frac{\partial Z_1}{\partial d_1} & \cdots & \frac{\partial Z_N}{\partial d_N} \end{array} \right], \tag{B.29}$$

$$\mathbf{A} = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_N \end{pmatrix}. \tag{B.30}$$

By (2.8) and (B.27), we have

$$\begin{aligned} \mathbf{F} &= \mathbb{E}\left[(\nabla_{\boldsymbol{\theta}} \ln L) (\nabla_{\boldsymbol{\theta}} \ln L)^H \right] \\ &= \frac{1}{\sigma^4}\mathbb{E}\left[(\mathbf{J}^H\boldsymbol{\xi} + (\mathbf{J}^H\boldsymbol{\xi})^*)(\boldsymbol{\xi}^H\mathbf{J} + (\boldsymbol{\xi}^H\mathbf{J})^*) \right] \\ &= \frac{2}{\sigma^2}\mathrm{Re}\{\mathbf{J}^H\mathbf{J}\}. \end{aligned} \tag{B.31}$$

Using (B.28), we have

$$F_{a,a}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\mathrm{Re}\{\mathbf{Z}^H\mathbf{Z}\}, \tag{B.32}$$

$$F_{d,d}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\mathrm{Re}\{\mathbf{A}^H\mathbf{D}^H\mathbf{D}\mathbf{A}\}, \tag{B.33}$$

$$F_{\phi_0,\phi_0}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\mathrm{Re}\{\mathbf{a}^H\mathbf{Z}^H\mathbf{Z}\mathbf{a}\}, \tag{B.34}$$

$$F_{a,d}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\mathrm{Re}\{\mathbf{Z}^H\mathbf{D}\mathbf{A}\}, \tag{B.35}$$

$$F_{a,\phi_0}(\boldsymbol{\theta}) = -\frac{2}{\sigma^2}\mathrm{Im}\{\mathbf{Z}^H\mathbf{Z}\mathbf{a}\}, \tag{B.36}$$

$$F_{d,\phi_0}(\boldsymbol{\theta}) = -\frac{2}{\sigma^2}\mathrm{Im}\{\mathbf{A}^H\mathbf{D}^H\mathbf{Z}\mathbf{a}\}. \tag{B.37}$$

Finally

$$\mathbf{F}(\boldsymbol{\theta}) = \begin{pmatrix} F_{a,a} & F_{a,d} & F_{a,\phi_0} \\ F_{a,d}^H & F_{d,d} & F_{d,\phi_0} \\ F_{a,\phi_0}^H & F_{d,\phi_0}^H & F_{\phi_0,\phi_0} \end{pmatrix}. \tag{B.38}$$

REFERENCES

[1] R. A. De Graaf, *In Vivo NMR Spectroscopy: Principles and Techniques*. John Wiley & Sons, 2013.

[2] R. A. Moats, T. Ernst, T. K. Shonk, and B. D. Ross, "Abnormal cerebral metabolite concentrations in patients with probable Alzheimer disease," *Magnetic resonance in medicine*, vol. 32, no. 1, pp. 110–115, 1994.

[3] D. P. Soares and M. Law, "Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications," *Clinical radiology*, vol. 64, no. 1, pp. 12–21, 2009.

[4] J. van der Veen, R. de Beer, P. Luyten, and D. van Ormondt, "Accurate quantification of in vivo 31P NMR signals using the variable projection method and prior knowledge," *Magnetic Resonance in Medicine*, vol. 6, no. 1, pp. 92–98, 1988.

[5] A. A. de Graaf and W. M. M. J. Bovée, "Improved quantification of in vivo 1H NMR spectra by optimization of signal acquisition and processing and by incorporation of prior knowledge into the spectral fitting," *Magnetic Resonance in Medicine*, vol. 15, pp. 305–319, 1990.

[6] S. W. Provencher, "Estimation of metabolite concentrations from localized in vivo proton NMR spectra," *Magnetic Resonance in Medicine*, vol. 30, no. 6, pp. 672–679, 1993.

[7] K. Young, V. Govindaraju, B. J. Soher, and A. A. Maudsley, "Automated spectral analysis i: Formation of a priori information by spectral simulation," *Magnetic Resonance in Medicine*, vol. 40, no. 6, pp. 812–815, 1998.

[8] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly, "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR in Biomedicine*, vol. 18, pp. 1–13, 2005.

[9] J.-B. Poullet, D. M. Sima, A. W. Simonetti, B. de Neuter, L. Vanhamme, P. Lemmerling, and S. van Huffel, "An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES," *NMR in Biomedicine*, vol. 20, no. 5, pp. 493–504, 2007.

[10] Q. Ning, C. Ma, and Z.-P. Liang, "Spectral estimation for magnetic resonance spectroscopic imaging with spatial sparsity constraints," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*. IEEE, 2015, pp. 1482–1485.

[11] Y. Bresler, S. Basu, and C. Couvreur, "Vector spaces and least squares methods for signal processing," February 2009, ECE513 lecture notes, UIUC.

[12] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal of Numerical Analysis*, vol. 10, no. 2, April 1973.

[13] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer, 2008, ch. 4.4, pp. 131–150.

[14] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *Information Theory, IEEE Transactions on*, vol. 36, no. 6, pp. 1285–1301, 1990.

[15] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *The Annals of Mathematical Statistics*, pp. 581–586, 1951.

[16] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3384–3389, 2010.

[17] G. Tang and A. Nehorai, "Lower bounds on the mean-squared error of low-rank matrix reconstruction," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4559–4571, 2011.

[18] W. G. Proctor and F. C. Yu, "The dependence of a nuclear magnetic resonance frequency upon chemical compound," *Physical Review*, vol. 77, no. 5, p. 717, 1950.

[19] W. C. Dickinson, "Dependence of the F19 nuclear resonance position on chemical compound," *Physical Review*, vol. 77, no. 5, p. 736, 1950.

[20] R. R. Ernst and W. A. Anderson, "Applications of Fourier transform spectroscopy to magnetic resonance," *Rev. Sci. Instrum.*, vol. 37, no. 1, pp. 93–102, 1966.

[21] P. C. Lauterbur, "Image formation by induced local interactions: examples employing nuclear magnetic resonance," *Nature*, vol. 242, no. 5394, pp. 190–191, 1973.

[22] H. Barkhuijsen, 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," *Journal of Magnetic Resonance*, vol. 61, no. 3, pp. 465–481, 1985.

[23] H. Barkhuijsen, R. de Beer, and D. van Ormondt, "Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals," *Journal of Magnetic Resonance*, pp. 553–557, 1987.

[24] W. W. F. Pijnappel, A. van den Boogaart, R. de Beer, and D. van Ormondt, "SVD-based quantification of magnetic resonance signals," *Journal of Magnetic Resonance*, vol. 97, no. 1, pp. 122–134, 1992.

[25] H. Chen, S. van Huffel, D. van Ormondt, and R. de Beer, "Parameter estimation with prior knowledge of known signal poles for the quantification of NMR spectroscopy data in the time domain," *Journal of Magnetic Resonance, Series A*, vol. 119, no. 2, pp. 225–234, 1996.

[26] P. Stoica, Y. Selén, N. Sandgren, and S. van Huffel, "Using prior knowledge in SVD-based parameter estimation for magnetic resonance spectroscopy-the ATP example," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 9, pp. 1568–1578, 2004.

[27] 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.

[28] R. Ogg, R. Kingsley, and J. Taylor, "WET, a T1- and B1-insensitive water-suppression method for in vivo localized 1H NMR spectroscopy," *J Magn Reson, B*, vol. 104, no. 1, pp. 1 – 10, 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S106418668471048X

[29] P. Le Roux, R. Gilles, G. McKinnon, and P. Carlier, "Optimized outer volume suppression for single-shot fast spin-echo cardiac imaging," *J Magn Reson Imaging*, vol. 8, pp. 1022–1032, 1998.

[30] J. H. Duyn, J. Gillen, G. Sobering, P. C. van Zijl, and C. T. Moonen, "Multisection proton MR spectroscopic imaging of the brain," *Radiology*, vol. 188, no. 1, pp. 277–282, 1993.

[31] C. Haupt, N. Schuff, M. Weiner, and A. Maudsley, "Removal of lipid artifacts in ^{1}h spectroscopic imaging by data extrapolation," *Magn Reson Med*, vol. 35, pp. 678–687, 1996.

[32] C. Ma, F. Lam, C. L. Johnson, and Z.-P. Liang, "Removal of nuisance signals from limited and sparse ^{1}H MRSI data using a union-of-subspaces model," *Magn Reson Med*, pp. 0–0, 2015. [Online]. Available: http://dx.doi.org/10.1002/mrm.25635

[33] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc IEEE Int Symp on Biomed Imag*, Arlington, VA, USA, 2007, pp. 988 – 991.

[34] J. P. Haldar and Z.-P. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *Proc IEEE Int Symp on Biomed Imag*, Rotterdam, Netherlands, 2010, pp. 716 – 719.

[35] K. L. Behar, D. L. Rothman, D. D. Spencer, and O. A. Petroff, "Analysis of macromolecule resonances in 1H NMR spectra of human brain," *Magnetic Resonance in Medicine*, vol. 32, no. 3, pp. 294–302, 1994.

[36] S. Posse, R. Otazo, S. R. Dager, and J. Alger, "MR spectroscopic imaging: principles and recent advances," *Journal of Magnetic Resonance Imaging*, vol. 37, no. 6, pp. 1301–1325, 2013.

[37] D. Graveron-Demilly, "Quantification in magnetic resonance spectroscopy based on semi-parametric approaches," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 27, no. 2, pp. 113–130, 2014.

[38] B. J. Soher, K. Young, and A. A. Maudsley, "Representation of strong baseline contributions in 1H MR spectra," *Magnetic Resonance in Medicine*, vol. 45, no. 6, pp. 966–972, 2001.

[39] Q. Ning, C. Ma, and Z.-P. Liang, "Towards short-TE MR spectroscopic imaging: Spectral decomposition and removal of baseline signals," in *Conference Abstract of IEEE EMBC, Chicago, USA*, 2014.

[40] D. Spielman, P. Webb, and A. Macovski, "Water referencing for spectroscopic imaging," *Magnetic Resonance in Medicine*, vol. 12, no. 1, pp. 38–49, 1989.

[41] D. C. Noll, J. Fessler, B. P. Sutton et al., "Conjugate phase MRI reconstruction with spatially variant sample density correction," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 3, pp. 325–336, 2005.

[42] X. Peng, H. Nguyen, J. Haldar, D. Hernando, X.-P. Wang, and Z.-P. Liang, "Correction of field inhomogeneity effects on limited k-space MRSI data using anatomical constraints," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 883–886.

[43] B. J. Soher, K. Young, A. Bernstein, Z. Aygula, and A. A. Maudsley, "GAVA: spectral simulation for in vivo MRS applications," *Journal of Magnetic Resonance*, vol. 185, no. 2, pp. 291–299, 2007.

[44] J. P. Haldar, D. Hernando, S.-K. Song, and Z.-P. Liang, "Anatomically constrained reconstruction from noisy data," *Magnetic Resonance in Medicine*, vol. 59, no. 4, pp. 810–818, 2008.

[45] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.

[46] J. Nocedal and S. Wright, "Numerical optimization, series in operations research and financial engineering," *Springer, New York, USA*, 2006.

[47] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[48] W. Guo, J. Qin, and W. Yin, "A new detail-preserving regularization scheme," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1309–1334, 2014.

[49] F. Lam, C. Ma, and Z.-P. Liang, "Performance analysis of denoising with low-rank and sparsity constraints," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on.* IEEE, 2013, pp. 1223–1226.