© 2014 Bo Zhao

FAST MRI WITH SPARSE SAMPLING: MODELS, ALGORITHMS, AND APPLICATIONS

BY

BO ZHAO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Zhi-Pei Liang, Chair
Professor Minh Do
Professor Farzad Kamalabadi
Associate Professor Brad Sutton

ABSTRACT

Conventional magnetic resonance imaging (MRI) methods are based on the Shannon-Nyquist sampling theorem. The number of required Nyquist samples grows exponentially with respect to the underlying physical dimension of the imaging problem, resulting in significant difficulty of achieving high resolution for higher-dimensional imaging applications. This research addresses such a problem from a sparse sampling perspective. We have proposed novel constrained imaging approaches, including imaging models and reconstruction algorithms, to enable high-quality reconstruction from highly undersampled data. The utility of the proposed techniques is demonstrated in two challenging higher-dimensional MRI applications, i.e., dynamic MRI and MR parameter mapping.

First, we propose a novel constrained image model, the joint low-rank and sparse model, to enable dynamic image reconstruction from highly undersampled $(\mathbf{k}, t)$-space data. Low-rank and sparse models are two low-dimensional signal structures, each of which parsimoniously models dynamic imaging data. Here, we integrate the two models into a single mathematical formulation. With pre-estimation of the temporal subspace for the low-rank model, the proposed formulation results in a convex optimization problem, for which we develop a globally convergent algorithm based on the half-quadratic regularization with continuation procedure. We systematically demonstrate the complementary advantages of incorporating both low-rank and sparsity models into the proposed formulation with a real-time cardiac imaging application.

Second, we extend the joint low-rank and sparsity model to accelerate an important class of quantitative MRI problems, i.e., MR parameter mapping. We specifically tailor the low-rank and sparse model to capture different signal/image characteristics in parameter mapping applications: the low-rank constraint captures the correlation among relaxation signals at

different spatial locations, whereas the sparsity constraint captures the edge structure shared by a sequence of co-registered, contrast-weighted MR images. We demonstrate the superior performance of the proposed method over two state-of-the-art methods that are based on a single low-rank or sparse model with in-vivo data.

Third, we address the MR parameter mapping problem from a different perspective with a novel model-based parameter estimation approach. Specifically, we propose a mathematical formulation that integrates an explicit parametric signal model from MR physics with sparsity constraints on the model parameters. It enables direct estimation of the parameters of interest from highly undersampled, noisy **k**-space data. We present an efficient greedy-pursuit algorithm to solve the resulting optimization problem. We also derive estimation-theoretic bounds to analyze the benefits of sparsity constraints and benchmark the performance of the proposed method. Both the theoretical properties and empirical performance of the proposed method are illustrated in a T_2 mapping application example.

To my family, for their endless love and support.

ACKNOWLEDGMENTS

I gratefully acknowledge people who have given me guidance and help throughout my graduate studies, without which this dissertation would not have been possible. First of all, I am deeply indebted to my advisor, Professor Zhi-Pei Liang, who gave me the valuable opportunity to study at the University of Illinois at Urbana-Champaign. I greatly appreciate his constant support, guidance and patience throughout my thesis research. His enthusiasm, philosophy, and vision have fundamentally shaped my way of conducting scientific research, and will significantly impact my later career. Furthermore, I would like to thank my other Ph.D. thesis committee members, Professors Minh Do, Farzard Kamalabadi, and Brad Sutton, for their value comments on my thesis research.

Special thanks also go to my former labmate Dr. Justin Haldar, who gave me tremendous guidance and inspiration during the early days of my graduate studies. Furthermore, I would like to particularly thank my labmate and friend, Fan Lam, for stimulating discussion, fruitful collaboration, and sincere friendship, since we joined the group together in 2008. I am also grateful to my labmates Anthony Christodoulou and Maojing Fu, and the former Beckman Institute research scientist Dr. Wenmiao Lu for research collaboration. Outside the University of Illinois, I want to express my great gratitude to Professor Chien Ho and Dr. Kevin Hitchens at Carnegie Mellon University for the research collaboration and help.

I am very lucky to have worked in a very nice research group, in which many current and former labmates, as well as visitors, have offered help and support to me. Especially, I would like to thank Chao Ma, Diego Hernando, Hien Nguyen, Cornelius Brinegar, Bryan Clifford, Ding Liu, Mohammed Sheikh, Qiang Ning, Xiaobo Qu, Xi Peng, Aiqi Sun, and Zhenghua Wu.

Finally, my deepest thanks go to my family for their endless love, encouragement, and

support. This dissertation is dedicated to them.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS ix

CHAPTER 1 INTRODUCTION 1
1.1 Problem Statement 1
1.2 Motivation 2
1.3 Main Results 3
1.4 Organization of the Dissertation 4

CHAPTER 2 BACKGROUND 6
2.1 Principles of MRI 6
2.1.1 Signal Generation and Detection 6
2.1.2 Spatial Information Encoding 9
2.1.3 Image Formation 11
2.2 Compressive Sensing 13
2.2.1 Sparsity and Compressibility 14
2.2.2 Sparse Reconstruction Algorithms 15
2.2.3 Performance Guarantees 19
2.2.4 Applications to MRI 21
2.3 Statistical Estimation Theory 23
2.3.1 Statistical Estimators 23
2.3.2 Cramér-Rao Lower Bound 25
2.3.3 Properties of Maximum Likelihood Estimators 26

CHAPTER 3 DYNAMIC IMAGE RECONSTRUCTION WITH JOINT LOW-RANK AND SPARSITY CONSTRAINTS 28
3.1 Introduction 28
3.2 Imaging Model 30
3.2.1 Data Model 30
3.2.2 Proposed Formulation 30
3.3 Solution Algorithm and Characteristics 34
3.3.1 Solution Algorithm 34
3.3.2 Convergence Analysis 38
3.4 Application to Real-Time Cardiac Imaging 41
3.4.1 Simulation Results 41
3.4.2 In Vivo Results 52

3.5 Discussion . . . 53
3.6 Summary . . . 57

CHAPTER 4 ACCELERATED MR PARAMETER MAPPING WITH JOINT LOW-RANK AND SPARSITY CONSTRAINTS . . . 58
4.1 Introduction . . . 58
4.2 Imaging Model . . . 60
4.2.1 Data Model . . . 60
4.2.2 Proposed Formulation . . . 61
4.3 Solution Algorithm . . . 63
4.4 Parameter Estimation . . . 66
4.5 In Vivo Results . . . 66
4.5.1 T_2 Mapping Application . . . 66
4.5.2 T_1 Mapping Application . . . 72
4.6 Discussion . . . 75
4.7 Summary . . . 79

CHAPTER 5 MODEL-BASED MR PARAMETER MAPPING WITH SPARSITY CONSTRAINTS . . . 80
5.1 Introduction . . . 80
5.2 Imaging Model . . . 80
5.2.1 Parametric Physical Model . . . 81
5.2.2 Sparsity Constraint . . . 82
5.3 Solution Algorithm . . . 83
5.4 Estimation-Theoretic Characterizations . . . 86
5.4.1 Cramér-Rao Bound . . . 87
5.4.2 Constrained Cramér-Rao Bound . . . 88
5.5 Results . . . 90
5.5.1 Without model-mismatch . . . 92
5.5.2 With model-mismatch . . . 99
5.6 Discussion . . . 101
5.7 Summary . . . 106
5.8 Miscellaneous Derivations . . . 106
5.8.1 Calculation of Gradients . . . 106
5.8.2 Calculation of Fisher Information Matrix . . . 108
5.8.3 Asymptotics of Oracle Estimator . . . 110

CHAPTER 6 CONCLUSIONS AND FUTURE WORK . . . 113
6.1 Conclusions . . . 113
6.2 Future Work . . . 114

REFERENCES . . . 116

LIST OF ABBREVIATIONS

ADMM	Alternating Direction Method of Multipliers
CoSaMP	Compressive Sampling Matching Pursuit
CRB	Cramér-Rao Bound
CS	Compressive Sensing
ECG	Electrocardiography
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
FOV	Field of View
GraSP	Gradient Support Pursuit
L-BFGS	Limited Memory Broyden-Flecher-Goldfarb-Shanno
MAP	Maximum A Posteriori
ML	Maximum Likelihood
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NE	Normalized Error
NLS	Nonlinear Least Squares
NMR	Nuclear Magnetic Resonance
NRMSE	Normalized Root Mean Squared Error
RF	Radiofrequency

PML	Penalized Maximum Likelihood
RPCA	Robust Principal Component Analysis
PS	Partial Separability
rNE	Region-Of-Interest Normalized Error
rNRMSE	Region-Of-Interest Normalized Root Mean Squared Error
ROI	Region of Interest
SENSE	Sensitivity Encoding
SNR	Signal-to-Noise Ratio
SP	Subspace Pursuit
SURE	Stein's Unbiased Risk Estimate
VARPRO	Variable Projection

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Image formation for MRI involves reconstruction of a desired image function $\rho(\mathbf{x})$ from samples collected in the spatial Fourier domain (often referred to as $\mathbf{k}$-space). The associated imaging equation can be written as [1]:

$$s(\mathbf{k}) = \int \rho(\mathbf{x}) \exp(-i2\pi\mathbf{k}\cdot\mathbf{x}) d\mathbf{x} + n(\mathbf{k}), \tag{1.1}$$

where $s(\mathbf{k})$ denotes the $\mathbf{k}$-space measured data, and $n(\mathbf{k})$ denotes the measurement noise.

Conventional MRI reconstructions are based on direct Fourier inversion of $\mathbf{k}$-space data that are acquired following the Shannon-Nyquist sampling theorem. But, in higher-dimensional imaging applications (e.g., spatiotemporal imaging or spatiospectral imaging), the number of required Nyquist samples grows exponentially with the physical dimension of the problem (the so-called "curse of dimensionality" for sampling [2]). Such an issue often hampers the practical utility of MRI to probing many biological processes (e.g., real-time imaging of a beating heart or natural speech production).

Developing fast MRI techniques has been a core research area since the inception of the field. In this research, we address the problem from a sparse sampling perspective. Specifically, we tackle two challenging higher-dimensional imaging problems with novel constrained imaging approaches. First, we consider image reconstruction for dynamic MRI, in which we aim at reconstructing a high-resolution spatiotemporal image function $\rho(\mathbf{x}, t)$ from highly undersampled $(\mathbf{k}, t)$-space data. Second, we consider the problem of accurately estimating parameter maps with highly accelerated MR relaxometry experiments.

1.2 Motivation

Since its invention by Dr. Paul Lauterbur in 1973 [3], MRI has been developed into a powerful tool to probe structural, functional, and metabolic information of biological systems. It has enabled a wide spectrum of applications, including cardiac imaging [4], phase-contrast blood flow imaging [5], dynamic speech imaging [6], dynamic contrast-enhanced cancer imaging [7], diffusion imaging [8], functional neuro-imaging [9], MR spectroscopic imaging [10] and MR elastography [11]. However, the practical utility of MRI has often been compromised by limited acquisition speed. Virtually all MRI applications can benefit from additional speedups, in particular higher-dimensional imaging problems.

Dynamic MRI is one class of imaging applications in which fast imaging techniques can play a critical role. For dynamic imaging, it is desirable to simultaneously achieve high spatial resolution and high temporal resolution. However, limited imaging speed often results in a very difficult trade-off between achievable spatial and temporal resolution for many dynamic MRI applications (e.g., 3D dynamic imaging of the whole heart). Although electrocardiography (ECG) gated acquisition and respiration control can help alleviate the problem, these techniques suffer from many well-known limitations. For example, ECG gating-based techniques heavily rely on the assumption that motion of interest is periodic, which constrains their effectiveness in a number of dynamic MRI applications (e.g., real-time imaging of natural speech production [12]).

Another higher-dimensional imaging application is MR parameter mapping, which aims at obtaining quantitative tissue properties from a sequence of images with different contrast-weightings. To ensure accurate estimates of relaxation parameters, a relatively long sequence of contrast-weighted images often has to be acquired. Furthermore, it is usually desirable to acquire parameter maps with high-resolution and broad volumetric coverage, which further contributes to prolonged imaging experiments. Achieving accurate quantitative MRI with highly accelerated acquisition has become a very active research area.

Over the past three decades, significant research efforts have been devoted to accelerating MRI, which has resulted in a large number of techniques, including efficient rapid imaging methods with advanced pulse sequences (e.g., [13–17]), parallel imaging techniques (e.g., [18–

22]), and constrained imaging approaches (e.g., [2, 23–31]). For emerging imaging problems with ever-increasing dimensions, it is often difficult to meet the acceleration requirement using a single type of fast MRI technology. Effective integration of novel constrained imaging approaches with advanced hardware and acquisition schemes could create unprecedented opportunities. In this thesis, we will present our exploration along this line.

1.3 Main Results

- We have proposed a novel constrained reconstruction method to enable image reconstruction from highly undersampled $(\mathbf{k}, t)$-space data for dynamic MRI. Specifically, we introduce a novel imaging model, i.e., the joint low-rank and sparse model, to parsimoniously represent dynamic imaging data. We further design specialized data acquisition schemes to pre-estimate the temporal subspace of the low-rank model. A convex formulation is proposed to integrate the low-rank/subspace constraint with sparsity constraint. We develop an efficient solution algorithm based on the half-quadratic regularization with continuation procedure to solve the resulting optimization problem. We evaluated the performance of the proposed method using both simulated and in-vivo real-time cardiovascular imaging data. We demonstrate the complementary advantages of low-rank and sparse constraints, and highlight the superior performance of imposing joint constraints. Related publications include Refs. [31–36].

- We have extended the joint low-rank and sparse model to accelerate MR parameter mapping. We specifically tailor the low-rank model and sparse model to capture different signal/image characteristics for MR parameter mapping applications: the low-rank structure captures the correlation among relaxation signals at different spatial locations, whereas the joint sparsity constraint captures the edge structure shared by a sequence of co-registered contrast-weighted MR images. We further integrate the proposed image model with multichannel acquisition to achieve an even higher level of acquisition. We demonstrate the superior performance of the proposed method over two state-of-the-art methods that are based on a single low-rank or sparse models using

experimental data from T_2 mapping of the human brain and T_1 mapping of the rat brain. Related publications include Refs. [37–39].

- We have proposed a novel constrained parameter estimation method for fast MR parameter mapping with sparse sampling. It utilizes a parametric signal model from spin physics and imposes sparsity constraints on the model parameters. With the proposed constrained imaging model, the parameter mapping reconstruction problem has been converted into a parameter estimation problem. We present an efficient greedy-pursuit algorithm to solve the resulting optimization problem. Estimation-theoretic bounds are also derived to analyze the advantages of using the sparsity constraints and benchmark the proposed method against the fundamental performance limit. The theoretical characterizations and empirical performance of the proposed method are illustrated in a T_2 mapping application example with computer simulations. Related publications include Refs. [40–42].

1.4 Organization of the Dissertation

The rest of this dissertation is organized as follows.

Chapter 2 provides a review of the necessary background for the subsequent chapters, covering Fourier-based MR image formation, compressive sensing theory, algorithms and their applications to MRI, and statistical estimation theory.

Chapter 3 presents the proposed method for dynamic image reconstruction with joint low-rank and sparsity constraints. It includes the problem formulation, solution algorithm, and an application example in real-time cardiovascular imaging.

Chapter 4 presents the proposed method for accelerated MR parameter mapping with joint low-rank and sparsity constraints, including the proposed formulation, solution algorithm, and two MR parameter mapping applications.

Chapter 5 presents the proposed method for fast model-based MR parameter mapping. It includes the proposed formulation, greedy-pursuit based solution algorithm, and estimation-theoretic characterization. Both theoretical and empirical results are shown to demonstrate

the performance of the proposed method.

Finally, Chapter 6 concludes this dissertation and points out several future research directions.

CHAPTER 2

BACKGROUND

In this chapter, we present the necessary background material for the subsequent chapters. We first review the principles of MRI in Section 2.1, covering signal generation and detection, spatial encoding, and image formation. In Section 2.2, we present a survey of compressive sensing (CS) and its applications in MRI. In Section 2.3, we summarize some of key results from statistical estimation theory.

2.1 Principles of MRI

2.1.1 Signal Generation and Detection

MRI is based on a physical phenomenon called nuclear magnetic resonance (NMR), which involves the interaction of nuclei with two different magnetic fields, a static magnetic field and a radiofrequency (RF) magnetic field. It is such interaction that gives rise to physically detectable NMR signals. From quantum mechanics, it is known that certain types of nuclei, which have odd atomic weights and/or odd atomic numbers, possess an intrinsic property called spin angular momentum. Nuclei with spin angular momentum can interact with external magnetic fields, which makes them relevant to NMR. Many nuclei in biological systems, such as ^{1}H, ^{32}P, and ^{23}Na, are NMR relevant. In particular, ^{1}H, as a constituent of the water molecule, is extremely abundant in various biological tissues, making it the most widely used spin-species in MRI.

At the microscopic level, the spin angular momentum $\mathbf{J}$ induces the magnetic moment, which is equal to

$$\boldsymbol{\mu} = \gamma \mathbf{J},$$

where γ is the gyromagnetic ratio (a constant for a specific spin-species). From quantum mechanics, the magnitude of $\boldsymbol{\mu}$ is equal to $\gamma\hbar\sqrt{\mathrm{I}\left(\mathrm{I}+1\right)}$, where $\hbar$ is Planck's constant divided by 2π, and I is the nuclear spin quantum number. The direction of $\boldsymbol{\mu}$ is random due to thermal motion. At the macroscopic level, if we look at the collective behavior resulting from an ensemble of N_s nuclei, the magnetization vector, defined as $\mathbf{M}=\sum_{n=1}^{N_s}\boldsymbol{\mu}_n$, is zero, because of the cancelation of magnetic moments associated with randomly oriented nuclei.

If an external static magnetic field is applied along the longitudinal direction, i.e., $\mathbf{B}_0 = B_0\mathbf{k}$, magnetic polarization occurs, which aligns randomly oriented spins. As a result, $\boldsymbol{\mu}$ can take only two possible orientations, either parallel or antiparallel with respect to $\mathbf{B}_0$. Furthermore, spins at the two directions are associated with different energy levels. According to the Boltzmann distribution, the population difference can be approximated by

$$N_+ - N_- \cong N_s\frac{\gamma\hbar B_0}{2KT_s},$$

where N_+ and N_- respectively represent the number of up spins and down spins, K is the Boltzmann constant, and T_s is the absolute temperature of the spin system. At the macroscopic level, the population difference results in a nonzero magnetization vector $\mathbf{M}_z$ pointing in the same direction as $\mathbf{B}_0$. For example, for the spin-$\frac{1}{2}$ system such as ^{1}H, the magnitude of $\mathbf{M}_z$ can be written as

$$M_z^0 = \frac{\gamma^2\hbar^2 B_0 N_s}{4KT_s}.$$

Note that $\mathbf{M}_z$ is not time-varying, which implies that it cannot be measured physically.

To excite the spin system, an RF magnetic field is applied in the perpendicular direction to $\mathbf{B}_0$, e.g.,

$$\mathbf{B}_1 = 2B_1^e\left(t\right)\cos\left(\omega_{\mathrm{rf}}t\right)\mathbf{i},$$

where $B_1^e\left(t\right)$ is the pulse envelope function, and ω_{rf} is the carrier frequency of the RF pulse. If ω_{rf} is tuned to be equal to the Larmor frequency of a specific spin species, i.e., $\omega_{\mathrm{rf}} = \omega_0$, the corresponding spin system is excited. Phenomenologically, the spin dynamics can be

described by the Bloch equation:

$$\frac{d\mathbf{M}}{dt} = \mathbf{M} \times \gamma\mathbf{B} - \frac{M_x\mathbf{i} + M_y\mathbf{j}}{T_2} - \frac{(M_z - M_z^0)}{T_1}, \tag{2.1}$$

where T_1 and T_2 respectively represent the longitudinal and transverse relaxation parameters, which are constants for a specific type of tissue. Considering that the duration of the RF pulse is much shorter than T_1 or T_2, the following simplified equation can be considered:

$$\frac{d\mathbf{M}}{dt} = \mathbf{M} \times \gamma\mathbf{B}, \tag{2.2}$$

where $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$, due to the presence of the main magnetic field and the RF pulse. By solving (2.2), we can analyze the spin dynamics during the excitation. It should be noted that (2.2) is more easily solved in the rotating frame. The related details can be found in [1]. A direct consequence of (2.2) is that the magnetization is tipped a certain angle away from its equilibrium, and precesses around the longitudinal direction.

After the RF excitation, by solving (2.1) with $\mathbf{B} = \mathbf{B}_0$, we can obtain the dynamics of magnetization at the longitudinal direction and transverse plane:

$$M_z(t) = M_z^0\left(1 - e^{t/T_1}\right), \tag{2.3}$$

$$M_{xy}(t) = M_x + iM_y = M_z^0 e^{-t/T_2} e^{-i\omega_0 t}. \tag{2.4}$$

Equations (2.3) and (2.4) indicate that the longitudinal magnetization goes through an exponential recovery process with a time constant T_1, while the transverse magnetization experiences an exponential decay process with a time constant T_2.

In (2.3) and (2.4), both M_z and M_{xy} are time-varying magnetization, which implies that they can be detected using receiver RF coils. By Faraday's law of induction, the voltage in the receiver coils can be expressed as

$$s_r(t) = -\frac{\partial}{\partial t}\int_{\text{volume}} \mathbf{B}_r(\mathbf{r}) \cdot \mathbf{M}(\mathbf{r}, t)\, d\mathbf{r}, \tag{2.5}$$

where $\mathbf{B}_r(\mathbf{r}) = B_{r,x}(\mathbf{r})\,\mathbf{i} + B_{r,y}(\mathbf{r})\,\mathbf{j} + B_{r,z}(\mathbf{r})\,\mathbf{k}$ denotes the receiver coil's sensitivity. Con-

sidering that $M_z(t)$ is a much slower time-varying magnetization compared to $M_{xy}(t)$, (2.5) can be approximated by

$$s_r(t) = -\frac{\partial}{\partial t}\int_{\text{volume}} \left[B_{r,x}(\mathbf{r}) M_x(\mathbf{r},t) + B_{r,y}(\mathbf{r}) M_y(\mathbf{r},t)\right] d\mathbf{r}. \tag{2.6}$$

Substituting (2.4) into (2.6), we can obtain the expression of the received signal:

$$s_r(t) = i\omega_0 e^{-i\omega_0 t}\int_{\text{volume}} B^*_{r,xy}(\mathbf{r}) M_{xy}(\mathbf{r},0)\, e^{-t/T_2(\mathbf{r})} d\mathbf{r}. \tag{2.7}$$

If we assume that the sensitivity of the receiver coil is uniform and further ignore the T_2 relaxation during the readout time, the measured time signal in (2.7) can be rewritten as

$$s_r(t) = Ce^{-i\omega_0 t}\int_{\text{volume}} M_{xy}(\mathbf{r},0)\, d\mathbf{r}, \tag{2.8}$$

where C is a system-specific scaling constant. Furthermore, through demodulation, we can obtain the baseband signal as follows:

$$s_r(t) = C\int_{\text{volume}} M_{xy}(\mathbf{r},0)\, d\mathbf{r}. \tag{2.9}$$

2.1.2 Spatial Information Encoding

As described above, the activated NMR signal can be measured physically using receiver RF coils. The measured signal in (2.9) contains the information from the whole imaging object, which does not have the capability of spatial localization. For tomographic imaging, a certain mechanism is needed for spatial information encoding. There are multiple ways of encoding spatial information into measured time signals. In the following, we review the most common one that is based on the use of linear gradient fields. Specifically, after the RF pulse, a linear gradient magnetic field

$$\triangle\mathbf{B}(\mathbf{r}) = (G_x x + G_y y + G_z z)\,\mathbf{k}, \tag{2.10}$$

is applied in the longitudinal direction. Substituting $\mathbf{B} = (B_0 + \triangle B)\,\mathbf{k}$ into (2.1), we can obtain the Bloch equation for the transverse component:

$$\frac{dM_{xy}(\mathbf{r},t)}{dt} = -\left\{\frac{1}{T_2(\mathbf{r})} + i\left[\omega_0 + \gamma\triangle B(\mathbf{r})\right]\right\} M_{xy}(\mathbf{r},t). \tag{2.11}$$

The solution of (2.11) can be written as

$$M_{xy}(\mathbf{r},t) = M^0 e^{-t/T_2(\mathbf{r})} e^{-i\omega_0 t} e^{-i\gamma t(G_x x + G_y y + G_z z)}. \tag{2.12}$$

Following the same signal detection principle as before, the measured signal at the receiver coil can be expressed as

$$s_r(t) = C\int M_{xy}(\mathbf{r}) \exp\left(-i\gamma t\left(G_x x + G_y y + G_z z\right)\right) d\mathbf{r}. \tag{2.13}$$

By defining $k_x = \gamma G_x t/2\pi$, $k_y = \gamma G_y t/2\pi$, and $k_z = \gamma G_z t/2\pi$, (2.13) can be rewritten as

$$\begin{aligned} s_r(t) &= C\int M_{xy}(\mathbf{r}) \exp\left(-i2\pi\left(k_x x + k_y y + k_z z\right)\right) d\mathbf{r}, \\ &= \int \rho(\mathbf{r}) \exp\left(-i2\pi\mathbf{k}(t)\cdot\mathbf{r}\right) d\mathbf{r}, \end{aligned} \tag{2.14}$$

where $\rho(\mathbf{r}) = CM_{xy}(\mathbf{r})$ is the underlying image function. Equation (2.14) establishes a one-to-one mapping between the samples of $s_r(t)$ and samples in the spatial Fourier transform of $\rho(\mathbf{r})$. It implies that by proper manipulation of the gradient field, we can acquire a set of time signals that can be mapped to the Fourier samples of $\rho(\mathbf{r})$, which yields the following Fourier imaging equation:

$$s(\mathbf{k}(t)) = \int \rho(\mathbf{r}) \exp(-i2\pi\mathbf{k}(t)\cdot\mathbf{r}) d\mathbf{r}, \tag{2.15}$$

where $s(\mathbf{k}(t))$ is the $\mathbf{k}$-space data.

2.1.3 Image Formation

The goal of image formation is to reconstruct $\rho(\mathbf{r})$ from $s(\mathbf{k})$. First, we make the two assumptions on the imaging process: a) the number of measured $\mathbf{k}$-space data is finite, and b) measurements are corrupted by noise. As a result, the imaging equation in (2.15) can be rewritten as

$$s(\mathbf{k}_m) = \int \rho(\mathbf{r}) \exp(-i2\pi \mathbf{k}_m \cdot \mathbf{r}) d\mathbf{r} + \eta(\mathbf{k}_m), \tag{2.16}$$

for $m = 1, \cdots, M$, where $\eta(\mathbf{k}_m)$ is the measurement noise.

Direct determination of $\rho(\mathbf{r})$, which is a continuous function, from a finite number of measurements is typically ill posed. Prior knowledge of $\rho(\mathbf{r})$ is needed. A number of useful assumptions can be utilized to make the reconstruction problem well-posed. Here, we assume that $\rho(\mathbf{r})$ lies in a finite dimensional space, in which $\rho(\mathbf{r})$ can be represented by a finite series expansion [43, 44]:

$$\rho(\mathbf{r}) = \sum_{n=1}^{N} \boldsymbol{\rho}_n \phi_n(\mathbf{r}), \tag{2.17}$$

where $\{\phi_n(\mathbf{r})\}_{n=1}^{N}$ is a given set of basis functions and N denotes the dimension of the space. There are many ways of choosing basis functions for (2.17). Here, we consider a widely used one, translations of a single voxel function $\phi(\mathbf{r})$, i.e.,

$$\phi_n(\mathbf{r}) = \phi(\mathbf{r} - \mathbf{r}_n), \tag{2.18}$$

for $n = 1, \cdots, N$, where $\mathbf{r}_n$ denotes the center of the nth translated voxel function. Typically, $\phi(\mathbf{r})$ is chosen to be a highly localized function, such as the Dirac delta function or a rectangular function.

Substituting (2.17) and (2.18) into (2.16) yields

$$s(\mathbf{k}_m) = \sum_{n=1}^{N} \mathbf{F}_{m,n} \boldsymbol{\rho}_n + \eta(\mathbf{k}_m), \quad m = 1, \cdots, M, \tag{2.19}$$

where $\mathbf{F}_{m,n} = \exp(-i2\pi \mathbf{k}_m \mathbf{r}_n) \Phi(\mathbf{k}_m)$, and $\Phi(\mathbf{k}_m)$ is the Fourier transform of $\phi(\mathbf{r})$ evaluated

at $\mathbf{k} = \mathbf{k}_m$. In the matrix form, (2.19) can written as

$$\mathbf{d} = \mathbf{F}\boldsymbol{\rho} + \boldsymbol{\eta}, \tag{2.20}$$

where $\mathbf{d} \in \mathbb{C}^M$ contains measured data, $\mathbf{F} \in \mathbb{C}^{M \times N}$ is the measurement matrix, $\boldsymbol{\rho} \in \mathbb{C}^N$ contains coefficients of the voxel functions, and $\boldsymbol{\eta} \in \mathbb{C}^N$ is the noise vector.

Note that $\mathbf{F}$ in (2.20) depends on the $\mathbf{k}$-space sampling trajectories and selection of basis functions. For example, if $\phi(\mathbf{x})$ is chosen as the Dirac delta function, and if we assume that $M = N$ samples are acquired on a uniform Cartesian grid in $\mathbf{k}$-space, $\mathbf{F}$ reduces to the standard discrete Fourier transform matrix. In this case, reconstruction of $\boldsymbol{\rho}$ can be simply accomplished using the fast Fourier transform (FFT). Generally, if $\mathbf{F}$ has a full column rank, (2.20) can be solved using the following least-squares (LS) formulation:

$$\hat{\boldsymbol{\rho}}_{\mathrm{LS}} = \arg\min_{\boldsymbol{\rho}} \|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|_2^2. \tag{2.21}$$

The solution of (2.21) can be written as

$$\begin{aligned} \hat{\boldsymbol{\rho}}_{\mathrm{LS}} &= \mathbf{F}^{\dagger}\mathbf{d} \\ &= (\mathbf{F}^H\mathbf{F})^{-1}\mathbf{F}^H\mathbf{d}, \end{aligned} \tag{2.22}$$

where $\dagger$ denotes the Moore-Penrose pseudo-inverse. It is worth making a few comments on the LS solution. First, from a parameter estimation perspective, under the assumption that the measurement noise is complex white Gaussian noise, the LS solution is optimal in the maximum likelihood sense. Second, note that a necessary condition for $\mathbf{F}$ being full column rank is that $M \geq N$, which implies that the number of samples is no less than the number of image voxels. Considering that MRI is a relatively slow imaging modality, acquiring a relatively large number of samples can lead to prolonged imaging experiments.

In many scenarios (e.g., reduced acquisition), the number of samples can be less than the number of image voxels, making the LS formulation ill-posed. Therefore, it is necessary to incorporate a prior information of $\boldsymbol{\rho}$ to solve (2.20). A classical approach is the so-called

Tikhonov regularization:

$$\min_{\boldsymbol{\rho}} \|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|_2^2 + \lambda \|\mathbf{W}\left(\boldsymbol{\rho} - \boldsymbol{\rho}_0\right)\|_2^2, \tag{2.23}$$

where λ is the regularization parameter, $\mathbf{W}$ is a pre-specified linear transformation (e.g., a finite-difference matrix), and $\boldsymbol{\rho}_0$ is an initial guess of $\boldsymbol{\rho}$. Equation (2.23) has a unique solution, which can be expressed as

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\rho}_0 + \left(\mathbf{F}^H\mathbf{F} + \lambda\mathbf{W}^H\mathbf{W}\right)^{-1}\left(\mathbf{d} - \mathbf{F}\boldsymbol{\rho}_0\right). \tag{2.24}$$

Computationally, (2.23) is usually solved using iterative algorithms, such as the conjugate gradient algorithm. In these algorithms, neither $\mathbf{F}$ nor $\mathbf{F}^H$ needs to be stored explicitly, since both of them can be effectively evaluated through FFT. Furthermore, note that (2.24) is a linear reconstruction procedure, which allows for easier characterization of the resolution and SNR of $\hat{\boldsymbol{\rho}}$. However, there are a number of limitations of directly using (2.23) for image reconstruction. The most serious one is that the Tikhonov regularization essentially assumes that $\boldsymbol{\rho}$ is drawn from a Gaussian random process, which is rarely realistic for natural images. Thus, the performance of (2.23) is rather limited in practice. In the next section, we present an emerging technique that can effectively address the limited data reconstruction problem.

2.2 Compressive Sensing

Compressive sensing (CS) [45, 46] is an emerging research area in signal processing and information theory, which studies the problem of solving an undetermined linear system:

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.25}$$

where $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{A} \in \mathbb{C}^{M\times N}$, $\mathbf{b} \in \mathbb{C}^M$, and $M < N$. This problem arises in a wide variety of real-world applications, such as medical imaging [28,47], computer vision [48], and radar [49]. In these applications, $\mathbf{A}$ often denotes an encoding matrix, $\mathbf{x}$ denotes a signal of interest, and $\mathbf{b}$ contains the measured data. As an example, in MRI, $\mathbf{A}$ is an undersampled Fourier

encoding matrix, $\mathbf{x}$ is the underlying image function, and $\mathbf{b}$ is $\mathbf{k}$-space measurement vector.

Generally, it is not possible to find a unique solution for (2.25) without invoking further assumptions on the structures of $\mathbf{A}$ and/or $\mathbf{x}$. Compressive sensing deals with recovering a special class of signals, called sparse signals. In terms of the encoding process, it relies heavily on the special design of sensing mechanisms, in which randomness is inherent. To decode sparse signals from undersampled data, nonlinear recovery algorithms are employed. In this section, we review some of theoretical and algorithmic advances in CS, followed by a brief discussion of the potentials and limitations of applying CS to MRI.

2.2.1 Sparsity and Compressibility

Sparsity, as a parsimonious signal characteristic, is found to be inherent in many natural or man-made signals, such as audio signals, images, videos, etc. These real-world signals typically can be transformed to some domain (or certain coordinate system), in which their representational coefficients are sparse or approximately sparse. For example, piecewise smooth images are usually approximately sparse in the wavelet domain, which provides the foundation for modern image compression.

Mathematically, $\mathbf{x}$ is K-sparse, if

$$\|\mathbf{x}\|_0 \leq K, \tag{2.26}$$

where $K < N$ and $\|\mathbf{x}\|_0$ is the ℓ_0 quasi-norm of $\mathbf{x}$, which counts the nonzero entries in $\mathbf{x}$, i.e.,

$$\|\mathbf{x}\|_0 = \sum_{n=1}^{N} \mathbf{1}_{\{\mathbf{x}_n \neq 0\}}. \tag{2.27}$$

For most real-world signals, they are rarely sparse, but they may be well approximated by sparse signals. The notion of compressibility is thus introduced for those signals. In the following, we introduce the weak-ℓ_p norm that is often used to measure the compressibility of a signal.

Definition 2.2.1. *For $0 < p < \infty$, the weak-ℓ_p norm is defined as [50]:*

$$\|\mathbf{x}\|_{w\ell_p} \triangleq inf\left\{r : |\mathbf{x}|_{(n)} \leq r \cdot n^{-1/p} \quad for \quad n = 1, 2, \ldots, N\right\}, \tag{2.28}$$

where $|\mathbf{x}|_{(n)}$ *denotes the component of* $\mathbf{x}$ *that has the largest* n*th magnitude.*

The smaller the weak-ℓ_p norm, the higher compressibility of $\mathbf{x}$ is expected. For example, it has been shown in [50] that if the weak-ℓ_p norm of $\mathbf{x}$ is small for $p < 2$, $\mathbf{x}$ can be well approximated by its K-sparse approximation x_K, since

$$\|\mathbf{x} - \mathbf{x}_K\|_1 \leq K^{1-1/p}\|\mathbf{x}\|_{w\ell_p}. \tag{2.29}$$

It is worth noting that signals by themselves may not be sparse or compressible. However, through some proper sparsifying transforms, they can be transformed to be sparse or compressible. Common sparsifying transforms include orthonormal transforms, tight frames, patch-based dictionaries, etc. For the sake of simplicity, hereafter we assume that $\mathbf{x}$ is sparse or compressible in the canonical coordinate system. Note that most of the theory and algorithms below can be extended to the case in which sparsifying transforms are involved.

2.2.2 Sparse Reconstruction Algorithms

Here, we consider a more practical measurement model in which noise perturbation is taken into account, i.e.,

$$\mathbf{A}\mathbf{x} + \mathbf{n} = \mathbf{b}, \tag{2.30}$$

where $\mathbf{n} \in \mathbb{C}^N$ is the noise vector.

Sparse reconstruction aims at finding the sparsest solution to (2.30). Mathematically, it can be formulated as

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \ & \|\mathbf{x}\|_0, \\ \text{s.t.} \ \ & \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon, \end{aligned} \tag{2.31}$$

where ϵ is related to the noise level.

Note that (2.31) is an NP hard problem. It is not feasible to solve it directly. However, there are a variety of algorithms that have been developed to solve variants of (2.31), which are computationally tractable and guaranteed to yield the same solution as (2.31) under

certain theoretical conditions. These algorithms can be roughly grouped into the two classes, optimization-based methods and greedy-based methods. In the following, we first review two representative algorithms that respectively belong to the two classes.

Optimization-Based Method

The algorithms in this class employ proper surrogate functions to replace the ℓ_0 norm in (2.31). The most common surrogate function is based on the ℓ_q norm ($q \leq 1$) of $\mathbf{x}$, which is defined as

$$\|\mathbf{x}\|_q = \left(\sum_{n=1}^{N} |\mathbf{x}_n|^q \right)^{1/q}. \tag{2.32}$$

For $q = 1$, the resulting optimization problem is convex, whereas $q < 1$ yields non-convex optimization problems. Although there are algorithms that have been developed to solve nonconvex relaxations [51], the convex relaxation generally leads to more efficient computational problems. Here, we focus on solving the convex relaxation of (2.31):

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1, \\ \text{s.t.} \quad & \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon. \end{aligned} \tag{2.33}$$

This formulation is often called basis pursuit denoising [52]. The solution of (2.33) is related to the solution of the following unconstrained optimization problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \tag{2.34}$$

which is easier to solve. It can be shown that there exists a Lagrangian multiplier λ_ϵ, with which the solution of (2.34) is the same as the one for (2.33).

Note that the cost function in (2.34) is convex but not differentiable. A number of algorithms have recently been developed to address this problem (e.g., the interior point methods [53] and first-order methods [54–58]). The first-order methods are generally much more computationally efficient and scalable than the interior point methods, and thus are widely used to solve large-scale problems such as image reconstruction problems considered

in this thesis. In the following, we review a representative first-order algorithm, called the alternating direction method of multipliers (ADMM) [57, 59, 60]. The ADMM algorithm first converts (2.34) into a constrained optimization problem through the following variable splitting scheme:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{g}\|_1, \\ &\text{s.t.} \quad \mathbf{g}=\mathbf{x}. \end{aligned} \tag{2.35}$$

The augmented Lagrangian function for (2.35) can be written as

$$L_\mu(\mathbf{x},\mathbf{g},\mathbf{y}) = \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{g}\|_1 + <\mathbf{y},\mathbf{g}-\mathbf{x}> + \frac{\mu}{2}\|\mathbf{g}-\mathbf{x}\|_2^2. \tag{2.36}$$

An alternating minimization scheme is used to minimize (2.36). The procedures can be summarized as follows:

$$\mathbf{g}^{(n+1)} = \arg\min_{\mathbf{g}} L_\mu(\mathbf{x}^{(n)},\mathbf{g},\mathbf{y}^{(n)}), \tag{2.37a}$$

$$\mathbf{x}^{(n+1)} = \arg\min_{\mathbf{x}} L_\mu(\mathbf{x},\mathbf{g}^{(n+1)},\mathbf{y}^{(n)}), \tag{2.37b}$$

$$\mathbf{y}^{(n+1)} = \mathbf{y}^{(n)} + \mu(\mathbf{g}^{(n+1)} - \mathbf{x}^{(n+1)}). \tag{2.37c}$$

For (2.37a), it is equivalent to

$$\mathbf{g}^{(n+1)} = \arg\min_{\mathbf{g}} \frac{\mu}{2}\|\mathbf{g}-\mathbf{x}^{(n)}+\frac{1}{\mu}\mathbf{y}^{(n)}\|_2^2 + \lambda\|\mathbf{g}\|_1, \tag{2.38}$$

which can be solved by the following soft-thresholding operation:

$$\mathbf{g}_j^{(n+1)} = \frac{\mathbf{t}_j}{|\mathbf{t}_j|}\max\left\{|\mathbf{t}_j| - \frac{\lambda}{\mu}, 0\right\}, \tag{2.39}$$

where $\mathbf{t} = \mathbf{x}^{(n)} - \frac{1}{\mu}\mathbf{y}^{(n)}$. For equation (2.37b), it is equivalent to

$$\mathbf{x}^{(n+1)} = \arg\min_{\mathbf{x}} \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 + \frac{\mu}{2}\|\mathbf{g}-\mathbf{x}^{(n)}+\frac{1}{\mu}\mathbf{y}^{(n)}\|_2^2, \tag{2.40}$$

which is a quadratic optimization, for which a number of off-the-shelf solvers can be used, such as the conjugate gradient algorithm. In terms of (2.37c), it is the update of the Lagrangian multiplier.

It is worthwhile to mention that choosing an initialization $\{\mathbf{x}^0, \mathbf{g}^0, \mathbf{y}^0\}$ for the ADMM algorithm is fairly flexible, since the above iterative procedures are guaranteed to converge to the global optimal solution of (2.34) from any initialization [59] as long as $\mu > 0$. With respect to the stopping criterion, the algorithm is often terminated when the relative change of the solution is less than some pre-specified threshold or the number of iterations exceeds the maximum number of iterations. For the penalty parameter μ, it generally determines the convergence rate of algorithm [57]. However, note that the optimal selection of μ is still an open problem that is worth systematic study [57].

We summarize the overall implementation of ADMM in Algorithm 1.

Algorithm 1: Alternating Direction Method of Multipliers (ADMM)

Input: $\mathbf{A}, \mathbf{b}$, λ, *and* μ
Initialization: $\mathbf{x}^{(0)}$, $\mathbf{g}^{(0)}$, *and* $\mathbf{y}^{(0)}$
while *"the stopping criterion does not satisfy"* **do**

1. Solving $\mathbf{g}^{(n+1)}$ by (2.39),
2. Solving $\mathbf{x}^{(n+1)}$ from (2.40),
3. $\mathbf{y}^{(n+1)} = \mathbf{y}^{(n)} + \mu(\mathbf{g}^{(n+1)} - \mathbf{x}^{(n+1)})$.

Output: $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(n_{max})}$

Greedy-Based Method

The second class of algorithms are the greedy-based methods. These algorithms consider an alternative formulation of (2.31):

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \\ &\text{s.t.} \quad \|\mathbf{x}\|_0 \leq K, \end{aligned} \tag{2.41}$$

where K is a pre-specified parameter that controls the sparsity level.

Here, we review a representative greedy algorithm called compressive sampling matching pursuit (CoSaMP) [61]. First of all, the CoSaMP algorithm is initialized with $\mathbf{x}^{(0)} = 0$. At the nth iteration, the algorithm computes

$$\mathbf{y}^{(n+1)} = \mathbf{A}^H \left(\mathbf{b} - \mathbf{A}\mathbf{x}^{(n)}\right), \tag{2.42}$$

from which the largest $2K$ supports is detected. The resulting support set is denoted as $\Omega = L_{2K}\left(\mathbf{y}^{(n+1)}\right)$, where $L_{2K}(\cdot)$ returns the index set associated with the largest $2K$ entries of a vector.

Then, it combines Ω with the support of $\mathbf{x}^{(n)}$ to form a new support set $\mathcal{T}$:

$$\mathcal{T} = \Omega \cup \operatorname{supp}\left\{\mathbf{x}^{(n)}\right\}. \tag{2.43}$$

Note that $\mathcal{T}$ has at most $3K$ entries. After that, CoSaMP solves a support-constrained least-squares problem as follows:

$$\begin{aligned} \tilde{\mathbf{x}}^{(n)} &= \arg\min \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2, \\ \text{s.t.} \quad & \mathbf{x}|_{\mathcal{T}^c} = 0. \end{aligned} \tag{2.44}$$

The last step of CoSaMP is to prune the support of $\tilde{\mathbf{x}}^{(n)}$, i.e.,

$$\hat{\mathbf{x}}^{(n)} = H_K\left(\tilde{\mathbf{x}}^{(n)}\right), \tag{2.45}$$

where $H_K(\cdot)$ is the operator that keeps the largest K entries of an vector, while setting the remaining entries to zero. The above procedures are repeated until the maximum number of iterations $n_{\max}$ is reached.

We summarize the overall implementation of CoSaMP in Algorithm 2.

2.2.3 Performance Guarantees

Many CS recovery algorithms have theoretical guarantees for accurate and robust recovery. These guarantees commonly require that the measurement matrix $\mathbf{A}$ satisfies some desirable

Algorithm 2: Compressive Sampling Matching Pursuit (CoSaMP)

Input: $\mathbf{A}, \mathbf{b}$ *and* K

Initialization: $\mathbf{x}^{(0)} = \mathbf{0}$

while *"repeat until* n *reaches* n_{max}*"* **do**

1. $\mathbf{y}^{(n)} = \mathbf{A}^H \left(\mathbf{b} - \mathbf{A}\mathbf{x}^{(n)}\right)$,
2. $\Omega = L_{2K}\left(\mathbf{y}^{(n)}\right)$,
3. $\mathcal{T} = \Omega \cup \operatorname{supp}\left\{\mathbf{x}^{(n)}\right\}$,
4. Solve $\tilde{\mathbf{x}}^{(n)}$ from (2.44),
5. $\hat{\mathbf{x}}^{(n)} = H_K\left(\tilde{\mathbf{x}}^{(n)}\right)$.

Output: $\hat{\mathbf{x}} = \mathbf{x}^{(n_{max})}$

properties, such as the restricted isometry property (RIP) [45] or incoherence property [46]. Here, we review the RIP property and its related performance guarantees.

Definition 2.2.2 (Restricted Isometry Property [45])**.** *An* $M \times N$ *matrix* $\mathbf{A}$ *has the* K*-restricted isometry property if there exists a smallest constant* δ_K *such that*

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2 \tag{2.46}$$

holds for all K *sparse signals.*

The RIP essentially requires that all $M \times K$ submatrices of $\mathbf{A}$ are close to an isometry such that distances are preserved when sparse signals are projected from a high dimensional space to a low dimensional space. A set of sufficient conditions can be derived using RIP, which guarantees the performance of CS recovery with (2.33). For example, it can be shown that in the absence of noise, if $\delta_{2K} < 1$, (2.31) has a unique K-sparse solution; if $\delta_{2K} < \sqrt{2} - 1$, the solution to (2.33) is exactly the same as the solution to (2.31). With noise perturbation, the performance guarantee is given in the following theorem.

Theorem 2.2.3 (Robust Recovery Guarantee [62])**.** *Assume that* $\delta_{2s} < \sqrt{2}-1$ *and* $\|\boldsymbol{\eta}\|_2 \leq \epsilon$.

The solution to (2.33) *satisfies*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 s^{-1/2} \|\mathbf{x} - \mathbf{x}_s\|_1 + C_1 \epsilon, \tag{2.47}$$

where $\mathbf{x}_s$ *is the best* s *term approximation of* $\mathbf{x}$*,* C_0 *and* C_1 *are two positive constants that depend on* δ_{2s}*.*

The first term in the right hand side of (2.47) is related to the sparse approximation error, i.e., $\|\mathbf{x} - \hat{\mathbf{x}}_s\|_1$, while the second term is related to the noise perturbation. This performance bound implies that the reconstruction error of (2.33) is bounded by the weighted sum of the modeling error and noise error. If there exists no modeling error and noise error, (2.47) implies that perfect recovery can be achieved. Although (2.47) provides a complete characterization for robust recovery of sparse signals, checking whether a specific, deterministic measurement matrix $\mathbf{A}$ satisfies RIP is an NP hard problem in general. However, if $\mathbf{A}$ is a random matrix whose entries are independently drawn from Gaussian or Bernoulli distributions, $\mathbf{\Psi}$ satisfies the RIP with high probability, provided $M = \mathcal{O}(K \log(N/K))$.

2.2.4 Applications to MRI

CS provides both a theoretical and an algorithmic framework for MRI reconstruction with sparsely-sampled data. In general, medical images are piecewise smooth, thus they are often highly compressible with proper sparsifying transforms (e.g. finite difference or wavelet transforms). Image reconstruction with sparsity constraint can be formulated as [28]:

$$\hat{\boldsymbol{\rho}}_{\text{CS}} = \arg\min_{\boldsymbol{\rho}} \|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|_2^2 + \lambda_1 \|\mathbf{W}\boldsymbol{\rho}\|_1 + \lambda_2 \text{TV}(\boldsymbol{\rho}), \tag{2.48}$$

where $\mathbf{W}$ denotes the wavelet transform, TV denotes the total variation regularization, and λ_1 and λ_2 are two regularization parameters. The optimization problem in (2.48) can be easily solved by the ADMM-based algorithm as reviewed before. Here, we use a simple example to illustrate the performance of CS reconstruction for MRI. Figure 2.1 shows reconstructions with two sets of undersampled data acquired at the acceleration factors (AF) of 2.5 and 5,

respectively. Along with the reconstructions, the error maps are also shown. As can be seen, CS reconstructions perform fairly well at both accelerations. However, at AF = 5, it is easy to see that some edge structures of images were smoothed out (in particular those pointed by the arrows). This observation reflects one common limitation of CS reconstruction for MRI. More generally, there are still a number of key challenges of applying CS. From a theoretical point of view, the theoretical conditions of CS cannot be verified for MRI applications. For example, it is not computationally tractable to compute the RIP constant with respect to a specific Fourier undersampling scheme. Furthermore, note that the CS reconstruction process is nonlinear in nature, which brings significant difficulty of characterizing the resolution and noise properties of reconstructions. As a result, in terms of evaluating reconstruction performance, it is still based on a case-by-case basis.

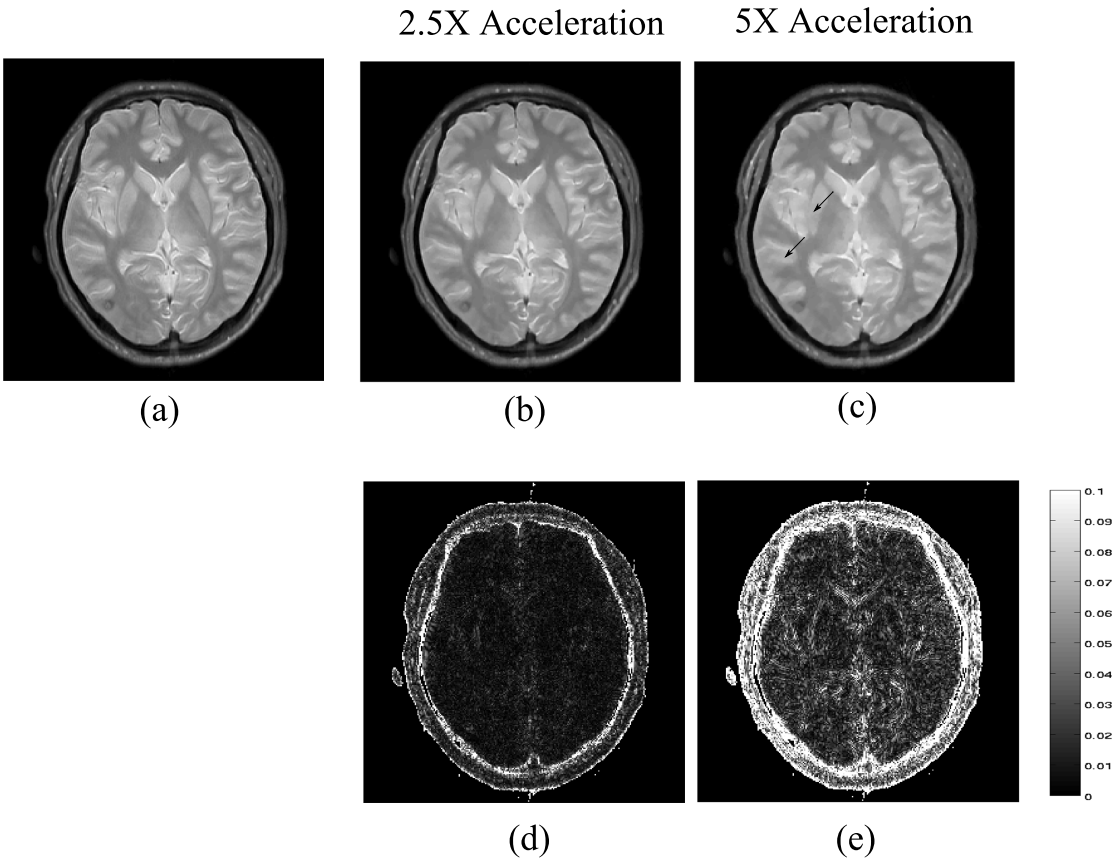


Figure 2.1: CS MRI reconstructions with two acceleration factors (AF): (a) ground truth, (b)-(c) CS reconstruction under AF = 2.5 and 5.0, (d)-(e) Error map of CS reconstruction under AF = 2.5 and 5.0.

2.3 Statistical Estimation Theory

In this section, we review some necessary background from statistical estimation theory, which will serve as the foundations for Chapter 5. In Section 2.3.1, we present two important estimators, the maximum likelihood (ML) estimator and the maximum a posterior (MAP) estimator. In Section 2.3.3, we review the Cramér-Rao lower bound (CRLB), followed by a summary of two important properties of ML estimators in Section 2.3.3.

2.3.1 Statistical Estimators

A key ingredient in statistical estimation is the data model, which indicates how observations are generated. The likelihood function $p(\mathbf{d};\boldsymbol{\rho})$, as a function of both observation $\mathbf{d}$ and unknown parameter $\boldsymbol{\rho}$, serves as such a generative model. For the ML estimation, it determines the unknown parameter that yields the highest probability of obtaining $\mathbf{d}$, i.e.,

$$\hat{\boldsymbol{\rho}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\rho}} p(\mathbf{d};\boldsymbol{\rho}). \tag{2.49}$$

A key assumption in the ML estimation is that $\boldsymbol{\rho}$ is a deterministic, unknown parameter rather than a random variable. Thus, the ML estimation is a frequentist inference approach.

The ML estimator can be determined by solving the optimization problem in (2.49). For example, with the imaging equation in (2.20) and the assumption that $\mathbf{n}$ is complex white Gaussian noise, i.e., $\mathbf{n} \sim N(0, \sigma^2\mathbf{I})$, the ML estimation can be formed as

$$\begin{aligned}
\hat{\boldsymbol{\rho}}_{\mathrm{ML}} &= \arg\max_{\boldsymbol{\rho}} p(\mathbf{d};\boldsymbol{\rho}) \\
&= \arg\min_{\boldsymbol{\rho}} \; -\log p(\mathbf{d};\boldsymbol{\rho}) \\
&= \arg\min_{\boldsymbol{\rho}} \; \frac{1}{\sigma^2}\|\mathbf{F}\boldsymbol{\rho} - \mathbf{d}\|_2^2 - M(\ln\pi + \ln\sigma^2) \\
&= \arg\min_{\boldsymbol{\rho}} \; \|\mathbf{F}\boldsymbol{\rho} - \mathbf{d}\|_2^2.
\end{aligned} \tag{2.50}$$

Equation (2.50) means that the ML estimator in this case is an optimal solution of a least-squares problem. The ML estimation is based on the data model and noise statistics. If a

relatively large number of observations is collected, it often leads to accurate estimates of unknown parameters.

However, when the number of observations is limited and/or the data SNR is low, the ML estimation may not be reliable. To improve estimation performance, incorporating prior knowledge of $\boldsymbol{\rho}$ is often needed. In this case, the estimation problem can be formulated within the Bayesian estimation framework, in which $\boldsymbol{\rho}$ is treated as a random variable with a certain prior distribution $p(\boldsymbol{\rho})$. An important class of Bayesian estimators is the MAP estimator, which chooses $\boldsymbol{\rho}_{\text{MAP}}$ that maximizes the posterior distribution $p(\boldsymbol{\rho}|\mathbf{d})$, i.e.,

$$\begin{aligned}\hat{\boldsymbol{\rho}}_{\text{MAP}} &= \arg\max_{\boldsymbol{\rho}} p(\boldsymbol{\rho}|\mathbf{d}) \\ &= \arg\min_{\boldsymbol{\rho}} -\log\frac{p(\boldsymbol{\rho})p(\mathbf{d};\boldsymbol{\rho})}{p(\mathbf{d})} \\ &= \arg\min_{\boldsymbol{\rho}} -\log p(\boldsymbol{\rho}) - \log p(\mathbf{d};\boldsymbol{\rho}),\end{aligned} \tag{2.51}$$

where $p(\mathbf{d})$ denotes the marginal distribution of the data. Considering that $p(\mathbf{d})$ is a constant irrelevant to the optimization problem, it is dropped from the last equation in (2.51). Note that $\hat{\boldsymbol{\rho}}_{\text{MAP}}$ represents a trade-off between the consistency with observations and the prior knowledge of $\boldsymbol{\rho}$.

We take the MR image reconstruction as an example to illustrate how MAP estimation can be formulated through the use of a prior information. First, we assume that $\boldsymbol{\rho}$ follows a complex Gaussian distribution with the mean $\boldsymbol{\rho}_0$ and covariance $\lambda\mathbf{I}$. In this case, the MAP estimator can be determined by solving

$$\hat{\boldsymbol{\rho}}_{\text{MAP}} = \arg\min_{\boldsymbol{\rho}} \|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|_2^2 + \frac{\sigma^2}{\lambda}\|\left(\boldsymbol{\rho} - \boldsymbol{\rho}_0\right)\|_2^2. \tag{2.52}$$

The MAP estimation in this case is equivalent to the Tikhonov regularization in (2.23). As mentioned before, natural images rarely follow a simple Gaussian distribution. Thus, applying (2.52) in a straightforward manner could lead to undesirable effects. For example, image reconstruction based on (2.52) often results in blurring of edges in images.

Alternatively, if $\boldsymbol{\rho}$ is assumed to follow a Laplacian distribution, the MAP estimator can

be determined by

$$\hat{\boldsymbol{\rho}}_{\mathrm{MAP}} = \arg\min_{\boldsymbol{\rho}} \|\mathbf{d} - \mathbf{F}\boldsymbol{\rho}\|_2^2 + \frac{\sigma^2}{\lambda} \| \left(\boldsymbol{\rho} - \boldsymbol{\rho}_0\right) \|_1. \tag{2.53}$$

Equation (2.53) is an ℓ_1 regularized least-squares problem. This estimator is commonly used in the cases that the signal has a compressible representation, as reviewed in Section 2.2. In MRI or other of other image recovery applications, this estimator or its variants can better preserve edges in images.

2.3.2 Cramér-Rao Lower Bound

In addition to constructing "optimal" estimators, statistical estimation theory is also concerned with understanding fundamental limits of parameter estimation. In the following, we present a widely used estimation-theoretic bound, the Cramér-Rao lower bound (CRLB), which provides a lower bound on covariance of any unbiased estimator.

We first define the Fisher Information Matrix (FIM) as follows:

$$\mathbf{I}_{\boldsymbol{\rho}} \triangleq \mathrm{E}\left[\left(\frac{\partial \ln p\left(\mathbf{d};\boldsymbol{\rho}\right)}{\partial \boldsymbol{\rho}}\right)\left(\frac{\partial \ln p\left(\mathbf{d};\boldsymbol{\rho}\right)}{\partial \boldsymbol{\rho}}\right)^T\right]. \tag{2.54}$$

Note that the FIM in (2.54) essentially indicates the curvature of the log-likelihood function. With the FIM, the CRLB can be described as follows. Under the following regularity condition:

$$E\left[\frac{\partial \ln p(\mathbf{d};\boldsymbol{\rho})}{\partial \boldsymbol{\rho}}\right] = 0 \quad \text{for all } \boldsymbol{\rho}, \tag{2.55}$$

the CRLB provides a lower bound on the covariance of any unbiased estimator $\hat{\boldsymbol{\rho}}$ [63]:

$$\mathrm{Cov}\left(\hat{\boldsymbol{\rho}}\right) - \mathbf{I}_{\boldsymbol{\rho}}^{-1} \geq 0, \tag{2.56}$$

where ≥ 0 is in the sense that the matrix is positive semidefinite. Equation (2.56) indicates that the Fisher information measures the amount of information that observations carry about the parameter $\boldsymbol{\rho}$. The higher the Fisher information, the better estimation accuracy that can be expected. Furthermore, for an unbiased estimator $\hat{\boldsymbol{\rho}}$ to attain the equality in

(2.56), a sufficient and necessary condition is that $\hat{\boldsymbol{\rho}}$ satisfies the following equation:

$$\frac{\partial \ln p(\mathbf{d};\boldsymbol{\rho})}{\partial \boldsymbol{\rho}} = \mathbf{I}_{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}). \tag{2.57}$$

An estimator that achieves the CRLB is called an efficient estimator.

As an example, we calculate the CRLB for the imaging equation in (2.20). Through simple matrix algebra, the FIM for (2.20) can be calculated as

$$\mathbf{I}_{\boldsymbol{\rho}} = \frac{1}{\sigma^2}(\mathbf{F}^H\mathbf{F}). \tag{2.58}$$

As a result, the CRLB can be expressed as

$$\mathrm{Cov}\,(\hat{\boldsymbol{\rho}}) \geq \sigma^2(\mathbf{F}^H\mathbf{F})^{-1}. \tag{2.59}$$

It is clear from (2.58) or (2.59) that with higher level of noise, less information is contained in observations, which leads to poorer estimation accuracy. Furthermore, it can be easily shown that the ML estimator in (2.22) is unbiased, and that it is also an efficient estimator that attains the CRLB.

2.3.3 Properties of Maximum Likelihood Estimators

There are a number of desirable properties associated with ML estimators. In the following, we review two important theorems that will be used in Chapter 5. The first one is about the asymptotic property of ML estimation stated as follows.

Theorem 2.3.1. *(Asymptotic Distribution of The ML Estimator [63]) Assuming that the likelihood function $p(\mathbf{d};\boldsymbol{\rho})$ satisfies the regularity condition in* (2.55)*, the ML estimator $\boldsymbol{\rho}_{\mathrm{ML}}$ is asymptotically distributed as*

$$\boldsymbol{\rho}_{\mathrm{ML}} \sim N(\boldsymbol{\rho}, \mathbf{I}_{\boldsymbol{\rho}}^{-1}), \tag{2.60}$$

where $\mathbf{I}_{\boldsymbol{\rho}}$ is the Fisher information matrix evaluated at the true parameter $\boldsymbol{\rho}$.

This theorem reveals several key asymptotic characteristics about the ML estimator. First,

as the number of observations goes to infinity, the limiting distribution of an ML estimator is simply a Gaussian, with its mean being the true parameter and the variance the inverse of the FIM. From the mean, it is clear that the ML estimator is asymptotically unbiased. Furthermore, since ML estimators asymptotically attain the CRLB, it is also efficient. Although this theorem holds in the asymptotic sense, it often provides a good approximation of the distribution of $\boldsymbol{\rho}_{\mathrm{ML}}$, when the number of observations is sufficiently large.

The second theorem is called the invariance property of the ML estimation, which involves estimating a function of $\boldsymbol{\rho}$.

Theorem 2.3.2. *(Invariance Property of ML Estimation [63]) The ML estimator of the parameter $\boldsymbol{\theta} = g(\boldsymbol{\rho})$ is given by*

$$\boldsymbol{\theta}_{\mathrm{ML}} = g(\boldsymbol{\rho}_{\mathrm{ML}}). \tag{2.61}$$

If g is not a invertible function, then $\boldsymbol{\theta}_{\mathrm{ML}}$ maximizes the following modified likelihood function $p_T(\mathbf{d}; \boldsymbol{\theta})$:

$$p_T(\mathbf{d}; \boldsymbol{\theta}) = \max_{\boldsymbol{\theta}=g(\boldsymbol{\rho})} p(\mathbf{d}; \boldsymbol{\theta}). \tag{2.62}$$

This theorem provides a simple way to find the ML estimator of $\boldsymbol{\theta} = g(\boldsymbol{\rho})$, given that $\boldsymbol{\rho}_{\mathrm{ML}}$ is already known.

CHAPTER 3

DYNAMIC IMAGE RECONSTRUCTION WITH JOINT LOW-RANK AND SPARSITY CONSTRAINTS

3.1 Introduction

In this chapter, we consider the dynamic image reconstruction problem with highly undersampled $(\mathbf{k}, t)$-space data. Dynamic MRI is a higher-dimensional imaging problem, which encompasses a wide variety of applications, including cardiac imaging [4], phase-contrast flow imaging [5], dynamic speech imaging [64], dynamic contrast-enhanced cancer imaging [7], time-resolved angiographic imaging [65], and functional neuroimaging [9]. The imaging equation for dynamic MRI can be written as

$$s(\mathbf{k}, t) = \int \rho(\mathbf{x}, t) \exp(-i2\pi\mathbf{k} \cdot \mathbf{x}) d\mathbf{x} + \eta(\mathbf{k}, t), \tag{3.1}$$

where $\rho(\mathbf{x}, t)$ is the desired image function, $s(\mathbf{k}, t)$ denotes the $(\mathbf{k}, t)$-space measured data, and $\eta(\mathbf{k}, t)$ denotes the measurement noise.

Conventional imaging methods often assume that $\rho(\mathbf{x}, t)$ is a support-limited function in the spatial-spectral domain and reconstruct $\rho(\mathbf{x}, t)$ from samples of $s(\mathbf{k}, t)$ acquired at the Nyquist rate. Due to the limited imaging speed of MRI, these methods usually face the inherent difficulty of simultaneously achieving both high spatial resolution and high temporal resolution. Furthermore, for a number of emerging applications (e.g., real-time imaging of natural speech production or real-time imaging of blood flow), it is rarely possible to collect a set of $(\mathbf{k}, t)$-space data that satisfy the classical sampling theorem. To effectively tackle these challenges, we here consider the problem of reconstructing $\rho(\mathbf{x}, t)$ from highly undersampled data, which frequently arises in accelerated dynamic imaging.

Although conventional reconstruction methods cannot handle sub-Nyquist data, it was

demonstrated very early that high-quality images can be reconstructed from undersampled data with appropriate constraints [25, 66]. Over the last few years, significant advances have been made in theory and algorithms on the use of low-dimensional signal structures to solve reconstruction problems with undersampled data, which results in a large number of new methods. These methods can be roughly grouped into the following classes: (a) methods utilizing generalized spatial and spectral support structure (e.g., [26, 27, 29, 67–69]), (b) methods utilizing sparse structure (e.g., [70–78]),(c) methods utilizing low-rank structure (e.g., [2,25,30,32,79–87]), (d) methods utilizing low-rank tensor structure (e.g., [88, 89]), (e) methods utilizing low-dimensional manifold structure (e.g., [90]), and (f) methods simultaneously utilizing multiple signal structures (e.g., [85, 91–93]).

In this chapter, we present a new constrained reconstruction method for dynamic MRI. It is based on a mathematical formulation utilizing a novel low-dimensional model, joint low-rank and sparse model, to enable reconstruction from highly undersampled $(\mathbf{k}, t)$-space data. The proposed formulation results in a convex optimization problem, for which we develop a globally convergent algorithm based on the half-quadratic regularization and continuation procedure. We systematically study the strength and limitations of using low-rank and sparse models individually for dynamic image reconstruction. We demonstrate that the proposed simultaneous low-rank and sparse model integrates the complementary advantages of individual low-rank and sparsity model by using the sparsity constraint to regularize low-rank constrained reconstruction. We illustrate the performance of the proposed method in a real-time cardiac imaging application.

The rest of this chapter is organized as follows. In Section 3.2, we present the proposed formulation; in Section 3.3, we present the reconstruction algorithm and its computational characteristics; in Section 3.4, we show representative results from both simulated and in-vivo real-time cardiac imaging data; in Section 3.5, we discuss the related issues, followed by a summary in Section 3.6.

3.2 Imaging Model

3.2.1 Data Model

We use a discrete image model in which $\rho(\mathbf{x}, t)$ can be completely represented by the following Casorati matrix [2]:

$$\mathbf{C} = \begin{bmatrix} \rho(\mathbf{x}_1, t_1) & \dots & \rho(\mathbf{x}_1, t_M) \\ \vdots & \ddots & \vdots \\ \rho(\mathbf{x}_N, t_1) & \dots & \rho(\mathbf{x}_N, t_M) \end{bmatrix} \in \mathbb{C}^{N \times M}, \tag{3.2}$$

which is formed from samples of $\rho(\mathbf{x}, t)$ on a grid of N spatial locations $\{\mathbf{x}_n\}_{n=1}^{N}$ and M time instants $\{t_m\}_{m=1}^{M}$. Furthermore, let $\boldsymbol{\rho} = \mathrm{vec}(\mathbf{C}) \in \mathbb{C}^{NM \times 1}$ be vector representation of $\rho(\mathbf{x}, t)$.

We also assume that D samples are acquired from $s(\mathbf{k}, t)$. Then, the imaging equation (3.1) can be written as

$$\mathbf{d} = \boldsymbol{\Phi}\boldsymbol{\rho} + \boldsymbol{\eta}, \tag{3.3}$$

where $\mathbf{d} \in \mathbb{C}^{D \times 1}$ contains the measured data, $\boldsymbol{\Phi} \in \mathbb{C}^{D \times NM}$ is the measurement matrix that models the operations of the spatial Fourier transform and $(\mathbf{k}, t)$-space sparse sampling, and $\boldsymbol{\eta} \in \mathbb{C}^{D \times 1}$ is the noise vector.

3.2.2 Proposed Formulation

In accelerated dynamic imaging, $D \ll NM$ and (3.3) is highly underdetermined. Here we propose a new constrained reconstruction method to solve this underdetermined system of equations, by utilizing a joint low-rank and sparsity structure via the following formulation:

$$\hat{\boldsymbol{\rho}} = \arg \min_{\boldsymbol{\rho} \in \mathbb{C}^{NM \times 1}} \|\mathbf{d} - \boldsymbol{\Phi}\boldsymbol{\rho}\|_2^2 + R_{\mathrm{r}}(\boldsymbol{\rho}) + R_{\mathrm{s}}(\boldsymbol{\rho}). \tag{3.4}$$

In this formulation, the penalty functions $R_{\mathrm{r}}(\cdot)$ and $R_{\mathrm{s}}(\cdot)$ are used to incorporate the low-rank and sparsity constraints respectively, as explained below.

Low-Rank Constraint

The low-rank constraint is due to the partial separability model [2], which represents $\rho(\mathbf{x}, t)$ as:

$$\rho(\mathbf{x}, t) = \sum_{\ell=1}^{L} u_\ell(\mathbf{x}) v_\ell(t), \tag{3.5}$$

where L is the order of the PS model, and $\{u_\ell(\mathbf{x})\}_{\ell=1}^{L}$ and $\{v_\ell(t)\}_{\ell=1}^{L}$ are sets of spatial and temporal functions, respectively. The PS model was proposed to capture spatiotemporal correlation often observed in dynamic image sequences [2]. It is easy to show that with (3.5), any set of functions of the form $\{\rho(\mathbf{x}_\ell, t)\}_{\ell=1}^{\hat{L}}$ are linearly dependent if $\hat{L} > L$. The PS constraint defined in (3.5) also implies that the rank of $\mathbf{C}$ is upper bounded by L [2]. As a consequence, $\mathbf{C}$ has at most $2L(M + N - L)$ real degrees of freedom [82], which is often much smaller than the total number of elements in $\mathbf{C}$. This intuitively explains why low-rank constraint has the capability to enable reconstruction from sub-Nyquist data.

There are several ways to enforce the PS or low-rank constraint. One way is to enforce it "implicitly" using surrogate functions. Representative surrogate functions include the nuclear norm, i.e., $R_{\mathrm{r}}(\boldsymbol{\rho}) = \lambda_r \|\mathbf{C}\|_*$, or the Schatten-p norm, i.e., $R_{\mathrm{r}}(\boldsymbol{\rho}) = \lambda_r \|\mathbf{C}\|_p^p$, which are the convex and nonconvex surrogates for rank($\mathbf{C}$), respectively. These implicit low-rank constraints have been investigated for dynamic image reconstruction in the previous work [85].

Another way is to explicitly enforce the low-rank constraint with a known L [2,30,32], by using

$$R_{\mathrm{r}}(\boldsymbol{\rho}) = \begin{cases} 0, & \text{if rank}(\mathbf{C}) \leq L, \\ \infty, & \text{else.} \end{cases} \tag{3.6}$$

Using (3.6) is equivalent to forcing the following matrix factorization:

$$\mathbf{C} = \mathbf{U}_{\mathrm{s}} \mathbf{V}_{\mathrm{t}}, \tag{3.7}$$

where $\mathbf{U}_{\mathrm{s}} \in \mathbb{C}^{N \times L}$ represents a basis for the spatial (or column) subspace of $\mathbf{C}$, and $\mathbf{V}_{\mathrm{t}} \in \mathbb{C}^{L \times M}$ contains a basis for the temporal (or row) subspace of $\mathbf{C}$. Note that $\mathbf{U}_{\mathrm{s}}$ and $\mathbf{V}_{\mathrm{t}}$ in

(3.7) are not unique, but the columns of $\mathbf{U}_\mathrm{s}$ span the same column subspace and rows of $\mathbf{V}_\mathrm{t}$ span the same row subspace.

An even stronger way to enforce the explicit low-rank constraints is to assume that $\mathbf{V}_\mathrm{t}$ in (3.7) is known, or can be accurately estimated prior to full image reconstruction. This assumption is generally valid when specialized data acquisition schemes are used to sample $(\mathbf{k}, t)$-space [2, 71, 79–83]. For example, if $s(\mathbf{k}, t)$ is fully sampled at the temporal Nyquist rate for a number of $\mathbf{k}$-space locations, then the dominant (rank-L) temporal subspace of these measurements can be extracted using singular value decomposition (SVD). Assuming that this subspace is representative of the dominant subspace of $\mathbf{C}$, $\mathbf{V}_\mathrm{t}$ can be obtained from the L dominant right singular vectors from the above SVD.

For the proposed formulation in (3.4), we choose to impose explicit rank constraint with a known $\mathbf{V}_\mathrm{t}$ (or estimated as described above). We further assume, without loss of generality, that $\mathbf{V}_\mathrm{t}$ has L orthonormal rows. Later, we will demonstrate in Section 3.4 that our proposed approach has both imaging and computational advantages over the method using implicit rank constraints [85].

Sparsity Constraint

Dynamic image sequences often have spatiotemporal features that lead to approximately sparse representations [71, 72]. Here we assume that $\Psi\boldsymbol{\rho}$ is sparse (or has a small ℓ_0 norm) under a certain sparsifying transform Ψ. Generally, the choice of Ψ is application-dependent. Here, we use the temporal Fourier transform as an example to derive our proposed algorithm. This sparsifying transform is particularly effective for quasi-periodic motion often arising in dynamic MRI. Directly enforcing the sparsity constraints through the ℓ_0 norm is not desirable, although greedy algorithms for ℓ_0 minimization exist [94]. An important development in compressed sensing is the establishment of a theoretical foundation for the use of surrogate cost functions that are more tractable to optimize. Specifically, the convex ℓ_1 norm is much easier to optimize and has proven optimality for sparsity-constrained inverse problems under certain conditions [45, 46], although nonconvex surrogate functions, such as the ℓ_q quasinorm $(0 < q < 1)$ [51], can also be used. For the proposed formulation in (3.4), we

use the ℓ_1 norm to impose the sparsity constraint and the corresponding penalty function is $R_{\mathrm{s}}(\boldsymbol{\rho}) = \lambda_{\mathrm{s}}\|\Psi\boldsymbol{\rho}\|_1$.

Reconstruction with Joint Low-Rank and Sparsity Constraints

Under the assumptions described above, the solution to (3.4) is given by $\hat{\mathbf{C}} = \hat{\mathbf{U}}_{\mathrm{s}}\mathbf{V}_{\mathrm{t}}$, where $\hat{\mathbf{U}}_{\mathrm{s}}$ is given by

$$\hat{\mathbf{U}}_{\mathrm{s}} = \arg\min_{\mathbf{U}_{\mathrm{s}}\in\mathbb{C}^{N\times L}} \|\mathbf{d} - \Omega(\mathbf{F}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{t}})\|_2^2 + \lambda\|\mathrm{vec}(\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{f}})\|_1, \tag{3.8}$$

where $\Omega : \mathbb{C}^{N\times M} \to \mathbb{C}^{D\times 1}$ represents the $(\mathbf{k}, t)$-space sampling operator, $\mathbf{F}_{\mathrm{s}} \in \mathbb{C}^{N\times N}$ is the spatial Fourier matrix, $\mathbf{V}_{\mathrm{f}} = \mathbf{V}_{\mathrm{t}}\mathbf{F}_{\mathrm{t}}$, and $\mathbf{F}_{\mathrm{t}} \in \mathbb{C}^{M\times M}$ is an orthonormal temporal Fourier matrix. Hereafter, we called the proposed solution to (3.8) as "PS-Sparse". It reduces to the basic PS/low-rank constrained reconstruction [2] (referred to as Basic-PS) when $\lambda = 0$, i.e.,

$$\hat{\mathbf{U}}_{\mathrm{s}} = \arg\min_{\mathbf{U}_{\mathrm{s}}\in\mathbb{C}^{N\times L}} \|\mathbf{d} - \Omega(\mathbf{F}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{t}})\|_2^2, \tag{3.9}$$

and to a basic $(\mathbf{x}, f)$-domain sparsity constrained reconstruction [71] (referred to as Basic-Sparse) if $L = M$ and $\mathrm{rank}(\mathbf{V}_{\mathrm{t}}) = M$, i.e.,

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}\in\mathbb{C}^{N\times M}} \|\mathbf{d} - \Omega(\mathbf{F}_{\mathrm{s}}\mathbf{C})\|_2^2 + \lambda\|\mathrm{vec}(\mathbf{C}\mathbf{F}_{\mathrm{t}})\|_1. \tag{3.10}$$

The proposed method incorporates both the low-rank and sparsity constraints into a single formulation, providing some desirable advantages over (3.9) and (3.10) summarized here (to be demonstrated in Section 3.4):

- In Basic-PS, the model order L cannot be high because it is limited by the number of measurements available. In principle, for the solution of the Lth order Basic-PS to be well-defined, each $\mathbf{k}$-space location has to have at least L temporal samples. In practice, many more measurements are often needed to avoid an ill-conditioned model fitting problem. The proposed method overcomes the limitations with the sparsity constraint serving as an effective regularizer.

- A major limitation of Basic-Sparse is that spatiotemporal blurring often appears when $(\mathbf{k}, t)$-space is highly undersampled. This is because many different sparse solutions can closely match the highly undersampled $(\mathbf{k}, t)$-space data. By imposing a PS constraint with $L \ll M$, the proposed method effectively exploits spatiotemporal correlations in the data to significantly reduce the set of candidate sparse solutions. This frequently yields better reconstruction results, removing blurring in the sparsity-constrained reconstructions.

3.3 Solution Algorithm and Characteristics

3.3.1 Solution Algorithm

In this section, we present a globally convergent, efficient solution algorithm to solve (3.8). The cost function of (3.8) is convex and coercive over $\mathbb{C}^{N\times L}$; therefore, there exists at least one optimal solution [95]. In most cases of interest, strict convexity also holds, which guarantees uniqueness of the optimal solution to (3.8). We focus here on finding any optimal solution. Note that (3.8) is a large-scale, non-smooth convex optimization problem for which a number of algorithms can be adapted (e.g., [54–58]). We develop an algorithm based on additive half-quadratic regularization [55, 96, 97] with continuation [55, 58]. The proposed algorithm is efficient and simple to implement, although computational efficiency could potentially be improved by considering other algorithms. In the following, we first present an overview of the proposed algorithm, and then describe each step of the algorithm in detail.

Summary of Algorithm

Because the regularization term in (3.8) is non-differentiable, solving (3.8) directly can be challenging. We approximate it with the following differentiable cost function:

$$\hat{\mathbf{U}}_{\mathrm{s}} = \arg \min_{\mathbf{U}_{\mathrm{s}}\in\mathbb{C}^{N\times L}} \|\mathbf{d} - \Omega(\mathbf{F}_{\mathrm{s}}\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{t}})\|_2^2 + \lambda \sum_{n=1}^{N}\sum_{m=1}^{M} \varphi((\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{f}})_{n,m}). \tag{3.11}$$

where $\varphi(t) : \mathbb{C} \to \mathbb{C}$ is the Huber function defined as

$$\varphi(t) = \begin{cases} |t|^2/2\alpha, & \text{if } |t| \leq \alpha, \\ |t| - \alpha/2, & \text{if } |t| > \alpha. \end{cases} \tag{3.12}$$

The parameter α controls the accuracy of the approximation of (3.8) by (3.11); Note that as $\alpha \to 0$, (3.11) approaches (3.8).

We will show that for a fixed α, solving (3.11) is equivalent to solving the following half-quadratic minimization problem:

$$\min_{\{\mathbf{U}_\mathrm{s},\ \mathbf{G}\}} \|\mathbf{d} - \Omega(\mathbf{F}_\mathrm{s}\mathbf{U}_\mathrm{s}\mathbf{V}_\mathrm{t})\|_2^2 + \frac{\lambda}{2\alpha}\|\mathbf{U}_\mathrm{s}\mathbf{V}_\mathrm{f} - \mathbf{G}\|_\mathrm{F}^2 + \lambda\|\mathrm{vec}(\mathbf{G})\|_1, \tag{3.13}$$

where $\mathbf{G}$ is an auxiliary matrix. Note that (3.13) is non-quadratic in $\mathbf{G}$ but quadratic in $\mathbf{U}_\mathrm{s}$.

Equation (3.13) can be solved by a simple alternating minimization procedure. Specifically, at the $(\ell+1)$th iteration, we first optimize (3.13) over $\mathbf{G}$ with a fixed $\mathbf{U}_\mathrm{s}^{(\ell)}$, i.e.,

$$\mathbf{G}^{(\ell+1)} = \arg\min_{\mathbf{G}\in\mathbb{C}^{N\times M}} \frac{1}{2\alpha}\left\|\mathbf{U}_\mathrm{s}^{(\ell)}\mathbf{V}_\mathrm{f} - \mathbf{G}\right\|_\mathrm{F}^2 + \|\mathrm{vec}(\mathbf{G})\|_1, \tag{3.14}$$

and then minimize (3.13) with respect to $\mathbf{U}_\mathrm{s}$ with a fixed $\mathbf{G}^{(\ell+1)}$, i.e.,

$$\mathbf{U}_\mathrm{s}^{(\ell+1)} = \arg\min_{\mathbf{U}_\mathrm{s}\in\mathbb{C}^{N\times L}} \|\mathbf{d} - \Omega(\mathbf{F}_\mathrm{s}\mathbf{U}_\mathrm{s}\mathbf{V}_\mathrm{t})\|_2^2 + \frac{\lambda}{2\alpha}\left\|\mathbf{U}_\mathrm{s}\mathbf{V}_\mathrm{f} - \mathbf{G}^{(\ell+1)}\right\|_\mathrm{F}^2. \tag{3.15}$$

The above alternating minimization procedure is repeated until $\mathbf{G}^{(\ell+1)}$ and $\mathbf{U}_\mathrm{s}^{(\ell+1)}$ converge. This iterative procedure can be viewed as a majorize-minimize algorithm for solving (3.11).

For a fixed non-zero α, the solution to (3.11) only yields an approximation to the solution of (3.8). To obtain a better solution, we apply a continuation scheme in which we gradually reduce α to zero. The continuation procedure starts with a large value of α, for which the optimization problem is approximately quadratic. In this case, the alternating minimization typically converges rapidly. For each subsequent continuation step, we decrease the value of α and use the previous solution as an initialization for the next round of optimization. This procedure is repeated until the cost function of (3.11) well approximates that in (3.8).

Equivalence of (3.11) and (3.13)

Note that the Huber function $\varphi(t)$ can be equivalently expressed as [97]

$$\varphi(t) = \min_{g\in\mathbb{C}} \frac{1}{2\alpha}|t-g|^2 + |g|, \tag{3.16}$$

where g is an auxiliary variable. With (3.16), the regularization term in (3.11) can be written as

$$\sum_{n=1}^{N}\sum_{m=1}^{M} \varphi(|\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{f}}|_{n,m}) = \min_{\mathbf{G}\in\mathbb{C}^{N\times M}} \frac{1}{2\alpha}\left\|\mathbf{U}_{\mathrm{s}}\mathbf{V}_{\mathrm{f}} - \mathbf{G}\right\|_{\mathrm{F}}^2 + \left\|\mathrm{vec}(\mathbf{G})\right\|_1. \tag{3.17}$$

Therefore, minimizing (3.11) is equivalent to minimizing (3.13).

Solution of (3.14)

Equation (3.14) is separable with respect to each entry of $\mathbf{G}$, and $\mathbf{G}^{(\ell+1)}$ can be obtained analytically through [98]

$$\mathbf{G}^{(\ell+1)} = \mathcal{S}(\mathbf{U}_{\mathrm{s}}^{(\ell)}\mathbf{V}_{\mathrm{f}}), \tag{3.18}$$

where $\mathcal{S}: \mathbb{C}^{N\times M} \to \mathbb{C}^{N\times M}$ is a soft-thresholding operator defined as

$$\mathcal{S}(\mathbf{Q})_{n,m} = \begin{cases} 0, & \text{if } |\mathbf{Q}_{n,m}| < \alpha, \\ \frac{\mathbf{Q}_{n,m}}{|\mathbf{Q}_{n,m}|}(|\mathbf{Q}_{n,m}| - \alpha), & \text{if } |\mathbf{Q}_{n,m}| \geq \alpha, \end{cases} \tag{3.19}$$

for any $\mathbf{Q} \in \mathbb{C}^{N\times M}$.

Solution of (3.15)

Equation (3.15) is a standard linear least squares problem, and its optimal solution is given by the following normal equations:

$$\left(\mathcal{A}^*\mathcal{A} + \frac{\lambda}{2\alpha}\mathcal{B}^*\mathcal{B}\right)\mathbf{U}_{\mathrm{s}}^{(\ell+1)} = \mathcal{A}^*(\mathbf{d}) + \frac{\lambda}{2\alpha}\mathbf{G}^{(\ell+1)}\mathbf{V}_{\mathrm{f}}^H, \tag{3.20}$$

where $\mathcal{A}(\mathbf{U}) \triangleq \Omega(\mathbf{F}_{\mathrm{s}}\mathbf{U}\mathbf{V}_{\mathrm{t}})$ and $\mathcal{B}(\mathbf{U}) \triangleq \mathbf{U}\mathbf{V}_{\mathrm{f}}$. It can be shown that $\mathcal{B}^*\mathcal{B}$ is positive definite when $\mathbf{V}_{\mathrm{t}}$ has full row rank. Since $\mathcal{A}^*\mathcal{A}$ is also positive semidefinite, the coefficient matrix corresponding to $\mathcal{A}^*\mathcal{A} + \frac{\lambda}{2\alpha}\mathcal{B}^*\mathcal{B}$ is positive definite, guaranteeing a unique solution to (3.20). Since we assumed that the rows of $\mathbf{V}_{\mathrm{t}}$ are orthonormal, it can be shown that the rows of $\mathbf{V}_{\mathrm{f}}$ are also orthonormal, and thus (3.20) can be rewritten as

$$\left(\mathcal{A}^*\mathcal{A} + \frac{\lambda}{2\alpha}\mathcal{I}\right)\mathbf{U}_{\mathrm{s}}^{(\ell+1)} = \mathcal{A}^*(\mathbf{d}) + \frac{\lambda}{2\alpha}\mathbf{G}^{(\ell+1)}\mathbf{V}_{\mathrm{f}}^H. \tag{3.21}$$

Equation (3.21) can be solved efficiently using iterative matrix solvers. Our implementation uses the conjugate gradient (CG) algorithm, initialized with $\mathbf{U}_{\mathrm{s}}^{(\ell)}$ to improve computational speed.

Efficient computation of (3.15) for Cartesian Fourier sampling

If $\mathbf{F}_{\mathrm{s}}$ is orthonormal, as is often the case with Cartesian Fourier sampling, (3.15) can be decoupled to enable much more efficient computation. Specifically, in this case, (3.15) can be converted into

$$\mathbf{U}_{\mathrm{k}}^{(\ell+1)} = \arg\min_{\mathbf{U}_{\mathrm{k}} \in \mathbb{C}^{N\times L}} \|\Omega(\mathbf{U}_{\mathrm{k}}\mathbf{V}_{\mathrm{t}}) - \mathbf{d}\|_2^2 + \frac{\lambda}{2\alpha}\left\|\mathbf{U}_{\mathrm{k}}\mathbf{V}_{\mathrm{f}} - \mathbf{B}^{(\ell+1)}\right\|_{\mathrm{F}}^2, \tag{3.22}$$

where $\mathbf{U}_{\mathrm{s}}^{(\ell+1)} = \mathbf{F}_{\mathrm{s}}^H\mathbf{U}_{\mathrm{k}}^{(\ell+1)}$ and $\mathbf{G}^{(\ell+1)} = \mathbf{F}_{\mathrm{s}}^H\mathbf{B}^{(\ell+1)}$. Equation (3.22) has a decoupled structure, which can be understood by rewriting

$$\Omega(\mathbf{U}_{\mathrm{k}}\mathbf{V}_{\mathrm{t}}) = \begin{bmatrix} \Omega_1(\mathbf{u}_1\mathbf{V}_{\mathrm{t}}) \\ \Omega_2(\mathbf{u}_2\mathbf{V}_{\mathrm{t}}) \\ \vdots \\ \Omega_N(\mathbf{u}_N\mathbf{V}_{\mathrm{t}}) \end{bmatrix}, \tag{3.23}$$

where $\mathbf{u}_n \in \mathbb{C}^{1\times L}$ represents the nth row of $\mathbf{U}_{\mathrm{k}}$, and Ω_n is the sampling operator that takes samples from the nth row of $\mathbf{U}_{\mathrm{k}}\mathbf{V}_{\mathrm{t}}$. The solution to (3.22) can be shown to be equivalent to

$$\mathbf{u}_n^{(\ell+1)} = \arg\min_{\mathbf{u}_n\in\mathbb{C}^{1\times L}} \|\Omega_n(\mathbf{u}_n\mathbf{V}_{\mathrm{t}}) - \mathbf{d}_n\|_2^2 + \frac{\lambda}{2\alpha}\left\|\mathbf{u}_n\mathbf{V}_{\mathrm{f}} - \mathbf{b}_n^{(\ell+1)}\right\|_{\mathrm{F}}^2, \tag{3.24}$$

for $n = 1, \ldots, N$, where $\mathbf{d}_n$ is the measured data from the nth row of $\Omega^*(\mathbf{d})$ and $\mathbf{b}_n^{(\ell)}$ is the nth row of $\mathbf{B}^{(\ell)}$. To obtain the solution to (3.24), we solve

$$\left(\mathcal{A}_n^*\mathcal{A}_n + \frac{\lambda}{2\alpha}\mathcal{I}\right)\mathbf{u}_n^{(\ell+1)} = \mathcal{A}_n^*(\mathbf{d}_n) + \frac{\lambda}{2\alpha}\mathbf{b}_n^{(\ell+1)}\mathbf{V}_{\mathrm{f}}^H, \tag{3.25}$$

for $n = 1, \ldots, N$, where $\mathcal{A}_n(\mathbf{u}) \triangleq \Omega_n(\mathbf{u}\mathbf{V}_{\mathrm{t}})$. Note that each of the normal equations in (3.25) has only L unknowns. Equation (3.25) implies that the coefficient matrix of the linear least squares problem (3.22) can be permuted to have a block diagonal structure. There are a total of N blocks, each of which has $L \times L$ elements. Such a block diagonal matrix guarantees that CG has fast convergence (CG converges to the optimal solution within L iterations [99], assuming exact arithmetic).

3.3.2 Convergence Analysis

In this subsection, we establish the following convergence properties for the proposed algorithm.

- As $\alpha \to 0$, the optimal solution of (3.13) converges to that of (3.8).
- For a fixed α, the sequence $\{(\mathbf{G}^{(\ell)}, \mathbf{U}_{\mathrm{s}}^{(\ell)})\}$ generated by the alternating minimization algorithm converges to an optimal solution of (3.13).
- Some entries of $\mathbf{G}$ have finite convergence, and the other entries of $\mathbf{G}$ and $\mathbf{U}_{\mathrm{s}}$ have q-linear (quotient-linear) convergence rates.

The first property follows from the fact that the continuation procedure belongs to a special case of the quadratic penalty method and thus has guaranteed global convergence (see Proposition 4.2.1 in [95]). The second and third convergence properties of the alternating

minimization procedure (i.e., (3.18) and (3.20)) for solving (3.13) are established in the following. We first convert (3.18) and (3.20) into a similar form to that used in [55], and then establish the nonexpansiveness of the linear operators in the alternating minimization. Based on these nonexpansiveness properties, we apply the theorems in [55] to obtain the finite and q-linear convergence results.

For convenience, the proposed alternating minimization for solving (3.13) is first summarized below:

$$\mathbf{G}^{(\ell+1)} = S(\mathbf{U}_{\mathrm{s}}^{(\ell)}\mathbf{V}_{\mathrm{f}}), \tag{3.26}$$

$$\mathbf{U}_{\mathrm{s}}^{(\ell+1)} = \left(\mathcal{A}^*\mathcal{A} + \frac{\lambda}{2\alpha}\mathcal{I}\right)^{-1}\left(\mathcal{A}^*(\mathbf{d}) + \frac{\lambda}{2\alpha}\mathbf{G}^{(\ell+1)}\mathbf{V}_{\mathrm{f}}^H\right). \tag{3.27}$$

We next convert (3.26) and (3.27) to a form similar to (3.4) and (3.5) in [55]. First, the matrix representation of $\mathcal{A}$ is denoted by $\mathbf{K} \in \mathbb{C}^{P\times NL}$, i.e.,

$$\mathrm{vec}(\mathcal{A}(\cdot)) = \mathbf{K}\mathrm{vec}(\cdot). \tag{3.28}$$

Second, we have

$$\mathrm{vec}\left(\left(\mathcal{A}^*\mathcal{A} + \frac{\lambda}{2\alpha}\mathcal{I}\right)(\cdot)\right) = \left(\mathbf{K}^H\mathbf{K} + \frac{\lambda}{2\alpha}\mathbf{I}_{NL}\right)\mathrm{vec}(\cdot). \tag{3.29}$$

Let $\mathbf{D} = (\mathbf{V}_{\mathrm{f}}^T \bigotimes \mathbf{I}_N) \in \mathbb{C}^{MN\times NL}$, where $\bigotimes$ is the Kronecker product of two matrices. Based on the property of the Kronecker product and the fact that $\mathbf{V}_{\mathrm{f}}\mathbf{V}_{\mathrm{f}}^H = \mathbf{I}_L$, it can be shown that the columns of $\mathbf{D}$ are orthonormal, i.e.,

$$\mathbf{D}^H\mathbf{D} = \mathbf{I}_{NL}. \tag{3.30}$$

Finally, we introduce the following two positive definite matrices,

$$\mathbf{M} = \mathbf{K}^H\mathbf{K} + \frac{\lambda}{2\alpha}\mathbf{D}^H\mathbf{D} = \mathbf{K}^H\mathbf{K} + \frac{\lambda}{2\alpha}\mathbf{I}_{NL} \tag{3.31}$$

and

$$\mathbf{T} = \frac{\lambda}{2\alpha}\mathbf{D}\mathbf{M}^{-1}\mathbf{D}^H. \tag{3.32}$$

It can be easily shown that the maximum singular value of $\mathbf{T}$ is less than 1.

Based on the above notation, (3.26) and (3.27) can be rewritten as follows:

$$\mathrm{vec}(\mathbf{G}^{(\ell+1)}) = s(\mathrm{vec}(\mathbf{U}_{\mathrm{s}}^{(\ell)}\mathbf{V}_{\mathrm{f}})) = s \circ h(\mathrm{vec}(\mathbf{G}^{(\ell)})), \tag{3.33}$$

$$\mathrm{vec}(\mathbf{U}_{\mathrm{s}}^{(\ell+1)}) = \mathbf{M}^{-1}\left(\frac{\lambda}{2\alpha}\mathbf{D}^H\mathrm{vec}(\mathbf{G}^{(\ell)}) + \mathbf{K}^H\mathbf{d}\right), \tag{3.34}$$

where s performs element-by-element soft-thresholding of a vector with the threshold value of α, $h(\mathbf{v}) = \mathbf{D}\mathbf{M}^{-1}(\frac{\lambda}{2\alpha}\mathbf{D}^H\mathbf{v} + \mathbf{K}^H\mathbf{d})$ for $\forall\ \mathbf{v} \in \mathbb{C}^{NM\times 1}$, and $\circ$ represents the composition of two operators.

The above alternating minimization procedures (4.9) and (3.34) have the same structure as (3.4) and (3.5) in [55]. Furthermore, we can establish that s and h in (4.9) have the same nonexpansiveness properties as those in Proposition 3.1 and 3.2 in [55], which were used to prove the following theorems for convergence. Specifically, for s, we have that $|s(\mathbf{a})_i - s(\mathbf{b})_i| \leq |\mathbf{a}_i - \mathbf{b}_i|$, for $i = 1, \ldots, NM$ and $\forall\ \mathbf{a}, \mathbf{b} \in \mathbb{C}^{NM\times 1}$, and $|s(\mathbf{a})_i - s(\mathbf{b})_i| = |\mathbf{a}_i - \mathbf{b}_i|$ if and only if $s(\mathbf{a})_i - s(\mathbf{b})_i = \mathbf{a}_i - \mathbf{b}_i$; for h, $\|h(\mathbf{w}) - h(\tilde{\mathbf{w}})\|_2 \leq \|\mathbf{w} - \tilde{\mathbf{w}}\|_2$, $\forall\ \mathbf{w}, \tilde{\mathbf{w}} \in \mathbb{C}^{NM\times 1}$, and $\|h(\mathbf{w}) - h(\tilde{\mathbf{w}})\|_2 = \|\mathbf{w} - \tilde{\mathbf{w}}\|_2$ if and only if $h(\mathbf{w}) - h(\tilde{\mathbf{w}}) = \mathbf{w} - \tilde{\mathbf{w}}$.

As a result, the global convergence property for the proposed method can be established based on Theorem 3.4 of [55]. Let $(\mathbf{G}^{\infty}, \mathbf{U}_{\mathrm{s}}^{\infty})$ denote an optimal solution of (3.13).

Theorem 3.3.1 (Theorem 3.4 [55]). *For any fixed $\alpha > 0$, the sequence $\{(\mathbf{G}^{(\ell)}, \mathbf{U}_{\mathrm{s}}^{(\ell)})\}$ generated by* (4.9) *and* (3.34) *from any starting point $(\mathbf{G}^{(0)}, \mathbf{U}_{\mathrm{s}}^{(0)})$ converges to $(\mathbf{G}^{\infty}, \mathbf{U}_{\mathrm{s}}^{\infty})$ of* (3.13).

Furthermore, we can apply Theorems 3.5 and 3.6 in [55] to characterize the convergence rate of $\mathbf{G}$ and $\mathbf{U}_{\mathrm{s}}$. The following theorems will make use of two new index sets defined as:

$$I = \{i : |\mathrm{vec}(\mathbf{U}_{\mathrm{s}}^{\infty}\mathbf{V}_{\mathrm{f}})_i| = |h(\mathrm{vec}(\mathbf{G}^{\infty}))_i| < \alpha\} \text{ and } E =$$
$$\{1, \ldots, NL\}\backslash I, \text{where } \backslash \text{ represents a set difference operation.}$$

Theorem 3.3.2 (Theorem 3.5 [55]). *For the sequence $\{(\mathbf{U}_{\mathrm{s}}^{(\ell)}, \mathbf{G}^{(\ell)})\}$ generated by* (4.9) *and* (3.34) *from any starting point $(\mathbf{U}_{\mathrm{s}}^{(0)}, \mathbf{G}^{(0)})$, there exists $\zeta_0 \leq \|\mathbf{G}^{(0)} - \mathbf{G}^{\infty}\|_{\mathrm{F}}^2/\omega^2 < +\infty$ such*

that $\forall i \in I$, $\mathrm{vec}(\mathbf{G}^{(\ell)})_i = \mathrm{vec}(\mathbf{G}^{\infty})_i$ *as long as* $\ell \geq \zeta_0$, *where*

$$\omega \triangleq \min_{i \in I} \{\alpha - |h(\mathrm{vec}(\mathbf{G}^{\infty}))_i|\} > 0.$$

Theorem 3.3.3 (Theorem 3.6 [55]). *Let* $\mathbf{T}_{EE} = [\mathbf{T}_{i,j}]_{i,j \in E}$ *be a* $|E| \times |E|$ *submatrix of* $\mathbf{T}$ *whose entries are taken from* $\mathbf{T}$ *according to E. The sequence* $\{(\mathbf{U}_{\mathrm{s}}^{(\ell)}, \mathbf{G}^{(\ell)})\}$ *generated by* (4.9) *and* (3.34) *satisfies*

$$\|\mathrm{vec}(\mathbf{G}^{(\ell+1)} - \mathbf{G}^{\infty})_E\|_2 \leq c_1 \|\mathrm{vec}(\mathbf{G}^{(\ell)} - \mathbf{G}^{\infty})_E\|_2,$$

$$\|\mathbf{M}^{\frac{1}{2}} \mathrm{vec}(\mathbf{U}_{\mathrm{s}}^{(\ell+1)} - \mathbf{U}_{\mathrm{s}}^{\infty})\|_2 \leq c_2 \|\mathbf{M}^{\frac{1}{2}} \mathrm{vec}(\mathbf{U}_{\mathrm{s}}^{(\ell)} - \mathbf{U}_{\mathrm{s}}^{\infty})\|_2,$$

for $\ell \geq \zeta_0$, *where* $c_1 = \sqrt{\sigma_{\max}((\mathbf{T}^2)_{EE})}$, $c_2 = \sqrt{\sigma_{\max}((\mathbf{T})_{EE})}$, $\sigma_{\max}(\cdot)$ *represents the maximum singular value of a matrix, and* $\mathbf{M}^{\frac{1}{2}}$ *is any positive semidefinite Hermitian square root of the Hermitian matrix* $\mathbf{M}$, *i.e.,* $\mathbf{M} = \mathbf{M}^{\frac{1}{2}}(\mathbf{M}^{\frac{1}{2}})^H = (\mathbf{M}^{\frac{1}{2}})^2$.

3.4 Application to Real-Time Cardiac Imaging

3.4.1 Simulation Results

A complex-valued numerical cardiac phantom was used for our simulation studies. The simulation was designed to emulate single-channel, real-time cardiac MR experiments with variable heart rate and variable respiratory rate. The phantom was created from real human cardiac MR data which were collected using retrospective ECG-gating during a single breath-hold and used to generate a time series of images representing a single prototype cardiac cycle. We applied multiple time-warps to this prototype to simulate heart rate variability, and the individual time-warped heart-beats were concatenated together to form a long image sequence spanning multiple cardiac cycles. Subsequently, respiration was modeled by applying an additional quasi-periodic spatial deformation [100] to this image sequence. The simulated cardiac and respiratory variability were consistent with statistics from the PhysioBank archive [101]. Acquisition parameters for our simulations included: acquisi-

tion matrix size = 200 × 256, field of view (FOV) = 273 mm × 350 mm, effective spatial resolution = 1.36 mm × 1.36 mm, slice thickness = 6 mm, and $T_{\mathrm{R}} = 3$ ms.

A Cartesian Fourier-based sampling scheme was used for our simulations, part of which is shown in Fig. 3.1(a). We acquire one **k**-space line per T_{R} in an interleaved fashion, alternating between sampling one line from central **k**-space and sampling one line from outer **k**-space. The lines from central **k**-space are acquired in a sequential order, while the lines acquired from outer **k**-space follow a random order. Let N_{c} and N_{p} respectively denote the number of lines for central and complete **k**-space. With this $(\mathbf{k}, t)$-space sampling pattern, one full set of **k**-space lines (called a "data frame") is acquired in a time interval of $2(N_{\mathrm{p}} - N_{\mathrm{c}})T_{\mathrm{R}}$. The total acquisition time T_{a} is $2N_{\mathrm{df}}(N_{\mathrm{p}} - N_{\mathrm{c}})T_{\mathrm{R}}$, with N_{df} being the total number of data frames.

In image reconstruction, we assume that the entire data acquisition interval can be divided into subintervals of length T_{m}, as illustrated in Fig. 3.1(b), during which temporal signal variations are negligible. Under this assumption, the measured data within each T_{m} interval can be treated as an instantaneous snapshot. This kind of temporal modeling is commonly used in real-time imaging, although other types of temporal modeling, such as the bandlimited model in [32], could also be used.

Using the above sampling and temporal modeling schemes, the measured data can be partitioned into two sets. One set contains the fully sampled central **k**-space data with high temporal resolution, and the other contains the sparsely sampled data with high spatial resolution. We determine $\mathbf{V}_{\mathrm{t}}$ from all the data in the fully sampled region of **k**-space as described in Section 3.2.2. Note that with this scheme for sampling and $\mathbf{V}_{\mathrm{t}}$ estimation, $\mathbf{V}_{\mathrm{t}}$ is automatically synchronized temporally with all the data used for image reconstruction. Also $\mathbf{V}_{\mathrm{t}}$ can represent general temporal variations (e.g., periodic or aperiodic) in the data.

Throughout all the simulations, we used $N_{\mathrm{c}} = 8$ and $N_{\mathrm{p}} = 200$ for data acquisition. We used $T_{\mathrm{m}} = 2N_{\mathrm{c}}T_{\mathrm{R}} = 16T_{\mathrm{R}}$ for the temporal modeling. As a consequence, 16 readout lines (i.e., 8 percent of the full k-space data) were measured at each reconstructed time point, and the reconstructions were performed at a frame rate of 21 images per second (i.e., an effective temporal resolution of 48 ms).

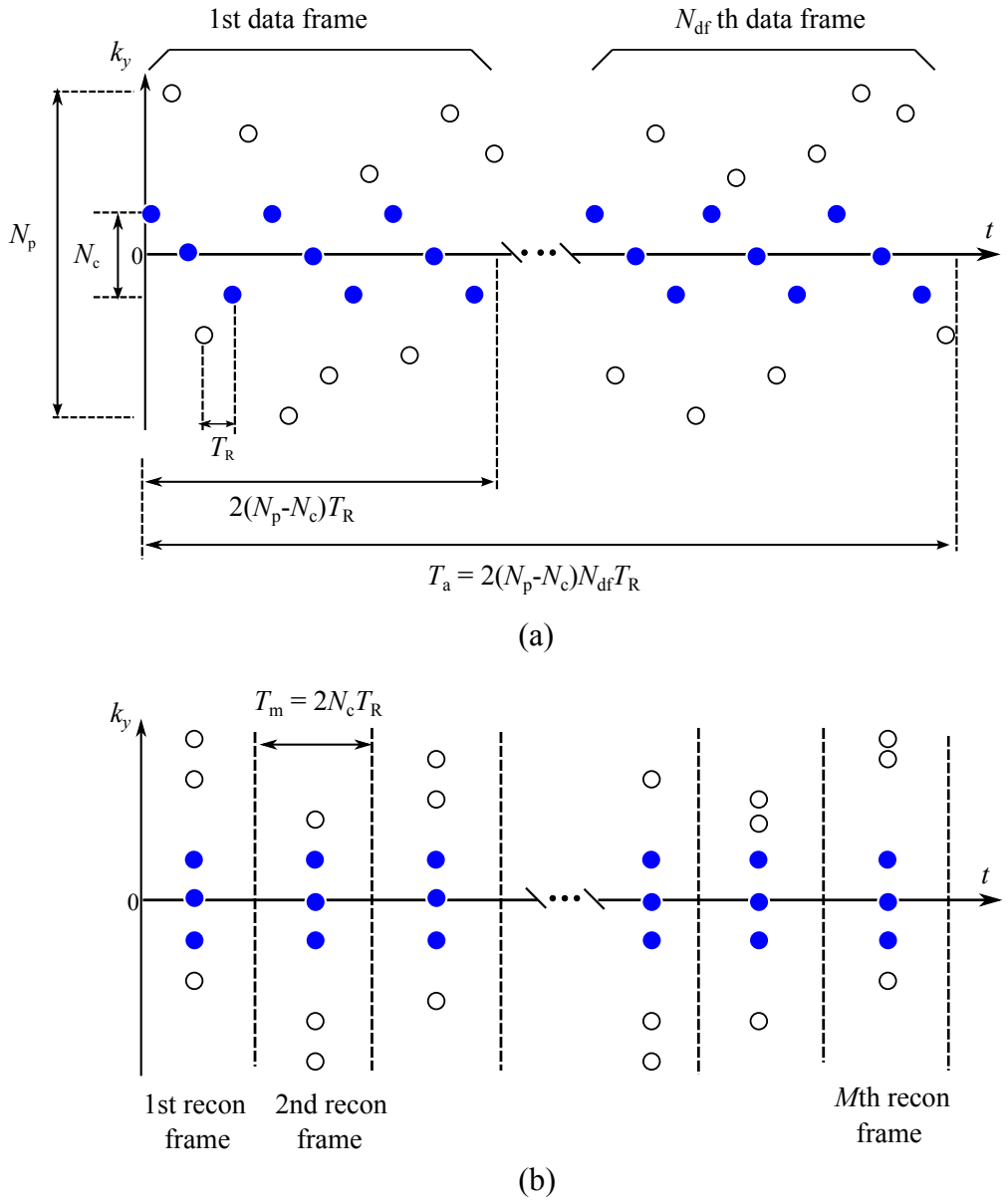


Figure 3.1: (a) $(\mathbf{k}, t)$-space sampling pattern and (b) temporal modeling used for image reconstruction. As shown in (a), we acquire one k-space line per T_R in an interleaved fashion, alternating between sampling one line from central **k**-space and one line from outer **k**-space. The lines acquired from central **k**-space follow a sequential order, while the lines acquired from outer **k**-space follow a random order. Using this sampling pattern, it takes $2(N_p - N_c)T_R$ to collect a full set of **k**-space lines, denoted as a data frame. The total data acquisition time T_a is $2(N_p - N_c)T_R N_{df}$ with a total number of N_{df} data frames. In image reconstruction, we assume that the entire data acquisition interval can be subdivided into intervals of length T_m, shown by dashed lines in (b), during which temporal signal variations are negligible.

General evaluation

We compared PS-Sparse with Basic-PS and Basic-Sparse, using simulated data for $T_{\mathrm{a}} = 35.3$ s (i.e., $N_{\mathrm{df}} = 32$). We performed Basic-PS reconstruction for $1 \leq L \leq 32$. With the proposed sampling scheme, the Basic-PS solution is unique for these values of L, since N_{df} (equal to the minimum number of times each $\mathbf{k}$-space line is sampled) is always greater than or equal to L. We also performed PS-Sparse reconstructions using the same data set for $1 \leq L \leq 800$. Note that in our simulation setup, $L = M = T_{\mathrm{a}}/T_{\mathrm{m}} = 800$ is the full model order, for which PS-Sparse corresponds to Basic-Sparse. We manually chose the regularization parameter λ for PS-Sparse at every 10 model orders (including the full model order) to minimize the following reconstruction error:

$$\text{error} = \frac{\left\| \mathbf{C}_{\mathrm{g}} - \hat{\mathbf{U}}_{\mathrm{s}} \mathbf{V}_{\mathrm{t}} \right\|_{\mathrm{F}}}{\| \mathbf{C}_{\mathrm{g}} \|_{\mathrm{F}}}, \tag{3.35}$$

where $\mathbf{C}_{\mathrm{g}} \in \mathbb{C}^{N \times M}$ is the gold standard.

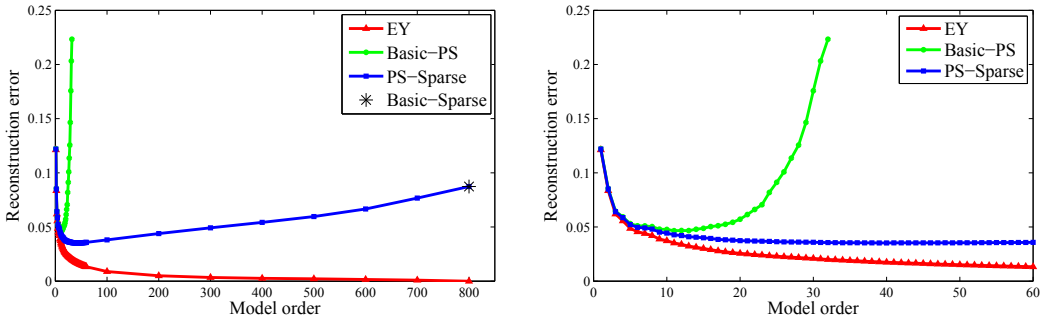


Figure 3.2: (a) Reconstruction error versus L (ranging from 1 to the full model order) with $N_{\mathrm{df}} = 32$ as calculated using (3.35), and (b) zoomed-in plot showing model orders from 1 to 60. Note that PS-Sparse is significantly more robust to changes in model order than Basic-PS. PS-Sparse also has smaller reconstruction error than Basic-Sparse except when the full model order is adopted (when PS-Sparse becomes Basic-Sparse).

As a comparison, Eckart-Young (EY) approximation [99], defined as

$$\hat{\mathbf{E}}_L = \arg \min_{\mathrm{rank}(\mathbf{E}) \leq L} \| \mathbf{E} - \mathbf{C}_{\mathrm{g}} \|_{\mathrm{F}}^2, \tag{3.36}$$

with L ranging from 1 to 800 was also computed from the gold standard using SVD. The rank-L EY approximation represents the best possible performance that any order-L PS model can achieve. The error of the EY approximations reflects the underlying model mismatch between the PS model and the ground truth.

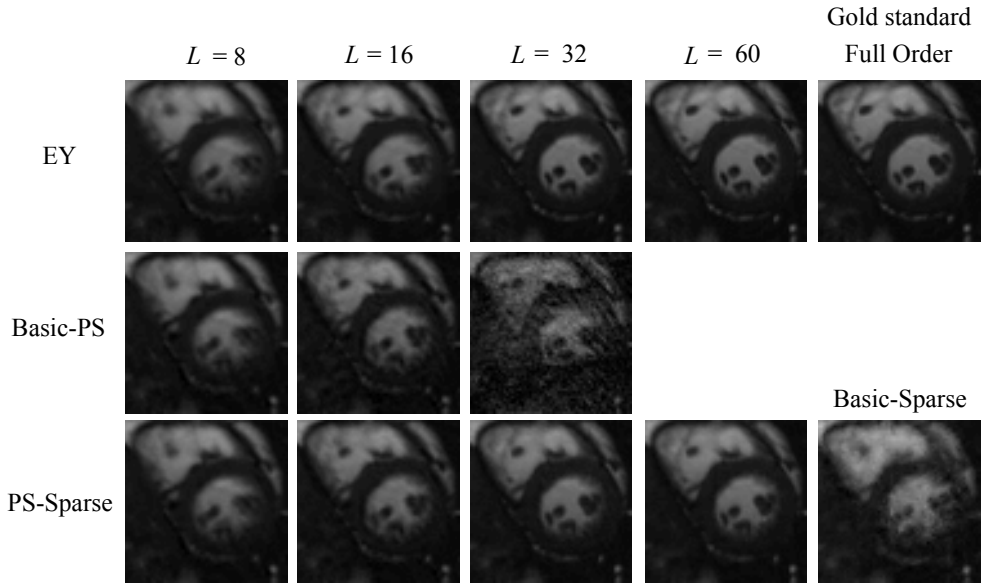


Figure 3.3: Reconstructions from the mid-systolic cardiac phase (zoomed in on the heart). Note that rank-L EY provides the best performance that Lth order Basic-PS and PS-Sparse could achieve, and that full-order EY is equivalent to the gold standard. Note the improvements of PS-Sparse reconstructions over Basic-PS reconstructions and the robustness of PS-Sparse to different model orders.

The reconstruction errors for these simulations are shown in Fig. 3.2. Note that when L is small, both Basic-PS and PS-Sparse have large reconstruction errors, due to the limited capability of the low-order PS models. As L increases, the reconstruction accuracy of Basic-PS and PS-Sparse improves, as expected. However, as we continue to increase L (e.g., beyond 16), the Basic-PS reconstructions quickly deteriorate because the inverse problem becomes increasingly ill-conditioned. This ill-conditioning problem is nicely overcome by PS-Sparse, with sparsity as a regularizer, as shown in Fig. 3.2 (b). As L approaches the full model, PS-Sparse reduces to Basic-Sparse and the reconstruction error becomes worse.

To illustrate the reconstruction improvement of PS-Sparse over Basic-PS and Basic-Sparse, a set of mid-systolic cardiac snapshots are shown in Fig. 3.3; the corresponding temporal variations from a vertical line through the left ventricle are shown in Fig. 3.4. At very low

model orders (such as $L = 8$), the PS constraint is too restrictive and both Basic-PS and PS-Sparse reconstructions suffer from severe blurring (e.g., in the endocardial border and papillary muscles). Increasing the model order to 16 benefits both reconstruction methods. The advantages of PS-Sparse over Basic-PS become clearer as the model order further increases. For example, at $L = 32$, the Basic-PS reconstruction has severe artifacts due to ill-conditioning, which are removed in the PS-Sparse reconstruction. The PS-Sparse reconstruction remains stable for much higher model orders (e.g., $L = 60$). When L equals the full model order, the PS-Sparse reconstruction reduces to the Basic-Sparse reconstruction, and image artifacts, such as spatiotemporal blurring, appear.

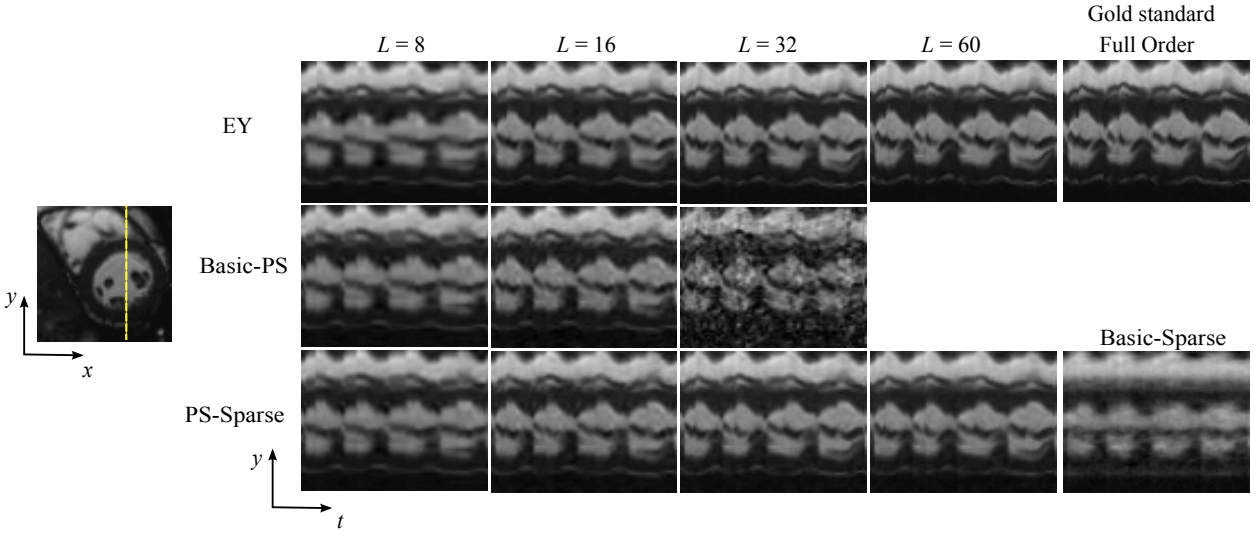


Figure 3.4: The temporal variations of a vertical line passing through the left ventricle (zoomed in on the heart), corresponding to reconstructions in Fig. 3.3. Note the ill-conditioning problem with Basic-PS with $L = 32$, and the blurring artifacts in the Basic-Sparse reconstruction.

To complement the results in Figs. 3.3 and 3.4, reconstructions of four different cardiac phases, ranging from end-systole to end-diastole in a cardiac cycle, are shown in Fig. 3.5. As can be seen, the PS-Sparse reconstructions are consistently better than the other reconstructions across all cardiac phases.

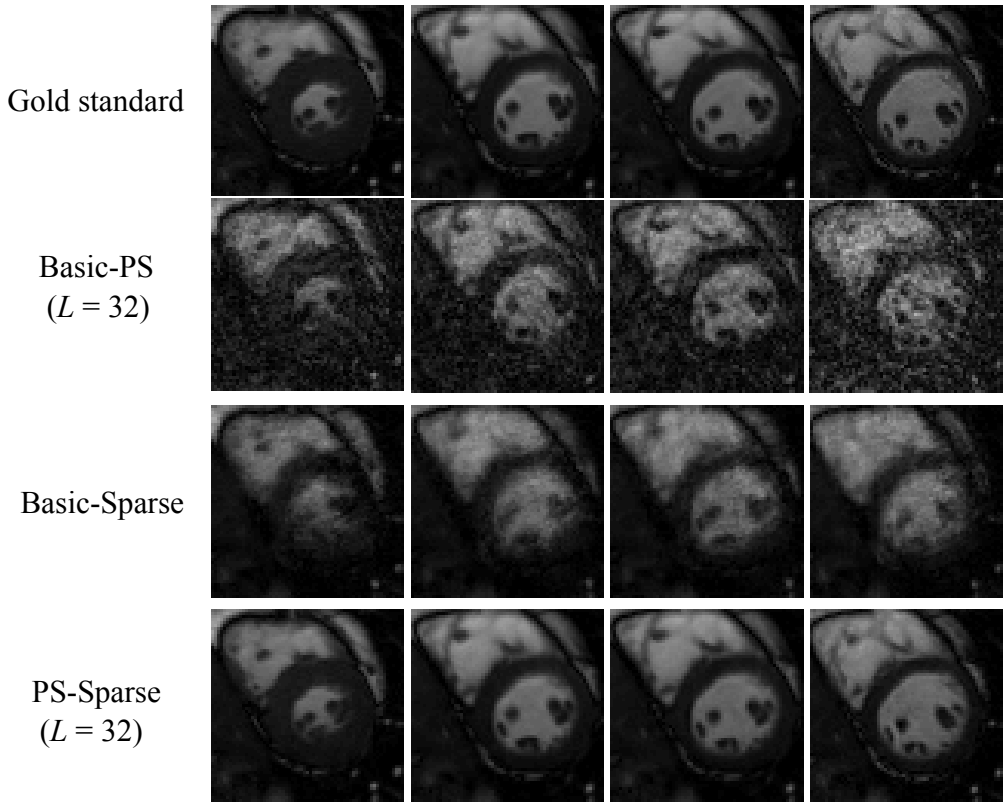


Figure 3.5: Reconstructions in four cardiac phases ranging from end-systole to end-diastole based on the same measurements used in Fig. 3.3. To avoid repetition, four time points shown here are different from the single time point shown in Fig. 3.3. Note the improved reconstruction quality of PS-Sparse over Basic-PS and Basic-Sparse.

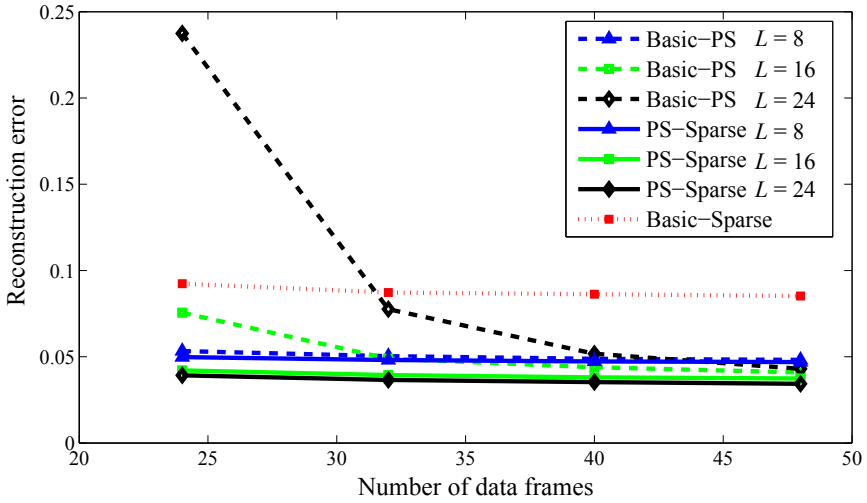


Figure 3.6: Reconstruction error versus the number of data frames for Basic-PS (dashed lines), Basic-Sparse (red dotted line), and PS-Sparse (solid lines). Note that PS-Sparse has lower reconstruction error than either Basic-PS (with the same model order) or Basic-Sparse at each data acquisition window length. In addition, the reconstruction error of PS-Sparse is relatively robust to N_{df} for both high and low model orders.

Influence of T_{a}

We have also evaluated PS-Sparse with respect to different data acquisition window lengths, which provides useful insight into the performance of PS-Sparse in accelerating imaging experiments. The simulation study was done with $T_{\mathrm{a}} = 26.5,\ 35.3,\ 44.2$, and 53.0 s (i.e., $N_{\mathrm{df}} = 24,\ 32,\ 40$ and 48). Image reconstruction was done using Basic-PS, Basic-Sparse, and PS-Sparse, with $L = 8,\ 16$, and 24 for Basic-PS and PS-Sparse. As before, we manually selected the regularization parameter of each method for each data acquisition window length to optimize their performance. Figure 3.6 shows the reconstruction errors as a function of N_{df}. The reconstruction errors of Basic-PS of all three model orders improve as N_{df} increases, as expected, because longer data acquisition windows resulted in better conditioning of the PS model fitting. This effect becomes more pronounced for higher-order PS models (e.g., $L = 24$). For Basic-Sparse, since the ratio of the number of measurements to the number of reconstructed spatiotemporal voxels remains the same for different acquisition window lengths, the reconstruction error of Basic-Sparse stays at about the same level for different values of N_{df}. For PS-Sparse, the reconstruction error is relatively less sensitive to N_{df} for

both high and low model orders, and is lower than that of either Basic-PS (with matched model order) or Basic-Sparse for a given N_{df}.

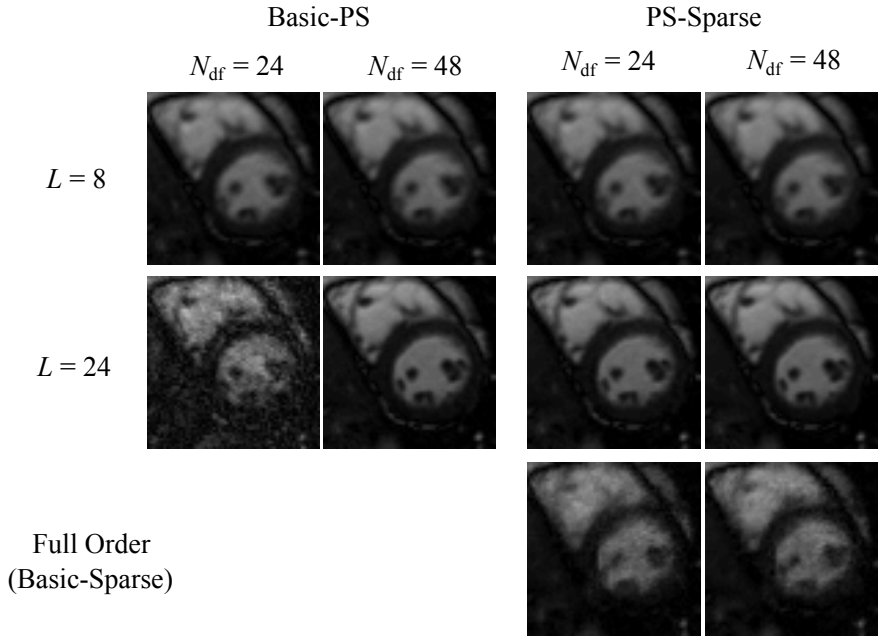


Figure 3.7: Reconstructions of the mid-diastolic cardiac phase (zoomed in on the heart) using $N_{\mathrm{df}} = 24$ and 48. Note the superior performance of PS-Sparse in the presence of limited measurements and also its robustness to the number of measurements.

Figure 3.7 shows the reconstructions of one mid-diastolic phase using $N_{\mathrm{df}} = 24$ and 48. When $L = 8$, the PS constraint causes significant blurring in both Basic-PS and PS-Sparse reconstructions, which did not improve as the length of the data acquisition window increased. For $L = 24$, Basic-PS reconstructed a higher quality image, though this only occurred with a relatively long data acquisition ($N_{\mathrm{df}} = 48$). With shorter data acquisition windows, the Basic-PS reconstructions have serious artifacts due to ill-conditioning problem. In contrast, PS-Sparse with the same model order reconstructs higher-quality images at both $N_{\mathrm{df}} = 24$ and 48. Although the Basic-Sparse reconstructions are relatively stable for $N_{\mathrm{df}} = 24$ and 48, they both suffer from significant blurring artifacts. To better illustrate the image artifacts, we show the error images of these reconstructions in Fig. 3.8, which were obtained by subtracting the reconstructions from the gold standard.

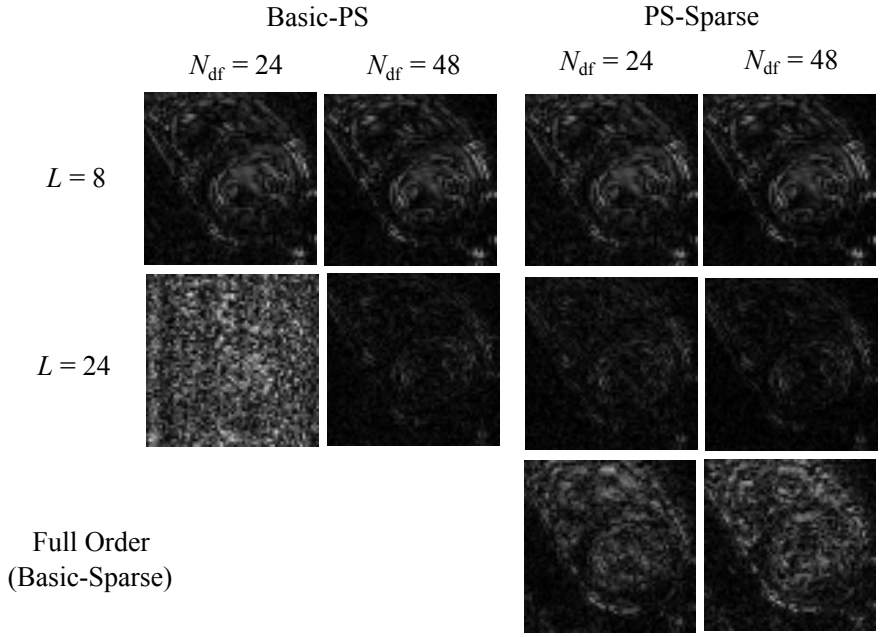


Figure 3.8: Error images for the reconstructions in Fig. 3.7, which were obtained by subtracting the reconstructions from the gold standard. Note that the error images have been scaled up for visualization purpose (the maximum value of the error images is one seventh of that in Fig. 3.7).

Comparison with implicit PS reconstruction

Lingala et al. recently proposed to use implicit PS constraints based on the Schatten p-norm along with a sparsity constraint for dynamic image reconstruction [85, 102]. The reconstruction problem was formulated as

$$\bar{\mathbf{C}} = \arg \min_{\mathbf{C} \in \mathbb{C}^{N \times M}} \|\mathbf{d} - \Omega(\mathbf{F}_{\mathrm{s}}\mathbf{C})\|_2^2 + \lambda_{\mathrm{s}} \|\mathrm{vec}(\mathbf{C}\mathbf{F}_{\mathrm{t}})\|_1 + \lambda_{\mathrm{r}} \|\mathbf{C}\|_p^p , \tag{3.37}$$

where $p = 1$ was used in [102], and $p < 1$ in [85].

We compared PS-Sparse with the above two methods, denoted as NN-Sparse and Sch-Sparse, respectively. We used $(\mathbf{x}, f)$-domain sparsity constraints for our comparisons for the sake of consistency, noting however that [85] used an alternative sparsity constraint based on spatiotemporal smoothness. The data was simulated with $T_{\mathrm{a}} = 26.5$ s (i.e., $N_{\mathrm{df}} = 24$). For PS-Sparse, we used $L = 32$ and manually selected λ. For NN-Sparse and Sch-Sparse, we manually chose the regularization parameters λ_{s} and λ_{r} to optimize their performance. We chose $p = 0.1$ for Sch-Sparse. Note that Sch-Sparse is nonconvex and subject to local

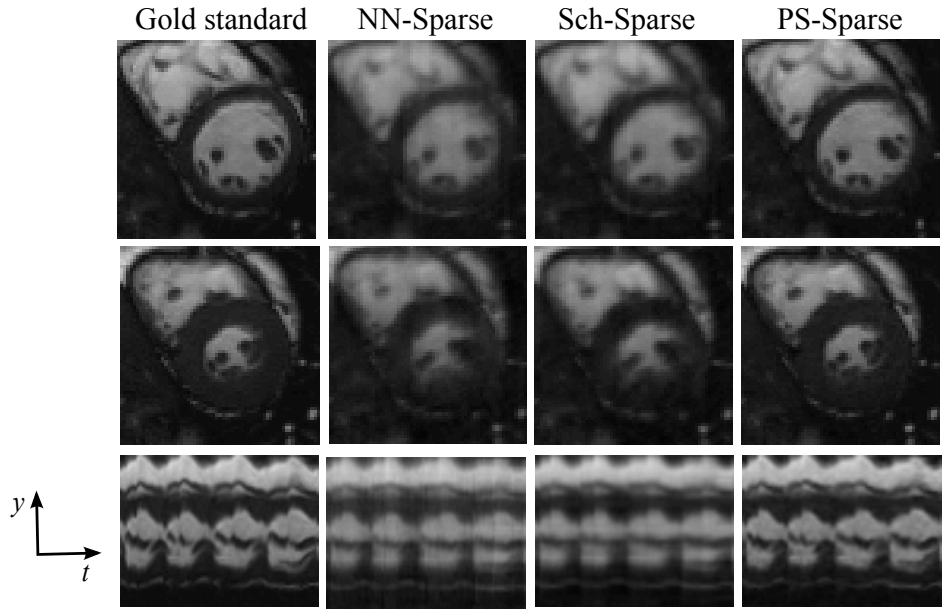


Figure 3.9: Comparison of NN-Sparse, Sch-Sparse, and PS-Sparse. The first two rows show end-diastolic and end-systolic cardiac images, and the last row shows the reconstructed temporal variations of a line passing through the left ventricle. Note that the PS-Sparse reconstructions have higher spatial and temporal fidelity than the NN-Sparse and Sch-Sparse reconstructions (e.g., the myocardial border and the motion of the papillary muscles are much better defined in the PS-Sparse reconstructions than in the NN-Sparse and Sch-Sparse counterparts).

minima, but that our results were insensitive to the initialization in this case. Figure 3.9 shows reconstructions from the end-diastolic and end-systolic cardiac phases. Although Sch-Sparse has slightly better performance over NN-Sparse, its reconstructed spatial and temporal images still suffer from significant blurring and artifacts. In contrast, the corresponding PS-Sparse reconstructions show higher spatial and temporal fidelity than both the NN-Sparse and Sch-Sparse reconstructions.

The superior performance of PS-Sparse can be attributed to the use of the explicit PS constraints. Since the explicit PS constraints have fewer degrees of freedom than the implicit PS constraints, the low-rank structure was enforced more strongly. It has been observed empirically that low-rank matrix recovery using explicit PS constraints can be more effective than the nuclear norm constraint when the number of measurements is small [103–105].

Computationally, PS-Sparse is more efficient than NN-Sparse and Sch-Sparse, especially in the case of Cartesian Fourier sampling. In addition to using explicit PS, PS-Sparse also employs a pre-determined temporal subspace, resulting in a much smaller number of degrees of freedom and a significantly simplified computational problem. In addition, the proposed solution algorithm avoids computation-intensive SVD evaluations used in the implementations of NN-Sparse [102] and Sch-Sparse [85]. We implemented these methods on a 2×Intel® Quad-Core Xeon® 3.16GHz workstation with 48GB RAM running MATLAB® R2011b on a Linux platform. In the experiments described above, PS-Sparse took 11 minutes while NN-Sparse and Sch-Sparse each took more than 2 hours based on the implementations described in [85, 102]. However, it should be noted that a faster augmented Lagrangian algorithm has been implemented on graphics processing units [106], leading to significant reductions in computation time compared to the implementations in [85, 102]. Our proposed method can also benefit from similar implementation, although in-depth evaluation of computational efficiency are beyond the scope of this proposal.

3.4.2 In Vivo Results

In this section, we show some representative results of applying PS-Sparse to real-time cardiac imaging. Experimental data were collected from rats on a Bruker (Billerica, MA)

Avance AV1 4.7 T magnet, equipped with a 40 G/cm shielded gradient set and a 4-channel array coil. A customized FLASH pulse sequence was used for data acquisition. Acquisition parameters for *in vivo* experiments included: $T_{\mathrm{R}} = 7.5$ ms, $T_{\mathrm{E}} = 2.4$ ms, acquisition matrix size = 128 × 128, FOV = 40 mm × 40 mm, effective spatial resolution = 0.31 mm × 0.31 mm, and slice thickness = 1.5 mm. A similar sampling pattern to Fig. 3.1 (a) was used with $N_{\mathrm{c}} = 1$. All experiments were performed in compliance with federal and institutional regulations and guidelines.

Image reconstruction was done using a sliding window method [107], Basic-PS, Basic-Sparse, and PS-Sparse. For Basic-PS, Basic-Sparse and PS-Sparse, we used $T_{\mathrm{m}} = 2T_{\mathrm{R}}$, (i.e., two readout lines for each reconstructed time point). The reconstructions were performed at a frame rate of 67 images per second. We manually chose $L = 16$ for both Basic-PS and PS-Sparse based on empirical evaluation of the reconstruction quality. Specifically, an initial estimate of the model order was selected using the method described in [108]. We performed image reconstruction for a range of model orders about this initial estimate, and chose the final model order based on visual evaluation of reconstruction quality. The regularization parameters for Basic-Sparse and PS-Sparse were chosen based on the discrepancy principle [109]. The noise variance for the discrepancy principle was estimated from the variance of the signal in high-frequency regions of $\mathbf{k}$-space where measurement noise is dominant [110].

Some representative reconstruction results are shown in Fig. 3.10. The sliding window (SW) reconstruction is included as a reference to illustrate the level of undersampling in $(\mathbf{k}, t)$-space as well as the noise level of the measured data. For Basic-PS and Basic-Sparse, the reconstructions suffer from poor conditioning and blurring artifacts respectively, which is consistent with our simulation results. As can be seen from the last row of Fig. 3.10, PS-Sparse produced much improved reconstructions in both magnitude and phase.

3.5 Discussion

The proposed method makes use of specialized data acquisition and processing to pre-estimate the temporal subspace of $\mathbf{C}$. It results in a simplified problem that can be solved efficiently. Pre-estimating the temporal subspace from a subset of the measured data (e.g.,

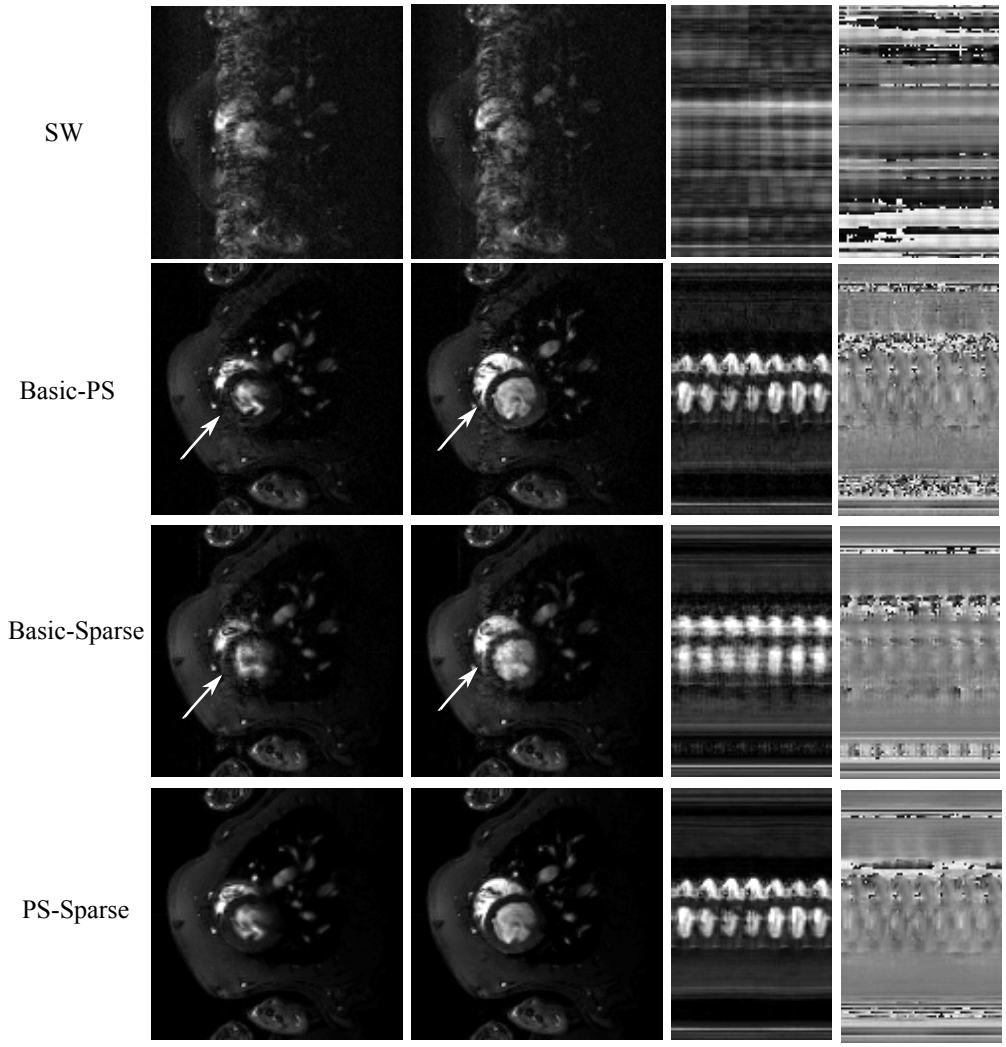


Figure 3.10: Reconstructions from an *in vivo* real-time cardiac MR experiment on rats. The first two columns show the reconstructions in two cardiac phases, and the last two columns show the reconstructed magnitude and phase of temporal variations for a line passing through the left ventricle. Consistent with the simulation results, PS-Sparse shows improved performance over Basic-PS and Basic-Sparse (e.g., regions marked by arrows). Furthermore, the PS-Sparse reconstructions are also capable of capturing the phase changes in dynamic image sequences.

a number of fully sampled central **k**-space points) is based on the assumption that this set of data and $\mathbf{C}$ share the same temporal subspace. In addition to the results shown in this chapter, this strategy has been successfully used in a number of other dynamic MRI contexts [2, 81, 111]. However, it is worth noting that the use of the PS/low-rank model itself does not require pre-estimation of $\mathbf{V}_\mathrm{t}$, and reconstruction accuracy can be further improved by using all, rather than just a subset, of the measured data to determine the temporal subspace [30, 32, 112]. The proposed method can be extended to simultaneously estimate $\mathbf{U}_\mathrm{s}$ and $\mathbf{V}_\mathrm{t}$ from all of the measured data. For example, the PowerFactorization algorithm can be used to perform this extension [30, 32, 103, 112], in which $\mathbf{U}_\mathrm{s}$ and $\mathbf{V}_\mathrm{t}$ are estimated in an alternating manner. Some preliminary investigations indicate that this approach can further improve robustness to low-rank modeling error and measurement noise, although the solution is computationally much more challenging.

Although [92, 93] are also based on the use of low-rank and sparsity constraints, they are associated with a different image model, in which the dynamic image is represented as a linear combination of a low-rank and a sparse component, i.e., a robust principal component analysis (RPCA) based model. In contrast, the proposed model here assumes that the dynamic image has both low-rank and sparsity characteristics simultaneously. In general, the proposed model has fewer degrees of freedom than the RPCA-based model, and thus it is expected to yield better performance than [92,93] for applications with highly undersampled data.

The real-time cardiac imaging examples shown in this chapter made use of $(\mathbf{x}, f)$-domain sparsity, which is a useful constraint for real-time cardiac imaging. Extending the proposed method to other types of sparsity constraints is mathematically straightforward. For example, we can reformulate the reconstruction problem with joint PS and spatial total variation (TV) constraints. The resulting optimization problem can be solved using the algorithm in [34]. Although the TV constraints did not necessarily improve reconstruction results in that case, further improvement could potentially be obtained with other sparsifying transforms.

Both PS-Sparse and Basic-PS use an explicit model order for image reconstruction. Generally, choice of the model order L must be made to balance representational power, the

required number of measurements, and noise sensitivity. Because of the additional sparsity constraints, PS-Sparse demonstrates improved robustness to the choice of the model order compared to Basic-PS. In practice, this makes model-order selection easier for PS-Sparse than for Basic-PS. However, in the absence of ground truth, defining meaningful quantitative metrics to evaluate the performance of PS-Sparse under different model orders is nontrivial and is an interesting open problem. Preliminary investigation on the use of information-theoretic model-selection methods for this purpose has been done in [30, 112], but further study is necessary to gain deeper insight.

PS-Sparse also requires the choice of the regularization parameter λ. In our simulation studies, the regularization parameters were manually tuned for optimal performance of the algorithms, but this was only possible because we knew the ground truth. Although automatic selection of optimal regularization parameters is still an open problem, a number of methods have been proposed [109], including: a) discrepancy principle-based method, b) L-curve method, and c) generalized cross-validation method. For processing our *in vivo* experimental data, we used the discrepancy principle-based method to select λ, which yielded good empirical results. It is also worth reiterating that in PS-Sparse, the sparsity constraint serves as a regularizer to stabilize the PS model. So a large range of λ values would produce similar reconstruction results, as long as stability is achieved. However, if λ is chosen too big, PS-Sparse would overemphasize the sparsity constraint, resulting in blurring artifacts.

We showed one representative sampling pattern in Fig. 3.1 for the proposed method. It is worth noting that the proposed method allows for flexible design of sparse sampling schemes. In terms of navigator data for the subspace estimation, they are not restricted to be on the Cartesian grid. For example, optimized navigation strategy with more efficient non-Cartesian trajectories, such as spiral trajectories, could lead to improved performance [12]. Furthermore, the sampling of the actual imaging data does not have to be random. With the pre-estimated temporal subspace, the proposed method results in reconstructions with similar accuracy level, even with imaging data acquired in a structured manner (e.g., using the lattice sampling). But, note that how to design an optimal sampling scheme for the proposed method remains an interesting open problem that requires further systematic research.

In this chapter, we focused on a proof-of-the concept study of the proposed method for real-time cardiac imaging. It is worth noting that the proposed method can also be successfully applied to other dynamic imaging applications, such as first-pass myocardial perfusion imaging [111, 113], dynamic speech imaging [12], and functional MRI [114].

3.6 Summary

In this chapter, we presented a new constrained reconstruction method based on the joint low-rank and sparsity model. The proposed method incorporates the complementary advantages of both constraints into a single formulation by using sparsity constraint to regularize low-rank model-based reconstruction. An efficient, globally convergent algorithm based on half-quadratic regularization with continuation was presented to solve the resulting optimization problem. The performance of the proposed method was evaluated using both simulated and in-vivo real-time cardiac imaging data. The results demonstrated that the proposed method yields significantly better performance than using low-rank and sparsity constraints individually.

CHAPTER 4

ACCELERATED MR PARAMETER MAPPING WITH JOINT LOW-RANK AND SPARSITY CONSTRAINTS

4.1 Introduction

In this and the next chapter, we consider another higher-dimensional imaging problem that often arises in quantitative MRI. Specifically, we address the image reconstruction problem with highly undersampled data for quantitative MR parameter mapping (e.g., T_1 mapping, T_2 mapping, and T_2^* mapping). MR relaxometry is arguably one of the earliest and also the most important MR contrast mechanism. It provides useful quantitative information for characterization of tissue properties [115]. It has demonstrated great potential in a wide variety of practical applications, including early diagnosis of neuro-degenerative diseases [116], measurement of iron overload in livers [117], evaluation of myocardial infarction [118], and quantification of labeled cells [119].

MR parameter mapping experiments involve acquiring a sequence of images $\{I_m(\mathbf{x})\}_{m=1}^{M}$ with variable contrast weightings. Each contrast-weighted image $I_m(\mathbf{x})$ is related to the measured $\mathbf{k}$-space data by

$$s_m(\mathbf{k}) = \int I_m(\mathbf{x}) \exp(-i2\pi\mathbf{k}\cdot\mathbf{x})d\mathbf{x} + n_m(\mathbf{k}), \tag{4.1}$$

where $n_m(\mathbf{k})$ denotes complex white Gaussian noise. The conventional approach samples $\mathbf{k}$-space at the Nyquist rate in the acquisition of each $s_m(\mathbf{k})$, from which the $I_m(\mathbf{x})$ are reconstructed, followed by parameter estimation from $\{I_m(\mathbf{x})\}_{m=1}^{M}$. To ensure accurate estimates of relaxation parameters, a relatively long sequence of contrast-weighted images often have to be acquired, which can lead to lengthy data acquisition, in particular for applications that require high spatial resolution and/or broad volume coverage.

Various fast imaging techniques can be adapted, or have been developed specifically, to accelerate parameter mapping experiments. For example, advanced pulse sequences (e.g., [17, 120]) and parallel imaging (e.g., [19–22]) are useful to improve the acquisition efficiency of parameter mapping experiments. Furthermore, a variety of model-based reconstruction methods have been proposed to enable accurate parameter mapping from undersampled data. They can be roughly grouped into two approaches:

- The first approach is to reconstruct $\{I_m(\mathbf{x})\}_{m=1}^{M}$ from undersampled data using various constraints, including temporal smoothness constraint [121], sparsity or structured sparsity constraint [122–127], low-rank constraint [128], and joint low-rank and sparsity constraint [39, 129]. Then voxel-by-voxel parameter estimation is performed to obtain parameter maps of interest.

- The second approach is to directly estimate the parameter map from the undersampled $\mathbf{k}$-space data, bypassing the image reconstruction step completely. This approach typically makes explicit use of a parametric signal model, and formulates the parameter estimation problem as a statistical parameter estimation problem, which allows for easier performance characterization. Representative reconstruction methods in this category include [42, 130–132].

In this chapter, we present a new constrained reconstruction method, which belongs to the first approach above, to accelerate MR parameter mapping. It extends our proposed reconstruction framework in the last chapter, but now sparse sampling is considered in the $\mathbf{k}$-parametric domain, i.e., $\mathbf{k}$-p space. Furthermore, we tailor the low-rank constraint and sparsity constraint to respectively capture different signal/image characteristics in MR parameter mapping. Specifically, low-rank constraint models the correlation of relaxation signals at different spatial locations, while the joint sparsity constraint models the edge structure shared by a sequence of coregistered contrast weighted images. The proposed formulation results in a convex optimization problem. An algorithm based on the alternating direction method of multipliers [57, 59, 60] is developed to efficiently solve the optimization problem. The performance of the proposed method was evaluated in in-vivo T_1 and T_2 mapping applications.

The rest of the chapter is organized as follows. In Section 4.2, we present the proposed formulation. In Section 4.3 and 4.4, we present the solution algorithms for image reconstruction and parameter estimation, respectively. In Section 4.5, we show some representative results from in-vivo experiments. Section 4.6 discusses associated issues with the proposed method, followed by a summary in Section 4.7

4.2 Imaging Model

4.2.1 Data Model

In this chapter, we consider a data model for practical MR parameter mapping experiments, which takes into account parallel imaging with multiple receiver coils. In this data model, a sequence of contrast-weighted images $\{I_m(\mathbf{x})\}_{m=1}^{M}$ are related to the measured $\mathbf{k}$-space data by

$$d_{m,c}(\mathbf{k}) = \int S_c(\mathbf{x}) I_m(\mathbf{x}) \exp(-2\pi \mathbf{k}\cdot\mathbf{x}) d\mathbf{x} + n_{m,c}(\mathbf{k}), \tag{4.2}$$

for $m = 1, \cdots, M$ and $c = 1, \cdots, N_c$, where $S_c(\mathbf{x})$ denotes the coil sensitivity profile for the c-th receiver coil, N_c denotes the number of coils, and $n_{m,c}(\mathbf{k})$ is assumed to be complex white Gaussian noise.

Furthermore, we use the discrete image model based on the Casorati matrix to represent the contrast-weighted image sequence:

$$\mathbf{C} = \begin{bmatrix} I_1(\mathbf{x}_1) & \dots & I_M(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ I_1(\mathbf{x}_N) & \dots & I_M(\mathbf{x}_N) \end{bmatrix} \in \mathbb{C}^{N\times M}. \tag{4.3}$$

Note that the first and second directions of $\mathbf{C}$ represent the spatial and parameter dimensions, respectively. Therefore, Eq. (4.2) can be rewritten as

$$\mathbf{d}_c = \Omega(\mathbf{F}\mathbf{S}_c\mathbf{C}) + \mathbf{n}_c, \tag{4.4}$$

for $c = 1, \cdots, N_c$, where $\mathbf{d}_c \in \mathbb{C}^P$ contains the measured data for the contrast-weighted image sequence from the c-th coil, $\Omega(\cdot) : \mathbb{C}^{N\times M} \to \mathbb{C}^P$ denotes the undersampling operator that sparsely acquires $\mathbf{k}$-space data for each contrast-weighted image and then concatenates them into the data vector $\mathbf{d}_c$, $\mathbf{F} \in \mathbb{C}^{N\times N}$ denotes the Fourier encoding matrix (e.g., the standard discrete Fourier transform matrix for the Cartesian case), $\mathbf{S}_c \in \mathbb{C}^{N\times N}$ is a diagonal matrix that contains the sensitivity map of the c-th coil, and $\mathbf{n}_c \in \mathbb{C}^P$ is the noise vector.

4.2.2 Proposed Formulation

If $\mathbf{d}_c$ contains only sparsely sampled data, direct Fourier inversion of the measured data generally incurs severe artifacts in reconstructed contrast-weighted images and parameter maps. Here, we propose a formulation that simultaneously enforces the low-rank and joint sparsity constraints to enable reconstruction of $\mathbf{C}$ from highly undersampled data, i.e.,

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}\in\mathbb{C}^{N\times M}} \sum_{c=1}^{Nc} \|\mathbf{d}_c - \Omega(\mathbf{F}\mathbf{S}_c\mathbf{C})\|_2^2 + \mathrm{R}_r(\mathbf{C}) + \mathrm{R}_s(\mathbf{C}), \tag{4.5}$$

where $\mathrm{R}_r(\cdot)$ denotes the low-rank constraint, and $\mathrm{R}_s(\cdot)$ denotes the joint sparsity constraint.

First, the low-rank constraint $\mathrm{R}_r(\mathbf{C})$ is based on the assumption that there is strong correlation of relaxation signals from different types of tissues, which leads to the partial separability/low-rank modeling of $\mathbf{C}$. Since, in the last chapter, explicit rank constraint has proven more powerful for image reconstruction from highly undersampled data, we here use this constraint for (4.5), i.e., $\mathbf{C} = \mathbf{U}\mathbf{V}$, where $\mathbf{U} \in \mathbb{C}^{N\times L}$, $\mathbf{V} \in \mathbb{C}^{L\times M}$, and $L \ll \min\{M, N\}$. Note that columns of $\mathbf{U}$ and rows of $\mathbf{V}$ span the spatial and parametric subspaces of $\mathbf{C}$, respectively. A stronger rank constraint can be enforced by pre-estimating $\mathbf{V}$ from some acquired auxiliary data using the principal component analysis or singular value decomposition [2, 79, 83, 128]. We adopt such rank constraint for $\mathrm{R}_r(\mathbf{C})$ in Eq. (4.5).

Secondly, the joint sparsity constraint $\mathrm{R}_s(\mathbf{C})$ is motivated by the assumption that sparse coefficients of different coregistered contrast-weighted images are often highly correlated. Such correlated sparse structure can be more effectively captured by joint sparse modeling [125, 127], since it not only enforces sparsity constraint for each individual image, but

also favors a shared sparse support for different contrast weighted images. Mathematically, joint sparse constraint can be enforced by the mixed ℓ_2/ℓ_1 norm, i.e., $\mathrm{R}_s(\mathbf{C}) = \|\mathbf{\Psi C}\|_{2,1}$, where $\|\mathbf{A}\|_{2,1} = \sum_{n=1}^{N} \|\mathbf{A}^{(n)}\|_2$, and $\mathbf{A}^{(n)}$ denotes the nth row of $\mathbf{A}$. Here, we chose $\mathbf{\Psi}$ as the finite difference operator to exploit the joint edge sparsity. For simplicity, we consider a two direction finite difference transform to obtain a concrete formulation, i.e., $\mathrm{R}_s(\mathbf{C}) = \lambda\|\mathbf{D}_x\mathbf{C}\|_{2,1} + \lambda\|\mathbf{D}_y\mathbf{C}\|_{2,1}$, where $\mathbf{D}_x$ and $\mathbf{D}_y$ represent the horizontal and vertical finite differences, respectively. Extensions to other forms of joint sparsity constraints can be mathematically straightforward.

With the above specific low-rank and joint sparsity constraints, Eq. (4.5) can be rewritten as

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}\in\mathbb{C}^{N\times M}} \sum_{c=1}^{N_c} \|\mathbf{d}_c - \Omega(\mathbf{F}\mathbf{S}_c\mathbf{U}\hat{\mathbf{V}})\|_2^2 + \lambda\|\mathbf{D}_x\mathbf{U}\hat{\mathbf{V}}\|_{2,1} + \lambda\|\mathbf{D}_y\mathbf{U}\hat{\mathbf{V}}\|_{2,1}, \tag{4.6}$$

and the image sequence can be reconstructed as $\hat{\mathbf{C}} = \hat{\mathbf{U}}\hat{\mathbf{V}}$, where $\hat{\mathbf{V}}$ denotes the estimated subspace from auxiliary data, and λ denotes the regularization parameter. Eq. (4.5) integrates both low-rank and joint sparse modelling of $\mathbf{C}$ into a unified mathematical framework. It is easy to see the connection of Eq. (4.6) with the two state-of-the-art methods that use either low-rank or joint sparsity constraint. Specifically, Eq. (4.6) reduces to the subspace-augmented low-rank constrained reconstruction (i.e., kt-PCA [128]) if $\lambda = 0$, and that Eq. (4.6) reduces to the joint sparsity constrained reconstruction [125] if $L = M$ (i.e., the full rank is used).

The complementary roles that low-rank and sparsity constraints play are comprehensively studied in the last chapter for dynamic MRI [31]. Here, for MR parameter mapping, the low-rank model provides strong power to represent the ensemble of relaxation signals of interest. However, low-rank constrained reconstruction (in particular with the pre-estimated $\hat{\mathbf{V}}$) often suffers from ill-conditioning issues with highly undersampled data, which can lead to severe image artifacts and SNR penalty. The joint sparsity constraint acts as not only an additional prior but also an effective regularizer to reduce image artifacts and enhance SNR. The benefits of simultaneously imposing the low-rank and joint sparsity constraints, over using each of these two constraints individually, will be demonstrated in Chapter 4.5.

4.3 Solution Algorithm

Note that Eq. (4.6) is a convex optimization problem with nonsmooth regularization, for which there are a number of numerical algorithms that can be used. Here, we describe an efficient, globally convergent algorithm based on the ADMM [57, 59, 60] to solve it. The algorithm consists of the following major steps. Firstly, Eq. (4.6) is converted into the following equivalent constrained optimization problem through variable splitting, i.e.,

$$\begin{aligned} \{\hat{\mathbf{U}}, \hat{\mathbf{G}}, \hat{\mathbf{H}}\} = \arg \min_{\mathbf{U},\mathbf{G},\mathbf{H}} \sum_{c=1}^{N_c} \|\mathbf{d}_c - \Omega(\mathbf{F}\mathbf{S}_c\mathbf{U}\hat{\mathbf{V}})\|_2^2 + \lambda\|\mathbf{G}\|_{2,1} + \lambda\|\mathbf{H}\|_{2,1}, \\ \text{s.t.} \qquad \mathbf{G} = \mathbf{D}_x\mathbf{U}\hat{\mathbf{V}} \quad \text{and} \quad \mathbf{H} = \mathbf{D}_y\mathbf{U}\hat{\mathbf{V}}. \end{aligned} \tag{4.7}$$

Secondly, the augmented Lagrangian function for Eq. (4.7) can be written as

$$\begin{aligned} L(\mathbf{U}, \mathbf{G}, \mathbf{H}, \mathbf{Y}, \mathbf{Z}) = \sum_{c=1}^{N_c} \|\mathbf{d}_c - \Omega(\mathbf{F}\mathbf{S}_c\mathbf{U}\hat{\mathbf{V}})\|_2^2 + \lambda\|\mathbf{G}\|_{2,1} + \lambda\|\mathbf{H}\|_{2,1} \\ + < \mathbf{Y}, \mathbf{G} - \mathbf{D}_x\mathbf{U}\hat{\mathbf{V}} > + < \mathbf{Z}, \mathbf{H} - \mathbf{D}_y\mathbf{U}\hat{\mathbf{V}} > \\ + \frac{\mu_1}{2}\|\mathbf{G} - \mathbf{D}_x\mathbf{U}\hat{\mathbf{V}}\|_F^2 + \frac{\mu_2}{2}\|\mathbf{H} - \mathbf{D}_y\mathbf{U}\hat{\mathbf{V}}\|_F^2, \end{aligned} \tag{4.8}$$

where $\mathbf{Y} \in \mathbb{C}^{N\times M}$ and $\mathbf{Z} \in \mathbb{C}^{N\times M}$ are the two Lagrangian multipliers, and $\mu_1, \mu_2 > 0$ are penalty parameters related to convergence speed of the algorithm [59].

Thirdly, Eq. (4.8) can be minimized through the following alternating direction method, i.e.,

$$\mathbf{G}_{k+1} = \arg \min_{\mathbf{G}} L(\mathbf{U}_k, \mathbf{G}, \mathbf{H}_k, \mathbf{Y}_k, \mathbf{Z}_k), \tag{4.9}$$

$$\mathbf{H}_{k+1} = \arg \min_{\mathbf{H}} L(\mathbf{U}_k, \mathbf{G}_{k+1}, \mathbf{H}, \mathbf{Y}_k, \mathbf{Z}_k), \tag{4.10}$$

$$\mathbf{U}_{k+1} = \arg \min_{\mathbf{U}} L(\mathbf{U}, \mathbf{G}_{k+1}, \mathbf{H}_{k+1}, \mathbf{Y}_k, \mathbf{Z}_k), \tag{4.11}$$

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_1(\mathbf{G}_{k+1} - \mathbf{D}_x\mathbf{U}_{k+1}\hat{\mathbf{V}}), \tag{4.12}$$

$$\mathbf{Z}_{k+1} = \mathbf{Z}_k + \mu_2(\mathbf{H}_{k+1} - \mathbf{D}_y\mathbf{U}_{k+1}\hat{\mathbf{V}}). \tag{4.13}$$

The solutions to the subproblems Eqs. (4.9) - (4.11) are described in the following subsec-

tions.

In practical implementation, we initialize $\mathbf{U}^{(0)}$ with the projection of zero-filling reconstruction onto the low-rank subspace spanned by $\hat{\mathbf{V}}$, while $\mathbf{G}^{(0)}$, $\mathbf{H}^{(0)}$, $\mathbf{Y}^{(0)}$, and $\mathbf{Z}^{(0)}$ were all initialized with zeros matrices. It should be noted that for the convex optimization problem in Eq. (4.6), the ADMM algorithm is guaranteed to have global convergence from any initializations. With respect to the penalty parameters, μ_1 is set equal to μ_2, considering that the finite differences of the horizontal and vertical directions are approximately at the same scale. Furthermore, we use the following stopping criteria, i.e.,

$$\max\{\frac{\|\mathbf{U}_{k+1}-\mathbf{U}_k\|_F}{\|\mathbf{U}_k\|_F}, \frac{\|\mathbf{G}_{k+1}-\mathbf{G}_k\|_F}{\|\mathbf{G}_k\|_F}, \frac{\|\mathbf{H}_{k+1}-\mathbf{H}_k\|_F}{\|\mathbf{H}_k\|_F}\} \leq \epsilon \tag{4.14}$$

and

$$k > K_{\max}, \tag{4.15}$$

where ϵ and $K_{\max}$ are the pre-defined tolerance parameter and maximum number of iterations, respectively. The algorithm is terminated until either Eq. (4.14) or Eq. (4.15) is satisfied.

Solutions to Eq. (4.9) - Eq. (4.11)

We present the specific procedures to solve the subproblems in Eq. (4.9) - Eq. (4.11). Note that Eq. (4.9) can be rewritten as

$$\begin{aligned}\mathbf{G}_{k+1} &= \arg\min_{\mathbf{G}} \frac{1}{2}\|\mathbf{G}-\mathbf{D}_x\mathbf{U}_k\hat{\mathbf{V}}+\frac{1}{\mu_1}\mathbf{Y}_k\|_F^2+\frac{\lambda}{\mu_1}\|\mathbf{G}\|_{2,1} \\ &= \arg\min_{\mathbf{G}} \frac{1}{2}\|\mathbf{G}-\mathbf{D}_x\mathbf{U}_k\hat{\mathbf{V}}+\frac{1}{\mu_1}\mathbf{Y}_k\|_F^2+\frac{\lambda}{\mu_1}\sum_{n=1}^{N}\|\mathbf{G}^{(n)}\|_2,\end{aligned} \tag{4.16}$$

where $\mathbf{G}^{(n)}$ denotes the nth row of $\mathbf{G}$. It can be shown that Eq. (4.16) is separable with respect to each row of $\mathbf{G}$. Solving Eq. (4.16) is equivalent to solving

$$\mathbf{G}_{k+1}^{(n)} = \arg\min_{\mathbf{G}^{(n)}} \frac{1}{2}\|\mathbf{G}^{(n)} - (\mathbf{D}_x^{(n)}\mathbf{U}_k\hat{\mathbf{V}} - \frac{1}{\mu_1}\mathbf{Y}^{(n)})\|_2^2 + \frac{\lambda}{\mu_1}\|\mathbf{G}^{(n)}\|_2, \tag{4.17}$$

for $n = 1, \cdots, N$. This problem admits a closed-form solution, which can be obtained via the following soft-thresholding operation, i.e.,

$$\mathbf{G}_{k+1}^{(n)} = \frac{\mathbf{T}_k^{(n)}}{\|\mathbf{T}_k^{(n)}\|_2}\max\{\|\mathbf{T}_k^{(n)}\|_2 - \frac{\lambda}{\mu_1}, \quad 0\}, \qquad n = 1, \cdots, N, \tag{4.18}$$

where $\mathbf{T}_k^{(n)} = \mathbf{D}_x^{(n)}\mathbf{U}_k\hat{\mathbf{V}} - \frac{1}{\mu_1}\mathbf{Y}_k^{(n)}$.

For the subproblem in Eq. (4.10), it can be solved by a very similar procedure as the above. Specifically, each row of $\mathbf{H}_{k+1}$ can be obtained as follows:

$$\mathbf{H}_{k+1}^{(n)} = \frac{\mathbf{Q}_k^{(n)}}{\|\mathbf{Q}_k^{(n)}\|_2}\max\{\|\mathbf{Q}_k^{(n)}\|_2 - \frac{\lambda}{\mu_2}, \quad 0\}, \qquad n = 1, \cdots, N, \tag{4.19}$$

where $\mathbf{Q}_k^{(n)} = \mathbf{D}_y^{(n)}\mathbf{U}_k\hat{\mathbf{V}} - \frac{1}{\mu_2}\mathbf{Z}_k^{(n)}$.

For the subproblem in Eq. (4.11), note that it can be written as

$$\begin{aligned}\mathbf{U}_{k+1} = \arg\min_{\mathbf{U}} \sum_{c=1}^{N_c} \|\mathbf{d}_c - \Omega(\mathbf{F}\mathbf{S}_c\mathbf{U}\hat{\mathbf{V}})\|_2^2 + \frac{\mu_1}{2}\|\mathbf{G}_{k+1} - \mathbf{D}_x\mathbf{U}\hat{\mathbf{V}} + \frac{1}{\mu_1}\mathbf{Y}_k\|_F^2 \\ + \frac{\mu_2}{2}\|\mathbf{H}_{k+1} - \mathbf{D}_y\mathbf{U}\hat{\mathbf{V}} + \frac{1}{\mu_2}\mathbf{Z}_k\|_F^2,\end{aligned} \tag{4.20}$$

which is a large-scale quadratic optimization problem. It can be efficiently solved by a number of numerical algorithms. Here, the conjugate gradient (CG) algorithm is applied, with $\mathbf{U}$ initialized by $\mathbf{U}_k$ from the last iteration. Furthermore, it should be noted that in the above CG iterations, the sampling operator Ω and the Fourier encoding matrix $\mathbf{F}$ do not have to be explicitly stored, since they can be evaluated via very fast operation or transformation.

4.4 Parameter Estimation

After reconstructing $\hat{\mathbf{C}}$, relaxation parameters of interest (e.g., T_1 or T_2 maps) can be easily determined voxel-by-voxel via solving a nonlinear least-squares (NLS) fitting problem, for which a number of algorithms can be used. But note that in contrast to a generic NLS problem, the problem here has separable structure [133] in the sense that the nonlinear contrast weighting model linearly depends on a subset of its unknowns, i.e., the proton density value. To take advantage of such special structure, we adopt the variable projection (VARPRO) algorithm [133–136], which has been shown to be more computationally efficient than generic nonlinear optimization algorithms [133, 135]. In our case, after variable projection, the optimization problem becomes one dimensional, so we can discretize the relaxation parameters of interest into a finite set of values, and apply VARPRO with one-dimensional grid search [134, 136], which is guaranteed to result in a global optimal solution.

4.5 In Vivo Results

In this section, we show respresentative results from both in-vivo T_2 mapping and T_1 mapping experiments to illustrate the performance of the proposed method.

To perform quantitative evaluation of different reconstruction methods, we use the following three metrics: a) voxelwise error $= (\boldsymbol{\gamma}_n - \hat{\boldsymbol{\gamma}}_n)/\boldsymbol{\gamma}_n$, where $\boldsymbol{\gamma}_n$ and $\hat{\boldsymbol{\gamma}}_n$ respectively denote the true and estimated relaxation parameter at the nth voxel, b) region-of-interest (ROI) error $= \|\boldsymbol{\gamma}_{\text{ROI}} - \hat{\boldsymbol{\gamma}}_{\text{ROI}}\|_2/\|\boldsymbol{\gamma}_{\text{ROI}}\|_2$, where $\boldsymbol{\gamma}_{\text{ROI}}$ and $\hat{\boldsymbol{\gamma}}_{\text{ROI}}$ respectively contain the true and estimated relaxation parameters in a specific ROI, and c) overall error $= \|\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}\|_2/\|\boldsymbol{\gamma}\|_2$, where $\boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}$ respectively denote the true and estimated relaxation parameter map that contains all image voxels.

4.5.1 T_2 Mapping Application

We evaluated the proposed method in an in vivo human brain T_2 mapping experiment on a healthy volunteer. The experiment was performed on a 3T Siemens Trio scanner (Siemens

Medical Solutions, Erlangen, Germany) equipped with a 12-channel receiver headcoil. A multi-echo spin-echo imaging sequence was used with 25 evenly spaced echoes (the first echo time $\mathrm{TE}_1 = 11.5$ ms and the echo spacing $\triangle\mathrm{TE} = 11.5$ ms). Other relevant imaging parameters were: repetition time (TR) = 3.11 s, field-of-view (FOV) = 180 mm $\times$ 240 mm, matrix size = 208×256, number of slices = 8, and slice thickness = 3 mm. A pilot scan with a rapid GRE sequence was also performed, from which the coil sensitivity maps $\mathbf{S}_c$ were estimated.

We performed retrospective undersampling of this fully sampled data, and Fig. 4.1 illustrates one representative sampling scheme in $\mathbf{k}$-p space, where the parametric dimension p refers to the echo number for the T_2 mapping experiments. Specifically, in this acquisition scheme, one central $\mathbf{k}$-space readout was fully acquired at all echo times and treated as training data, from which we estimate $\hat{\mathbf{V}}$ using the principal component analysis [2, 79, 128]. To measure the sparse sampling level, the acceleration factor (AF) is defined as MN/P. Specifically, the four AFs, namely, AF = $2.8, 4.1, 6.0,$ and 8.0, were considered for this set of data. The T_2 map estimated from the fully sampled data was treated as a reference, with which we evaluated the performance of different reconstruction methods. We performed slice-by-slice reconstructions from undersampled data using the proposed method. The rank L was selected as 3, and the regularization parameter λ was empirically optimized by visual inspection.

To demonstrate the benefits of imposing simultaneous low-rank and joint sparsity constraints, we also performed low-rank based reconstruction (i.e., kt-PCA [128]) and joint sparsity based reconstruction [125, 127] (denoted as joint sparse hereafter). For kt-PCA, we used the same sampling pattern as the one used for the proposed method. However, for joint-sparse, since such sampling scheme often leads to sub-optimal performance, a variable density random sampling pattern with densely acquired central $\mathbf{k}$-space [125] was adopted. Furthermore, we used the same rank/model order for kt-PCA as for the proposed method. For joint sparse, the regularization parameter was manually optimized by visual inspection. All three reconstruction methods shared the same set of sensitivity maps. After reconstructions, T_2 maps were estimated using VARPRO based on a mono-exponential T_2 relaxation model. A discrete set of T_2 values, i.e., $\{1, \cdots, 500\}$ ms, was used as the search grid for

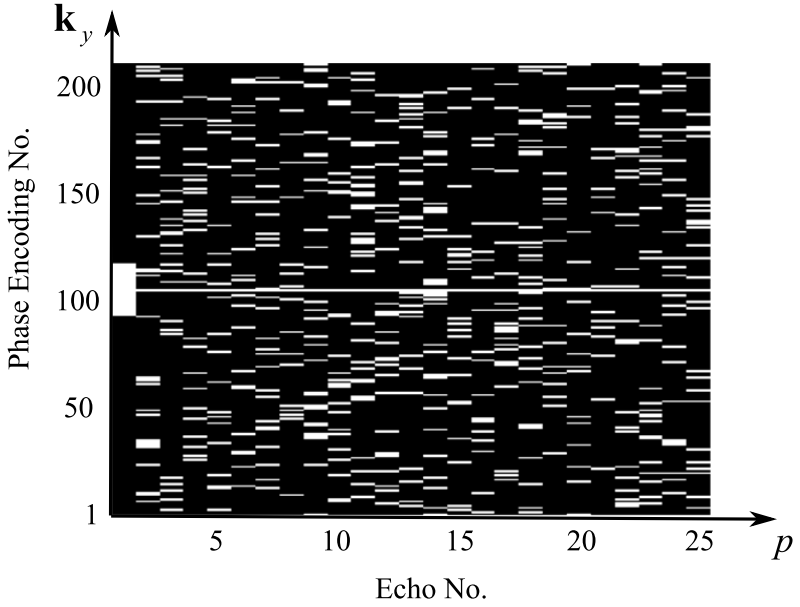


Figure 4.1: Representative undersampling patterns in $\mathbf{k}_y$-p space for T_2 mapping used in the proposed method. The white bars and black bars respectively denote the acquired and unacquired **k**-space readouts. For both experiments, we acquire one central **k**-space readout at each acquisition parameter and use this set of data as training data, whereas we sparsely acquire data in other regions of $\mathbf{k}_y$-p space with a uniform random sampling pattern and use such data as imaging data. Furthermore, to enhance SNR, we densely acquire the low resolution data at the first TE for T_2 mapping experiments.

VARPRO, with which the resolution of T_2 values is 1 ms.

For all of image reconstruction, we used the initialization and stopping criteria described in the Theory section. Specifically, $\epsilon = 5e^{-4}$ and $K_{\max} = 50$ were respectively set in Eqs. (4.14) and (4.15) for the stopping criteria. Under these conditions, the ADMM algorithm typically converged within 20 iterations, although the specific number of iterations depended on the number of measurements acquired. The above image reconstruction was performed on a workstation with a 3.47GHz dual-hex-core Intel Xeon processor X5690, 96 GB RAM, Linux system and Matlab R2012a. The computation time is within 7 minutes for all the T_2 mapping reconstruction (using the multi-channel data). After reconstruction, the VARPRO algorithm was used to estimate the T_2 maps, which took around 6 seconds.

Figure 4.2 shows the reconstructed T_2 maps of the human brain of slice 4 from the T_2 mapping data using joint sparse, kt-PCA, and the proposed method at two AFs (i.e., AF = 4.1 and 8.0). Along with reconstructions, the corresponding voxelwise error maps are also shown with the overall errors indicated in the top left corner of the images. As can be seen, at a moderate acceleration AF = 4.1, all three methods perform fairly well both qualitatively and quantitatively, although the proposed method yields noticeably better performance than the other two methods. As AF is increased to 8.0, the performance of joint sparse and kt-PCA dramatically degrades. Qualitatively, the edge structures of the T_2 map obtained by joint sparse are severely smoothed out, while the T_2 map obtained by kt-PCA is corrupted by severe artifacts induced by the ill-conditioning issue. Quantitatively, the T_2 values from joint sparse and kt-PCA also become much less accurate at AF = 8.0. In contrast, by simultaneously enforcing low-rank and joint sparsity constraints, the proposed method has much better preserved features and significantly reduced artifacts compared to the other two methods. It also yields much more accurate T_2 values.

Figure 4.3 shows the reconstructed T_2 maps using the proposed method at the highest AF (i.e., AF = 8.0) for slices 2, 3, 6, and 8, along with corresponding voxelwise error maps and overall errors. As can be seen, the proposed method has consistent performance across different slice locations at such a high AF.

Figure 4.4 shows the ROI error versus AF for the ROI, which was chosen from a region of the white matter from the human data (marked in Fig. 4.2). As can be seen, this figure

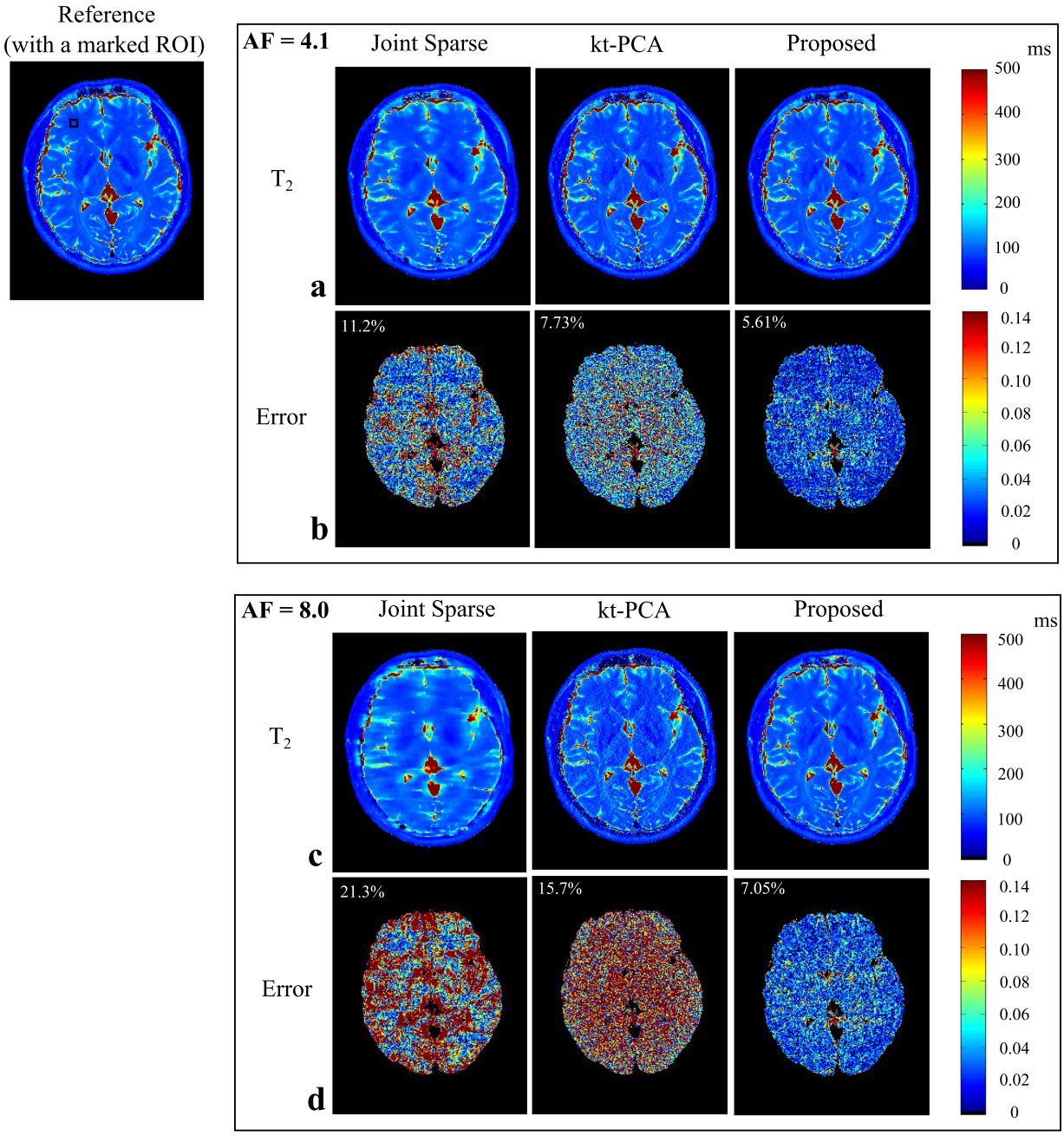


Figure 4.2: Reconstructed T_2 maps of the human brain and associated errors for slice 4 at different AFs. **a**, **b**: Row (a) shows the reconstructed T_2 maps from joint sparse, kt-PCA, and the proposed method at AF = 4.1 and Row (b) shows the voxelwise error maps and the overall errors for the reconstructions in Row (a). **c**, **d**: Row (c) shows the reconstructed T_2 maps from joint sparse, kt-PCA, and the proposed method at AF = 8.0 and Row (d) shows the voxelwise error maps and the overall errors for the reconstructions in Row (c).

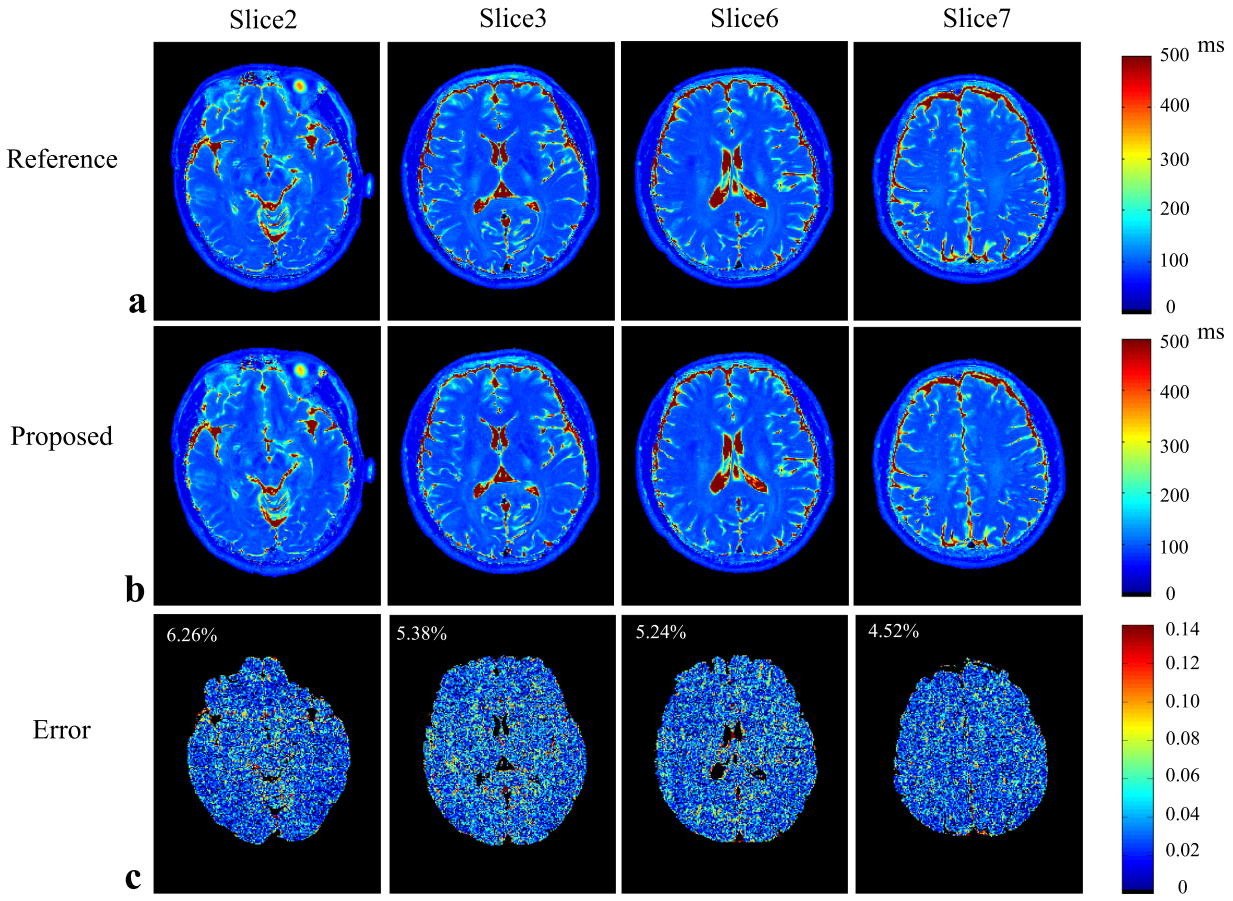


Figure 4.3: Reconstructed T_2 maps of the human brain and associated errors for slices 2, 3, 6, and 7 at AF $= 8.0$ using the proposed method. **a**: the reference T_2 maps of the four slices, **b**: the reconstructed T_2 maps from the proposed method of these slices, and **c**: the voxelwise error maps and the overall errors for the reconstructions in **b**.

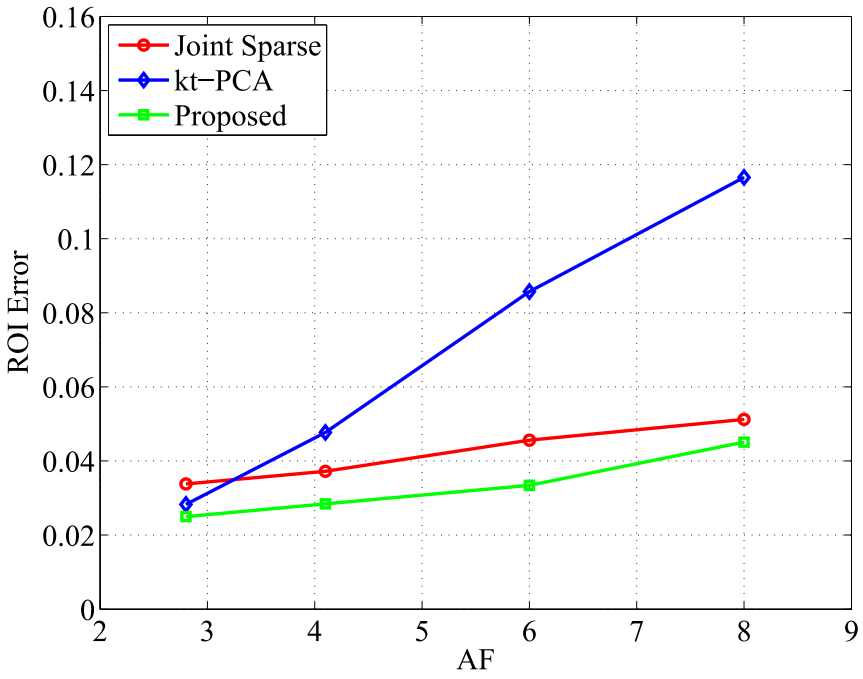


Figure 4.4: The ROI error with respect to AF. The error plot for a ROI in the white matter of the human brain (marked in Fig. 4.2).

.

further illustrates the improved accuracy by using the proposed method. Furthermore, note that although the performance of all three methods degrades as AF increases, the proposed method has improved robustness over the other two methods with respect to the change of AF, which again demonstrates the benefits offered by simultaneously using two constraints.

4.5.2 T_1 Mapping Application

We also evaluated the proposed method using an in vivo rat brain T_1 mapping experiment. The experiment was approved by the Carnegie Mellon University Institutional Animal Care and Use Committee. The experiment was performed on a Bruker Avance III 7T scanner (Bruker Biospin, Billerica, MA) with a single receiver coil using a saturation recovery spin echo sequence with 16 evenly spaced repetition times from 200 ms to 8520 ms. Other relevant imaging parameters were: FOV = 32 mm × 32 mm, matrix size = 128 × 128, flip angle = 90°, number of slices = 1, and slice thickness = 2 mm.

Similar to the human in-vivo data, we performed reconstructions of contrast-weighed

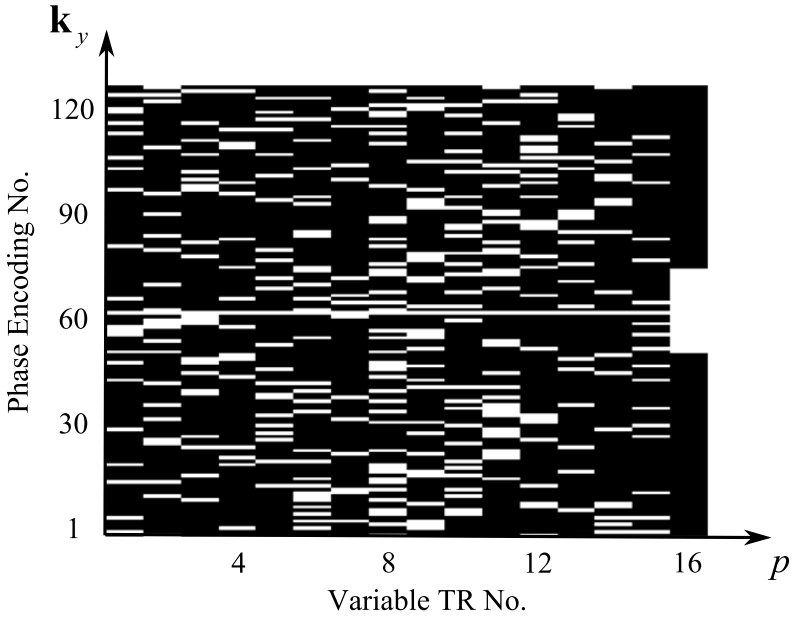


Figure 4.5: Representative undersampling patterns in $\mathbf{k}_y$-p space for T_1 mapping experiments used in the proposed method. For the SNR consideration, we densely acquire the low resolution data at the last TR for T_1 mapping experiments.

image sequences from retrospectively undersampled data with kt-PCA, joint sparse and the proposed method. For the T_1 experiments, one representative $\mathbf{k}$-p space sparse sampling scheme as shown in Fig. 4.5 was used for kt-PCA and the proposed method, whereas a variable density random sampling pattern was used for joint sparse. For kt-PCA and the proposed method, we again acquired one single central $\mathbf{k}$-space readout at every TR as training data to estimate $\hat{\mathbf{V}}$. Three different acceleration factors were considered, i.e., AF $= 3.0, 4.0$ and 5.0. Similarly, the rank $L = 3$ was used for kt-PCA and the proposed method, and the regularization parameters were optimized for the proposed method and joint sparsity reconstruction based on visual inspection. After reconstruction, T_1 maps were estimated using VARPRO based on a mono-exponential T_1 relaxation model. A search grid of T_1 values $\{1, \cdots, 3000\}$ ms was used.

Figure 4.6 shows the reconstructed T_1 maps of the rat brain, the corresponding voxelwise error maps, and the overall errors from the T_1 mapping data using joint sparse, kt-PCA, and the proposed method at AF $= 3.0$ and 5.0. Consistent with the results shown in the T_2 mapping example, the proposed method improves over the other two methods, both qualitatively and quantitatively. Furthermore, Fig. 4.7 shows the ROI error versus AF for

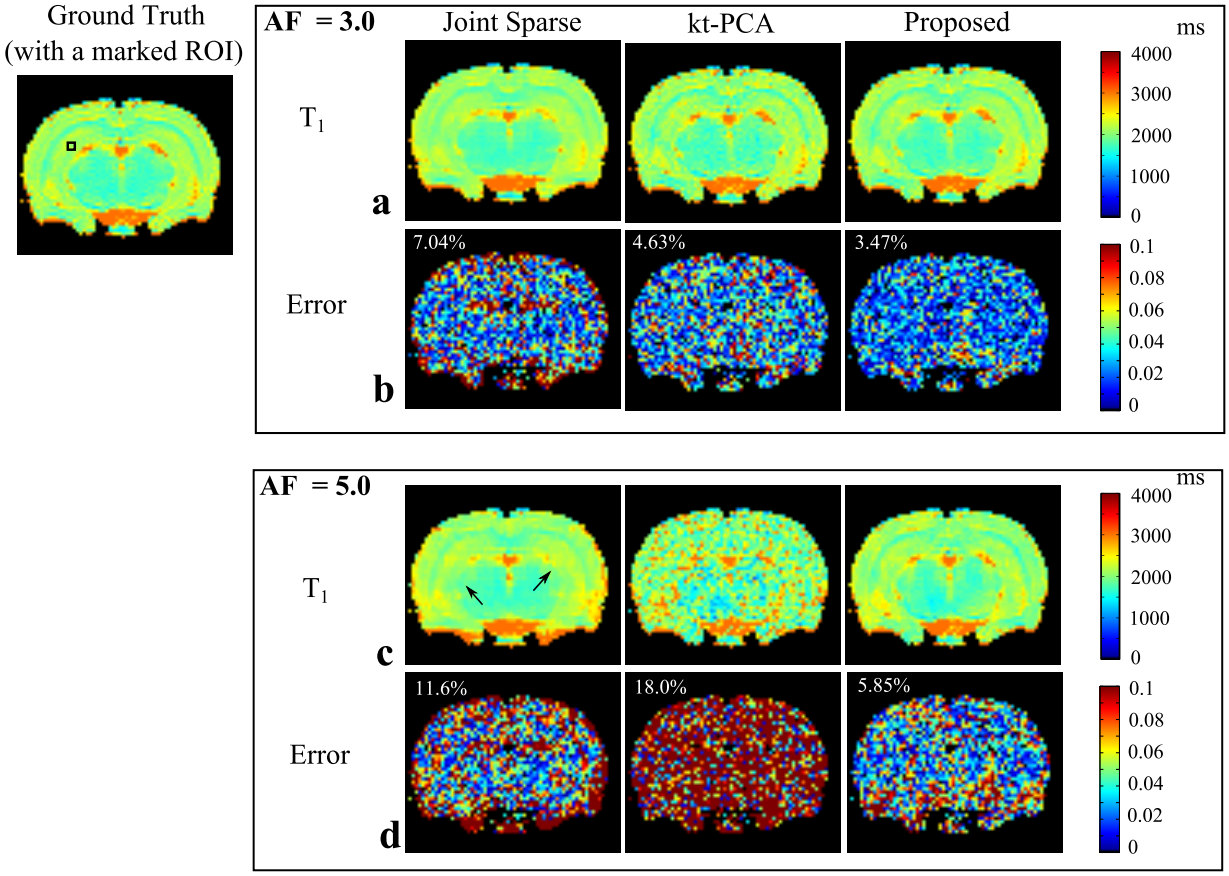


Figure 4.6: Reconstructed T_1 maps of the rat brain and associated errors at different AFs. **a**, **b**: Row (a) shows the reconstructed T_1 maps using joint sparse, kt-PCA, and the proposed method at AF = 3.0 and Row (b) shows the corresponding voxelwise error maps and the overall errors for the reconstructions in Row (a). **c**, **d**: Row (c) shows the reconstructed T_1 maps using joint sparse, kt-PCA, and the proposed method at AF = 5.0 and Row (d) shows the corresponding voxelwise error maps and the overall errors for the reconstructions in Row (c).

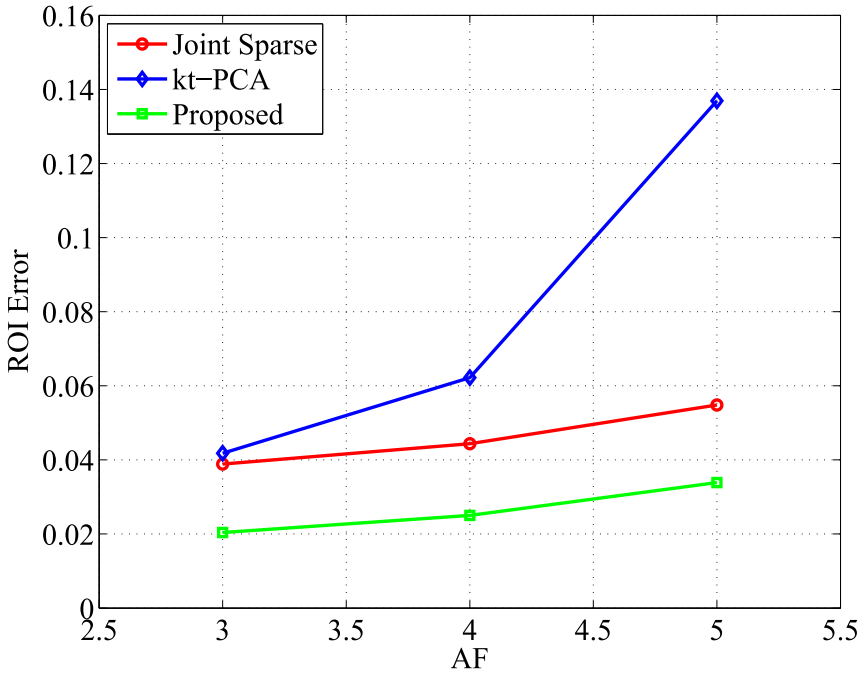


Figure 4.7: The ROI error with respect to AF. The error plot for a ROI in the hippocampus of the rat brain (marked in Fig. 4.6).

.

the ROI (chosen from a region of a region of the hippocampus from the rat data set (marked in Fig. 4.6). Similar to Fig. 4.4, this figure illustrates the superior performance of the proposed over the other methods over different AFs.

4.6 Discussion

The effectiveness of integrating low-rank and joint sparsity constraints for accelerated parameter mapping has been demonstrated in the last section. It is worthwhile to make further comments on some points. First of all, parameter subspaces estimated from limited training data can accurately capture the underlying relaxation process. For example, for parameter mapping applications in [128] and our work, $\hat{\mathbf{V}}$ estimated from only a single central $\mathbf{k}$-space readout results in accurate parameter values. Since this amount of training data typically only comprises a small portion of the total number of measurements, acquiring training data does not significantly compromise the overall acceleration of parameter map-

ping experiments. Note that an alternative is to estimate the subspace from an ensemble of relaxation signals generated using a pre-assumed signal model with a range of parameters [129]. However, data-driven subspaces from acquired data can be more faithful to capture the underlying relaxation process, and they can also provide better robustness to potential signal model mismatches (e.g., multi-exponential relaxation).

It is worth noting that with $\hat{\mathbf{V}}$ pre-estimated, the sampling of the imaging data for the proposed method is very flexible. This can be demonstrated by the results from the T2 mapping example. Specifically, we investigated the three sampling schemes shown in Fig. 4.8 for collecting imaging data, i.e., (a) uniform random sampling (same as the one used in Fig. 4.1), (b) lattice sampling, and (c) block sampling. To enhance SNR, we acquired the low-resolution k-space data associated with the first TE for all the three sampling patterns. The reconstruction results using these sampling patterns are shown in Fig. 4.9. As can be seen, given the same amount of data (i.e., at AF = 8), the proposed method results in reconstructions with similar level of accuracy with three distinct sampling patterns, although the reconstruction result with the random sampling is slightly better. From this illustrative example, it can be seen that the proposed method does not have to rely on either random sampling or a specific design of interleaving sampling pattern. Thus, it is fairly flexible to design different acquisition schemes for the proposed method. However, regarding the optimal sampling design, it is still an open problem that is worth further systematic research, although the "optimal sampling" might not lead to substantial improvement based on the results shown here.

The proposed method requires selecting the rank L. Theoretically, a proper rank is determined by the number of distinct tissue types. Practically, as shown in [128, 129] and the last section, $L = 3$ enables accurate T_1 or T_2 mapping with a mono-exponential signal model. Note, however, that the optimal choice of L may be different for other parameter mapping applications (e.g., multi-exponential models). A useful way to select L is to first adjust it with some reference data set, and then translate the optimally tuned rank to experiments with similar imaging protocols.

In addition to selecting L, the proposed method also requires to choose the regularization parameter λ. Here we empirically chose it based on visual inspection, which leads to good

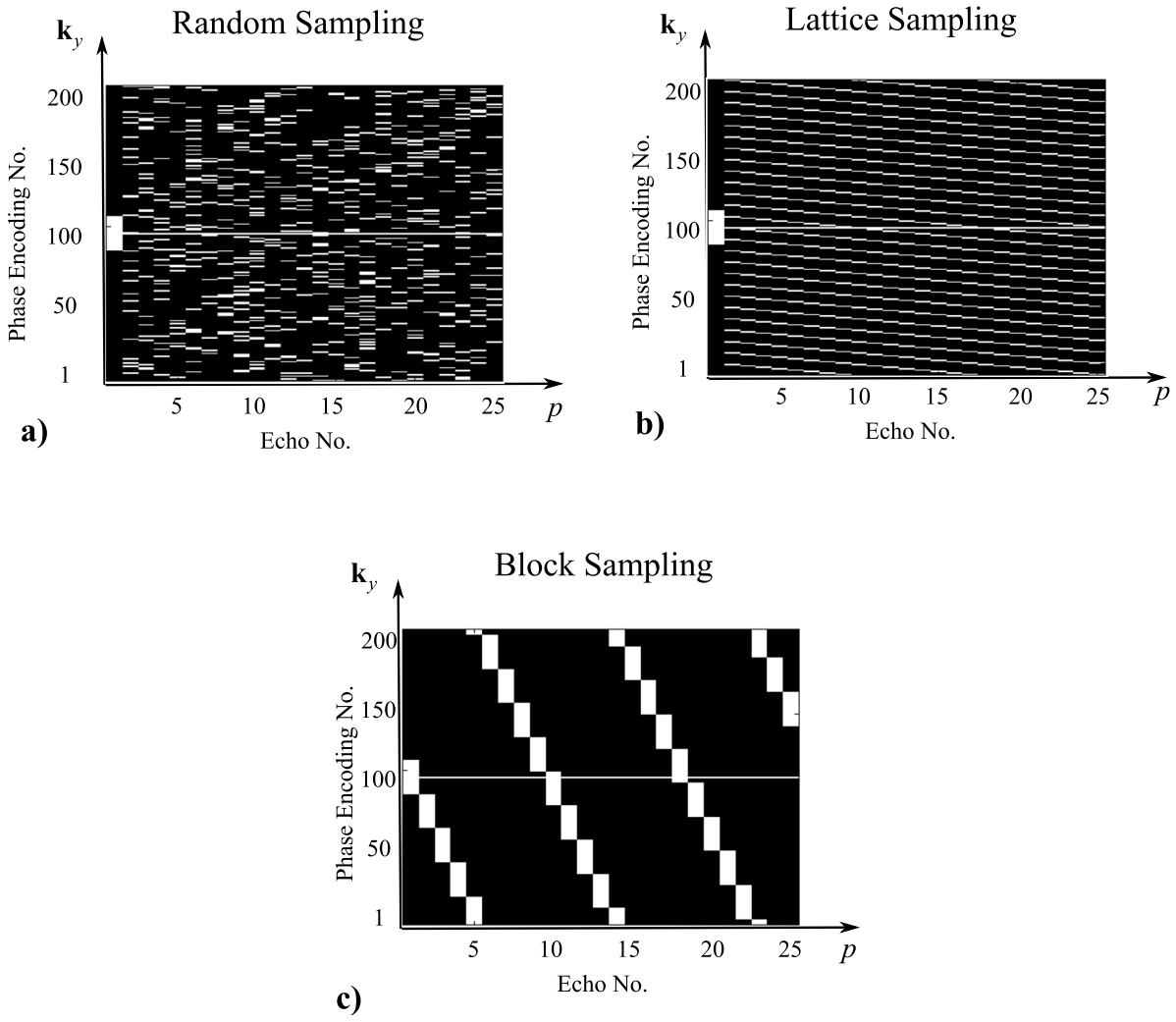


Figure 4.8: Three $\mathbf{k}_y$-p space sampling patterns for the proposed method. The white bar denotes the acquired $\mathbf{k}$-space readout, while the black bar denotes the unacquired k-space readout. (a) Uniform random sampling, (b) Lattice sampling, and (c) Block sampling.

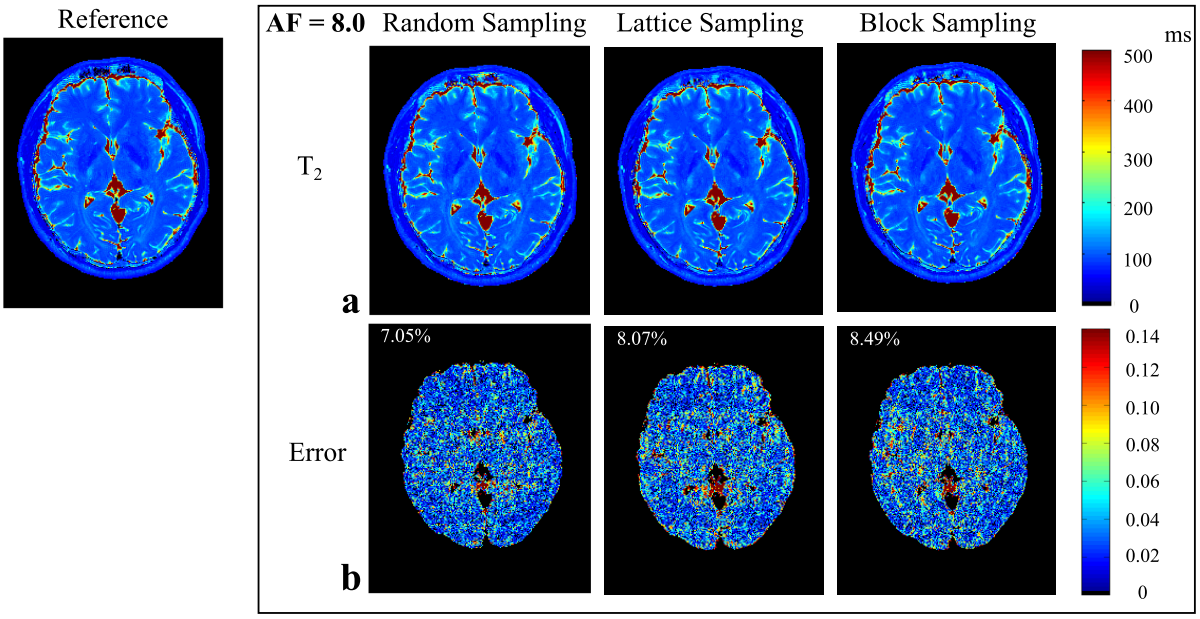


Figure 4.9: Three $\mathbf{k}_y$-p space sampling patterns for the proposed method. The white bar denotes the acquired $\mathbf{k}$-space readout, while the black bar denotes the unacquired k-space readout. (a) Uniform random sampling, (b) Lattice sampling, and (c) Block sampling.

empirical results. A number of alternative methods, such as the L-curve [109] and SURE (Stein's unbiased risk estimate)-based scheme [137], can also be useful to help choose an optimized λ. Furthermore, since in the proposed method, the joint sparsity constraint is mainly used as a regularizer to stabilize the reconstruction problem, a relatively large range of λ values would result in reconstructions with similar level of accuracy, as long as stability is achieved.

We integrated the proposed signal model with the SENSE (Sensitivity Encoding)-based parallel imaging technique and estimated the sensitivity maps from pilot scan. For the brain imaging applications considered in this chapter, since there is no severe motion, the proposed method provided good accuracy. In the case of significant motion between the pilot scan and the parameter mapping scan, inaccurate sensitivity maps can degrade the performance. However, in this case, self-calibration based parallel imaging (e.g., self-calibrated SENSE), which have improved robustness to motion, can be used together with the proposed model. Alternatively, we can always perform channel-by-channel reconstruction of image sequences, and then estimate parameter maps from sum-of-square reconstructions.

4.7 Summary

In this chapter, a new constrained reconstruction method is proposed to accelerate MR parameter mapping. It effectively integrates low-rank constraint with joint sparsity constraint into a unified mathematical formulation. With data-driven parameter subspace pre-estimation, the proposed formulation results in a convex optimization problem, which is solved by an efficient algorithm based on ADMM. Representative results from two sets of in vivo data demonstrate that the proposed method significantly improves, both qualitatively and quantitatively, over state-of-the-art methods that only use low-rank constraint or joint sparsity constraint, when parameter mapping experiments are highly accelerated.

CHAPTER 5

MODEL-BASED MR PARAMETER MAPPING WITH SPARSITY CONSTRAINTS

5.1 Introduction

In this chapter, we address the parameter mapping problem from a different perspective. Specifically, we tackle the problem from an estimation viewpoint by formulating the problem as a parameter estimation problem. We propose a new mathematical formulation, which integrates the parametric physical model with sparsity constraints on the model parameters. It enables direct estimation of the parameters of interest from highly undersampled, noisy data. Furthermore, we develop an efficient greedy-based algorithm to solve the resulting sparsity constrained parameter estimation problem. We also derive the estimation-theoretic bounds to analyze the advantages of using sparsity constraints and benchmark the proposed method against the fundamental performance limit.

The rest of the chapter is organized as follows. In Section 5.2 and 5.3, we present the imaging model and the reconstruction algorithm, respectively. In Section 5.4, we present the estimation-theoretic characterization for the proposed method, followed by the representative results in Section 5.4. In Section 5.6, we discuss related issues with respect to the proposed method, and then summarize this chapter in Section 5.7.

5.2 Imaging Model

For simplicity, we consider a discrete image model, in which $\mathbf{I}_m$ is a $N \times 1$ vector denoting a contrast weighted image. For each image, a finite number of measurements, denoted as

$\mathbf{d}_m \in \mathbb{C}^{P_m}$, are collected. In this setting, the imaging equation (4.1) can be written as[1]

$$\mathbf{d}_m = \mathbf{F}_m \mathbf{I}_m + \mathbf{n}_m, \tag{5.1}$$

for $m = 1, \cdots, M$, where $\mathbf{F}_m \in \mathbb{C}^{P_m \times N}$ denotes the undersampled Fourier measurement matrix, and $\mathbf{n}_m \in \mathbb{C}^{P_m}$ denotes the complex white Gaussian noise with variance σ^2.

5.2.1 Parametric Physical Model

In parameter mapping, due to spin physics, the parameter-weighted images $I_m(\mathbf{x})$ can be expressed as

$$I_m(\mathbf{x}) = \rho(\mathbf{x})\phi(\theta(\mathbf{x}), \boldsymbol{\gamma}_m)\mathrm{e}^{j\psi_m(\mathbf{x})}, \tag{5.2}$$

where $\rho(\mathbf{x})$ represents the spin density distribution, $\theta(\mathbf{x})$ is the desired parameter map, $\phi(\theta(\mathbf{x}), \boldsymbol{\gamma}_m)$ is a contrast weighting function, $\psi_m(\mathbf{x})$ denotes the phase distribution, and $\boldsymbol{\gamma}_m$ contains the user-specified parameters for a given data acquisition sequence (e.g., echo time T_E, repetition time T_R, and flip angle α).

The exact mathematical form of (5.2) is generally known for a chosen parameter mapping experiment [130–132]. For example, for a variable flip angle T_1-mapping experiment, (5.2) can be written as follows [120]:

$$I_m(\mathbf{x}) = \rho(\mathbf{x})\frac{\sin\alpha_m \left[1 - e^{-T_R/T_1(\mathbf{x})}\right]}{1 - \cos\alpha_m e^{-T_R/T_1(\mathbf{x})}}\mathrm{e}^{j\psi_m(\mathbf{x})}, \tag{5.3}$$

where $T_1(\mathbf{x})$ is the parameter map of interest, α_m and T_R are pre-selected data acquisition parameters. We can, therefore, assume that ϕ is a known function in (5.2). Furthermore, we assume that the phase distribution $\{\psi_m(\mathbf{x})\}_{m=1}^M$ is known or can be estimated accurately prior to parameter map reconstruction (e.g., [130–132, 138]). Although both $\rho(\mathbf{x})$ and $\theta(\mathbf{x})$ are unknown parameters in the model, $\theta(\mathbf{x})$ is used mainly for tissue characterization in many applications. Hereafter, we assume that $\theta(\mathbf{x})$ is of primary interest, and $\rho(\mathbf{x})$ is treated as a nuisance parameter.

[1] In this model, we ignore the time-varying relaxation effects between different $\mathbf{k}$-space samples from the same $\mathbf{I}_m$.

After discretization, (5.2) can be written as

$$\mathbf{I}_m = \mathbf{\Psi}_m \mathbf{\Phi}_m(\boldsymbol{\theta})\boldsymbol{\rho}, \tag{5.4}$$

where $\boldsymbol{\theta} \in \mathbb{R}^N$ contains the parameter values of interest, $\boldsymbol{\rho} \in \mathbb{R}^N$ contains the spin density values, $\mathbf{\Phi}_m \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $[\mathbf{\Phi}_m]_{n,n} = \phi\left(\boldsymbol{\theta}_n, \boldsymbol{\gamma}_m\right)$, $\boldsymbol{\theta}_n$ denotes the parameter value at the nth voxel, and $\mathbf{\Psi}_m \in \mathbb{C}^{N\times N}$ is a diagonal matrix containing the phase of $\mathbf{I}_m$. Note that in (5.4), $\mathbf{I}_m$ linearly depends on $\boldsymbol{\rho}$, but nonlinearly depends on $\boldsymbol{\theta}$.

Substituting (5.4) into (5.1) yields

$$\mathbf{d}_m = \mathbf{F}_m \mathbf{\Psi}_m \mathbf{\Phi}_m(\boldsymbol{\theta})\boldsymbol{\rho} + \mathbf{n}_m. \tag{5.5}$$

Based on (5.5), we can determine $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ directly from the measured data $\{\mathbf{d}_m\}_{m=1}^M$ without reconstructing $\{I_m(\mathbf{x})\}_{m=1}^M$. Under the assumption that $\{\mathbf{n}_m\}_{m=1}^M$ are white Gaussian noise, the ML estimation of $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ is given as follows [130–132]:

$$\{\hat{\boldsymbol{\rho}}_{\mathrm{ML}}, \hat{\boldsymbol{\theta}}_{\mathrm{ML}}\} = \arg\min_{\boldsymbol{\rho},\boldsymbol{\theta}} \sum_{m=1}^{M} \left\|\mathbf{d}_m - \mathbf{F}_m \mathbf{\Psi}_m \mathbf{\Phi}_m(\boldsymbol{\theta})\boldsymbol{\rho}\right\|_2^2. \tag{5.6}$$

From (5.6), it can be seen that by introducing the explicit signal model, the number of unknowns reduces from MN to $2N$. A necessary condition for (5.6) to be well-posed is that $\sum_{m=1}^M P_m \geq 2N$, i.e., the total number of measurements is no less than the total number of unknowns in the model [130].

5.2.2 Sparsity Constraint

It is well known that in parameter mapping, the values of $\boldsymbol{\theta}$ are tissue-dependent. Since the number of tissue types is relatively small compared to the number of voxels, we can apply a sparsity constraint to $\boldsymbol{\theta}$ with an appropriate sparsifying transform to incorporate this prior information. Similarly, we can also impose a sparsity constraint on the spin density vector $\boldsymbol{\rho}$. Enforcing the sparsity constraints on $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ leads to the following constrained ML

estimation:

$$\begin{aligned}\{\hat{\boldsymbol{\rho}}_{\text{sML}},\, \hat{\boldsymbol{\theta}}_{\text{sML}}\} = \ & \arg\min_{\boldsymbol{\rho},\boldsymbol{\theta}} \sum_{m=1}^{M} \|\mathbf{d}_m - \mathbf{F}_m \boldsymbol{\Psi}_m \boldsymbol{\Phi}_m(\boldsymbol{\theta})\boldsymbol{\rho}\|_2^2 \,, \\ & \text{s.t. } \|\mathbf{W}_1\boldsymbol{\theta}\|_0 \le K_1 \text{ and } \|\mathbf{W}_2\boldsymbol{\rho}\|_0 \le K_2.\end{aligned} \tag{5.7}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are two chosen sparsifying transforms (e.g., wavelet transforms) for $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, respectively, and K_1 and K_2 are the corresponding sparsity levels. For simplicity, we assume that both $\mathbf{W}_1$ and $\mathbf{W}_2$ are orthonormal transforms.[2] Under this assumption, we can solve the following equivalent formulation:

$$\begin{aligned}\{\hat{\mathbf{c}}, \hat{\mathbf{u}}\} = \arg\min_{\mathbf{c},\mathbf{u}} & \sum_{m=1}^{M} \left\|\mathbf{d}_m - \mathbf{F}_m \boldsymbol{\Psi}_m \boldsymbol{\Phi}_m(\mathbf{W}_1^T\mathbf{c})\mathbf{W}_2^T\mathbf{u}\right\|_2^2 \,, \\ & \text{s.t.} \quad \|\mathbf{c}\|_0 \le K_1 \text{ and } \|\mathbf{u}\|_0 \le K_2,\end{aligned} \tag{5.8}$$

where $\mathbf{c} = \mathbf{W}_1\boldsymbol{\theta}$ and $\mathbf{u} = \mathbf{W}_2\boldsymbol{\rho}$ contain the transform domain coefficients for $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, respectively. Note that (5.8) is a constrained ML estimation problem with explicit sparsity constraints. Later in the chapter, we will demonstrate the benefits of incorporating sparsity in this formulation, both theoretically and empirically.

5.3 Solution Algorithm

Note that (5.8) is a nonlinear optimization problem with a smooth, non-convex cost function and explicit sparsity constraints. For this problem, it can be shown that there exists at least one optimal solution based on the Weierstrass extreme value theorem [95], although the uniqueness of the solution can be generally difficult to establish. With respect to solving this type of problem, a number of greedy pursuit algorithms have been recently developed (e.g., [139–141]). These algorithms are mostly generalizations of greedy algorithms for compressive sensing with nonlinear measurements. Here, we adapt the Gradient Support Pursuit (GraSP) algorithm [141] to solve (5.8). GraSP is an extension of compressive sampling matching

[2]Since both $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ are real, for simplicity, $\mathbf{W}_1$ and $\mathbf{W}_2$ are assumed to be real-valued orthonormal transforms.

pursuit (CoSaMP) [61] or subspace pursuit (SP) [142]. It is an iterative algorithm that utilizes the gradient of the cost function to identify the candidate support of the sparse coefficients, and then solves the nonlinear optimization problem constrained to the identified support.

Specifically, the procedures of GraSP for solving (5.8) can be summarized in Algorithm 1. We denote the cost function in (5.8) as $\Xi(\mathbf{c}, \mathbf{u})$, and the solution at the nth iteration as $\{\mathbf{c}^{(n)}, \mathbf{u}^{(n)}\}$. At the $(n+1)$th iteration, we first compute the partial derivative of Ξ with respect to $\mathbf{c}$ and $\mathbf{u}$ evaluated at $\{\mathbf{c}^{(n)}, \mathbf{u}^{(n)}\}$ (see Section 5.8.1 for the gradient calculation). We denote the values of these partial derivatives as $\mathbf{g}_\mathbf{c} = \frac{\partial \Xi}{\partial \mathbf{c}}|_{\{\mathbf{c}^{(n)}, \mathbf{u}^{(n)}\}}$ and $\mathbf{g}_\mathbf{u} = \frac{\partial \Xi}{\partial \mathbf{u}}|_{\{\mathbf{c}^{(n)}, \mathbf{u}^{(n)}\}}$.

Secondly, we identify the support sets $\mathcal{Z}_\mathbf{c}$ and $\mathcal{Z}_\mathbf{u}$ that are associated with the $2K_1$ largest entries of $\mathbf{g}_\mathbf{c}$ and the $2K_2$ largest entries of $\mathbf{g}_\mathbf{u}$, respectively, i.e., $\mathcal{Z}_\mathbf{c} = \text{supp}(\mathbf{g}_\mathbf{c}^{2K_1})$ and $\mathcal{Z}_\mathbf{u} = \text{supp}(\mathbf{g}_\mathbf{u}^{2K_2})$. Intuitively, under the assumption that $\|\hat{\mathbf{c}}\|_0 \leq K_1$ and $\|\hat{\mathbf{u}}\|_0 \leq K_2$, minimization of $\Xi(\mathbf{c}, \mathbf{u})$ over $\mathcal{Z}_\mathbf{c}$ and $\mathcal{Z}_\mathbf{u}$ would lead to the most effective reduction in the cost function value.

Thirdly, we merge $\mathcal{Z}_\mathbf{c}$ with $\text{supp}(\mathbf{c}^{(n)})$ to form a combined support set $\mathcal{T}_\mathbf{c}$ for $\mathbf{c}$. Similarly, we merge $\mathcal{Z}_\mathbf{u}$ with $\text{supp}(\mathbf{u}^{(n)})$ to form a combined support set $\mathcal{T}_\mathbf{u}$ for $\mathbf{u}$. It is easily shown that $|\mathcal{T}_\mathbf{c}| \leq 3K_1$, and $|\mathcal{T}_\mathbf{u}| \leq 3K_2$. We then minimize $\Xi(\mathbf{c}, \mathbf{u})$ over $\mathcal{T}_\mathbf{c}$ and $\mathcal{T}_\mathbf{u}$. This optimization problem is a support-constrained nonlinear optimization problem, i.e.,

$$\begin{aligned} \min_{\mathbf{c},\mathbf{u}} \sum_{m=1}^{M} \left\| \mathbf{d}_m - \mathbf{F}_m \mathbf{\Psi}_m \mathbf{\Phi}_m (\mathbf{W}_1^T \mathbf{c}) \mathbf{W}_2^T \mathbf{u} \right\|_2^2, \\ \text{s.t.} \quad \mathbf{c}|_{\mathcal{T}_\mathbf{c}^c} = 0 \ \text{ and } \ \mathbf{u}|_{\mathcal{T}_\mathbf{u}^c} = 0. \end{aligned} \tag{5.9}$$

Equation (5.9) can be easily converted into an unconstrained optimization problem by variable change. More specifically, let $\mathbf{c} = \mathbf{E}_\mathbf{c}\mathbf{c}_t$ and $\mathbf{u} = \mathbf{E}_\mathbf{u}\mathbf{u}_t$ where $\mathbf{c}_t \in \mathbb{R}^{|\mathcal{T}_\mathbf{c}|}$ and $\mathbf{u}_t \in \mathbb{R}^{|\mathcal{T}_\mathbf{u}|}$ contain the coefficients on the support $\mathcal{T}_\mathbf{c}$ and $\mathcal{T}_\mathbf{u}$, respectively, and $\mathbf{E}_\mathbf{c} \in \mathbb{R}^{N \times |\mathcal{T}_\mathbf{c}|}$ and $\mathbf{E}_\mathbf{u} \in \mathbb{R}^{N \times |\mathcal{T}_\mathbf{u}|}$ are two sub-matrices of the $N \times N$ identity matrix whose columns are selected according to $\mathcal{T}_\mathbf{c}$ and $\mathcal{T}_\mathbf{u}$, respectively. The resulting unconstrained optimization problem becomes:

$$\min_{\mathbf{c}_t,\mathbf{u}_t} \sum_{m=1}^{M} \left\| \mathbf{d}_m - \mathbf{F}_m \mathbf{\Psi}_m \mathbf{\Phi}_m (\mathbf{W}_1^T \mathbf{E}_\mathbf{c} \mathbf{c}_t) \mathbf{W}_2^T \mathbf{E}_\mathbf{u} \mathbf{u}_t \right\|_2^2. \tag{5.10}$$

Note that (5.10) is a nonlinear least-squares problem. Since the cost function of this unconstrained optimization problem is continuous and coercive, there exists at least one optimal solution [95], although the optimal solution may not be unique. To solve this problem, a number of numerical algorithms can be used. Here, we apply the quasi-Newton limited memory Broyden-Flecher-Goldfarb-Shanno (L-BFGS) algorithm [143]. This algorithm only needs the gradient evaluation, and the Hessian matrix is approximated by the gradient information (see Section 5.8.1 for the gradient of the cost function in (5.10)). Note that in our case, the scaling of two variables $\mathbf{u}$ and $\mathbf{c}$ can be quite distinct. With proper scaling compensation (e.g., through a similar procedure as used in [131, 132]), it has been empirically observed that the accuracy and convergence speed of the L-BFGS algorithm can be significantly improved.

Finally, after obtaining the solution $\bar{\mathbf{c}}$ and $\bar{\mathbf{u}}$ to (5.9), we only keep the largest K_1 entries of $\mathbf{c}$ and the largest K_2 entries of $\mathbf{u}$, and set other entries to zero, i.e., $\mathbf{c}^{(n+1)} = \bar{\mathbf{c}}^{K_1}$ and $\mathbf{u}^{(n+1)} = \bar{\mathbf{u}}^{K_2}$. The above procedure is repeated until the relative change of Ξ between two consecutive iterations is smaller than some pre-specified threshold, or the number of iterations exceeds a pre-specified maximum number of iterations.

Computationally, the algorithm is relatively efficient: at each iteration it only involves gradient evaluation, support detection and solving a support-constrained optimization problem. For the support-constrained optimization problem, its computational complexity is smaller than (5.6) due to the reduced number of unknowns. It has been shown in [141] that under certain theoretical conditions (generalized restricted isometry property for linear measurement model), GraSP has guaranteed performance. However, note that verifying these theoretical conditions in the context of this problem can be very difficult.

Due to the non-convex nature of the optimization problem, the solution of the GraSP is dependent on the initialization $\{\mathbf{c}^{(0)}, \mathbf{u}^{(0)}\}$. In Section 5.5, we will discuss the initialization used in the algorithm, which is closely related to the applied data acquisition scheme. This initial condition consistently yields good empirical results, although other initializations may potentially yield better performance.

Algorithm 3: The GraSP algorithm for solving (5.8).

Input: $\Xi(\mathbf{c}, \mathbf{u})$ *and sparsity levels* K_1 *and* K_2
Initialization: $\{\mathbf{c}^{(0)}, \mathbf{u}^{(0)}\}$
while *"not converged"* **do**

1. evaluate the partial gradient: $\mathbf{g}_{\mathbf{c}} = \frac{\partial \Xi}{\partial \mathbf{c}}|_{\{\mathbf{c}^{(n)}, \mathbf{u}^{(n)}\}}$, and $\mathbf{g}_{\mathbf{u}} = \frac{\partial \Xi}{\partial \mathbf{u}}|_{\{\mathbf{c}^{(n)}, \mathbf{u}^{(n)}\}}$
2. identify directions: $\mathcal{Z}_{\mathbf{c}} = \text{supp}(\mathbf{g}_{\mathbf{c}}^{2K_1})$, and $\mathcal{Z}_{\mathbf{u}} = \text{supp}(\mathbf{g}_{\mathbf{u}}^{2K_2})$
3. merge supports: $\mathcal{T}_{\mathbf{c}} = \mathcal{Z}_{\mathbf{c}} \cup \text{supp}(\mathbf{c}^{(n)})$, and $\mathcal{T}_{\mathbf{u}} = \mathcal{Z}_{\mathbf{u}} \cup \text{supp}(\mathbf{u}^{(n)})$
4. minimize over support: $\{\bar{\mathbf{c}}, \bar{\mathbf{u}}\} = \underset{\mathbf{c},\mathbf{u}}{\text{argmin}}\ \Xi(\mathbf{c}, \mathbf{u})$ s.t. $\mathbf{c}|_{\mathcal{T}_{\mathbf{c}}^c} = 0$ and $\mathbf{u}|_{\mathcal{T}_{\mathbf{u}}^c} = 0$
5. prune estimate: $\mathbf{c}^{(n+1)} = \bar{\mathbf{c}}^{K_1}$ and $\mathbf{u}^{(n+1)} = \bar{\mathbf{u}}^{K_2}$

5.4 Estimation-Theoretic Characterizations

In this section, we derive a constrained CRB to characterize the proposed estimator (5.7) (or (5.8)). The constrained CRB provides a lower bound on the covariance of any locally unbiased estimator under constraints [144, 145]. It has been widely used to characterize various constrained parameter estimation problems (e.g., [146–148]). Specifically, we here calculate the sparsity constrained CRB [146] to analyze the benefits of incorporating the sparsity constraints into the parameter estimation problem. We can also use it as a benchmark to evaluate the performance of the proposed method.

In the rest of this section, we first derive the CRB on the estimated parameter map $\hat{\boldsymbol{\theta}}$ in the unconstrained setting,[3] then extend it to consider the incorporation of the sparsity constraints, and finally we use these bounds to characterize the performance of the ML estimator or sparsity constrained ML estimator. Considering the consistency between the unconstrained and constrained case, we derive both bounds on the sparse coefficients in the

[3]It is worth noting that we can also obtain the performance bounds on $\boldsymbol{\rho}$ by using a similar procedure. We omit such derivations here, since $\boldsymbol{\rho}$ is the nuisance parameter of the model.

transform domain.

5.4.1 Cramér-Rao Bound

Note that since $\hat{\boldsymbol{\theta}} = \mathbf{W}_1^T \hat{\mathbf{c}}$, $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$ can be written as

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \mathbf{W}_1^T \mathrm{Cov}(\hat{\mathbf{c}}) \mathbf{W}_1. \tag{5.11}$$

To obtain the bounds on $\mathrm{Cov}(\hat{\boldsymbol{\theta}})$, it suffices to derive the bounds on $\mathrm{Cov}(\hat{\mathbf{c}})$.

Given the data model, i.e.,

$$\mathbf{d}_m = \mathbf{F}_m \boldsymbol{\Psi}_m \boldsymbol{\Phi}_m (\mathbf{W}_1^T \mathbf{c}) \mathbf{W}_2^T \mathbf{u} + \mathbf{n}_m, \tag{5.12}$$

the unconstrained CRB on $\hat{\mathbf{c}}$ can be written as follows [63]:

$$\mathrm{Cov}(\hat{\mathbf{c}}) \geq \mathbf{Z}, \tag{5.13}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{E}_N & \mathbf{0}_N \end{bmatrix} \mathbf{J}^{\dagger} \begin{bmatrix} \mathbf{E}_N \\ \mathbf{0}_N \end{bmatrix}, \tag{5.14}$$

$\mathbf{E}_N$ is the $N \times N$ identity matrix, $\mathbf{J}$ is the FIM (see Section 5.8.2 for a detailed derivation of the FIM) for the model in (5.12), and $\dagger$ denotes the Moore-Penrose pseudo-inverse. Substituting (5.13) into (5.11) yields the following unconstrained CRB on $\hat{\boldsymbol{\theta}}$:

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{W}_1^T \mathbf{Z} \mathbf{W}_1. \tag{5.15}$$

Taking the diagonal entries of the matrices on the both sides of (5.15), we can obtain a lower bound for the variance of each voxel in $\hat{\boldsymbol{\theta}}$

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}_n) \geq [\mathbf{W}_1^T \mathbf{Z} \mathbf{W}_1]_{n,n}. \tag{5.16}$$

Due to the large dimensionality of $\mathbf{u}$ and $\mathbf{c}$, it is much more computationally efficient to

evaluate $\mathbf{J}^\dagger$ in (5.14) using the inversion of its submatrices. Let the partitioned $\mathbf{J}$ be

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\mathbf{c},\mathbf{c}} & \mathbf{J}_{\mathbf{c},\mathbf{u}} \\ \mathbf{J}_{\mathbf{u},\mathbf{c}} & \mathbf{J}_{\mathbf{u},\mathbf{u}} \end{bmatrix}. \tag{5.17}$$

Using the pseudo-inverse of the partitioned Hermitian matrix [149], $\mathbf{J}^\dagger$ can be written as follows[4]:

$$\mathbf{J}^\dagger = \begin{bmatrix} \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger + \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger \mathbf{J}_{\mathbf{c},\mathbf{u}} \mathbf{V}^\dagger \mathbf{J}_{\mathbf{u},\mathbf{c}} \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger & -\mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger \mathbf{J}_{\mathbf{c},\mathbf{u}} \mathbf{V}^\dagger \\ -\mathbf{V}^\dagger \mathbf{J}_{\mathbf{u},\mathbf{c}} \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger & \mathbf{V}^\dagger, \end{bmatrix} \tag{5.18}$$

where $\mathbf{V} = \mathbf{J}_{\mathbf{u},\mathbf{u}} - \mathbf{J}_{\mathbf{u},\mathbf{c}} \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger \mathbf{J}_{\mathbf{c},\mathbf{u}}$. By substituting (5.18) into (5.14), we can obtain the simplified expression for $\mathbf{Z}$:

$$\mathbf{Z} = (\mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger + \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger \mathbf{J}_{\mathbf{c},\mathbf{u}} \mathbf{V}^\dagger \mathbf{J}_{\mathbf{u},\mathbf{c}} \mathbf{J}_{\mathbf{c},\mathbf{c}}^\dagger). \tag{5.19}$$

5.4.2 Constrained Cramér-Rao Bound

With the sparsity constraints on $\mathbf{c}$ and $\mathbf{u}$, i.e.,

$$\|\mathbf{c}\|_0 \leq K_1 \quad \text{and} \quad \|\mathbf{u}\|_0 \leq K_2, \tag{5.20}$$

(assuming K_1 and K_2 are given), and the data model (5.12), the sparsity constrained CRB for any locally unbiased estimator $\hat{\mathbf{c}}$ can be expressed as follows [146]:

$$\mathrm{Cov}(\hat{\mathbf{c}}) \geq \mathbf{Z}_s, \tag{5.21}$$

where

$$\mathbf{Z}_s = \begin{bmatrix} \mathbf{E}_N & \mathbf{0}_N \end{bmatrix} \mathbf{A} (\mathbf{A}^T \mathbf{J} \mathbf{A})^\dagger \mathbf{A}^T \begin{bmatrix} \mathbf{E}_N \\ \mathbf{0}_N \end{bmatrix}, \tag{5.22}$$

[4]The formula here has already taken into account the case that the FIM is singular. This can happen when the null signal intensity appear in the background. In such case, only the parameter values within the support of the imaging object are estimatable [150, 151].

$\mathbf{A} = \begin{bmatrix} \mathbf{E_c} & \mathbf{0} \\ \mathbf{0} & \mathbf{E_u} \end{bmatrix}$, $\mathbf{E_c}$ and $\mathbf{E_u}$ are two sub-matrices of $\mathbf{E}_N$ whose columns are selected based on the support of $\mathbf{c}$ and $\mathbf{u}$, respectively. Based on (5.11) and (5.21), the constrained CRB on $\hat{\boldsymbol{\theta}}$ can be expressed as

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{W}_1^T \mathbf{Z}_s \mathbf{W}_1, \tag{5.23}$$

and the corresponding voxelwise variance bound can be expressed as:

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}_n) \geq [\mathbf{W}_1^T \mathbf{Z}_s \mathbf{W}_1]_{n,n}. \tag{5.24}$$

Similar to the unconstrained case, we can further simplify the calculation of $\mathbf{Z}_s$ using the block matrix pseudo-inversion. Let the partitioned $\mathbf{A}^T\mathbf{J}\mathbf{A}$ be

$$\mathbf{A}^T\mathbf{J}\mathbf{A} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix},$$

where $\mathbf{G}_{11} = \mathbf{E}_\mathbf{c}^T \mathbf{J}_{\mathbf{cc}} \mathbf{E}_\mathbf{c}$, $\mathbf{G}_{12} = \mathbf{E}_\mathbf{c}^T \mathbf{J}_{\mathbf{cu}} \mathbf{E}_\mathbf{u}$, $\mathbf{G}_{21} = \mathbf{G}_{12}^T$, and $\mathbf{G}_{22} = \mathbf{E}_\mathbf{u}^T \mathbf{J}_{\mathbf{uu}} \mathbf{E}_\mathbf{u}$. Again, with the pseudo-inverse of the partitioned Hermitian matrix [149], $(\mathbf{A}^T\mathbf{J}\mathbf{A})^\dagger$ can be expressed as

$$(\mathbf{A}^T\mathbf{J}\mathbf{A})^\dagger = \begin{bmatrix} \mathbf{G}_{11}^\dagger + \mathbf{G}_{11}^\dagger \mathbf{G}_{12} \mathbf{Q}_s^\dagger \mathbf{G}_{21} \mathbf{G}_{11}^\dagger & -\mathbf{G}_{11}^\dagger \mathbf{G}_{12} \mathbf{Q}_s^\dagger \\ -\mathbf{Q}_s^\dagger \mathbf{G}_{21} \mathbf{G}_{11}^\dagger & \mathbf{Q}_s^\dagger \end{bmatrix}, \tag{5.25}$$

where $\mathbf{Q}_s = \mathbf{G}_{22} - \mathbf{G}_{21}\mathbf{G}_{11}^\dagger \mathbf{G}_{12}$. Substituting (5.25) into (5.22) yields the following simplified expression for $\mathbf{Z}_s$:

$$\mathbf{Z}_s = \mathbf{E}_\mathbf{c} (\mathbf{G}_{11}^\dagger + \mathbf{G}_{11}^\dagger \mathbf{G}_{12} \mathbf{Q}_s^\dagger \mathbf{G}_{21} \mathbf{G}_{11}^\dagger) \mathbf{E}_\mathbf{c}^T, \tag{5.26}$$

where $\mathbf{Q}_s = \mathbf{G}_{22} - \mathbf{G}_{21}\mathbf{G}_{11}^\dagger \mathbf{G}_{12}$.

Under the assumption that there is no model mismatch in the data model (5.12), and that the ML estimator $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ is unbiased, the unconstrained CRB can provide a lower bound on the covariance or MSE of $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ in (5.6) [152]. Note that this bound can also be asymptotically achieved by the ML estimator [63], i.e., when $\sum_{m=1} P_m \to \infty$. Similarly, under the assumption that there is no model mismatch in either the data model (5.12) or the sparsity

model (5.20), and that the constrained ML estimator $\hat{\boldsymbol{\theta}}_{\text{sML}}$ is locally unbiased, (5.23) can also be used to characterize the covariance or MSE of $\hat{\boldsymbol{\theta}}_{\text{sML}}$ in (5.7).

Furthermore, under no mismatch in the signal model and data model, it can be shown [144, 146] that

$$\mathbf{W}_1^T \mathbf{Z} \mathbf{W}_1 \ \geq \mathbf{W}_1^T \mathbf{Z}_s \mathbf{W}_1, \tag{5.27}$$

which confirms the benefits of the sparsity constraints. In the next section, we will calculate the above bounds in a specific application to further illustrate this point.

5.5 Results

In this section, we show representative simulation results from T_2 mapping of the brain to illustrate both the estimation-theoretic bounds and the empirical performance of the proposed method.

The following metrics are used to evaluate the performance of different reconstruction methods: a) normalized error (NE) at the nth voxel:

$$\text{NE}_n = \frac{1}{\boldsymbol{\theta}_n} \left| \boldsymbol{\theta}_n - \hat{\boldsymbol{\theta}}_n \right|, \tag{5.28}$$

where $\hat{\boldsymbol{\theta}}_n$ denotes the reconstructed parameter value at the nth voxel,

b) ROI normalized error (rNE):

$$\text{rNE} = \frac{1}{\boldsymbol{\theta}_{\text{ROI}}} \left| \boldsymbol{\theta}_{\text{ROI}} - \hat{\boldsymbol{\theta}}_{\text{ROI}} \right|, \tag{5.29}$$

where $\boldsymbol{\theta}_{\text{ROI}}$ and $\hat{\boldsymbol{\theta}}_{\text{ROI}}$ respectively denote the averaged parameter values within the ROI from the ground truth and estimator,

c) normalized root-mean-squared-error (NRMSE) at the nth voxel:

$$\text{NRMSE}_n = \frac{1}{\boldsymbol{\theta}_n} \sqrt{\text{E}\left[\left| \boldsymbol{\theta}_n - \hat{\boldsymbol{\theta}}_n \right|^2 \right]}, \tag{5.30}$$

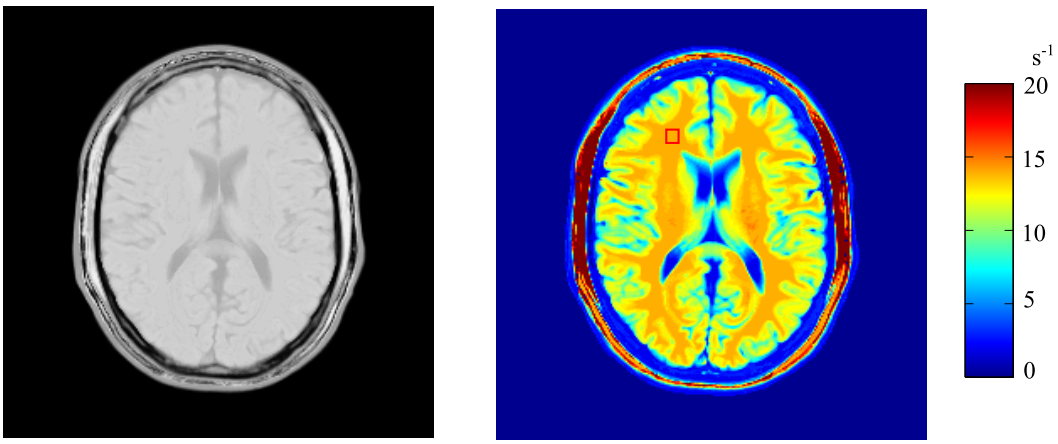


Figure 5.1: Numerical brain phantom: a) the spin density map, and b) the R_2 map (with a marked ROI in the white matter).

d) ROI normalized root-mean-squared-error (rNRMSE) in a ROI:

$$\text{rNRMSE} = \sqrt{\frac{1}{N_{\text{ROI}}} \sum_{n \in \text{ROI}} \text{NRMSE}_n^2}. \tag{5.31}$$

The first two metrics are used to evaluate the accuracy for one noise realization in a simulation, whereas the last two metrics are used to evaluate the accuracy in Monte Carlo simulations.[5]

A realistic numerical brain phantom from the Brainweb database [153] was used to simulate a single-channel, multi-echo spin-echo T_2 mapping experiment. For this experiment, the contrast weighting function is $\phi(\theta, \gamma_m) = \exp(-T_{Em} R_2)$, where $\theta = R_2$ is the transverse relaxation rate (i.e., the reciprocal of T_2), and $\gamma_m = T_{Em}$ denotes the mth echo time. For the spin-echo imaging sequence, the phase distribution matrix $\mathbf{\Psi}_m$ can be assumed to be an identity matrix [132], for $m = 1, \cdots, M$. The spin density map and R_2 map of the phantom are shown in Fig. 5.1.

The imaging sequence consists of $M = 16$ equally spaced echoes with the first echo time $T_{E1} = 12.5$ ms and the echo spacing $\triangle T_E = 9.5$ ms. The acquisition matrix is 256×256, and the spatial resolution is 1×1 mm^2. We performed retrospective undersampling with

[5] In Monte Carlo simulations, the expectations in (5.30) and (5.31) are replaced by the empirical mean.

a sampling pattern that acquires full **k**-space samples for the first echo time and sparse **k**-space data (with a uniform random sampling pattern) for the remaining echo times. The acceleration factor (AF) is defined as $NM/\sum_{m=1}^{M} P_m$, i.e., the total number of voxels in the image sequence divided by the total number of measurements, to measure the undersampling levels. The SNR is defined as $20\log_{10}(s/\sigma)$, where s denotes the average signal intensity within a region of the gray matter in the T_2-weighted image with the longest echo time, and σ denotes the noise variance.

Simulations were performed in the following two settings. In the first setting, we used a data set that has a) a monoexponential T_2 relaxation, and b) $\boldsymbol{\rho}$ and R_2 maps that are sparse in the wavelet domain. Specifically, we synthesized the $\boldsymbol{\rho}$ and R_2 maps using their largest 20% wavelet coefficients with the Daubechies 4 wavelet transform and three-level decomposition. The T_2-weighted image sequence $\{I_m(\mathbf{x})\}_{m=1}^{M}$ was then generated using the monoexponential T_2 relaxation model. This scenario is referred to as the simulation setting without model mismatch. In the second setting, we used the original brain phantom from the Brainweb database. This phantom was simulated in a way that both signal model mismatch (multiexponential relaxation caused by partial volume effect [153]) and sparsity model mismatch (the wavelet coefficients are just compressible) exist. This scenario is referred to as the simulation setting with model mismatch.

5.5.1 Without model-mismatch

We evaluated the performance of the following three methods: the ML estimator in (5.6), the proposed estimator in (5.7), and an oracle estimator that assumes complete knowledge of the support of the wavelet coefficients of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$. Mathematically, the oracle estimator $\{\hat{\boldsymbol{\rho}}_{\mathrm{o}}, \hat{\boldsymbol{\theta}}_{\mathrm{o}}\}$ can be defined as $\hat{\boldsymbol{\rho}}_{\mathrm{o}} = \mathbf{W}_2\hat{\mathbf{u}}_{\mathrm{o}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{o}} = \mathbf{W}_1\hat{\mathbf{c}}_{\mathrm{o}}$ with

$$\begin{aligned}\{\hat{\mathbf{c}}_{\mathrm{o}}, \hat{\mathbf{u}}_{\mathrm{o}}\} = \underset{\mathbf{c},\mathbf{u}}{\operatorname{argmin}} \sum_{m=1}^{M} \left\|\mathbf{d}_m - \mathbf{F}_m\boldsymbol{\Psi}_m\boldsymbol{\Phi}_m(\mathbf{W}_1^T\mathbf{c})\mathbf{W}_2^T\mathbf{u}\right\|_2^2, \\ \text{s.t.} \quad \mathbf{c}|_{\mathcal{T}_1^c} = 0 \ \text{ and } \ \mathbf{u}|_{\mathcal{T}_2^c} = 0,\end{aligned} \tag{5.32}$$

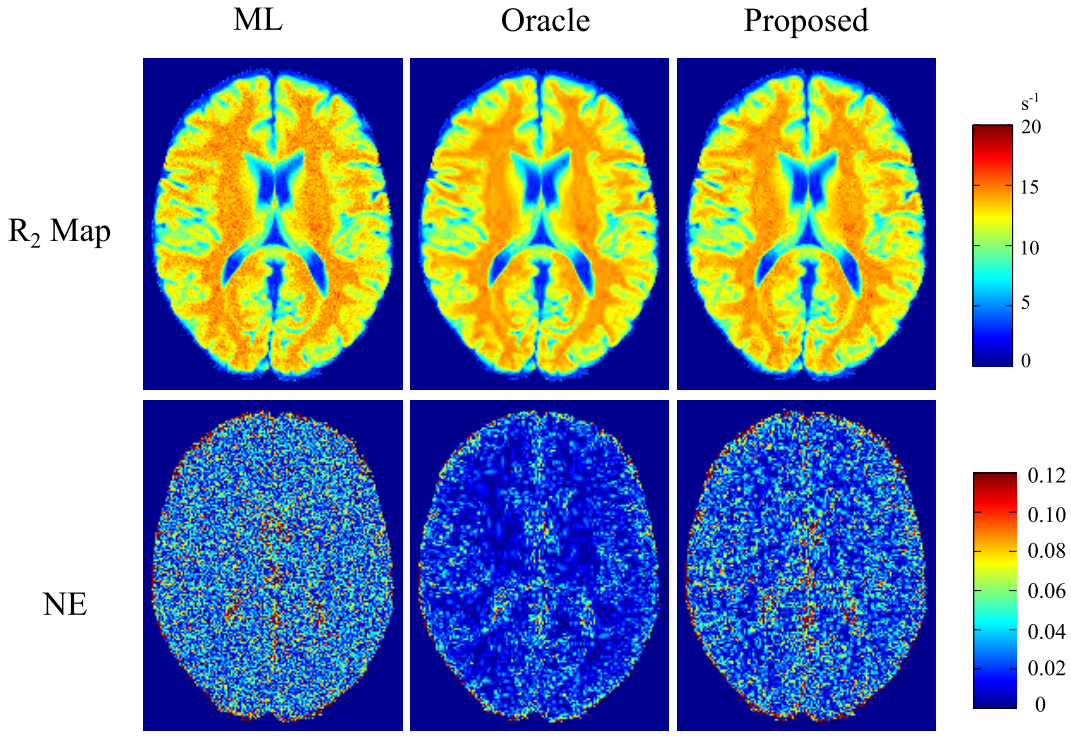


Figure 5.2: Reconstructed R_2 maps and corresponding NE maps for the simulations without model mismatch, from the ML estimator, the oracle estimator, and the proposed estimator, at AF $= 4$ and SNR $= 30$ dB.

where $\mathcal{T}_1$ and $\mathcal{T}_2$ represent the true support sets of the wavelet coefficients of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, respectively. Although the oracle estimator is generally impossible to implement in practice, its performance can indicate the best performance that the proposed estimator could achieve. It can be shown that the oracle estimator asymptotically achieves the constrained CRB, and that it is asymptotically unbiased (see Section 5.8.3).

For the ML estimator, $\boldsymbol{\rho}$ was initialized with the image reconstructed from the fully sampled data at the first echo, and R_2 was initialized with zero. For the oracle estimator, $\mathbf{u}_{\text{o}}$ was initialized with the largest $0.2N$ wavelet coefficients of the reconstructed image from the first echo, and $\mathbf{c}_{\text{o}}$ was initialized with zero. For the proposed method, the same initializations as the oracle estimator were used. It was observed empirically that the above initializations consistently yielded good reconstruction results, although other initializations may result in further improvement. In terms of sparsity levels, we set $K_1 = K_2 = 0.2N$, i.e., the proposed estimator knows the true sparsity level, but it does not have any knowledge of the support

sets of the sparse wavelet coefficients.

To numerically determine the ML estimator and the oracle estimator, the L-BFGS algorithm was used, in which the maximum number of iterations was set to be 150 for both estimators. For the proposed estimator, we used the GraSP algorithm as described in Section 5.3. Specifically, for the subproblem (5.10) in GraSP, we applied the L-BFGS algorithm as well, for which the maximum number of iteration was also set to be 150. With respect to the stopping criterion of GraSP, we terminated the algorithm, when either a) the maximum iteration is larger than 20, or b) the relative change of the cost function value Ξ between two consecutive iterations is less than $1e^{-4}$.

We performed reconstructions using the three methods for AF = 4 and SNR = 30 dB. The reconstructed R_2 maps and their corresponding NE are shown in Fig. 5.2.[6] As can be seen qualitatively, the proposed method reduced noise corruption, comparing to the R_2 reconstructed by the ML estimator. Quantitatively, the proposed method also has better accuracy than the ML estimator that only takes advantage of the contrast-weighting signal model, although the proposed method is inferior to the oracle estimator, as expected. For a ROI in the white matter (labeled in Fig. 5.1), the rNEs are 2.13%, 1.17%, 1.50% for the ML, oracle, and proposed estimators, consistent with the comparison shown in Fig. 5.2.

We conducted MC simulations (with 200 trials) to investigate the statistical properties of the three estimators empirically. Figure 5.3 illustrates the empirical bias and variance of each estimator. The bias from the ML estimator and the oracle estimator are almost negligible, confirming that these two estimators are asymptotically unbiased in theory. The proposed estimator has larger bias than the ML. Compared to the oracle, it has also larger bias due to its error in support detection. In terms of variance, the proposed estimator has much smaller values than the ML estimator, due to the sparsity constraints. Furthermore, we see that for all three estimators, their variances are much larger than their corresponding squared biases, which implies the MSE is dominated by the variance component in the current simulation setting.

In addition to the above MC study, we calculated the estimation-theoretic bounds using

[6]Note that the background, skull, and scalp are not region of interest for our study, and thus they were set to zeros for all the results shown in this section.

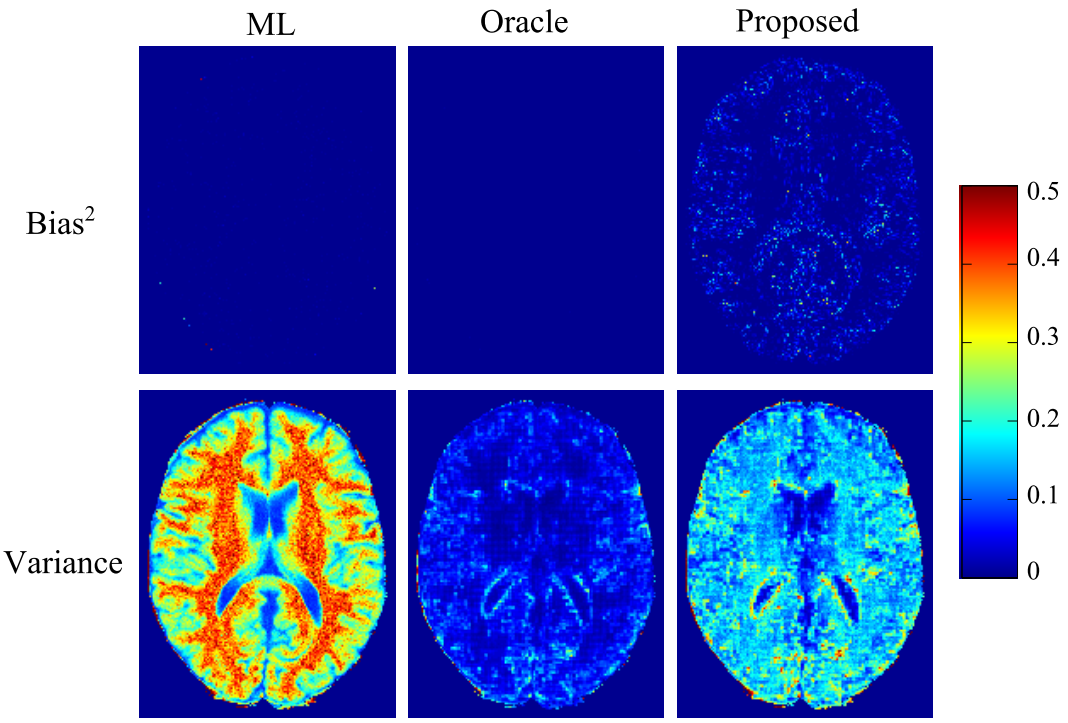


Figure 5.3: Squared bias and variance of the ML estimator, the oracle estimator, and the proposed estimator at AF = 4 and SNR = 30 dB for the simulations without model mismatch.

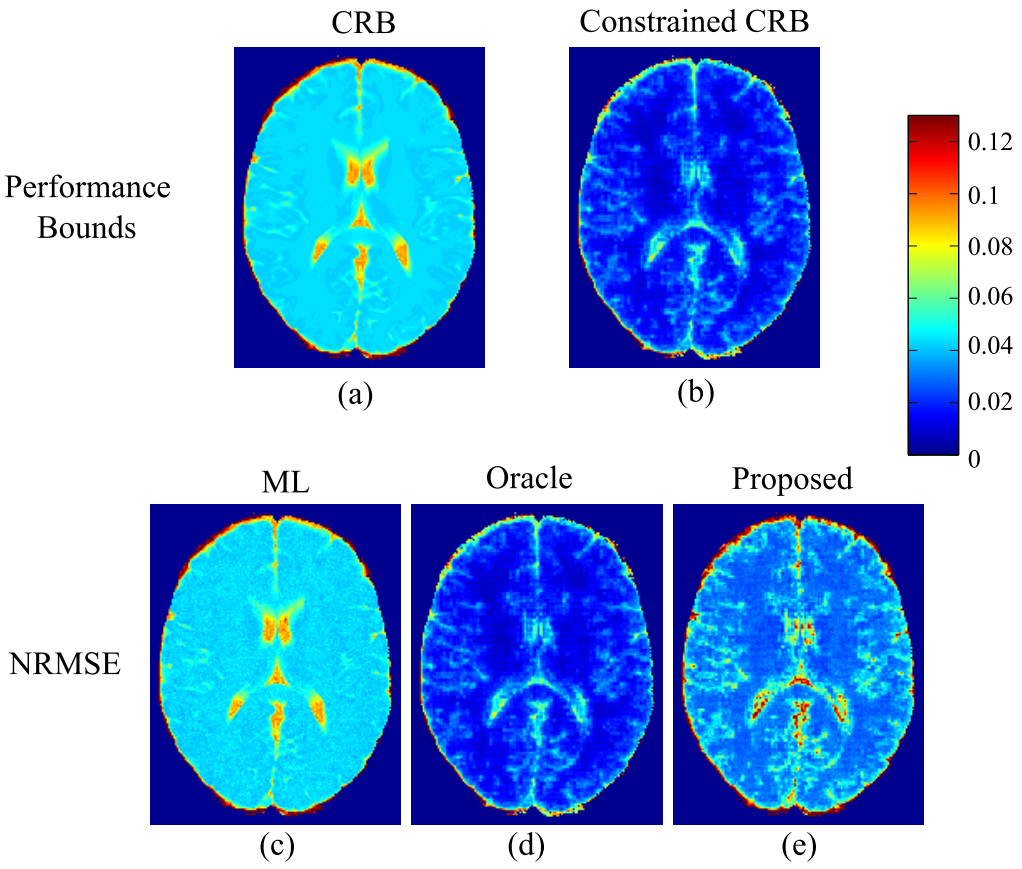


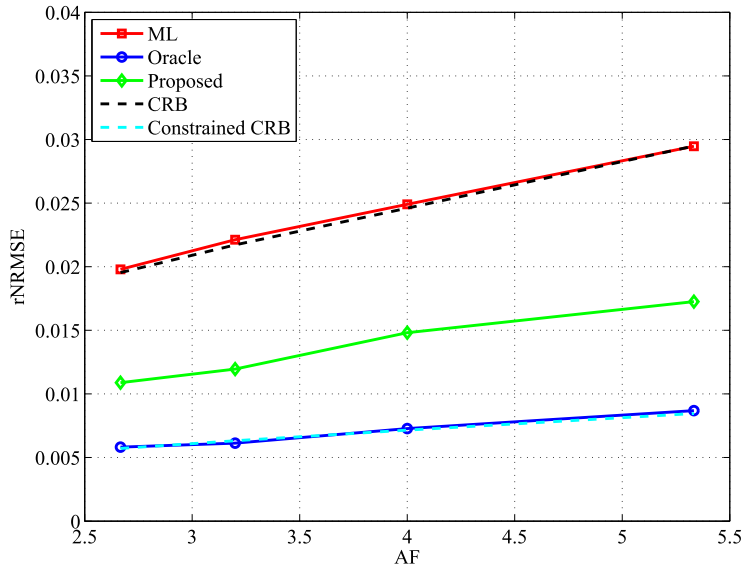
Figure 5.4: Estimation-theoretic bounds and NRMSE maps for the simulation setting without model mismatch: a) normalized CRB map, b) normalized constrained CRB map, c)-e) NRMSE maps of the ML, oracle, and proposed estimators.

(5.16) and (5.24) to analyze the benefits of sparsity constraints. The two bounds also inform the performance limits of the ML estimator and the proposed estimator in the absence of modeling errors. We also calculated the NRMSE of the three estimators from the MC study. To compare the performance bounds with the NRMSE at the right scale, we use the following normalized CRB and CCRB:

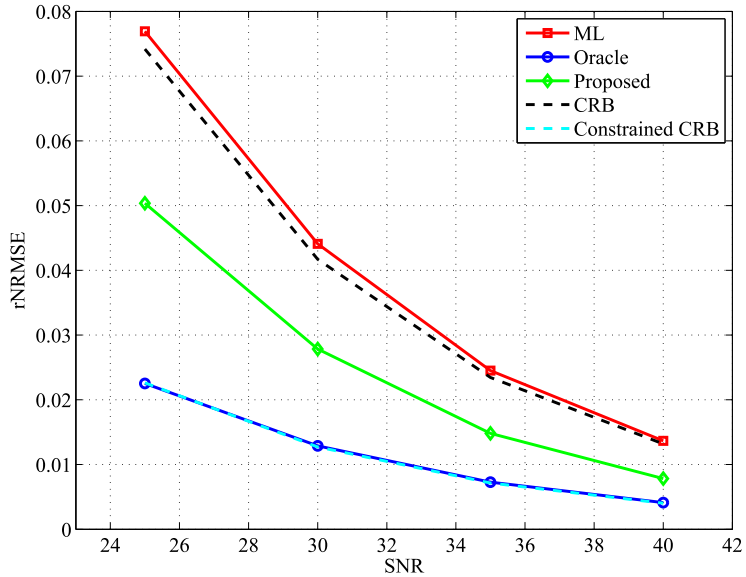
$$\begin{aligned} \text{nCRB}_n &= \frac{1}{\boldsymbol{\theta}_n}\sqrt{[\mathbf{W}_1^T\mathbf{Z}\mathbf{W}_1]_{n,n}}, \\ \text{nCCRB}_n &= \frac{1}{\boldsymbol{\theta}_n}\sqrt{[\mathbf{W}_1^T\mathbf{Z}_s\mathbf{W}_1]_{n,n}}. \end{aligned} \tag{5.33}$$

Figure 5.4 shows the performance bounds and NRMSE maps. As can be seen, the normalized CRB predicts the NRMSE of the ML estimator well, while the normalized CCRB accurately captures the NRMSE of the oracle estimator. Both theoretical results and empirical study clearly demonstrate the benefits of incorporating the sparsity constraints. Take a square ROI in the white matter (labeled in Fig. 5.1) for example. From the normalized CRB and CCRB maps, it can be seen that by incorporating the sparsity constraints with $K_1 = K_2 = 0.2N$, the reconstruction error at best decreases from 4.17% to 1.27%. As one practical estimator, the proposed method attains the reconstruction error at around 2.78%, which is better than the ML estimator.

We studied the performance bounds and empirical performance of the three estimators at different undersampling sampling levels. The MC simulations were performed for AF = 2.67, 3.2, 4, and 5.33, all with SNR = 35 dB. The rNRMSE (with the ROI labeled in Fig. 5.1) was computed. We also calculated the performance bounds at the above undersampling levels. Figure 5.5(a) shows the plot of the rNRMSE and the normalized performance bounds with respect to AF. As expected, both performance bounds and the empirical performance of the three estimators become worse when AF increases. Consistent with the previous results, the two performance bounds well predict the empirical performance of the ML and oracle estimators, respectively. The proposed estimator improves over the ML estimator at all AFs. Note that the rNRMSE of the proposed estimator at AF = 5.33 is smaller than that of the ML estimator at AF = 2.67, i.e., the proposed estimator achieves even better performance than the ML estimator when it only uses half the **k**-space data of the ML estimator.



(a)



(b)

Figure 5.5: The rNRMSE and estimation-theoretic bounds with respect to different AF or SNR: a) rNRMSE and normalized performance bounds versus AF, b) rNRMSE and normalized performance bounds versus SNR.

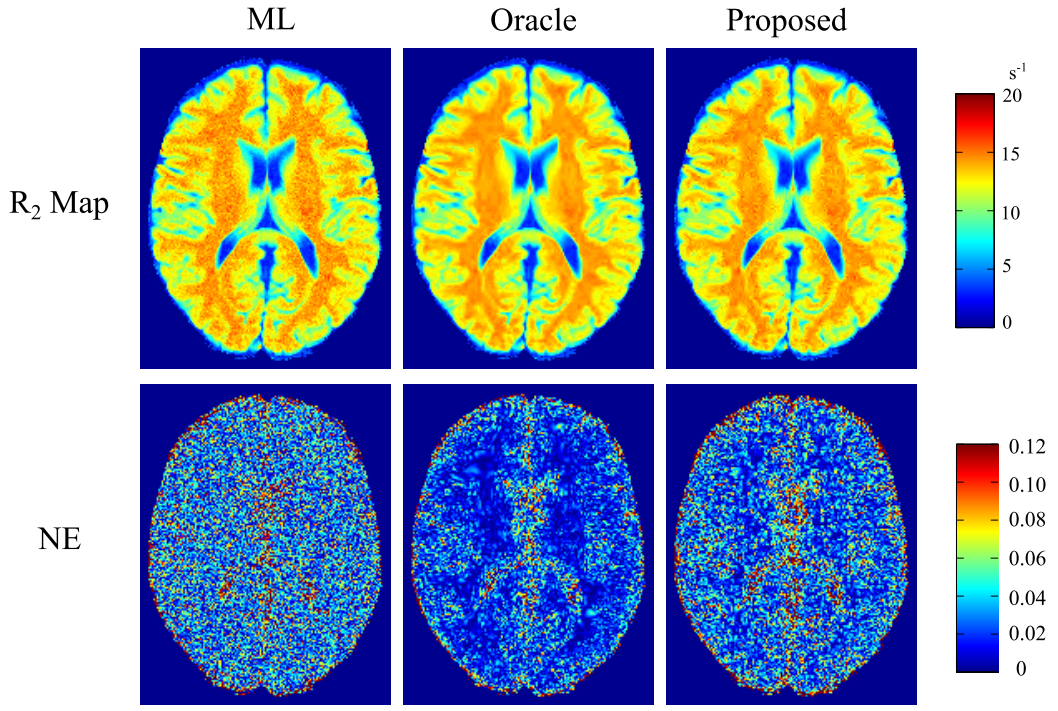


Figure 5.6: Reconstructed R_2 maps for the simulations with model mismatch, from the ML estimator, oracle estimator, and proposed estimator, under AF = 4 and SNR = 30 dB.

We also studied the performance bounds and empirical performance of the three estimators at different SNR levels, i.e., SNR = 25, 30, 35, and 40 dB, all with AF = 4. The plot of rNRMSE (with the same ROI) and the normalized performance bounds with respect to SNR are shown in Fig. 5.5(b). It can be seen that the proposed estimator has improved noise robustness over the ML estimator, both theoretically and empirically, at all tested SNR levels. Note that the proposed estimator achieves a similar rNRMSE at SNR = 35 dB to what the ML estimator does at SNR = 40 dB.

5.5.2 With model-mismatch

We also performed simulations to study the performance of the proposed method with model mismatch. In this simulation setting, the proposed estimator and the oracle will be affected by the modeling errors from both the monoexponential model and sparse approximation, while the ML estimator is only influenced by the monoexponential model mismatch.

For the proposed method, we empirically set the sparsity levels $K_1 = K_2 = 0.2N$, which

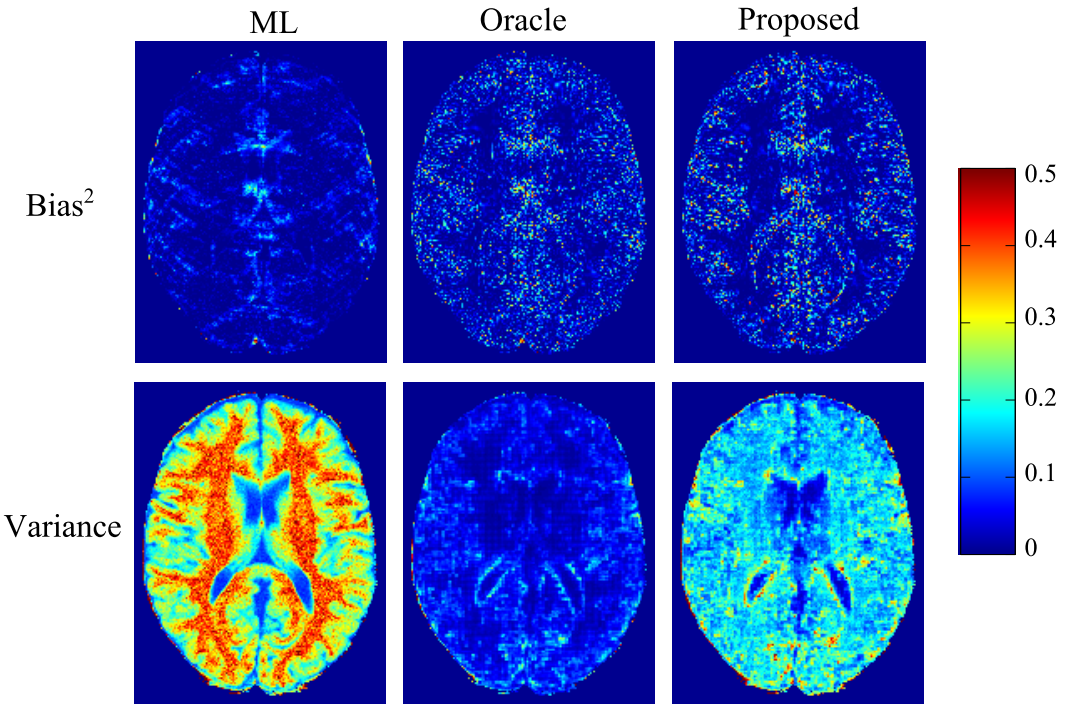


Figure 5.7: Squared bias and variance for the ML estimator, oracle estimator, and the proposed estimator from the simulations with model mismatch.

is the same as the previous simulations. For the oracle estimator, we set $\mathcal{T}_1$ and $\mathcal{T}_2$ as two support sets that contain the largest $0.2N$ wavelet coefficients of $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$, respectively. Furthermore, for the numerical solvers of three estimators, we followed the same implementation as described for the simulations without model mismatch.

We performed reconstructions using the three methods under AF = 4 and SNR = 30 dB. The reconstructed R_2 maps, along with the corresponding NE maps, are shown in Fig. 5.6. This figure demonstrates the improvement of the oracle and the proposed estimator over the ML estimator. For the ROI in the white matter, the rNEs are 2.41%, 1.47%, and 1.71% for the ML, oracle, and the proposed estimator.

Figure 5.7 shows the bias-variance analysis of the three estimators from the MC simulations (with 200 trials) under AF = 4 and SNR = 30 dB, and Fig. 5.8 shows the corresponding NRMSE maps. With the sparsity constraints, the oracle and proposed estimators have significantly reduced variance and NRMSE than the ML estimator. Compared to the case without modeling error, the three estimators exhibit similar level of variance. However, in terms of biases, the three estimators all have increased values. It is worth noting that the

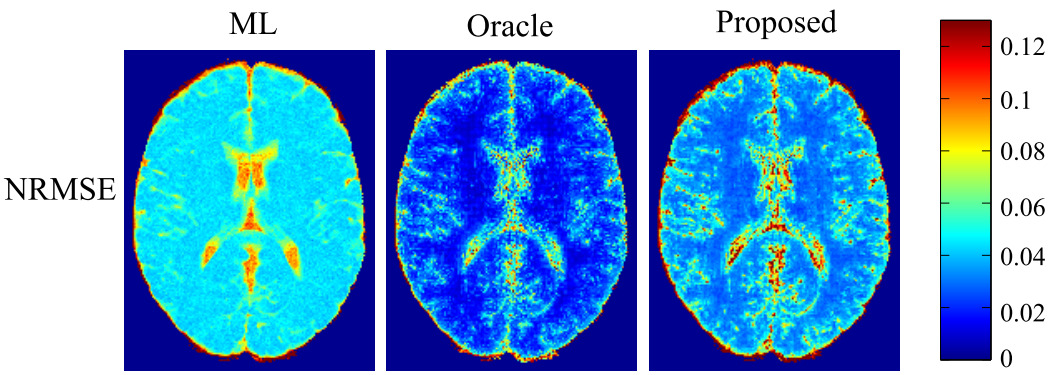


Figure 5.8: NRMSE maps for the ML estimator, oracle estimator, and the proposed estimator from the simulations with model mismatch.

squared bias of the ML estimator exhibits an aliasing-like pattern, which is different from those of the oracle and the proposed estimator. This is due to the different modeling errors that they have.

We again performed MC simulations to evaluate the three estimators at different undersampling levels, i.e., AF $= 2.67,\ 3.2,\ 4,$ and 5.33, with SNR $= 35$ dB. We also performed the MC simulations at different noise levels, i.e., SNR $= 25, 30, 35,$ and 40 dB, with AF $= 4$. The plot of the rNRMSE (with the same ROI) with respect to AF, as well as rNRMSE with respect to SNR, is shown in Fig. 5.9 (a) and (b), respectively. This figure further demonstrates that the proposed method improves over the ML estimator even in the presence of modeling errors, at different undersampling levels or noise levels.

5.6 Discussion

The proposed method imposes sparsity constraints on both the relaxation parameter map $\boldsymbol{\theta}$ and the proton density map $\boldsymbol{\rho}$, which extends our early work [40] that only enforces the sparsity constraint on $\boldsymbol{\theta}$. Although $\boldsymbol{\rho}$ is assumed to be the nuisance parameter, we observed that imposing the sparsity constraint on $\boldsymbol{\rho}$ slightly improved the reconstruction accuracy for $\boldsymbol{\theta}$. Furthermore, with the additional sparsity constraint, the computation and memory usage for calculating the constrained CRB is significantly reduced, since the matrices in (5.26) have much smaller sizes.

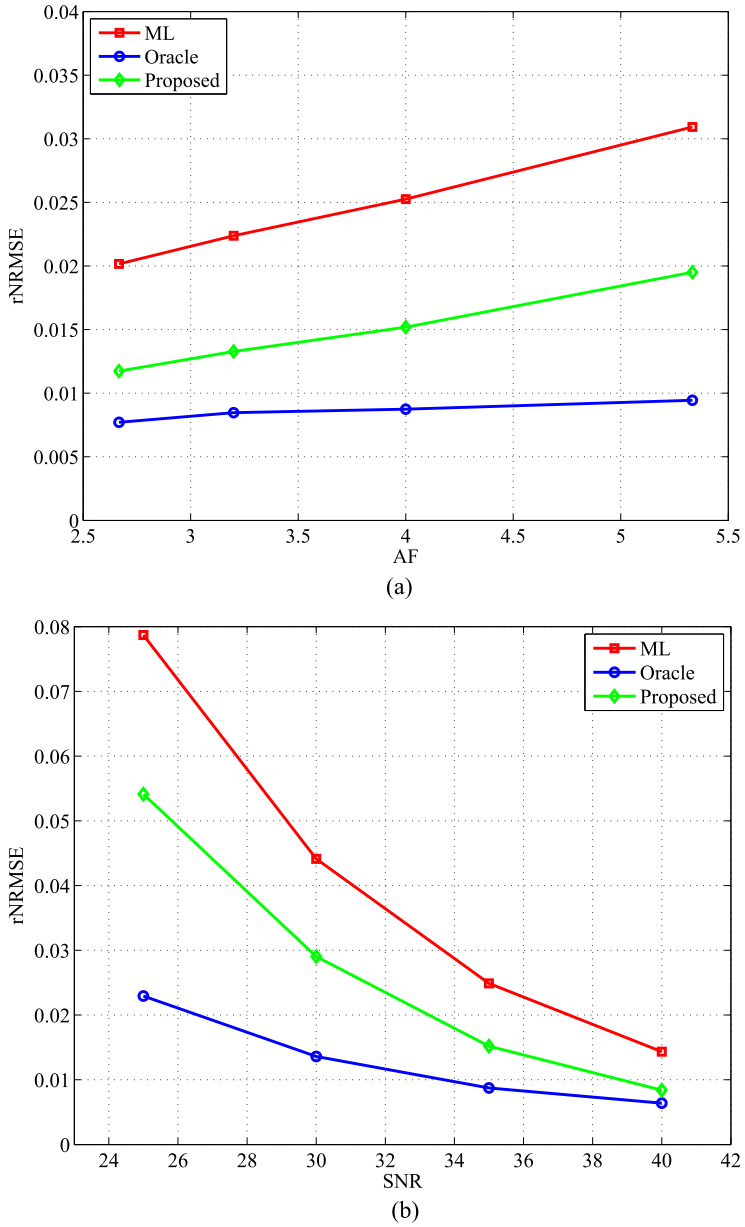


Figure 5.9: The rNRMSE plot at different AF and SNR for the simulations with model mismatch: a) the rNRMSE versus AF, and b) the rNRMSE versus SNR.

The proposed estimator needs to specify the sparsity levels K_1 and K_2 in (5.8), which will impact the sparse approximation error and noise amplification of the resulting parameter maps. It is also directly related to the trade-off between bias and variance of the proposed estimator. If K_1 and K_2 are too small, severe sparse approximation error can incur significant bias that dominates MSE. On the contrary, if K_1 and K_2 are too large, significant variance due to noise amplification can diminish the benefit of the sparsity constraints and the performance of the proposed method can be very close to that of the ML estimator. Here, we manually set K_1 and K_2, with the reference of the underlying ground truth, to balance the bias and variance of the proposed estimator. For some specific applications (e.g., T_2 mapping for brains), the level of sparse approximation error might be roughly known, which could assist the sparsity level selection. Generally, choosing sparsity levels in a principled way, in the absence of the ground truth, is still an open problem, which requires further systematic study.

The parametric signal model that assumes a monoexponential relaxation process has been used in the proposed method. In some practical applications, a multi-exponential signal model may be more accurate [154–156]. The proposed method, including the algorithm and performance bounds, can be generalized to multiexponential signal models, although the optimization problem will be more difficult. Furthermore, in some specific applications (e.g., variable flip angle T_1 mapping), it would be beneficial to take into account of imperfections in data acquisition [157], such as B_1 field inhomogeneity. These imperfections can also be incorporated into the signal model used in the proposed method, although the estimation problem could have increased noise sensitivity due to the increased number of unknowns.

The proposed method predetermines the phase distribution $\{\Psi_m\}_{m=1}^M$ and incorporates it into the signal model in (5.2). In many applications, the phase distribution can be accurately pre-estimated from auxiliary data. This way of formulating the problem simplifies the algorithm and calculation of the performance bounds. However, note that the proposed method can be generalized such that the phase distribution can be treated as unknown. One straightforward extension is to reformulate the problem with the separate real and imaginary components of $\boldsymbol{\rho}$, and modify the algorithm and performance bound calculation accordingly. Alternatively, we can also directly deal with complex $\boldsymbol{\rho}$ in the estimation problem. Note

that this extension requires careful complexification in solving the nonlinear least-squares problem in (5.10) (see [158] and references therein) and calculating estimation-theoretic bounds [159, 160].

The proposed method imposes explicit sparsity constraints through the ℓ_0 quasi-norm. Alternatively, sparsity constraints can also be enforced using other penalty functions (such as the ℓ_1, or total-variation regularizations). With those sparsity constraints, the reconstruction problem can be formulated as a penalized maximum likelihood (PML) estimation problem. In [40], we have presented an investigation along this direction. Preliminary results demonstrate that these two different methods of imposing sparsity constraints can lead to distinct bias-variance trade-offs under different SNR regimes.

The accuracy of the GraSP algorithm for the proposed method depends critically on the accuracy of solving the nonconvex subproblem (5.10), which involves the multiplicative coupling of two optimization variables $\mathbf{u}$ and $\mathbf{c}$. A poor solution from (5.10) can lead the GraSP algorithm to an inaccurate reconstruction. However, in our case, we have empirically observed that with a proper initialization and scaling compensation, the L-BFGS algorithm can converge to a reasonable stationary point, leading to a fairly accurate final reconstruction.

Although the proposed method improves over the ML estimation, there is still a relatively large gap between its empirical performance and the constrained CRB when SNR is low or the undersampling level is high. There are ways to potentially improve the proposed method, and drive its performance closer to the constrained CRB. One possibility is to use a solution algorithm of better accuracy than the current GraSP algorithm. Viable candidates to be studied include [139, 140]. Another possibility is to use stronger sparsity constraints (e.g., the structured sparsity) on $\mathbf{u}$ and $\mathbf{c}$ for the estimation problem. For example, the tree structured sparsity constraints [161] on the wavelet coefficients could be incorporated into (5.8). Note that following a similar procedure in [161], the GraSP algorithm could be also extended to solve the resulting new constrained estimation problem.

Regarding the computational efficiency, it takes about 16 minutes to run the GraSP algorithm for the data set in Section 5.5 on a workstation with a 3.47GHz dual-hex-core Intel Xeon processor X5690, 96 GB RAM, Linux system and Matlab R2012a. Since, at each iteration, the running time for the gradient evaluation and support operation is almost negligible

compared to solving the subproblem (5.10) by the L-BFGS algorithm, the overall efficiency of the algorithm may be considerably improved by using faster numerical solvers for (5.10). The majorization-minimization based algorithms [162] might be a viable alternative, since these algorithms can decompose the original optimization problem into a series of simplified problems, which may lead to acceleration. Also, note that (5.10) is a separable nonlinear least-squares problem (i.e., when $\mathbf{c}_t$ is fixed, it becomes a linear least-squares problem with respect to $\mathbf{u}_t$), thus the variable projection algorithm [133] can be utilized to solve (5.10), which might result in better convergence.

The theoretical analysis with the sparsity constrained CRB relies on the assumption in (5.20). Note that if the coefficients $\mathbf{c}$ and $\mathbf{u}$ are only approximately sparse (or compressible), sparsity constrained CRB is not applicable in theory, although it may still provide some useful insights in practice. In this case, different types of performance bounds may be needed, depending on the class of signal models considered. For example, if we assume certain compressible prior distributions [163] on $\mathbf{c}$ and $\mathbf{u}$, the parameter estimation problem can be posed within the Bayesian estimation framework and the Bayesian CRB [164–166] could be used for performance characterization.

For the sake of simplicity, we used the orthonormal wavelet transform as an example to illustrate the performance of our proposed method. But note that both the algorithm and theoretical analysis can be extended to incorporate other types of sparse representations, such as overcomplete sparsifying transforms or learned dictionaries, which may lead to better performance. Also note that if overcomplete sparsifying transforms are used for $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, although (5.7) is no longer equivalent to (5.8), the synthesis form of the formulation (5.8) still applies. Furthermore, extension to analysis-type sparsifying transforms would also be an interesting problem for future investigation. Some preliminary work has been done using the finite difference within the ℓ_1-PML formulation [40], and it might also be interesting to explore the extension to the ℓ_0 constrained formulation.

In this chapter, we focus on presenting the algorithmic and theoretical aspects of the proposed method with computer simulations. Systematic experimental studies are needed to further evaluate the practical utility of the proposed method. In this case, some practical issues also need to be taken into account, such as the generalization to multichannel

acquisitions [167] and compensation of effects from sequence imperfection [168].

5.7 Summary

This chapter presented a novel constrained reconstruction method to accelerate MR parameter mapping with sparse sampling. It directly reconstructs parameter maps from highly undersampled, noisy **k**-space data, utilizing a parametric physical model while imposing sparsity constraints on the parameter values. An efficient greedy-pursuit algorithm was developed to solve the underlying optimization problem. The properties and performance of the proposed method were analyzed theoretically using estimation-theoretic bounds and also illustrated empirically using a T_2 mapping application example.

5.8 Miscellaneous Derivations

5.8.1 Calculation of Gradients

In this section, we present a detailed derivation for the gradients $\frac{\partial \Xi}{\partial \mathbf{c}}$ and $\frac{\partial \Xi}{\partial \mathbf{u}}$, and also the gradient for the cost function in (5.10). First, we derive $\frac{\partial \Xi}{\partial \mathbf{c}}$ and $\frac{\partial \Xi}{\partial \mathbf{u}}$. Recall that $\Xi(\mathbf{c}, \mathbf{u})$ denotes the cost function in (5.8). For simplicity, we introduce an auxiliary variable $\tilde{\mathbf{I}}_m$ defined as

$$\tilde{\mathbf{I}}_m = \mathbf{\Phi}_m(\boldsymbol{\theta})\boldsymbol{\rho}. \tag{5.34}$$

Using the chain rule of the derivative, we have

$$\begin{aligned}
\frac{\partial \Xi}{\partial \mathbf{c}} &= \sum_{m=1}^{M} \left(\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{c}}\right)^T \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}}\right)^T \frac{\partial \Xi}{\partial \tilde{\mathbf{I}}_m} \\
&= \sum_{m=1}^{M} \mathbf{W}_1 \frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}} \mathbf{q}_m \\
\frac{\partial \Xi}{\partial \mathbf{u}} &= \sum_{m=1}^{M} \left(\frac{\partial \boldsymbol{\rho}}{\partial \mathbf{u}}\right)^T \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\rho}}\right)^T \frac{\partial \Xi}{\partial \tilde{\mathbf{I}}_m} \\
&= \sum_{m=1}^{M} \mathbf{W}_2 \boldsymbol{\Phi}_m (\mathbf{W}_1^T \mathbf{c}) \mathbf{q}_m,
\end{aligned} \tag{5.35}$$

where $\frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}}$ is a diagonal matrix with $\left[\frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}}\right]_{n,n} = \boldsymbol{\rho}_n \frac{\partial \phi}{\partial \theta}|_{\{\theta_n, \gamma_m\}}$, and

$$\mathbf{q}_m = \mathrm{Re}\left\{ (\mathbf{F}_m \boldsymbol{\Psi}_m)^H \left[\mathbf{F}_m \boldsymbol{\Psi}_m \tilde{\mathbf{I}}_m - \mathbf{d}_m \right] \right\}. \tag{5.36}$$

We then derive the gradient for (5.10). Denoting the cost function value of (5.10) by $\tilde{\Xi}$, its derivative with respect to $\mathbf{c}_t$ and $\mathbf{u}_t$ can be written as

$$\begin{aligned}
\frac{\partial \tilde{\Xi}}{\partial \mathbf{c}_t} &= \left(\frac{\partial \mathbf{c}}{\partial \mathbf{c}_t}\right)^T \frac{\partial \tilde{\Xi}}{\partial \mathbf{c}} \\
&= \sum_{m=1}^{M} \hat{\mathbf{E}}^T \mathbf{W}_1 \frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}} \mathbf{q}_m, \\
\frac{\partial \tilde{\Xi}}{\partial \mathbf{u}_t} &= \left(\frac{\partial \mathbf{u}}{\partial \mathbf{u}_t}\right)^T \frac{\partial \tilde{\Xi}}{\partial \mathbf{u}} \\
&= \sum_{m=1}^{M} \tilde{\mathbf{E}}^T \mathbf{W}_2 \boldsymbol{\Phi}_m \left(\mathbf{W}_1^T \mathbf{c}\right) \mathbf{q}_m.
\end{aligned} \tag{5.37}$$

5.8.2 Calculation of Fisher Information Matrix

In this section, we derive the FIM for the data model (5.12). Based on the definition of FIM [63], the partitioned FIM is formed as follows:

$$\begin{aligned}\mathbf{J} &= \begin{bmatrix} \mathbf{J}_{\mathbf{c},\mathbf{c}} & \mathbf{J}_{\mathbf{c},\mathbf{u}} \\ \mathbf{J}_{\mathbf{u},\mathbf{c}} & \mathbf{J}_{\mathbf{u},\mathbf{u}} \end{bmatrix} \\ &= -\begin{bmatrix} \mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{c}^2}\right] & \mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{u}\partial \mathbf{c}}\right] \\ \\ \mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{c}\partial \mathbf{u}}\right] & \mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{u}^2}\right] \end{bmatrix},\end{aligned} \tag{5.38}$$

where $\mathbf{d} = [\mathbf{d}_1^H, \cdots, \mathbf{d}_M^H]^H$. In the following, we derive the expression for each submatrix of $\mathbf{J}$.

Assuming $\mathbf{n}_m$ is complex white Gaussian noise, i.e., $\mathbf{n}_m \sim N(0, \sigma^2)$, the log-likelihood function $\ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})$ can be expressed as

$$\begin{aligned}\ln p(\mathbf{d}; \mathbf{u}, \mathbf{c}) &= \sum_{m=1}^{M} \ln p(\mathbf{d}_m; \mathbf{u}, \mathbf{c}) \\ &= -\frac{1}{\sigma^2} \sum_{m=1}^{M} \|\mathbf{F}_m \mathbf{\Psi}_m \mathbf{\Phi}_m(\mathbf{W}_1^T \mathbf{c}) \mathbf{W}_2^T \mathbf{u} - \mathbf{d}_m\|_2^2 \\ &\quad - \left(\ln \pi + \ln \sigma^2\right) \sum_{m=1}^{M} P_m.\end{aligned} \tag{5.39}$$

The first-order derivative of $\ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})$ with respect to $\boldsymbol{\rho}$ and $\mathbf{c}$ can be expressed as:

$$\begin{aligned}\frac{\partial \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \mathbf{u}} &= -\frac{2}{\sigma^2} \sum_{m=1}^{M} \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}}\right)^T \mathbf{q}_m, \\ \frac{\partial \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \mathbf{c}} &= -\frac{2}{\sigma^2} \sum_{m=1}^{M} \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}}\right)^T \mathbf{q}_m,\end{aligned} \tag{5.40}$$

where

$$\begin{aligned}\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}} &= \frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\rho}} \frac{\partial \boldsymbol{\rho}}{\partial \mathbf{u}} = \boldsymbol{\Phi}_m \left(\mathbf{W}_1^T \mathbf{c}\right) \mathbf{W}_2^T \in \mathbb{R}^{N \times N}, \\ \frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}} &= \frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{c}} = \frac{\partial \tilde{\mathbf{I}}_m}{\partial \boldsymbol{\theta}} \mathbf{W}_1^T \in \mathbb{R}^{N \times N},\end{aligned} \tag{5.41}$$

and also recall that $\mathbf{q}_m$ is defined in (5.36).

Taking the second-order derivatives yields

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \mathbf{u}^2} &= -\frac{2}{\sigma^2} \sum_{m=1}^{M} \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}}\right)^T \mathrm{Re}\left\{\mathbf{G}_m\right\} \frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}}, \\ \frac{\partial^2 \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \mathbf{u} \partial \mathbf{c}} &= -\frac{2}{\sigma^2} \sum_{m=1}^{M} \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}}\right)^T \mathrm{Re}\left\{\mathbf{G}_m\right\} \frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}} \\ &\quad - \frac{2}{\sigma^2} \sum_{m=1}^{M} \mathbf{W}_1 \mathrm{Diag}\left\{\mathbf{q}_m\right\} \boldsymbol{\Lambda}_m \mathbf{W}_2^T, \\ \frac{\partial^2 \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \mathbf{c} \partial \mathbf{u}} &= -\frac{2}{\sigma^2} \sum_{m=1}^{M} \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}}\right)^T \mathrm{Re}\left\{\mathbf{G}_m\right\} \frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}} \\ &\quad - \frac{2}{\sigma^2} \sum_{m=1}^{M} \mathbf{W}_2 \mathrm{Diag}\left\{\mathbf{q}_m\right\} \boldsymbol{\Lambda}_m \mathbf{W}_1^T, \\ \frac{\partial^2 \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \mathbf{c}^2} &= -\frac{2}{\sigma^2} \sum_{m=1}^{M} \left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}}\right)^T \mathrm{Re}\left\{\mathbf{G}_m\right\} \frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}} \\ &\quad - \frac{2}{\sigma^2} \sum_{m=1}^{M} \mathbf{W}_1 \mathrm{Diag}\left\{\mathbf{q}_m\right\} \mathbf{H}_m \mathbf{W}_1^T,\end{aligned} \tag{5.42}$$

where $\mathbf{G}_m = \left(\mathbf{F}_m \boldsymbol{\Psi}_m\right)^H \mathbf{F}_m \boldsymbol{\Psi}_m$, $\boldsymbol{\Lambda}_m$ is a diagonal matrix with $\left[\boldsymbol{\Lambda}_m\right]_{n,n} = \frac{\partial \phi}{\partial \theta}|_{(\mathfrak{d}_n, \gamma_m)}$, $\mathbf{H}_m$ is a diagonal matrix with $\left[\mathbf{H}_m\right]_{n,n} = \boldsymbol{\rho}_n \frac{\partial^2 \phi}{\partial \theta^2}|_{(\mathfrak{d}_n, \gamma_m)}$, and $\mathrm{Diag}\left\{\mathbf{a}\right\}$ converts a vector $\mathbf{a}$ into a diagonal matrix such that $\left[\mathrm{Diag}\left\{\mathbf{a}\right\}\right]_{n,n} = \mathbf{a}_n$.

Evaluating the expectation of the expressions in (5.42) with respect to $\mathbf{n}_m$ yields

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{u}^2}\right] &= -\frac{2}{\sigma^2}\sum_{m=1}^{M}\left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}}\right)^T \mathrm{Re}\{\mathbf{G}_m\}\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}},\\
\mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{c}\partial \mathbf{u}}\right] &= -\frac{2}{\sigma^2}\sum_{m=1}^{M}\left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}}\right)^T \mathrm{Re}\{\mathbf{G}_m\}\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}},\\
\mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{u}\partial \mathbf{c}}\right] &= -\frac{2}{\sigma^2}\sum_{m=1}^{M}\left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}}\right)^T \mathrm{Re}\{\mathbf{G}_m\}\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{u}},\\
\mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{d};\mathbf{u},\mathbf{c})}{\partial \mathbf{c}^2}\right] &= -\frac{2}{\sigma^2}\sum_{m=1}^{M}\left(\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}}\right)^T \mathrm{Re}\{\mathbf{G}_m\}\frac{\partial \tilde{\mathbf{I}}_m}{\partial \mathbf{c}},
\end{aligned} \tag{5.43}$$

In obtaining (5.43), we use the fact that

$$\mathrm{E}[\mathbf{q}_m] = \mathbf{0}, \tag{5.44}$$

based on the assumption that $\mathbf{n}_m$ is white Gaussian noise.

5.8.3 Asymptotics of Oracle Estimator

In this section, we establish the following asymptotic properties for the oracle estimator in (5.32).

Theorem 1. *Assuming that both the data model* (5.12) *and sparsity model* (5.20) *hold, and that* $\mathbf{n}_m$ *is complex white Gaussian noise, the oracle estimator* $\{\hat{\mathbf{u}}_o, \hat{\mathbf{c}}_o\}$ *in* (5.32) *is asymptotically unbiased. Furthermore, the covariance of* $\{\hat{\mathbf{u}}_o, \hat{\mathbf{c}}_o\}$ *also asymptotically achieves the constrained CRB, i.e.,* $\mathbf{A}(\mathbf{A}^T\mathbf{J}\mathbf{A})^{\dagger}\mathbf{A}^T$, *where* $\mathbf{A} = \begin{bmatrix} \mathbf{E_c} & \mathbf{0} \\ \mathbf{0} & \mathbf{E_u} \end{bmatrix}$ *is defined in* (5.22).

Proof: First, note that the oracle estimator in (5.32) can be determined by solving the following unconstrained optimization problem, i.e.,

$$\min_{\mathbf{u}_s,\mathbf{c}_s} \sum_{m=1}^{M} \left\|\mathbf{d}_m - \mathbf{F}_m\mathbf{\Psi}_m\mathbf{\Phi}_m(\mathbf{W}_1^T\mathbf{E_c}\mathbf{c}_s)\mathbf{W}_2^T\mathbf{E_u}\mathbf{u}_s\right\|_2^2, \tag{5.45}$$

through variable change $\mathbf{u} = \mathbf{E}_{\mathbf{u}}\mathbf{u}_s$ and $\mathbf{c} = \mathbf{E}_{\mathbf{c}}\mathbf{c}_s$. The optimal solution of (5.45), $\{\hat{\mathbf{u}}_s, \hat{\mathbf{c}}_s\}$, is the ML estimator with the data model:

$$\mathbf{d}_m = \mathbf{F}_m \mathbf{\Psi}_m \mathbf{\Phi}_m(\mathbf{W}_1^T \mathbf{E}_{\mathbf{c}}\mathbf{c}_s)\mathbf{W}_2^T \mathbf{E}_{\mathbf{u}}\mathbf{u}_s + \mathbf{n}_m. \tag{5.46}$$

Based on the invariance property of the ML estimation (Theorem 7.4 in [63]), the oracle estimator, given by

$$\begin{bmatrix} \hat{\mathbf{u}}_o \\ \hat{\mathbf{c}}_o \end{bmatrix} = \mathbf{A} \begin{bmatrix} \hat{\mathbf{u}}_s \\ \hat{\mathbf{c}}_s \end{bmatrix}, \tag{5.47}$$

is the ML estimator of $\{\mathbf{u}, \mathbf{c}\}$ for the model (5.12) and (5.20), and thus it is asymptotically unbiased.

Furthermore, it is known that the ML estimator $\{\hat{\mathbf{u}}_s, \hat{\mathbf{c}}_s\}$ asymptotically achieves the CRB [63], i.e.,

$$\mathrm{Cov}\left(\begin{bmatrix} \hat{\mathbf{u}}_s \\ \hat{\mathbf{c}}_s \end{bmatrix}\right) \overset{a}{=} \mathbf{J}_s^{\dagger} \tag{5.48}$$

where $\mathbf{J}_s$ is the FIM for the data model (5.46), defined by

$$\mathbf{J}_s = \mathrm{E}\left[\left(\frac{\partial \ln p(\mathbf{d}; \mathbf{u}_s, \mathbf{c}_s)}{\partial \left[\mathbf{u}_s^T, \mathbf{c}_s^T\right]^T}\right)\left(\frac{\partial \ln p(\mathbf{d}; \mathbf{u}_s, \mathbf{c}_s)}{\partial \left[\mathbf{u}_s^T, \mathbf{c}_s^T\right]^T}\right)^T\right], \tag{5.49}$$

which can be rewritten as

$$\begin{aligned}
\mathbf{J}_s &= \mathrm{E}\left[\left(\mathbf{A}^T \frac{\partial \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \left[\mathbf{u}^T, \mathbf{c}^T\right]^T}\right)\left(\mathbf{A}^T \frac{\partial \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \left[\mathbf{u}^T, \mathbf{c}^T\right]^T}\right)^T\right] \\
&= \mathbf{A}^T \mathrm{E}\left[\left(\frac{\partial \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \left[\mathbf{u}^T, \mathbf{c}^T\right]^T}\right)\left(\frac{\partial \ln p(\mathbf{d}; \mathbf{u}, \mathbf{c})}{\partial \left[\mathbf{u}^T, \mathbf{c}^T\right]^T}\right)^T\right]\mathbf{A} \\
&= \mathbf{A}^T \mathbf{J} \mathbf{A}.
\end{aligned} \tag{5.50}$$

Finally, taking the covariance on the both sides of (5.47) yields

$$\begin{aligned}\mathrm{Cov}\left(\begin{bmatrix}\hat{\mathbf{u}}_o\\ \hat{\mathbf{c}}_o\end{bmatrix}\right) &= \mathbf{A}\mathrm{Cov}\left(\begin{bmatrix}\hat{\mathbf{u}}_s\\ \hat{\mathbf{c}}_s\end{bmatrix}\right)\mathbf{A}^T \\ &\stackrel{a}{=} \mathbf{A}(\mathbf{A}\mathbf{J}\mathbf{A}^T)^{\dagger}\mathbf{A}^T, \end{aligned} \tag{5.51}$$

which establishes the second half of the theorem.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

With the ever-increasing dimension of MR imaging applications, sparse sampling has become a critical means to break the curse of dimensionality. In this dissertation, we address the problem of accelerating high-dimensional MRI with constrained imaging approaches. We have developed novel imaging models and reconstruction algorithms to enable high-quality MRI with highly undersampled data. The key contributions of this dissertation are summarized as follows.

- We have proposed a novel constrained image model, i.e., the joint low-rank and sparsity model, to enable dynamic image reconstruction from highly undersampled $(\mathbf{k}, t)$-space data. With the design of specialized data acquisition schemes, we pre-estimate the temporal subspace for the low-rank model. A convex formulation has been proposed to integrate the low-rank constraint, subspace constraint, with sparsity constraint. We have developed a globally convergent algorithm to efficiently solve the resulting optimization problem. The performance of the proposed method has been systematically evaluated in the real-time cardiac imaging application, demonstrating that the significantly improved performance can be achieved by using the proposed joint constraints compared to the state-of-the-art methods using an individual low-rank or sparsity constraint.

- We have extended the proposed joint low-rank and sparsity model to accelerate quantitative MR parameter mapping. We have specifically adapted the model to capture different signal/image characteristics for parameter mapping applications: the low-

rank constraint is used to capture the correlation of relaxation signals associated with different types of tissues, whereas the joint sparsity constraint is used to model the sparse support shared by a sequence of co-registered contrast-weighted images. Furthermore, we have integrated the proposed image model with multichannel data acquisition, achieving an even higher level of acceleration. The superior performance of the proposed method has been demonstrated two in-vivo MR parameter mapping applications.

- We have proposed a novel model-based parameter estimation approach to accelerate MR parameter mapping. With the use of the parametric physical model and sparsity constraints, we have reformulated the problem as a constrained parameter estimation problem with limited, noisy data. We have developed an efficient greedy-pursuit algorithm to solve the resulting optimization problem, and also derived the estimation-theoretical bounds to characterize the performance of the proposed estimator. We have illustrated the theoretical properties and empirical performance of the proposed method with a T_2 mapping application example.

6.2 Future Work

There are several directions that are worth exploring in the future research.

- We have demonstrated the benefits of enforcing the joint low-rank and sparsity constraints with empirical results from simulated and in-vivo imaging data. It is worthwhile to gain better understanding about this simultaneously structured signal model from a theoretical perspective. For example, it is useful to understand the information-theoretic or estimation-theoretic limit of the model for sparse sampling.

- In this dissertation, all of the proposed methods are associated with nonlinear reconstruction processes. To the imaging community, establishing the resolution and noise properties for these nonlinear reconstruction procedures is critical, but is still an open problem. With a breakthrough in this direction, we would be able to effectively se-

lect appropriate constrained imaging models based on the desired resolution and noise characteristics of imaging experiments.

- For the parameter estimation for quantitative MRI, we focused on the use of a simplified physical model to study the properties of the proposed method. For certain imaging applications, it is necessary to incorporate other factors from imaging physics into the signal model. It would be useful to investigate the performance of the proposed method with those complex physical models.

- We have derived the estimation-theoretic bounds to study the estimation performance for MR parameter mapping. With the aid of these powerful tools, we would be able to "optimally" design quantitative MR imaging experiments, which can have the capability of collecting the most informative samples.

- This dissertation is focused on a proof-of-concept study of our proposed constrained reconstruction methods. Systematic assessment of the clinical utility of these methods with respect to specific clinical applications would be an important step forward.

REFERENCES

[1] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. New York: IEEE Press, 2000.

[2] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2007, pp. 988–991.

[3] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature.*, vol. 242, pp. 190–191, 1973.

[4] J. P. Finn, K. Nael, V. Deshpande, O. Ratib, and G. Laub, "Cardiac MR imaging: State of the technology," *Radiology.*, vol. 241, pp. 338–354, 2006.

[5] M. Markl, A. Frydrychowicz, S. Kozerke, M. Hope, and O. Wieben, "4D flow MRI," *J. Magn. Reson. Imaging*, vol. 36, pp. 1015–1036, 2012.

[6] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," *IEEE Signal Process. Mag.*, vol. 25, pp. 123–132, 2008.

[7] P. L. Choyke, A. J. Dwyer, and M. V. Knopp, "Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging," *J. Magn. Reson. Imaging*, vol. 17, pp. 509–520, 2003.

[8] P. J. Basser and D. K. Jones, "Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review," *NMR Biomed.*, vol. 15, pp. 456–467, 2002.

[9] R. B. Buxton, *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge: Cambridge University Press, 2002.

[10] S. Posse, R. Otazo, S. R. Dager, and J. Alger, "MR spectroscopic imaging: Principles and recent advances," *J. Magn. Reson. Imaging*, vol. 37, pp. 1301–1325, 2013.

[11] R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," *Science*, vol. 269, pp. 1854–1857.

[12] M. Fu, B. Zhao, C. Carignan, R. K. Shosted, J. L. Perry, D. P. Kuehn, Z.-P. Liang, and B. P. Sutton, "High-resolution dynamic speech imaging with joint low-rank and sparsity constraints," *Magn. Reson. Med.*, 2014, in press.

[13] P. Mansfield, "Multi-planar image formation using NMR spin echoes," *J. Phy. C: Solid State Phys*, vol. 10, p. L55, 1977.

[14] J. Hennig, A. Nauerth, and H. Friedburg, "RARE imaging: A fast imaging method for clinical MR," *Magn. Reson. Med.*, vol. 67, pp. 823 – 833, 1986.

[15] A. Haase, J. Frahm, D. Matthaei, W. Hanicke, and K.-D. Merboldt, "FLASH imaging. rapid NMR imaging using low flip-angle pulses," *J. Magn. Reson.*, vol. 67, no. 2, pp. 258 – 266, 1986.

[16] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. Burlington: Elsevier Academic Press, 2004.

[17] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. Sunshine, J. Duerk, and M. Griswold, "Magnetic resonance fingerprinting," *Nature*, vol. 495, pp. 187–192, 2013.

[18] D. Sodickson and W. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, no. 4, pp. 591–603, 1997.

[19] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, pp. 952–962, 1999.

[20] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, pp. 1202–1210, 2002.

[21] M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magn. Reson. Med.*, vol. 64, no. 2, pp. 457–471, 2010.

[22] L. Ying and Z.-P. Liang, "Parallel MRI using phased array coils," *IEEE Signal Process. Mag.*, vol. 27, pp. 90–98, 2010.

[23] Z.-P. Liang, F. E. Boada, R. T. Constable, E. M. Haacke, P. C. Lauterbur, and M. R. Smith, "Constrained reconstruction methods in MR imaging," *Rev. Magn. Reson. Med.*, vol. 4, pp. 67 – 185, 1992.

[24] Z.-P. Liang and P. C. Lauterbur, "An efficient method for dynamic magnetic resonance imaging," *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 677–686, 1994.

[25] Z.-P. Liang, H. Jiang, C. P. Hess, and P. C. Lauterbur, "Dynamic imaging by model estimation," *Int. J. Imaging Syst. Tech.*, vol. 8, pp. 551–557, 1997.

[26] B. Madore, G. H. Glover, and N. J. Pelc, "Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI," *Magn. Reson. Med.*, vol. 42, pp. 813–828, 1999.

[27] J. Tsao, P. Boesiger, and K. P. Pruessmann, "k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations," *Magn. Reson. Med.*, vol. 50, pp. 1031–1042, 2003.

[28] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.

[29] N. Aggarwal and Y. Bresler, "Patient-adapted reconstruction and acquisition dynamic imaging method (PARADIGM) for MRI," *Inverse Probl.*, vol. 24, p. 045015, 2008.

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

[31] B. Zhao, J. Haldar, A. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (**k**, t) -space data with joint partial separability and sparsity constraints," *IEEE Trans. Med. Imaging*, vol. 31, pp. 1809–1820, 2012.

[32] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, "Low rank matrix recovery for real-time cardiac MRI," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2010, pp. 996–999.

[33] B. Zhao, J. Haldar, and Z.-P. Liang, "PSF model-based reconstruction with sparsity constraint: Algorithm and application to real-time cardiac MRI," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 3390–3393.

[34] B. Zhao, J. Haldar, A. Christodoulou, and Z.-P. Liang, "Further development of image reconstruction from highly undersampled (**k**, t)-space data with joint partial separability and sparsity constraints," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2011, pp. 1593–1596.

[35] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (k, t)-space with joint partial separability and sparsity constraints," in *Proc. Int. Symp. Magn. Reson. Med.*, 2011, p. 4375.

[36] B. Zhao, A. Sun, K. Ma, R. Li, A. G. Christodoulou, C. Yuan, and Z.-P. Liang, "Real-time phase contrast cardiovascular flow imaging with joint low-rank and sparsity constraints," in *Proc. Int. Symp. Magn. Reson. Med.*, 2014, p. 743.

[37] B. Zhao, W. Lu, and Z.-P. Liang, "Highly accelerated parameter mapping with joint partial separability and sparsity constraints," in *Proc. Int. Symp. Magn. Reson. Med.*, 2012, p. 2233.

[38] B. Zhao, T. K. Hitchens, A. G. Christodoulou, F. Lam, Y. L. Wu, C. Ho, and Z.-P. Liang, "Accelerated 3D UTE relaxometry for quantification of iron-oxide labeled cells," in *Proc. Int. Symp. Magn. Reson. Med.*, 2013, p. 2455.

[39] B. Zhao, W. Lu, T. K. Hitchens, F. Lam, C. Ho, and Z.-P. Liang, "Accelerated MR parameter mapping with low-rank and sparsity constraints," *Magn. Reson. Med.*, 2014, in press.

[40] B. Zhao, F. Lam, W. Lu, and Z.-P. Liang, "Model-based MR parameter mapping with sparsity constraint," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2013, pp. 1–4.

[41] B. Zhao, F. Lam, W. Lu, and Z.-P. Liang, "Model-based MR parameter mapping with sparsity constraint," in *Proc. Int. Symp. Magn. Reson. Med.*, 2013, p. 2459.

[42] B. Zhao, F. Lam, and Z.-P. Liang, "Model-based MR parameter mapping with sparsity constraints: Parameter estimation and performance bounds," *IEEE Trans. Med. Imaging*, vol. 33, pp. 1832–1844, 2014.

[43] Y. Censor, "Finite series-expansion reconstruction methods," *Proc. IEEE*, vol. 71, pp. 409–419, 1983.

[44] J. Fessler, "Model-based image reconstruction for MRI," *IEEE Signal Process. Mag.*, vol. 27, pp. 81–89, 2010.

[45] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, pp. 4203–4215, 2005.

[46] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289–1306, 2006.

[47] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Med Phys.*, vol. 37, pp. 5113–5125, 2010.

[48] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans Pattern Anal Mach Intell.*, vol. 31, pp. 210–227, 2009.

[49] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 57, pp. 2275–2284, 2009.

[50] A. C. Gilbert, M. Stauss, J. A. Tropp, and V. R., "One sketch for all: fast algorithms for compressed sensing," in *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, 2007, pp. 237–246.

[51] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[52] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[53] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," *IEEE J. Sel. Top. Sign. Proces*, vol. 1, pp. 606–617, 2007.

[54] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM J. Imaging Sciences*, vol. 1, pp. 143–168, 2008.

[55] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imaging Sciences*, vol. 1, pp. 248–272, 2008.

[56] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, pp. 183–202, 2009.

[57] J. Yang and Y. Zhang, "Alternating direction algorithm for ℓ_1 problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, pp. 250–278, 2011.

[58] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sciences*, vol. 4, pp. 1–39, 2011.

[59] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundation and Trends in Machine Learning*, vol. 3, pp. 1–122, 2011.

[60] S. Ramani and J. Fessler, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans. Med. Imaging*, vol. 30, pp. 694–706, 2011.

[61] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Anal.*, vol. 26, pp. 301–321, 2009.

[62] E. Candès and T. Tao, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.

[63] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River: Prentice Hall, 1993.

[64] A. Niebergall, S. Zhang, E. Kunay, G. Keydana, M. Job, M. Uecker, and J. Frahm, "Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction," *Magnetic Resonance in Medicine*, vol. 69, pp. 477–485, 2013.

[65] M. R. Prince, T. M. Grist, and J. F. Debatin, *3D Contrast MR Angiography.* New York: Springer-Verlag, 1999.

[66] N. P. Willis and Y. Bresler, "Lattice-theoretic analysis of time-sequential sampling of spatiotemporal signals–Part I," *IEEE Trans. Info. Theory*, vol. 43, pp. 190–207, 1997.

[67] D. Xu, K. F. King, and Z.-P. Liang, "Improved k-t SENSE by adaptive regularization," *Magn. Reson. Med.*, vol. 57, pp. 918–930, 2007.

[68] C. A. Mistretta, "Undersampled radial MR acquisition and highly constrained back projection (HYPR) reconstruction: Potential medical imaging applications in the post-Nyquist era," *J. Magn. Reson. Imaging*, vol. 29, pp. 501–516, 2009.

[69] B. Sharif, J. A. Derbyshire, A. Z. Faranesh, and Y. Bresler, "Patient-adaptive reconstruction and acquisition in dynamic imaging with sensitivity encoding (PARADISE)," *Magn. Reson. Med.*, vol. 64, pp. 501–513, 2010.

[70] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. Int. Symp. Magn. Reson. Med.*, 2006, p. 2420.

[71] H. Jung, J. C. Ye, and E. Y. Kim, "Improved k-t BLAST and k-t SENSE using FOCUSS," *Phys. Med. Biol.*, vol. 52, pp. 3201–3226, 2007.

[72] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, pp. 365–373, 2008.

[73] J. Trzasko, C. Haider, E. Borisch, N. Campeau, J. Glockner, S. Riederer, and A. Manduca, "Sparse-CAPR: Highly accelerated 4D CE-MRA with parallel imaging and non-convex compressive sensing," *Magn. Reson. Med.*, pp. 1019–1032, 2011.

[74] J. V. Velikina, K. M. Johnson, W. F. Block, and A. A. Samsonov, "Design of temporally constrained compressed sensing methods for accelerated dynamic MRI," in *Proc. Int. Symp. Magn. Reson. Med.*, 2010, p. 4865.

[75] G. Adluru, C. McGann, P. Speier, E. G. Kholmovski, A. Shaaban, and E. Dibella, "Acquisition and reconstruction of undersampled radial data for myocardial perfusion magnetic resonance imaging," *J. Magn. Reson.*, vol. 29, pp. 466–473, 2009.

[76] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic MR imaging," in *Proc. IEEE Int. Conf. Image Process.*, 2009, pp. 3045–3048.

[77] D. Liang and L. Ying, "Compressed-sensing dynamic MR imaging using partially known support," in *Proc. IEEE Eng. Med. Biol. Conf.*, 2010, pp. 2829–2832.

[78] M. Usman, C. Prieto, T. Schaeffter, and P. G. Batchelor, "k-t group sparse: A method for accelerating dynamic MRI," *Magn. Reson. Med.*, vol. 66, pp. 1163–1176, 2011.

[79] A. Sen Gupta and Z.-P. Liang, "Dynamic imaging by temporal modeling with principle component analysis," in *Proc. Int. Symp. Magn. Reson. Med.*, 2001, p. 10.

[80] J. Ji and Z.-P. Liang, "High resolution cardiac magnetic resonance imaging: A model-based approach," in *Proc. IEEE Eng. Med. Biol. Conf.*, 2001, pp. 2268–2271.

[81] C. Brinegar, S. S. Schmitter, N. N. Mistry, G. A. Johnson, and Z.-P. Liang, "Improving temporal resolution of pulmonary perfusion imaging in rats using the partially separable functions model," *Magn. Reson. Med.*, vol. 64, pp. 1162–1170, 2010.

[82] J. P. Haldar and Z.-P. Liang, "Low-rank approximations for dynamic imaging," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2011, pp. 1052–1055.

[83] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: Temporally constrained k-t BLAST reconstruction using principle component analysis," *Magn. Reson. Med.*, vol. 62, pp. 706–716, 2009.

[84] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis-based four-dimensional computed tomography," *Phys. Med. Biol.*, vol. 56, pp. 3181–3198, 2011.

[85] S. G. Lingala, Y. Hu, E. Dibella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imaging*, vol. 30, pp. 1042–1054, 2011.

[86] J. Trzasko, A. Manduca, and E. Borisch, "Local versus global low-rank promotion in dynamic MRI series reconstruction," in *Proc. Int. Symp. Magn. Reson. Med.*, 2011, p. 4371.

[87] A. Christodoulou, H. Zhang, B. Zhao, T. Hitchens, C. Ho, and Z.-P. Liang, "High-resolution cardiovascular MRI by integrating parallel imaging with low-rank and sparse modeling," *IEEE Trans. Biomed Eng.*, vol. 60, pp. 3083 – 3092, 2013.

[88] J. Trzasko and A. Manduca, "A unified tensor regression framework for calibrationless dynamic, multi-channel MRI reconstruction," in *Proc. Int. Symp. Magn. Reson. Med.*, 2013, p. 603.

[89] Y. Yu, J. Jin, F. Liu, and S. Crozier, "Multidimensional compressed sensing MRI using tensor decomposition-based sparsifying transform," *PLoS ONE*, vol. 9, p. e98441, 2014.

[90] S. Poddar, S. Lingala, and M. Jacob, "Joint recovery of under sampled signals on a manifold: Application to free breathing cardiac MR," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 6904 – 6908.

[91] A. Majumdar, "Improved dynamic {MRI} reconstruction by exploiting sparsity and rank-deficiency," *Magn Reson Imaging*, vol. 31, pp. 789 – 795, 2013.

[92] R. Otazo, E. Cands, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, 2014, in press.

[93] B. Tremoulheac, N. Dikaios, D. Atkinson, and S. Arridge, "Dynamic MR image reconstruction - separation from undersampled (k, t)-space via low-rank plus sparse prior," *IEEE Trans. Med. Imaging*, vol. 33, pp. 1689–1701, 2014.

[94] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, pp. 948–958, 2010.

[95] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont: Athena Scientific, 1999.

[96] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, pp. 932–946, 1995.

[97] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, pp. 937–966, 2005.

[98] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Info. Theory*, vol. 41, pp. 613–627, 1995.

[99] L. N. Trefethen and D. B. III, *Numerical linear algebra*. Philadelphia: SIAM, 1997.

[100] Y.-C. Tsai, H.-D. Lin, Y.-C. Hu, C.-L. Yu1, and K.-P. Lin, "Thin-plate spline technique for medical image deformation," *J. Med. Biol. Eng.*, vol. 20, no. 4, pp. 203–210, 2000.

[101] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. 215–220, 2000.

[102] S. G. Lingala, Y. Hu, and M. Jacob, "Real-time cardiac MRI using low-rank and sparsity penalties," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2010, pp. 988–991.

[103] J. P. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using PowerFactorization," *IEEE Signal Process. Lett.*, vol. 16, pp. 584–587, 2009.

[104] K. Lee and Y. Bresler, "ADMiRA: Atomic decomposition for minimum rank approximation," *IEEE Trans. Info. Theory*, pp. 4402–4416, 2010.

[105] W. Dai and O. Milenkovic, "SET: An algorithm for consistent matrix completion," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010.

[106] S. G. Lingala, Y. Hu, E. Dibella, and M. Jacob, "Accelerated first pass cardiac perfusion MRI using improved k-t SLR," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2011, pp. 1280–1283.

[107] J. A. d' Arcy, D. J. Collins, I. J. Rowland, A. R. Padhanni, and M. O. Leach, "Applications of sliding window reconstruction with Cartesian sampling for dynamic contrast enhanced MRI," *NMR Biomed.*, pp. 174–183, 2002.

[108] M. O. Ulfarsson and V. Solo, "Dimension estimation in noisy PCA with SURE and random matrix theory," *IEEE Trans. Signal Process.*, vol. 56, pp. 5804–5816, 2008.

[109] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia: SIAM, 2002.

[110] U. Sümbül, J. M. Santos, and J. M. Pauly, "Improved time series reconstruction for dynamic magnetic resonance imaging," *IEEE Trans. Med. Imaging*, vol. 28, pp. 1093–1104, 2009.

[111] A. G. Christodoulou, B. Zhao, H. Zhang, C. Ho, and Z.-P. Liang, "Four-dimensional MR cardiovascular imaging: Method and applications," in *Proc. IEEE Eng. Med. Biol. Conf.*, 2011, pp. 3732–3735.

[112] J. P. Haldar, "Constrained imaging: Denoising and sparse sampling," Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign, Urbana, IL, 2011.

[113] A. G. Christodoulou, H. Zhang, B. Zhao, Q. Ye, T. K. Hitchens, C. Ho, and Z.-P. Liang, "High resolution 3D first pass myocardial perfusion imaging," in *Proc. Int. Symp. Magn. Reson. Med.*, 2012, p. 1151.

[114] F. Lam, B. Zhao, Y. Liu, Z.-P. Liang, M. Weiner, and N. Schuff, "Accelerated fMRI using low-rank model and sparsity constraints," in *Proc. Int. Symp. Magn. Reson. Med.*, 2013, p. 2620.

[115] H.-L. Margaret Cheng, N. Stikov, N. R. Ghugre, and G. A. Wright, "Practical medical applications of quantitative MR relaxometry," *J. Magn. Reson. Imaging*, vol. 36, pp. 805–824, 2012.

[116] S. Baudrexel, L. Nurnberger, U. Rub, C. Seifried, J. C. Klein, T. Deller, H. Steinmetz, R. Deichmann, and R. Hilker, "Quantitative mapping of T_1 and T_2^* discloses nigral and brainstem pathology in early Parkinson's disease," *NeuroImage*, vol. 51, pp. 512 – 520, 2010.

[117] J. S. Hankins, M. B. McCarville, R. B. Loeffler, M. P. Smeltzer, M. Onciu, F. A. Hoffer, C.-S. Li, W. C. Wang, R. E. Ware, and C. M. Hillenbrand, "R_2^* magnetic resonance imaging of the liver in patients with iron overload," *Blood*, vol. 113, pp. 4853–4855.

[118] D. R. Messroghli, K. Walters, S. Plein, P. Sparrow, M. G. Friedrich, J. P. Ridgway, and M. U. Sivananthan, "Myocardial T_1 mapping: Application to patients with acute and chronic myocardial infarction," *Magn. Reson. Med.*, vol. 58, pp. 34–40, 2007.

[119] W. Liu, H. Dahnke, J. Rahmer, E. K. Jordan, and J. A. Frank, "Ultrashort T_2^* relaxometry for quantitation of highly concentrated superparamagnetic iron oxide (SPIO) nanoparticle labeled cells," *Magn. Reson. Med.*, vol. 61, pp. 761–766, 2009.

[120] S. C. Deoni, B. K. Rutt, and T. M. Peters, "Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state," *Magn. Reson. Med.*, vol. 49, no. 3, pp. 515–526, 2003.

[121] J. V. Velikina, A. L. Alexander, and A. Samsonov, "Accelerating MR parameter mapping using sparsity-promoting regularization in parametric dimension," *Magn. Reson. Med.*, in press.

[122] M. Doneva, P. Bornert, H. Eggers, C. Stehning, J. Senegas, and A. Mertins, "Compressed sensing reconstruction for magnetic resonance parameter mapping," *Magn. Reson. Med.*, vol. 64, pp. 1114–1120, 2010.

[123] L. Feng, R. Otazo, H. Jung, J. H. Jensen, J. C. Ye, D. K. Sodickson, and D. Kim, "Accelerated cardiac T_2 mapping using breath-hold multiecho fast spin-echo pulse sequence with k-t FOCUSS," *Magn. Reson. Med.*, vol. 65, pp. 1661–1669, 2011.

[124] B. Bilgic, V. K. Goyal, and E. Adalsteinsson, "Multi-contrast reconstruction with Bayesian compressed sensing," *Magn. Reson. Med.*, vol. 66, pp. 1601–1615, 2011.

[125] A. Majumdar and R. K. Ward, "Accelerating multi-echo T_2 weighted MR imaging: Analysis prior group-sparse optimization," *J. Magn. Reson. Imaging*, vol. 210, pp. 90 – 97, 2011.

[126] J. Yuan, D. Liang, F. Zhao, Y. Li, Y. X. Xiang, and L. Ying, "k-t ISD compressed sensing reconstruction for T_1 mapping: A study in rat brains at 3T," in *Proc. Int. Symp. Magn. Reson. Med.*, 2012, p. 4197.

[127] J. Huang, C. Chen, and L. Axel, "Fast multi-contrast MRI reconstruction," in *Proc. MICCAI*, 2012, pp. 281–288.

[128] F. H. Petzschner, I. P. Ponce, M. Blaimer, P. M. Jakob, and F. A. Breuer, "Fast MR parameter mapping using k-t principal component analysis," *Magn. Reson. Med.*, vol. 66, pp. 706–716, 2011.

[129] C. Huang, C. G. Graff, E. W. Clarkson, A. Bilgin, and M. I. Altbach, "T_2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing," *Magn. Reson. Med.*, vol. 67, pp. 1355–1366, 2012.

[130] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Super-resolution reconstruction of MR image sequences with contrast modeling," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2009, pp. 266–269.

[131] K. Block, M. Uecker, and J. Frahm, "Model-based iterative reconstruction for radial fast spin-echo MRI," *IEEE Trans. Med. Imaging*, vol. 28, pp. 1759 –1769, 2009.

[132] T. J. Sumpf, M. Uecker, S. Boretius, and J. Frahm, "Model-based nonlinear inverse reconstruction for T_2 mapping using highly undersampled spin-echo MRI," *J. Magn. Reson. Imaging*, vol. 34, pp. 420–428, 2011.

[133] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Probl.*, vol. 19, pp. R1–R26, 2003.

[134] J. P. Haldar, J. Anderson, and S. W. Sun, "Maximum likelihood estimation of T1 relaxation parameters using VARPRO," 2007, p. 41.

[135] J. Trasko, P. M. Mostardi, S. J. Riederer, and A. Manduca, "A simplified nonlinear fitting strategy for estimating T1 from variable flip angle sequences," 2011, p. 4561.

[136] J. K. Barral, E. Gudmundson, N. Stikov, M. Etezadi-Amoli, P. Stoica, and D. G. Nishimura, "A robust methodology for in vivo T1 mapping," *Magn. Reson. Med.*, vol. 64, pp. 1057–1067, 2010.

[137] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans. Image Process.*, vol. 21, pp. 3659–3672, 2012.

[138] W. Li, M. Griswold, and X. Yu, "Fast cardiac T_1 mapping in mice using a model-based compressed sensing method," *Magn. Reson. Med.*, vol. 68, pp. 1127–1134, 2012.

[139] A. Beck and Y. C. Eldar, "Sparsity constrained nonlinear optimization: Optimality conditions and algorithms," *SIAM J. Optim.*, in press.

[140] T. Blumensath, "Compressed sensing with nonlinear observations and related nonlinear optimization problems," *IEEE Trans. Inf. Theory*, vol. 59, pp. 3466–3474, 2013.

[141] S. Bahmani, B. Raj, and P. Boufounos, "Greedy sparsity-constrained optimization," *Journal of Machine Learning Research*, vol. 14, pp. 807–841, 2013.

[142] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2230–2249, 2009.

[143] J. Nocedal and S. J. Wright, *Numerical Optimization (2nd Edition)*. New York, NY: Springer, 2006.

[144] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, pp. 1285–1301, 1990.

[145] P. Stoica and B. Ng, "On the Cramer-Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, pp. 177–179, 1998.

[146] Z. Ben-Haim and Y. C. Eldar, "The Cramer-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, pp. 3384–3389, 2010.

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

[148] G. Tang and A. Nehorai, "Constrained Crámer-Rao bound on robust principal component analysis," *IEEE Trans. Signal Process.*, vol. 59, pp. 5070–5076, 2011.

[149] C. A. Rohde, "Generalized inverses of partitioned matrices," *J. Soc. Indust. Appl. Math.*, vol. 13, pp. 1033–1035, 1965.

[150] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, pp. 87–90, 2001.

[151] Z. Ben-Haim and Y. C. Eldar, "On the Cramer-Rao bound with a singular fisher information matrix," *IEEE Signal Process. Lett.*, vol. 16, pp. 453–456, 2009.

[152] Y. C. Eldar, "Rethinking biased estimation: Improving maximum likelihood and the Cramér-Rao bound," *Foundations and Trends in Signal Processing*, vol. 1, pp. 305–449, 2008.

[153] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imaging*, vol. 17, pp. 463–468, 1998.

[154] A. MacKay, C. Laule, I. Vavasour, T. Bjarnason, S. Kolind, and B. Madler, "Insights into brain microstructure from the T_2 distribution," *Magn. Reson. Imaging*, vol. 24, pp. 515–525, 2006.

[155] C. R. McCreary, T. A. Bjarnason, V. Skihar, J. R. Mitchell, V. W. Yong, and J. F. Dunn, "Multiexponential T_2 and magnetization transfer MRI of demyelination and remyelination in murine spinal cord," *NeuroImage*, vol. 45, pp. 1173–1182, 2008.

[156] D. Reiter, R. Roque, P. Lin, T. Bjarnason, S. Kolind, and B. Madler, "Mapping proteoglycan-bound water in cartilage: improved specificity of matrix assessment using multiexponential transverse relaxation analysis," *Magn. Reson. Med.*, vol. 65, pp. 377–384, 2011.

[157] A. K. Funai, "Regularized estimation of main and rf field inhomogeneity and longitudinal relaxation rate in magnetic resonance imaging," Ph.D. dissertation, Univ. of Michigan, Ann Arbor, MI, 2011.

[158] L. Sorber, M. Barel, and L. Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM J. Optimization*, vol. 22, pp. 879–898, 2012.

[159] A. Van Den Bos, "A Cramer-Rao lower bound for complex parameters," *IEEE Trans. Signal Process.*, vol. 42, no. 10, p. 2859, 1994.

[160] A. Jagannatham and B. Rao, "Cramer-Rao lower bound for constrained complex parameters," *IEEE Signal Process. Lett.*, vol. 11, pp. 875–878, 2004.

[161] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, pp. 1982–2001, 2010.

[162] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist*, vol. 58, pp. 30–37, 2004.

[163] R. Gribonval, V. Cevher, and M. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, pp. 5016–5034, 2012.

[164] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. New York: Wiley-IEEE Press, 2007.

[165] P. Chatterjee and P. Milanfar, "Is denoising dead?" *IEEE Trans. Image Process.*, vol. 19, pp. 895–911, 2010.

[166] R. Prasad and C. Murthy, "Cramér-Rao-type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, pp. 622–632, 2013.

[167] J. D. Trzasko, P. M. Mostardi, S. J. Riederer, and A. Manduca, "Estimating T_1 from multichannel variable flip angle SPGR sequences," *Magn. Reson. Med.*, vol. 69, pp. 1787–1794, 2013.

[168] R. M. Lebel and A. H. Wilman, "Transverse relaxometry with stimulated echo compensation," *Magn. Reson. Med.*, vol. 64, pp. 1005–1014, 2010.