CONSTRAINED IMAGING: DENOISING AND SPARSE SAMPLING

BY

JUSTIN P. HALDAR

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, 2011

Urbana, Illinois

Doctoral Committee:

Professor Zhi-Pei Liang, Chair
Professor Yoram Bresler
Associate Professor Paul Scott Carney
Assistant Professor Bradley P. Sutton

Abstract

Magnetic resonance imaging (MRI) is a powerful tool for studying the anatomy, physiology, and metabolism of biological systems. Despite the fact that MRI was introduced decades ago and has already revolutionized medical imaging, current applications are still far from utilizing the full potential of the MR signal. Traditional MRI data acquisition and image reconstruction methods are based on simple Fourier inversion, leading to undesirable trade-offs between image resolution, signal-to-noise ratio (SNR), and data acquisition time. Classical approaches to addressing these trade-offs have relied on improved imaging hardware and more efficient pulse sequences. In contrast, our work addresses the limitations of MR using relatively less-explored signal processing approaches, which have recently become practical because of increasing computational capabilities. This dissertation concerns the use of constrained imaging models to guide the design of both data acquisition and image reconstruction, leading to improved imaging performance in the context of both noise-limited and resolution-limited scenarios.

To address noise limitations for high-resolution imaging, we introduce a quasi-Bayesian edge-preserving smoothness prior for modeling correlated image sequences. The prior models the correlated edge structures that are observed in the image sequence, and is used within a penalized maximum likelihood framework to reduce image noise while preserving high-resolution anatomical structure. In contrast to many constrained imaging methods, we demonstrate that the proposed method is relatively simple to analyze and is robust to model inaccuracy when reconstruction parameters are chosen appropriately. Resolution and SNR analysis shows that the proposed formulations lead to substantial improvements in SNR with only a moderate decrease in spatial resolution. An examination of resolution and SNR trade-offs is presented, which serves as a guide for the

optimal design of data acquisition and image reconstruction procedures in this context.

To address limited spatial resolution in high-SNR scenarios, we design specialized data acquisition and image reconstruction procedures to enable image reconstruction from sparsely-sampled data. Specifically, we leverage prior information that the image has sparse or low-rank structure to significantly reduce sampling requirements in two different contexts. In the first context, we assume that the image is sparse in a known transform domain, and develop a novel non-Fourier data acquisition scheme to enable high-quality reconstruction from undersampled data. The second context is specific to spatiotemporal imaging, and it is assumed that the temporal evolution of the spatiotemporal image is highly correlated at different spatial positions. This correlation leads to the formulation of a novel low-rank matrix recovery problem, which we demonstrate can be solved efficiently and effectively using special algorithms.

Applications of the proposed techniques are illustrated with simulated and experimental data from a variety of different MR imaging scenarios.

To my mother

Acknowledgments

This dissertation would not have been possible without the help and support from a large number of people. Though I can't list all of them, I would like to express my thanks to everyone who helped me get to this point – I am indebted to you all for the inspiration you have given me.

I am grateful to my advisor, Professor Zhi-Pei Liang, for his support and guidance throughout my graduate studies, and for providing me with so many opportunities for scientific, professional, and personal growth. I would also like to thank the other members of my doctoral committee — Professors Yoram Bresler, Brad Sutton, and Scott Carney — for their time, effort, and help, and for their influence on my training and research.

I am thankful for my labmates at the Beckman Institute — especially Diego Hernando, Chao Ma, Hien Nguyen, Bo Zhao, Fan Lam, Anthony Christodoulou, Xiaobo Qu, Dan Xu, Tim Deller, Yong Yue, Loan Vo, David Huang, John Stastny, and Zhubin Gahvari — for the friendly and stimulating work environment.

I am also grateful to Dr. Victor Song for hosting me at the Washington University in St. Louis, and for teaching me about animal MRI during our collaborations. Thanks also to Joong Hee Kim, Jeff Anderson, Matt Budde, Qing Wang, William Tu, Yong Wang, Gordon Xu, and everyone else, for helping to make those visits fun and productive.

I am also very appreciative of the various other friends and collaborators I have had the privilege to interact with at the University of Illinois and elsewhere. Special thanks go to Grace Chen for her constant encouragement, support, and inspiration.

I would like to acknowledge generous financial support of my graduate studies from the National Science Foundation, the Beckman Institute, the Department of Electrical and Computer

Engineering, and the University of Illinois.

Finally, I am grateful to my parents for their unfailing love, support, and patience, and for all the sacrifices they made so that I could have more.

Table of Contents

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
ADM	Alternating Directions Minimization
AIC	Akaike Information Criterion
ALS	Alternating Least-Squares
BIC	Bayesian Information Criterion
CG	Conjugate Gradient
CS	Compressed Sensing
DCE	Dynamic Contrast Enhanced
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DTI	Diffusion Tensor Imaging
DW	Diffusion Weighted
emf	Electromotive Force

EPI Echo-Planar Imaging

EW Equivalent Width

EY Eckart-Young

FARM Fast Alternating-subsets descent algorithm for Rank-constrained Matrix recovery

FFT Fast Fourier Transform

FLASH Fast Low-Angle Shot

FOV Field of View

FPE Final Prediction Error

FWHM Full-Width at Half-Maximum

FWTM Full-Width at Tenth-Maximum

IRPF Incremented Rank PowerFactorization

MAP Maximum A Posteriori

MDL Minimum Description Length

MIL Matrix Inversion Lemma

ML Maximum-Likelihood

MMSE Minimum Mean-Squared Error

MNLS Minimum-Norm Least-Squares

MR Magnetic Resonance

MRF Markov Random Field

MRI Magnetic Resonance Imaging

NIPALS Nonlinear Iterative Partial Least Squares

NMR Nuclear Magnetic Resonance

NN Nuclear Norm

NNM Nuclear Norm Minimization

NSP Nullspace Property

PML Penalized Maximum-Likelihood

PS Partial Separability

RC Rayleigh Criterion

RF Radiofrequency

RIP Restricted Isometry Property

rRIP Rank-Restricted Isometry Property

SC Sparrow Criterion

SD Standard Deviation Criterion

SNR Signal-to-Noise Ratio

SRF Spatial Response Function

SVD Singular Value Decomposition

TPSF Transform Point Spread Function

TV Total Variation

Chapter 1

Introduction

1.1 Problem Statement

Data acquisition in conventional MRI is typically modeled as sampling in the spatial Fourier domain (also called *k-space*):

$$d_m = \int \rho(\mathbf{x})\, e^{-i2\pi \mathbf{k}_m \cdot \mathbf{x}} d\mathbf{x} + \eta_m,\ m = 1, 2, \ldots, M. \tag{1.1}$$

In this expression, M denotes the number of measured *k*-space samples, $\{\mathbf{k}_m\}_{m=1}^{M}$ is the set of *k*-space sampling locations, $\{d_m\}_{m=1}^{M}$ is the set of measured data, $\rho(\mathbf{x})$ is the image function of interest, and $\{\eta_m\}_{m=1}^{M}$ is the set of additive measurement noise perturbations. In practice, the choice of M represents a balance between several factors:

- **Experiment duration.** MRI data acquisition is a time-consuming process, with the total experiment duration typically proportional to the amount of acquired data samples. As a result, the use of small M can improve the speed of imaging experiments.

- **Image resolution.** MR image resolution is a function of the Fourier-domain sampling pattern, and traditional high-resolution reconstructions generally require extended *k*-space sampling and large M.

- **Image signal-to-noise ratio.** The standard method of reducing the impact of additive noise in reconstructed MR images is to perform data averaging, which also increases M.

This research addresses the problem of reconstructing $\rho(\mathbf{x})$ from $\{d_m\}_{m=1}^{M}$ when the number of acquired data samples M is too small to yield acceptable reconstructions using conventional Fourier reconstruction algorithms. The use of small M can dramatically reduce experiment duration, and can lead to improved temporal resolution in dynamic imaging applications. In contrast to standard methods, we use prior information about the imaging experiment to develop new imaging models that are capable of dealing with limited and/or noisy data. The characteristics of these new models influence the choice of MR data acquisition and image reconstruction strategies. The methods described in this work focus on two specific small-M imaging contexts:

1. We consider constrained denoising approaches to reduce noise limitations in high-resolution imaging. In this case, we propose and investigate the use of prior information derived from coregistered anatomical reference images to improve image signal-to-noise ratio (SNR), while simultaneously preserving high-resolution image features.

2. We consider approaches for reconstructing MR images from sparsely-sampled data with high SNR. In particular, we propose and investigate new approaches that leverage the prior information that MR images are often highly structured (i.e., static MR images are generally sparse in appropriately-chosen transform domains, and spatiotemporal MR images often satisfy special linear-dependence relationships along the temporal dimension) to enable high-quality reconstructions from significantly undersampled data.

1.2 Motivation

The theory and methods underlying MR studies have been developing for decades. Important early landmarks leading up to the development of MRI include development in the 1930s of the first magnetic resonance method for measuring the magnetic characteristics of nuclei by Isador I. Rabi (who received the 1944 Nobel Prize in Physics), the first measurements of MR signal from bulk matter in the 1940s by Felix Bloch and Edward M. Purcell (who shared the 1952 Nobel Prize

in Physics), and the development of pulsed Fourier transform and multidimensional MR spectroscopy methods by Richard R. Ernst (who received the 1991 Nobel Prize in Chemistry). MRI was first demonstrated in the early 1970s by Paul C. Lauterbur [386], who shared the 2003 Nobel Prize in Physiology or Medicine with Sir Peter Mansfield. Since that time, MRI has evolved into a powerful non-invasive imaging tool that can probe the structure, biochemistry, and function of living biological tissues. MRI has a number of advantages relative to other existing medical imaging modalities [173, 406, 442, 646]. For example, unlike modalities such as X-ray computed tomography, positron emission tomography, and single photon emission computed tomography, MRI does not require the use of ionizing radiation. In addition, unlike ultrasound and many optical techniques, MRI is less constrained by penetration depth limitations at the magnetic field strengths that are commonly used in current clinical practice. Furthermore, the MR signal is sensitive to a wide range of different physical phenomena, allowing MR experiments to study biological tissues from a wide range of different perspectives through the careful manipulation of various intrinsic and extrinsic contrast mechanisms. As a result, MRI can enable the early detection of pathology before the onset of chronic disease, can facilitate new approaches to personalized medical treatment based on imaging-guided interventions and/or quantitative assays of the unique biological parameters of each patient, and can be used to explore fundamental questions about the nature of biological systems. However, despite this amazing potential and decades of revolutionary progress, modern MR techniques are still quite far from reaching the limits of what can be extracted from the MR signal.

The main limitations of MRI are its relatively low sensitivity and slow data acquisition speed. In practice, these limitations often mean that it is necessary to choose between low-resolution data, noisy reconstructed images, prolonged experiments, limited MR contrast encoding, or some combination thereof. Long experiment duration is problematic for several reasons: (1) extended experiments are uncomfortable for live imaging subjects, since it is often important to remain still during data acquisition to reduce motion artifacts in the reconstructed images; (2) long data acquisition times can limit the spatiotemporal resolution in studies of dynamic processes; (3) imaging

time on MR scanners is expensive, and long experiments limit subject throughput; (4) in clinical situations, fast availability of imaging findings can be critical to the treatment planning and prognosis of emergency room patients. Thus, while the MR signal can be manipulated to encode a wealth of multidimensional biophysical and biochemical information, practical considerations have prevented modern MR experiments from realizing their full potential. As a result, the ability to reconstruct high-quality images from noisy and/or sparsely-sampled data would significantly enhance current and future MRI applications.

1.3 Main Results

- We have proposed and evaluated a new constrained approach to reconstructing MR images when correlated reference images are available. In contrast to conventional constrained approaches, which typically use anatomical prior information to achieve super-resolution reconstruction from low-resolution data, the proposed method is most effective at improving image SNR while preserving anatomical structure within noisy high-resolution data. We show that the proposed method is easy to characterize, and an analysis of the resolution and noise characteristics of the proposed method demonstrates that substantial improvements in SNR can be achieved with only a moderate decrease in spatial resolution. Formulations are presented for the reconstruction of single images in the presence of high-quality, high-resolution reference information, and for the joint reconstruction of a coregistered sequence of noisy correlated images. The power of the technique is demonstrated using simulated and experimental data. Related publications include Refs. [24, 25, 232, 267, 269, 270, 274, 276–280, 283–286, 378, 593].

- We have proposed a new sparsely-sampled data acquisition scheme for reconstructing images that are sparse in a known transform domain. In contrast to conventional MRI, which makes use of Fourier encoding, we make use of a specialized non-Fourier encoding scheme that is better aligned with existing theoretical literature on signal reconstruction using spar-

sity constraints (i.e., *compressed sensing* [117, 180]). Simulation and experimental results demonstrate that the method can significantly reduce data sampling requirements relative to standard Fourier-based compressed sensing reconstruction techniques in high-SNR scenarios. Related publications include Refs. [273, 275].

- We have proposed a new matrix-recovery approach to reconstructing spatiotemporal images from sparsely-sampled data. In particular, we use the partial separability (PS) model to model the strong temporal correlations found in many spatiotemporal imaging applications [399]. The use of the PS model allows us to reformulate image reconstruction as the reconstruction of a low-rank matrix. We describe some theoretical characteristics of this new problem, propose efficient new algorithms, and demonstrate with simulated and experimental data that the combined PS/matrix-recovery approach provides a new, powerful way to reconstruct spatiotemporal images from limited data. Related publications include Refs. [268, 281, 282].

1.4 Organization of the Dissertation

This dissertation is organized as follows:

Chapter 2 presents background material that will be helpful for understanding the subsequent chapters of the report. It contains a high-level overview of MR physics, incorporating descriptions of basic signal generation and detection, spatial encoding techniques, and noise. The chapter also includes a brief review of basic image reconstruction techniques for data sampled in the Fourier domain, and a brief review of matrix rank.

Chapter 3 describes a new quasi-Bayesian image model for modeling image sequences with correlated edge structures. The formulation is described, and a convergent algorithm to solve the resulting optimization problem is presented. In addition, the resolution and SNR properties of the resulting reconstruction scheme are characterized, and the optimal trade-off between data acquisition resolution and SNR in is discussed. Finally, a number of simulation and experimental

reconstruction results are presented to demonstrate the effectiveness of the proposed scheme at improving SNR while preserving high-resolution image features.

Chapter 4 presents a new compressed sensing scheme for reconstructing images from sparsely-sampled high-SNR data, using the prior knowledge that images are sparse in a known transform domain. The chapter describes the rationale for and design of a specialized non-Fourier encoding scheme for this context. Simulation and experimental results are shown to demonstrate that this data acquisition scheme, when coupled with an appropriate sparsity-promoting reconstruction scheme, can significantly reduce data-sampling requirements in certain contexts relative to Fourier-based schemes.

Chapter 5 presents a new reconstruction scheme for sparsely-sampled spatiotemporal imaging data. The chapter introduces a novel and flexible formulation of the spatiotemporal imaging inverse problem in terms of the recovery of a low-rank matrix from sparsely-sampled data, and an efficient algorithm is described to solve the resulting optimization problem. Results are shown that illustrate the effectiveness of rank constraints for achieving high-resolution reconstructions from highly-undersampled data.

Finally, Chapter 6 provides conclusions.

Chapter 2

Background

This chapter provides fundamental concepts needed for later chapters. A brief discussion of the physics of the MRI experiment is presented in Sec. 2.1, while Sec. 2.2 provides a short overview of basic Fourier MRI reconstruction approaches. Section 2.3 presents a brief review of matrix rank.

2.1 The Magnetic Resonance Imaging Experiment

Due to the complexity of MR physics, this section will present an abbreviated high-level view of MRI signal generation and detection. In particular, though MRI has its foundation in quantum mechanics, we will adopt a semi-classical treatment in this work, which is generally quite accurate for describing the ensemble behavior of the large collections of nuclei that are present in macroscopic objects [290]. Deeper perspectives on MR physics can be found in Refs. [2, 46, 263, 394, 406, 577] and similar texts, though it should be noted that an end-to-end quantum mechanical description of MRI does not yet exist [309].

This section will review MRI signal generation and detection in Section 2.1.1, spatial encoding techniques in Section 2.1.2, and MR noise characteristics in Section 2.1.3.

2.1.1 Signal Generation and Detection

MRI is possible because of the nuclear magnetic resonance (NMR) phenomenon, which involves the interaction of atomic nuclei with magnetic fields. The NMR phenomenon is itself dependent on the quantum mechanical property known as *spin*, an intrinsic form of angular momentum possessed by elementary particles. Many biologically-important atomic nuclei possess non-zero spin

(e.g., ^{1}H, ^{31}P, ^{23}Na, etc.), which helps to explain why MRI has become such a powerful tool for biological research and medical practice.

Atomic nuclei with non-zero spin quantum numbers generate microscopic magnetic fields about themselves, each of which can characterized by a magnetic moment vector $\boldsymbol{\mu} = \mu_x\hat{\mathbf{i}} + \mu_y\hat{\mathbf{j}} + \mu_z\hat{\mathbf{k}}$, where $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ are the unit vectors for the standard Cartesian coordinate system. The total magnetic field generated by a collection of nuclear spins is described through the *bulk magnetization vector* $\mathbf{M} = M_x\hat{\mathbf{i}} + M_y\hat{\mathbf{j}} + M_z\hat{\mathbf{k}}$, with

$$\mathbf{M} = \sum_i \boldsymbol{\mu}_i, \tag{2.1}$$

where the ith nucleus has magnetic moment $\boldsymbol{\mu}_i$.

Signal detection in practical NMR and MRI experiments relies on a non-zero bulk magnetization. However, at thermal equilibrium and in the absence of a strong external magnetic field, the $\{\boldsymbol{\mu}_i\}$ are randomly and incoherently oriented due to random thermal perturbations, such that $\mathbf{M} = \mathbf{0}$. On the other hand, in the presence of a strong uniform main magnetic field $\mathbf{B}_0$, the different orientations of the $\{\boldsymbol{\mu}_i\}$ with respect to $\mathbf{B}_0$ are associated with different energy levels. As a result, individual magnetic moments will have preferred orientations, leading to a nonzero bulk magnetization along the direction of the applied field. Thus, a magnet that can provide a strong homogeneous magnetic field is an important component of most MR experiments. In keeping with the standard NMR/MRI literature, we will assume that $\mathbf{B}_0$ is oriented along the z direction, i.e., that $\mathbf{B}_0 = B_0\hat{\mathbf{k}}$, and will denote the thermal-equilibrium value of the bulk magnetization vector as $\mathbf{M}_0 = M_z^0\hat{\mathbf{k}}$.

An important aspect of the bulk magnetization is that the energy differences between the different orientations of the $\{\boldsymbol{\mu}_i\}$ are generally very small, meaning that the orientations of the $\{\boldsymbol{\mu}_i\}$ are still largely incoherent at thermal equilibrium and that $M_z^0 \ll \sum_i \|\boldsymbol{\mu}_i\|_{\ell_2}$.[1] In particular, the

[1] We use the notation $\|\cdot\|_{\ell_2}$ to denote the ℓ_2 norm, which measures the Euclidean length of a vector. An explicit mathematical definition of this norm is presented later in this chapter.

value of $\sum_i \|\boldsymbol{\mu}_i\|_{\ell_2}$ for a nucleus with spin quantum number I is

$$\sum_i \|\boldsymbol{\mu}_i\|_{\ell_2} = \gamma \hbar N_s \sqrt{I(I+1)}, \tag{2.2}$$

while the value of M_z^0 is

$$M_z^0 = \frac{\gamma^2 \hbar^2 B_0 N_s I(I+1)}{3 k_B T_s}, \tag{2.3}$$

where γ is a constant known as the *gyromagnetic ratio* that is unique to each species of nucleus, $\hbar$ is Planck's constant (6.6×10^{-34} J-s) divided by 2π, N_s is the total number of spins in the system, k_B is Boltzmann's constant (1.38×10^{-23} J/K), and T_s is the absolute temperature (K) of the system. For ^{1}H imaging ($I = 1/2$ and $\gamma/2\pi = 42.58$ MHz/T) at room temperature ($T_s = 300$ K) and the relatively large magnetic field strength of $B_0 = 9.4$ T, the ratio of M_z^0 to $\sum_i \|\boldsymbol{\mu}_i\|_{\ell_2}$ is approximately 2×10^{-5}. While the energy differences between the different orientations increase with increasing B_0 and decreasing temperature, modern MRI is still widely regarded as having relatively low sensitivity.

In the presence of an external magnetic field, the bulk magnetization vector is known to behave according to the *Bloch equations*, which can be written as

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B} - \frac{\left(M_x \hat{\mathbf{i}} + M_y \hat{\mathbf{j}}\right)}{T_2} - \frac{\left(M_z - M_z^0\right) \hat{\mathbf{k}}}{T_1}, \tag{2.4}$$

where $\mathbf{B}$ is the total magnetic field and T_1 and T_2 are relaxation constants. In the presence of the static field $\mathbf{B} = \mathbf{B}_0$, the magnetization will remain at its thermal equilibrium value. However, by applying a carefully tailored time-varying additional magnetic field $\mathbf{B}_1(t)$ (typically oriented in the transverse $x - y$ plane) such that

$$\mathbf{B}(t) = \mathbf{B}_0 + \mathbf{B}_1(t), \tag{2.5}$$

the movement of the bulk magnetization vector can be controlled so that $\mathbf{M}$ is forced to tip into

the plane transverse to the main magnetic field. This process of forcing $\mathbf{M}$ into the transverse plane is known as *excitation*. The $\mathbf{B}_1$ field is often referred to as the *radiofrequency* (RF) *pulse*, since it typically takes the form of a radiofrequency modulated signal. If the significant transverse components of $\mathbf{B}$ are removed after excitation, then $\mathbf{M}$ will precess clockwise about the positive z axis as described through Eq. (2.4). In particular, a magnetization vector $\mathbf{M}(t)$ starting at $\mathbf{M}(0) = M_x(0)\hat{\mathbf{i}} + M_y(0)\hat{\mathbf{j}} + M_z(0)\hat{\mathbf{k}}$ at time $t = 0$ will evolve in the presence of $\mathbf{B} = B_z\hat{\mathbf{k}}$ according to

$$\begin{aligned} M_{xy}(t) &= M_{xy}(0)e^{-i\gamma B_z t}e^{-t/T_2} \\ M_z(t) &= M_z(0)e^{-t/T_1} + M_z^0\left(1 - e^{-t/T_1}\right), \end{aligned} \tag{2.6}$$

where we have introduced the complex phasor notation $M_{xy}(t) = M_x(t) + iM_y(t)$, and it is assumed that $B_z = B_0 + \Delta B$, with $\Delta B \ll B_0$ such that the thermal equilibrium magnetization M_z^0 is not significantly perturbed.

As a result of Faraday's law of induction,[2] the changes in the magnetic flux caused by the precessing magnetization vector $\mathbf{M}$ will generate an electromotive force (emf) through a coil placed in proximity to the sample. In particular, the emf $\mathcal{E}(t)$ induced in the coil is given by

$$\begin{aligned} \mathcal{E}(t) &= -\frac{d}{dt}\int \mathbf{B}^r(\mathbf{x}) \cdot \mathbf{M}(\mathbf{x}, t)d\mathbf{x} \\ &= -\frac{d}{dt}\int \left\{\text{real}\left[B_{xy}^{r*}(\mathbf{x})M_{xy}(\mathbf{x}, t)\right] + B_z^r(\mathbf{x})M_z(\mathbf{x}, t)\right\} d\mathbf{x}, \end{aligned} \tag{2.7}$$

where $\mathbf{M}(\mathbf{x}, t)$ is the local bulk magnetization per unit volume at position $\mathbf{x} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$ and time t, $\mathbf{B}^r(\mathbf{x}) = B_x^r(\mathbf{x})\hat{\mathbf{i}} + B_y^r(\mathbf{x})\hat{\mathbf{j}} + B_z^r(\mathbf{x})\hat{\mathbf{k}}$ is the spatially-varying receive field of the detection coil,[3] $B_{xy}^r(\mathbf{x}) = B_x(\mathbf{x}) + iB_y(\mathbf{x})$ is a complex phasor, and where * is used to denote complex

[2] In this description, we are largely following the somewhat simplified description by Hoult [312], which relies on Faraday's law of induction and requires a loop receiver coil. However, it is possible to use Lorenz reciprocity to derive expressions that work for arbitrary coil geometries – see, e.g., Ref. [564] for more detail.

[3] The receive field is related through the principle of reciprocity to the magnetic field that would be generated by a unit of current at the Larmour frequency passing through the coil.

conjugation. Substituting Eq. (2.6) into this expression yields

$$\begin{aligned}\mathcal{E}(t) = \int \Bigg\{ &-\text{real}\left[\left(i\gamma B_z + \frac{1}{T_2}\right) B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x},t)\right] \\ &+ \frac{M_z(\mathbf{x},0)}{T_1} B_z^r(\mathbf{x}) e^{-t/T_1} - \frac{M_z^0(\mathbf{x})}{T_1} B_z^r(\mathbf{x}) e^{-t/T_1} \Bigg\} d\mathbf{x}.\end{aligned} \tag{2.8}$$

Since γB_z is generally orders of magnitude larger than $1/T_2$ and $1/T_1$, many terms appearing in the integral in Eq. (2.8) can be neglected, and the signal $s(t)$ measured by a receiver coil from magnetization evolving according to Eq. (2.6) will be approximately given by

$$\begin{aligned}s(t) &\approx C \int \gamma B_z \text{imag}\left[B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x},t)\right] d\mathbf{x} \\ &= C \int \frac{\gamma B_z}{2i} \left[B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x},0) e^{-i\gamma B_z t} e^{-t/T_2} - B_{xy}^{r}(\mathbf{x}) M_{xy}^{*}(\mathbf{x},0) e^{i\gamma B_z t} e^{-t/T_2}\right] d\mathbf{x},\end{aligned} \tag{2.9}$$

where C is a constant of proportionality depending on the measurement system. In practice, the signal $s(t)$ is generally demodulated by multiplication with a complex sinusoid at frequency γB_0, leading to

$$\begin{aligned}&2is(t)e^{i\gamma B_0 t} = \\ &\quad C \int \gamma B_z \left[B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x},0) e^{-i\gamma \Delta B t} e^{-t/T_2} - B_{xy}^{r}(\mathbf{x}) M_{xy}^{*}(\mathbf{x},0) e^{i\gamma(2B_0+\Delta B)t} e^{-t/T_2}\right] d\mathbf{x},\end{aligned} \tag{2.10}$$

where ω is the demodulation frequency. This demodulated signal can then be low-pass filtered to eliminate the component oscillating at frequencies near $\gamma\left(2B_0 + \Delta B\right)$, resulting in a signal $\bar{s}(t)$ given by

$$\begin{aligned}\bar{s}(t) &= C \int \gamma B_z B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x},0) e^{-i\gamma \Delta B t} e^{-t/T_2} d\mathbf{x} \\ &\approx \gamma B_0 C \int B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x},0) e^{-i\gamma \Delta B t} e^{-t/T_2} d\mathbf{x}.\end{aligned} \tag{2.11}$$

Finally, the signal $\bar{s}(t)$ is sampled for further digital processing on a computer.

2.1.2 Spatial Encoding Techniques

There are several mechanisms available for spatial encoding in MRI (see Ref. [356] for a detailed review in the context of spectroscopy), though the most common forms of this involve designing ΔB in Eq. (2.11) to have spatial and temporal variations. Examples of this include the gradient-encoding techniques first described by Lauterbur [386] in which the spatial position is encoded into the resonance frequency $\gamma \Delta B$ through the use of a spatially-varying ΔB, and sensitive point/field focusing methods [67, 154, 249, 306, 565] in which the magnetic fields are generated in such a way that only a single spatial position contributes significantly to the measured signal.

Spatial encoding through the use of linear gradient fields during data acquisition, as in Lauterbur's original experiment [386], has become the most widely-used spatial encoding mechanism in MRI, and has a very convenient Fourier-domain interpretation [376, 614]. In this approach, the magnetic field ΔB is constructed such that

$$\Delta B(\mathbf{x}) = G_x x + G_y y + G_z z. \tag{2.12}$$

In this case, defining $\mathbf{k}(t)$ as

$$\mathbf{k}(t) = \frac{\gamma}{2\pi} \begin{bmatrix} G_x t \\ G_y t \\ G_z t \end{bmatrix}, \tag{2.13}$$

Eq. (2.11) can be rewritten as

$$\bar{s}(t) = \gamma B_0 C \int B_{xy}^{r*}(\mathbf{x}) M_{xy}(\mathbf{x}, 0) e^{-i2\pi \mathbf{k}(t)\cdot\mathbf{x}} e^{-t/T_2} d\mathbf{x}. \tag{2.14}$$

Under the assumption that the effects of relaxation are negligible during data acquisition, this expression can also be written as

$$\int B_{xy}^{r*}(\mathbf{x}) \rho(\mathbf{x}) e^{-i2\pi \mathbf{k}(t)\cdot\mathbf{x}} d\mathbf{x}, \tag{2.15}$$

where $\rho(\mathbf{x})$ is a scaled, relaxation-weighted version[4] of the original $M_{xy}(\mathbf{x}, 0)$, which itself is a function of the nuclear spin distribution. If we assume that the receive field is spatially homogeneous such that $B_{xy}^{r*}(\mathbf{x}) = 1$, then Eq. (2.15) matches the Fourier-transform relationship between the observed signal and the image $\rho(\mathbf{x})$ given in Eq. (1.1). This type of Fourier relationship also holds in the more general case where G_x, G_y, and G_z (the linear gradients) have temporal variations. In this case, Eq. (2.15) is still valid, with $\mathbf{k}(t)$ now defined as

$$\mathbf{k}(t) = \frac{\gamma}{2\pi} \begin{bmatrix} \int_0^t G_x(\tau) d\tau \\ \int_0^t G_y(\tau) d\tau \\ \int_0^t G_z(\tau) d\tau \end{bmatrix}. \tag{2.16}$$

In practice the maximum gradient strength and slew rate are limited, due to both hardware and regulatory constraints. In addition, it is often desirable from the perspective of SNR to use small gradients, since large gradients increase the frequency bandwidth of the desired signal, and increasing the bandwidth of the detection system leads to increased measurement noise as described in Sec. 2.1.3. Thus, while pulse sequences exist that try to acquire data as fast as possible (including echo-planar imaging (EPI) [433], spiral imaging [7], fast low-angle shot (FLASH) imaging [265], burst imaging [300], and several others [46]), *k*-space trajectories are still traversed relatively slowly. As a result, the need to have the data acquisition window be short relative to relaxation and other physical effects often means that multiple iterations of the excitation and encoding procedure are necessary to acquire sufficient Fourier data for high-resolution reconstruction. The relatively time-consuming nature of this procedure is one of the primary limitations of MRI. Due to the widespread use of Fourier acquisition in MRI, most of this dissertation will focus on image reconstruction from Fourier-encoded data. However, the methods we will discuss can easily be extended to model non-Fourier aspects of data acquisition, such as undesired spatial inhomogeneity of the B_0 field [212, 595] and spatial inhomogeneity of the receive field of the de-

[4]The relaxation parameters T_1 and T_2 tend to vary spatially in real experiments, and this provides one of the many useful sources of MR image contrast.

tection coil [519]. The remainder of this subsection will briefly discuss some of the other spatial encoding methods that have been proposed for MR.

The Fourier encoding scheme described above created a structured magnetic field inhomogeneity $\Delta B(\mathbf{x})$ in which the magnetic field strength varied linearly as a function of spatial position. Techniques have also been proposed that make use of nonlinear gradient fields for spatial encoding during data acquisition. Examples include the use of nonlinear gradients for phase-scrambling applications [445, 641], Fresnel and O-space MRI using quadratic gradient fields [331, 332, 590, 665], and PatLoc imaging using general curvilinear fields [227, 301, 561].

Another approach to spatial encoding is to introduce spatial variations in the magnetic field during the RF pulse used for excitation, such that the response of $\mathbf{M}(\mathbf{x}, t)$ to the RF pulse has spatial dependence. The most common form of this kind of RF encoding is slice selection, in which only a thin two-dimensional (2D) slice of a three-dimensional (3D) object is excited by the RF pulse [310, 387, 434], while the remaining 2D localization is still achieved by conventional gradient-based Fourier encoding. However, a wide variety of other spatially-selective excitation or saturation based methods have been proposed and provide significant flexibility to the design of MR experiments [27, 65, 68, 69, 83, 151, 221, 234, 245, 297, 311, 342, 351, 443, 444, 455, 486, 496, 507, 547, 640, 678]. Classes of spatial encoding techniques also exist in which data is acquired simultaneously or interleaved with RF irradiation. These techniques include continuous wave MRI [199], frequency-swept MRI [329, 644], and stochastic NMR imaging [62–64, 338, 473, 543].

A fundamentally different approach to obtaining spatially-localized information from an object is to use receiver coils with spatially-localized detection sensitivities $\mathbf{B}^r(\mathbf{x})$ (i.e., *local coils* or *surface coils*) [5, 363]. Parallel imaging methods represent an evolution of this approach, simultaneously acquiring data from multiple coils in parallel. In this case, spatial information is encoded into the measured data by the spatially-varying sensitivity profiles of each coil, and if the $\mathbf{B}^r(\mathbf{x})$ parameters for each coil can be calibrated accurately and the coils are sufficiently linearly independent, then images can be obtained through the application of suitable signal processing algorithms [124, 125, 256, 257, 298, 326, 327, 337, 358, 379, 380, 408, 411, 517, 519, 520, 525, 542, 580, 669].

In practice, applications that use localized/parallel coils to encode spatial information frequently combine sensitivity-based spatial encoding with the other MR spatial encoding schemes described above.

2.1.3 Noise

As a result of the limited sensitivity of MR (cf. Sec. 2.1.1), noise is a major limitation in practical MRI experiments. Noise in MRI comes from multiple sources, including thermal fluctuations, physiological effects, and signal perturbations due to system instability (see discussion in Refs. [98,311,313,407,425,465]). The main type of noise that we will consider for this dissertation is thermal (Johnson-Nyquist) noise [343, 480]. The thermal noise in MRI predominantly comes from the random thermal motion of charge carriers in the receiver electronics (e.g., electrons) and within the sensitive region of the receiver coil (e.g., charged ions like sodium and dipolar molecules like water in the sample) [518]. For most biological studies, the charged ions in the sample are the dominant source of noise fluctuations.

Moving charged particles interact electromagnetically with the signal detector, resulting in fluctuations of the measured signal voltage. Due to the nature of this thermal noise, the statistics of the noise samples (i.e., $\{\eta_m\}_{m=1}^{M}$ from Eq. (1.1)) are expected to be modeled well by an additive white complex Gaussian distribution (with independent and identically distributed real and imaginary components). It has been shown experimentally that this distribution accurately describes the distribution of experimental data, as long as the temporal filters used in the demodulation step (cf. Eqs. (2.10) and (2.11)) and in the analog-to-digital conversion process maintain a uniform frequency response over the frequency range of interest [451]. Several factors influence the standard deviation of the measurement noise, including the bandwidth of the temporal filters (which is usually adjusted based on the spatial extent of the sample and the maximum magnitude of the gradient fields G_x, G_y, and G_z used during data acquisition), the temperature of the sample and electronics, and the number of moving charged particles in the sensitive region of the receive coil.

To avoid confusion, we should mention that many descriptions of noise in the MRI litera-

ture [9, 10, 16, 122, 171, 260, 299, 365, 366, 441, 447, 453, 574] describe the effects of noise using the non-central Chi and Rician distributions [539], rather than the Gaussian distribution. We emphasize that MR data in *k*-space is Gaussian, meaning that any linear image reconstruction will also have Gaussian-distributed voxels. The Rician distribution arises from taking the magnitude of Gaussian-distributed images, while the non-central Chi distribution arises in parallel imaging when magnitude images of the same subject from different receiver coils are combined using a sum-of-squares procedure. In both cases, these operations result in a loss of information. While the magnitude of MR images is often taken for display purposes, this specific processing step is not a necessary part of image reconstruction or a necessary step for the display of images (see, e.g., competing phase-correction methods [47, 80, 81, 241, 412, 450, 475, 515]). As a result, while many approaches exist for denoising MR images with non-Gaussian distributions [9, 10, 29, 36, 66, 216, 295, 374, 415, 431, 441, 476, 510, 575, 605], the discussion presented in this dissertation will be limited to the Gaussian case, which is much easier to analyze and for which more significant gains can be obtained [660].

In addition to thermal noise, a noise source that appears in the context of biological imaging is physiological noise [98]. Physiological noise includes effects resulting from subject motion such as bulk subject movement, cardiac motion and pulsation, and respiration. These effects can be significant and important in certain applications (e.g., fMRI), but are difficult to analyze because we still do not have a good quantitative understanding of physiological noise [98]. In this dissertation, we will either neglect the effects of physiological noise or assume that it also leads to additive white Gaussian measurement noise.

Finally, other sources of noise include quantization in the analog-to-digital converters used for sampling the MR signal [39], spin noise in the bulk magnetization vector $\mathbf{M}$ [393], external sources of RF signal (e.g., radio stations), and instabilities in the main magnetic field [465].

2.2 Fourier Image Reconstruction

The previous section derived the physical basis of the Fourier imaging equation given in Eq. (1.1), which we repeat below:

$$d_m = \int \rho(\mathbf{x})\, e^{-i2\pi \mathbf{k}_m \cdot \mathbf{x}} d\mathbf{x} + \eta_m, \;\; m = 1, 2, \ldots, M. \tag{2.17}$$

In this section, we review some of the standard techniques that are available for reconstructing $\rho(\mathbf{x})$ from Fourier data. While there are many constrained reconstruction approaches that leverage strong prior information about the images being reconstructed (e.g., [400, 454, 475]), this section will focus primarily on general methods for Fourier inversion. Finitely-sampled Fourier inversion is well-known to be an *ill-posed* inverse problem, where an inverse problem is said to be well-posed (in the sense of Hadamard) if it has the following three characteristics (see Refs. [48, 289, 601, 625] for a precise mathematical definition):

1. At least one solution exists.

2. The solution is unique.

3. The solution is stable with respect to small perturbations in the data.

As we will see, the inverse problem corresponding to Eq. (2.17) generally fails to have a unique solution, and additional constraints are necessary to make the problem well-posed.

To begin, we will make some mathematical assumptions to simplify the discussion. We will assume that $\rho(\mathbf{x})$ is a complex function in D-dimensional space, i.e., that $\rho : \mathbb{R}^D \to \mathbb{C}$, where D is typically 2 or 3, corresponding to 2D and 3D imaging problems, respectively. We will additionally assume that $\rho(\mathbf{x})$ is supported on a set $\Omega \subseteq \mathbb{R}^D$ (i.e., $\rho(\mathbf{x}) = 0$ for $\mathbf{x} \notin \Omega$) of positive measure, and that $\rho(\mathbf{x})$ has finite $\mathcal{L}_1$ and $\mathcal{L}_2$ norms, where the $\mathcal{L}_p$ norms for functions on Ω are defined for $p \geq 1$ as

$$\|\rho(\mathbf{x})\|_{\mathcal{L}_p} = \left(\int_\Omega |\rho(\mathbf{x})|^p \, d\mathbf{x} \right)^{1/p}. \tag{2.18}$$

Under these conditions, it can be shown [353, 385] that the Fourier transform $s(\mathbf{k})$ of $\rho(\mathbf{x})$ is well-defined, with

$$s(\mathbf{k}) = \int_{\Omega} \rho(\mathbf{x}) e^{-i2\pi\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}, \tag{2.19}$$

and that an inverse Fourier transform exists that maps $s(\mathbf{k})$ back to $\rho(\mathbf{x})$, with

$$\rho(\mathbf{x}) = \int_{\mathbb{R}^D} s(\mathbf{k}) e^{i2\pi\mathbf{k}\cdot\mathbf{x}} d\mathbf{k}. \tag{2.20}$$

While the relationship in Eq. (2.20) provides one way of recovering an image from Fourier data, direct use of this relationship is impractical because it requires knowledge of $s(\mathbf{k})$ for every $\mathbf{k} \in \mathbb{R}^D$, an uncountably infinite set of points. However, sampling density requirements can be considerably reduced if it is additionally assumed that the support set Ω is closed and bounded. In this case, it can be shown that $\rho(\mathbf{x})$ can be recovered from a countably infinite set of points [12, 44, 56, 123, 189, 353, 382, 385, 437]. Theoretical sampling density requirements for exact reconstruction have been established through the Whittaker-Nyquist-Kotel'nikov-Shannon sampling theorem in the context of periodic lattice sampling [12, 44, 56, 189, 353, 385, 437], and by the Beurling-Landau theory for general non-uniform sampling [12, 44, 45, 123, 382]. An example result is that if

$$R_i \geq \sup_{\mathbf{x}\in\Omega} (2|x_i|), \text{ for } i = 1, 2, \ldots, D, \tag{2.21}$$

then

$$\rho(\mathbf{x}) = \frac{1}{|\mathbf{R}|} \mathbf{1}_{\{\mathbf{x}\in\Omega\}} \sum_{\mathbf{n}\in\mathbb{Z}^D} s\left(\mathbf{R}^{-1}\mathbf{n}\right) e^{i2\pi\mathbf{R}^{-1}\mathbf{n}\cdot\mathbf{x}}, \tag{2.22}$$

where $\mathbb{Z}^D$ is the space of D-dimensional integer vectors, $\mathbf{1}_{\{\mathbf{x}\in\Omega\}}$ is the indicator function for Ω, and $\mathbf{R}$ is a diagonal matrix with ith diagonal entry equal to R_i. If Ω is a hyper-rectangle centered at the origin with sides aligned with the coordinate axes (a common assumption in MRI experiments), then choosing the smallest possible R_i values in Eq. (2.21) leads to *Nyquist rate* sampling in Eq. (2.22).[5] Cartesian *k*-space trajectories with points spaced at the Nyquist rate are very common

[5]The Nyquist rate has a more general definition for an arbitrary support set Ω. See, e.g., Ref. [437].

in practical MRI experiments.

Equation (2.22) brings us one step closer to addressing Eq. (2.17), but still requires knowledge of $s(\mathbf{k})$ at infinitely many points. In general, the need for infinite sampling to achieve perfect reconstruction is unavoidable without making additional assumptions. There are three common classes of methods that are used for reconstructing noisy, finitely-sampled Fourier data as in Eq. (2.17) [207], which we describe in the following three subsections.

2.2.1 Conjugate Phase Reconstruction

The first class of methods ignores the fact that data is finitely sampled, and directly makes approximations of analytic reconstruction formulas such as Eqs. (2.20) or (2.22). In this case, the standard approach is to estimate the image using a *conjugate phase* [424] reconstruction

$$\hat{\rho}(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x}\in\Omega\}} \sum_{m=1}^{M} w_m d_m e^{i2\pi\mathbf{k}_m\cdot\mathbf{x}}, \tag{2.23}$$

where $\{w_m\}$ are weighting coefficients. Equation (2.23) can be viewed as a Riemann-sum approximation of Eq. (2.20), with the $\{w_m\}$ adjusted based on the local density of Fourier samples in the vicinity of each measurement. The reconstruction can also be viewed as a weighted, truncated version of the infinite summation in Eq. (2.22) if the original Fourier samples lie on a Cartesian lattice. Reconstruction using Eq. (2.23) is generally very computationally efficient, since the fast Fourier transform (FFT) [485] can be used for Cartesian-sampled data, and various approximations for Eq. (2.23) exist that leverage the FFT for non-Cartesian data [37, 191, 214, 252, 333, 334, 377, 488, 509, 532, 560, 567].

2.2.2 Hilbert Space Reconstruction for Square-Integrable Functions on Ω

The second reconstruction approach is obtained by choosing a solution with finite $\mathcal{L}_2$ norm and with support on Ω (we will use $\mathcal{L}_2(\Omega)$ to denote the Hilbert space [417] of images satisfying these constraints) that matches the data as closely as possible. For example, we could choose a

reconstructed image from the set Γ, the set of all maximally data consistent solutions with respect to the Euclidean distance metric, i.e.,

$$\begin{aligned}\Gamma &= \left\{ \rho\left(\mathbf{x}\right) \in \mathcal{L}_2(\Omega) : \sum_{m=1}^{M} \left| d_m - \int_{\Omega} \rho\left(\mathbf{x}\right) e^{-i2\pi \mathbf{k}_m \cdot \mathbf{x}} d\mathbf{x} \right|^2 \text{ is minimal} \right\} \\ &= \left\{ \rho\left(\mathbf{x}\right) \in \mathcal{L}_2(\Omega) : \mathcal{F}^* \mathcal{F} \rho = \mathcal{F}^* \mathbf{d} \right\},\end{aligned} \tag{2.24}$$

where $\mathcal{F}^* : \mathbb{C}^M \rightarrow \mathcal{L}_2(\Omega)$ is the adjoint of $\mathcal{F}$,[6] $\mathbf{d}$ is the length-M vector with mth entry equal to d_m, and the second equality is a standard result from optimization in Hilbert spaces [417]. The images in Γ are all *least-squares* solutions to the inverse problem, since they all have minimal sum-of-squares error with respect to the measured data. However, one practical difficulty is that the set Γ generally contains an infinite number of images, and we would need to specify some additional constraints to choose a unique reconstruction. The reason that Γ is infinitely large is because we have only sampled a finite amount of data, which means that there are generally an infinite number of images in the *null space* $\mathcal{N}(\mathcal{F})$ of our sampled Fourier transform operator $\mathcal{F} : \mathcal{L}_2(\Omega) \rightarrow \mathbb{C}^M$. This null space $\mathcal{N}(\mathcal{F})$ is defined as

$$\mathcal{N}(\mathcal{F}) = \left\{ \rho\left(\mathbf{x}\right) \in \mathcal{L}_2(\Omega) : \int_{\Omega} \rho\left(\mathbf{x}\right) e^{-i2\pi \mathbf{k}_m \cdot \mathbf{x}} d\mathbf{x} = 0 \text{ for all } m \in \{1, 2, \ldots, M\} \right\}. \tag{2.25}$$

It can be shown that Γ is the linear variety

$$\Gamma = \left\{ \rho\left(\mathbf{x}\right) \in \mathcal{L}_2(\Omega) : \rho_0\left(\mathbf{x}\right) + z\left(\mathbf{x}\right), \text{ with } z\left(\mathbf{x}\right) \in \mathcal{N}(\mathcal{F}) \right\}, \tag{2.26}$$

where $\rho_0\left(\mathbf{x}\right)$ is any least-squares solution [417]. A standard heuristic approach to choosing a unique reconstruction from this linear variety is to use the pseudo-inverse solution, which selects the unique image from Γ with minimum $\mathcal{L}_2$-norm [48, 49, 207, 289, 417, 618, 625]. This minimum-

[6] Note that we use * with different meanings in different contexts. For functions and scalars, we use * to denote complex conjugation, while for operators, we use * to denote adjoint. See Ref. [417] for a formal definition of the adjoint operator.

norm least-squares (MNLS) solution is given by

$$\begin{aligned}\hat{\rho}(\mathbf{x}) &= \mathcal{F}^{*}\left(\mathcal{F}\mathcal{F}^{*}\right)^{\dagger}\mathbf{d} \\ &= \mathbf{1}_{\{\mathbf{x}\in\Omega\}}\sum_{m=1}^{M} c_m e^{i2\pi\mathbf{k}_m\cdot\mathbf{x}},\end{aligned} \tag{2.27}$$

where † denotes the matrix pseudo-inverse [248, 417, 531], c_m is the mth element of the vector $\mathbf{c}$, $\mathbf{c} = \left(\mathcal{F}\mathcal{F}^{*}\right)^{\dagger}\mathbf{d}$, and where the elements of the $M \times M$ matrix $\mathcal{F}\mathcal{F}^{*}$ are given by

$$\left[\mathcal{F}\mathcal{F}^{*}\right]_{m_1 m_2} = \int_{\Omega} e^{i2\pi\left(\mathbf{k}_{m_1}-\mathbf{k}_{m_2}\right)\cdot\mathbf{x}}. \tag{2.28}$$

Note that the reconstruction obtained with Eq. (2.27) is very similar to that obtained with the approximation of analytic Fourier inversion given by Eq. (2.23); indeed, in the case when $\mathcal{F}\mathcal{F}^{*}$ is diagonal (i.e., when the Fourier sampling functions $\left\{e^{-i2\pi\mathbf{k}_m}\right\}_{m=1}^{M}$ are orthogonal over Ω), the pseudo-inverse reconstruction is a special case of Eq. (2.23).

2.2.3 Reconstructions Using Finite-Dimensional Linear Image Models

The final class of methods represents the image in terms of a finite-dimensional linear model, i.e.,

$$\rho(\mathbf{x}) = \sum_{n=1}^{N} \rho_n \phi_n(\mathbf{x}), \tag{2.29}$$

where N is the dimension of the linear model, and $\{\phi_n(\mathbf{x})\}$ are functions in $\mathcal{L}_2(\Omega)$. While some approaches use specialized anatomically-adapted basis functions (e.g., [32, 76, 119, 121, 319, 400, 403, 405, 592, 611]), basis functions modulated by prior image-phase information [101, 262, 307, 400, 673], or basis functions defined in *k*-space [223, 459, 544, 567], a more common approach in the imaging literature is to use translations of a single "voxel function" $\phi(\mathbf{x})$ to represent the image, i.e.,

$$\phi_n(\mathbf{x}) = \phi(\mathbf{x} - \mathbf{x}_n), \tag{2.30}$$

where $\{\mathbf{x}_n\}$ specify the voxel locations (e.g., [127, 168, 207, 289, 364, 512, 592]). Standard choices of $\phi(\mathbf{x})$ include the Dirac delta function or an appropriate polynomial spline function [616]. Under this parameterization,

$$\begin{aligned}\int_\Omega \rho(\mathbf{x})\, e^{-i2\pi\mathbf{k}_m\cdot\mathbf{x}} d\mathbf{x} &= \sum_{n=1}^{N} \rho_n \int_\Omega \phi(\mathbf{x}-\mathbf{x}_n)\, e^{-i2\pi\mathbf{k}_m\cdot\mathbf{x}} d\mathbf{x} \\ &= \Phi(\mathbf{k}_m) \sum_{n=1}^{N} \rho_n e^{-i2\pi\mathbf{k}_m\cdot\mathbf{x}_n}, \\ &= \mathbf{F}\boldsymbol{\rho},\end{aligned} \tag{2.31}$$

where $\Phi(\mathbf{k})$ is the Fourier transform of $\phi(\mathbf{x})$, $\mathbf{F}$ is the $M \times N$ matrix with elements

$$[\mathbf{F}]_{mn} = \Phi(\mathbf{k}_m)\, e^{-i2\pi\mathbf{k}_m\cdot\mathbf{x}_n}, \tag{2.32}$$

and $\boldsymbol{\rho}$ is the length-N vector of ρ_n coefficients. With a finite-dimensional parametric model, it becomes possible to use statistically-motivated reconstructions such as *maximum-likelihood* (ML) estimation that make use of the probability distribution of the measurements to help determine the unknown model parameters [354, 458, 514]. Given the assumption of complex white Gaussian noise measurements and assuming $\boldsymbol{\rho}$ is known, the probability distribution of $\mathbf{d}$ is given by

$$p(\mathbf{d};\boldsymbol{\rho}) = \frac{1}{\pi^M \sigma^{2M}} e^{-\frac{1}{\sigma^2}\|\mathbf{F}\boldsymbol{\rho}-\mathbf{d}\|_{\ell_2}^2}, \tag{2.33}$$

where σ is the standard deviation of the complex noise samples, and the ℓ_p norm of a length-M vector is defined for $p \geq 1$ as

$$\|\mathbf{d}\|_{\ell_p} = \left(\sum_{m=1}^{M} |d_m|^p\right)^{1/p}. \tag{2.34}$$

In statistics, $p(\mathbf{d};\boldsymbol{\rho})$ is frequently called the *likelihood function*.

The ML estimate of $\boldsymbol{\rho}$ is obtained by maximizing Eq. (2.33) with respect to $\boldsymbol{\rho}$. Maximizing the likelihood is equivalent to minimizing the negative log-likelihood, so that the ML solution is

obtained by solving the simple least-squares problem

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \|\mathbf{F}\boldsymbol{\rho} - \mathbf{d}\|_{\ell_2}^2 . \tag{2.35}$$

The ML solution is only unique when $\mathbf{F}^H\mathbf{F}$ has full rank (a necessary condition for this is that $M \geq N$), in which case

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= \mathbf{F}^H \left(\mathbf{F}\mathbf{F}^H\right)^\dagger \mathbf{d} \\ &= \left(\mathbf{F}^H\mathbf{F}\right)^{-1} \mathbf{F}^H\mathbf{d}. \end{aligned} \tag{2.36}$$

When the ML solution is unique, it satisfies a number of desirable properties. In particular, $\hat{\boldsymbol{\rho}}$ is a linear function of the measured data, is an unbiased estimate of $\boldsymbol{\rho}$, and has the smallest possible variance among all possible unbiased estimators [354, 458, 514].[7] The ML solution is also quite similar in form to the MNLS reconstruction of Eq. (2.27) and the conjugate phase reconstruction of Eq. (2.23).

In practice, $\mathbf{F}^H\mathbf{F}$ may not have full rank or may be poorly conditioned (i.e., the solution is not very stable with respect to small perturbations in the data). In these cases, a penalized maximum-likelihood (PML) estimate is often computed [164, 207, 625]. PML estimates are solutions to a regularized ML problem of the form

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \|\mathbf{F}\boldsymbol{\rho} - \mathbf{d}\|_{\ell_2}^2 + R\left(\boldsymbol{\rho}\right), \tag{2.37}$$

where $R\left(\boldsymbol{\rho}\right)$ is a penalty (regularization) function that is used to ensure that the reconstruction is both unique and stable with respect to noise perturbations. The choice of $R\left(\boldsymbol{\rho}\right)$ often reflects prior information about the image being reconstructed, and the objective of regularization is generally

[7] These facts follow from the Gauss-Markov theorem [354], which states that the least-squares solution (when it exists and is unique) is the best (i.e., minimum covariance) linear unbiased estimator for any problem where the data measurements are linear in the unknown parameters and the noise samples (not necessarily Gaussian or identically distributed) are zero-mean, uncorrelated, and homoscedastic. More general versions of the Gauss-Markov theorem can be found in Refs. [521, 531].

to make reconstructions "more regular" with respect to the class of images under consideration. Most existing approaches for MRI use $R(\boldsymbol{\rho})$ that reflect the prior information that medical images are often spatially smooth (with occasional discontinuities) [50, 131, 163, 237, 546] or are sparse or compressible in appropriately-chosen transform domains [117, 137, 180, 291, 420, 484].

In some cases, the PML reconstruction can have a Bayesian interpretation. Bayesian estimation methods assume that the unknown parameters $\boldsymbol{\rho}$ are random variables with some known prior probability distribution $p(\boldsymbol{\rho})$, and use this prior to extract the posterior distribution of the image conditioned on the observed data $p(\boldsymbol{\rho}; \mathbf{d})$ using Bayes' rule [354, 514]:

$$p(\boldsymbol{\rho}; \mathbf{d}) = \frac{p(\mathbf{d}; \boldsymbol{\rho}) p(\boldsymbol{\rho})}{p(\mathbf{d})}, \tag{2.38}$$

where $p(\mathbf{d})$ is the probability of observing the measured data. Due to the inclusion of the prior distribution, this posterior distribution contains more information than the likelihood alone, which can lead to improved parameter estimates relative to ML. Common approaches to Bayesian image estimation using this posterior distribution are to compute the *minimum mean-squared error* (MMSE) estimate

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= \min_{\boldsymbol{\theta}} E\left[\left| \boldsymbol{\theta} - \boldsymbol{\rho} \right|^2 ; \mathbf{d} \right] \\ &= \min_{\boldsymbol{\theta}} \int \left| \boldsymbol{\theta} - \boldsymbol{\rho} \right|^2 p(\boldsymbol{\rho}; \mathbf{d}) d\boldsymbol{\rho} \\ &= E\left[\boldsymbol{\rho}; \mathbf{d} \right], \end{aligned} \tag{2.39}$$

or the *maximum a posteriori* (MAP) estimate

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= \max_{\boldsymbol{\rho}} p(\boldsymbol{\rho}; \mathbf{d}) \\ &= \max_{\boldsymbol{\rho}} p(\mathbf{d}; \boldsymbol{\rho}) p(\boldsymbol{\rho}) \\ &= \min_{\boldsymbol{\rho}} - \ln p(\mathbf{d}; \boldsymbol{\rho}) - \ln p(\boldsymbol{\rho}) \\ &= \min_{\boldsymbol{\rho}} \| \mathbf{F}\boldsymbol{\rho} - \mathbf{d} \|_{\ell_2}^2 - \sigma^2 \ln p(\boldsymbol{\rho}). \end{aligned} \tag{2.40}$$

As can be seen, the MAP estimate in Eq. (2.40) has the same general form as the PML estimate in Eq. (2.37). Generally, PML estimates can have a Bayesian interpretation if there exists a normalization constant Z (often called the *partition function*, following the nomenclature of statistical mechanics) such that

$$p(\boldsymbol{\rho}) = \frac{1}{Z} e^{\frac{-R(\boldsymbol{\rho})}{\sigma^2}} \tag{2.41}$$

is a valid probability distribution.

2.2.4 Resolution and Noise Characteristics of Conventional Reconstructions

The conjugate phase, MNLS, and ML solutions are all linear reconstructions, and can be unified under the following linear reconstruction formula:

$$\hat{\rho}(\mathbf{x}) = \sum_{m=1}^{M} g_m(\mathbf{x}) d_m, \tag{2.42}$$

where $\{g_m(\mathbf{x})\}$ is a set of spatially-varying reconstruction coefficients. It is easy to show by substitution of Eq. (2.17) into Eq. (2.42) that

$$\hat{\rho}(\mathbf{x}) = \int \rho(\mathbf{y}) h_{\mathbf{x}}(\mathbf{y}) d\mathbf{y} + \bar{\eta}_{\mathbf{x}}, \tag{2.43}$$

with

$$h_{\mathbf{x}}(\mathbf{y}) = \sum_{m=1}^{M} g_m(\mathbf{x}) e^{-i2\pi \mathbf{k}_m \cdot \mathbf{y}}, \tag{2.44}$$

where $h_{\mathbf{x}}(\mathbf{y})$ is called the *spatial response function* (SRF) , and $\bar{\eta}_{\mathbf{x}}$ is a zero-mean Gaussian random process with covariance

$$C_{\bar{\eta}}(\mathbf{x}, \mathbf{x}') = E[\bar{\eta}_{\mathbf{x}} \bar{\eta}^*_{\mathbf{x}'}] = \sigma^2 \sum_m g_m(\mathbf{x}) g^*_m(\mathbf{x}'). \tag{2.45}$$

For a fixed image location $\mathbf{x}$, the variance of the reconstructed image value is thus given by

$$C_{\bar{\eta}}(\mathbf{x},\mathbf{x}) = E\left[|\bar{\eta}_{\mathbf{x}}|^2\right] = \sigma^2 \sum_m |g_m(\mathbf{x})|^2. \tag{2.46}$$

The relationship in Eq. (2.43) shows that the estimated value of the image at a given spatial location is actually a weighted average of the true image, plus additional noise. To achieve high resolution, the SRF $h_{\mathbf{x}}(\mathbf{y})$ should be highly concentrated about the spatial location $\mathbf{x}$. However, the extent to which this is possible is governed by both the *k*-space sampling trajectory and the particular choice of reconstruction coefficients.

To illustrate the effects of the number of measured *k*-space samples on the resolution and SNR of the reconstruction, we will consider a simple one-dimensional (1D) example in which $\Omega = [-1/2, 1/2]$, M is an even number, and $\{\mathbf{k}_m\}_{m=1}^M = \{-M/2, -M/2+1, \ldots, M/2-1\}$ is a truncated version of a periodic lattice satisfying Eq. (2.21) with the property that the corresponding Fourier sampling kernels $e^{-i2\pi\mathbf{k}_m\cdot\mathbf{x}}$ are orthogonal over Ω. In this case, the MNLS reconstruction has the form of a conjugate phase reconstruction given by

$$\hat{\rho}(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x}\in\Omega\}} \sum_{m=1}^{M} d_m e^{i2\pi\mathbf{k}_m\cdot\mathbf{x}}. \tag{2.47}$$

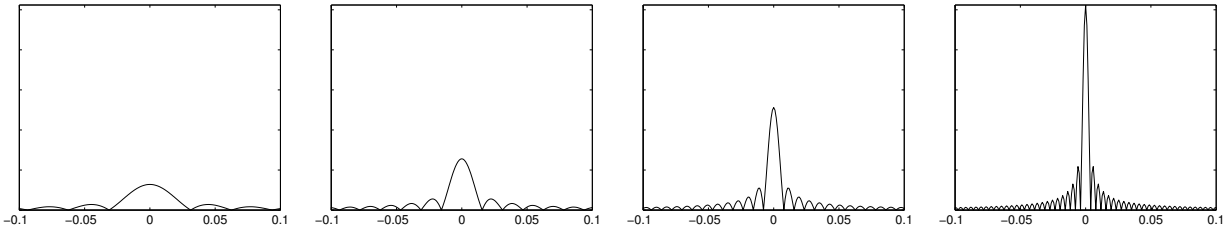
The corresponding SRF has the form of a spatially-windowed Dirichlet kernel

$$h_{\mathbf{x}}(\mathbf{y}) = \mathbf{1}_{\{\mathbf{x}\in\Omega\}} \frac{\sin(\pi M(\mathbf{x}-\mathbf{y}))}{\sin(\pi(\mathbf{x}-\mathbf{y}))} e^{-i\pi(\mathbf{x}-\mathbf{y})}. \tag{2.48}$$

An important characteristic of this SRF is that it becomes more concentrated around $\mathbf{x}$ (i.e., the resolution improves) as the number of Fourier samples M increases, as illustrated in Fig. 2.1.

The covariance function $C_{\bar{\eta}}(\mathbf{x},\mathbf{x}')$ for this case has a similar form to that of the SRF

$$C_{\bar{\eta}}(\mathbf{x},\mathbf{x}') = \mathbf{1}_{\{\mathbf{x},\mathbf{x}'\in\Omega\}} \frac{\sin(\pi M(\mathbf{x}-\mathbf{x}'))}{\sin(\pi(\mathbf{x}-\mathbf{x}'))} e^{-i\pi(\mathbf{x}-\mathbf{x}')}, \tag{2.49}$$



(a) $M = 32$ (b) $M = 64$ (c) $M = 128$ (d) $M = 256$

Figure 2.1: The magnitude of SRFs computed using Eq. (2.48) with $\mathbf{x} = 0$ for different numbers of *k*-space samples M. As the number of samples increases, so does the spatial resolution.

and the noise variance at a particular spatial location $\mathbf{x}$ in the reconstruction is given by

$$\sigma_{\mathbf{x}}^2 = C_{\bar{\eta}}(\mathbf{x}, \mathbf{x}) = M\sigma^2. \tag{2.50}$$

This simple example has two important characteristics that are generally true in more complicated linear reconstruction scenarios:

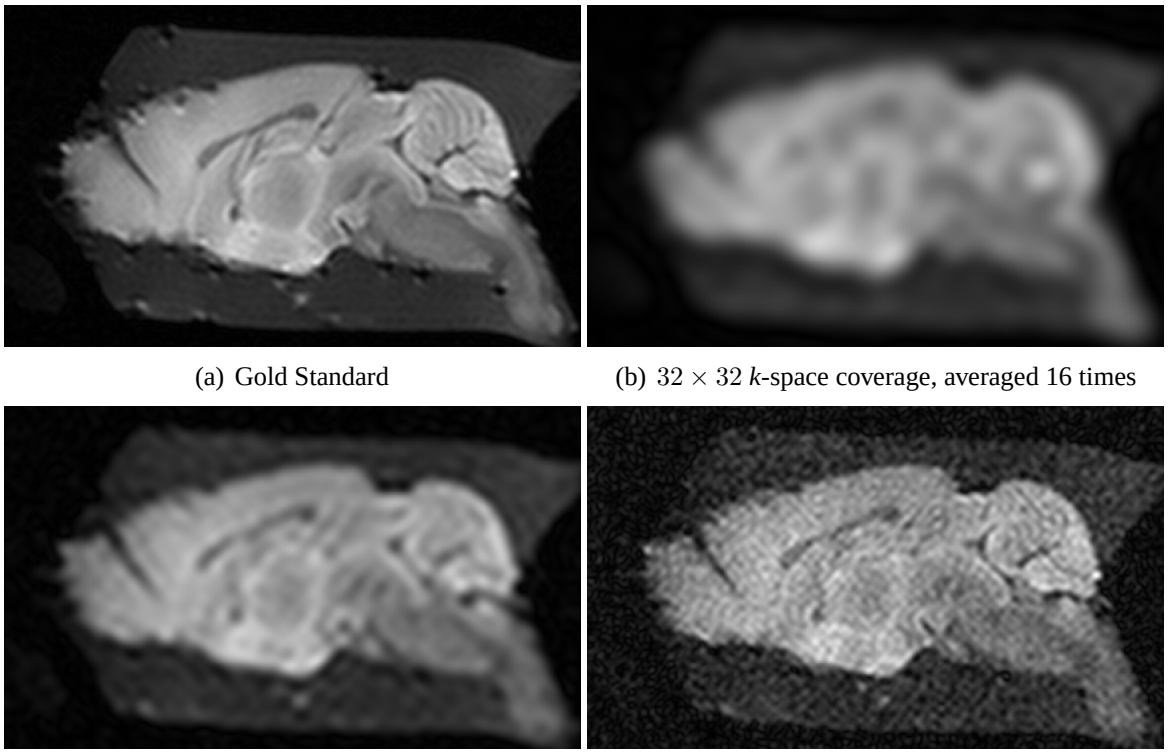
1. The spatial resolution of a reconstructed image is limited by the *k*-space coverage of the experiment, and can only improve by increasing M and collecting additional *k*-space data with non-redundant information.

2. The noise variance in the reconstruction increases as additional non-redundant *k*-space samples are incorporated into the reconstruction.

As a result of these characteristics, for a fixed M, the design of the *k*-space sampling pattern typically represents a trade-off between spatial resolution and SNR, as shown in Figs. 2.2 and 2.3.

2.3 Matrix Rank

This section presents a brief review of matrix rank, which will be particularly helpful for understanding the material in Ch. 5. More complete descriptions of matrix rank can be found in standard texts covering matrices and linear algebra [248, 308, 458]. Formally, the rank of an $N \times M$ matrix



(a) Gold Standard

(b) 32×32 *k*-space coverage, averaged 16 times

(c) 64×64 *k*-space coverage, averaged 4 times

(d) 128×128 *k*-space coverage, no averaging

Figure 2.2: Illustration of the trade-off between spatial resolution and SNR with fixed M for conventional linear reconstructions. (a) Gold standard image of a fixed mouse brain. (b-d) Simulated noisy MNLS/conjugate phase reconstructions with different *k*-space trajectories. The support set Ω was assumed to be rectangular, and the Fourier samples are arranged on a rectangular grid at the Nyquist rate. In particular, the reconstructions in (b), (c), and (d) used the central 32×32, 64×64, and 128×128 samples from the Nyquist grid. As *k*-space coverage increases, the spatial resolution improves while the SNR decreases.

$\mathbf{A}$ is defined as the largest number of linearly independent columns of $\mathbf{A}$, which is also equal to the largest number of linearly independent rows of $\mathbf{A}$. The rank satisfies the following properties:

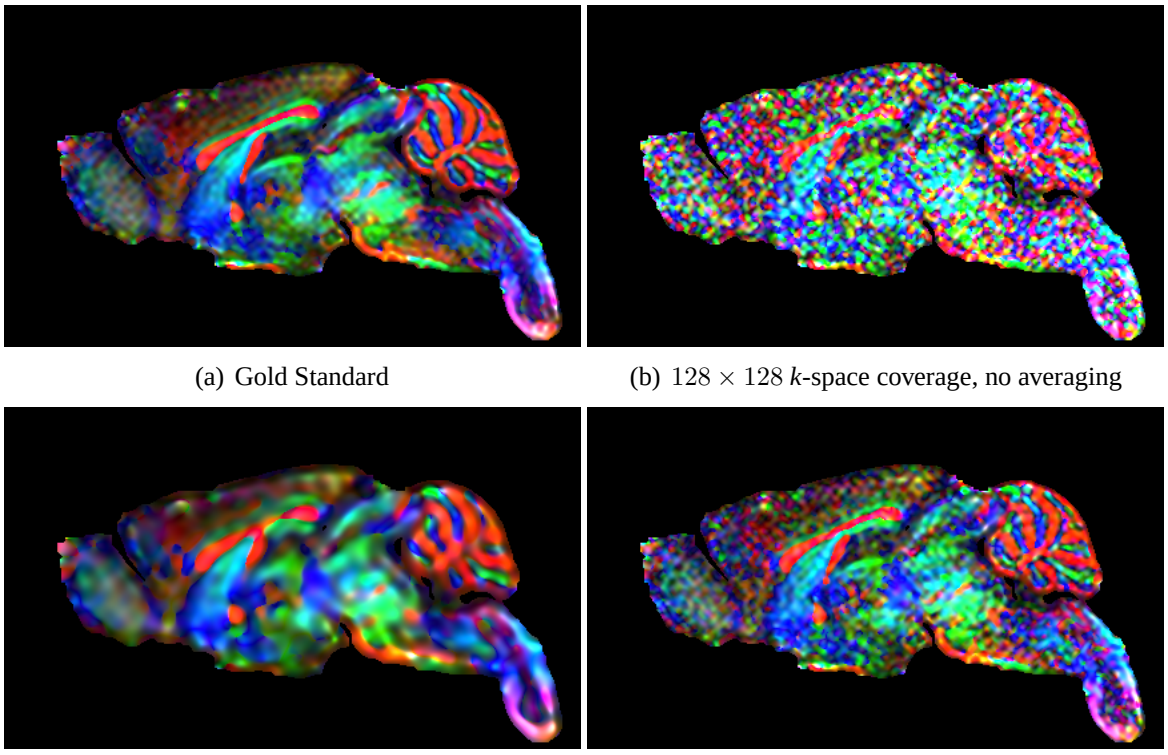
1. $\operatorname{rank}(\mathbf{A}) \leq \min(N, M)$. $\mathbf{A}$ is said to be full-rank if $\operatorname{rank}(\mathbf{A}) = \min(N, M)$.

2. $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}\left(\mathbf{A}^H\right)$.

3. For two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size, $\operatorname{rank}(\mathbf{A} + \mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B})$.

4. For a matrix $\mathbf{C}$ with N columns, $\operatorname{rank}(\mathbf{CA}) \leq \min(\operatorname{rank}(\mathbf{C}), \operatorname{rank}(\mathbf{A}))$.

5. For a full-rank $N \times N$ matrix $\mathbf{C}$ and a full-rank $M \times M$ matrix $\mathbf{D}$, $\operatorname{rank}(\mathbf{CAD}) = \operatorname{rank}(\mathbf{A})$.



(a) Gold Standard

(b) 128×128 *k*-space coverage, no averaging

(c) 64×64 *k*-space coverage, averaged 4 times

(d) 128×128 *k*-space coverage, averaged 8 times

Figure 2.3: Illustration of the trade-off between spatial resolution and SNR with fixed M for conventional linear reconstructions in the context of a quantitative parameter estimation experiment. (a) Gold standard quantitative diffusion tensor parameter map (color-coded fractional anisotropy) [460] estimated from a sequence of diffusion-weighted MR images. (b-d) Simulated noisy reconstructions with different acquisition strategies. The reconstruction procedures and sampling trajectories match those of Fig. 2.2. (b) The high-resolution parameter map is too noisy to be useful without additional averaging. (c) SNR can be improved by acquiring data at lower resolution, though this comes at the expense of a significant loss of anatomical detail. (d) Alternatively, the high-resolution data can be averaged to improve SNR while maintaining resolution, though this comes at the expense of a significant increase in data acquisition time.

An important fact about matrices is that they can be represented using the singular value decomposition (SVD). In particular, if $\mathbf{A}$ is an $N \times M$ matrix, then there exists a set of $\min(M, N)$ different $N \times 1$ orthonormal vectors $\{\mathbf{p}_n\}_{n=1}^{N}$, a set of $\min(M, N)$ different $M \times 1$ orthonormal vectors $\{\mathbf{q}_m\}_{m=1}^{M}$, and a uniquely-defined set of $\min(M, N)$ non-negative real numbers σ_k (known as "singular values") satisfying $\sigma_i \geq \sigma_j$ when $i < j$, such that

$$\mathbf{A} = \sum_{k=1}^{\min(M,N)} \sigma_k \mathbf{p}_k \mathbf{q}_k^H = \mathbf{P}\boldsymbol{\Sigma}\mathbf{Q}^H. \tag{2.51}$$

In this expression, the $N \times \min(M,N)$ matrix $\mathbf{P}$ has its kth column equal to $\mathbf{p}_k$, the $M \times \min(M,N)$ matrix $\mathbf{Q}$ has its kth column equal to $\mathbf{q}_k$, and the $\min(M,N) \times \min(M,N)$ matrix $\mathbf{\Sigma}$ is diagonal with kth diagonal entry equal to σ_k.

The SVD provides useful insight into matrix rank. In particular, a matrix with rank L will have exactly L non-zero singular values, such that

$$\mathbf{A} = \sum_{k=1}^{\min(M,N)} \sigma_k \mathbf{p}_k \mathbf{q}_k^H = \sum_{k=1}^{\text{rank}(\mathbf{A})} \sigma_k \mathbf{p}_k \mathbf{q}_k^H. \tag{2.52}$$

In addition, the SVD provides a convenient structure for computing optimal low rank approximations. Consider the optimization problem

$$\hat{\mathbf{A}}_R = \arg \min_{\substack{\mathbf{A}_R \in \mathbb{C}^{N \times L} \\ \text{rank}(\mathbf{A}_R) \leq R}} \|\mathbf{A}_R - \mathbf{A}\|_F, \tag{2.53}$$

where R is assumed to be less than or equal to $\min(M,N)$, and $\|\cdot\|_F$ is the standard matrix Frobenius norm defined as

$$\|\mathbf{A}\|_F = \sqrt{\sum_{n=1}^{N} \sum_{m=1}^{M} |[\mathbf{A}]_{nm}|^2} = \sqrt{\sum_{k=1}^{\text{rank}(\mathbf{A})} \sigma_k^2}. \tag{2.54}$$

The solution to this problem, known as the Eckart-Young approximation [192], is given by

$$\hat{\mathbf{A}}_R = \sum_{k=1}^{R} \sigma_k \mathbf{p}_k \mathbf{q}_k^H, \tag{2.55}$$

and the optimal approximation error is given by

$$\left\| \hat{\mathbf{A}}_R - \mathbf{A} \right\|_F = \sqrt{\sum_{k=R+1}^{\text{rank}(\mathbf{A})} \sigma_k^2}. \tag{2.56}$$

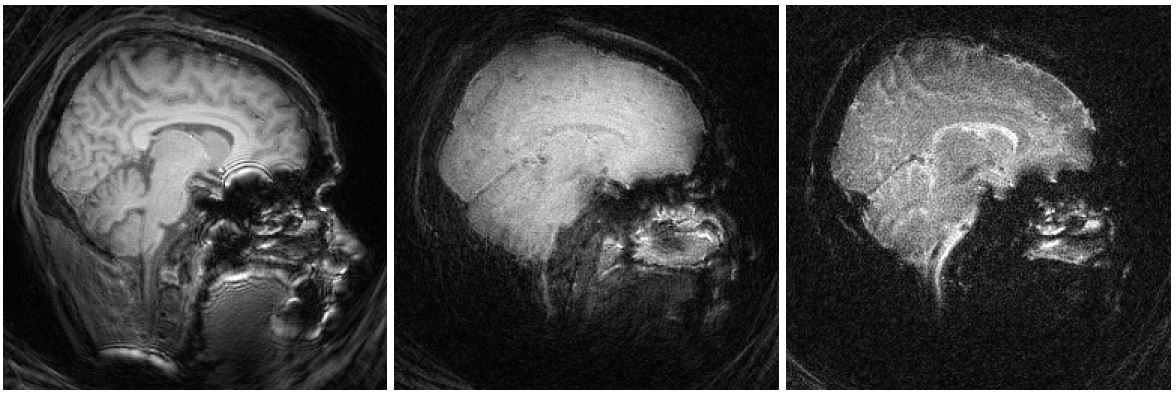
From this, it can be seen that any matrix with a large number of small singular values has the potential to be accurately approximated with a low-rank matrix.

Chapter 3

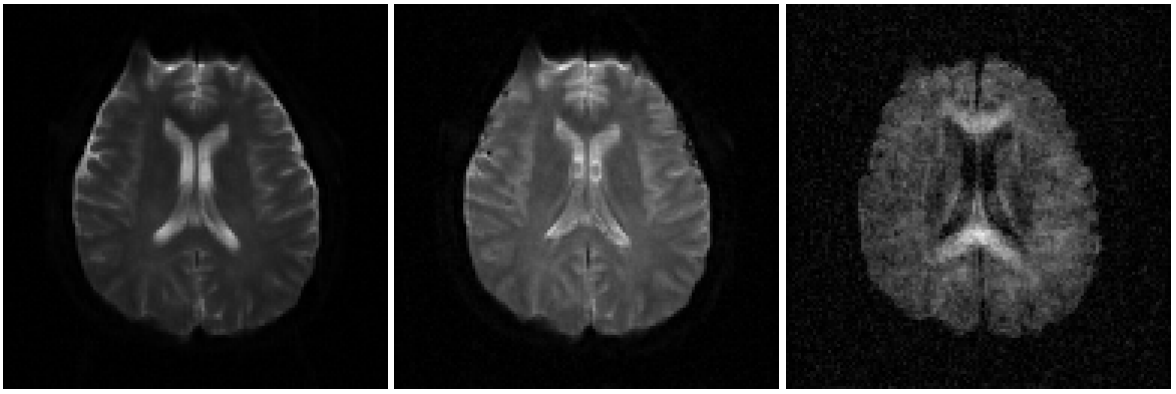
Reconstructing Correlated Images with a Shared Edge Prior

Contrast in MR images is dependent on both the physical properties of the imaging subject and on the experimental parameters used to acquire data. In many exams, coregistered images of the same anatomy are acquired with different contrast, providing different perspectives on the underlying characteristics of the imaging subject. However, as described in the previous chapter, the acquisition of many different high-resolution images can be time-consuming, due to the relatively slow data acquisition process and the trade-off between data acquisition time, image resolution, and SNR. A simple way to improve SNR with fixed spatial resolution is to perform signal averaging, though this comes at the expense of longer data acquisition time. A more common alternative is to simply reduce *k*-space coverage, which will simultaneously reduce data acquisition time and improve the resulting image SNR, though this comes at the cost of significant partial volume artifacts in the reconstructed images.

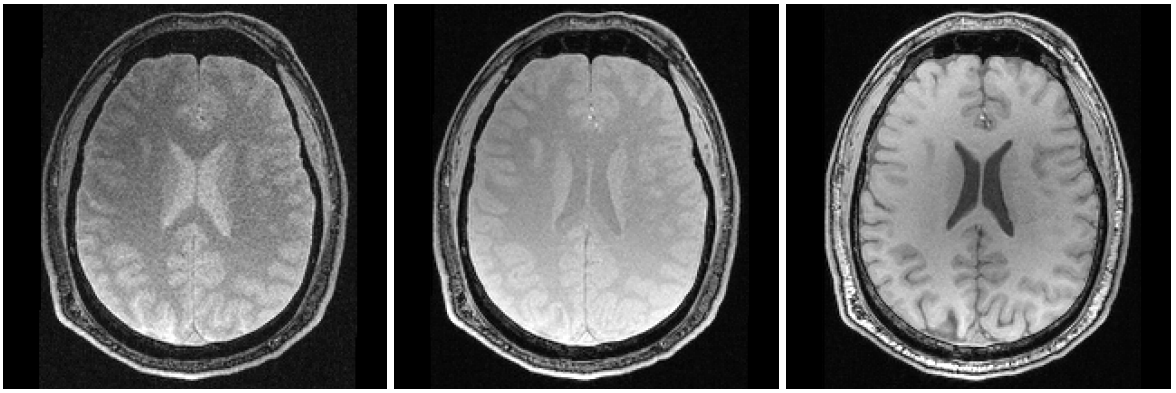
In practice, it can also be possible to use side information from related reference images to alleviate resolution and noise concerns. The main assumption of this kind of constrained image reconstruction approach is that coregistered images from the same subject will frequently be highly correlated. This large amount of correlation can be observed visually in example image sequences shown in Fig. 3.1. The correlation observed in these images is not coincidental, and is the result of the fact that the observed MR signal intensity is a function of the microstructural and chemical characteristics of the tissue. If one type of tissue has significantly different characteristics from another, this can easily lead to consistent observed MR differences between the tissues under a variety of different MR contrast mechanisms. Thus, while the intensities of different features are quite distinct, the images demonstrate, for example, highly-correlated edge structures and wavelet



(a) Multi-Echo Spiral Data



(b) Diffusion Weighted Data



(c) Variable Flip-Angle FLASH Data

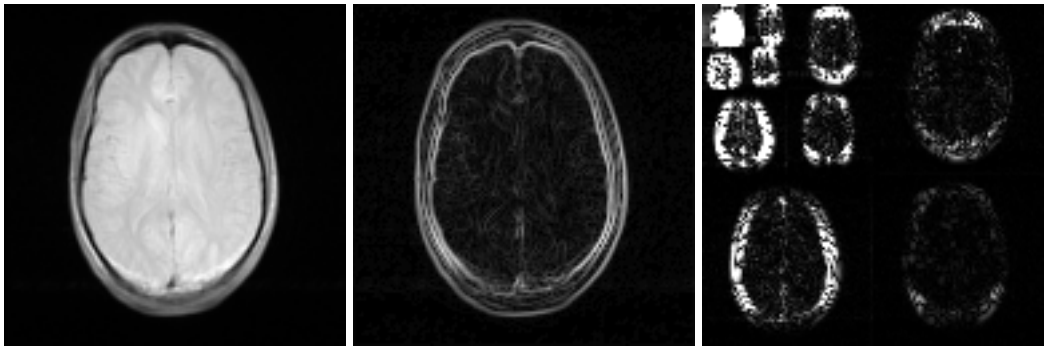
Figure 3.1: Coregistered image sequences with different contrast parameters. (a) Multi-echo spiral data (courtesy of Dr. B. Sutton); data acquisition parameters are described in [593]. (b) Diffusion-weighted image data (courtesy of Dr. X. Zhou); data acquisition parameters are described in [677]. (c) Variable flip-angle FLASH data (courtesy of Dr. N. Schuff). Though the contrast of each image is different, there is significant visual similarity between the different images of the same subject.

coefficients, and this can mean that there is an opportunity to leverage information from one image to help improve reconstruction of another. An example illustrating the high degree of edge and wavelet correlation is shown in Figs. 3.2 and 3.3.

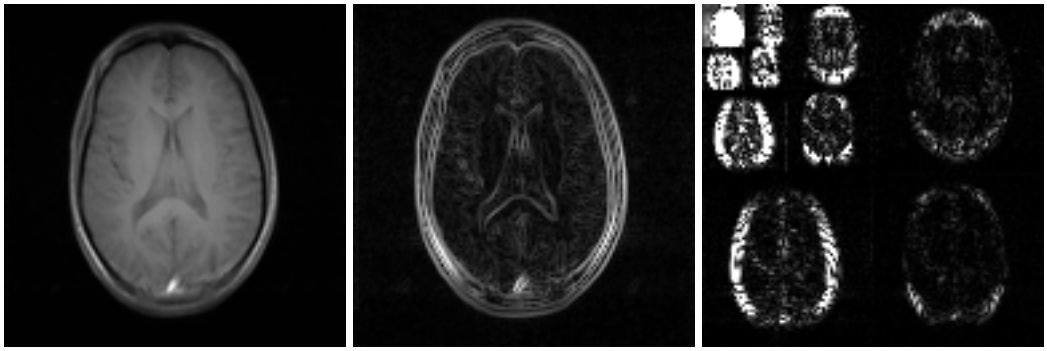
There are two general classes of signal processing methods that have been proposed to use this kind of prior information in the context of Nyquist-sampled data:[1]

1. **Super-Resolution Methods.** This class of methods is based on the idea that high resolution MR images with certain types of contrast are naturally faster to acquire than others. As a result, high-resolution prior information from these images can be fused with low-resolution high-SNR data from slower MR experiments to yield high-resolution reconstructions. The constraints used by these approaches are derived from coregistered high-resolution anatomical datasets, and include information about the locations of different tissues within the field of view (FOV) , edge locations, image phase, and image support [32, 101, 119, 121, 146, 166, 197, 233, 262, 271, 307, 319, 325, 335, 370–372, 400, 403, 405, 464, 475, 512, 592, 611, 673]. Similar super-resolution approaches have been proposed in other imaging modalities, including emission tomography [13, 20, 30, 42, 70, 72, 76, 93, 128, 136, 139, 145, 210, 242, 243, 304, 316, 344, 375, 389, 410, 416, 478, 479, 481, 489, 530, 551, 579, 581, 582, 597, 629, 674], transmission tomography [671], susceptibility/impedance imaging [162, 244], and optical tomography [33, 71, 91, 158, 186, 261, 330, 395, 409, 477]. For this class of methods, the quality of the reconstruction depends heavily on the quality of the constraints that are imposed. If the prior information leads to an accurate signal model with a small number of unknown parameters compared to the number of measured data samples, very high quality reconstructions can be achieved. However, if the constraints are not strong enough, then reconstruction performance improvement may be limited. In addition, if the constraints are inaccurate, then the modeling inaccuracies could result in images with significant artifacts and limited practical utility.

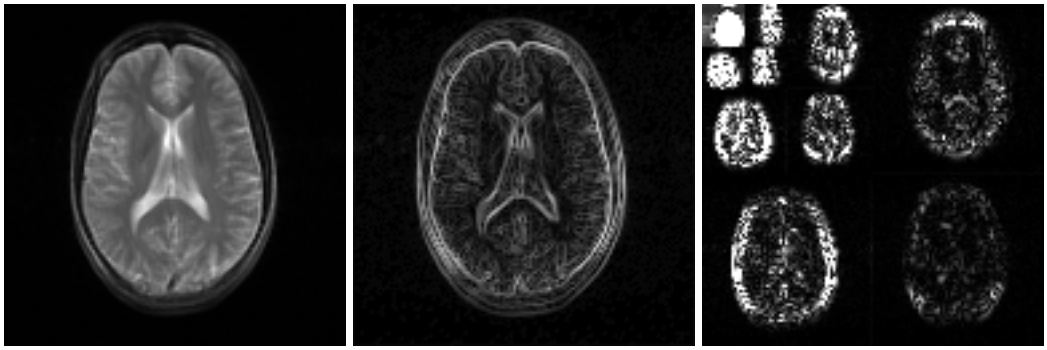
[1] We note that there are other types of constrained methods that have been proposed when sampling is not performed at the Nyquist rate, though we will leave discussion of these to subsequent chapters.



(a) Proton Density Weighted Image

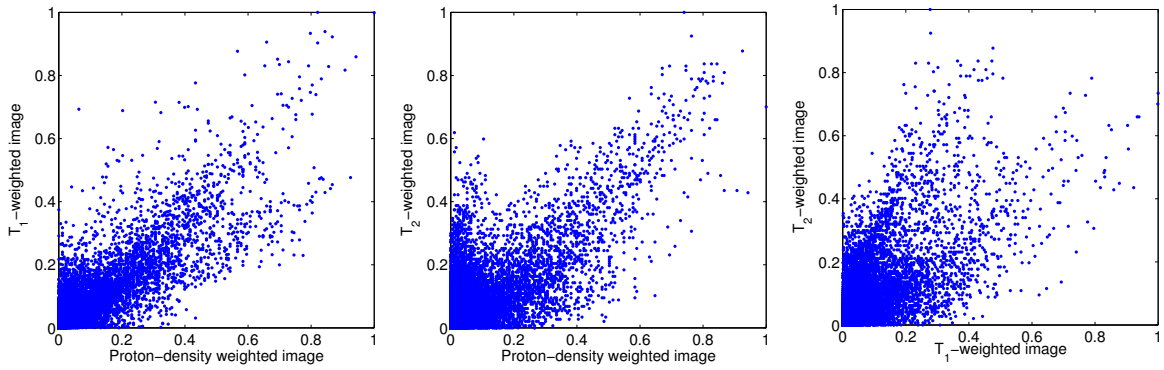


(b) T_1 Weighted Image



(c) T_2 Weighted Image

Figure 3.2: Three different MR brain images of the same anatomy in the same subject, but with acquisition parameters adjusted to yield different image contrast. The first column shows the original images, while the second and third columns show the corresponding edge structures (computed using finite differences) and Daubechies-4 wavelet coefficients, respectively. There is a significant degree of correlation between the image edge structures and wavelet coefficients.



(a) $r = 0.863$ (b) $r = 0.767$ (c) $r = 0.720$

Figure 3.3: Joint histograms of the finite-difference edge magnitudes for the images shown in Fig. 3.2. The correlation coefficient (Pearson's r) between the edge magnitudes of each image pair is given below each subfigure [541], indicating the strong level of statistical correlation.

2. **Denoising Methods.** Denoising methods have the general goal of removing noise contamination from high-resolution images. While many approaches exist that can independently denoise each image in the image sequence [9, 10, 14, 29, 36, 94, 95, 149, 152, 176, 200, 246, 295, 318, 374, 415, 422, 430, 431, 440, 461, 476, 499, 504, 510, 546, 550, 575, 639, 650, 651, 653, 660, 670], it is generally more powerful to use approaches that leverage correlated image structure in the image sequence. Approaches for the joint denoising of a sequence of images can again be divided into two groups:

   (a) **Parametric Contrast Models.** Techniques in the first group begin by using a constrained image model in which the signal intensity variations between different images in the sequence are modeled for each voxel in terms of a small number (relative to the number of images in the sequence) of contrast parameters. Given this model, denoising is achieved by directly enforcing that the reconstructed images obey the contrast model, and/or by imposing prior information (e.g., spatial smoothness) directly on the spatial maps of the contrast parameters. The body of work on parameter-space denoising is large; some representative examples from the MR literature include regularized field inhomogeneity mapping [215, 303], regularized MR spectroscopic imaging [372, 552], regularized pharmacokinetic parameter estimation [558], regularized

relaxation parameter estimation [161], regularized apparent diffusion coefficient estimation [628], regularized probability and orientation distribution function reconstruction [22, 361, 448, 527], and various algorithms for tensor smoothing with application to diffusion tensor MRI (DTI) [31, 36, 134, 142, 148, 167, 204, 216, 287, 637, 643, 650]. However, techniques that regularize in a parametric space usually use relatively simple contrast models to ensure the tractability of the estimation procedure, and thus often lack the ability to cope with model mismatch or fully represent the complex inter-image signal characteristics seen in real biological tissues.

(b) **General Image Models.** The second group of denoising techniques aims to reduce noise in the reconstructed image sequence without imposing a strict contrast model on the image sequence, and is thus more broadly applicable. These techniques are also more amenable to fitting multiple potential contrast models in cases of model ambiguity. These image-domain techniques typically rely on generalizations of methods developed for the reconstruction and restoration of single images. Example methodologies include the joint filtering of vector-valued images using geometric partial differential equations (PDEs) and related concepts (see [613] for review of the image processing literature and [135, 172, 239, 266, 449] for MR application examples), filtering using vector order statistics (see review in [419]), statistical thresholding/filtering in an appropriate transform domain (e.g., see [21, 23, 43, 75, 100, 324, 418, 438, 534, 555, 599, 648] and references), joint reconstruction using Bayesian/regularized statistical optimality criteria to impose spatial smoothness (e.g., see [61, 79, 88, 159, 226, 296, 441, 457, 556, 562, 605, 617, 645, 666] and references), and methods based on non-local statistics [66, 225, 432, 650].

This chapter proposes a new sequence-based constrained imaging scheme that can be applied in both super-resolution and denoising contexts, though we will demonstrate that the scheme is

more effective when used for denoising.[2] The specific form of the reconstruction formulation is based on the following considerations:

1. Images from typical MRI experiments usually consist of "smooth" regions separated by edges. Smoothness constraints can be used to reduce the effects of noise, although edges need to be preserved to avoid amplifying partial volume effects.

2. The edge structures seen in different frames of an image have strong correlation. For example, object support boundaries will exist in every image, regardless of the image contrast. Edge structures should be imposed in a joint fashion.

3. The reconstruction should not be overly sensitive to incorrect prior information.

4. The proposed method should be easily characterizable in terms of the trade-off between resolution and SNR.

Traditional joint-reconstruction formulations for super-resolution and denoising only incorporate the first two items of this list. In contrast to such schemes, our proposed method is not very sensitive to incorrect prior information, and is easily characterized in terms of the trade-off between resolution and SNR. An analysis of reconstruction properties indicates that collecting extended *k*-space information (at the expense of lower SNR) and performing denoising is more advantageous with this approach than attempting to achieve super-resolution.

We formulate the joint reconstruction in the context of statistical estimation using a Markov random field (MRF) smoothness prior [50, 51, 237] that models shared edge structures to achieve image reconstruction with joint feature-preserving regularization, similar to several existing methods [13, 30, 93, 128, 136, 145, 158, 162, 166, 186, 210, 226, 242, 243, 296, 304, 361, 370, 372, 375, 389, 410, 457, 479, 489, 527, 562, 617, 637, 645, 666, 674]. Galatsanos *et al.* [226] previously proposed

[2] Some of the text and figures in this chapter have been previously published in [280, 283], and are copyright of the IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

In addition, some of the figures in this chapter have been previously published in [274], and are copyright of the ISMRM.

reconstructing correlated image sequences using a weighted multidimensional Laplacian regularization operator to impose both spatial and inter-image intensity correlation between the different images from the sequence. However, this form of smoothing can cause signal leakage between different image frames. Schultz and Stevenson [562] proposed a spatial edge-preserving MRF model that couples the reconstruction of different image frames by using prior knowledge about the specific size of the spatial discontinuities that should be observed in each frame. While this procedure works well when the prior edge information is accurate, this type of smoothing can lead to biased results in practice when the discontinuity sizes are estimated poorly. A more widely-used class of methods simply applies smoothness constraints that are spatially-adapted to reduce smoothing at the locations of suspected edges [13, 30, 93, 128, 136, 145, 158, 162, 166, 186, 210, 242, 243, 296, 304, 361, 370, 372, 375, 389, 410, 457, 479, 489, 527, 617, 637, 645, 666, 674]. Some of these methods assume that edge information has been pre-estimated using edge-detection or tissue-classification methods [13, 30, 93, 128, 145, 158, 162, 186, 210, 304, 370, 372, 375, 410, 479], while others jointly estimate image voxel values and edge locations from the collection of images [136, 166, 242, 243, 296, 361, 389, 457, 489, 527, 617, 637, 645, 666, 674].

When jointly estimating the image values and the correlated image edge structure, most of the existing methods use a multi-image extension of the line process prior proposed by Geman and Geman [237] to model edges and the coupling between edge locations in different image frames. However, due to the complexity of this kind of prior, the proposed optimization procedures have frequently used either a computationally intensive simulated annealing procedure or a greedy algorithm that provides only local convergence. Recently, convex optimization methods have also been proposed for joint estimation of images and coupled edge locations [296, 361, 617, 645, 666]. While this leads to fast, globally-optimal computations, the resolution and SNR characteristics of these methods have not been characterized in depth.

Similar to the methods described in the preceding paragraph, our proposed approach performs joint estimation of the images and their edge structures using a line process prior. Our proposed prior and the resulting optimization algorithm are extended from the Geman and Reynolds [235]

formulation of multiplicative half-quadratic regularization (see [54] for review of the early literature, and [15, 131, 163, 328, 471, 472] and references for examples of more recent developments). Like the previous convex formulations described in Refs. [296, 361, 617, 645, 666], our proposed reconstruction problem is also convex, and optimization can be performed using globally convergent iterative reconstruction algorithms. In addition, the proposed prior has a form that is somewhat similar to certain cost functionals that appear in literature on joint sparse approximation [147, 606], Bayesian multivariate modeling of wavelet coefficients [140, 570, 596, 626], and multivariate median filtering [647]. Connections can also be made to the literature on joint PDE filtering of vector-valued images; this is not surprising, given the similarities between geometric PDE filtering and half-quadratic regularization [55, 598]. Our proposed half-quadratic formulation with a shared line process prior can provide new insight into these existing methods, and our proposed scheme offers several distinct advantages for MR imaging applications.

Before introducing our proposed method to leverage this correlated information, we first describe a line process approach to modeling edges.

3.1 Line Processes for Modeling Image Edge Structures

In a seminal paper [237], Geman and Geman introduced a powerful framework for the Bayesian restoration of images, utilizing simply-structured MRFs to invoke complicated image priors. One such MRF image model uses the prior information that natural images are mostly smooth; that is, voxels which are spatially adjacent to each other typically have similar values, and spatial smoothing/averaging can improve SNR without significantly degrading image features. However, images also contain significant edge structures which should not be ignored; when neighboring voxels are very dissimilar, it is likely that an edge structure exists between them, and smoothness constraints should not be strongly enforced. Because of this, Geman and Geman suggested using an image model in which adjacent voxels are related to each other through new auxiliary variables called *line processes*, which are used to model the edge structure of the image (see Fig. 3.4 for illustra-

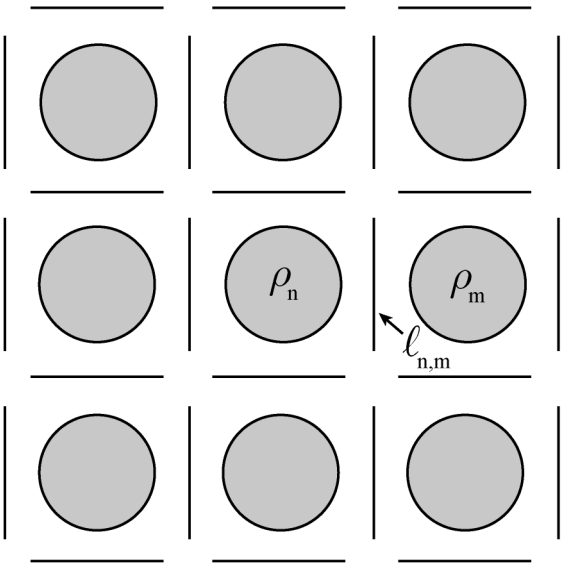


Figure 3.4: An image model with line processes, where voxels are shown as circles and line processes are shown as horizontal and vertical lines. For each pair of adjacent voxels, there is a corresponding line process variable defined between them, which is used to model the edge structure of the image. In this example, the voxel coefficients ρ_n and ρ_m will be encouraged to be similar, unless the corresponding line process variable $\ell_{n,m}$ indicates that an edge is present between them.

tion). The resulting MAP estimation problem for this image model thus requires joint estimation of the image and its edge structure, with the explicit edge modeling allowing for the preservation of important edge features.

In the context of MR imaging with a finite-dimensional discrete model as in Eq. (2.30) (with voxels spaced on a uniform Cartesian grid), the MAP estimation problem with explicit line processes can be posed as

$$\begin{aligned}
\left\{\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\ell}}\right\} &= \arg\max_{\{\boldsymbol{\rho},\boldsymbol{\ell}\}} p(\boldsymbol{\rho}, \boldsymbol{\ell}; \mathbf{d}) \\
&= \arg\max_{\{\boldsymbol{\rho},\boldsymbol{\ell}\}} p(\mathbf{d}; \boldsymbol{\rho}, \boldsymbol{\ell}) p(\boldsymbol{\rho}, \boldsymbol{\ell}) \\
&= \arg\max_{\{\boldsymbol{\rho},\boldsymbol{\ell}\}} p(\mathbf{d}; \boldsymbol{\rho}) p(\boldsymbol{\rho}, \boldsymbol{\ell}) \\
&= \arg\max_{\{\boldsymbol{\rho},\boldsymbol{\ell}\}} p(\mathbf{d}; \boldsymbol{\rho}) p(\boldsymbol{\rho}; \boldsymbol{\ell}) p(\boldsymbol{\ell}) \\
&= \arg\min_{\{\boldsymbol{\rho},\boldsymbol{\ell}\}} -\ln p(\mathbf{d}; \boldsymbol{\rho}) - \ln p(\boldsymbol{\rho}; \boldsymbol{\ell}) - \ln p(\boldsymbol{\ell}),
\end{aligned} \tag{3.1}$$

where $\boldsymbol{\ell}$ is the vector of all the line process variables, $p(\boldsymbol{\rho}, \boldsymbol{\ell}; \mathbf{d})$ is the posterior distribution of the image and the line process variables given the observed data, $p(\mathbf{d}; \boldsymbol{\rho}, \boldsymbol{\ell})$ is the probability of the data given the image and the line process variables, $p(\mathbf{d}; \boldsymbol{\rho})$ is the conditional probability of the data given the image model as introduced in Chapter 2, $p(\boldsymbol{\rho}, \boldsymbol{\ell})$ is the prior distribution of the image and line process variables, $p(\boldsymbol{\rho}; \boldsymbol{\ell})$ is the conditional distribution of the image variables given the line process variables, and $p(\boldsymbol{\ell})$ is the marginal distribution of the line process variables. In writing these equations, we have assumed that when conditioned on the image, the data is independent of the line process variables, which is quite reasonable given that line processes are simply components of our mathematical model rather than physical objects.

In this work, we choose $p(\boldsymbol{\rho}; \boldsymbol{\ell})$ to be a smoothness prior of the form

$$-\ln p(\boldsymbol{\rho}; \boldsymbol{\ell}) = \sum_{n=1}^{N} \sum_{\substack{m \in \Delta_n \\ m > n}} \ell_{n,m} \left| \rho_n - \rho_m \right|^2 + \varepsilon \sum_{n=1}^{N} \left| \rho_n \right|^2 , \tag{3.2}$$

where Δ_n is the set of all voxels belonging to the *neighborhood* of the nth voxel[3] [237] and $\ell_{n,m}$ is the line process variable between the nth and mth voxels. Conditioned on $\boldsymbol{\ell}$, the negative log-prior of $\boldsymbol{\rho}$ is a quadratic function, and thus has the form of a multivariate Gaussian distribution. Note that the covariance of this multivariate distribution is itself a random variable, due to the stochastic model being used for $\boldsymbol{\ell}$, which implies that $p(\boldsymbol{\rho}, \boldsymbol{\ell})$ has the form of a *Gaussian scale-mixture* [18].

The first term in Eq. (3.2) encourages smooth variation between adjacent voxels, with the strength of the smoothness constraint dependent on the value of the line process variables. In locations where the line process variables are large, smoothness is imposed quite heavily, while smoothness is not imposed heavily in locations where the line process variables have small values. As a result, line processes with small values can be considered as corresponding to locations where edge structures are quite probable and where discontinuities will be preserved. The second term in this expression is unrelated to edge structures, but is used to ensure that $p(\boldsymbol{\rho}; \boldsymbol{\ell})$ is

[3] In this work, we focus on the first-order or nearest-neighbor system [237], for which Δ_n is the set of four pixels that are vertically or horizontally adjacent to pixel n.

a valid probability distribution; if $\varepsilon = 0$, then $\boldsymbol{\rho}$ does not have a mean-value and $p(\boldsymbol{\rho};\boldsymbol{\ell})$ is not integrable. In practice, having a strictly proper prior is not essential, and we will generally use $\varepsilon = 0$, since imposing a mean-value on $\boldsymbol{\rho}$ can bias the reconstruction toward zero. As a result, the approach we will present later is only quasi-Bayesian, and is more appropriately considered as a PML approach. This choice to deviate from a Bayesian reconstruction is reasonable for several reasons: 1) true Bayesian priors for the joint distribution of MR images with different contrast are frequently unavailable; 2) even when an accurate prior is available, it has been shown theoretically that Bayesian MAP reconstructions often deviate substantially from both the data acquisition model and the prior model [470]; 3) the resolution and noise characteristics of the reconstructed image can be described easily using the PML interpretation. Allowing $\varepsilon = 0$, Eq. (3.1) can be rewritten (after combination with Eq. (2.33) and Eq. (3.2)) as

$$\left\{\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\ell}}\right\} = \arg\min_{\{\boldsymbol{\rho},\boldsymbol{\ell}\}} \frac{1}{\sigma^2} \left\|\mathbf{F}\boldsymbol{\rho} - \mathbf{d}\right\|_{\ell_2}^2 + \sum_{n=1}^{N} \sum_{\substack{m \in \Delta_n \\ m > n}} \ell_{n,m} \left|\rho_n - \rho_m\right|^2 - \ln p(\boldsymbol{\ell}). \tag{3.3}$$

The prior specified in Eq. (3.3) provides a reasonable way of statistically modeling and preserving the edge structure of real images, and is quite prevalent in the image reconstruction and enhancement literature. However, the statistical edge model is not complete without specification of the line process prior $p(\boldsymbol{\ell})$. In Geman and Geman's original conception, the line process variables were binary (i.e., either an edge exists or it does not) and interacting (i.e., $p(\boldsymbol{\ell})$ imposed statistical dependence relationships on line processes in nearby spatial locations) [237]. While priors with interacting line processes are very powerful and flexible, optimization of Eq. (3.1) with interacting line processes can be more computationally intensive than with non-interacting line processes, and it can be hard in general to design interacting line process priors that perform significantly better than non-interacting line process priors [54, 328]. As a result, following initial work by Blake and Zisserman [57], many recent edge-preserving regularization methods use non-interacting line process models [54, 131, 235]. With a non-interacting line process model, $p(\boldsymbol{\ell})$ the

$\ell_{n,m}$ values are assumed to be independent such that

$$\ln p(\boldsymbol{\ell}) = \sum_{n=1}^{N} \sum_{\substack{m \in \Delta_n \\ m > n}} \ln p(\ell_{n,m}). \tag{3.4}$$

In this case, and under suitable conditions on $\ln p(\ell_{n,m})$, Eq. (3.3) can be simplified as

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \frac{1}{\sigma^2} \|\mathbf{F}\boldsymbol{\rho} - \mathbf{d}\|_{\ell_2}^2 + \sum_{n=1}^{N} \sum_{\substack{m \in \Delta_n \\ m > n}} \Psi_{nm}\left(|\rho_n - \rho_m|\right), \tag{3.5}$$

where

$$\Psi_{nm}(t) = \inf_{\ell_{n,m}} \left(\ell_{n,m} t^2 - \ln p(\ell_{n,m})\right). \tag{3.6}$$

Conditions on $\ln p(\ell_{n,m})$ under which Eq. (3.5) is valid are discussed in Refs. [131, 235, 328, 471], and we will elaborate on one specific set of conditions in the sequel.

Note that Eq. (3.5) has the same form as the PML reconstruction in Eq. (2.37), despite the fact that line processes were not involved in the construction of Eq. (2.37). This illustrates the fact that many PML image reconstruction approaches have a line process interpretation, even if the line process variables are not represented explicitly [54, 131, 163, 235, 472]. An example of this is that the ℓ_1 norm (Laplace) regularization functional [137, 420, 546] (cf. Chapter 4) can be obtained using the line process model when using the following (improper and unnormalizable) prior for each $\ell_{n,m}$ [54]:

$$p(\ell_{n,m}) \propto e^{-\frac{1}{4\ell_{n,m}}} \mathbf{1}_{\{0 < \ell_{n,m}\}}. \tag{3.7}$$

In practice, it is common for recent regularized reconstruction methods to start directly from Eq. (3.5) using an appropriate choice of $\Psi_{nm}(\cdot)$, rather than starting directly from the formulation involving line processes.

To encourage smooth reconstructions, an important feature of $\Psi_{nm}(\cdot)$ is that it should be non-decreasing, so that larger image discontinuities are never favored over smaller image discontinuities unless there is significant evidence for the discontinuity from the data. An additional desirable

feature of $\Psi_{nm}(\cdot)$ is that it has the ability to preserve edges to some extent. Edge-preserving regularization has been considered by a number of authors [54, 131, 163, 469], and one of the defining features of edge-preserving priors is that they do not significantly penalize large edges from forming, i.e., that $\Psi(t)$ does not grow too rapidly as $t \to \infty$. Generally, this is often assumed to mean that $\Psi(t)$ grows slower than a quadratic function for large values of t. The quadratic function is well known to over-smooth edges, as a result of the fact that the rapid growth of the quadratic penalty function makes it difficult for reconstructions to contain significant edge structures.

3.2 Joint Reconstruction of Images with a Shared Line Process Prior

The previous section introduced the line process model for describing edge structures in natural images. In this section, we introduce a new scheme for joint-reconstruction of coregistered images using a shared line process edge model. Before introducing our proposed reconstruction formulation, we will first establish some notation. We consider the case where a length-Q sequence of coregistered images is acquired with different contrasts from a static object. We let the length-M_q vector $\mathbf{d}^q$ represent the acquired *k*-space data for the qth image frame, let the length-M_q vector $\boldsymbol{\eta}^q$ represent the noise (assumed to be independent and identically distributed Gaussian noise with variance σ_q^2) in the data for the qth image, let the $M_q \times N$ matrix $\mathbf{F}_q$ represent the Fourier acquisition operator, and let the length-N vector $\boldsymbol{\rho}^q$ represent the corresponding image. Note that the *k*-space trajectory is allowed to be different for each image.

Our proposed reconstruction involves solving the following optimization problem:

$$\left\{\hat{\boldsymbol{\rho}}^1, \hat{\boldsymbol{\rho}}^2, \ldots, \hat{\boldsymbol{\rho}}^Q\right\} = \underset{\{\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q\}}{\arg\min} \sum_{q=1}^{Q} \alpha_q^2 \left\|\mathbf{F}_q \boldsymbol{\rho}^q - \mathbf{d}^q\right\|_2^2 + \sum_{n=1}^{N} \sum_{\substack{m \in \Delta_n \\ m > n}} \Psi\left(\sqrt{\sum_{q=1}^{Q} \beta_q^2 \left|\rho_m^q - \rho_n^q\right|^2}\right), \tag{3.8}$$

where $\Psi : [0, \infty) \to [0, \infty)$ is an appropriate regularization function that is used to encourage

smooth reconstructions, and $\{\alpha_q\}_{q=1}^Q$ and $\{\beta_q\}_{q=1}^Q$ are user-selected real-valued positive weighting coefficients. We will assume for simplicity that $\Psi(t)$ is continuously differentiable. In addition, to have edge-preserving regularization (as discussed in the previous subsection), we will assume that $\Psi'(t)/(2t)$ is non-increasing and that $\lim_{t\to\infty} \Psi(t)/(t^2) = 0$ (i.e., $\Psi(t)$ never grows faster than a quadratic function, and grows significantly slower than a quadratic function as $t \to \infty$).

In order to have a line process interpretation, we will additionally impose that $\Psi(\sqrt{t})$ is concave on $(0, \infty)$ [235, 492], and will assume that the limits $T = \lim_{t\to\infty} \Psi'(t)/(2t)$ and $L = \lim_{t\to 0^+} \Psi'(t)/(2t)$ are well defined. In this case, it can be proven [26] that there exists a convex and decreasing function $J(\cdot)$ such that

$$\Psi(t) = \inf_{T\le\ell\le L} \left(\ell t^2 + J(\ell)\right). \tag{3.9}$$

In addition, for fixed t, the value of ℓ that optimizes Eq. (3.9) is $\ell = \Psi'(t)/(2t)$ [26].

As a result of this relationship, Eq. (3.8) has the line process interpretation

$$\left\{\hat{\boldsymbol{\rho}}^1, \hat{\boldsymbol{\rho}}^2, \ldots, \hat{\boldsymbol{\rho}}^Q\right\} = \underset{\{\boldsymbol{\rho}^1,\boldsymbol{\rho}^2,\ldots,\boldsymbol{\rho}^Q\}}{\arg\min} \inf_{\boldsymbol{\ell}} \mathcal{C}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q, \boldsymbol{\ell}), \tag{3.10}$$

where

$$\begin{aligned}\mathcal{C}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q, \boldsymbol{\ell}) = \sum_{q=1}^{Q} \alpha_q^2 \left\|\mathbf{F}_q\boldsymbol{\rho}^q - \mathbf{d}^q\right\|_2^2 + \sum_{q=1}^{Q}\sum_{n=1}^{N}\sum_{\substack{m\in\Delta_n\\ m>n}} \beta_q^2 \ell_{n,m} \left|\rho_m^q - \rho_n^q\right|^2 \\ + \sum_{n=1}^{N}\sum_{\substack{m\in\Delta_n\\ m>n}} J(\ell_{n,m}).\end{aligned} \tag{3.11}$$

Note that the line process variables in this expression are shared among all Q images. Thus, this formulation has the ability to capture the correlated nature of the edge structures in the image sequence. If $\ell_{n,m}$ is small (indicating an edge could be present), then the reconstruction will avoid smoothing any of the images significantly across that possible edge. On the other hand, if $\ell_{n,m}$ is large, then smoothness will be imposed at that location for every image in the sequence.

Due to the line process interpretation, optimization of Eq. (3.8) can be performed efficiently using the following optimization procedure (this procedure is frequently called a multiplicative half-quadratic optimization algorithm [15, 131, 163, 235, 472], though it also has a gradient linearization interpretation, a quasi-Newton interpretation, a fixed-point method interpretation, and a majorize-minimize interpretation [207, 471, 625]):

1. Set iteration number $j = 0$, and set the estimated image sequence equal to an initial guess $\left\{\hat{\boldsymbol{\rho}}^1_{(j)}, \hat{\boldsymbol{\rho}}^2_{(j)}, \ldots, \hat{\boldsymbol{\rho}}^Q_{(j)}\right\}$ (e.g., set all voxel coefficient values based on an initial noisy Fourier reconstruction).

2. At the jth iteration, define line process variables for each voxel pair as

$$\boldsymbol{\ell}^{(j)} = \arg\inf_{\boldsymbol{\ell}} \mathcal{C}(\hat{\boldsymbol{\rho}}^1_{(j)}, \hat{\boldsymbol{\rho}}^2_{(j)}, \ldots, \hat{\boldsymbol{\rho}}^Q_{(j)}, \boldsymbol{\ell}), \tag{3.12}$$

   i.e., set

$$\ell^{(j)}_{n,m} = \frac{\Psi'(t^{(j)}_{n,m})}{2t^{(j)}_{n,m}} \tag{3.13}$$

   where $t^{(j)}_{n,m}$ is given by

$$t^{(j)}_{n,m} = \sqrt{\sum_{q=1}^{Q} \beta_q^2 \left|\hat{\rho}^q_{m,(j)} - \hat{\rho}^q_{n,(j)}\right|^2}, \tag{3.14}$$

   and $\rho^q_{m,(j)}$ denotes the mth element of $\hat{\boldsymbol{\rho}}^q_{(j)}$. Note that this step is an explicit line process calculation, and that the value of the line process depends on a weighted average of the edge strengths in all Q images, where the weighting is defined through the user-selected β_q parameters.

3. Update the image sequence according to

$$\left\{\hat{\boldsymbol{\rho}}^1_{(j+1)}, \hat{\boldsymbol{\rho}}^2_{(j+1)}, \ldots, \hat{\boldsymbol{\rho}}^Q_{(j+1)}\right\} = \underset{\{\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q\}}{\arg\min}\ \mathcal{C}(\hat{\boldsymbol{\rho}}^1 \hat{\boldsymbol{\rho}}^2, \ldots, \hat{\boldsymbol{\rho}}^Q, \boldsymbol{\ell}^{(j)}). \tag{3.15}$$

This optimization problem is separable, such that the solution for each $\hat{\boldsymbol{\rho}}^q_{(j+1)}$ is given by

$$\begin{aligned}
\hat{\boldsymbol{\rho}}^q_{(j+1)} &= \arg\min_{\boldsymbol{\rho}^q} \|\mathbf{F}_q\boldsymbol{\rho}^q - \mathbf{d}^q\|^2_{\ell_2} + \sum_{n=1}^{N}\sum_{\substack{m\in\Delta_n \\ m>n}} \frac{\beta_q^2}{\alpha_q^2}\ell^{(j)}_{n,m}\left|\rho^q_m - \rho^q_n\right|^2 \\
&= \arg\min_{\boldsymbol{\rho}^q} \|\mathbf{F}_q\boldsymbol{\rho}^q - \mathbf{d}^q\|^2_{\ell_2} + \sum_{n=1}^{N}\sum_{\substack{m\in\Delta_n \\ m>n}} \frac{\beta_q^2}{\alpha_q^2}\ell^{(j)}_{n,m}\left|\rho^q_m - \rho^q_n\right|^2 \\
&= \arg\min_{\boldsymbol{\rho}^q} \|\mathbf{F}_q\boldsymbol{\rho}^q - \mathbf{d}^q\|^2_{\ell_2} + \frac{\beta_q^2}{\alpha_q^2}\left\|\mathrm{diag}(\sqrt{\ell^{(j)}_{n,m}})\mathbf{D}\boldsymbol{\rho}^q\right\|^2_{\ell_2} \\
&= \left(\mathbf{F}_q^H\mathbf{F}_q + \frac{\beta_q^2}{\alpha_q^2}\mathbf{D}^H\mathrm{diag}(\ell^{(j)}_{n,m})\mathbf{D}\right)^{-1}\mathbf{F}_q^H\mathbf{d}^q,
\end{aligned} \tag{3.16}$$

where $\mathbf{D}$ is the matrix that computes finite differences between neighboring voxels, the diagonal matrices $\mathrm{diag}\left(\ell^{(j)}_{n,m}\right)$ and $\mathrm{diag}(\sqrt{\ell^{(j)}_{n,m}})$ have diagonal elements equal to the different components of $\ell^{(j)}_{n,m}$ and $\sqrt{\ell^{(j)}_{n,m}}$ (arranged in the same order as the finite-differences in $\mathbf{D}$), respectively, and we have assumed that the matrix being inverted has full rank (which necessarily implies that the nullspace of $\mathbf{D}$ has trivial intersection with the nullspace of $\mathbf{F}_q$ for each q). This full rank condition will generally be satisfied if $\frac{\beta_q^2}{\alpha_q^2}\ell^{(j)}_{n,m}$ is always a positive number at every spatial location in every image, and if the center of *k*-space is always sampled. Alternatively, the condition that $\frac{\beta_q^2}{\alpha_q^2}\ell^{(j)}_{n,m}$ is positive is not necessary when $\mathbf{F}_q$ has a trivial nullspace.

On a practical note, the matrices $\mathbf{F}_q$ and $\mathbf{D}$ are often very large, and working with them directly would require large amounts of memory and processing time. However, iterative optimization algorithms which only require computation of multiplications with these matrices can be done efficiently; $\mathbf{F}$ is often related to the discrete Fourier transform due to the form of Eq. (2.32), and $\mathbf{D}$ is sparse. Thus, $\hat{\boldsymbol{\rho}}^q_{(j+1)}$ can be determined efficiently using iterative algorithms like the conjugate gradient (CG) method [305] or LSQR [491]. Explicitly, each iteration will require multiplication of vectors with $\mathbf{F}_q^H\mathbf{F}_q$ and $\mathbf{D}^H\mathrm{diag}(\ell^{(j)}_{n,m})\mathbf{D}$. When the acquired data lies on a Cartesian grid matched to the size of the field of view, then

multiplications by $\mathbf{F}_q$ and $\mathbf{F}_q^H$ can be performed efficiently using the FFT. Even when the acquired data is not Cartesian, multiplication by $\mathbf{F}_q^H \mathbf{F}_q$ can be performed efficiently using a simple convolution [212, 627] that can be implemented efficiently using the FFT algorithm. Multiplications with $\mathbf{D}$ and $\mathbf{D}^H$ are also computationally simple due to sparsity.

The Q different optimization problems in Eq. (3.16) are completely independent from each other, and thus can be solved in parallel if appropriate parallel computation hardware is available. However, if the matrices being inverted happen to be the same for all q (which will occur if $\mathbf{F}_q$ and $\frac{\beta_q^2}{\alpha_q^2}$ are the same for all q), then use of an iterative algorithm that takes advantage of this shared structure (e.g., the algorithm described in Ref. [129]) might be more computationally efficient. In addition, we have previously shown that the computation of each individual matrix inversion in Eq. (3.16) can also be significantly accelerated by using parallel processing on graphics processing units [593, 661].

4. Increment j. Repeat steps 2 and 3 until convergence is achieved.

By construction, the value of the cost function in Eq. (3.10) is bounded below by zero and is non-increasing as the iterations proceed. As a result, the value of the cost function is guaranteed to converge. In addition, stronger assumptions on $\Psi(\cdot)$ can ensure global convergence of the iterates to a global minimum. For example, Delaney and Bresler [163] have proven that the following additional conditions will ensure global convergence of this half-quadratic algorithm to a unique optimal solution:

1. $\Psi(t)$ is strictly convex.

2. The nullspace of $\frac{\beta_q}{\alpha_q}\mathbf{D}$ has trivial intersection with the nullspace of $\mathbf{F}_q$ for each q.

3. $\theta(t) = \Psi(\sqrt{t})$ is a twice continuously differentiable strictly concave function, with $\theta(0) = 0$, $\theta'(0) = 1$, and $0 < \theta'(t) \leq 1$ for all $t \in [0, \infty)$.

Other conditions that can ensure good convergence properties for this algorithm can be found in Refs. [15, 131, 472].

3.2.1 Modes of Operation

Different modes of operation can be achieved through the choice of the $\{\alpha_q\}_{q=1}^{Q}$ and $\{\beta_q\}_{q=1}^{Q}$ parameters. The value of β_q determines the influence of the qth image on the estimated shared line-process variables. Large values of β_q imply that the spatial finite-differences computed from the qth image should have a significant role in the estimation of the shared line-process variables, while small values of β_q imply the opposite. In addition, the value of $\frac{\beta_q^2}{\alpha_q^2}$ will determine the trade-off between spatial resolution and image SNR in the reconstructed images, and α_q should be selected to optimize this trade-off for the particular application context. Large values of α_q can be used to impose that the qth reconstructed image should be highly data consistent (i.e., high resolution with limited SNR improvement), while smaller values of α_q permit more significant data inconsistency for improved denoising. In general, the selection of $\{\alpha_q\}_{q=1}^{Q}$ and $\{\beta_q\}_{q=1}^{Q}$ will be different based on whether the proposed method is being used for super-resolution or for denoising.

For the super-resolution context, we note that it is generally very difficult to accurately infer high-resolution edge structure from low-resolution data unless very strong additional assumptions are made.[4] As a result, when the acquisition includes images with very different resolution characteristics, it is better to rely on the high-resolution images to estimate the edge structures, and β_q can be set very small for the low resolution images. In the super-resolution context, it is often the case that the high-resolution "reference" images also have relatively high SNR. In this case, α_q can be set very high for the high resolution images, to emphasize data fidelity for these high-quality images. In the limiting case where $\frac{\beta_q^2}{\alpha_q^2}$ approaches zero for the high-resolution images, and β_q approaches zero for the low-resolution images (with appropriately scaled α_q such that $\frac{\beta_q^2}{\alpha_q^2}$ is non-zero), then the iterative algorithm described in the previous section will converge after a single iteration if the $\mathbf{F}_q$ matrices for the high-resolution images have trivial nullspaces. For this scenario, reconstruction of the low-resolution images is perfectly linear with respect to the corresponding

[4] Note that in many cases, attempting to infer high-resolution edge structure from low-resolution data can lead to image reconstructions possessing sharp edge features, but with significantly distorted image geometry compared to the true high-resolution image. This issue is discussed briefly in Ref. [274], in the context of a special type of reference-based super-resolution reconstruction, where line processes are estimated both from the reference images and from low-resolution noisy data.

low-resolution data.

For the denoising context, we will assume that the Q different datasets all have similar resolution. In this case, it is useful to use every image to infer the shared edge structure, meaning that β_q should be greater than zero for all q. Setting β_q to be exactly the same for all images would be one simple approach. However, this simple approach could be problematic in some contexts, since it will give unequal weight to images which are scaled differently from one another. We generally choose β_q inversely proportional to the average magnitude of the signal within the region of interest of the qth image, though it could also be adjusted manually if other prior information is available on which images should play more or less significant roles in determining the joint edge structure. The choice of α_q in the denoising context can depend on the experimental objectives. From a PML point of view, it would be optimal to set $\alpha_q = 1/\sigma_q$, since the data fidelity criterion in this case would exactly equal the negative log-likelihood of the measured data (neglecting irrelevant additive constants). In this case, the amount of smoothing applied to each image will depend on the amount of noise present in each image. However, when $\mathbf{F}_q$ is the same for all images, it can also be useful to choose α_q such that $\frac{\beta_q^2}{\alpha_q^2}$ is the same for all images. In this case, the reconstruction matrix from Eq. (3.16) will be identical for all images, leading to uniform reconstruction characteristics between the different images from the sequence.

3.3 Characteristics

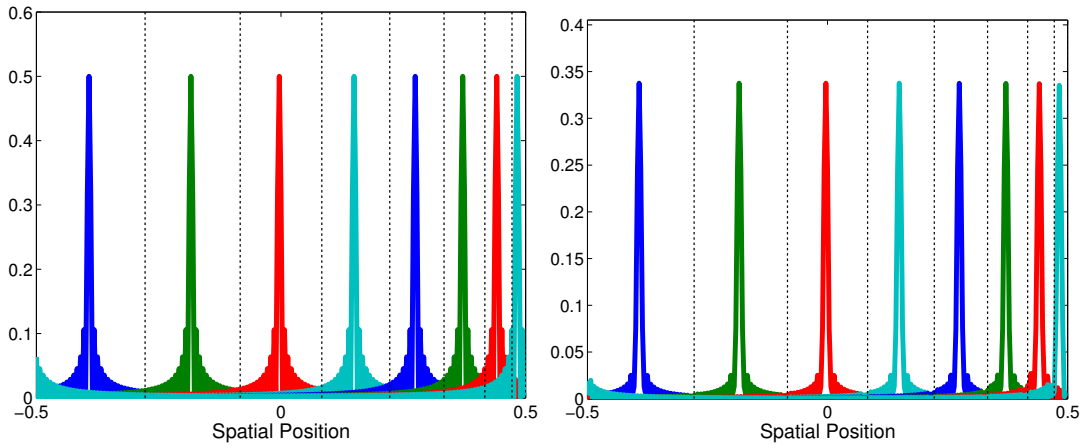
The proposed method can reduce image noise and Gibbs ringing artifacts though the use of spatial smoothing, and is applicable to both super-resolution and denoising contexts. However, as a result of spatial smoothing, it is intuitive that there could also naturally be some loss in spatial resolution relative to traditional Fourier reconstruction. In contrast to most constrained reconstruction methods, the loss of spatial resolution and the corresponding gain in SNR are easy to analyze with the proposed method. These issues, as well as other important characteristics of the proposed algorithm, are discussed in the next subsections.

3.3.1 Resolution and Noise Characteristics of the Proposed Method

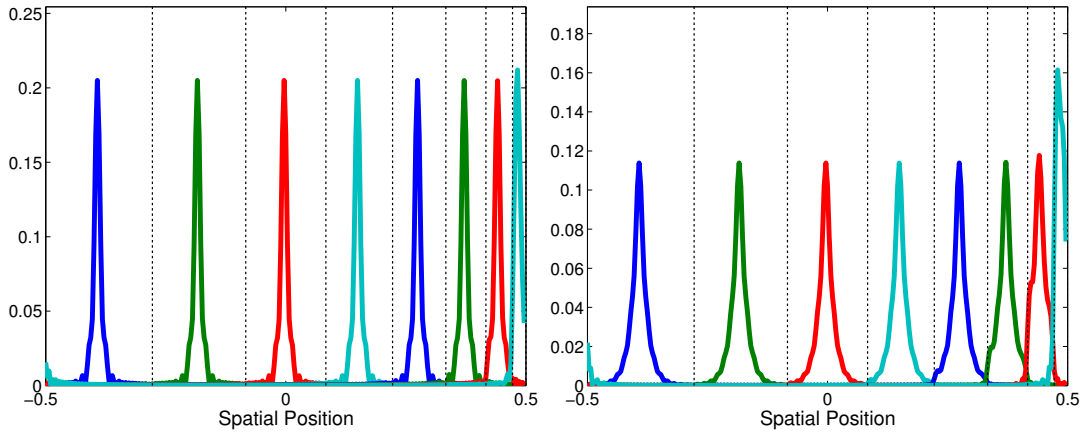
As illustrated by Eq. (3.16), for a given estimated edge map, the estimated image $\hat{\boldsymbol{\rho}}^q$ can be viewed as a linear transform of the measured data $\mathbf{d}^q$, even though the reconstruction procedure itself is nonlinear. This linear perspective provides an analytic way to analyze the resolution of the reconstructed image using the SRF as in Eq. (2.43). Analysis using the SRF is generally more powerful than linearized local perturbation-based resolution analyses that have been developed for non-linear reconstruction in emission tomography [8, 213, 524, 588, 631], since the SRF precisely describes how a reconstructed voxel relates to the original true image. In addition, due to the use of a shared edge-map, for all images with the same $\mathbf{F}_q$ and the same value of $\frac{\beta_q^2}{\alpha_q^2}$, the SRFs for each reconstructed voxel will be identical over the different images. This latter point is particularly important for quantitative studies, since the validity of any parametric voxel-by-voxel model-fitting procedure would be compromised if the same voxel in different images does not correspond to the same spatial spin population.

Linear noise analysis, as described previously in Eq. (2.46), is also possible if it is assumed that the estimated line processes $\boldsymbol{\ell}$ are not a function of the noise, and can be treated as deterministic variables rather than random variables. The ability to analytically characterize both resolution and noise is very powerful, and provides ways to understand the performance characteristics of the reconstruction algorithm that typically are not available for general non-linear denoising and super-resolution approaches. In addition, it becomes possible to choose $\frac{\beta_q^2}{\alpha_q^2}$ based on the desired trade-off between resolution and SNR in each image.

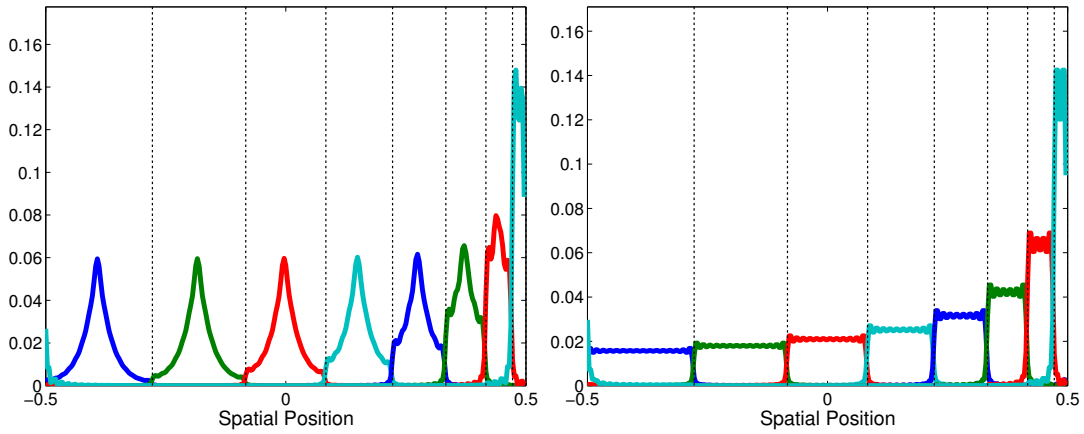
To understand the characteristics of reconstructions using Eq. (3.8), we first start by analyzing the performance characteristics of weighted smoothness reconstructions as in Eq. (3.16), assuming that $\boldsymbol{\ell}$ is deterministic. As a first step, consider the SRFs shown in Fig. 3.5, which are formed from regularized reconstruction of 1D Cartesian Fourier data sampled uniformly at the Nyquist rate. The dashed vertical lines correspond to edge locations, at which $\ell_{n,m}$ has been set equal to 0. At all other spatial locations, $\ell_{n,m}$ has been set equal to 1. The different subfigures correspond to



(a) $\gamma = 1$ (b) $\gamma = 2$



(c) $\gamma = 4$ (d) $\gamma = 8$



(e) $\gamma = 16$ (f) $\gamma = 32$

Figure 3.5: SRFs calculated at several different spatial locations in the image, using a weighted smoothness-regularized reconstruction with binary edge weights. The dashed lines indicate spatial locations where $\ell_{n,m} = 0$. The different subfigures correspond to different amounts of regularization, with increasing regularization corresponding to increasing γ. The $\gamma = 1$ case corresponds to standard conjugate phase reconstruction.

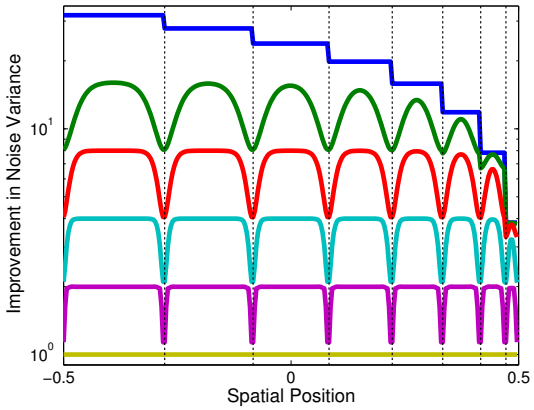


Figure 3.6: Improvement in SNR as a function of spatial location corresponding to the same binary-edge regularized reconstructions considered in Fig. 3.5. The six different curves correspond, from bottom to top, to $\gamma = 1$, 2, 4, 8, 16, and 32.

different values of $\frac{\beta_q^2}{\alpha_q^2}$, which have been adjusted to achieve different levels of SNR improvement in the image. We use the symbol γ to denote the *equivalent number of averages*, which we define as the number of averages that would have been necessary to acquire to achieve a similar level of SNR improvement in smooth image regions. Notably, with little regularization, the SRFs are not significantly modified. However, as regularization is increased, the SRFs become broader (i.e., resolution is degraded), and the SRFs are largely shift-invariant unless they are sufficiently close to edge structures. However, the shapes of the SRFs begin to adapt as they encounter edge locations, and this adaptation leads to small leakage of signal across the imposed edge locations. In the limit as the regularization grows very large, the reconstructed image will become a piecewise constant image, with the value of each piece of the image approximately equal to the average value of the image within each image compartment, and with the different compartments separated by the boundary locations.

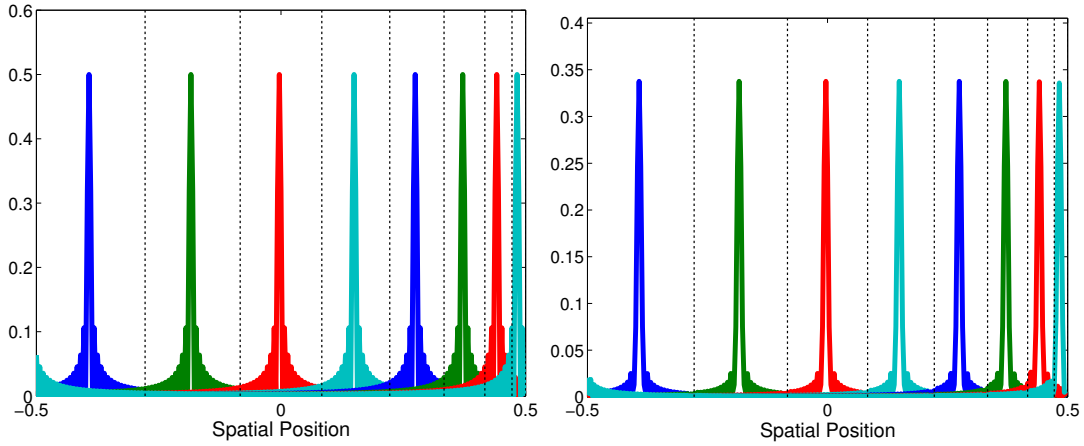
The corresponding noise plot is shown in Fig. 3.6. As can be seen, noise improvement is fairly uniform, except near edge locations where the reduction in noise variance is not as significant as it was in regions that were far from estimated edge locations. This result is not surprising, since less smoothing (i.e., less spatial averaging) is applied in these regions to avoid signal leakage across the known edge. Another notable feature is that the achievable improvement in SNR is limited by

the size of the image feature. This observation is consistent with theoretical characterizations of highly-constrained compartmental image models [321, 404], for which the SNR is also known to be highly dependent on the size of the compartment. Looking at both Figs. 3.5 and 3.6, we observe that increasing regularization has the effect of improving SNR at the expense of spatial resolution, as expected. However, the use of edge weights can help to ensure that partial volume artifacts are minimal.

It is also insightful to look at the reconstructions that are obtained when $\ell_{n,m} = 0.1$ instead of 0 at edge locations, since the reconstruction behavior changes considerably for non-zero edge weights. The resolution and noise characteristics in this case are shown in Figs. 3.7 and 3.8, respectively. Unlike in the previous case, a small amount of leakage is allowed across edge locations if the regularization is strong enough. However, use of non-zero $\ell_{n,m}$ allows the improvement in noise variance to be more spatially uniform.

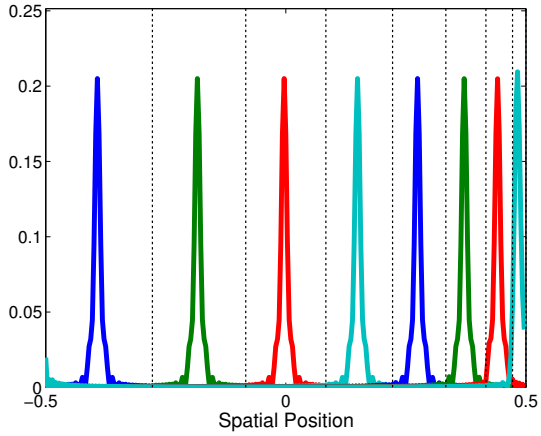
In the previous examples, we demonstrated the behavior of weighted-smoothness regularization when edge weights were spatially separated from each other. Different behavior is observed when contiguous regions of the image have small $\ell_{n,m}$ values. This is illustrated in Figs. 3.9 and 3.10 for the case when edge locations receive a weight of $\ell_{n,m} = 0$. While the SRFs and noise variances behave similarly to those in Figs. 3.5 and 3.6 outside of the spatially-contiguous "edge region," the resolution and noise properties are extremely poor inside this region. Figures 3.11 and 3.12 show the behavior when edge locations receive a weight of $\ell_{n,m} = 0.1$, and the situation is much better in this case. In particular, we still observe a loss of resolution and improvement in SNR in the "edge region," though the changes are not as significant as those observed in the rest of the reconstruction.

In the scenarios we have considered so far, the Fourier data resolution was higher than that of the smallest separation between edge locations in the reconstructed image. In order to address the issue of super-resolution reconstruction, it is also of interest to examine the case where the features of interest are small with respect to the resolution of the data. One such case is shown in Fig. 3.13. As can be seen, when the data resolution is lower than that of the edge information, the

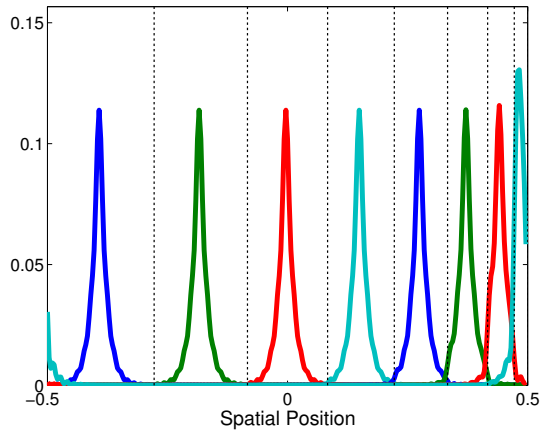


(a) $\gamma = 1$

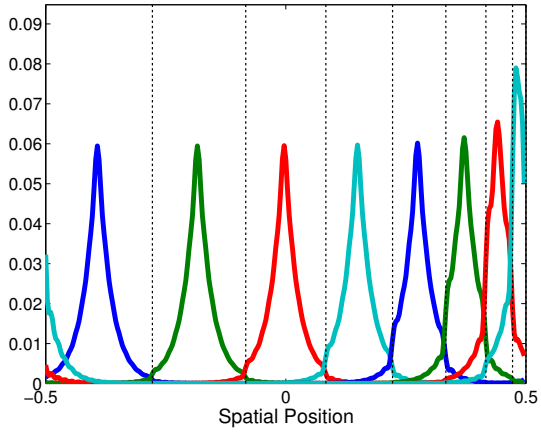
(b) $\gamma = 2$



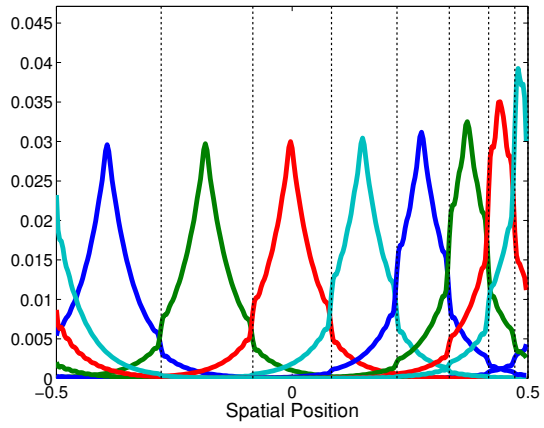
(c) $\gamma = 4$



(d) $\gamma = 8$



(e) $\gamma = 16$



(f) $\gamma = 32$

Figure 3.7: SRFs calculated at several different spatial locations in the image, using a weighted smoothness-regularized reconstruction with non-binary edge weights. The dashed lines indicate spatial locations where $\ell_{n,m} = 0.1$. The different subfigures correspond to different amounts of regularization, with increasing regularization corresponding to increasing γ. The $\gamma = 1$ case corresponds to standard conjugate phase reconstruction.

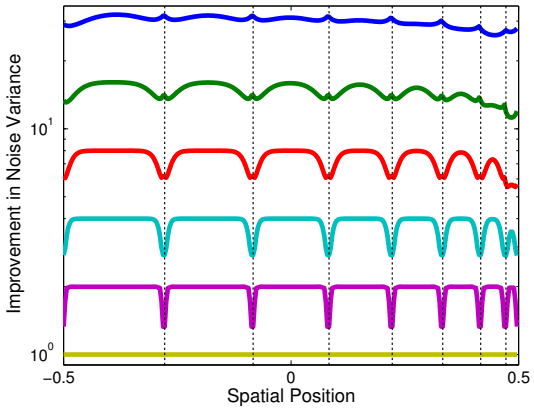
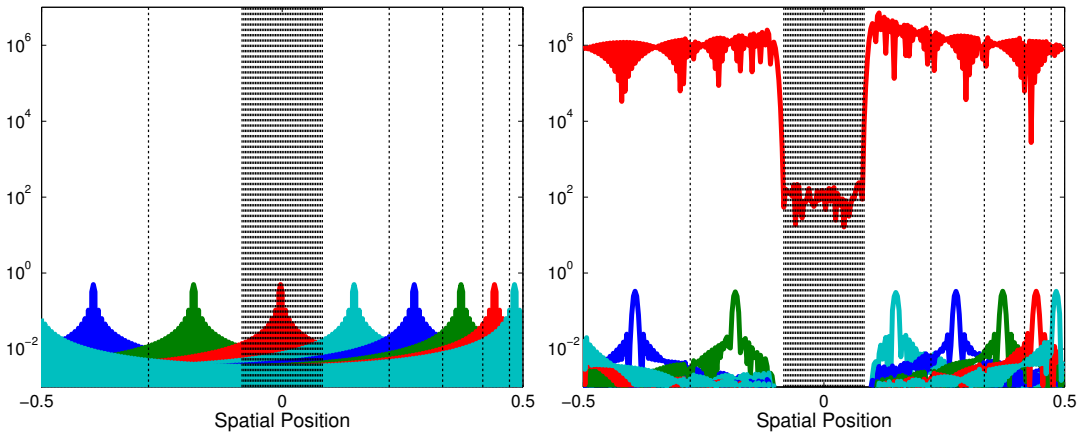


Figure 3.8: Improvement in SNR as a function of spatial location corresponding to the same regularized reconstructions with non-binary edge weights as considered in Fig. 3.7. The six different curves correspond, from bottom to top, to $\gamma = 1$, 2, 4, 8, 16, and 32.

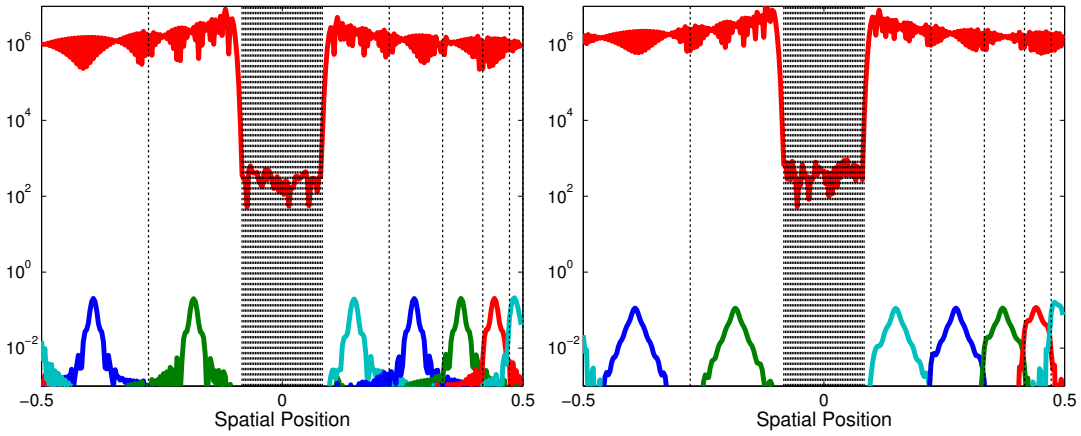
edge constraints have limited ability to suppress signal leakage. In addition, the small compartments actually have noise amplification rather than noise reduction. The implication of this is that unless the prior edge constraints are very accurate and can be imposed in a very strong way, use of the proposed technique for the reconstruction of sub-resolution features is not recommended. Additionally, this result also implies that smoothness-based SNR improvement should only be attempted when the spatial image regions of interest are large with respect to the resolution of the data.

Finally, while the results we have shown so far are all 1D examples, the results extend naturally to 2D. This is illustrated with a simple example in Figs. 3.14 and 3.15. These figures show that, as expected based on the previous 1D results, the SRFs for voxels far from edge locations demonstrate the standard trade-off between resolution and SNR, while SRFs for voxels near edge locations adapt to avoid signal leakage across edge locations. Notably, the SRFs from regularized reconstruction of 32×32 Fourier data can have both smaller noise variance and smaller full-width at half-maximum (FWHM) than the SRFs from the MNLS reconstructions from 16×16 data, even if the 16×16 data is averaged four times. The issue of resolution and SNR trade-offs in data acquisition is significant, and will be discussed in more detail in the next subsection.

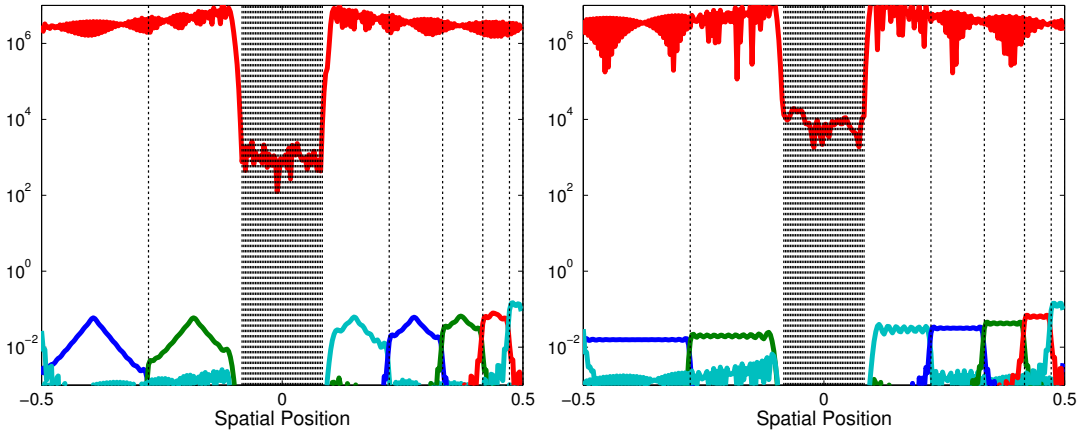
To summarize the results of this subsection:



(a) $\gamma = 1$ (b) $\gamma = 2$



(c) $\gamma = 4$ (d) $\gamma = 8$



(e) $\gamma = 16$ (f) $\gamma = 32$

Figure 3.9: SRFs calculated at several different spatial locations in the image, using a weighted smoothness-regularized reconstruction with binary edge weights, and a spatially-contiguous "edge" region. The dashed lines indicate spatial locations where $\ell_{n,m} = 0$. The different subfigures correspond to different amounts of regularization, with increasing regularization corresponding to increasing γ. The $\gamma = 1$ case corresponds to standard conjugate phase reconstruction.

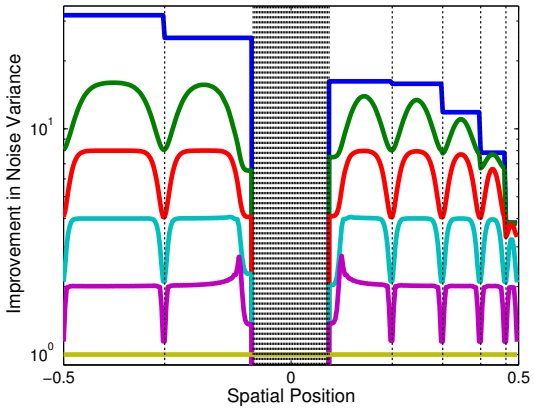


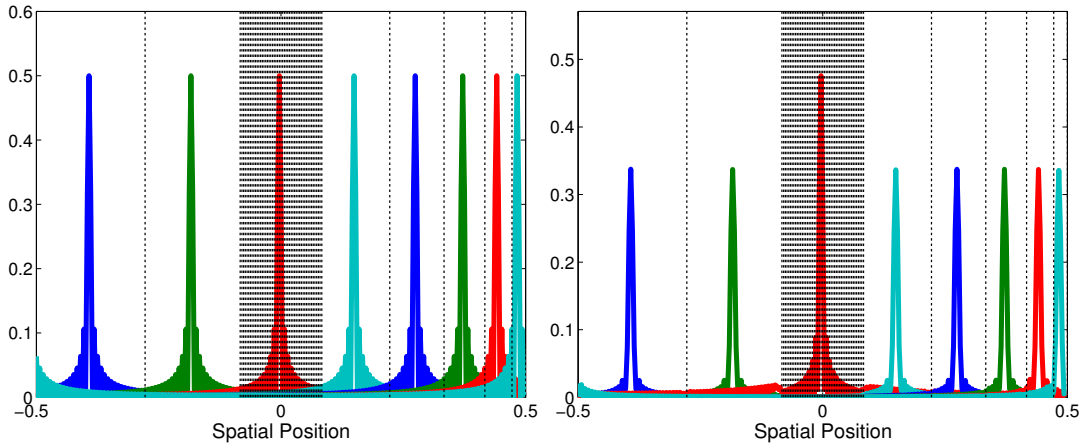
Figure 3.10: Improvement in SNR as a function of spatial location corresponding to the same binary-edge regularized reconstructions (with a spatially-contiguous "edge" region) considered in Fig. 3.9. The six different curves correspond, from bottom to top, to $\gamma = 1$, 2, 4, 8, 16, and 32.

- Weighted smoothness priors provide a mechanism for improving image SNR through spatial smoothing, while preventing signal leakage and partial volume artifacts across known edges.

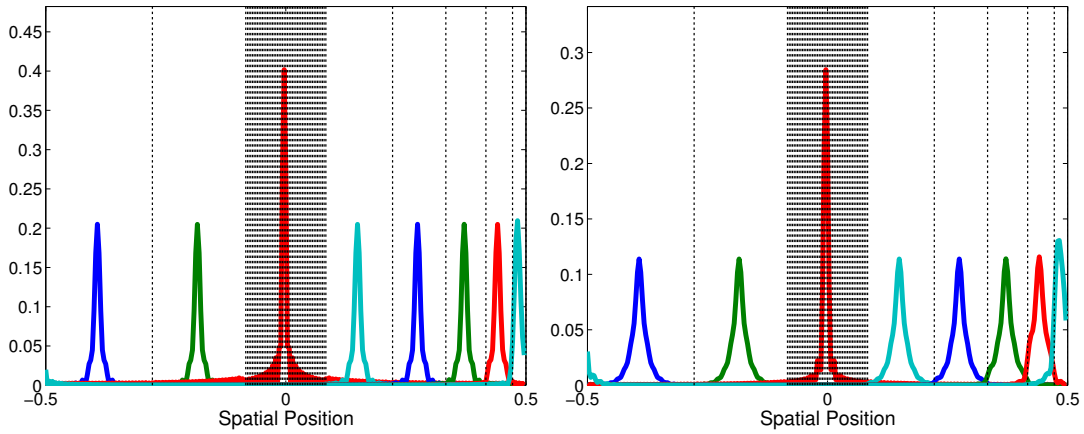
- Use of weighted smoothness priors is most effective when the resolution of the data is higher than that of the smallest image feature to be preserved. The proposed regularization scheme has limited capabilities for avoiding signal leakage with very low resolution data, unless the smoothness model can be imposed very strictly [319, 321, 404]. Even when the model can be imposed strictly, there is generally an amplification of noise rather than a reduction of noise.

- Due to conditioning problems, it is generally not useful to have large contiguous regions of the image with $\ell_{n,m} = 0$.

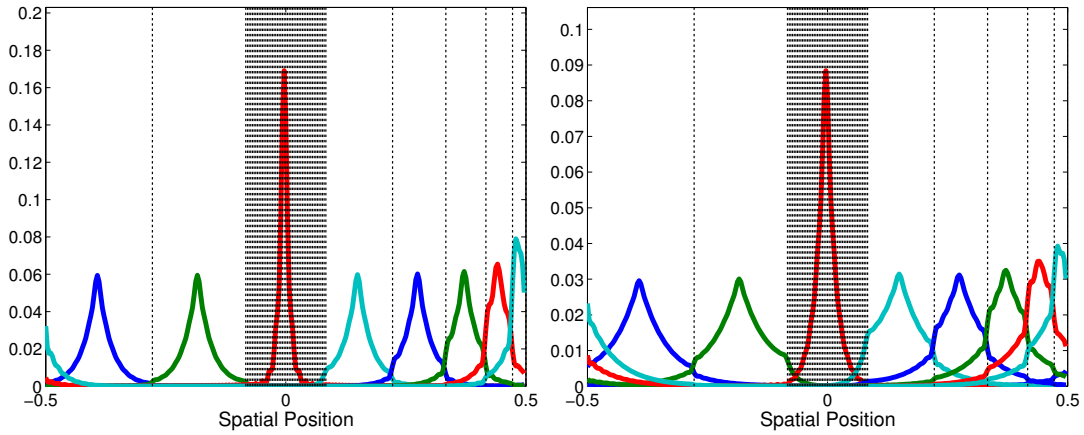
- Spatial smoothing of high-resolution data can lead to reconstructions with both higher resolution and higher SNR than standard reconstruction of low-resolution data (averaged several times for equivalent acquisition time).



(a) $\gamma = 1$ (b) $\gamma = 2$



(c) $\gamma = 4$ (d) $\gamma = 8$



(e) $\gamma = 16$ (f) $\gamma = 32$

Figure 3.11: SRFs calculated at several different spatial locations in the image, using a weighted smoothness-regularized reconstruction with non-binary edge weights, and a spatially-contiguous "edge" region. The dashed lines indicate spatial locations where $\ell_{n,m} = 0.1$. The different subfigures correspond to different amounts of regularization, with increasing regularization corresponding to increasing γ. The $\gamma = 1$ case corresponds to standard conjugate phase reconstruction.

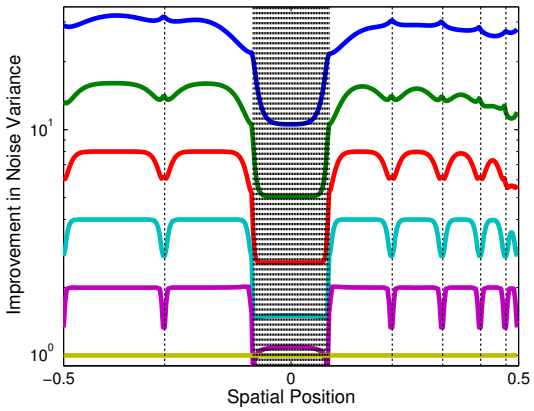
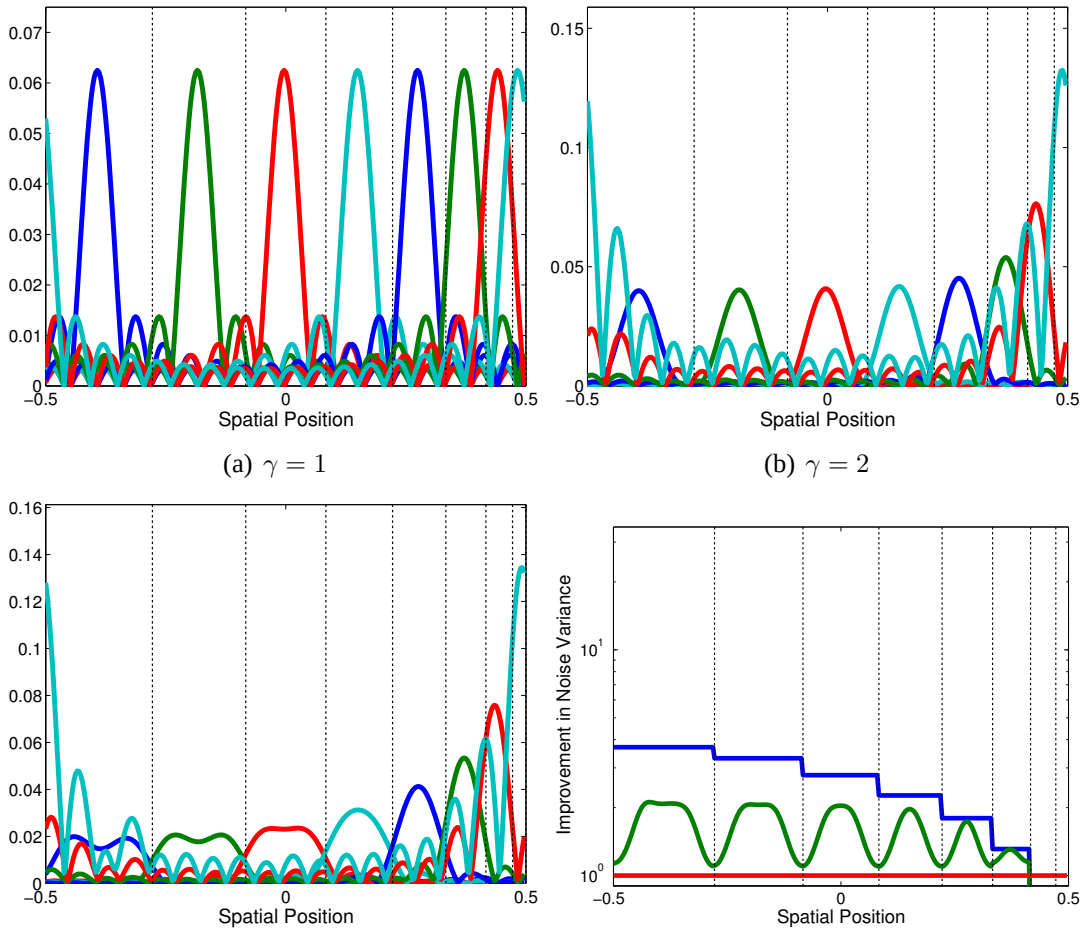


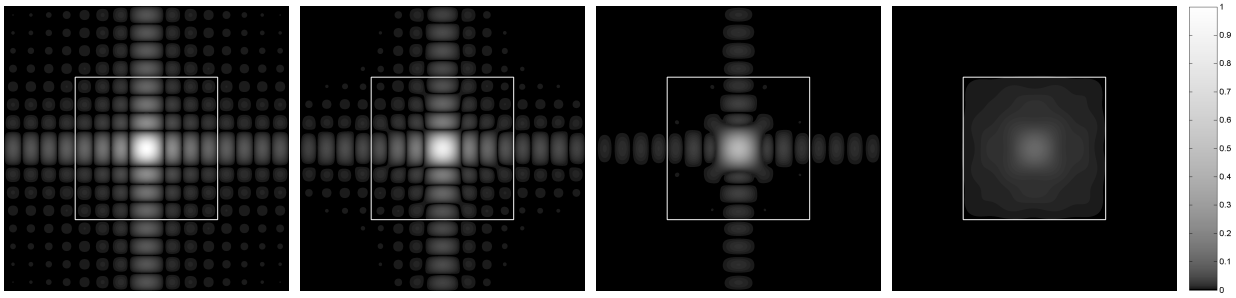
Figure 3.12: Improvement in SNR as a function of spatial location corresponding to the same regularized reconstructions with non-binary edge weights (and a spatially-contiguous "edge" region) as considered in Fig. 3.11. The six different curves correspond, from bottom to top, to $\gamma = 1$, 2, 4, 8, 16, and 32.



(a) $\gamma = 1$ (b) $\gamma = 2$

(c) $\gamma = 4$ (d) Noise Characteristics

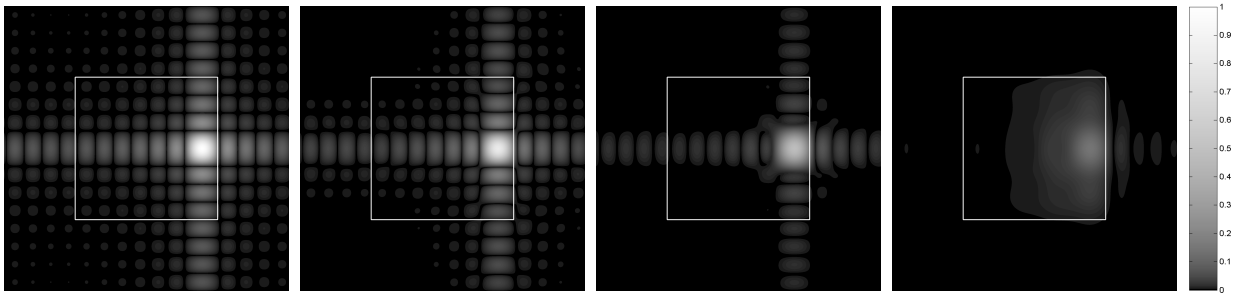
Figure 3.13: (a)-(c) SRFs calculated at several different spatial locations in the image, using a weighted smoothness-regularized reconstruction with binary edge weights, with low-resolution Fourier data. Note that signal leakage is unavoidable for the very high-resolution signal compartments. (d) The noise variance only improves with regularization for image compartments that are larger than the native resolution of the data.



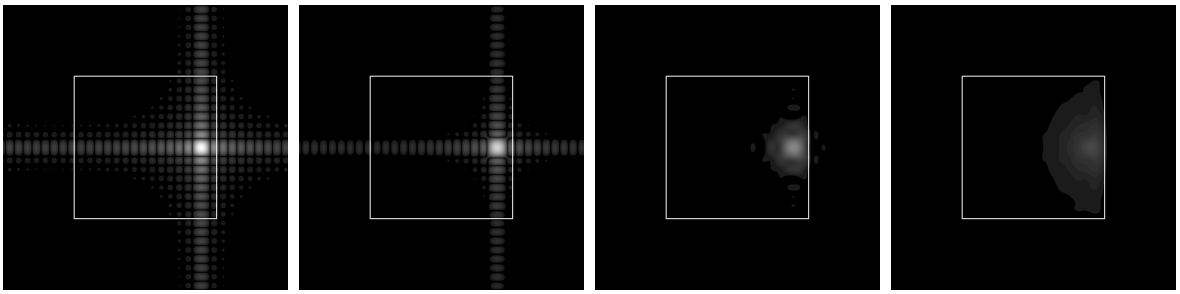
(a) MNLS (b) $\lambda = 0.0001$ (c) $\lambda = 0.001$ (d) $\lambda = 0.01$

(e) MNLS (f) $\lambda = 0.0001$ (g) $\lambda = 0.001$ (h) $\lambda = 0.01$

Figure 3.14: Regularized reconstruction SRFs for different values of $\lambda = \frac{\beta_q^2}{\alpha_q^2}$, using binary edge weights located along the white square. (a-d) 16×16 Cartesian sampling at the Nyquist rate. (e-h) 32×32 Cartesian sampling at the Nyquist rate. SRFs for voxels far from edge locations demonstrate the standard trade-off between resolution and SNR.



(a) MNLS (b) $\lambda = 0.0001$ (c) $\lambda = 0.001$ (d) $\lambda = 0.01$



(e) MNLS (f) $\lambda = 0.0001$ (g) $\lambda = 0.001$ (h) $\lambda = 0.01$

Figure 3.15: Regularized reconstruction SRFs for different values of $\lambda = \frac{\beta_q^2}{\alpha_q^2}$, using binary edge weights located along the white square. (a-d) 16×16 Cartesian sampling at the Nyquist rate. (e-h) 32×32 Cartesian sampling at the Nyquist rate. SRFs for voxels near edge locations adapt to avoid signal leakage across edge locations.

3.3.2 The Trade-off Between Resolution and SNR

In the previous subsection, we demonstrated that the proposed regularized reconstruction method gives us the flexibility to choose the trade-off between resolution and SNR of the reconstructed images. As a result, an important question is: how should data be acquired to optimize this trade-off? The optimal design of MRI sampling strategies has been considered for a long time, and there are two main strategies that have been advocated in the context of Nyquist-sampled Fourier data:

S1. Sample *k*-space over a region corresponding to the desired nominal resolution, and reconstruct images using basic inverse Fourier-transform methods. Use additional time for averaging.

S2. Sample *k*-space using the same total number of measurements as in S1, but cover a larger region of *k*-space (i.e., encode higher-resolution information) at the expense of reduced signal averaging. Reconstruct images using methods that reduce noise levels, necessarily reducing image resolution in the process.

These strategies have been compared in previous work [99, 194, 497], and the general consensus has been that S1 is more efficient than S2. Edelstein *et al.* [194] and Parker and Gullberg [497] demonstrate that, for fixed data acquisition time, simply averaging adjacent voxels within a high-resolution reconstructed image yields lower SNR than if data was directly collected at lower resolution. One limitation of these analyses is that they do not consider the effect of this voxel-averaging procedure on the SRF, which can provide both quantitative and qualitative measures of resolution. Buxton's analysis of the SNR/resolution trade-off [99] makes use of the *equivalent width* (EW) of the SRF to quantify image resolution (see [346] for additional discussion of this resolution metric), and shows that the trade-off between SNR and the EW is optimized when spatial smoothing is not applied. However, as will be described later in this subsection, there are many alternative measures of resolution based on the SRF, and the EW metric does not necessarily correspond as well with qualitative perceptual assessment compared to many of these other resolution metrics.

Despite these existing pessimistic characterizations of uniformly-sampled data with higher-than-nominal resolution, there are still indications in the literature that there can be benefits to strategy S2. For example, leveraging imaging physics, Li *et al.* [396] demonstrate that in MR spectroscopic imaging studies, higher-resolution acquisition followed by spatial smoothing can mitigate the loss in SNR associated with line-broadening due to intravoxel B_0 field inhomogeneity. Similarly, Triantafyllou *et al.* [604] demonstrate that this strategy can also reduce the effects of physiological noise in fMRI studies. Finally, as we have described in the previous subsection, there can be a theoretical advantage in SNR/resolution efficiency for oversampling and filtering of uniformly-sampled Fourier data with respect to the FWHM resolution metric in the context of our regularized reconstruction method.

To examine the trade-offs between resolution and SNR, we will focus on a simple abstraction of our proposed reconstruction scheme. In particular, we will focus on the special 1D case where $\mathbf{F}_q$ corresponds to the M low-frequency rows from an $N \times N$ 1D unitary discrete Fourier transform (DFT) matrix (N and M both even numbers), and where $\ell_{n,m} = 1$ at every spatial location. In addition, we will assume periodic (toroidal) boundary conditions for the smoothness prior. In this case, it can be shown that the solution to Eq. (3.16) has the form of the conjugate phase reconstruction in Eq. (2.23), with

$$w_m = \frac{1}{1 + 4\frac{\beta_q^2}{\alpha_q^2} \sin^2\left(\pi \frac{\mathbf{k}_m}{N}\right)}. \tag{3.17}$$

This result indicates that shift-invariant quadratic smoothness regularization can be equivalent to traditional windowed Fourier reconstruction.[5] For simplicity, and because the proposed reconstruction behaves like windowing in spatial regions that are far from edge locations, we will provide an analysis of the resolution and noise trade-offs for standard windowed Fourier reconstruction.

[5]Interested readers should note that Ref. [283] derives similar expressions that are applicable in the context of slightly more general regularization and data acquisition schemes.

Resolution Metrics

We described resolution properties of linear reconstructions in Ch. 2, though we did not specify any quantitative resolution metrics because there is no universally accepted way of quantifying resolution [165]. However, many heuristic approaches for measuring resolution based on the SRF have been proposed in the literature. In one-dimension, popular choices include:

- **Rayleigh Criterion (RC).** The RC [533] measures resolution by the distance from the maximum of the SRF to its first minimum. This definition was originated in the context of optical imaging systems; for MRI, we will use the distance from the maximum of the SRF to its first zero-crossing to define the RC, such that

$$h_0\left(\mathrm{RC}\right) = 0. \tag{3.18}$$

- **Sparrow Criterion (SC).** The SC [584] measures resolution by identifying the distance between two identical point sources at which the "valley" that forms between them is no longer visible. Mathematically, the SC is the smallest distance such that

$$\frac{\partial^2}{\partial x^2}\left[h_0\left(x+\frac{\mathrm{SC}}{2}\right)+h_0\left(x-\frac{\mathrm{SC}}{2}\right)\right]\Bigg|_{x=0} = 0. \tag{3.19}$$

- **Full-Width at Half-Maximum (FWHM).** . The FWHM [314] is one of the most common measures of resolution in imaging, and is defined as twice the largest distance from the center of the SRF to a spatial location where the magnitude of the SRF is at least half of its maximum value. For the smooth SRFs that can be obtained from finite Fourier data, the FWHM is defined as the largest distance within the FOV such that

$$\left|h_0\left(\frac{\mathrm{FWHM}}{2}\right)\right| = \frac{1}{2}. \tag{3.20}$$

- **Full-Width at Tenth-Maximum (FWTM).** Similar to the FWHM, the FWTM is defined as twice the largest distance from the center of the SRF to a spatial location where the magnitude of the SRF is at least one-tenth of its maximum value. With finite Fourier data, the FWTM is defined as the largest distance within the FOV such that

$$\left| h_0 \left(\frac{\text{FWTM}}{2} \right) \right| = \frac{1}{10}. \tag{3.21}$$

- **Standard Deviation Criterion (SD).** The SD [630] measures resolution using the second moment of the SRF. The original definition was developed for strictly positive SRFs. For this dissertation, the SD will be defined (assuming the SRF is centered at $x = 0$ such that the first moment is zero) as

$$\text{SD} = \sqrt{\int_{\text{FOV}} x^2 \left| h_0 \left(x \right) \right| dx}. \tag{3.22}$$

- **Equivalent Width (EW).** The EW [99, 346] is equal to the width of a rectangle that has the same integral and same maximum value as the SRF. In particular,

$$\text{EW} = \frac{\int_{\text{FOV}} h_0 \left(x \right) dx}{h_0 \left(0 \right)}. \tag{3.23}$$

For a fixed 1D *k*-space trajectory (with 256 Nyquist-rate samples that symmetrically cover the low-frequency region of *k*-space), the trade-off between spatial resolution and image noise variance is illustrated in Fig. 3.16 for strategy S2 with the apodization window described in Eq. (3.17) with $N = 256$, as well as five other common window functions common in signal processing applications [292]: the Gaussian window, the Kaiser-Bessel window, the Tukey window, the Hann window, and the Dolph-Chebyshev window. In addition to this, the figure also shows the trade-off between spatial resolution and image noise variance that would be obtained with strategy S1, where *k*-space coverage is reduced to provide additional time for data averaging. Notably, except

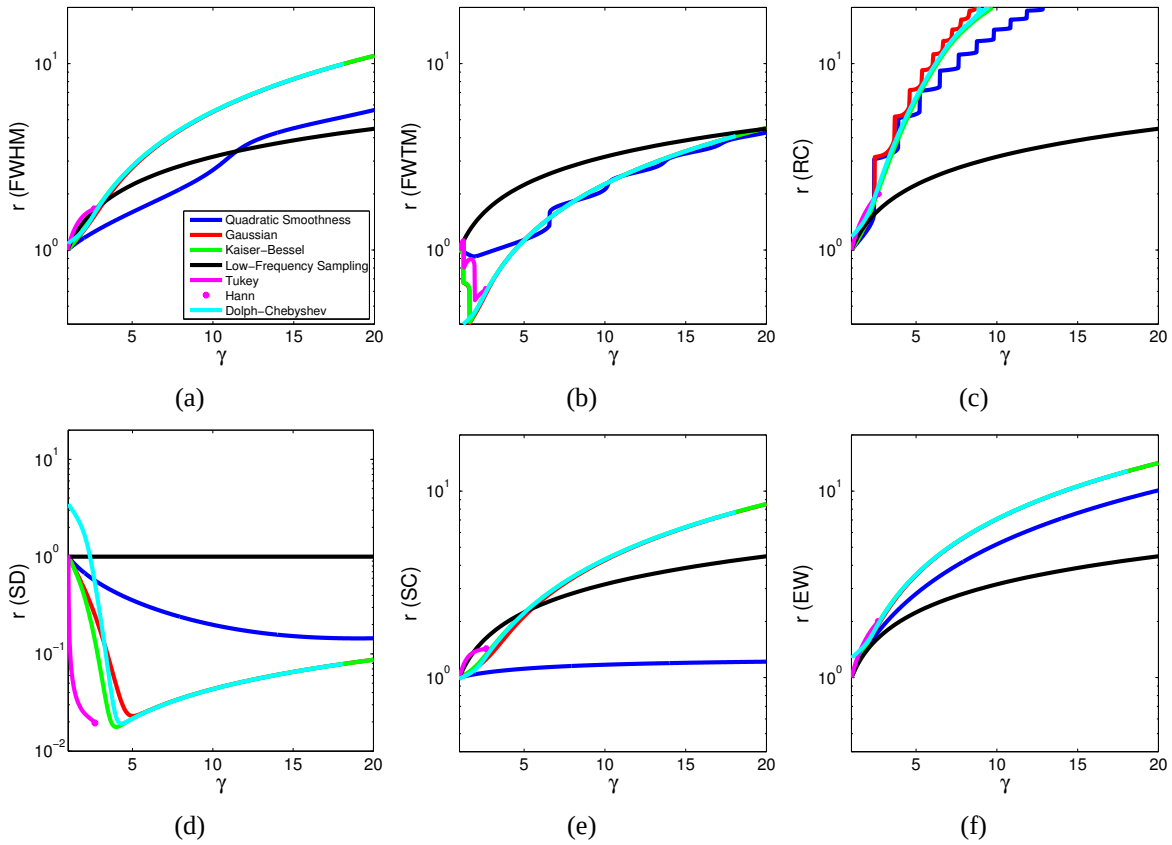


Figure 3.16: Resolution r versus the equivalent number of averages γ for several different apodization functions. The black line (Low-Frequency Sampling) does not correspond to apodization, but rather the standard Fourier reconstruction scheme where resolution can be traded for additional data acquisition time.

for the EW metric, most of the window functions have some operating points where they have higher resolution (with respect to the appropriate resolution metric) and lower noise variance than the standard low-frequency Fourier sampling scheme. This improvement in SNR efficiency can be observed more clearly in Fig. 3.17, where we have taken the curves from Fig. 3.16 and normalized them by the curve obtained by standard low-frequency Fourier sampling. For fixed γ, this provides a measure of the resolution-efficiency of strategy S2 relative to strategy S1, where a resolution efficiency larger than 1 indicates that the S2 reconstruction has better resolution than S1 for the same SNR. These plots indicate that it is possible to achieve higher spatial resolution (except under the EW metric) for the same SNR using S2. However, this advantage is not maintained for arbitrary amounts of SNR improvement; for example, it is no longer resolution efficient under the RC to improve SNR more than about $\gamma = 2$ or 3 with any of the apodization windows that were considered. These results extend to higher dimensions in a nearly geometric fashion, with γ between 4 and 9 being reasonably efficient for 2D imaging scenarios. However, it is also important to realize that the performance curves shown in this subsection will vary as the image model and the data acquisition scheme are varied.

Example comparisons between S1 and S2 with simulated and real data are shown in Figs. 3.18 and 3.19, respectively. In these 2D cases, the S2 reconstruction has a roughly $1.4\times$ advantage in resolution (FWHM) compared to the S1 reconstruction, despite having identical SNR and data acquisition time.

Given the results presented in this subsection, a natural question is whether or not it is possible to improve on the resolution/SNR efficiency of smoothness-based regularization by careful optimization of apodizing window functions under appropriately-chosen resolution metrics. Our preliminary experience using stochastic optimization suggests a positive answer – for a fixed SNR, it is possible to achieve a better FWHM (for example) than that obtained using smoothness regularization. On the negative side, additional constraints are necessary to avoid undesirable SRF characteristics. For example, direct minimization of the SRF FWHM leads to an SRF with very poor side-lobe and tail characteristics. As a result, further investigation is necessary before op-

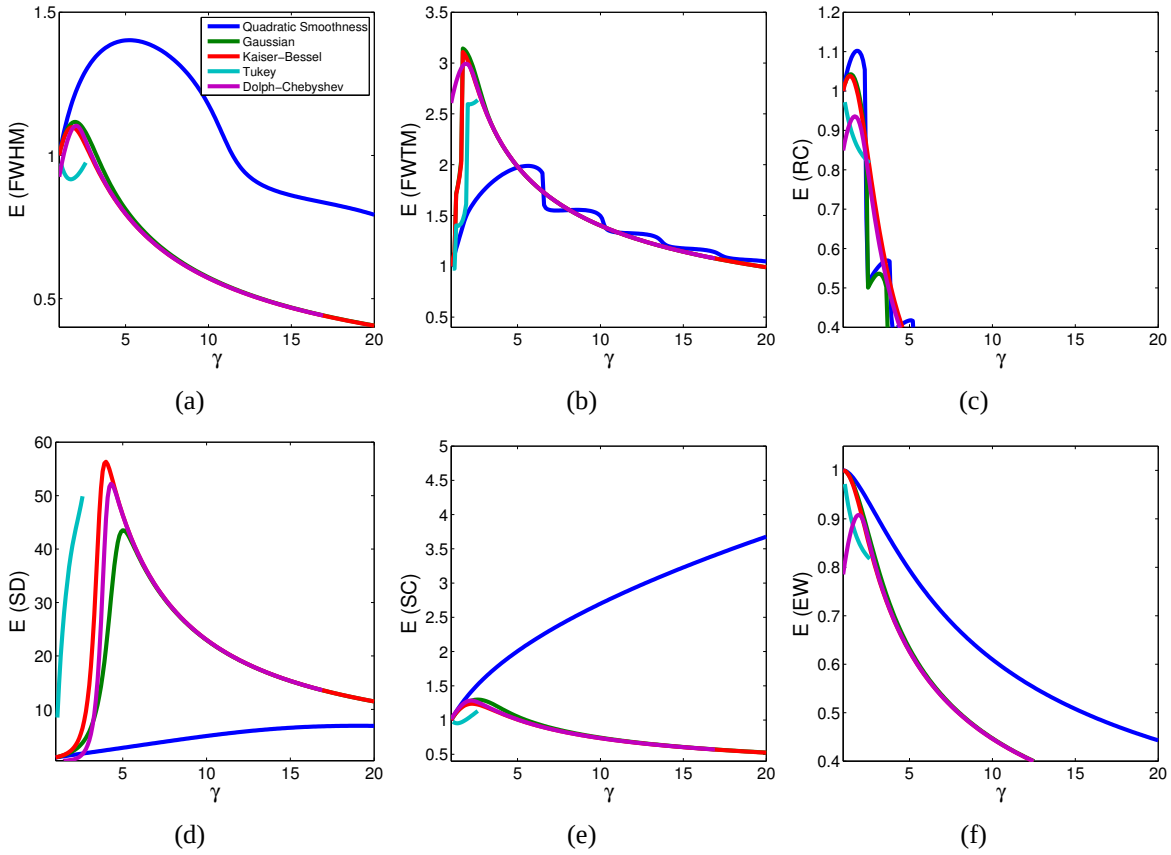
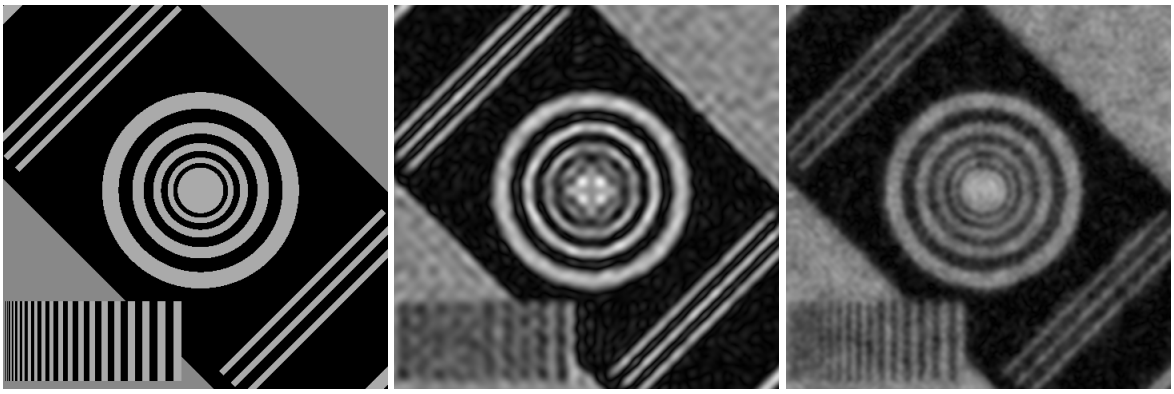


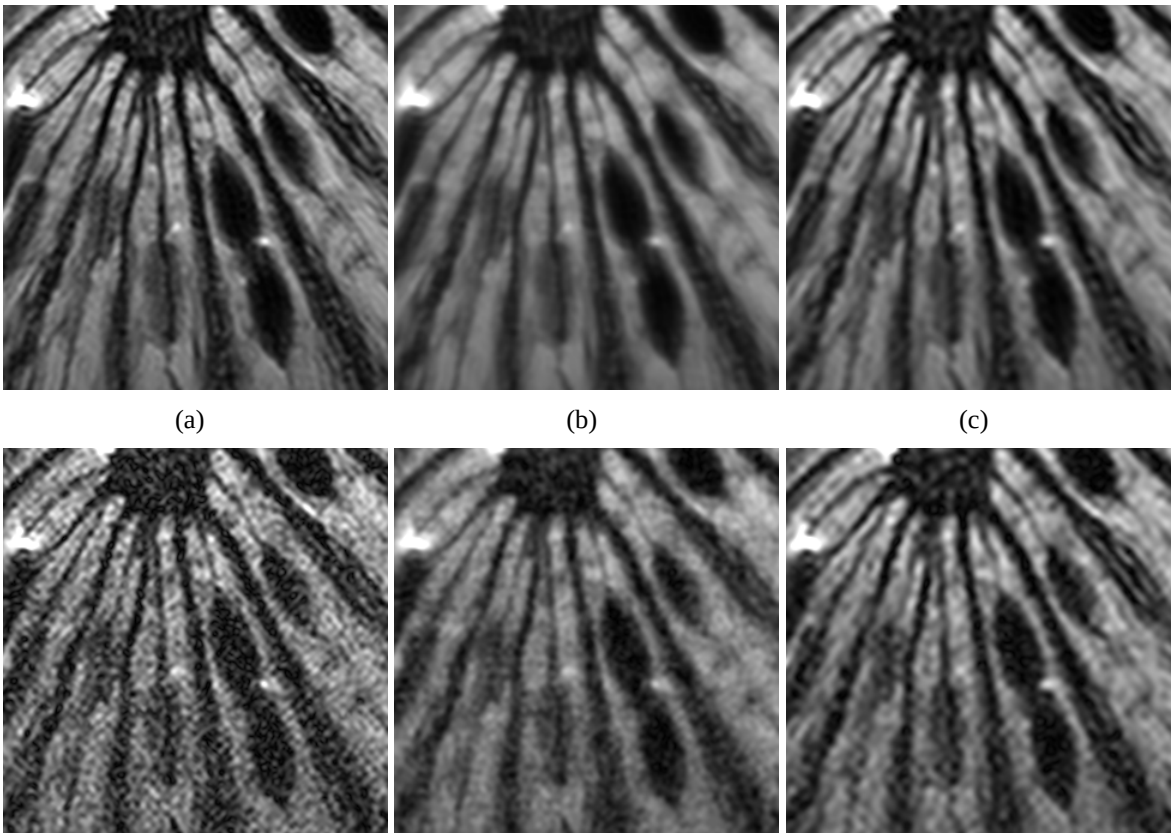
Figure 3.17: The resolution efficiency E for fixed SNR of windowed reconstruction versus standard low-frequency Fourier reconstruction.



(a) (b) (c)

Figure 3.18: Comparison of S1 versus S2 acquisition and reconstruction strategies, with equal data acquisition time assumed for both strategies. (a) Gold standard. (b) S1 image. (c) S2 image with $2\times$ higher-than-nominal *k*-space sampling. The S1 and S2 reconstructions have the same SNR, though the S2 reconstruction has significantly better resolution (FWHM) than the S1 reconstruction.

timized window functions could pose significant competition to schemes based on smoothness regularization.



(a) (b) (c)

(d) (e) (f)

Figure 3.19: Comparison of S1 versus S2 acquisition and reconstruction strategies using experimental MR images of a section of kiwi fruit. The top row shows high-SNR images reconstructed from real data, while simulated Gaussian noise has been added to the *k*-space data for the bottom row of images to reduce the SNR. (a,d) Standard Fourier reconstruction from high-resolution data. (b,e) S2 image using shift-invariant quadratic regularization ($\gamma = 4$). (c,f) S1 image with image resolution reduced by a factor of 2 along each dimension relative to (a). As before, the S1 and S2 reconstructions have the same SNR, though the S2 reconstruction has better resolution (FWHM) than the S1 reconstruction. The difference in resolution is most apparent when comparing the visibility and reconstructed width of the small line features running through the images.

3.3.3 Choosing $\Psi(\cdot)$

The previous subsections showed that considerable improvements in SNR were achievable at a relatively moderate loss of resolution, and that regularized reconstruction of high-resolution data can be significantly more efficient than simple averaging of low-resolution Fourier data. However, the key assumption for this resolution analysis to be applicable was that $\boldsymbol{\ell}$ is largely deterministic, such that the linear noise analysis can be applied to our nonlinear reconstruction. One way to achieve such an $\boldsymbol{\ell}$ is to have $\Psi'(t)/2t$ to be approximately constant for all t, and a perfect choice for this is

$$\Psi(t) = t^2. \tag{3.24}$$

Unfortunately, this choice is not edge preserving, and leads to $\ell_{n,m} = 1$ at all spatial locations. However, due to the sparsity of edges in natural images, we can obtain a mostly-constant line process map by using a Huber function [322] that transitions from quadratic to linear:

$$\begin{aligned} \Psi(t) &= \begin{cases} t^2, & t \leq \xi \\ 2\xi t - \xi^2, & t > \xi \end{cases} \\ &= \inf_{0<\ell\leq 1} \left(\ell t^2 + \frac{\xi^2}{\ell} - \xi^2 \right), \end{aligned} \tag{3.25}$$

where ξ is the parameter of the Huber function that controls where the transition takes place. The Huber function is convex (though not strictly convex), and is common in the edge-preserving regularization literature (e.g., see [54, 468, 472] and references for discussion). The corresponding expression used in Eq. (3.13) is

$$\frac{\Psi'(t)}{2t} = \begin{cases} 1, & t \leq \xi \\ \frac{\xi}{t}, & t > \xi \end{cases}. \tag{3.26}$$

Note that the Huber function has the ability to reduce the degree of spatial smoothing across the edges for which $t > \xi$, since the corresponding line process variables will get smaller and smaller as the argument of the Huber function grows large. Some of the edge-preserving characteristics of

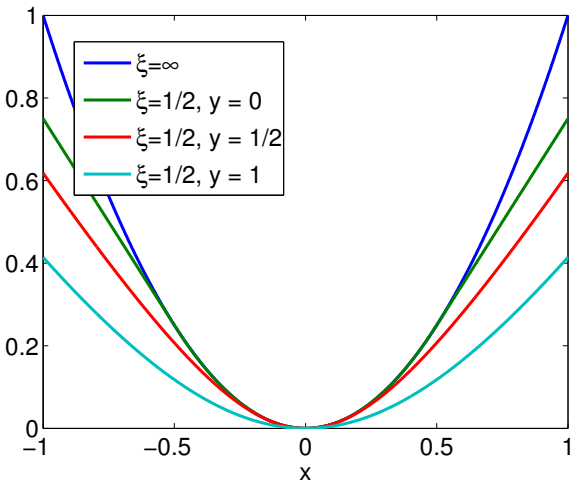
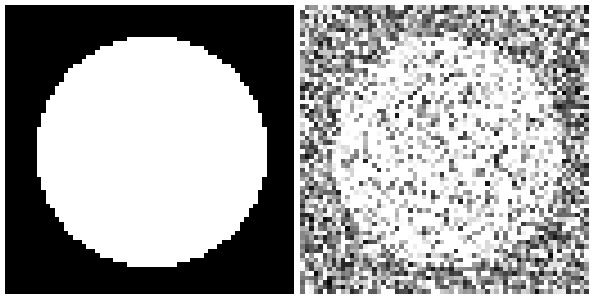


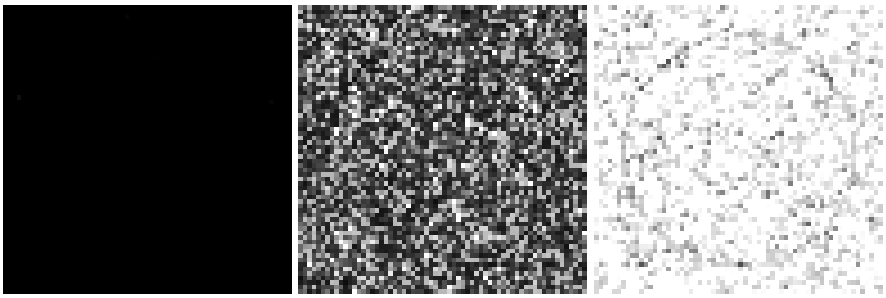
Figure 3.20: Illustration of some of the characteristics of our proposed regularization function based on the Huber function. The figure shows plots of $\Psi(\sqrt{|x|^2 + |y|^2}) - \Psi(\sqrt{|y|^2})$ as a function of x for different values of the Huber function parameter ξ and the variable y. In the context of our proposed reconstruction scheme, x and y would correspond to finite differences from different images, but from the same spatial location. As a result, x and y will share a line-process variable. For illustrative purposes, we assume that the value of y is fixed, and thus subtract out its contribution to the cost function for different values of x. When $\xi = \infty$, the Huber function is equivalent to the quadratic penalty in Eq. (3.24). In this case, the line-process variable always takes value one, and the cost function increases rapidly as the magnitude of x increases, heavily penalizing large edge values. As a result, the optimal reconstruction will generally not include significant edge features due to their large cost. In contrast, we are able to obtain edge preserving behavior when ξ is finite, which manifests as smaller cost function values for large values of x. In addition, when the value of y is large (i.e., when a correlated image also has a large edge value at the same spatial location), the cost function for x will decrease to enable x to more easily take on a large value.

the Huber function are illustrated in Fig. 3.20.

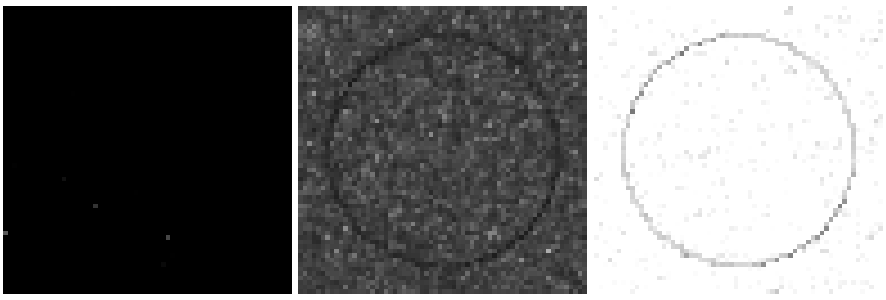
When the finite differences for a given spatial location lead to a value of t that is smaller than ξ, then the Huber function does not identify that location as an edge. Thus, if ξ is chosen to be larger than the expected noise contribution to the finite-difference computation of Eq. (3.14) and if the image is predominantly smooth, then the line process variables will be largely independent of the noise, and the previously described linear noise analysis will be valid. An empirical illustration of the proposed method's ability to achieve approximately noise-independent line process estimates with appropriately-chosen ξ is shown in Fig. 3.21, and Fig. 3.22 illustrates that after choosing ξ in this manner, we can accurately predict the SNR improvement using the previously-described linear



(a) Gold standard (b) Noisy image

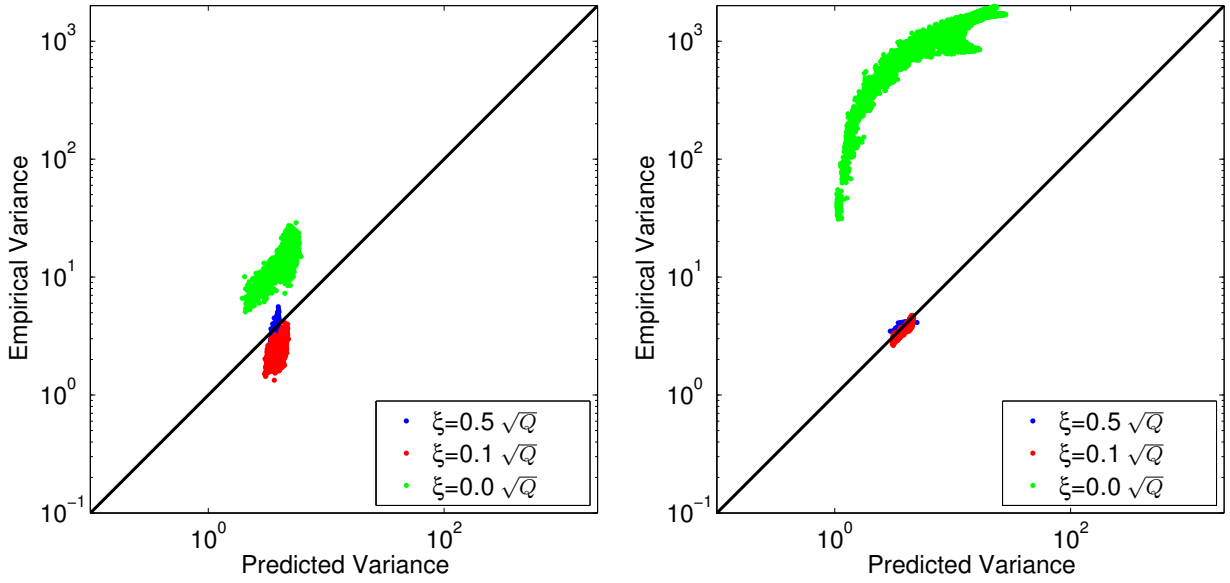


(c) $Q = 1, \xi = 0$ (d) $Q = 1, \xi = 0.1\sqrt{Q}$ (e) $Q = 1, \xi = 0.5\sqrt{Q}$



(f) $Q = 10, \xi = 0$ (g) $Q = 10, \xi = 0.1\sqrt{Q}$ (h) $Q = 10, \xi = 0.5\sqrt{Q}$

Figure 3.21: Empirical simulations were performed to illustrate that the proposed method with appropriately-designed Huber function regularization leads to predictable noise characteristics. For the simulation, noisy datasets were simulated for a noise-free image comprised of a homogeneous disc object on an empty background. The noise-free image is illustrated in (a), while a representative noisy Fourier reconstruction is shown in (b). (c-e) Estimated line process variables when reconstructing a single image (i.e., $Q = 1$) with the proposed method for different values of ξ. Dark values imply that the algorithm has identified a significant edge at the given spatial location. (f-h) Estimated line process variables when reconstructing ten different noisy realizations of the same image (i.e., $Q = 10$, with α_q and β_q the same for all q) with the proposed method for different values of ξ. When ξ is small, virtually all spatial locations are identified as edges. On the other hand, choosing ξ larger than the noise level allows the proposed method to better differentiate between real edge structures and noise, and the ability of the proposed method to robustly identify edge structures improves with increasing Q.



(a) $Q = 1$ (b) $Q = 10$

Figure 3.22: Illustration that appropriately-chosen values of ξ allow us to predict the noise variance in reconstructed images using the "deterministic" reconstruction matrix in Eq. (3.16) and the linear noise analysis from Eq. (2.46). The simulations described in Fig. 3.21 were repeated 500 times, with different noise realizations in each trial, and the empirical variance of each reconstructed pixel was computed. This was compared with the predicted variance for each pixel computed with Eq. (2.46), after assuming the reconstruction matrix in Eq. (3.16) was independent of the data noise. The figures show scatterplots comparing the empirical variance to the predicted variance. These scatterplots illustrate that when ξ is large enough that the line-process variables are not heavily influenced by the noise, the empirical pixel variance matches very well with the predicted pixel variances. However, if ξ is chosen too small, then the influence of noise on the reconstruction matrix in Eq. (3.16) makes it difficult to use Eq. (2.46) to predict the SNR improvement using the proposed method.

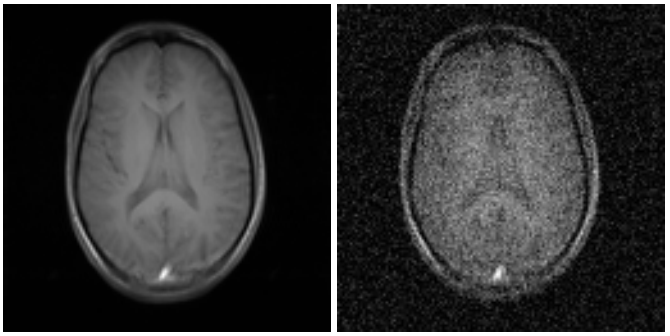
noise analysis. In addition to achieving approximately noise-independent ℓ with appropriately chosen ξ, the use of the Huber function also gives roughly uniform spatial smoothness weights in the smooth regions of the image, giving a fairly uniform improvement in SNR and uniform loss of resolution within these regions.

While choosing large values of ξ improves our ability to reject noise in the estimated line process variables, it also can lead to rejection of real edge structures. As a result, ξ should generally be set large enough that the edge map is approximately deterministic, but no larger. The consequences of different choices of ξ are illustrated with a real MR brain image in Fig. 3.23.

We note that allowing ξ to approach 0 is equivalent to using one form of the multichannel TV prior [79, 666]. However, as we have illustrated, our proposed Huber function prior can yield significantly better reconstruction characteristics for MR imaging problems.

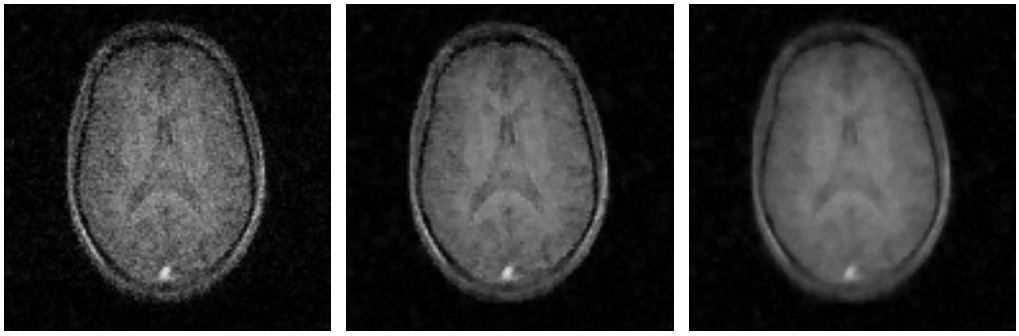
Convergence and Uniqueness Characteristics for the Huber Function

A nice feature of the Huber function is that it is convex, meaning that the cost function in Eq. (3.8) is also convex [74]. Convexity implies that any local minimum of Eq. (3.8) is also a global minimum, and that globally optimal solutions can be obtained using standard convex optimization methods [74]. However, it should be noted that the Huber function does not satisfy the strict convexity and differentiability constraints required for the global convergence guarantees of Delaney and Bresler [163] for the multiplicative half-quadratic algorithm. Even so, the algorithm is still guaranteed (by construction) to monotonically decrease the cost function in Eq. (3.8), and it can be proven that the sequence of iterates of the multiplicative half-quadratic algorithm will converge when using a Huber regularization penalty [472]. A formal proof that the multiplicative half-quadratic algorithm will always converge to a global minimum with Huber regularization has remained elusive, partly due to the lack of strict convexity/concavity and the fact that the Huber function is not twice continuously-differentiable at ξ [15, 131, 163, 328, 472]. However, it is still possible to test for global convergence to a global minimum. In particular, the first-order necessary and sufficient optimality condition for an optimal solution to the differentiable convex cost



(a) Gold standard (b) Noisy image

(c) $\xi = 10^{-4}$ (d) $\xi = 10^{-3}$ (e) $\xi = 10^{-2}$



(f) $\xi = 10^{-4}$ (g) $\xi = 10^{-3}$ (h) $\xi = 10^{-2}$

Figure 3.23: Illustration of the effects of different choices of ξ. In this simulation, we have reconstructed a noisy version of the T_1-weighted image from Fig. 3.2, assuming that the line-process variables have been estimated independently from the coregistered proton-density and T_2-weighted images from that figure. (a) Gold standard image. (b) Gold standard image with additional simulated noise. (c-e) Different line-process maps estimated from the proton-density and T_2-weighted images with different values of ξ. As expected, smaller values of ξ are able to better capture fine structural image features, though they also lead to increased noise sensitivity. (f-g) Reconstructed images using the proposed method. When ξ is too large, denoising is still achieved, though many image details are not preserved by the line processes. On the other hand, when ξ is too small, many noise features are identified as actual image structure, and are falsely preserved. As a result, ξ should be chosen to balance these two issues, and to ensure that the reconstructed image has predictable noise properties.

function in Eq. (3.8) is that the gradient of Eq. (3.8) is equal to 0 [74], i.e., that

$$2\left(\alpha_q^2 \mathbf{F}_q^H \mathbf{F}_q + \beta_q^2 \mathbf{D}^H \mathrm{diag}\left(\ell_{n,m}\right) \mathbf{D}\right) \hat{\boldsymbol{\rho}}_q - 2\alpha_q^2 \mathbf{F}_q^H \mathbf{d}^q = \mathbf{0} \tag{3.27}$$

for $q = 1, \ldots, Q$, where $\mathrm{diag}\left(\ell_{n,m}\right)$ is the diagonal matrix with diagonal elements equal to the optimal line-process values for the $\{\hat{\boldsymbol{\rho}}_q\}$, as computed using Eqs. (3.13) and (3.14). This equation provides a convenient means for testing for the global optimality of a numerical solution to Eq. (3.8). Despite the lack of proven global convergence guarantees for the multiplicative half-quadratic algorithm, global convergence has always been observed in practice. It should also be noted that the multiplicative half-quadratic algorithm is not the only algorithm that can be used to minimize Eq. (3.8), and some of the alternative algorithms have global convergence guarantees; these issues will be discussed in Sec. 3.7. Regardless of the lack of theoretical global convergence guarantees, the multiplicative half-quadratic procedure is of significant interest because it provides the linear reconstruction interpretation in Eq. (3.16), which was the foundation of our resolution and noise analysis.

An important question from the perspective of optimization is whether Eq. (3.8) has a unique solution. It can be difficult to prove uniqueness of a solution for arbitrary $\{\mathbf{F}_q\}$, $\{\mathbf{D}\}$, $\{\alpha_q\}$, and $\{\beta_q\}$; however, we will demonstrate that a unique solution exists in the context of two commonly-appearing special cases:

1. Eq. (3.8) has a unique solution when, for each q, $\mathbf{F}_q$ has a trivial nullspace and $\alpha_q > 0$. This case is commonly obtained in the denoising context described in Sec. 3.2.1. In this case, the data fidelity terms in Eq. (3.8) will be strictly convex, such that the overall cost function is also strictly convex. Strict convexity implies that there exists a unique globally-optimal solution [74].

2. Eq. (3.8) has a unique solution for the super-resolution mode of operation described in Sec. 3.2.1. In particular, we assume that for each q corresponding to a high-resolution dataset, $\mathbf{F}_q$ has a trivial nullspace, $\beta_q > 0$, and $\frac{\beta_q^2}{\alpha_q^2}$ is set arbitrarily close to zero. In this

case, the high-resolution images are reconstructed uniquely as

$$\hat{\boldsymbol{\rho}}_q = \left(\mathbf{F}_q^H \mathbf{F}_q\right)^{-1} \mathbf{F}_q^H \mathbf{d}^q. \tag{3.28}$$

Assuming also that β_q is set arbitrarily close to zero for the remaining images, then the line-process variables depend only on the high-resolution reconstructed images from Eq. (3.28), and are uniquely determined through Eq. (3.13). Finally, the remaining images are reconstructed using Eq. (3.16), and have unique solutions as long as the nullspaces of $\mathbf{F}_q$ and $\frac{\beta_q}{\alpha_q}\mathbf{D}$ have trivial intersection.

Nonconvex Alternatives to the Huber Function

One of the motivating factors for choosing $\Psi(\cdot)$ to be the Huber function was convexity, since convexity implies that global optimization can be achieved in a straightforward way. However, it should also be noted that nonconvex regularization functionals can be beneficial, because they can be even more edge-preserving (i.e., they can impose a smaller penalty on large edge values) than the Huber function [163]. In general, it is often difficult to guarantee global optimization of a nonconvex cost functional without resorting to time-consuming stochastic algorithms like simulated annealing [237]. However, for the super-resolution mode of operation for our proposed method, the ease of obtaining a unique globally-optimal reconstruction is independent of the convexity of $\Psi(\cdot)$. As a result, the use of the following nonconvex $\Psi(\cdot)$ can be beneficial for the super-resolution problem:[6]

$$\Psi(t) = \begin{cases} t^2, & t \leq \xi \\ \frac{2\xi^{2-\nu}}{\nu} t^\nu - \frac{2\xi^2}{\nu} + \xi^2, & t > \xi \end{cases}, \tag{3.29}$$

[6]This nonconvex $\Psi(\cdot)$ could also be useful for the denoising problem; however, it would be much more difficult to ensure global optimality in this case, which leads us to prefer the Huber function for the denoising problem.

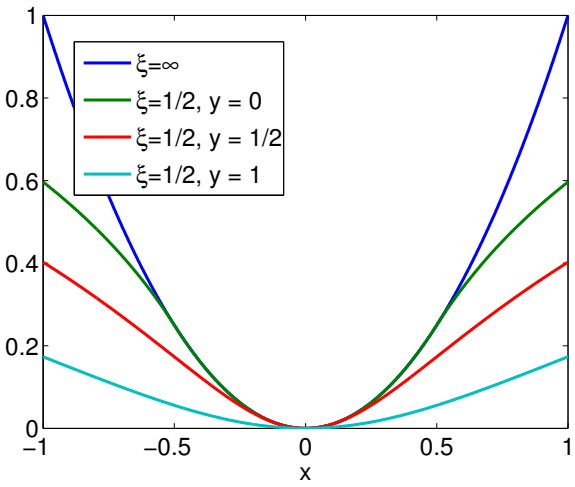
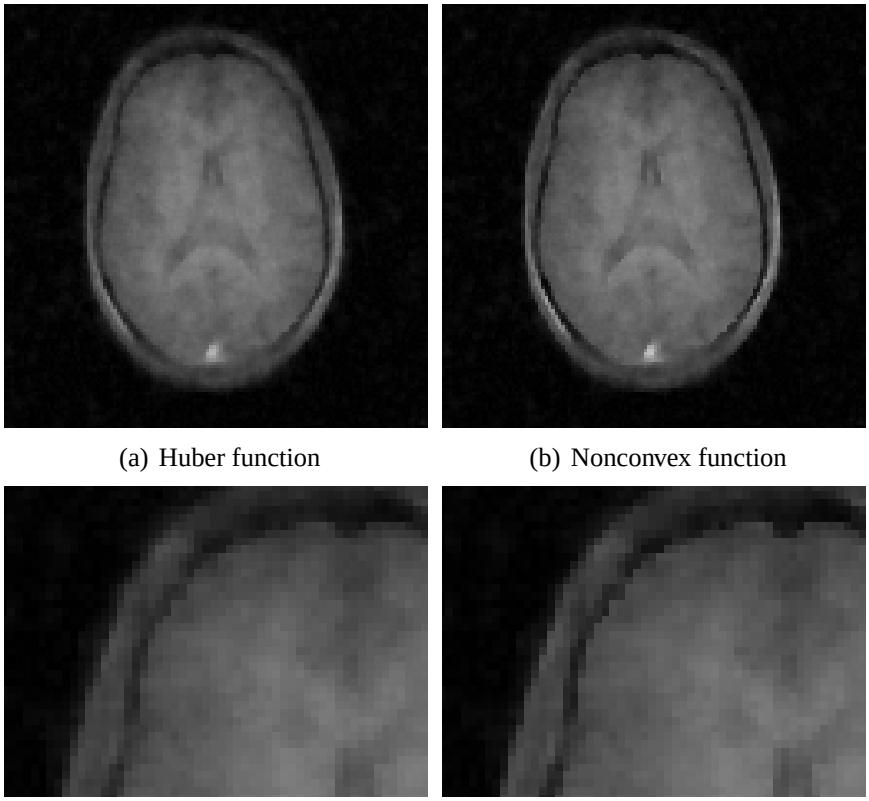


Figure 3.24: Illustration of some of the characteristics of the nonconvex $\Psi(\cdot)$ function in Eq. (3.29) with $\nu = 0.5$. The figure shows plots of $\Psi(\sqrt{|x|^2 + |y|^2}) - \Psi(\sqrt{|y|^2})$ as a function of x for different values of the Huber function parameter ξ and the variable y. Comparing this cost function with the traditional Huber function shown in Fig. 3.20, it is clear that the nonconvex cost function penalizes large edge values much less than the Huber function would.

where $\nu \in (0, 1)$ is a small positive constant. This penalty function approaches the Huber function as $\nu \to 1$, and satisfies

$$\frac{\Psi'(t)}{2t} = \begin{cases} 1, & t \leq \xi \\ \left(\frac{\xi}{t}\right)^{2-\nu}, & t > \xi \end{cases} . \tag{3.30}$$

Comparing $\Psi'(t)/(2t)$ for this function with Eq. (3.26) for the Huber function, it is apparent that the estimated line-process variables are much smaller for the nonconvex penalty when t is large, particularly as ν approaches 0. As a result, the nonconvex cost function is significantly more tolerant to large edge-values than the Huber function, and thus can be even better at avoiding partial volume effects when imposing spatial smoothness to enhance image SNR. The cost function is illustrated in Fig. 3.24, while the potential for improved performance using the cost function is illustrated in Fig. 3.25.



(a) Huber function (b) Nonconvex function

(c) Huber function (d) Nonconvex function

Figure 3.25: Illustration of the difference between reconstruction using the Huber function and the nonconvex penalty from Eq. (3.29). This simulation was exactly the same as the simulation described in Fig. 3.23. Reconstructions were performed using (a) the Huber function and (b) the nonconvex penalty with a vanishingly-small value of ν, both with $\xi = 10^{-2}$. Examining the contour separating the extracranial tissues from the brain parenchyma in the ROI shown in (c,d), it is clear that the reconstruction using the nonconvex penalty demonstrates somewhat less edge-blurring and reduced partial volume artifacts, as expected.

3.4 Application Examples

The previous section illustrated that the proposed approach is easy to characterize, and that it is efficient to regularize high-resolution noisy data. We have reconstructed many different datasets with this technique, and examples include MR spectroscopic imaging [269,270,274,276,277,286], sodium and oxygen imaging [24,25], and diffusion imaging [232,267,278–280,284,285,378]. For simplicity, we show just a few examples in this thesis.

3.4.1 Phantom Experiments

Phantom data was acquired on a Varian INOVA 14.1T MR system to illustrate the benefit of the proposed reconstruction for standard imaging experiments. In these experiments, a phantom was imaged using two phase-encoded spin-echo sequences, one with TE = 23 ms and TR = 1000 ms, and the other with TE = 40 ms and TR = 200 ms. The short TR experiment suffers from SNR problems at high resolution, though the data was acquired five times faster than the long TR experiment. Figure 3.26 shows the results of using the long TR image to constrain reconstruction of the short TR image. In this case, the algorithm was used in super-resolution mode, where $\{\alpha_q\}$ and $\{\beta_q\}$ were adjusted so that the line-process values depended only on the reference image, and so that the reference image was perfectly data consistent. The proposed method successfully mitigates the effects of noise while mostly preserving the resolution of small image features. Importantly, this experiment confirms our previous observation that the proposed method is more effective for denoising high-resolution data than it is for achieving super-resolution reconstruction.

3.4.2 Mouse Brain Diffusion Experiment

Diffusion-weighted (DW) MRI experiments can be used to characterize the random microscopic thermal Brownian motion of water molecules [388]. The diffusion process in biological tissues is highly sensitive to tissue microstructure, making DW MRI a powerful clinical tool for the detection and characterization of various pathologies. DW MRI has been found particularly relevant in

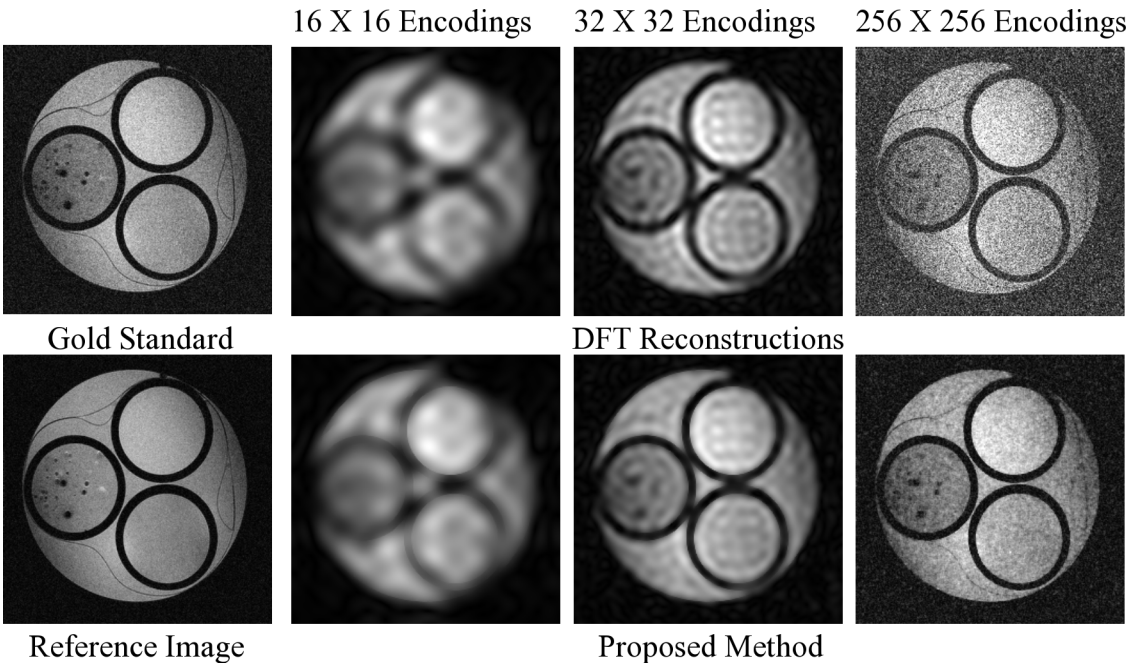


Figure 3.26: Experimental phantom reconstruction results. A short TR spin-echo sequence was used to acquire noisy *k*-space data, while data from a long TR spin-echo sequence was used as a reference image to generate anatomical constraints. The gold standard image was acquired by averaging the short TR experiment 8 times. The proposed reconstruction from the high-resolution noisy data has both high SNR and high resolution. This is in contrast to the reconstructions using low-resolution data, which have high SNR but reveal limited information regarding the small features of the image.

studies of the central nervous system, and has proven to be valuable in the assessment of trauma, ischemia, cancer and neurodegenerative diseases [388]. However, quantitative DW MRI experiments are often very time consuming, due to low SNR and the need to have multiple DW images for quantitative assessment of diffusion characteristics.

DW MRI experiments were performed on a euthanized juvenile mouse to illustrate the performance of our proposed method in this context. Experiments used a 14.1 T scanner, and sagittal images were acquired with a standard spin-echo DW MRI pulse sequence (TE = 32 ms, TR = 1000 ms, δ = 8 ms, Δ = 20 ms). Twelve different diffusion weightings were applied, with diffusion-weighting factors (b-values) ranging from 0 to $10,000$ s/mm^2. All diffusion gradients were applied along the dorso-ventral direction. Each DW image was encoded using a 128×256 *k*-space sampling grid, with samples spaced evenly at the Nyquist rate corresponding to a 1.9 cm $\times$ 3.8 cm field of view, and the slice thickness was 0.25 mm. The experiment used a transmit/receive volume RF coil. Four averages were acquired, but were saved separately. Reconstruction was performed using the denoising mode of operation. Results are shown in Fig. 3.27, and demonstrate significantly improved SNR. In particular, the SNR has improved equivalent to averaging four times in the smooth regions of the image, while the average voxel size (FWHM) has only degraded from $0.15 \times 0.15 \times 0.25$ mm^3 to $0.17 \times 0.17 \times 0.25$ mm^3. Figure 3.28 illustrates the effect of varying the Huber function parameter ξ on the estimated line process variables and the resulting level of SNR improvement, showing results consistent with the analysis in Sec. 3.3.3.

Figure 3.29 provides a comparison of our proposed method with two other common denoising schemes. The implementations and properties of these other reconstructions are summarized below:

- **BayesShrink thresholding in a decorrelated transform domain**. We applied the method described in Ref. [23]. This method first applies transforms to the image sequence to obtain a set of decorrelated image coefficients. Decorrelation is achieved between the different image frames by applying the empirical Karhunen-Loeve transform, while decorrelation is achieved in the spatial dimension by applying the wavelet transform. After decorrelation,

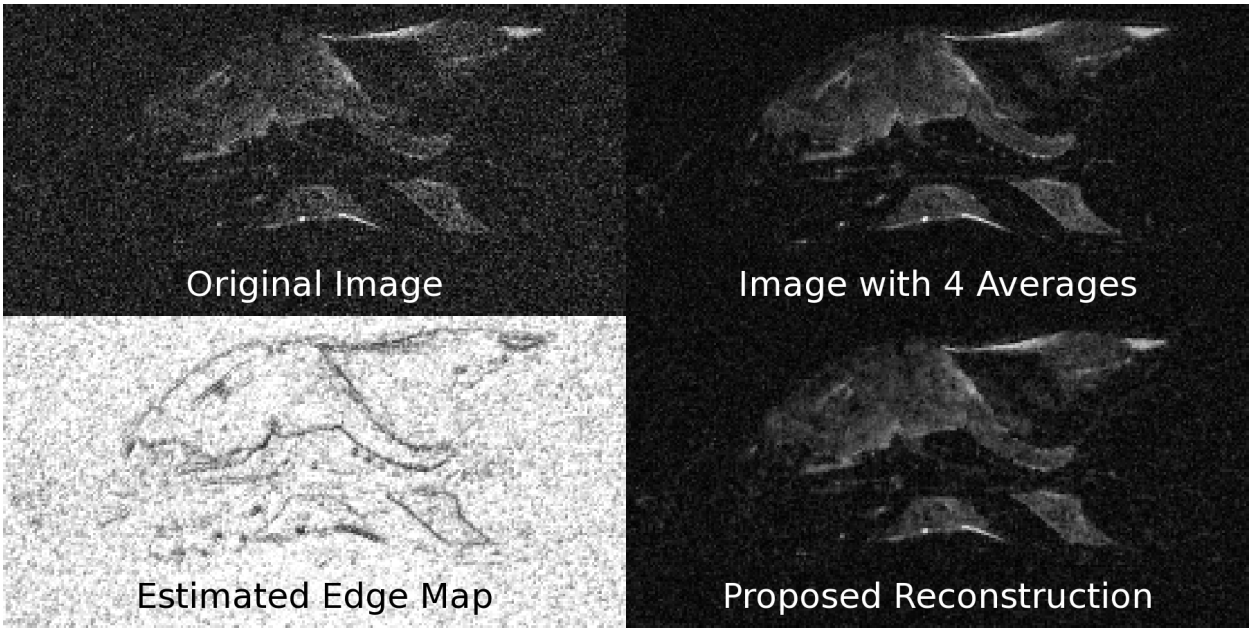


Figure 3.27: One image frame from a 12-frame diffusion-weighted mouse-brain sequence. The proposed method improves image SNR, while preserving important structural features.

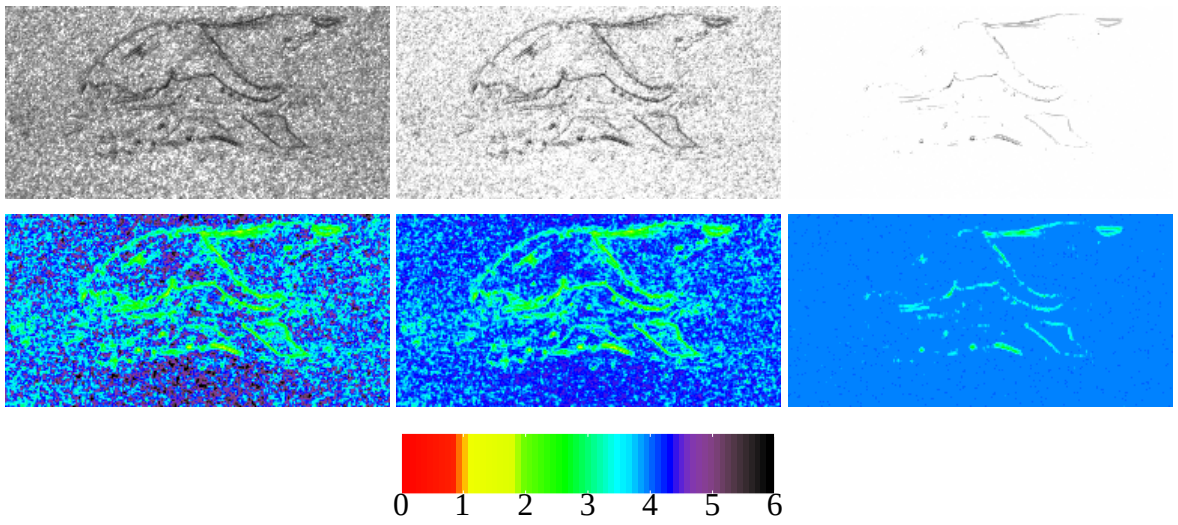


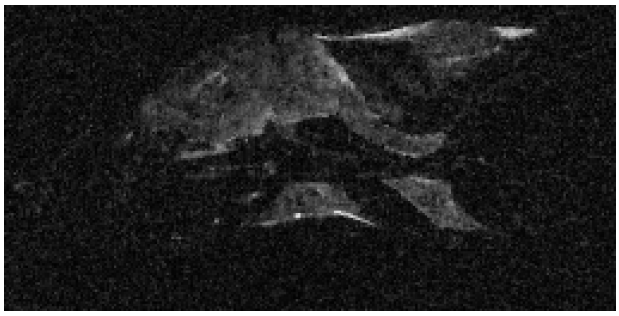
Figure 3.28: The effects of changing the parameter ξ in the Huber function. The image columns from left to right show increasing values of ξ. The top row shows the estimated $\ell_{n,m}$ variables, while the bottom row shows the equivalent number of averages for each voxel. Rather than directly evaluating the reconstruction noise using Eq. (2.46), reconstruction noise was estimated empircally using Monte Carlo methods (i.e., reconstruction of multiple simulated complex Gaussian noise vectors) to accelerate computations. As ξ increases, $\boldsymbol{\ell}$ becomes less sensitive to noise and the SNR improvement becomes more homogeneous over the FOV. On the other hand, image features are no longer as clearly evident in the estimated edge map.

denoising is performed using the BayesShrink adaptive thresholding algorithm [130]. This method has the capability of preserving high-resolution features. However, the resolution and noise characteristics of this method are not easily characterized. We observe that this method has successfully reduced the visually-apparent image noise, though structural image features and image contrast have also been adversely affected.

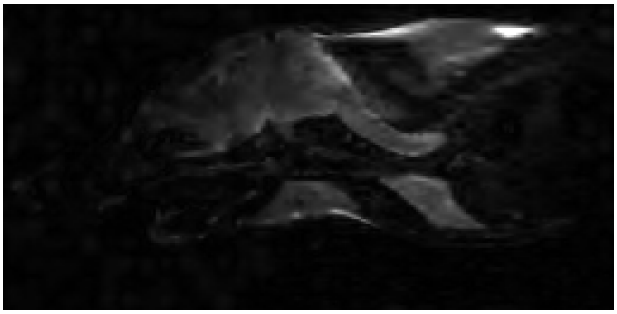
- **Coherence-enhancing PDE filter**. We implemented the PDE filtering approach described in Ref. [172]. The evolution of the PDE was halted when the variance from non-signal regions of the image was similar to that of the proposed method. The differences between the PDE filtered images and the proposed reconstructions are somewhat subtle; however, close examination shows that high-resolution features are more accurately preserved using the proposed method. In addition, the noise and resolution properties of the proposed reconstruction are easily and precisely characterized in terms of the SRF, while no convenient mechanisms exist for constructing SRFs in the PDE reconstruction. While point-spread functions of the PDE reconstruction can easily be computed due to the linearity of the PDE evolution at each discretized time step, this is not as useful as the SRF for characterizing resolution and noise.

3.4.3 T_1 Estimation Simulations

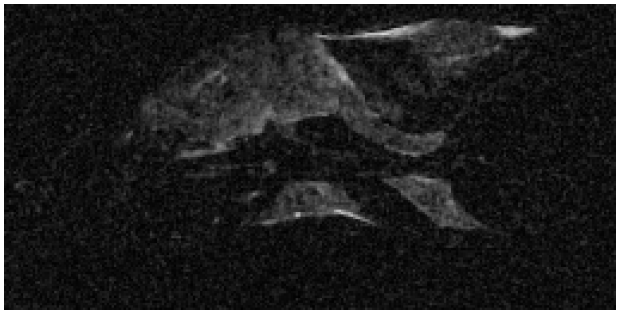
This experiment considered reconstruction of nine images from a variable flip-angle T_1 measurement experiment. The original images had high SNR, so additional simulated noise was added. The proposed approach was applied ($3\times$ improvement in noise variance) in the denoising mode of operation, and a denoising scheme based on principal component hard-thresholding [100] was also implemented for comparison. The results are shown in Fig. 3.30. While the principal component denoising approach gives very visually impressive denoising results for the image, the parameter estimate quality is actually worse than it was for the original noisy data. This illustrates the importance of having characterizable reconstructions; many denoising algorithms might be able to



(a)



(b)



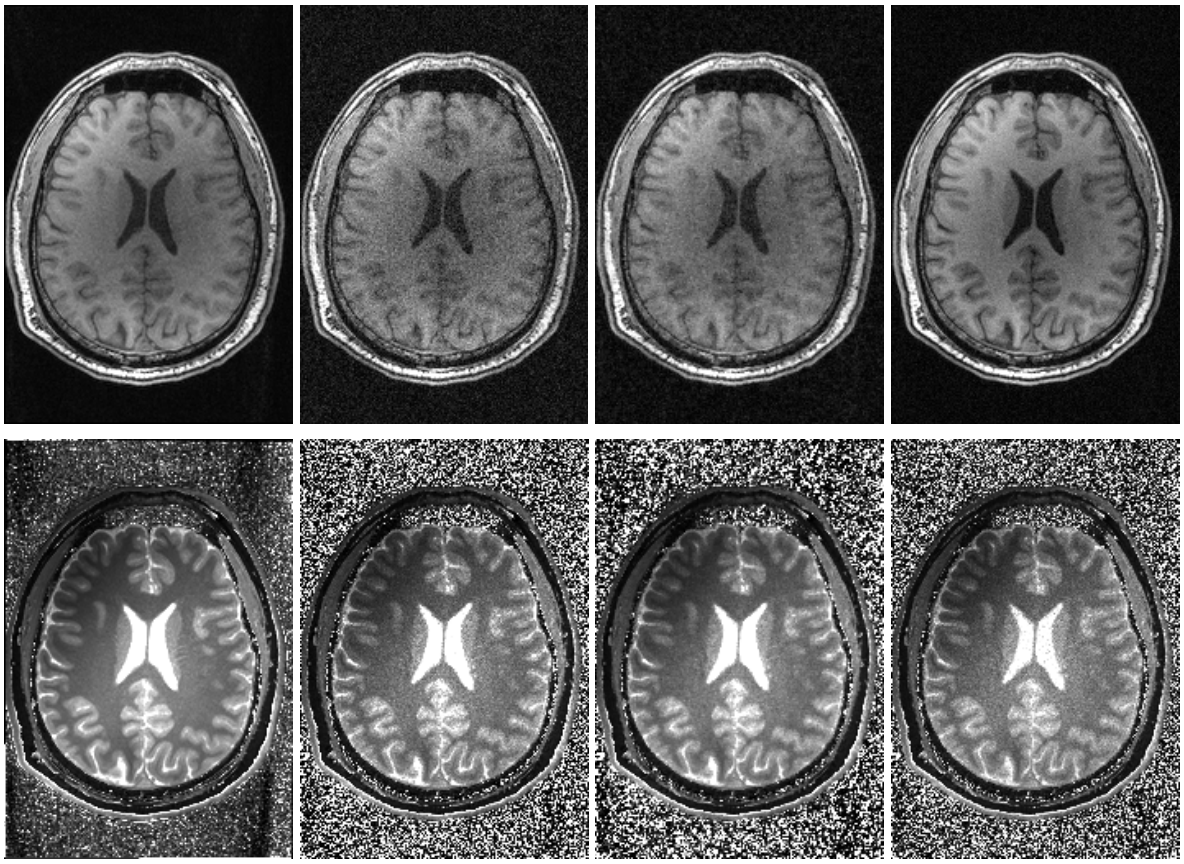
(c)

Figure 3.29: Comparison of the proposed method with other common denoising methods. (a) Proposed method. (b) BayesShrink thresholding. (c) Coherence-enhancing PDE filtering.

produce visually-attractive images, but characterization is important for proper understanding of quantitative parameter estimates that are obtained using the denoised data.

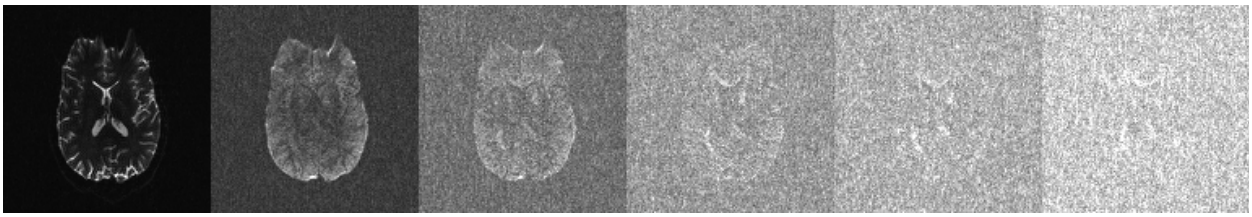
3.4.4 Human Brain Diffusion Tensor Experiment

This subsection demonstrates the performance of the proposed method in the context of *in vivo* human brain diffusion tensor imaging [460]. Diffusion tensor imaging is the most widely-used form of quantitative DW MRI. In this experiment, a total of 155 different DW images were acquired at 2 mm $\times$ 2 mm $\times$ 2 mm resolution on a 3T scanner, with b-values ranging from 0 s/mm^2 to 5,000 s/mm^2. Acquisition used 5/8ths partial Fourier sampling [101, 262, 307, 400, 475, 673], and an array of 32 different receiver coils. The proposed reconstruction method was used in the denoising mode of operation, but with additional modifications to the Fourier sampling operator to impose the phase constraints necessary for partial Fourier reconstruction [284]. The images corresponding to different receiver coils were reconstructed separately and combined in post-processing, though

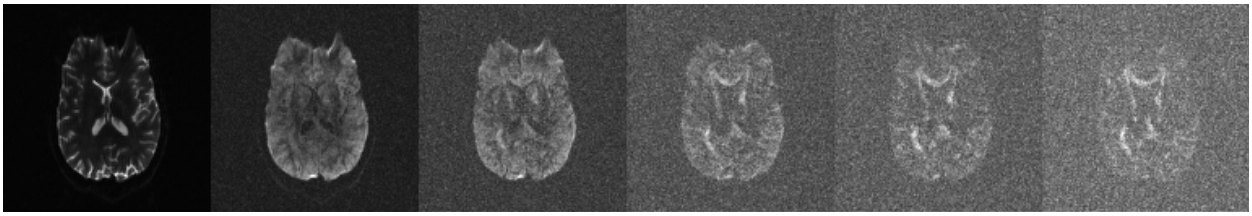


(e) Gold Standard Image (f) Noisy Image (g) Proposed Denoising (h) Principle Component Denoising

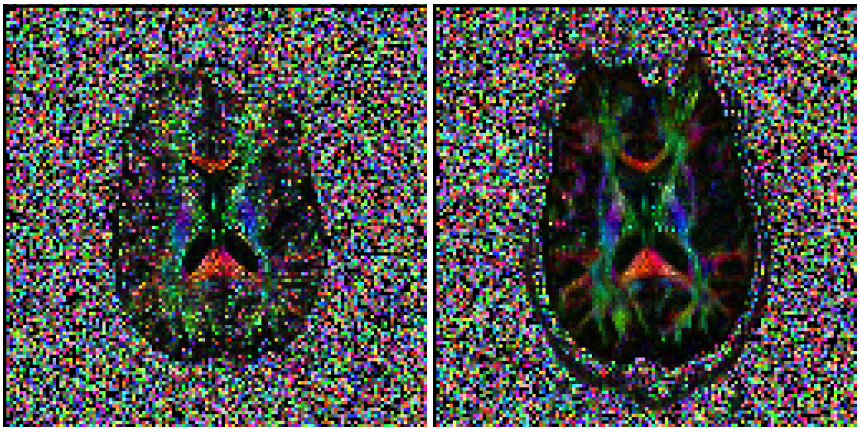
Figure 3.30: Results from the T_1 denoising simulation. The top row shows one image from the nine-image sequence, while the bottom row shows T_1 parameter estimates. The noisy parameter map has 12.0% error in signal regions and the parameter map after the proposed denoising has 8.9% error, while the parameter map after principal component denoising has 14.0% error.



(a)



(b)



(c) (d)

Figure 3.31: Results from the *in vivo* human brain diffusion tensor imaging experiment. (a) Standard reconstructions. (b) Proposed reconstructions. (c) Color-coded fractional anisotropy map estimated from the standard reconstructions [460]. (d) Color-coded fractional anisotropy map estimated from the proposed reconstructions.

reconstruction assumed that the different coil images possessed a shared edge structure. Results of this experiment are shown in Fig. 3.31, and demonstrate significantly improved image SNR and improved estimation of the diffusion tensor characteristics.

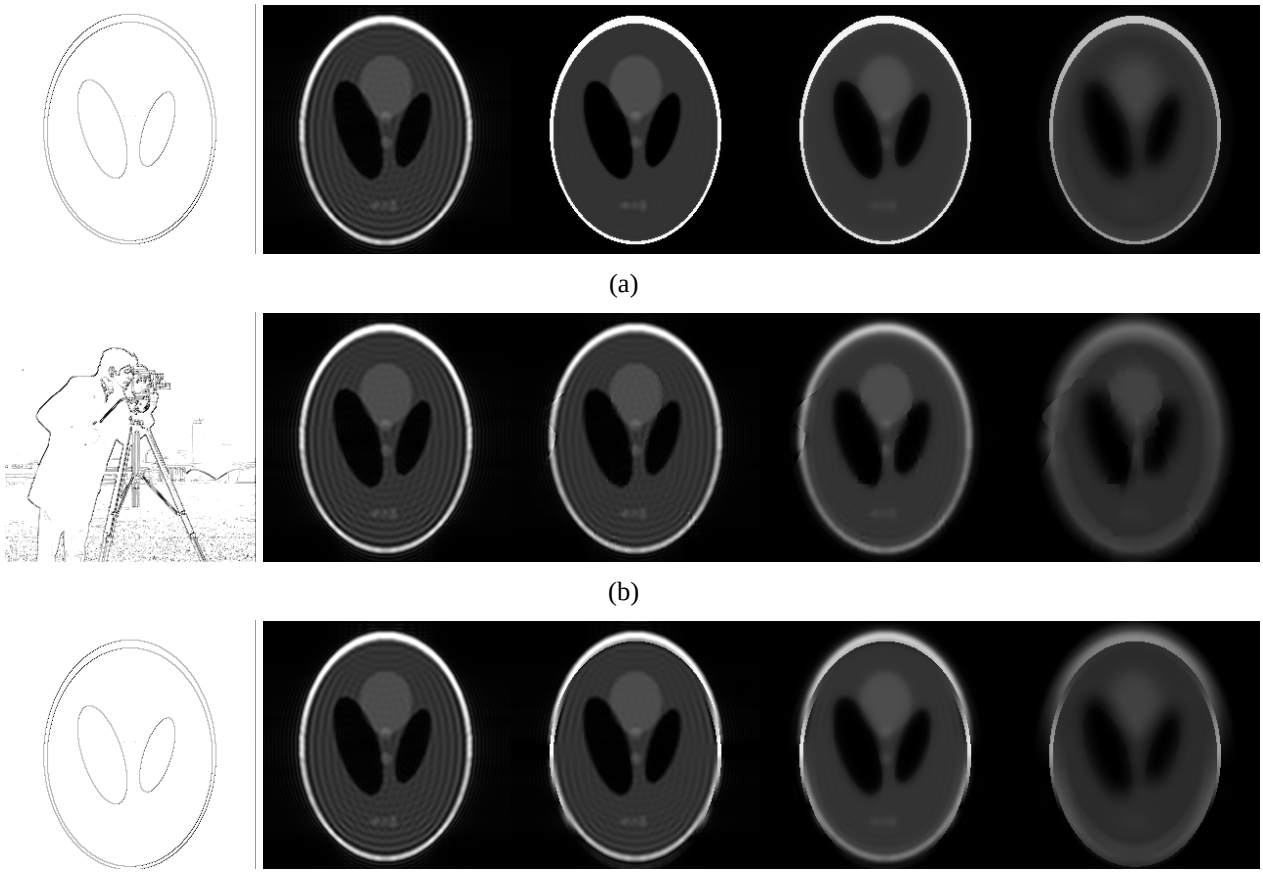
3.5 Performance in the Presence of Inaccurate and Incomplete Edge Information

A natural concern for the proposed method is how it behaves in the presence of inaccurate and incomplete edge information. The analysis of SRFs would lead to the conclusion that inaccurate edge information should be somewhat benign. While the method will not be able to avoid partial volume artifacts across unknown edge structures, the trade-off between resolution and SNR will still be efficient. In addition, the method is also fairly resilient to false edge structures. In this case, smoothing will not be performed across the false edge, leading to a slight reduction in SNR improvement, though false edges will generally not manifest as false image structures unless $\frac{\beta_q^2}{\alpha_q^2}$ is so large that there is a significant loss in spatial resolution. Note that as $\frac{\beta_q^2}{\alpha_q^2}$ becomes larger and larger, the reconstructed images are encouraged to become more and more piecewise constant, with discontinuities at the image boundaries. As a result, incorrect boundary information could be very detrimental in this extreme scenario. Illustrations of the resilience of the proposed method to inaccurate edge information are shown in Figs. 3.32-3.35.

3.6 Optimal Averaging Designs

We showed in Sec. 3.3.2 that our proposed reconstruction was efficient with respect to the resolution/SNR trade-off when used with a high-resolution data acquisition strategy with uniform-density *k*-space coverage. However, it has also been shown previously that data filtering of uniformly-sampled data is not as SNR efficient as if the acquisition strategy is modified to match the characteristics of the filter function, and significant SNR/resolution efficiency advantages for filtered reconstruction can be achieved with variable-density sampling/averaging [90, 253, 323, 436, 498, 513, 586, 589].

In this section, we derive an optimal data-averaging strategy for general fixed linear image reconstruction problems. This strategy can be used to further enhance the SNR of our proposed

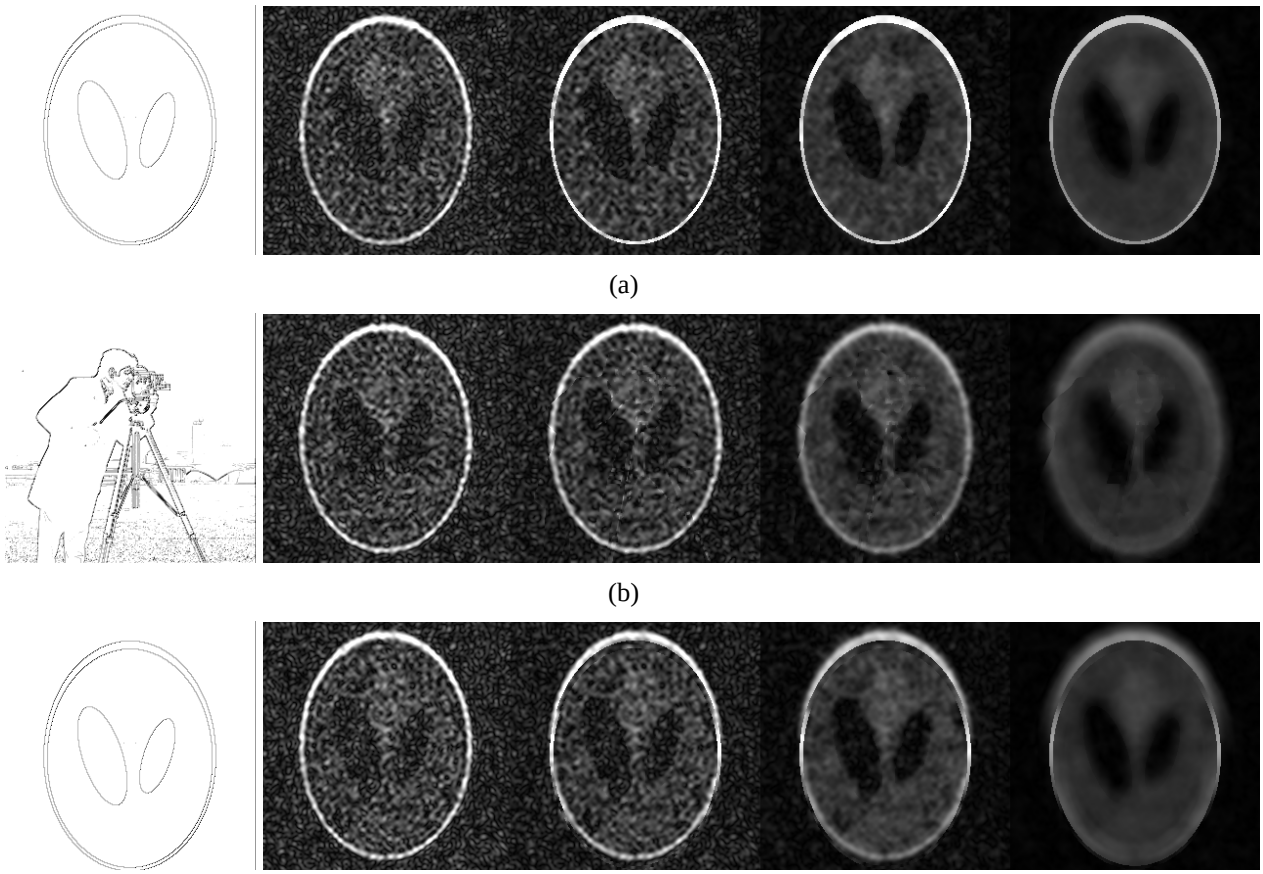


(a)

(b)

(c)

Figure 3.32: Reconstruction of the noiseless Shepp-Logan phantom with (a) accurate edge information, (b) completely inaccurate edge information, and (c) misregistered edge information. The left-most column of these subfigures shows the line-process variables used for reconstruction, which were assumed to be given and were not estimated jointly with the images. The remaining images show reconstructions using the proposed method, with increasing values of $\frac{\beta_q^2}{\alpha_q^2}$ as we move from left to right.



(a)

(b)

(c)

Figure 3.33: Reconstruction of the noisy Shepp-Logan phantom with (a) accurate edge information, (b) completely inaccurate edge information, and (c) misregistered edge information. The leftmost column of these subfigures shows the line-process variables used for reconstruction, which were assumed to be given and were not estimated jointly with the images. The remaining images show reconstructions using the proposed method, with increasing values of $\frac{\beta_q^2}{\alpha_q^2}$ as we move from left to right.

(a)

(b)

(c)

Figure 3.34: Reconstruction of the noiseless *Cameraman* image with (a) accurate edge information, (b) completely inaccurate edge information, and (c) misregistered edge information. The leftmost column of these subfigures shows the line-process variables used for reconstruction, which were assumed to be given and were not estimated jointly with the images. The remaining images show reconstructions using the proposed method, with increasing values of $\frac{\beta_q^2}{\alpha_q^2}$ as we move from left to right.

(a)

(b)

(c)

Figure 3.35: Reconstruction of the noisy *Cameraman* image with (a) accurate edge information, (b) completely inaccurate edge information, and (c) misregistered edge information. The left-most column of these subfigures shows the line-process variables used for reconstruction, which were assumed to be given and were not estimated jointly with the images. The remaining images show reconstructions using the proposed method, with increasing values of $\frac{\beta_q^2}{\alpha_q^2}$ as we move from left to right.

method when there is time for averaging during data acquisition. In our previous discussion, we assumed that each measurement sample in $\mathbf{d}^q$ was acquired exactly once, and that the noise variance was equal for all elements of $\mathbf{d}^q$. This assumption will be relaxed in this subsection, so that the mth element of $\mathbf{d}^q$ is assumed to be averaged ω_m times, and thus has an m-dependent variance of σ_q^2/ω_m. Our goal in this section will be to optimize $\{\omega_m\}$ to minimize the covariance matrix of the reconstructed $\hat{\boldsymbol{\rho}}_q$.

The problem of designing an experiment to optimize the covariance of a reconstruction is well studied in the field of statistics [205, 521, 576]. The existing literature generally considers the case where an ML reconstruction is obtained. However, in this case, the ML solution is generally an explicit function of the averaging strategy $\{\omega_m\}$, which makes optimization difficult. In the absence of special problem structure, optimal designs have generally been found numerically with iterative convex optimization techniques [74].

In this section, we observe that certain optimal averaging designs have simple closed-form solutions when the reconstruction matrix is linear and independent of the noise covariance. As a result, assume that a particular reconstructed image is obtained from measured data according to the linear reconstruction

$$\hat{\boldsymbol{\rho}}_q = \mathbf{G}\mathbf{d}^q, \tag{3.31}$$

where $\mathbf{G}$ is an $N \times M$ reconstruction matrix. In the context of our proposed method, it is reasonable to take $\mathbf{G}$ as the reconstruction matrix from Eq. (3.16), though the treatment in this section will be written for the general case where $\mathbf{G}$ is allowed to be arbitrary. For most ML or PML methods, the $\mathbf{G}$ reconstruction matrix would be chosen to be a function of $\{\omega_m\}$. However, in the context of MRI reconstruction, allowing $\mathbf{G}$ to be dependent on $\{\omega_m\}$ would also mean that the reconstruction SRFs would change as a function of $\{\omega_m\}$. This would be problematic, since modification of the averaging strategy could lead to undesirable unpredicted changes in the SRF characteristics. As a result, we will assume that we have already designed a reconstruction with desirable SRF characteristics for a given set of line-process variables (e.g., the matrix equation in Eq. (3.16)), and

will modify the averaging scheme to improve the SNR characteristics under this reconstruction. Note that the optimal averaging scheme is dependent on the estimated line-process variables. This is not a problem in the super-resolution mode of our proposed method, since the data for the high-resolution reference images providing edge information can be acquired prior to the data for the low-resolution images. This issue is more problematic for the denoising mode. In this case, it is reasonable to optimize the acquisition scheme for future experiments based on the empirical distribution of line processes that have been estimated in previous experiments.

We begin by assuming that the M measurements in $\mathbf{d}^q$ are partitioned into P subsets Θ_p of size $|\Theta_p|$. In addition, we assume that each subset of measurements Θ_p must be acquired simultaneously, so that each element of Θ_p must be allocated the same number of averages. This allows us to accommodate the conventional case in MRI, where multiple *k*-space samples are measured together after each RF excitation. Rather than dealing with integer-valued averaging schemes and the $\{\omega_m\}$ variables, we instead optimize the fractional amount of averaging effort $\chi_p \in (0,1)$ that is assigned to the subset Θ_p, under the constraint that $\sum_{p=1}^{P} \chi_p = 1$. Once the χ_p are obtained, ω_m is obtained by setting $\omega_m \approx \chi_{p(m)}/|\Theta_{p(m)}| \sum_{m=1}^{M} \omega_m$, where $p(m)$ returns the index of the subset Θ_p containing the mth data sample. Optimization of the real-valued averaging effort is known as *approximate design*, and avoids the need to solve the difficult integer-programming problems associated with exact designs. Efficient algorithms for the practical discretization of approximate designs are presented, e.g., in Ch. 12 of Ref. [521]. Greedy algorithms also exist for exact designs (see, e.g., Refs. [205, 537, 576]), but are beyond the scope of this work.

We will assume that $\mathbf{d}^q$ can be written as

$$\mathbf{d}^q = \begin{bmatrix} \mathbf{d}_1^q \\ \vdots \\ \mathbf{d}_P^q \end{bmatrix}, \tag{3.32}$$

where each $\mathbf{d}_p^q$ represents the length-$|\Theta_p|$ vector of data samples belonging to the subset Θ_p. With

this representation, the covariance matrix of the $\mathbf{d}^q$ is given by

$$E\left[\mathbf{d}^q\left(\mathbf{d}^q\right)^H\right] \propto \begin{bmatrix} \frac{\sigma_q^2}{\chi_1}\mathbf{I}_{|\Theta_1|} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \frac{\sigma_q^2}{\chi_P}\mathbf{I}_{|\Theta_P|} \end{bmatrix}, \tag{3.33}$$

where $\mathbf{I}_{|\Theta_p|}$ denotes the $|\Theta_p| \times |\Theta_p|$ identity matrix for each p. In addition, the covariance matrix of $\hat{\boldsymbol{\rho}}_q$ is given by

$$\begin{aligned} E\left[\hat{\boldsymbol{\rho}}_q\hat{\boldsymbol{\rho}}_q^H\right] &= E\left[\mathbf{G}\mathbf{d}^q\left(\mathbf{d}^q\right)^H\mathbf{G}^H\right] \\ &= \mathbf{G}E\left[\mathbf{d}^q\left(\mathbf{d}^q\right)^H\right]\mathbf{G}^H \\ &\propto \mathbf{G}\begin{bmatrix} \frac{\sigma_q^2}{\chi_1}\mathbf{I}_{|\Theta_1|} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \frac{\sigma_q^2}{\chi_P}\mathbf{I}_{|\Theta_P|} \end{bmatrix}\mathbf{G}^H. \end{aligned} \tag{3.34}$$

Assuming that $\mathbf{G}$ is also partitioned according to the subsets Θ_p as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \dots & \mathbf{G}_P \end{bmatrix}, \tag{3.35}$$

then the reconstructed image covariance can be simplified as

$$E\left[\hat{\boldsymbol{\rho}}_q\hat{\boldsymbol{\rho}}_q^H\right] \propto \sigma_q^2\sum_{p=1}^{P}\frac{1}{\chi_p}\mathbf{G}_p\mathbf{G}_p^H. \tag{3.36}$$

Our choice of $\{\chi_p\}$ is based on our desire to minimize the covariance of $\hat{\boldsymbol{\rho}}_q$. However, there are generally many different ways of defining an ordering of covariance matrices [521]. For this work, we will say that one covariance matrix is smaller than another if it has a smaller trace, where the trace of a matrix is equal to the sum of its diagonal elements. The trace of the covariance matrix is equal to the sum of the variances of each reconstructed voxel, and is a standard criterion

in optimal experiment design [205, 521, 576].[7] The trace criterion is often also called the *sum-of-squared errors*, the *average variance*, or the *A-optimality* criterion. In the present context, the trace criterion can be expanded as

$$\begin{aligned}\text{Trace}\left(E\left[\hat{\boldsymbol{\rho}}_q\hat{\boldsymbol{\rho}}_q^H\right]\right) &\propto \text{Trace}\left(\sum_{p=1}^{P}\frac{1}{\chi_p}\mathbf{G}_p\mathbf{G}_p^H\right)\\ &= \sum_{p=1}^{P}\frac{1}{\chi_p}\text{Trace}\left(\mathbf{G}_p\mathbf{G}_p^H\right)\\ &= \sum_{p=1}^{P}\frac{1}{\chi_p}\text{Trace}\left(\mathbf{G}_p^H\mathbf{G}_p\right).\end{aligned} \tag{3.37}$$

Note that $\text{Trace}\left(\mathbf{G}_p^H\mathbf{G}_p\right)$ can be expressed as

$$\text{Trace}\left(\mathbf{G}_p^H\mathbf{G}_p\right) = \sum_{k=1}^{|\Theta_p|}\|\mathbf{g}_{pk}\|_{\ell_2}^2, \tag{3.38}$$

where $\mathbf{g}_{pk}$ denotes the kth column of the matrix $\mathbf{G}_p$. As a result of this expression, $\text{Trace}\left(\mathbf{G}_p^H\mathbf{G}_p\right)$ can be computed in a straightforward manner.

3.6.1 Optimal Allocation of Averages

The optimal averaging design problem has a closed form solution when $\mathbf{G}$ is independent of $\{\chi_p\}$. This is made explicit in the following theorem.

Theorem 3.1. *For the optimal averaging problem described in the previous subsection, a closed form solution for the optimal* $\{\chi_p\}$ *is given by*

$$\hat{\chi}_p = \frac{\sqrt{\text{Trace}\left(\mathbf{G}_p^H\mathbf{G}_p\right)}}{\sum_{s=1}^{P}\sqrt{\text{Trace}\left(\mathbf{G}_s^H\mathbf{G}_s\right)}}, \tag{3.39}$$

[7] Introducing a weighting function into the this criterion can also be useful if noise perturbation would more problematic in certain spatial regions of the image than in others. This extension is straightforward, and is omitted from our discussion for simplicity.

for $p = 1, 2, \ldots, P$.

Proof. The optimal solution is derived using Lagrange multipliers, similar to the approach taken in [436, 498]. In particular, define the Lagrangian for the constrained optimization problem as

$$\mathcal{L}\left(\chi_1, \chi_2, \ldots, \chi_P, \lambda\right) = \sum_{p=1}^{P} \frac{1}{\chi_p} \text{Trace}\left(\mathbf{G}_p^H \mathbf{G}_p\right) + \lambda \left(\sum_{p=1}^{P} \chi_p - 1 \right), \tag{3.40}$$

where λ is a Lagrange multiplier that is adjusted to enforce the constraint that $\{\chi_p\}$ should sum to one. Differentiating the Lagrangian with respect to each χ_p for $p = 1, 2, \ldots, P$ and equating with 0, we find that

$$\hat{\chi}_p = \frac{\sqrt{\text{Trace}\left(\mathbf{G}_p^H \mathbf{G}_p\right)}}{\sqrt{\lambda}} \tag{3.41}$$

for each p. Choosing λ such that the constraint is satisfied results in Eq. (3.39). □

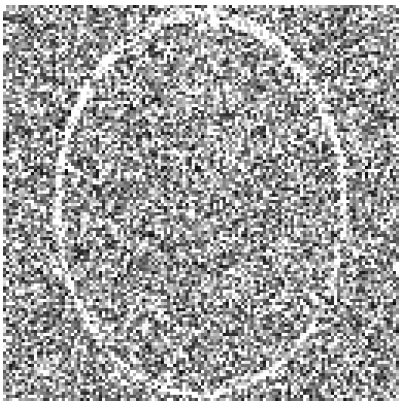
Optimal averaging designs have been considered previously in the context of windowed Fourier reconstruction where each *k*-space sampling location was allowed to be sampled independently (i.e., when $M = P$) [436, 498]. Our results coincide with these existing results for that case, though they can also be applied in much more general linear reconstruction scenarios.

We note that our expression for the optimal averaging design could also be derived using the methods described by Pukelsheim and Torsney [522], who derived closed-form optimal averaging designs for a special class of optimal linear statistical estimation problems. Applying their results to our problem requires the construction of a specialized optimal estimation problem, under which it can be proven that the $\mathbf{G}$ reconstruction matrix is independent of $\{\chi_p\}$. We do not present the details of this construction here, though the construction can be derived by manipulating the results of Zyskind's paper [680], which describes contexts where optimal statistical linear estimation is independent of the averaging scheme. Though the derivations of Pukelsheim and Torsney are quite interesting, their derivations are significantly more complicated than those we presented here, and rely heavily on results from the subdifferential calculus of convex analysis.

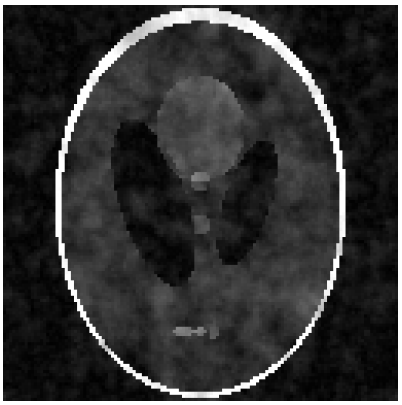
3.6.2 Application Example

We illustrate the benefits of the proposed averaging designs in the context of our proposed linear image reconstruction method with noisy Fourier data. Simulation is performed using the Shepp-Logan phantom shown in Fig. 3.36(a). Data collection was performed by sampling the Fourier transform of this image at the Nyquist rate on a 128×128 Cartesian grid. Simulated white Gaussian noise was added to the data, and several different averaging schemes were applied. With uniform averaging (two averages at each *k*-space location) and conventional unregularized Fourier reconstruction, we obtain the noisy image shown in Fig. 3.36(b). Reconstruction was also performed using the proposed method, where we assume the super-resolution mode of operation where the line-site variables are derived completely from reference images instead of being jointly estimated with the image from the noisy data. In this simulation, we assume that the prior information about the image edge structure is known perfectly, and choose the line-process variables in a binary way such that smoothness is never imposed across edge structures, and is imposed in a uniform way at locations that do not demonstrate edges. As expected, the use of regularization significantly improves SNR, even with uniform averaging (Fig. 3.36(c)). Two different averaging protocols that can be available for typical MRI experiments were also optimized for this regularized reconstruction. In 1D averaging, each of the sampling subsets Θ_p corresponds to one of the 128 different rows of the sampling grid – this corresponds to the standard phase-encoded MRI experiment, in which each line of *k*-space is acquired simultaneously. In 2D averaging, each of the 128^2 sampling locations can be averaged independently. Fig. 3.36(d-e) shows regularized reconstructions, using the optimized averaging patterns from Fig. 3.36(f-g) (shown on the 128×128 sampling grid). Optimized 1D averaging and 2D averaging improve the average variance of the regularized reconstruction by a factor of 1.76 and 2.71, respectively, as compared to uniform averaging, and significantly improve experimental efficiency. In addition, the optimized averaging schemes significantly reduce spatial correlations in the reconstructed noise fields.

(a) True image



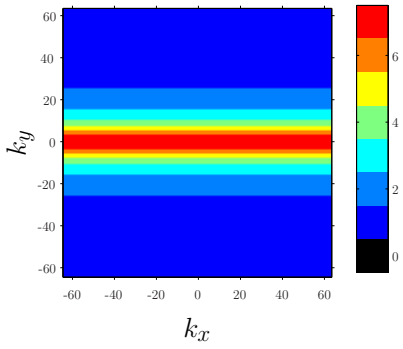
(b) Unregularized



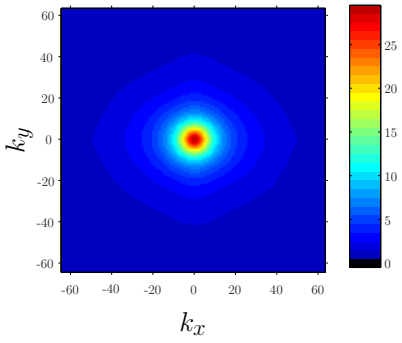
(c) Uniform averaging

(d) Optimized 1D averaging

(e) Optimized 2D averaging



(f) 1D averaging design



(g) 2D averaging design

Figure 3.36: Example of the benefits for using an optimized averaging scheme. (a) Noiseless image. (b) Noisy image with unregularized Fourier reconstruction. (c-e) Regularized reconstructions with different averaging schemes. (f-g) Optimized Fourier averaging patterns. The colorscale indicates the number of averages acquired at each Fourier sampling location.

3.7 Alternative Algorithms

The algorithm we described in Sec. 3.2 for minimizing Eq. (3.8) was based on the multiplicative form of half-quadratic regularization. In practice, however, there are multiple other algorithms that could be applied for minimizing Eq. (3.8), particularly when the $\Psi(\cdot)$ regularization functional is chosen as the Huber function as discussed in Sec. 3.3.3. In this section, we will describe and compare several alternative algorithms that could be used for minimizing Eq. (3.8) with Huber-function regularization. In particular, we will describe the implementation of an additive half-quadratic algorithm in Sec. 3.7.1, the implementation of Nesterov's algorithm in Sec. 3.7.2, and the implementation of a method for accelerating the traditional multiplicative half-quadratic algorithm in Sec. 3.7.3. Comparisons of all the algorithms are given in Sec. 3.7.4.

3.7.1 Additive Half-Quadratic Algorithm

The multiplicative half-quadratic algorithm described in Sec. 3.2 was based on the line-process representation of the potential function $\Psi(\cdot)$ as in Eq. (3.6). In contrast, additive half-quadratic algorithms are based on an alternative representation of $\Psi(\cdot)$ [15, 236, 328, 472]. In particular, it is assumed that $\Psi(\cdot)$ can be expanded as

$$\Psi(t) = \inf_{b} \left\{ (t-b)^2 + U(b) \right\}, \tag{3.42}$$

where $U(\cdot)$ is an appropriate functional. Unfortunately, the traditional additive half-quadratic algorithm does not extend naturally to our proposed prior with shared line processes [472]. However, Wang *et al.* [634] have recently described a new additive half-quadratic algorithm that can be adapted for solving Eq. (3.8). In addition, Wang *et al.* have proven that this approach is globally convergent, and have computed explicit convergence rates.

To explain the approach, we first observe that the expression $\sqrt{\sum_{q=1}^{Q} \beta_q^2 \left|\rho_m^q - \rho_n^q\right|^2}$ appearing

in Eq. (3.8) can be equivalently written as

$$\sqrt{\sum_{q=1}^{Q} \beta_q^2 \left|\rho_m^q - \rho_n^q\right|^2} = \left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}\right\|_{\ell_2}, \tag{3.43}$$

where $\mathbf{D}_{mn}$ is a $Q \times NQ$ weighted finite-differencing operator, and

$$\boldsymbol{\rho}_{\text{tot}} = \begin{bmatrix} \boldsymbol{\rho}^1 \\ \vdots \\ \boldsymbol{\rho}^q \end{bmatrix}. \tag{3.44}$$

Next, assuming that $\Psi(\cdot)$ is the Huber function from Eq. (3.25), we are inspired by Ref. [634] to rewrite $\Psi(\left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}\right\|_{\ell_2})$ as

$$\Psi(\left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}\right\|_{\ell_2}) = \inf_{\mathbf{b}_{mn}\in\mathbb{C}^q} \left\{\left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}} - \mathbf{b}_{mn}\right\|_{\ell_2}^2 + 2\xi\left\|\mathbf{b}_{mn}\right\|_{\ell_2}\right\}. \tag{3.45}$$

It is straightforward to derive that the optimizing value of $\mathbf{b}_{mn}$ in this expression is given by

$$\mathbf{b}_{mn} = \begin{cases} \frac{(\left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}\right\|_{\ell_2} - \xi)}{\left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}\right\|_{\ell_2}} \mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}, & \left\|\mathbf{D}_{mn}\boldsymbol{\rho}_{\text{tot}}\right\|_{\ell_2} > \xi \\ \mathbf{0}, & \text{else.} \end{cases} \tag{3.46}$$

Using this relationship, it is clear that the optimal reconstructed images in Eq. (3.8) can also be obtained by solving

$$\left\{\hat{\boldsymbol{\rho}}^1, \hat{\boldsymbol{\rho}}^2, \ldots, \hat{\boldsymbol{\rho}}^Q\right\} = \underset{\{\boldsymbol{\rho}^1,\boldsymbol{\rho}^2,\ldots,\boldsymbol{\rho}^Q\}}{\arg\min} \inf_{\{\mathbf{b}_{mn}\}} \mathcal{G}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q, \{\mathbf{b}_{mn}\}) \tag{3.47}$$

where

$$\mathcal{G}(\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q, \{\mathbf{b}_{mn}\} = \sum_{q=1}^{Q} \alpha_q^2 \left\| \mathbf{F}_q \boldsymbol{\rho}^q - \mathbf{d}^q \right\|_2^2 + \sum_{n=1}^{N} \sum_{\substack{m \in \Delta_n \\ m>n}} \left(\left\| \mathbf{D}_{mn} \boldsymbol{\rho}_{\text{tot}} - \mathbf{b}_{mn} \right\|_{\ell_2}^2 + 2\xi \left\| \mathbf{b}_{mn} \right\|_{\ell_2} \right). \tag{3.48}$$

The iterative half-quadratic algorithm, when $\Psi(\cdot)$ is the Huber function from Eq. (3.25), proceeds as follows:

1. Set iteration number $j = 0$, and set the estimated image sequence equal to an initial guess $\left\{ \hat{\boldsymbol{\rho}}_{(j)}^1, \hat{\boldsymbol{\rho}}_{(j)}^2, \ldots, \hat{\boldsymbol{\rho}}_{(j)}^Q \right\}$ (e.g., set all voxel coefficient values based on an initial noisy Fourier reconstruction).

2. At the jth iteration, compute the auxiliary vector for each voxel pair as

$$\mathbf{b}_{mn}^{(j)} = \begin{cases} \frac{\left(\left\| \mathbf{D}_{mn} \boldsymbol{\rho}_{\text{tot}}^{(j)} \right\|_{\ell_2} - \xi \right)}{\left\| \mathbf{D}_{mn} \boldsymbol{\rho}_{\text{tot}}^{(j)} \right\|_{\ell_2}} \mathbf{D}_{mn} \boldsymbol{\rho}_{\text{tot}}^{(j)}, & \left\| \mathbf{D}_{mn} \boldsymbol{\rho}_{\text{tot}}^{(j)} \right\|_{\ell_2} > \xi \\ \mathbf{0}, & \text{else.} \end{cases} \tag{3.49}$$

3. Update the image sequence according to

$$\left\{ \hat{\boldsymbol{\rho}}_{(j+1)}^1, \hat{\boldsymbol{\rho}}_{(j+1)}^2, \ldots, \hat{\boldsymbol{\rho}}_{(j+1)}^Q \right\} = \underset{\{\boldsymbol{\rho}^1, \boldsymbol{\rho}^2, \ldots, \boldsymbol{\rho}^Q\}}{\arg\min} \mathcal{G}(\hat{\boldsymbol{\rho}}^1 \hat{\boldsymbol{\rho}}^2, \ldots, \hat{\boldsymbol{\rho}}^Q, \{\mathbf{b}_{mn}^{(j)}\}). \tag{3.50}$$

This optimization problem is separable, such that the solution for each $\hat{\boldsymbol{\rho}}_{(j+1)}^q$ is given by

$$\begin{aligned} \hat{\boldsymbol{\rho}}_{(j+1)}^q &= \arg\min_{\boldsymbol{\rho}^q} \left\{ \alpha_q^2 \left\| \mathbf{F}_q \boldsymbol{\rho}^q - \mathbf{d}^q \right\|_{\ell_2}^2 + \left\| \beta_q \mathbf{D} \boldsymbol{\rho}^q - \mathbf{b}_q^{(j)} \right\|_{\ell_2}^2 \right\} \\ &= \left(\mathbf{F}_q^H \mathbf{F}_q + \frac{\beta_q^2}{\alpha_q^2} \mathbf{D}^H \mathbf{D} \right)^{-1} \left(\mathbf{F}_q^H \mathbf{d}^q + \frac{\beta_q}{\alpha_q^2} \mathbf{D}^H \mathbf{b}_q^{(j)} \right), \end{aligned} \tag{3.51}$$

where $\mathbf{D}$ was previously defined in Sec. 3.2, $\mathbf{b}_q^{(j)}$ is the vector comprised of the qth entries of the $\{\mathbf{b}_{mn}^{(j)}\}$ vectors, and it is assumed that the nullspaces of $\mathbf{F}_q$ and $\frac{\beta_q^2}{\alpha_q^2}\mathbf{D}$ have trivial inter-

section. As in Sec. 3.2, the matrix inversion in this expression can be computed efficiently using iterative algorithms like CG or LSQR. However, it should also be noted that the matrix $\left(\mathbf{F}_q^H \mathbf{F}_q + \frac{\beta_q^2}{\alpha_q^2}\mathbf{D}^H\mathbf{D}\right)$ frequently has special structure that enables even faster inversion. In particular, $\mathbf{F}_q^H \mathbf{F}_q$ is often a circulant matrix when using Cartesian *k*-space sampling, and $\mathbf{D}^H\mathbf{D}$ is also circulant when the finite-differencing scheme assumes periodic boundary conditions. Circulant structure implies that $\left(\mathbf{F}_q^H \mathbf{F}_q + \frac{\beta_q^2}{\alpha_q^2}\mathbf{D}^H\mathbf{D}\right)$ is diagonalized by the unitary DFT matrix, meaning that the matrix inversion in Eq. (3.51) can be computed noniteratively using a small number of FFT operations [207, 625].

4. Increment j. Repeat steps 2 and 3 until convergence is achieved.

Similar to multiplicative half-quadratic optimization, this additive half-quadratic algorithm monotonically reduces the cost function in Eq. (3.8). Moreover, this half-quadratic algorithm has guaranteed global convergence to a global minimum of Eq. (3.8).

In general, it is known that additive half-quadratic algorithms tend to converge more slowly per iteration than multiplicative half-quadratic algorithms [472]. However, when the matrix inversion can exploit circulant structure, the computational effort for each iteration of additive half-quadratic algorithms is significantly smaller than that for multiplicative half-quadratic algorithms. As a result, additive half-quadratic algorithms can lead to significantly reduced total computation times in these cases.

3.7.2 Nesterov's Algorithm

Our description of Nesterov's algorithm follows the description given in Ref. [38]. Nesterov's algorithm is a first-order optimization method with an optimal convergence rate, that can be used to minimize arbitrary smooth convex functions. In particular, assume that we wish to find a solution to

$$\min_{\mathbf{x}\in\mathcal{Q}_p} f(\mathbf{x}), \tag{3.52}$$

with $\mathbf{x} \in \mathbb{R}^N$, where f is a smooth convex function and the primal feasible set $\mathcal{Q}_p$ is convex. The function f is assumed to be differentiable. In addition, the gradient $\nabla f(\mathbf{x})$ is assumed to be Lipschitz such that

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_{\ell_2} \leq L \|\mathbf{x} - \mathbf{y}\|_{\ell_2}, \tag{3.53}$$

where L is some upper bound on the Lipschitz constant. Given two scalar sequences $\{\kappa_k\}$ and $\{\tau_k\}$, Nesterov's algorithm takes the following form:

Initialize $\mathbf{x}_0$. For $k \geq 0$,

1. Compute $\nabla f(\mathbf{x}_k)$.

2. Compute $\mathbf{y}_k$:

$$\mathbf{y}_k = \arg\min_{\mathbf{x} \in \mathcal{Q}_p} \frac{L}{2} \|\mathbf{x} - \mathbf{x}_k\|_{\ell_2}^2 + \langle \nabla f(\mathbf{x}_k), \mathbf{x} - \mathbf{x}_k \rangle, \tag{3.54}$$

   where $< \cdot, \cdot >$ denotes the standard ℓ_2 inner-product for $\mathbb{R}^N$.

3. Compute $\mathbf{z}_k$:

$$\mathbf{z}_k = \arg\min_{\mathbf{x} \in \mathcal{Q}_p} \frac{L}{\Gamma_p} p_p(\mathbf{x}) + \sum_{i=0}^{k} \kappa_i \langle \nabla f(\mathbf{x}_i), \mathbf{x} - \mathbf{x}_i \rangle. \tag{3.55}$$

4. Update $\mathbf{x}_k$:

$$\mathbf{x}_{k+1} = \tau_k \mathbf{z}_k + (1 - \tau_k) \mathbf{y}_k. \tag{3.56}$$

5. Increment k.

Stop when a given termination criterion is valid.

The function $p_p(\mathbf{x})$ is a strongly convex function over $\mathcal{Q}_p$ with convexity parameter Γ_p. Following Ref. [38], we will use

$$p_p(\mathbf{x}) = \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}\|_{\ell_2}^2, \tag{3.57}$$

where $\tilde{\mathbf{x}} \in \mathcal{Q}_p$ is an initial guess of the solution. Note that $\Gamma_p = 1$ in this case.

At each iteration, $\mathbf{y}_k$ is the current guess of the optimal solution. If $\kappa_k = (k+1)/2$ and $\tau_k = 2/(k+3)$, then $\mathbf{y}_k$ has been proven to converge to

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{Q}_p} f(\mathbf{x})$$

with the convergence rate

$$f(\mathbf{y}_k) - f(\mathbf{x}^*) \leq \frac{4Lp_p(\mathbf{x}^*)}{(k+1)^2\Gamma_p}.$$

However, it should be noted that Nesterov's algorithm does not necessarily converge monotonically.

While Nesterov's algorithm is designed for real-valued problems, it is straightforward to write Eq. (3.8) as a real-valued optimization problem with twice as many optimization variables (i.e., one real-valued variable for each of the real and imaginary parts of the complex random variable). Simplifying the algorithm derived using the real-valued problem formulation, it is possible to equivalently derive a complex-valued algorithm that makes use of the same update equations in Eqs. (3.54)-(3.56), and we will focus on this description for the remainder of this subsection. Note that $\mathcal{Q}_p = \mathbb{C}^{QN}$ for the problem in Eq. (3.8), and the complex vectors that we compute in Nesterov's algorithm will all have length QN.

Finalizing the description of Nesterov's algorithm for Eq. (3.8) is a simple matter of specifying the gradient of Eq. (3.8), specifying L, and describing the update equations in Eqs. (3.54) and (3.55):

- For each q, the gradient of Eq. (3.8) with respect to $\boldsymbol{\rho}_q$ is given by the left-hand side of Eq. (3.27). The full gradient of the cost function can be obtained by stacking these individual length-N gradient vectors into a single length-QN vector.

- Applying the triangle inequality to the gradient expression in Eq. (3.27), it can be shown that

an upper bound on the Lipschitz constant for Eq. (3.8) is given by

$$L = \max_q \left\{ 2\alpha_q^2 \left\| \mathbf{F}_q \right\|_2^2 \right\} + \max_q \left\{ \beta_q^2 \right\} 2 \left\| \mathbf{D} \right\|_2^2, \tag{3.58}$$

where $\|\cdot\|_2$ denotes the spectral norm. While the spectral norm is easy to compute analytically for certain matrices with special structure, there also exist fast iterative methods for computing the spectral norm of arbitrary matrices [248].

- The update equation for $\mathbf{y}_k$ in Eq. (3.54) is

$$\mathbf{y}_k = \mathbf{x}_k - \frac{\nabla f\left(\mathbf{x}_k\right)}{L}. \tag{3.59}$$

- The update equation for $\mathbf{z}_k$ in Eq. (3.55) is

$$\mathbf{z}_k = \mathbf{x}_k - \frac{\sum_{i=0}^{k} \kappa_i \nabla f\left(\mathbf{x}_k\right)}{L}. \tag{3.60}$$

3.7.3 Accelerated Multiplicative Half-Quadratic Algorithm

Ramani and Fessler have proposed an accelerated algorithm for solving the matrix inversion in Eq. (3.16) that appears in the multiplicative half-quadratic algorithm [528]. Their method is based on the observation that if we iterate

$$\mathbf{z}_{(k+1)} = \left(\mathbf{L} + \frac{\beta_q^2}{\alpha_q^2} \mathbf{D}^H \mathrm{diag}(\ell_{n,m}^{(j)}) \mathbf{D} \right)^{-1} \left(\mathbf{F}_q^H \mathbf{d}^q + \left(\mathbf{L} - \mathbf{F}_q^H \mathbf{F}_q \right) \mathbf{z}_{(k)} \right) \tag{3.61}$$

over k, then $\mathbf{z}_{(k)}$ converges to the optimal $\hat{\boldsymbol{\rho}}_{(j+1)}^q$ in Eq. (3.16). The matrix $\mathbf{L}$ appearing in this equation is any $N \times N$ Hermitian invertible matrix for which $\mathbf{L} - \mathbf{F}_q^H \mathbf{F}_q$ is positive definite. Ramani and Fessler use the matrix inversion lemma (MIL) to expand the matrix inversion appearing in

Eq. (3.61) as

$$\left(\mathbf{L} + \frac{\beta_q^2}{\alpha_q^2}\mathbf{D}^H \text{diag}(\ell_{n,m}^{(j)})\mathbf{D}\right)^{-1} = \mathbf{L}^{-1} - \mathbf{L}^{-1}\mathbf{D}^H \left(\text{diag}(\ell_{n,m}^{(j)})^{-1} + \mathbf{D}\mathbf{L}^{-1}\mathbf{D}^H\right)^{-1} \mathbf{D}\mathbf{L}^{-1}. \quad (3.62)$$

As a result, the solution to Eq. (3.16) can be obtained by iterating the following procedure over k:

1. Compute

$$\mathbf{b}_{(k)} = \mathbf{L}^{-1}\left(\mathbf{F}_q^H \mathbf{d}^q - \mathbf{F}_q^H \mathbf{F}_q \mathbf{z}_{(k)}\right) + \mathbf{z}_{(k)}. \quad (3.63)$$

2. Use an iterative algorithm like CG or LSQR to solve

$$\left(\text{diag}(\ell_{n,m}^{(j)})^{-1} + \mathbf{D}\mathbf{L}^{-1}\mathbf{D}^H\right)\mathbf{a}_{(k)} = \mathbf{D}\mathbf{b}_{(k)} \quad (3.64)$$

   for the vector $\mathbf{a}_{(k)}$.

3. Set

$$\mathbf{z}_{(k+1)} = \mathbf{b}_{(k)} - \mathbf{L}^{-1}\mathbf{D}^H \mathbf{a}_{(k)}. \quad (3.65)$$

Though this procedure is more complicated than directly applying CG or LSQR to Eq. (3.16), Ramani and Fessler argue that this method will hopefully converge after fewer iterations due to similarities with a preconditioning strategy; see Ref. [528] for further details.

A key component of this algorithm is that $\mathbf{L}$ should be chosen such that it is easily inverted. Ramani and Fessler choose the matrix to take the form $\mathbf{L} = \mathbf{F}_q^H \mathbf{F}_q + c\mathbf{D}^H\mathbf{D}$, where c is a small positive constant (we use $c = 0.001$), and it is assumed that the nullspaces of $\mathbf{F}_q$ and $\mathbf{D}$ have trivial intersection. Note that this matrix has the same form as the matrix appearing in the additive half-quadratic algorithm, and thus is similarly easy to invert using FFTs when data is sampled on a Cartesian *k*-space grid and when the finite-differencing operation assumes periodic boundary conditions.

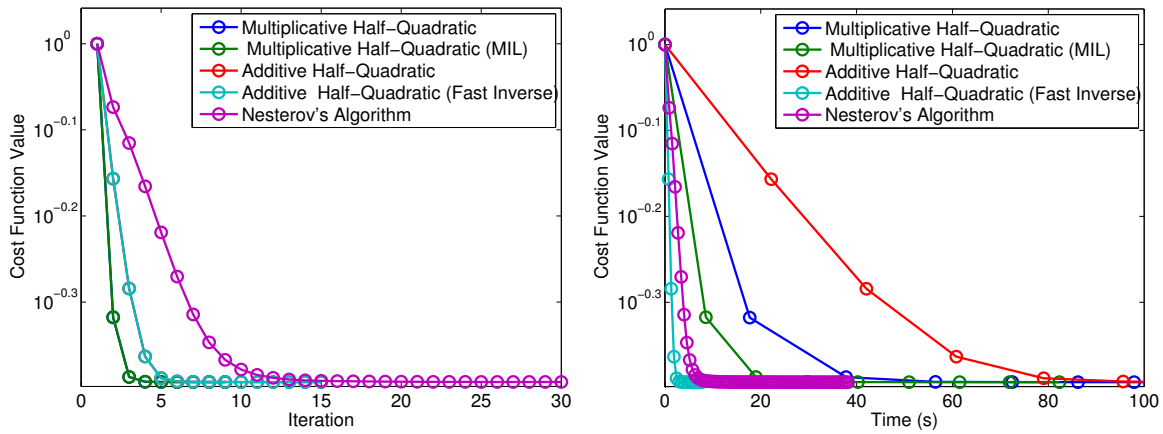


Figure 3.37: Computational performance of the different minimization algorithms with the DW mouse dataset. (a) Convergence as a function of iteration number. (b) Convergence as a function of time.

3.7.4 Algorithm Comparisons

The three different algorithms described in the preceding subsections were all implemented in Matlab (Mathworks, Natick, MA), and compared to the multiplicative half-quadratic algorithm described in Sec. 3.2. Algorithm performance was systematically tested on two different datasets, using a Linux-based workstation with two dual-core Xeon 3.60 GHz processors and 8 GB of RAM.

The first dataset was the DW mouse brain dataset described previously in Sec. 3.4.2. For this dataset, circulant matrix structure can be exploited for the additive half-quadratic algorithm and the multiplicative half-quadratic algorithm with MIL-based acceleration. In our performance evaluation, the additive half-quadratic algorithm was evaluated with both standard matrix inversion and fast matrix inversion. The results of the performance evaluation with this dataset are shown in Fig. 3.37. It is observed that the standard and the MIL-accelerated multiplicative half-quadratic minimization algorithms demonstrate the fastest per-iteration convergence, while Nesterov's algorithm demonstrates the slowest per-iteration convergence. However, due to the low computational effort for each iteration, both Nesterov's algorithm and the additive half-quadratic minimization algorithm using the fast matrix inverse significantly outperform the other algorithms in terms of raw speed. As expected, the MIL-based acceleration of the multiplicative half-quadratic algo-

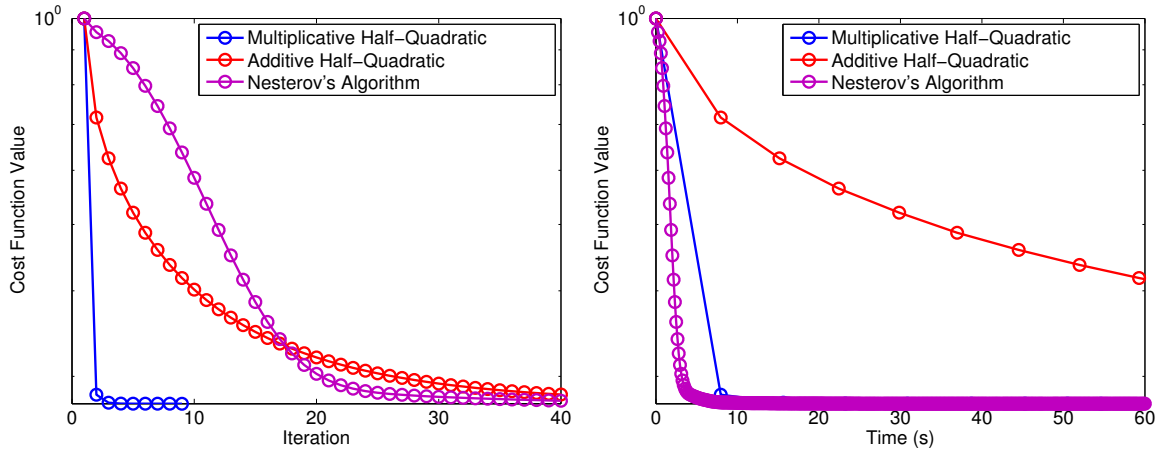


Figure 3.38: Computational performance of the different minimization algorithms with the simulated phantom dataset. (a) Convergence as a function of iteration number. (b) Convergence as a function of time.

rithm leads to reduced computation time. In addition, the multiplicative half-quadratic algorithm is significantly faster than the additive half-quadratic algorithm when the additive half-quadratic algorithm does not take advantage of circulant matrix structure.

The second dataset used was simulated from $Q = 10$ different 96×96 phantom images. Data was simulated based on a real EPI trajectory, which included ramp sampling. Ramp sampling means that circulant matrix structure cannot be used for the MIL-accelerated multiplicative half-quadratic algorithm or the additive half-quadratic algorithm. As a result, these accelerated algorithms were not included in the performance evaluation. The results of this experiment are shown in Fig. 3.38. As before, the multiplicative half-quadratic algorithm demonstrated the fastest per-iteration convergence. In addition, Nesterov's algorithm maintained a clear advantage in terms of raw speed.

This performance evaluation indicates that alternative algorithms can be preferable to the standard multiplicative half-quadratic algorithm for minimizing Eq. (3.8). Nesterov's algorithm is particularly attractive, since it can be applied for arbitrary *k*-space sampling patterns and is consistently fast. When circulant structure can be exploited during the iterative procedure, the additive half-quadratic algorithm also presents attractive features. It should be noted, however, that these

algorithms could perform differently when optimization is performed on different computational platforms or when using different datasets, and that the choice of algorithm should ideally be made on a case-by-case basis. We also remark that fast algorithms for regularized MR imaging methods continue to be developed, and new emerging algorithms could possibly outperform the algorithms we have evaluated here. A particularly promising class of new algorithms is based on the augmented Lagrangian formulation [529].

3.8 Conclusion and Summary

In this chapter, we have described and analyzed a new approach to reconstructing a series of correlated images. We formulated an edge preserving prior based on shared non-interacting line processes, and showed that the solution to the resulting PML optimization problem could be found efficiently using fast algorithms. In addition, we demonstrated that the method is easily characterized when reconstruction parameters are chosen appropriately, and this enables the user to directly control the trade-off between resolution and SNR. We performed an analysis of this trade-off and found that regularized reconstruction of high-resolution data was more efficient than simple reconstruction of averaged low-resolution data. Finally, we demonstrated the potential of the proposed approach with a series of simulation and experimental examples.

Chapter 4

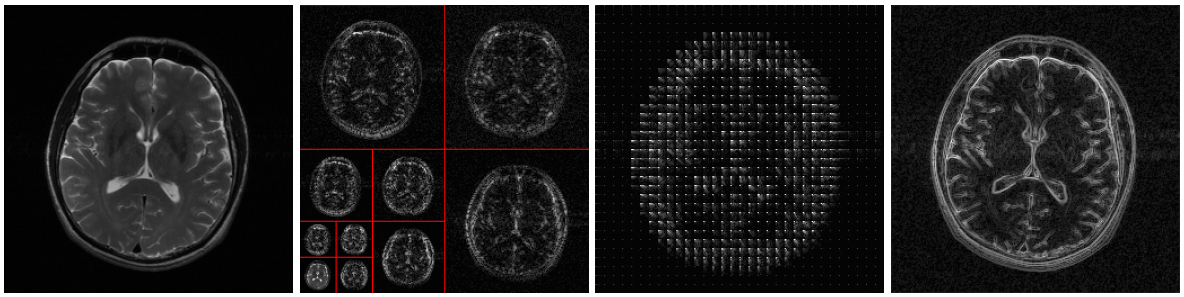
Compressed-Sensing MRI with Random Encoding

In this chapter, we address the problem of reconstructing MR images from highly-undersampled high-SNR data.[1] In particular, we make use of a relatively new approach to reconstructing signals from limited data known as compressed sensing/compressive sampling (CS) [108, 109, 115, 116, 177, 206, 622]. CS has generated significant interest in the signal processing community because of its potential to leverage signal sparsity to enable robust signal reconstruction from much fewer data samples than would be suggested by conventional sampling theory.[2] In the context of MRI, this type of undersampling could have the potential to significantly accelerate imaging experiments.

As described in Sec. 2.2, the necessary sampling density requirements for conventionally-reconstructed Fourier-encoded MRI are dependent on the spatial support of the imaging subject. In particular, images with small support can be reconstructed with fewer Fourier samples (for a given spatial resolution) than images with larger support, and as a result, strong prior information about the image support can be used to accelerate data acquisition [6, 92, 233, 320, 426, 464, 511, 554, 571, 610]. Intuitively, the necessary sampling rate in CS is similarly governed by prior information on the limited support of the image. However, rather than using precise information about the known support of the image in the spatial domain, CS assumes only that the image will

[1] Some of the text and figures in this chapter have been previously published in [273], and are copyright 2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

[2] CS is a rapidly developing field with a rapidly developing body of literature. Due to this rapid development, it is common for new CS results to be distributed online prior to journal publication. A repository of CS preprints and papers is currently maintained by the Digital Signal Processing group at Rice University (`http://dsp.rice.edu/cs`, and is regularly updated with new CS-related manuscripts. The Nuit Blanche blog (`http://nuit-blanche.blogspot.com`) is another good resource, and provides daily updates on CS-related preprints, papers, and ideas.



(a) Original Image (b) Wavelet Transform (c) DCT Transform (d) Finite Differences

Figure 4.1: Typical medical images have sparse representation in appropriately chosen transform domains. For example, the (a) brain image has relatively few significant coefficients in the (b) wavelet (the coefficients shown here were calculated using a Daubechies-4 wavelet transform) and (c) discrete cosine transform (DCT) domains. (d) In addition, due to the piecewise-smooth nature of this image, it also has a small number of significant coefficients using a transform that computes finite differences between adjacent pixels.

be sparse (i.e., the signal has few non-zero coefficients) or compressible (i.e., the signal has few non-negligible coefficients) in a known transform domain. And, unlike support-limited reconstruction, traditional CS does not assume prior knowledge regarding the exact locations of the non-zero transform-domain coefficients;[3] instead, the theoretical results indicate that robust reconstructions are possible even when the transform-domain support is estimated jointly with the image from sparsely-sampled data.

It is well known that natural images, such as those seen in MRI, are highly compressible in appropriately-chosen transform domains (frequently, this is the result of the images having piecewise-smooth structure [170,196,429,484,658]). The transform compressibility of MR images is illustrated with a typical brain image in Fig. 4.1. As a result of image compressibility, MRI has been viewed as a promising application for CS, and several MRI reconstruction schemes inspired by CS theory have been reported in the literature (see, e.g., Refs. [11, 60, 228, 240, 317, 350, 362, 397, 420, 421, 557, 568, 608, 668]). In addition, sparsity-exploiting reconstruction methods for limited-data scenarios existed in the MR literature for a long time, outside of the context of the new CS theory (e.g., Refs. [34, 78, 82, 264, 401, 439, 500, 505, 508, 553, 578, 662]).

[3] Fast MRI acquisition for the case where transform-domain support is known *a priori* has been explored previously [119–121, 495, 619, 678]. In addition, some recent developments in CS have addressed the case where the image support is partially known, though this will not be a focus of this chapter; see, e.g., Refs. [336, 620, 621].

The theoretical CS literature typically assumes finite-dimensional linear signal models (cf. the discussion of such models in MRI from Sec. 2.2.3), with the data acquisition model given by

$$\mathbf{d} = \mathbf{E}\boldsymbol{\rho} + \boldsymbol{\eta}, \tag{4.1}$$

where $\boldsymbol{\rho}$ is a length-N signal vector of interest, $\mathbf{d}$ is a length-M data vector, $\mathbf{E}$ is an $M \times N$ encoding matrix with $M \ll N$, and $\boldsymbol{\eta}$ is a length-M noise vector. Since $M < N$, the matrix $\mathbf{E}$ has a non-trivial nullspace, and there are an infinite number of possible reconstructions that are all maximally consistent with the measured data (cf. Sec. 2.2); as result, the inverse problem is ill-posed, and additional constraints must be imposed for there to be a unique reconstruction.

CS makes two key assumptions to ensure that the inverse problem becomes well-posed: (1) the signal vector $\boldsymbol{\rho}$ is sparse or compressible in a given linear transform domain, and (2) the observation matrix $\mathbf{E}$ satisfies certain mathematical conditions with respect to this transformation. Let $\boldsymbol{\Psi}$ be a sparsifying transform matrix such that the vector

$$\mathbf{c} = \boldsymbol{\Psi}\boldsymbol{\rho} \tag{4.2}$$

is known to be sparse or compressible. The various existing CS reconstruction procedures often find estimates $\hat{\boldsymbol{\rho}}_{\mathrm{CS}}$ of $\boldsymbol{\rho}$ by solving

$$\text{minimize } R_{\mathrm{s}}\left(\boldsymbol{\Psi}\hat{\boldsymbol{\rho}}_{\mathrm{CS}}\right) \text{ subject to } \|\mathbf{E}\hat{\boldsymbol{\rho}}_{\mathrm{CS}} - \mathbf{d}\|_{\ell_2}^2 \leq \varepsilon^2, \tag{4.3}$$

where $R_{\mathrm{s}}(\cdot)$ is a functional that promotes sparsity, and the parameter ε controls the allowed level

of data discrepancy (usually chosen based on an estimate of the noise variance[4]). There are several different choices for $R_{\mathrm{s}}(\cdot)$ that have been proposed in the literature [607]. The most intuitive choice is to use $R_{\mathrm{s}}(\cdot) = \|\cdot\|_{\ell_0}$, where the ℓ_0-"norm"[5] counts the number of non-zero components of its argument. Use of the ℓ_0-norm is desirable in that it results in a solution that has maximum sparsity among all possible solutions that are consistent with the data. However, using the ℓ_0-norm has one major limitation: the resulting optimization problem is non-convex and very difficult to solve (i.e., it is NP-hard [466]). While various greedy algorithms exist for ℓ_0-norm minimization, another common approach is to choose an $R_{\mathrm{s}}(\cdot)$ that makes optimization of Eq (4.3) more tractable. In particular, a common choice is the ℓ_1-norm (previously defined in Eq. (2.34)), resulting in

$$\text{minimize } \|\mathbf{\Psi}\hat{\boldsymbol{\rho}}_{\mathrm{CS}}\|_{\ell_1} \text{ subject to } \|\mathbf{E}\hat{\boldsymbol{\rho}}_{\mathrm{CS}} - \mathbf{d}\|_{\ell_2}^2 \leq \varepsilon^2. \tag{4.4}$$

The ℓ_1 norm is the tightest convex relaxation of the ℓ_0 norm, and its convexity implies that global optimization of Eq. (4.4) can be achieved through standard convex programming methods. In addition, there are many situations in which the solution to the ℓ_1-norm minimization problem is equivalent or nearly-equivalent to the solution of the ℓ_0-norm minimization problem [178, 179]. The fact that ℓ_1-norm minimization frequently leads to sparse solutions can be explained using geometric arguments [291, 600], as illustrated with the example shown in Fig. 4.2.

The accuracy of CS reconstruction using Eq. (4.4) can be guaranteed if $\mathbf{E}$ and $\mathbf{\Psi}$ satisfy certain

[4]The constraint that $\|\mathbf{E}\hat{\boldsymbol{\rho}}_{\mathrm{CS}} - \mathbf{d}\|_{\ell_2}^2$ is no larger than an upper bound of the total noise $\|\mathbf{E}\boldsymbol{\rho} - \mathbf{d}\|_{\ell_2}^2 = \|\boldsymbol{\eta}\|_{\ell_2}^2$ is frequently called the *discrepancy principle* [48, 207, 289, 625], and is a common method used for choosing regularization parameters in ill-posed problems. When an upper bound on $\|\boldsymbol{\eta}\|_{\ell_2}^2$ is estimated well, using the discrepancy principle can help ensure that the reconstructed solution $\hat{\boldsymbol{\rho}}_{\mathrm{CS}}$ will have similar level of data mismatch to the true unknown signal that we are trying to reconstruct. In CS theory, use of the discrepancy principle is also motivated by the fact that reconstruction characterizations can be derived that can guarantee the quality of reconstructed answers in certain cases. For white complex Gaussian noise with variance σ^2, note that $\frac{2}{\sigma^2}\|\boldsymbol{\eta}\|_{\ell_2}^2$ is chi-squared distributed with $2M$ degrees of freedom [73]. As a result, $\|\boldsymbol{\eta}\|_{\ell_2}^2$ has a mean of $M\sigma^2$, and a standard deviation of $\sigma\sqrt{2M}$. Due to the central limit theorem, the chi-squared distribution is approximately Gaussian when M is large (e.g., when $M > 50$ [73]). Thus, the distribution is also peaked about its mean when M is large, such that $\|\boldsymbol{\eta}\|_{\ell_2}^2 \approx \sigma^2 M$ with high probability.

[5]Strictly speaking, this functional is not a true norm, because it does not satisfy the triangle inequality.

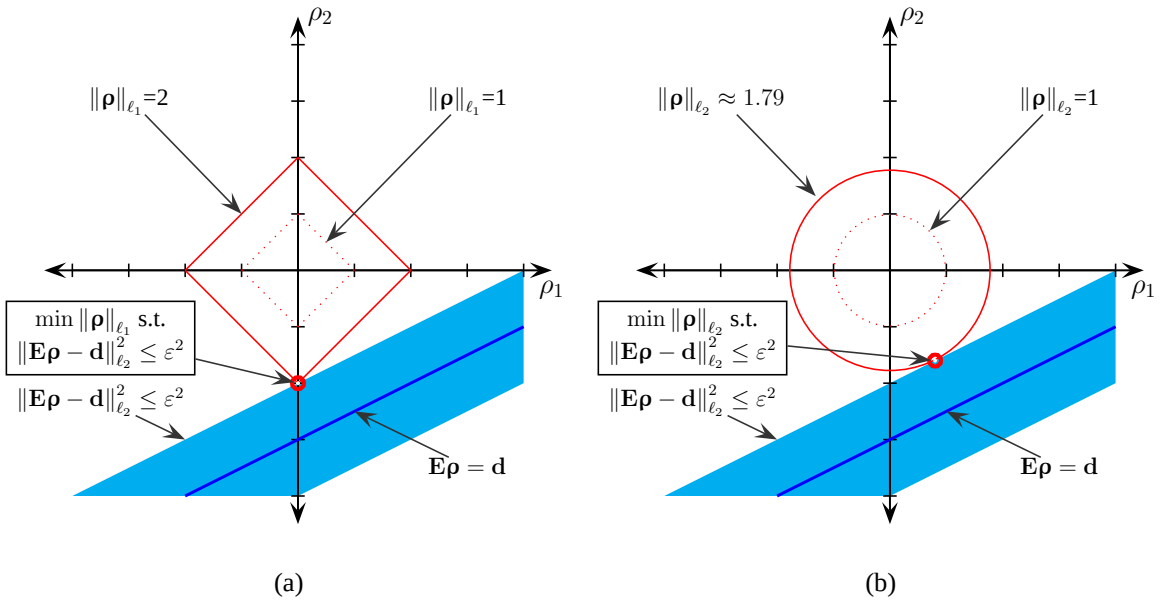


Figure 4.2: Illustrative problem geometries for (a) ℓ_1-norm minimization and (b) ℓ_2-norm minimization. The blue lines correspond to the linear variety of points satisfying $\mathbf{E}\boldsymbol{\rho} = \mathbf{d}$, with $\mathbf{E} = [-1 \;\; 2]$ and $\mathbf{d} = -2$. The shaded light-blue regions correspond to the sets of points satisfying $\|\mathbf{E}\boldsymbol{\rho} - \mathbf{d}\|_{\ell_2}^2 \leq \varepsilon^2$, with $\varepsilon^2 \approx 0.894$. The red curves in (a) and (b) are isocontours of the ℓ_1- and ℓ_2-norms, respectively. The optimal solution to the ℓ_1-norm minimization problem can be obtained by finding the smallest value κ such that the curve defined by $\|\boldsymbol{\rho}\|_{\ell_1} = \kappa$ intersects the set of points satisfying $\|\mathbf{E}\boldsymbol{\rho} - \mathbf{d}\|_{\ell_2}^2 \leq \varepsilon^2$. The solution to the ℓ_1-norm minimization is sparse in this case, since only one of the entries of the optimal $\boldsymbol{\rho}$ is nonzero. This sparsity is the result of the fact that the isocontours of the ℓ_1-norm are "pointy," extending further along the coordinate axes than along other directions. Notice that the ℓ_2-norm isocontours are spherical rather than pointy. As a result, the ℓ_2-norm does not prefer solutions along particular orientations, and ℓ_2-norm minimization generally does not yield sparse solutions.

mathematical conditions.[6] For example, consider the case where $\mathbf{\Psi}$ is a square, invertible matrix, and define $\mathbf{\Phi} = \mathbf{E\Psi^{-1}}$.[7] In this case, the performance of CS reconstruction can be guaranteed if $\mathbf{\Phi}$ satisfies appropriate restricted isometry properties (RIPs) [103, 108, 110, 116, 157, 219], incoherence properties [41, 104, 107, 181, 182], or nullspace properties (NSPs) [144, 183, 254]. While NSPs provide necessary and sufficient conditions for accurate CS in the absence of noise, this chapter will focus on RIPs, which can provide some of the strongest existing performance guarantees for stable and accurate reconstruction in the presence of noise [103, 116, 219]. To define the RIP, first let α_s and β_s denote the largest and smallest coefficients, respectively, such that

$$\alpha_s \|\mathbf{x}\|_{\ell_2}^2 \leq \|\mathbf{\Phi x}\|_{\ell_2}^2 \leq \beta_s \|\mathbf{x}\|_{\ell_2}^2 \tag{4.5}$$

is true for all vectors $\mathbf{x}$ with at most s non-zero entries. A simple generalization of the results in Ref. [110] yields that the best possible[8] *restricted isometry constant* of order s is given by

$$\delta_s = \frac{\beta_s - \alpha_s}{\beta_s + \alpha_s}. \tag{4.6}$$

The performance guarantees for CS reconstruction with Eq. (4.4) improve as δ_s gets smaller. For example, Candès [110] shows that if $\delta_{2s} < \sqrt{2} - 1 \approx 0.414$ and in the absence of noise, the solution to Eq. (4.4) with $\varepsilon = 0$ perfectly recovers any sparse vector with fewer than s non-zeros. In the more general setting with noise and a compressible $\mathbf{c}$, a trivial modification of the results in Ref. [110] shows that if $\delta_{2s} < \sqrt{2}-1$ and if the noise obeys $\|\boldsymbol{\eta}\|_{\ell_2}^2 \leq \varepsilon^2$, then the CS reconstruction

[6] Performance guarantees for other CS reconstruction formulations also exist (see references in Ref. [607]). These guarantees are sometimes better than those that exist for ℓ_1-minimization (e.g., [132, 133, 220, 548, 609, 622, 652]), though the methods that do better than ℓ_1-norm minimization typically require additional prior information and/or nonconvex optimization.

[7] Discussion of the case where $\mathbf{\Psi}$ is a more general matrix can be found in Ref. [111].

[8] The restricted isometry constant as defined in Ref. [110] is the smallest number δ_s such that Eq. (4.5) holds with $\alpha_s = 1 - \delta_s$ and $\beta_s = 1 + \delta_s$ for all vectors $\mathbf{x}$ with at most s non-zero entries. This definition of δ_s is not invariant with respect to rescaling of $\mathbf{\Phi}$, despite the fact that the solution to Eq. (4.4) would remain exactly the same (other than scaling) under this problem transformation. Equation (4.6) represents the minimal value of δ_s over the set of all possible rescalings of $\mathbf{\Phi}$.

$\hat{\mathbf{c}}_{\text{cs}} = \boldsymbol{\Psi}\hat{\boldsymbol{\rho}}_{\text{CS}}$ satisfies

$$\underbrace{\left\|\hat{\mathbf{c}}_{\text{cs}} - \mathbf{c}\right\|_{\ell_2}}_{\text{Reconstruction Error}} \leq \underbrace{C_0 s^{-1/2}\left\|\mathbf{c} - \mathbf{c}_s\right\|_{\ell_1}}_{\text{Compression Error}} + \underbrace{C_1 \varepsilon/\sqrt{\xi}}_{\text{Noise Error}}, \tag{4.7}$$

where $\mathbf{c}_s$ is the optimal s-term approximation of $\mathbf{c}$ [110], $\xi = \frac{1}{2}(\alpha_s + \beta_s)$, and C_0 and C_1 are dependent on δ_{2s}. Recent improvements on this result have been made that provide similar guarantees for stable and accurate reconstruction, but are valid under the weaker conditions that $\delta_s < 0.307$ [103] or $\delta_{2s} < 0.4734$ [219].[9]

For an arbitrary pair of matrices $\mathbf{E}$ and $\boldsymbol{\Psi}$, it is often computationally infeasible to calculate practically-useful guarantees on the quality and robustness of the CS reconstruction with Eq. (4.4). As a result, joint optimization of $\mathbf{E}$ and $\boldsymbol{\Psi}$ for optimal performance in the context of specific reconstruction scenarios is an even more challenging problem. Therefore, a common practice has been to construct CS matrices based on randomization, since certain randomized data acquisition schemes have a high probability of possessing good CS properties [108,116,184,545], and because the known deterministic constructions of CS matrices have significantly worse CS properties than randomized matrices [169].[10] Notably for Fourier-encoded MRI, if $\boldsymbol{\Psi}$ is an identity matrix and M and N are large, then CS reconstruction is guaranteed to be robust with high probability if $\mathbf{E}$ is a randomly undersampled discrete Fourier transform operator [116, 545].[11] However, Fourier encoding is not necessarily well-suited to CS reconstruction with arbitrary $\boldsymbol{\Psi}$. For example, Lustig *et al.* [420] have demonstrated that using slice-selective excitation as an additional encoding mechanism can improve CS reconstruction in 3D imaging with compressibility in a wavelet basis. As

[9] Most exisiting RIP-based results guaranteeing the performance of CS have assumed that all matrices and vectors involved are purely real-valued. For MRI, however, the complex case is of more practical interest. Foucart's recent work has explicitly demonstrated that the same kinds of performance guarantees are also valid for complex vectors and matrices [219].

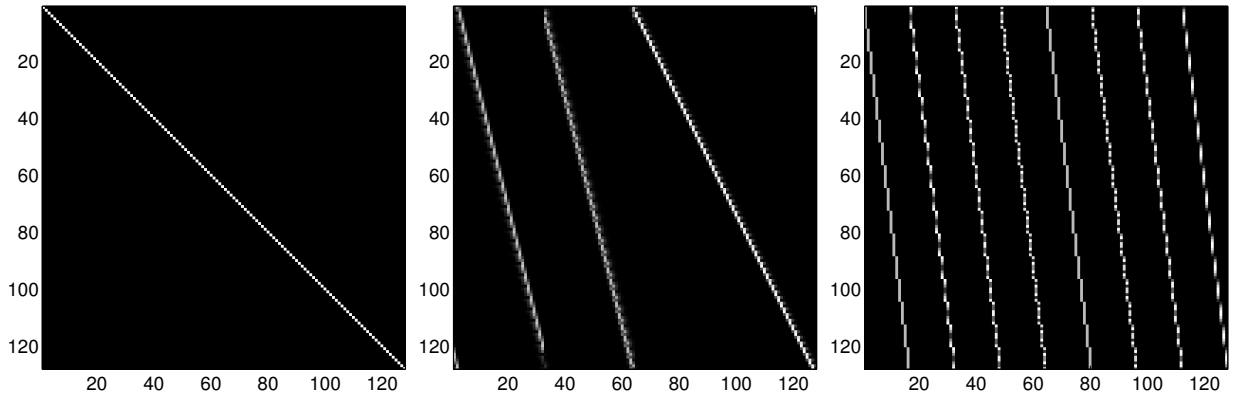
[10] Note, however, that if performance requirements are relaxed such that sparse-recovery is guaranteed for *most* sparse vectors instead of *all* sparse vectors $\mathbf{c}$, then certain high-quality deterministic constructions are also known to exist [105].

[11] Interestingly, random and/or nonuniform undersampled Fourier encoding had been proposed much earlier for certain MR imaging scenarios with sparse or otherwise simplified image models, unrelated to CS-based guarantees [34,82,355,439,500,505,508,553].

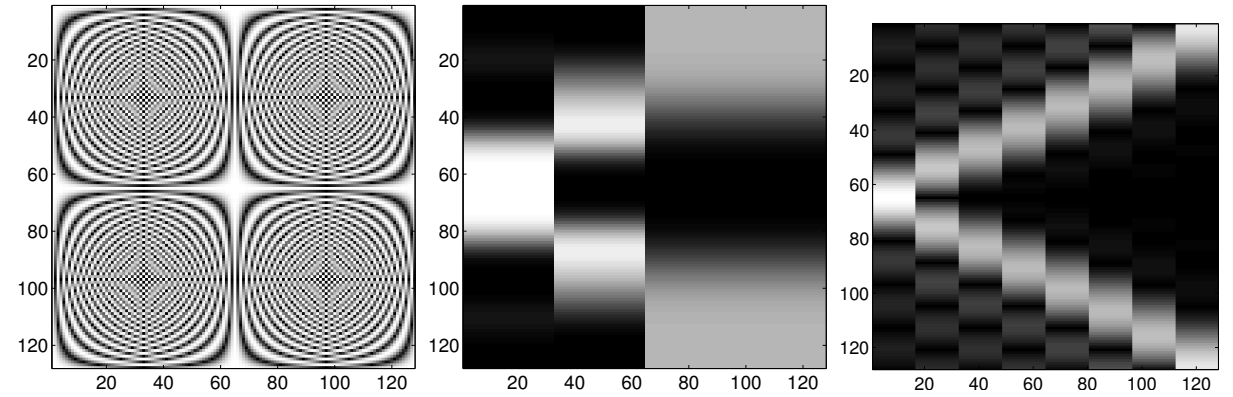
a result, the use of other non-Fourier encoding schemes for CS-MRI could also potentially yield benefits.

In this work, we investigate the use of random encoding for CS-MRI. This choice is motivated by the insight from the CS literature that if the entries of $\mathbf{E}$ are chosen independently from a Gaussian distribution and M and N are large, then there is a high probability that the RIP will be satisfied for *any* unitary matrix $\mathbf{\Psi}$ [108]. In addition, random Gaussian $\mathbf{E}$ matrices have been shown to be nearly optimal with respect to other encoding schemes for CS, and can be obtained without significant computational effort. This leads Candès and Tao to describe Gaussian measurements as a "universal encoding strategy" [109]. Many useful transforms for compressing medical images are unitary, including the identity transform, various wavelet transforms, the discrete cosine transform (DCT), and the discrete Fourier transform. Recent results also suggest that Gaussian measurements can often lead to good CS reconstructions even when $\mathbf{\Psi}$ is not unitary [111]. An objective of this chapter is to evaluate the utility of random encoding for practical MR imaging problems.

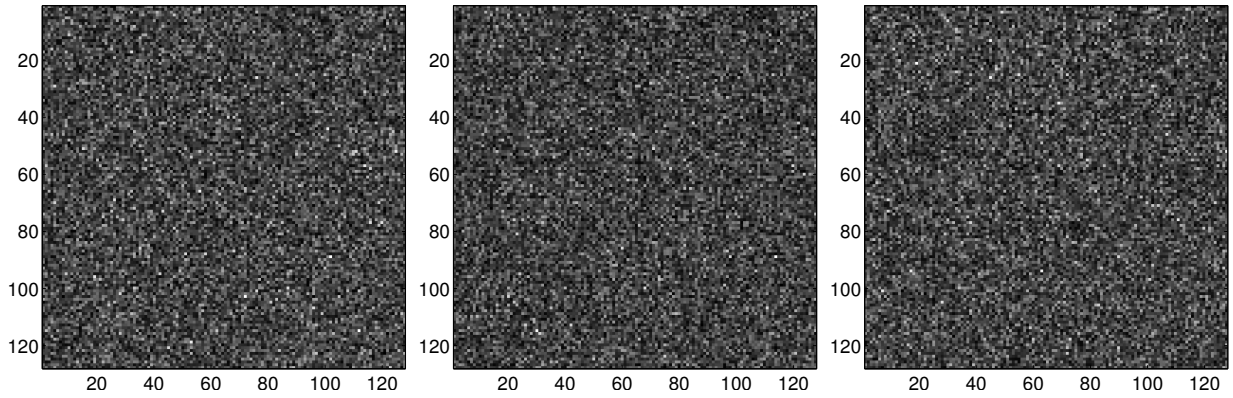
The use of a random matrix $\mathbf{E}$ instead of a Fourier matrix can have a dramatic effect on the structure of the $\mathbf{\Phi}$ matrix. This is illustrated in Fig. 4.3, which shows $\mathbf{\Phi}$ matrices for several unitary bases when the matrix $\mathbf{E}$ is either a "fully-sampled" (i.e., square and invertible) Fourier matrix or a "fully-sampled" random Gaussian matrix. When $\mathbf{\Psi}$ corresponds to the identity transform, then the use of Fourier encoding causes the $\mathbf{\Phi}$ matrix to have rows with significant energy in every entry. Practically, this means that every measurement contains significant information about every transform coefficient. However, this is not the case with the wavelet and block DCT transforms, where the signal energy for each row is significantly more concentrated. A concentration of signal energy in the Fourier domain is expected for these transforms: wavelet basis functions are well-known to be highly localized in both the spatial domain and the Fourier domain [429], and the cosine functions used to form the block DCT basis functions are also naturally localized in the Fourier domain. The consequence of this concentration is that low-frequency Fourier measurements provide very strong information about the transform coefficients corresponding to the low-resolution image ba-



(a) $\mathbf{\Psi}^{-1}$ for the identity transform (b) $\mathbf{\Psi}^{-1}$ for the wavelet transform (c) $\mathbf{\Psi}^{-1}$ for the DCT



(d) $\mathbf{F\Psi}^{-1}$ for the identity transform (e) $\mathbf{F\Psi}^{-1}$ for the wavelet transform (f) $\mathbf{F\Psi}^{-1}$ for the DCT



(g) $\mathbf{G\Psi}^{-1}$ for the identity transform (h) $\mathbf{G\Psi}^{-1}$ for the wavelet transform (i) $\mathbf{G\Psi}^{-1}$ for the DCT

Figure 4.3: The effects of different encoding strategies on the $\mathbf{\Phi}$ matrix. The top row shows the magnitudes of the 128×128 $\mathbf{\Psi}^{-1}$ matrices for (a) the identity transform, (b) the 1D Daubechies-4 wavelet transform, and (c) the 1D block DCT. The second row shows the magnitudes of $\mathbf{F\Psi}^{-1}$ for these three transforms, where $\mathbf{F}$ is the 128×128 DFT matrix (the magnitude of $\mathbf{F\Psi}^{-1}$ for the identity transform is a matrix that has every entry equal to 1, which is not very interesting to look at; as a result, (d) shows the magnitude of the real part of $\mathbf{F\Psi}^{-1}$). The bottom row shows the magnitudes of $\mathbf{G\Psi}^{-1}$ for these three transforms, where $\mathbf{G}$ is a 128×128 random complex Gaussian matrix.

sis functions, but relatively little information about the transform coefficients corresponding to the high-resolution basis functions. Similarly, high-frequency Fourier measurements contain limited information about the transform coefficients corresponding to low-resolution basis functions. This type of structure indicates that if CS-MRI with Fourier encoding is used for a wavelet- or DCT-compressible image, the accuracy at which high- and low-resolution image features can be reconstructed will likely be limited by the number of acquired high- and low-frequency *k*-space measurements, respectively. In contrast to the case with Fourier measurements, none of the $\mathbf{\Phi}$ matrices corresponding to a Gaussian $\mathbf{E}$ matrix have rows with concentrated energy. As a result, each row of the $\mathbf{\Phi}$ matrix for a random Gaussian $\mathbf{E}$ can simultaneously encode information about image features at every resolution scale, which suggests that CS-MRI with random encoding could have very different reconstruction characteristics compared to CS-MRI with Fourier encoding. This hypothesis is confirmed by our empirical results, which are described later in this chapter.

A preliminary account of this work was first presented in Ref. [275], and related work on CS-MRI with random and other non-Fourier encoding has subsequently been performed by other authors [398, 523, 566, 632, 649, 656]. While we focus here on MRI, the results we present could also provide insight into the utility of similar randomized encoding schemes with CS reconstruction in the context other imaging modalities, including coded-aperture computed tomography [494], radio interferometry [523], and coded-aperture or moving random exposure optical imaging [188, 435, 572].

4.1 CS-MRI with Random Encoding

The proposed random encoding scheme is achieved using tailored spatially-selective RF excitation pulses. Non-Fourier encoding schemes using selective excitation have been investigated previously (see, e.g., Refs. [83, 151, 342, 455, 496] and the discussion in Sec. 2.1.2), though outside of the context of CS-MRI. In contrast to these previous works, we use selective excitation to implement an encoding scheme similar to the "universal" encoding suggested by the CS literature [109].

MR data acquisition with slice-selective excitation (cf. Sec. 2.1.2) and uniform receive coil sensitivity[12] can be written generally as

$$d_m = \int w_m(\mathbf{x})\,\rho(\mathbf{x})\exp(-\imath 2\pi \mathbf{k}_m \cdot \mathbf{x})d\mathbf{x} + \eta_m, \quad m = 1,\ldots,M, \tag{4.8}$$

where $w_m(\mathbf{x})$ represents the effects of RF excitation for the mth sample. In conventional Fourier encoding, the RF excitation profile is designed in such a way that $w_m(\mathbf{x})$ is a constant. In this work, we allow $w_m(\mathbf{x})$ to vary with m and $\mathbf{x}$, as described in the next two subsections.

To connect with the CS formulation in Eq. (4.1), we first approximate Eq. (4.8) using a discrete voxel-based image model as in Eq. (2.30). Under this parameterization, Eq. (4.8) can be written as Eq. (4.1), with the $M \times N$ matrix $\mathbf{E}$ defined as

$$[\mathbf{E}]_{mn} = \int w_m(\mathbf{x})\,\phi(\mathbf{x}-\mathbf{x}_n)\exp(-\imath 2\pi \mathbf{k}_m \cdot \mathbf{x})d\mathbf{x}. \tag{4.9}$$

In the following two subsections, we describe two schemes for designing $\mathbf{E}$ to achieve random encoding.

4.1.1 Ideal Random Encoding

Ideally, we would like to have excitation profiles such that the matrix entries in Eq. (4.9) are drawn independently from a Gaussian distribution. One way to achieve this would be to have $w_m(\mathbf{x})$ be approximately constant within each voxel to minimize intravoxel signal dephasing, and choose the value of $w_m(\mathbf{x})$ at the center of each voxel randomly from a complex[13] Gaussian distribution.

[12] In principle, $w_m(\mathbf{x})$ could also be used to absorb the effects of using a receive coil with spatially non-uniform sensitivity, and this would be important to do when doing parallel imaging with an array of receiver coils (e.g., as in Ref. [519]). To simplify the notation and discussion, we assume for this chapter that only a single receiver coil is used for data acquisition and that any non-uniformity in the receive B_1 field is treated as a part of the image function $\rho(\mathbf{x})$.

[13] We choose the complex Gaussian distribution because complex random Gaussian matrices typically have better conditioning than real random Gaussian matrices [193].

Mathematically, this excitation profile can be described, in the 2D imaging case, as

$$w_m(x,y) = \sum_{q=-Q/2}^{Q/2-1} \sum_{p=-P/2}^{P/2-1} \gamma_{qpm} \Pi(x-q)\,\Pi(y-p), \tag{4.10}$$

where $\Pi(\cdot)$ is a rectangular window function with unit width, and each γ_{qpm} is a realization of a complex Gaussian random variable. In Eq. (4.10), we have assumed without loss of generality that the image voxel positions $\{\mathbf{x}_n\}_{n=1}^{N}$ lie on a $Q \times P$ Cartesian grid, normalized so that the distance between adjacent voxels is 1. With excitation profiles generated according to Eq. (4.10) and if $\phi(\mathbf{x})$ is chosen to be a Dirac delta function, the matrix $\mathbf{E}$ will have the desired Gaussian distribution for any $\mathbf{k}_m$.

However, there are a couple of practical limitations to implementing this scheme with a distinct excitation profile for each measurement sample. First, making $w_m(x,y)$ distinct for each m would mean that only a single sample is obtained for each excitation, thereby wasting the free precession period that is used for data acquisition in conventional Fourier schemes. Second, high-resolution multidimensional excitation profiles are difficult to achieve using current excitation hardware, due to practical constraints on pulse length. We next describe a practical alternative to this ideal random encoding scheme.

4.1.2 Practical Implementation

To make random encoding more practical, we consider a modification based on the conventional spin-warp imaging sequence shown in Fig. 4.4(a). In spin-warp imaging, each excitation is followed by phase encoding, and a full frequency-encoded line passing through the center of *k*-space is read out after the signal is refocused by a 180° pulse. In this manner, Cartesian coverage of *k*-space is obtained, with the total number of excitations given by the total number of phase encodings.

Our proposed modification of conventional spin-warp imaging replaces phase encoding by random 1D spatially-selective excitation, and is shown in Fig. 4.4(b). In particular, assuming that

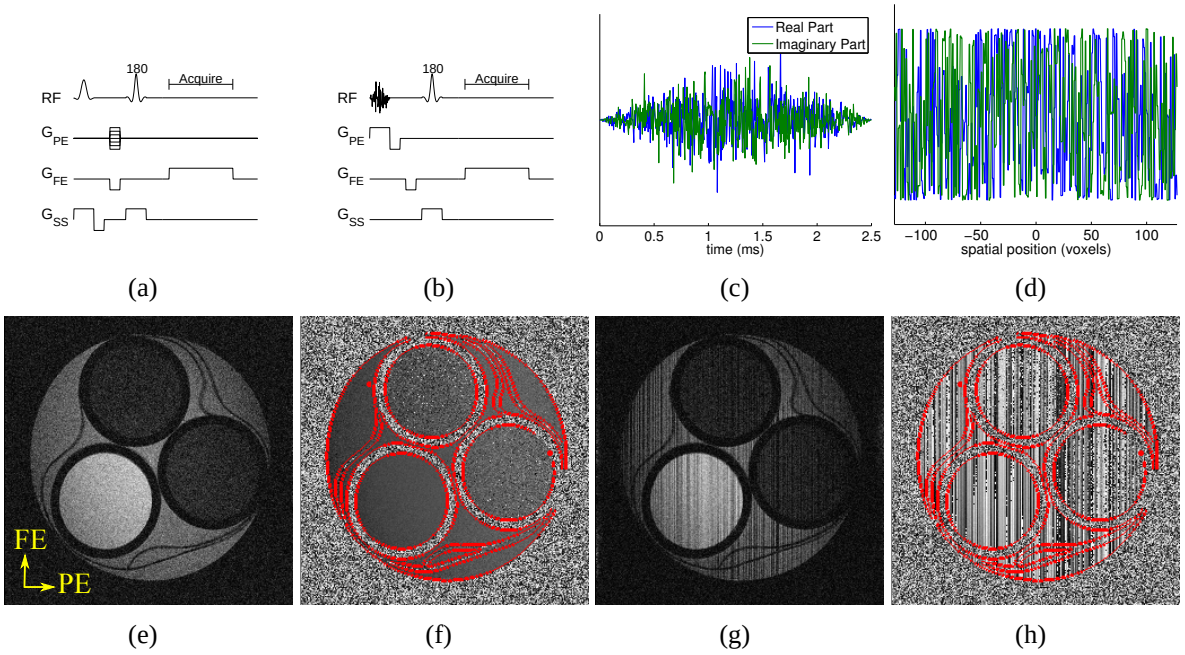


Figure 4.4: (a) The conventional Fourier-encoded spin-warp sequence, and (b) the proposed 1D random-encoding sequence. G_{PE}, G_{FE}, and G_{SS} represent the gradients along the phase encoding, frequency encoding, and slice select dimensions, respectively. Also shown are (c) a typical random-encoding RF pulse and (d) its corresponding excitation profile. The impact of random encoding is depicted with real experimental data in (e-h). The (e) magnitude and (f) phase of a phantom acquired with standard excitation and full Fourier encoding, as compared to the (g) magnitude and (h) phase of the same phantom acquired with random-encoding excitation and full Fourier encoding. The frequency encoding (FE) and phase encoding (PE) directions for these images are labeled in (e).

x is the phase encoding dimension and y is the frequency encoding dimension, we use

$$w_m(x,y) = \sum_{q=-Q/2}^{Q/2-1} \gamma_{qm} \Pi(x-q), \tag{4.11}$$

where γ_{qm} are Gaussian distributed as before, and $w_m(x,y)$ is the same for all samples from the same excitation.

The RF pulses used to achieve the 1D excitation profiles from Eq. (4.11) are designed using the small tip-angle approximation [502], such that the excitation RF pulse waveform can be generated by taking the Fourier transform of the desired 1D excitation profile. An example RF pulse and the corresponding excitation profile are shown in Fig. 4.4(c-d).

This form of random encoding requires the use of RF pulses for both spatial encoding and slice selection. Given the limitations of current multidimensional excitation technology, this necessitates the use of multiple pulses in practice. This limitation is common to other two-dimensional non-Fourier encoding schemes that use spatially-selective excitation (e.g., [455, 496]), though it can be overcome if the RF encoding is applied only along the third dimension of a three-dimensional experiment (e.g., [151, 420]). In addition, the use of varying excitation angles can complicate steady-state behavior [659]. This issue is also present for other similar non-Fourier encoding techniques, and is generally overcome by using small flip angles and relatively long repetition times [506]. Use of random encoding outside of this regime can mean that data acquisition is nonlinear and no longer accurately modeled by Eq. (4.8). The use of nonlinear random encoding does not fall within the scope of conventional CS or this chapter; however, preliminary empirical investigations of nonlinear random encoding can be found in Ref. [656], in which ℓ_1 regularization is used in the context of a parametric nonlinear signal model.

4.2 Evaluation

Experiments and simulations were performed to investigate the properties of random encoding for CS-MRI. In all cases, we compared three different data acquisition schemes with a fixed number M of data samples:

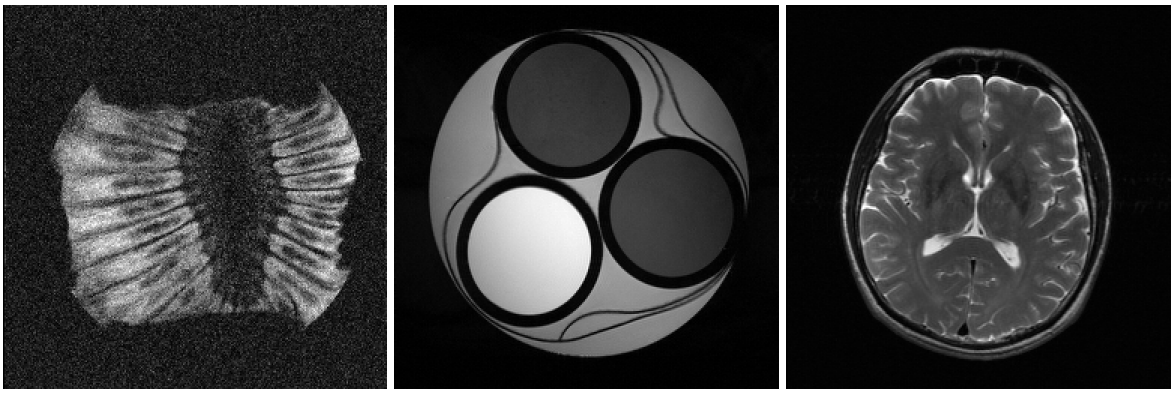
- *Random Encoding*. The proposed practical random encoding scheme with 1D spatially-selective RF excitations, as described in Sec. 4.1.2.

- *Fourier Encoding 1 (FE1)*. This scheme uses Fourier encoding with the standard spin-warp sequence from Fig. 4.4(a). The phase encoding locations are evenly spaced at the Nyquist rate, and cover the low frequency portion of *k*-space.

- *Fourier Encoding 2 (FE2)*. Similar to FE1, FE2 uses Fourier encoding with the standard spin-warp sequence. However, the phase-encoding locations are chosen randomly from the Nyquist grid according to a discretized Gaussian distribution centered at low-frequency *k*-space. This type of variable-density random sampling scheme performs empirically better than sampling *k*-space uniformly at random, and is consistent with both the prior knowledge that the typical images seen in MRI have energy concentrated at low-frequencies and the existing CS-MRI literature [420, 500, 635].

4.2.1 Experiments

The three different encoding schemes were implemented on a 14.1 T magnet system (Oxford Instruments, Abingdon, UK) interfaced with a Unity console (Varian, Palo Alto, CA, USA). The flip angle for FE1 and FE2 encoding and the root mean square flip angle for random encoding was 5°, with an RF pulse duration of 2.5 ms. The field of view was 3 cm $\times$ 3 cm, the slice thickness was 4 mm, and the sequence timing parameters were TE/TR = 26/500 ms. Data was collected for reconstruction on a 256$\times$256 voxel grid using two different test objects: a compartmental phantom



(a) (b) (c)

Figure 4.5: The fully Fourier-encoded image of the section of kiwi fruit from a real experiment is shown in (a). High-SNR images of the compartmental phantom and the brain image used for simulations are shown in (b) and (c), respectively.

and a section of kiwi fruit. The estimated SNR[14] for full 256×256 Fourier encoded data was approximately 4 for the compartmental phantom image (shown in Fig. 4.4(e)), and approximately 6 for the kiwi fruit image (shown in Fig. 4.5(a)).

Due to non-ideal experimental conditions (e.g., $\mathbf{B}_0$ and $\mathbf{B}_1$ inhomogeneity), the experimentally achieved excitation profiles used for random encoding did not match exactly with the designed profiles. As such, the excitation profile of each pulse was calibrated using prescans. Specifically, a fully-Fourier encoded image $\rho_{\text{cal}}(x, y)$ was acquired for each of the spatially-selective excitation pulses (one such image is shown in Fig. 4.4(g-h)). From these images, the γ_{qm} parameters for each excitation profile (recall Eq. (4.11)) were derived by solving the least squares problem

$$\hat{\gamma}_{qm} = \arg\min_{\gamma_{qm}} \sum_{p=-P/2}^{P/2-1} \left| \gamma_{qm} \rho_{\text{ref}}(q, p) - \rho_{\text{cal}}(q, p) \right|^2 , \tag{4.12}$$

where $\rho_{\text{ref}}(x, y)$ is an image acquired using traditional excitation pulses. This calibration procedure is somewhat coarse, since we ignore any potential excitation inhomogeneity along the frequency-encoding direction, though this choice leads to improved noise robustness compared

[14]Noise variances were empirically estimated from background regions of fully-sampled Fourier-encoded reference images that were free of visible artifacts, while signal levels were computed using the average value of the reference images in signal-containing regions of interest. The estimated SNR was calculated as the ratio between the signal level and the noise standard deviation.

to voxel-by-voxel estimation. In addition, while acquiring data for this calibration procedure is time consuming, the procedure could be simplified through direct mapping of the $\mathbf{B}_1$ transmit field and more accurate modeling of the excitation physics.

CS reconstructions were performed by solving

$$\text{minimize}\, \text{TV}\left(\hat{\boldsymbol{\rho}}_{\text{CS}}\right) \text{ subject to } \left\|\mathbf{E}\hat{\boldsymbol{\rho}}_{\text{CS}} - \mathbf{d}\right\|_{\ell_2}^2 \leq \varepsilon^2, \tag{4.13}$$

where ε was chosen according to an estimate of the expected data error due to noise (i.e., $\varepsilon^2 = M\sigma^2$, where σ^2 is the estimated noise variance and M is the number of measurements), and $\text{TV}\left(\boldsymbol{\rho}\right)$ is the total variation (TV) [546] cost functional that penalizes the ℓ_1 norm of the magnitude of the image gradient. Penalizing the image gradient is very common for CS reconstruction of MR images (e.g., [60,362,420,608]), since medical images are often approximately piecewise smooth, though it should be noted that the magnitude of the image gradient is a nonlinear transformation of the image and cannot be represented by a matrix $\boldsymbol{\Psi}$. Reconstructions were obtained using a version of Nesterov's algorithm as described in Ref. [38], with minor modifications to handle complex images. In particular, Ref. [38] approximates the nonsmooth ℓ_1 norm as $\|\cdot\|_{\ell_1} \approx \psi(\cdot)/(2\xi) + \xi^2$, where $\psi(\cdot)$ is the differentiable Huber function from Eq. (3.25). Subsequently, optimization is performed using Nesterov's algorithm as described previously in Sec. 3.7.2. A continuation approach is used, where the Huber function parameter ξ is initially set large and gradually reduced towards 0 during the optimization. When ξ is large, the cost functional being minimized is nearly quadratic, and Nesterov's algorithm converges rapidly. As $\xi \to 0$, the convergence speed of the algorithm decreases, though the Huber-function approximation of the ℓ_1 norm also becomes more and more accurate.

Instead of directly solving Eq. (4.13), Ref. [38] solves the Lagrangian form of the problem:

$$\text{minimize}\, \text{TV}\left(\hat{\boldsymbol{\rho}}_{\text{CS}}\right) + \lambda\left(\left\|\mathbf{E}\hat{\boldsymbol{\rho}}_{\text{CS}} - \mathbf{d}\right\|_{\ell_2}^2 - \varepsilon^2\right), \tag{4.14}$$

where λ is a Lagrange multiplier that is adjusted to satisfy the Karush-Kuhn-Tucker conditions for Eq. (4.13). The specific implementation of the algorithm described in Ref. [38] directly solves Eq. (4.13) for the special case when $\mathbf{E}$ is a submatrix of a unitary transform. In this case, the optimal Lagrange multiplier associated with the inequality constraint in Eq. (4.13) has a closed form expression that can be adjusted automatically during the iterative procedure [38]. While the encoding matrix is a submatrix of a unitary transform when Cartesian Fourier encoding is used, $\mathbf{E}$ does not have this property for random encoding. As a result, we manually adjust λ when reconstructing data acquired with random encoding. In most practical cases of interest (i.e., when $\|\mathbf{d}\|_{\ell_2} > \varepsilon$), λ should be chosen such that $\|\mathbf{E}\hat{\boldsymbol{\rho}}_{\mathrm{CS}} - \mathbf{d}\|_{\ell_2}^2 = \varepsilon^2$, which will ensure that the solution to Eq. (4.14) is equivalent to the solution of Eq. (4.13) [38]. Selection of λ to satisfy this condition is straightforward, since the data fidelity of the solution to Eq. (4.14) is monotonically decreasing with increasing λ. Note that the $\mathbf{E}$ matrix associated with random encoding has very similar structure to the encoding matrix used in SENSE parallel imaging reconstruction [519], except that RF excitation profiles are used in place of receiver coil sensitivity profiles. As a result, multiplication with $\mathbf{E}$ and its conjugate transpose can be performed efficiently using fast Fourier transforms [519], and these techniques were used to accelerate computations in the present context.

Reconstructions using experimental data from the low-SNR compartmental phantom and the higher-SNR kiwi fruit are shown in Figs. 4.6 and 4.7, respectively. With FE1, the CS reconstruction looks very similar to what would be obtained from conventional zero-padded reconstruction of low-frequency data, with accurate contrast information for low-resolution features, but also with significant blurring and distortion of the object geometry. With FE2, contrast is less accurate than with FE1, though the high-resolution image features are reconstructed better with FE2 than with FE1 with sufficient data. Results using random encoding indicate that it is possible to use this new scheme for CS-MRI, and that random encoding yields reconstructions with different characteristics than what are obtained with more traditional Fourier-based schemes. The figures suggest that random RF excitation can encode both high- and low-resolution image structures reasonably well, leading to a more balanced trade-off between contrast and resolution. Notably, some of the high-

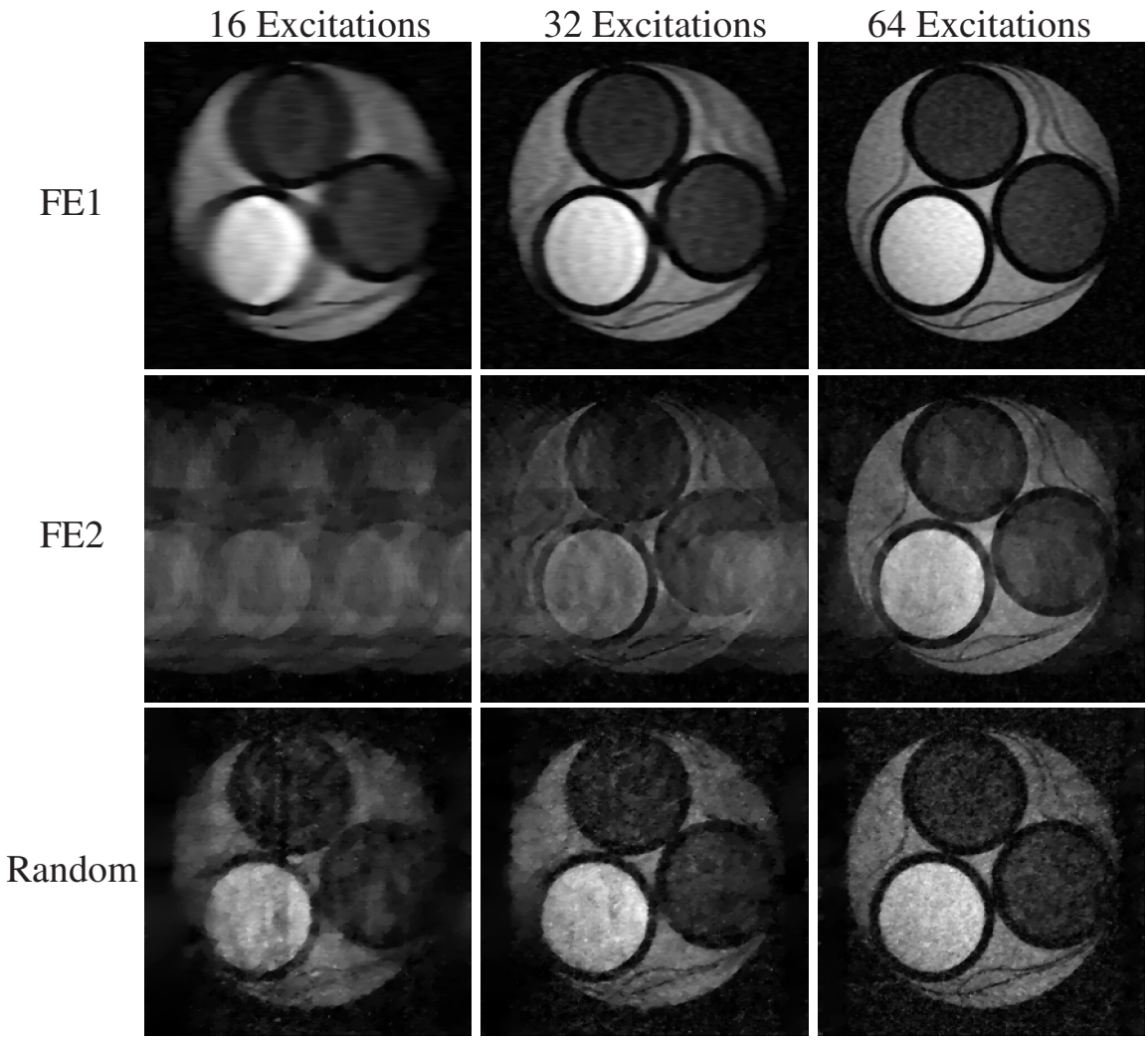


Figure 4.6: CS-MRI reconstructions from real experimental data from the compartmental phantom. Each row represents a different encoding scheme, while each column represents a different amount of measured data. These reconstructions demonstrate that CS-MRI with random encoding is feasible, and has different characteristics than either FE1 (which samples low-frequency *k*-space) or FE2 (which uses randomized *k*-space phase-encoding locations).

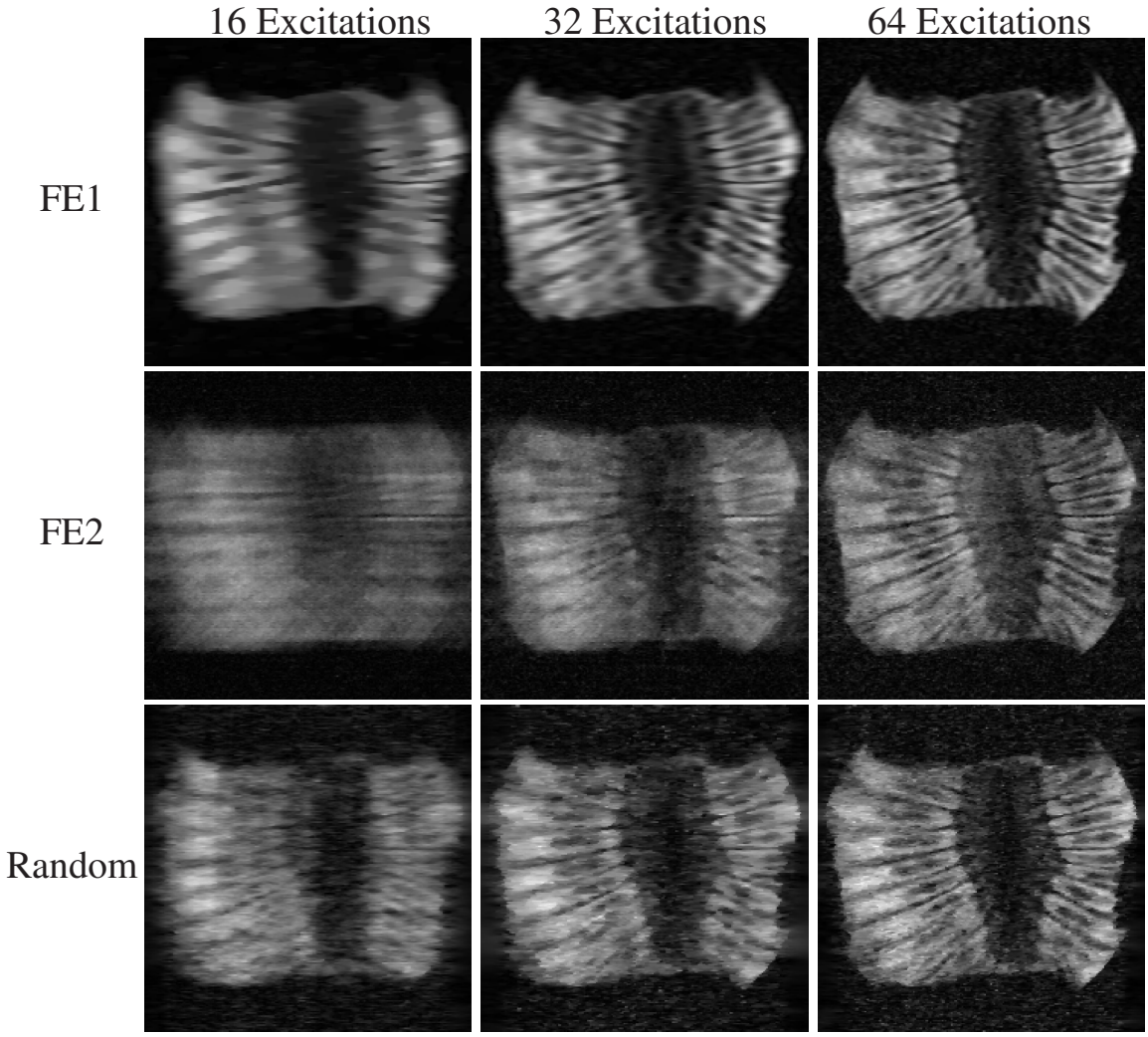


Figure 4.7: CS-MRI reconstructions of real experimental data from the section of kiwi fruit. Each row represents a different encoding scheme, while each column represents a different amount of measured data. As before, random encoding enables visualization of both low- and high-resolution image features with very limited data.

resolution image geometry is visible using random encoding with only 16 excitations (e.g., the geometry of the circular compartments in Fig. 4.6 and some of the fine edge structures in Fig. 4.7), while these features are significantly distorted with the other two schemes.

Similar to FE1 and FE2, reconstructions with random encoding become more accurate with increasing data. However, different from reconstructions with highly-undersampled FE1 and FE2 acquisitions (which can demonstrate significant geometry and/or large-scale contrast errors), the artifacts resulting from very limited random encoding data are more similar to the artifacts that might be observed from image compression (i.e., the loss of contrast for smaller image features). In addition, we should note that random encoding reconstructions also contain some artifacts that are not found in FE1 or FE2 reconstructions, and which could be attributed to noise, non-Gaussian excitation profiles, and/or errors in the calibration of the excitation profiles.

4.2.2 High-SNR Simulations

Compartmental Phantom

Simulations were also performed to illustrate performance when noise perturbations and calibration errors are minimal. The first set of simulations used a high-SNR image of the compartmental phantom as a gold standard, used nominal Gaussian excitation profiles, and incorporated simulated noise that was significantly weaker than that observed with the experimental data (the SNR was 80 with respect to the image from full 256×256 Fourier encoded data, which is shown in Fig. 4.5(b)). Figures 4.8 and 4.9 show representative results from these simulations. The improved SNR and nominal excitation profiles have led to improved reconstruction quality for all schemes, but with random encoding demonstrating a distinct advantage relative to the other schemes. The relative errors are shown in Table 4.1, where relative error is defined as

$$\text{Relative Error} = \frac{\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_{\text{cs}}\|_{\ell_2}}{\|\boldsymbol{\rho}\|_{\ell_2}}, \tag{4.15}$$

Table 4.1: Relative reconstruction errors for the high-SNR simulations using the compartmental phantom.

	Relative Error		
Encoding Scheme	16 Excitations	32 Excitations	64 Excitations
FE1	0.249	0.149	0.086
FE2	0.713	0.394	0.265
Random	**0.245**	**0.121**	**0.053**
Random (real profiles)	0.252	0.133	0.065
Random (2D profiles)	0.127	0.079	0.048

and serves as a measure of similarity between the reconstructed image $\hat{\boldsymbol{\rho}}_{\text{cs}}$ and the gold-standard image $\boldsymbol{\rho}$. For these simulations, random encoding outperformed both FE1 and FE2 in relative error at all investigated undersampling levels. As with the experimental results, it was observed that the distribution of errors with random encoding CS-MRI reconstructions was more evenly distributed between low- and high-resolution features than with FE1 or FE2.

Figure 4.10 shows results from additional random encoding simulations (relative errors for these are also shown in Table 4.1), where the excitation profiles were chosen to either be the empirically measured excitation profiles from the real experiment ("real profiles") or ideal two-dimensional profiles ("2D profiles") as in Eq. (4.10). As in the previous simulations, the SNR with respect to fully-encoded Fourier data was 80, and one frequency encoding line was acquired per excitation. The results with the real profiles are very similar to the results with the nominal profiles, and illustrate that it is not necessary to have perfectly white Gaussian γ_{qm} excitation profile parameters to have good reconstruction results. The results using 2D profiles in Fig. 4.10 demonstrate significantly improved performance relative to 1D random encoding, and indicate that even better results could be achieved if high-resolution multi-dimensional RF excitation techniques become more practical.

The quality of reconstructed images using random encoding can also be affected by errors in the encoding matrix $\mathbf{E}$ due to miscalibration of the RF excitation profiles. Theoretical analysis of Eq. (4.4) when $\mathbf{E}$ contains errors has been presented recently by Herman and Strohmer [302]. These results indicate that stable and accurate CS reconstructions can still be guaranteed with

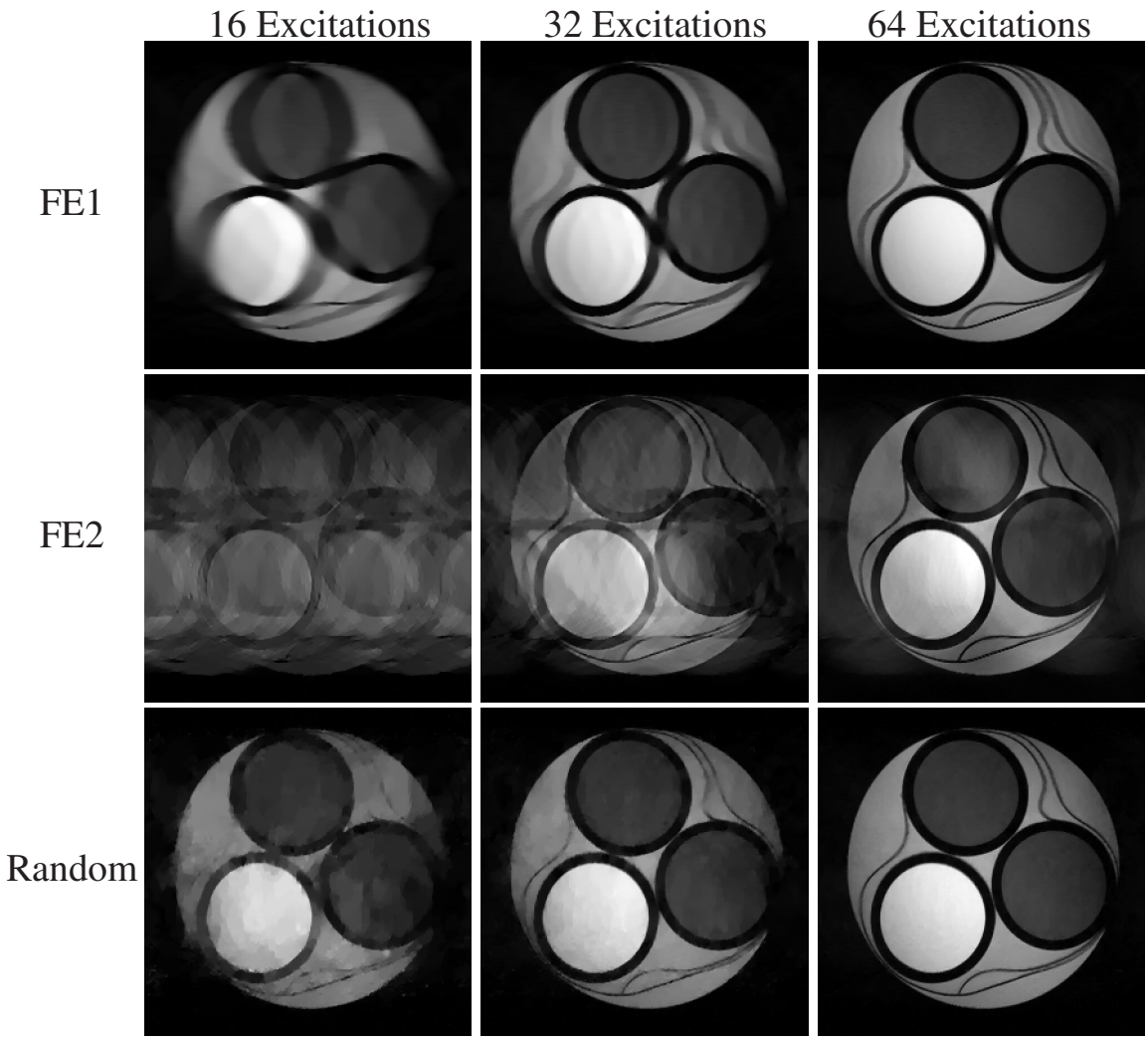


Figure 4.8: CS-MRI reconstructions from high-SNR simulations of the compartmental phantom. Each row represents a different encoding scheme, while each column represents a different amount of measured data. Relative to the experimental data, the improved SNR leads to better reconstructions for all encoding schemes. Reasonably accurate reconstruction was obtained using random encoding with only 32 excitations, while the Fourier encoding schemes required more data to achieve the same accuracy.

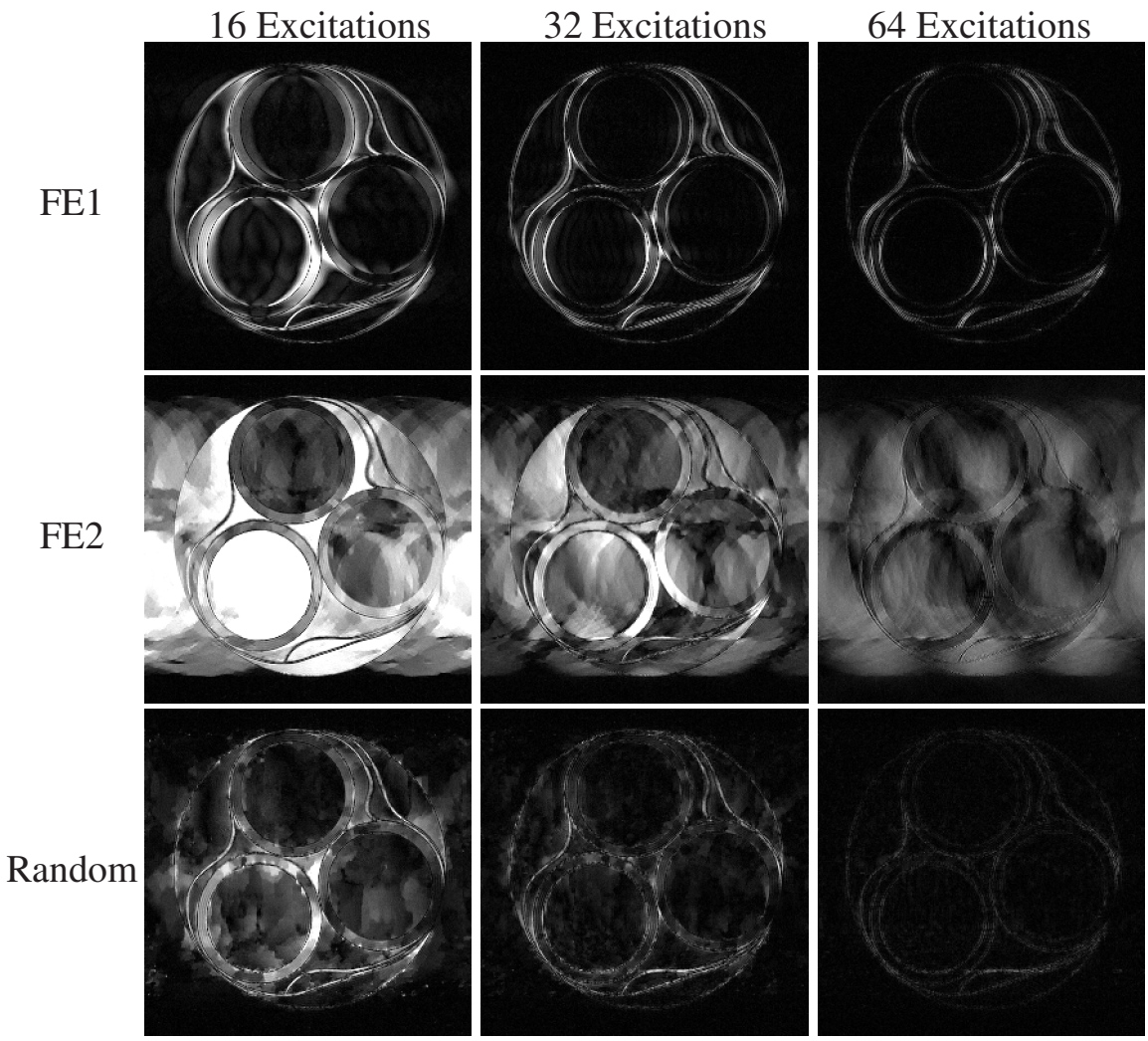


Figure 4.9: Error images (i.e., the difference between the gold standard and the reconstruction) corresponding to the high-SNR simulation results shown in Fig. 4.8. Each row represents a different encoding scheme, while each column represents a different amount of measured data. The error images have been scaled up by a factor of 3 for improved visualization.

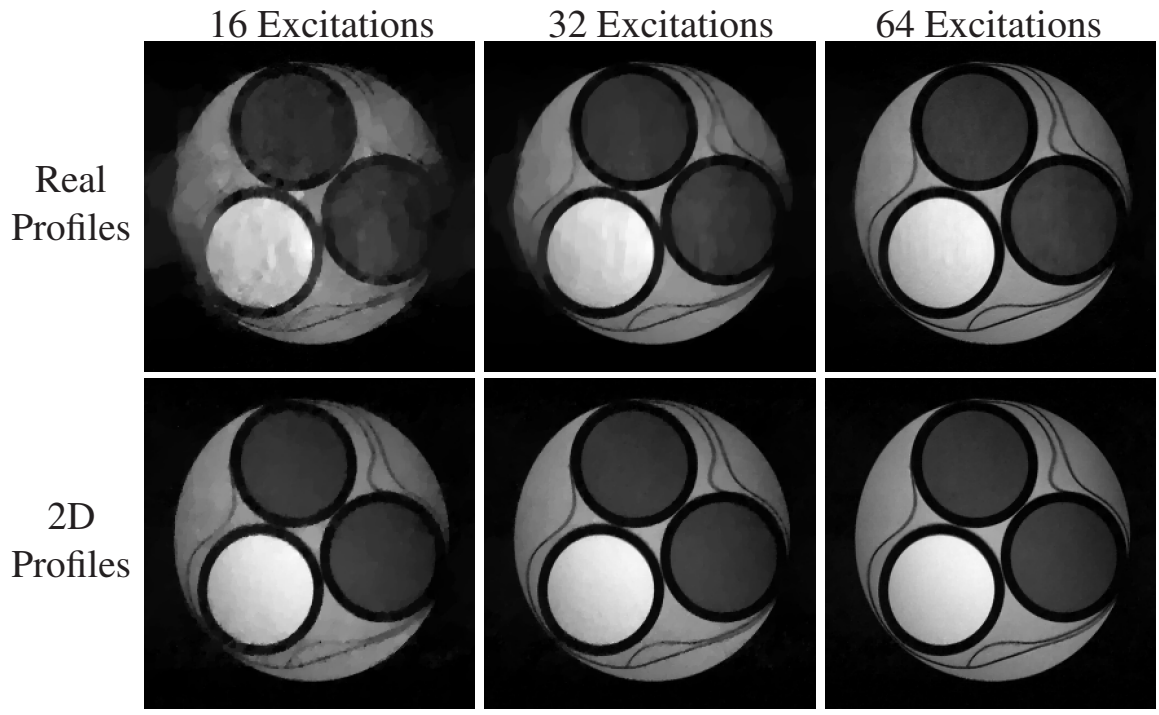
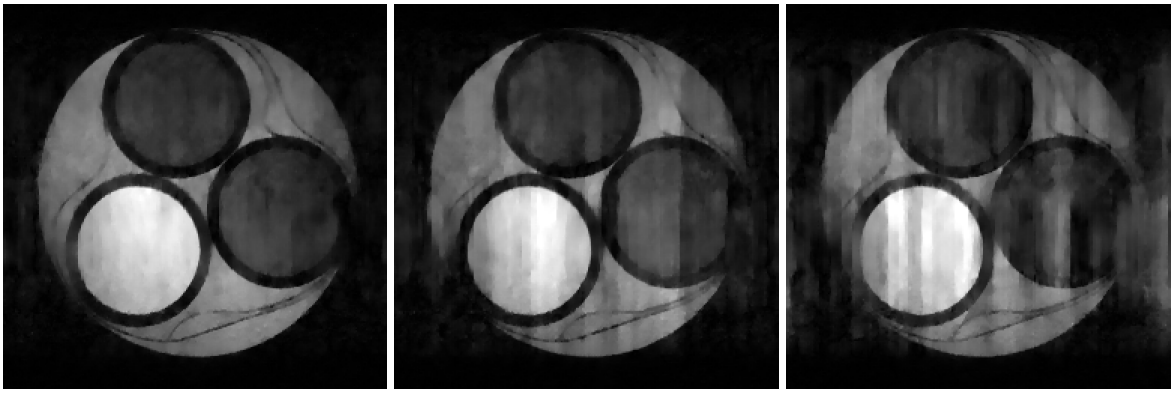
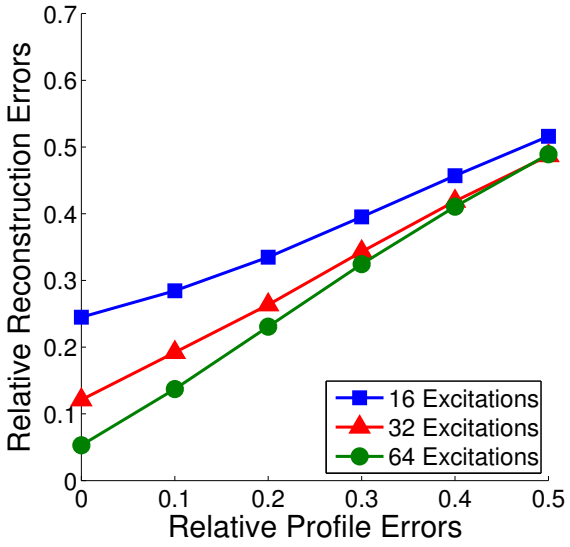


Figure 4.10: CS-MRI reconstructions from high-SNR random encoding simulations of the compartmental phantom. The top row shows results using the calibrated excitation profiles from a real experiment, while the bottom row shows results using random two-dimensional excitation profiles.

a noisy $\mathbf{E}$, under the assumptions that the true measurement matrix satisfies an appropriate RIP condition and that the magnitude of the perturbation is not too large. In particular, the theoretical analysis and numerical simulations in Ref. [302] suggest that the stability of $\hat{\boldsymbol{\rho}}_{\text{cs}}$ should scale linearly with the amount of perturbation to the system matrix. Simulation studies were performed to examine the effects of RF profile miscalibration. High-SNR data was simulated using standard 1D random encoding with nominal Gaussian RF profiles, and the RF profile parameters γ_{qm} used for reconstruction were perturbed by Gaussian noise. Results of these simulations are shown in Fig. 4.11. These results suggest a linear relationship between reconstruction error and calibration error, as might be expected based on the theoretical analysis [302].



(a) 10% Profile Error (b) 20% Profile Error (c) 30% Profile Error



(d)

Figure 4.11: Simulated random-encoding reconstruction results in the presence of miscalibration of the RF excitation profiles. (a)-(c) Representative reconstructions from 32 excitations in the presence of increasing levels of calibration error. (d) The total reconstruction error (shown averaged over 5 realizations) is observed to grow linearly with respect to the calibration error.

Brain Phantom

High-SNR simulations were also performed with the brain image shown in Fig. 4.5(c). The simulations in this case had the same noise level as the high-SNR compartmental phantom simulations. Representative reconstructions from 96 excitations and either TV or a wavelet (Daubechies-4) sparsifying transform are shown in Figs. 4.12 and 4.13. The TV-based reconstructions used Nesterov's algorithm, as described previously. However, for the wavelet-based reconstructions, we used the primal-based alternating directions minimization (ADM) algorithm as described in Ref. [667] for solving the constrained optimization problem in Eq. (4.4). This form of the ADM algorithm was observed to have faster convergence and higher numerical stability than Nesterov's algorithm for these problems. Representative relative reconstruction errors for a range of undersampling levels are listed in Table 4.2.

The brain image has lower compressibility than the compartmental phantom, and is thus more challenging for CS-MRI and required a larger amount of data for accurate reconstruction. In addition, the performance advantage (in terms of relative error) of random encoding relative to FE1 and FE2 was less substantial than it was with the compartmental phantom simulations. This was particularly true using the wavelet-based constraint, which was significantly less effective than the TV constraint for all encoding schemes. However, the spatial distributions of error for both TV and wavelet sparsity are still consistent with what was observed previously. In particular, the errors for FE1 encoding are concentrated around the high-resolution features of the image, while there are significant contrast errors for low-resolution image features with FE2 encoding. The distribution of errors with random encoding is intermediate between the FE1 and FE2 cases, with the errors somewhat more uniformly distributed between low- and high-resolution image features. These characteristics have been observed consistently in both simulations and experiments, and are important to note when choosing an encoding scheme for a particular imaging scenario, since different features will have more or less importance depending on the application.

Table 4.2: Relative reconstruction errors for the high-SNR simulations using the brain image.

Encoding Scheme (Sparsifying Transform)	Relative Error		
	64 Excitations	96 Excitations	128 Excitations
FE1 (TV)	0.179	0.119	0.076
FE2 (TV)	0.184	0.113	0.074
Random (TV)	**0.154**	**0.090**	**0.055**
FE1 (wavelet)	**0.228**	**0.154**	0.117
FE2 (wavelet)	0.330	0.191	0.133
Random (wavelet)	0.251	0.158	**0.099**

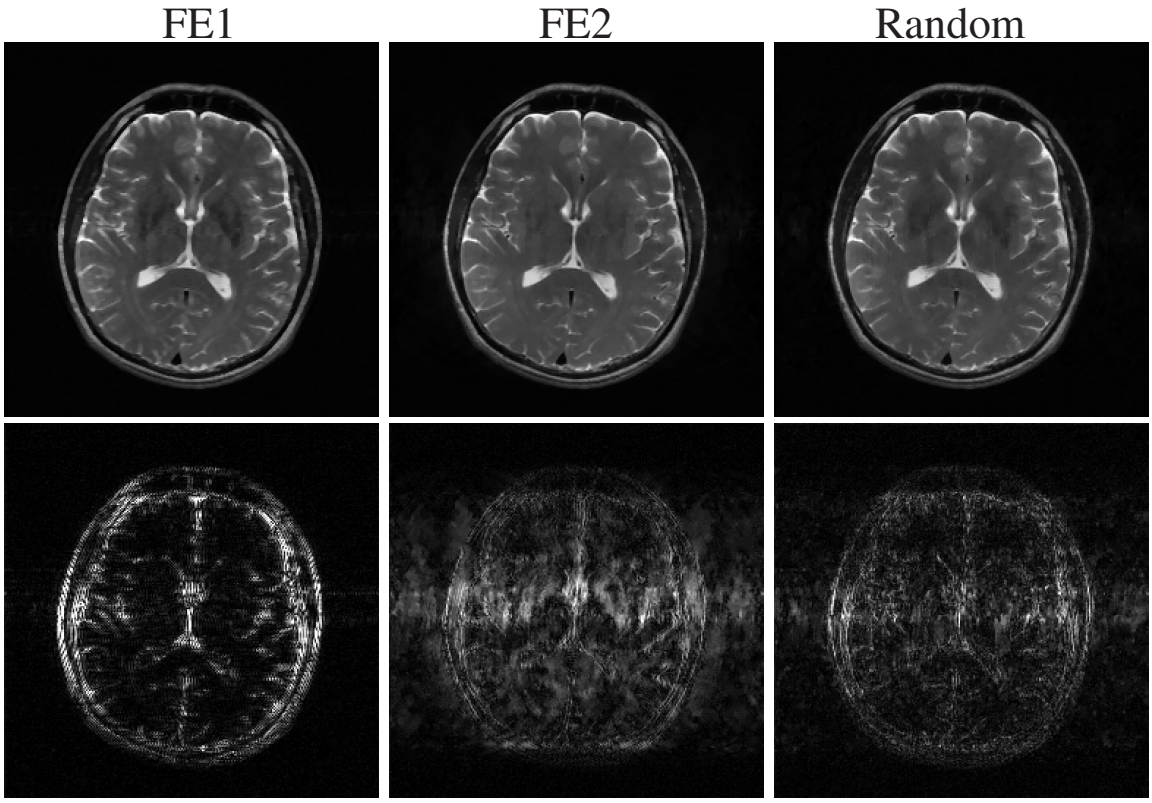


Figure 4.12: Simulated CS-MRI reconstructions of the compressible brain image from 96 excitations, with a TV penalty. The top row shows the reconstructions themselves, while the bottom row shows the differences (scaled up by a factor of 6) between the reconstructions and the gold standard.

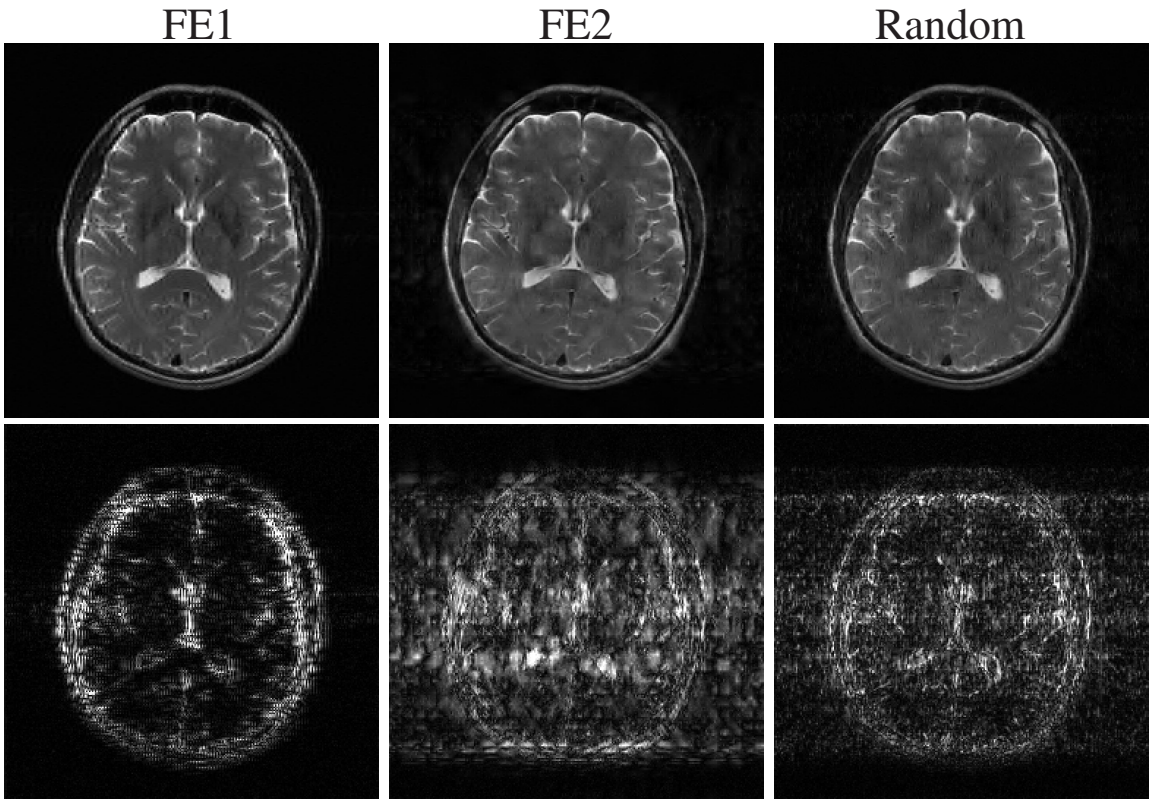


Figure 4.13: Simulated CS-MRI reconstructions of the compressible brain image from 96 excitations, with a Daubechies-4 wavelet penalty. The top row shows the reconstructions themselves, while the bottom row shows the differences (scaled up by a factor of 6) between the reconstructions and the gold standard.

4.2.3 Monte Carlo Simulations

Monte Carlo simulations were also performed to study the reconstruction and noise properties of random encoding relative to FE1 and FE2. In these simulations, reconstructions were performed using an image with a sparse gradient (the Shepp-Logan phantom) and an image with a compressible gradient (the MR brain image shown in Fig. 4.5(c)). Simulations were performed 50 times for each combination of six data undersampling levels (8, 16, 32, 64, 128, and 256 excitations), the three different encoding schemes (FE1, FE2, and random encoding), and seven different noise levels (SNRs ranging from 1 to 80 with respect to full 256×256 Fourier encoding). The random elements of the simulation (i.e., the sampling locations for FE2 encoding, the excitation profiles for random encoding, and the noise) were different for each trial. To improve the computational speed for these 12,600 reconstructions, each reconstruction made use of a simplified one-dimensional TV penalty that only penalized the ℓ_1 norm of the difference between adjacent voxel values along the phase-encoding dimension. Since the frequency encoding dimension was fully sampled, this modified TV penalty means that the optimal two-dimensional 256×256 CS reconstruction could be performed using 256 independent smaller one-dimensional CS reconstructions, one for each line of the image. This simplification allows reconstructions to be performed much more rapidly than if standard TV was used, and additionally means that the $\mathbf{E}$ matrix for each subproblem has the ideal "universal" distribution. To solve these one-dimensional CS problems, we used the CVX software package by Grant, Boyd, and Ye (`http://www.stanford.edu/~boyd/cvx/`).

Results from the Monte Carlo simulations using the brain image and the sparse Shepp-Logan phantom are shown in Figs. 4.14 and 4.15, respectively. Images are generally more compressible using a two-dimensional transform rather than a one-dimensional transform, leading to slightly lower performance for these simulations compared to those in the previous subsection. However, the relative performance characteristics of the different encoding schemes with one-dimensional sparsity constraints are consistent with the behavior observed with two-dimensional constraints. For both images in the Monte Carlo simulations, the relative error decreases as the amount of ac-

quired data increases, and FE1 encoding was generally superior to both FE2 and random encoding in cases with very limited data or with high levels of noise. FE2 encoding consistently outperforms FE1 encoding with high-SNR data when the number of measurements is large. Random encoding can outperform both FE1 and FE2 encoding, though this only occurs with high-SNR data, and the advantage of random encoding over the Fourier-based schemes disappears as the number of measurements M becomes comparable to the number of voxels N. One way of understanding this phenomenon is to consider the case of fully-sampled data (i.e., $M = N$) with standard reconstruction, where the reconstructed image is obtained by $\hat{\boldsymbol{\rho}} = \mathbf{E}^{-1}\mathbf{d}$. In this case, the discrete Fourier transform (DFT) matrix is unitary, which means that the noise in the data will not be amplified by $\mathbf{E}^{-1}$. In contrast, fully-sampled random encoding matrices will generally have worse condition numbers than the DFT matrix [193], resulting in more significant noise amplification.

Similar Monte Carlo simulations imposing a one-dimensional Daubechies-4 wavelet-based sparsity constraint are shown in Figs. 4.16 and 4.17, and have similar characteristics to the one-dimensional TV-based simulations. Notably, the regimes for which random encoding outperforms the Fourier-based schemes (in terms of relative error) are different for the Shepp-Logan phantom compared to the brain image, and are also different for different sparsifying transforms (i.e., the one-dimensional TV and wavelet transforms and the two-dimensional transforms considered in the previous subsection). This further suggests that the choice between the use of random encoding versus a Fourier encoding scheme should be made carefully based on the constraints of each application.

4.3 Discussion

4.3.1 Performance Guarantees

The use of random encoding in this work was motivated by the desire to improve restricted isometry constants and improve the theoretical characterization of CS-MRI reconstruction. As men-

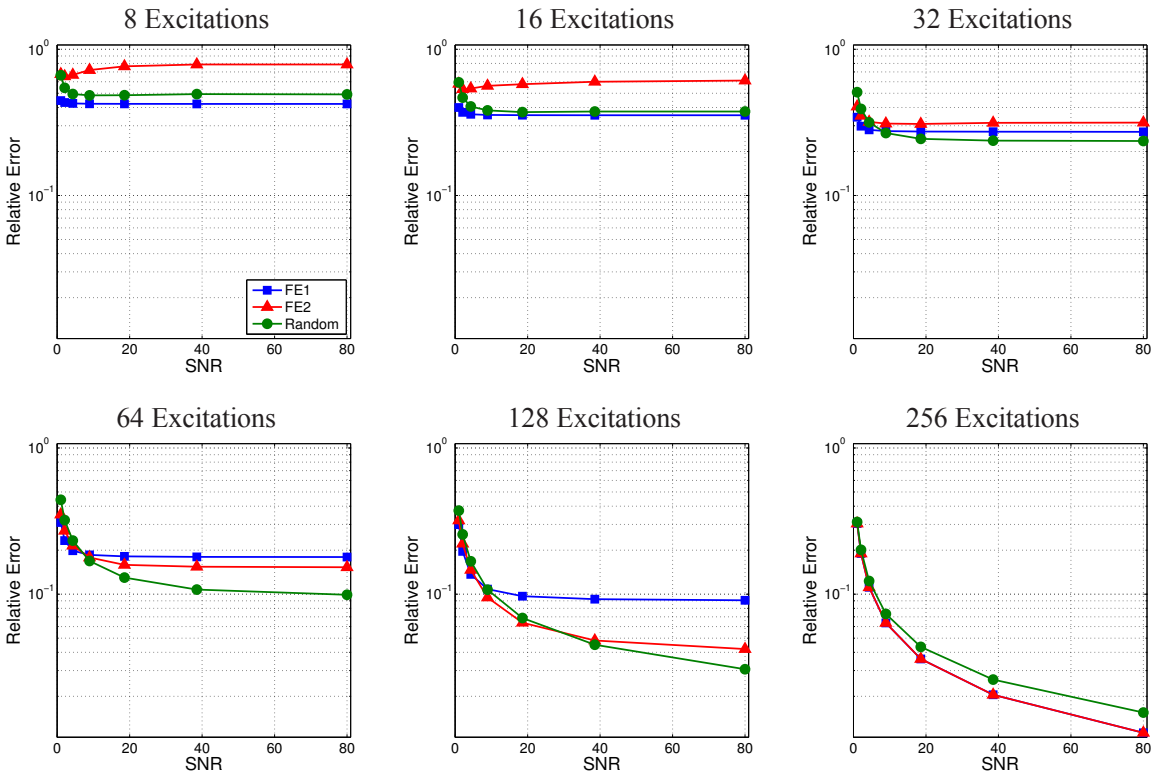


Figure 4.14: Plots showing the median relative error as a function of SNR from the Monte Carlo simulations with the brain image and TV-based sparsity constraints. In all cases, the relative error decreases as the amount of acquired data increases. FE1 encoding was generally superior in cases with very limited data or with high levels of noise. However, for moderate noise and sufficient data acquisition, random encoding performed better than the other two schemes, and FE2 outperforms FE1. For fully-encoded data, the SNR efficiency of the Fourier schemes allows them to dominate the random encoding scheme.

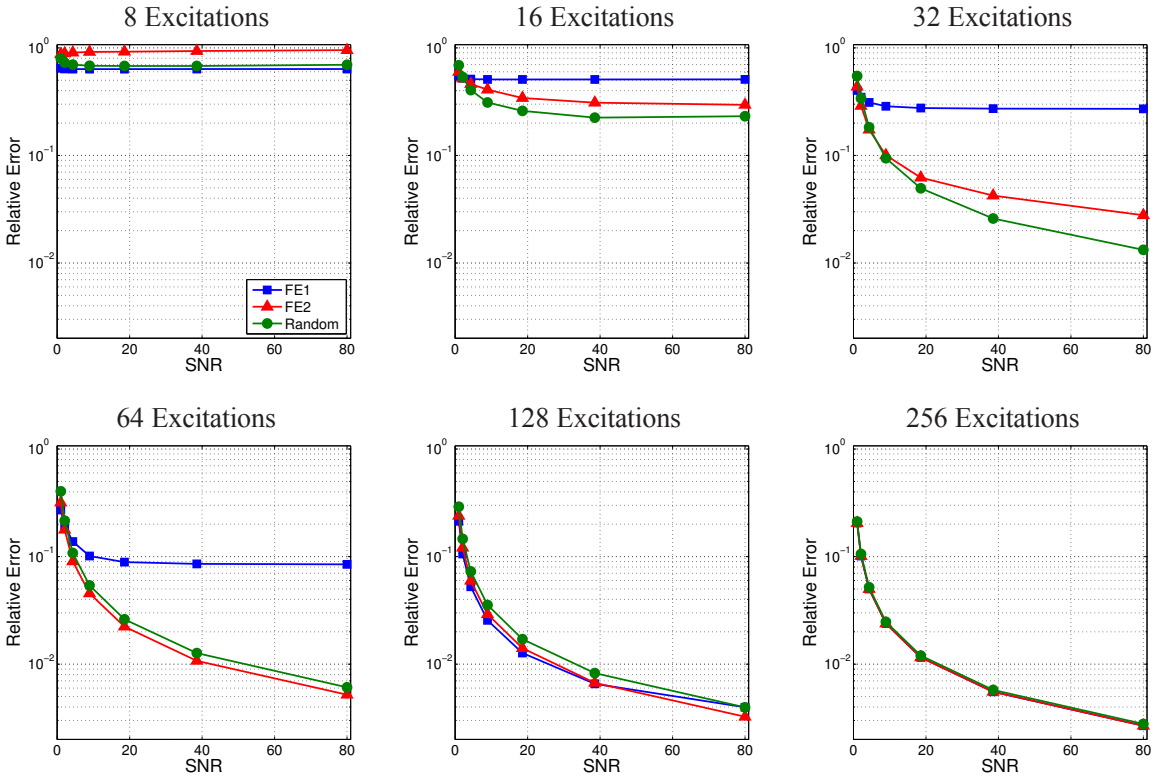


Figure 4.15: Plots showing the median relative error as a function of SNR from the Monte Carlo simulations with the Shepp-Logan phantom and TV-based sparsity constraints. The trends are similar to those observed for the compressible brain image, though for the same number of measurements, smaller relative error is generally achieved with this sparse image. Notably, the regime for which random encoding outperforms the Fourier-based schemes is different than it was with the brain image.

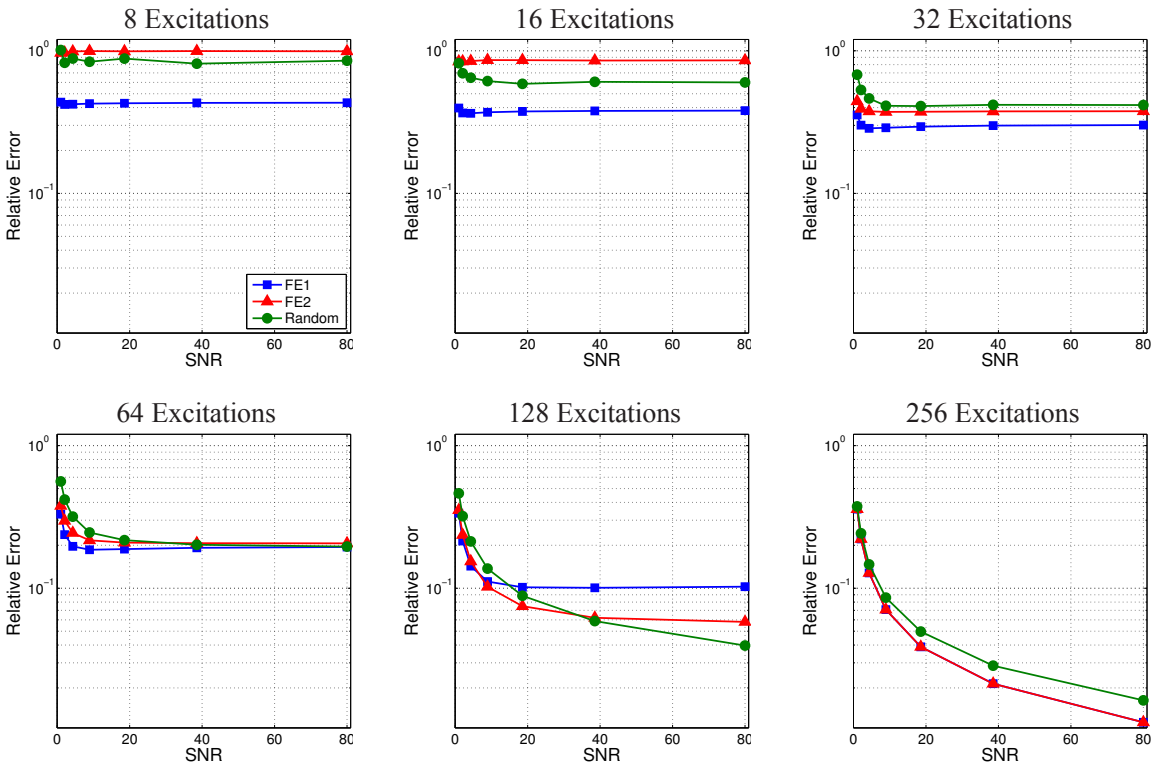


Figure 4.16: Plots showing the median relative error as a function of SNR from the Monte Carlo simulations with the brain image and wavelet-based sparsity constraints.

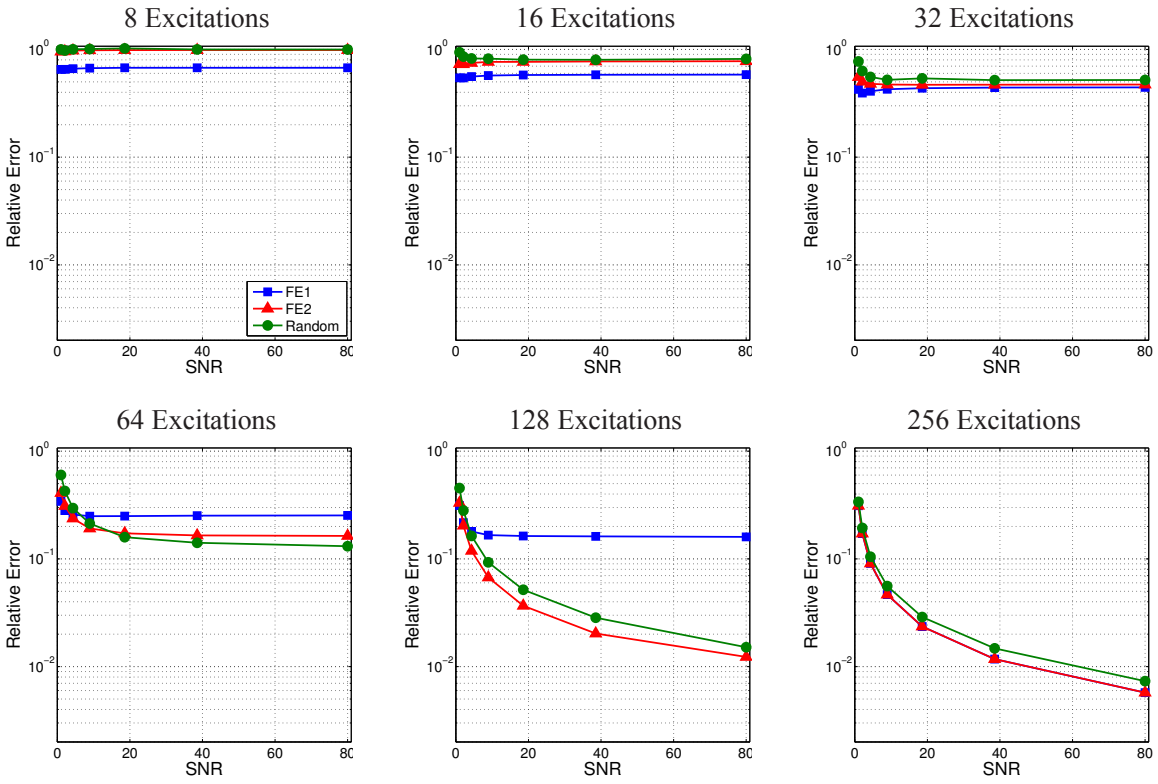


Figure 4.17: Plots showing the median relative error as a function of SNR from the Monte Carlo simulations with the Shepp-Logan phantom and wavelet-based sparsity constraints.

tioned in the introduction, it is generally computationally infeasible to compute the restricted isometry constants.[15] However, it is relatively straightforward to calculate the δ_1 restricted isometry constant for a matrix $\mathbf{\Phi}$ using Eq. (4.6) with

$$\alpha_1 = \min_i \|\mathbf{\Phi}_i\|_{\ell_2}^2 \text{ and } \beta_1 = \max_i \|\mathbf{\Phi}_i\|_{\ell_2}^2, \tag{4.16}$$

where the vectors $\mathbf{\Phi}_i$ are the columns of $\mathbf{\Phi}$.

Besides RIPs, there are also incoherence conditions on $\mathbf{\Phi}$ that can guarantee good CS performance and are significantly easier to compute [41, 104, 107, 181, 182]. While these incoherence-based guarantees are generally weaker than RIP-based guarantees, they have been used previously in the design of CS-MRI encoding schemes [420] and in other contexts [195]. For example, Lustig *et al.* [420] suggested that the maximum of the *transform point spread function* (TPSF) be used to characterize the incoherence of a sampling scheme, with more incoherent sampling schemes characterized as better for CS reconstruction. The TPSF has the form

$$\text{TPSF}\{i;j\} = \frac{\mathbf{\Phi}_j^H \mathbf{\Phi}_i}{\|\mathbf{\Phi}_i\|_{\ell_2} \left\|\mathbf{\Phi}_j\right\|_{\ell_2}}, \tag{4.17}$$

and is somewhat representative of the level of ambiguity between the ith and jth transform coefficients. Ideally, the TPSF should be small when $i \neq j$. The maximum of the TPSF is equal to the *mutual incoherence* μ:

$$\mu = \max_{i \neq j} |\text{TPSF}\{i;j\}|, \tag{4.18}$$

which can be used to generate another set of CS performance guarantees [104, 181, 182]. For example, if $\|\mathbf{c}\|_{\ell_0} < \frac{1}{4}(1/\mu + 1)$ and if the columns of $\mathbf{\Phi}$ are normalized to unit length, then the

[15] Currently, the only known way to compute δ_s coefficients is to enumerate and test the singular values of the set of all matrices formed by selecting s columns of Φ. This procedure is not practical even for small problems, due to the combinatorial nature of the computation. Several practical procedures have been proposed to place bounds on restricted isometry constants (and related computationally-intractable performance guarantees for CS) using a variety of different approaches [155, 156, 185, 347, 390, 494]. While such approaches are significantly easier than direct computation of δ_s, the resulting optimization problems can still be quite challenging to compute for problem sizes of interest.

Table 4.3: Representative δ_1 restricted isometry constants and mutual incoherence μ values for different encoding schemes, different amounts of acquired data, and different image grid sizes. Calculations were performed using a Daubechies-4 wavelet basis.

Grid Size	Encoding Scheme	δ_1		μ	
		32 Excitations	64 Excitations	32 Excitations	64 Excitations
256×256	FE1	0.9996	0.9876	0.9005	0.7833
	FE2	0.7636	0.7093	0.7147	0.6665
	Random	**0.5070**	**0.3567**	**0.5594**	**0.4179**
128×128	FE1	0.9874	0.7706	0.7840	0.7425
	FE2	**0.2655**	**0.0847**	0.6201	**0.3612**
	Random	0.4592	0.4129	**0.5619**	0.3635

solution to Eq. (4.4) is guaranteed to satisfy (see Thm. 3.1 in Ref. [182])

$$\|\hat{\mathbf{c}}_{\text{cs}} - \mathbf{c}\|_{\ell_2}^2 \leq \frac{4\varepsilon^2}{1 - \mu\left(4\|\mathbf{c}\|_{\ell_0} - 1\right)}, \tag{4.19}$$

where $\|\mathbf{x}\|_{\ell_0}$ is defined as the number of non-zero entries of $\mathbf{x}$.

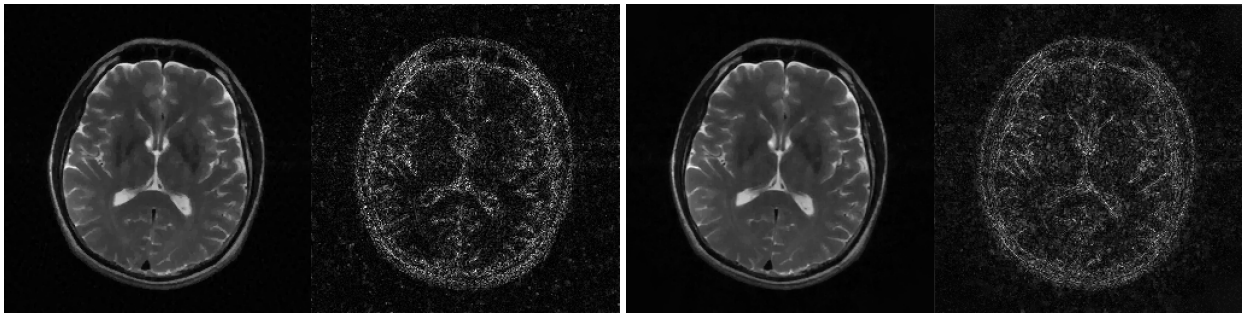
Table 4.3 shows representative values of δ_1 and μ for the three encoding schemes we have considered and using a $\mathbf{\Psi}$ matrix corresponding to a Daubechies-4 wavelet transform. Values are shown for reconstruction of both a 256×256 image and a 128×128 image. For the 256×256 case, both δ_1 and μ are smaller for random encoding than for FE1 and FE2. However, it is also important to note that δ_1 is never less than 0.307 for any encoding scheme, and only is only less than 0.4734 for random encoding with 64 frequency encoding lines. Since it is always true that $\delta_t \geq \delta_s$ when $t \geq s$, this implies that the current RIP-based guarantees for CS performance cannot be applied to the other measurement matrices, even for signals that have only one non-zero entry. Similar to what was observed with δ_1, μ is also smallest for random encoding at this image resolution. However, the characterization given by Eq. (4.19) can only be applied for non-zero vectors $\mathbf{c}$ when $\mu < 1/3$, so the observed μ values give no useful guarantees for any of the encoding schemes. Despite this, CS empirically works much better than what the theoretical bounds might suggest, and it is promising that random encoding yields the smallest δ_1 and μ values.

It is also important to note that the good theoretical properties for "universal" encoding are

somewhat dependent on the problem size, with a higher probability of good RIPs as M and N grow large [108]. We have observed that the superiority of the μ and δ_1 values for random encoding is also dependent on the problem size. For example, with a 128×128 image, we have observed that FE2 has consistently better δ_1 values relative to random encoding, which is the opposite of the behavior observed with 256×256 images. However, it is observed that μ does not follow the same trend as δ_1 for this 128×128 case, and that μ can still be smaller for random encoding than for FE2 (cf. Table 4.3).

4.3.2 Non-Cartesian Acquisitions and Multidimensional Undersampling

For both Fourier and random encoding, we have focused on 2D Cartesian *k*-space sampling patterns with undersampling along a single dimension to keep the discussion as short and simple as possible. In practice, however, several CS-MRI studies have shown good results when using non-Cartesian Fourier sampling patterns and/or multidimensional undersampling schemes (e.g., Refs. [11, 60, 228, 317, 350, 362, 397, 420, 421, 557, 568, 608, 668]). We note that non-Cartesian and multidimensionally-undersampled forms of random encoding are also possible, though there are several ways of implementing such schemes. For example, a naive approach to non-Cartesian random encoding would be to maintain the same 1D spatially-selective excitation scheme as in Section 4.1.2, but replace standard frequency encoding with a non-Cartesian readout. A more complicated implementation could change the orientation of RF encoding for each excitation pulse in combination with a non-Cartesian readout. Preliminary simulations using both of these schemes with Fourier encoding along radial lines indicate further performance improvements [275], and an illustrative example is shown in Fig. 4.18. However, a detailed investigation of multidimensional encoding schemes is left for future work.



(a) Radial Fourier encoding. Relative error = 0.078. (b) Radial random encoding. Relative error = 0.074

Figure 4.18: TV-based reconstructions (left) and error images (right) of the brain image using simulated non-Cartesian data acquisition strategies. (a) Radial Fourier encoding with 64 excitations. (b) Radial random encoding (both the frequency encoding and RF encoding orientations were rotated) with 64 excitations. Error images are shown scaled up by a factor of 10. High-quality reconstructions can be obtained with significantly fewer excitations when non-Cartesian trajectories are used for data acquisition.

4.3.3 Other Considerations

A number of extensions to this work are possible. Some of the practical limitations of the proposed approach are specific to the use of selective RF excitation to achieve non-Fourier encoding. In particular, it can be difficult to achieve high-resolution multi-dimensional RF encoding pulses, some of the flexibility in the choice of sequence timings and flip angles is limited by the need to avoid disrupting steady-state behavior, and it is necessary to calibrate the excitation pulses. One approach to overcoming these limitations is to use other non-Fourier encoding mechanisms in place of or in addition to RF-based encoding. Initial work along these lines includes the use of a multi-channel array of receiver coils with randomized sensitivity profiles [566] and the use of nonlinear encoding gradients [523, 649].

Another limitation of current compressed sensing theory is that the spatial resolution and noise characteristics of the reconstruction are difficult to characterize theoretically, and having access to such characterizations is important for interpreting reconstructed images and for choosing data acquisition/reconstruction parameters. While error bounds like Eq. (4.7) and Eq. (4.19) can be derived when the measurement operator satisfies certain mathematical properties, good fidelity with respect to the ℓ_2-norm does not necessarily imply good visual fidelity [636]. In addition,

these bounds provide no information about how the error is distributed spatially (e.g., we do not have easy characterizations of resolution in terms of the point-spread functions or spatial response functions described for linear reconstruction in Chapter 3) or about the statistical distribution of the reconstructed image with respect to the noise distribution.

Approximate resolution characterizations for general nonlinear image reconstruction problems (outside of the context of CS reconstruction) have been studied empirically and theoretically by a number of authors [8, 213, 524, 587, 588], usually by studying how the reconstruction changes when a small localized image feature is added to a gold-standard image. These studies indicate that the resolution characteristics in nonlinear reconstruction problems are generally spatially-variant and highly data-dependent. Our preliminary experience has shown that this is also the case for CS reconstruction, both with and without random encoding. However, many of the fast approximate resolution characterization methods described in [8, 213, 524, 587, 588] are not applicable to CS reconstruction, due to the fact that the $\mathbf{E}$ matrix in CS is underdetermined and the ℓ_1-norm is not differentiable at the origin.

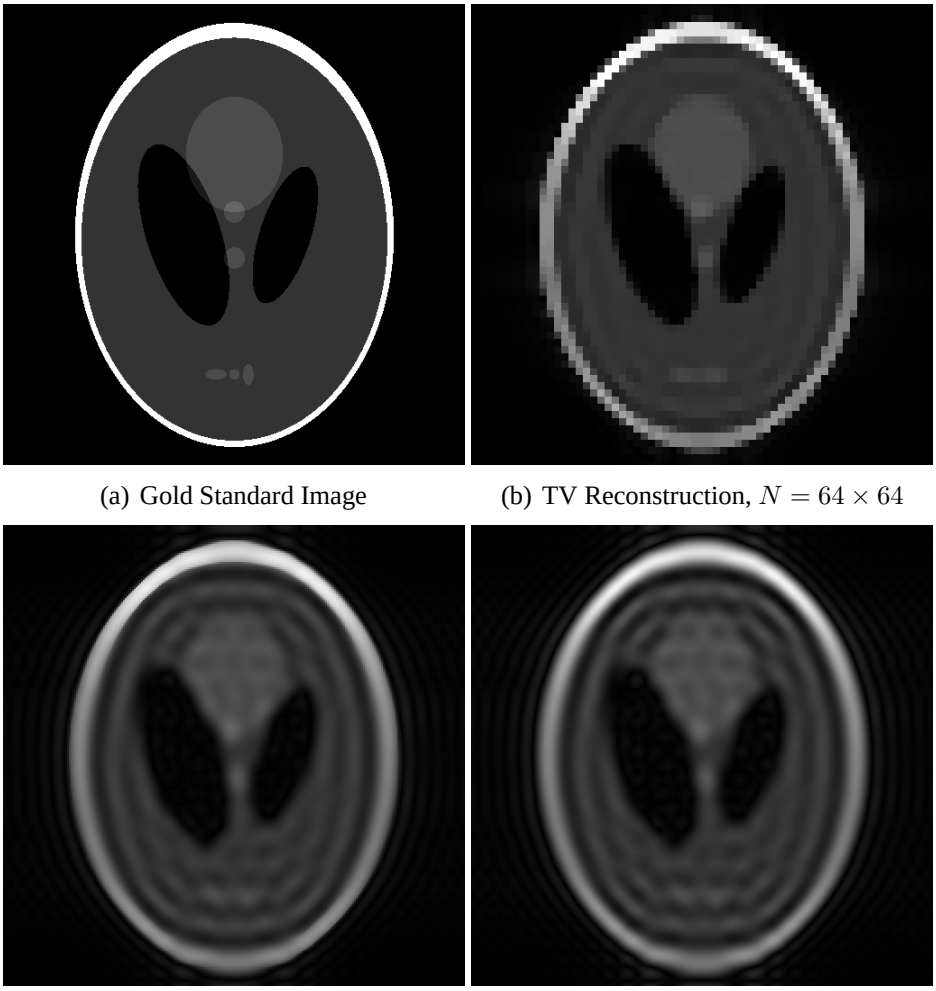
Noise characterizations for CS reconstruction can be obtained by brute-force Monte Carlo simulation methods, where the same image is reconstructed repeatedly under different noise realizations. However, due to the time consuming nature of CS reconstruction, this kind of approach is generally not computationally practical for real reconstruction problems. Approximate fast noise characterizations for general nonlinear reconstructions have also been studied [8, 208, 211, 524, 588]. However, as with the previously described fast resolution characterization methods, these methods cannot be applied to CS reconstruction due to the non-differentiability of the cost functional and/or the underdetermined nature of the $\mathbf{E}$ matrix. As a result, fast and easy characterization of the resolution and noise characteristics of CS-MRI remain important open problems.

A final open problem relates to the issue of choosing an appropriate image discretization for CS-MRI reconstruction. While real MRI images are continuous functions, this chapter made use of the finite-dimensional voxel-based image model described in Eq. (2.30) so that CS theory and

algorithms for finite-dimensional discrete inverse problems could be used in the MR context. However, this choice also introduces modeling errors. While modeling errors can often be characterized easily for linear reconstruction methods where point-spread functions or spatial-response functions can be derived (cf. Chapter 3), these modeling errors could be more pernicious for the nonlinear reconstruction schemes typical of CS-MRI. In this context the choice of the number of voxels N that are used in the image discretization could be an important issue. Frequently in MR, the choice of N is made based on the *k*-space coverage of the data – for example, if the *k*-space samples all fall within a region defined by a 64×64 Nyquist-rate grid, the reconstruction is often performed using a 64×64 voxel grid. This choice generally leads to smaller problem sizes and faster computation. However, an alternative approach would be to reconstruct with a large N (i.e., a very high-resolution voxel grid); this choice might seem intuitively better than the previous choice, since the discrete approximation of the continuous integral equation could be more accurate. However, some theoretical and empirical analysis has shown that high-resolution discretizations might not actually be able to correct modeling errors [138], and that as N grows large, the image prior and/or the reconstructed image do not always converge to a useful limit [384]. Figure 4.19 shows a simulation that illustrates some of the complex trade-offs in selecting the voxel grid size in CS-MRI. In practice, virtually all existing empirical evaluations of CS-MRI algorithms are guilty of what is known as an "inverse crime" – using a finite-dimensional image as the gold standard for comparison when solving a discretely-formulated inverse problem, even though the true images encountered in practical applications will be infinite dimensional [352]. While the results of these "criminal" experiments are still insightful, the characteristics of CS-MRI reconstruction in the presence of image modeling error are still not completely understood.

4.4 Conclusion and Summary

This chapter introduced a random encoding scheme for CS-MRI, replacing traditional phase encoding with RF encoding using randomized excitation profiles. This random scheme is conceptu-



(a) Gold Standard Image (b) TV Reconstruction, $N = 64 \times 64$

(c) TV Reconstruction, $N = 256 \times 256$ (d) TV Reconstruction, $N = 512 \times 512$

Figure 4.19: TV-based reconstructions of the Shepp-Logan phantom with different N. Simulated data was generated on a 32×32 Nyquist-rate grid in *k*-space using the analytic form of the Fourier transform for the Shepp-Logan phantom [493]. As N increases, the image model becomes more similar to that of a continuous image. However, while some image features are reconstructed more realistically for larger N, ringing artifacts also become more and more apparent as the grid size increases. The presence of ringing artifacts illustrates that the TV-based reconstruction is not able to accurately extrapolate the high-frequency *k*-space information from only low-resolution data. As a result, while larger N leads to better modeling of the real imaging process, it remains unclear how N should be chosen to yield good performance in practical experiments.

ally similar to the “universal” encoding schemes suggested by the CS literature, and simulations and experiments reveal that it has the potential to outperform Fourier-based schemes in certain high-SNR scenarios. However, our study also indicates that the random encoding scheme fails to satisfy the theoretical sufficient conditions for stable and accurate CS reconstruction in many scenarios of interest. Therefore, there is still no general theoretical performance guarantee for CS-MRI, with or without random encoding. As a result, the practical utility of CS methodology for MRI should be evaluated carefully for each application.

Chapter 5

Low-Rank Matrix Recovery for Spatiotemporal Imaging

The reconstruction schemes considered in the previous chapters were developed for contexts in which the image itself did not change significantly during the course of the experiment. In this chapter, we consider the reconstruction of spatiotemporal images $\rho(\mathbf{x}, t)$ that change over time.[1] Such experiments have a wide variety of applications, including dynamic cardiac imaging [217], dynamic contrast-enhanced (DCE) imaging [141, 373], interventional imaging [58], and fMRI [98].

In the context of Fourier imaging, spatiotemporal data acquisition can be modeled as [663]

$$s(\mathbf{k}_p, t_p) = \int_\Omega \rho(\mathbf{x}, t_p) \exp(-\imath 2\pi \mathbf{k}_p \cdot \mathbf{x}) \, d\mathbf{x} + \eta_p, \tag{5.1}$$

for $p = 1, 2, \ldots, P$, where $s(\mathbf{k}_p, t_p)$ is the pth measurement sample that takes place at *k*-space location $\mathbf{k}_p$ and time t_p, η_p is the corresponding noise perturbation, and P is the total number of sampled data points. In practice, a given *k*-space location might be sampled at multiple different times, meaning that the set of points $\{\mathbf{k}_p\}_{p=1}^P$ has repeated elements. As a result, we will introduce a new set of Q distinct points $\{\bar{\mathbf{k}}_q\}_{q=1}^Q$ representing the set of unique elements of $\{\mathbf{k}_p\}_{p=1}^P$. Similarly, we will introduce a new set of M distinct points $\{\bar{t}_m\}_{m=1}^M$ representing the set of unique elements of $\{t_p\}_{p=1}^P$.

For most medical imaging applications, it is commonly assumed that $\rho(\mathbf{x}, t)$ is a "conventional support-limited signal," i.e., that it is spatially support limited to a closed and bounded set Ω

[1] Some of the text and figures in this chapter have been previously published in [268, 281, 282], and are copyright of the IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

such that $\rho(\mathbf{x}, t) \approx 0$ for $\mathbf{x} \notin \Omega$ (cf. Sec. 2.2), and that it is temporally bandlimited such that $\tilde{\rho}(\mathbf{x}, f) \approx 0$ for $|f| > f_{\max}$, where

$$\tilde{\rho}(\mathbf{x}, f) = \int \rho(\mathbf{x}, t) \exp(-\imath 2\pi f t)\, dt, \tag{5.2}$$

and $f_{\max}$ is a finite constant. Similar to the case for static imaging (cf. Sec. 2.2), this spatial-spectral support-limited model means that dynamic images $\rho(\mathbf{x}, t)$ can be reconstructed from Nyquist-rate samples acquired in the k-t domain.

For many applications, it is desirable to reconstruct $\rho(\mathbf{x}, t)$ with high spatial and temporal resolution. However, as discussed in Sec. 2.1, MR spatial encoding is a relatively slow process due to physical and physiological constraints. While fast imaging sequences and parallel imaging technology can enable high-resolution two-dimensional image formation from data acquired in less than one second, many interesting physiological processes (e.g., human cardiac motion [563]) occur over significantly shorter time scales, and further acceleration is necessary for these scenarios. As a result, sampling k-t space at the Nyquist rate can place practical limits on the achievable spatiotemporal resolution.

There are many different signal processing approaches that have been proposed to overcome Nyquist limits in dynamic MRI applications (e.g., Refs. [6, 84–87, 92, 143, 187, 206, 228, 250, 251, 281, 282, 341, 348–350, 357, 373, 399, 402, 405, 420, 421, 426, 428, 454, 482, 501, 503, 516, 526, 540, 559, 569, 571, 610, 612, 615, 619, 622, 624, 664, 675, 676] and their references). Due to the relatively high SNR in many dynamic imaging scenarios, sparse-sampling approaches have featured prominently in the literature. These approaches sample both high- and low-frequency k-space data below the Nyquist rate, and use prior information to avoid reconstruction artifacts due to undersampling. Sliding window/interpolation/view-sharing methods [187, 373, 540, 619] make the assumption that certain regions of k-space change relatively slowly over time, so that missing data can be accurately interpolated from nearby time-points. Other techniques assume that the motion of the image between subsequent time points can be accurately predicted, leading

to a reduced set of parameters that need to be estimated [209, 348, 349, 482, 516]. Several other approaches leverage prior information to construct tailored spatial-spectral models of the image, which similarly simplifies the inverse problem [6,92,357,402,426,501,526,571,610,612]. Finally, there are techniques that make use of constraints on image structure derived directly from the data itself [84, 87, 206, 228, 250, 251, 281, 341, 348–350, 399, 405, 420, 421, 428, 454, 503, 569, 615, 622, 624, 664, 675]. Many of these techniques use the observation that dynamic images often demonstrate sparse structure in an appropriately-chosen transform domain [206, 228, 348–350, 420, 421, 622] (cf. the discussion of compressed sensing in the previous chapter) or demonstrate "low-rank" structure [84–87,143,250,251,281,282,341,399,428,503,559,569,624,664,675,676].[2]

In this chapter, we will focus on the *partial-separability* (PS) model [84–87, 143, 281, 282, 341, 399, 559, 569, 664, 675, 676], which can be used to capture low-rank structure. Similar to the sparsity constraints considered in the previous chapter, the use of rank constraints is based on the idea that spatiotemporal images frequently have considerably more structure than being support-limited. In particular, low-rank structure refers to the fact that the temporal behavior of $\rho(\mathbf{x}, t)$ at different spatial locations is often highly correlated. Indeed, the correlation is often so strong that we can observe approximate linear dependence relationships in *any* set of functions of the form $\{\rho(\mathbf{x}_\ell, t)\}_{\ell=1}^{L}$ for relatively small values of L [399]. Equivalently, low-rank structure reflects the observation that the temporal variations of $\rho(\mathbf{x}, t)$ are often approximately limited to a low-dimensional subspace.

It is interesting to note that this kind of low-dimensional structure is not specific to dynamic MRI. For example, it has been shown that high-quality low-rank representations exist for many operators and functions that appear in mathematics and physics problems [52, 53, 456]. In addition, low-rank or approximately low-rank structure is also frequently found in any high-dimensional dataset containing interrelated variables. In many cases, the correlation in such high-dimensional datasets can be described in terms of a much smaller set of latent variables, i.e., unobserved

[2] Also see Refs. [230,231] for an example of the use of rank constraints for dynamic imaging outside of the context of MRI.

variables that can still convey the essential features of the data [35]. This structure means that high-dimensional datasets can often be approximated using a small number of low-dimensional subspaces (with significantly fewer degrees of freedom) without significant loss of information. Similar to how sparsity has been used for compressing images (cf. the previous chapter), the approximation of high-dimensional data using a lower-dimensional subspace has been widely used for feature extraction and compression in a wide variety of fields [35, 238, 345, 367, 655]. As a result, the problem of reconstructing approximately low-rank signals from limited data has been considered by a large number of authors from many different communities, including chemistry [19], computer vision [96, 160, 573, 603], systems theory [201], and collaborative filtering (i.e., database-enabled recommendation systems where products/services are recommended to a user based on ratings from users with similar tastes) [1, 40, 369, 538, 594], among others. This has also motivated the recent development of theoretical results for low-rank reconstruction methods (e.g., see Refs. [1, 112–114, 118, 359, 360, 391, 392, 452, 535, 536]), many of which build upon the earlier CS theory.

Low-rank structure can be captured by the PS-based spatiotemporal image model, which is given by

$$\rho\left(\mathbf{x},t\right)=\sum_{\ell=1}^{L}\rho_{\ell}\left(\mathbf{x},t\right)=\sum_{\ell=1}^{L}u_{\ell}\left(\mathbf{x}\right)v_{\ell}\left(t\right), \tag{5.3}$$

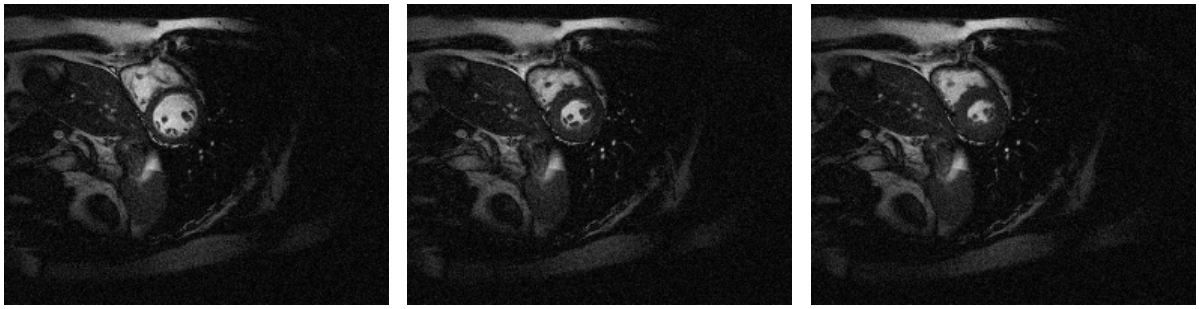
where L is the model-order. In this expression, the $\{\rho_{\ell}\left(\mathbf{x},t\right)\}_{\ell=1}^{L}$ are separable functions of $\mathbf{x}$ and t; i.e., they can be factored as $\rho_{\ell}\left(\mathbf{x},t\right)=u_{\ell}\left(\mathbf{x}\right)v_{\ell}\left(t\right)$, where $\{u_{\ell}\left(\mathbf{x}\right)\}_{\ell=1}^{L}$ and $\{v_{\ell}\left(t\right)\}_{\ell=1}^{L}$ are sets of signal dependent spatial and temporal functions, respectively. Given this model, the temporal signals $\{\rho(\mathbf{x}_{b},t)\}_{b=1}^{B}$ will have linear dependence for any set of points $\{\mathbf{x}_{b}\}_{b=1}^{B}$ if $B>L$. The PS model can be viewed as a generalization of the earlier DIME model for representing periodic dynamic images [402], which has the same form as Eq. (5.3), but restricted the functions $\{v_{\ell}\left(\cdot\right)\}_{\ell=1}^{L}$ to be complex sinusoids. Allowing general functions $\{v_{\ell}\left(\cdot\right)\}_{\ell=1}^{L}$ greatly expands the model class, and the set of PS-representable functions has been shown to be dense in the Hilbert space of square-integrable spatiotemporal functions [399].

While strict partial separability (i.e., $L = 1$) applies only to a "small" set of signals, higher-order partial separability ($L > 1$) significantly enhances the representational power of the model and has proven useful in a number of imaging applications. For example, low-order partially-separable representations have been used for image analysis and/or image reconstruction in the context of dynamic cardiac MRI [87, 642], dynamic MRI and emission tomographic imaging of contrast kinetics [75, 84, 198, 399], relaxation experiments [100, 175], diffusion experiments [21, 175, 490], fMRI [17, 350, 599, 648], and spectroscopic imaging [549], among others.[3] Figures 5.1 and 5.2 illustrate the low-rank characteristics of typical dynamic MR datasets.

In this chapter, we present a novel rank-constrained matrix framework for imaging with the PS model.[4] The proposed method is quite flexible and relies only on the use of rank constraints; this is in contrast to previous PS work, which generally made use of significant additional constraints on both data acquisition and image reconstruction [399]. This chapter is organized as follows. In Sec. 5.1, we establish some notation and pose spatiotemporal image reconstruction in terms of the recovery of an unknown matrix. In Sec. 5.2, we introduce the use of rank constraints, and describe several different approaches to enhanced image reconstruction by leveraging low-rank structure. In Sec. 5.3, we describe a novel algorithm that we have proposed for low-rank matrix reconstruction: Incremented Rank PowerFactorization (IRPF). Section 5.4 presents some application examples for our proposed formulation, while additional discussion is provided in Sec. 5.5. Finally, we conclude and summarize this chapter in Sec. 5.6.

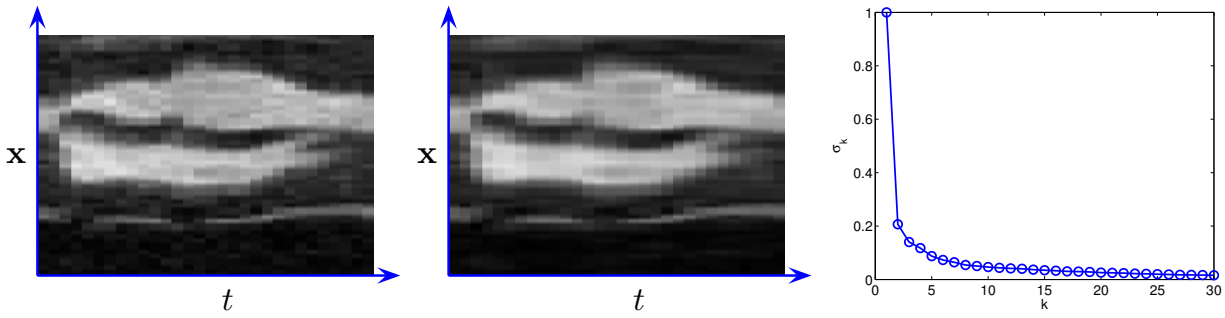
[3] Low-rank PS structure has also been exploited in the context of static (i.e., non-dynamic) imaging [119, 120, 427, 678], where the image was modeled as being approximately partially-separable along different spatial dimensions (e.g., $\rho(\mathbf{x}) \approx \sum_{\ell=1}^{L} \psi_\ell(x)\,\phi_\ell(y)$ in the 2D case with $\mathbf{x} = [x, y]^T$). However, typical images are frequently less separable (i.e., require larger L for accurate representation) along two different spatial dimensions than they are using spatiotemporal separability as in Eq. (5.3). As a result, the gains from using rank constraints are often more significant for dynamic imaging applications. Low-rank structure has also been used previously for interpolating missing *k*-space data [174]. However, the modeling assumptions for this case did not involve PS, instead assuming that *k*-space was linearly predictable.

[4] The review of matrix rank given previously in Sec. 2.3 will be helpful for understanding the material presented in this chapter.



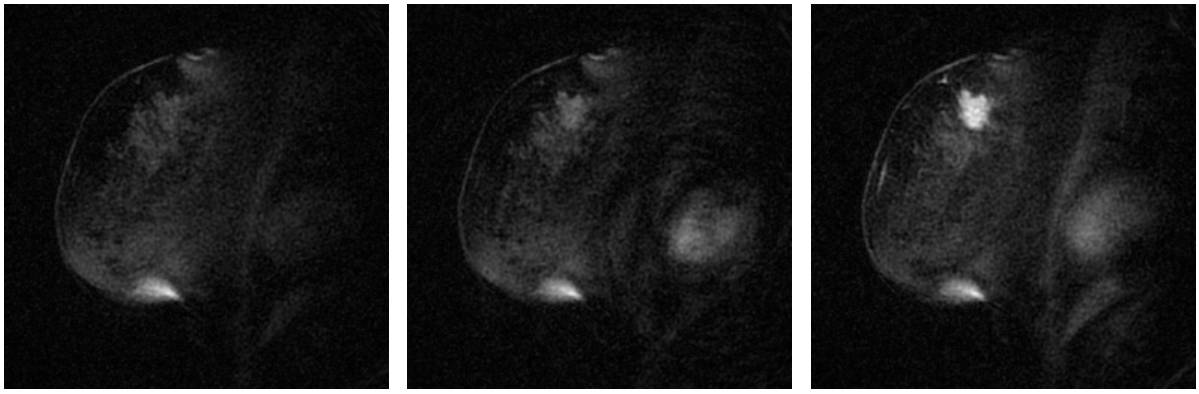
(a) Snapshot images

(b) Cardiac region of interest from the snapshot images in (a)

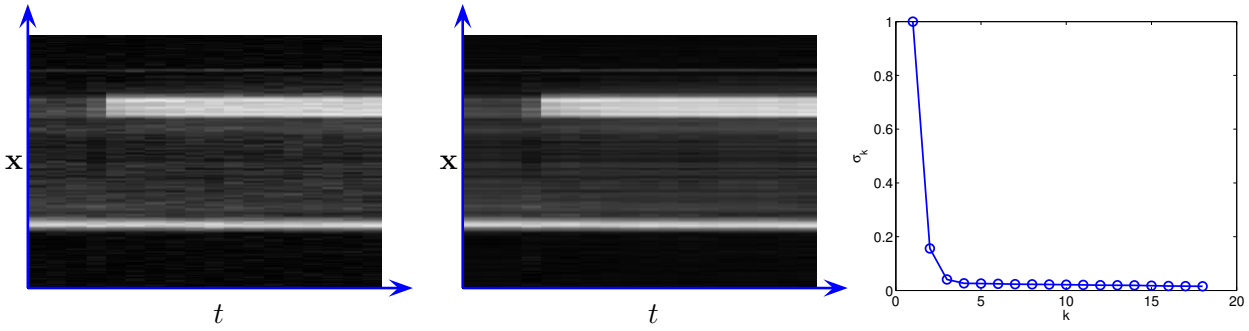


(c) Spatiotemporal profile (d) Rank-8 approximation (e) Singular Value Distribution

Figure 5.1: Low rank characteristics in cardiac cine imaging. (a) Three snapshot images from a 30-frame retrospectively-gated cardiac imaging experiment (courtesy of Dr. Peter Kellman). (b) Cardiac region of interest from the snapshot images in (a). (c) Spatiotemporal image corresponding to a vertical line passing through the heart. (d) Spatiotemporal image for the optimal rank-8 Eckart-Young approximation of the spatiotemporal matrix representation of this dataset (see Secs. 2.3 and 5.1 for details). The rank-8 approximation has less than 27% of the degrees of freedom of the full-rank model, but still captures the salient features of the cardiac motion. (e) The singular value distribution for this dataset (cf. Sec. 2.3). The large number of small singular values indicates that the spatiotemporal data matrix is approximately low-rank.



(a) Snapshot Images



(b) Spatiotemporal Profile (c) Rank-5 Approximation (d) Singular Value Distribution

Figure 5.2: Low rank characteristics in DCE breast imaging. (a) Three snapshot images from an 18-frame imaging experiment. (b) Spatiotemporal image corresponding to a vertical line passing through the contrast-enhancing tumor. (c) Spatiotemporal image for the optimal rank-5 Eckart-Young approximation of the spatiotemporal matrix representation of this dataset (see Secs. 2.3 and 5.1 for details). The rank-5 approximation has less than 28% of the degrees of freedom of the full-rank model, but still captures the important characteristics of the contrast kinetics. (d) The singular value distribution for this dataset (cf. Sec. 2.3). The large number of small singular values indicates that the spatiotemporal data matrix is approximately low-rank.

5.1 Matrix Recovery

This section establishes additional notation and describes the formulation of spatiotemporal image reconstruction as a general matrix-recovery problem. To simplify the discussion, we will focus on a finite-dimensional image model (cf. Sec. 2.2.3) where $\rho(\mathbf{x}, t)$ is represented through its samples on a grid of N spatial locations and M time points.[5] In this case, the image can be represented by the $N \times M$ Casorati matrix

$$\mathbf{C} = \begin{bmatrix} \rho(\mathbf{x}_1, \bar{t}_1) & \cdots & \rho(\mathbf{x}_1, \bar{t}_M) \\ \cdots & \ddots & \cdots \\ \rho(\mathbf{x}_N, \bar{t}_1) & \cdots & \rho(\mathbf{x}_N, \bar{t}_M) \end{bmatrix}, \tag{5.4}$$

where we have assumed without loss of generality that the set of time-points selected for representing the image is the same as the set of unique sampling times $\{\bar{t}_m\}_{m=1}^{M}$.

Given the representation of the spatiotemporal image in terms of the Casorati matrix, Eq. (5.1) can be written as

$$\mathbf{s} = \Xi(\mathbf{F}\mathbf{C}) + \boldsymbol{\eta}, \tag{5.5}$$

where $\mathbf{s}$ is the $P \times 1$ vector of data samples $s(\mathbf{k}_p, t_p)$, $\boldsymbol{\eta}$ is the $P \times 1$ vector of noise samples η_p, $\mathbf{F}$ is a $Q \times N$ matrix with entries $[\mathbf{F}]_{q,n} = \exp\left(-\imath 2\pi \bar{\mathbf{k}}_q \cdot \mathbf{x}_n\right)$, and $\Xi(\cdot) : \mathbb{C}^{Q \times M} \to \mathbb{C}^P$ is a linear sampling operator satisfying $[\Xi(\mathbf{S})]_p = [\mathbf{S}]_{q(p),m(p)}$, where $q(p) = \left\{q \in [1, \ldots, Q] : \bar{\mathbf{k}}_q = \mathbf{k}_p\right\}$, and $m(p) = \{m \in [1, \ldots, M] : \bar{t}_m = t_p\}$.

The matrix recovery problem is to estimate the unknown matrix $\mathbf{C}$ from the measurement vector $\mathbf{s}$.

[5] An alternative discretization scheme is presented in [281, 399], in which the data domain (i.e., k-t space) is discretized instead of the image domain. Image-domain discretization can facilitate the use of additional constraints (e.g., spatial regularization), while data-domain discretization can be more natural (due to the discrete nature of the sampling process) and lead to simpler reconstruction algorithms. Many other discretizations are also possible.

5.1.1 Unconstrained Matrix-Based Reconstruction

Without imposing any additional constraints, the maximum likelihood estimator of $\mathbf{C}$ under a white Gaussian noise model is given by the solution to the standard linear least-squares problem:

$$\hat{\mathbf{C}}_{\mathrm{ML}} = \arg \min_{\mathbf{C} \in \mathbb{C}^{N \times M}} \left\| \Xi \left(\mathbf{F}\mathbf{C} \right) - \mathbf{s} \right\|_{\ell_2}^2 . \tag{5.6}$$

The least-squares optimality condition for a global minimizer is that $\hat{\mathbf{C}}_{\mathrm{ML}}$ satisfies the normal equations:

$$\mathbf{F}^H \Xi^* \left(\Xi \left(\mathbf{F}\hat{\mathbf{C}}_{\mathrm{ML}} \right) \right) - \mathbf{F}^H \mathbf{S} = 0, \tag{5.7}$$

where $\Xi^* \left(\cdot \right) : \mathbb{C}^P \rightarrow \mathbb{C}^{Q \times M}$ is the adjoint of $\Omega \left(\cdot \right)$ (akin to a zero-filling operation), and $\mathbf{S} = \Xi^* \left(\mathbf{s} \right)$. Equation (5.7) is separable in the columns of $\hat{\mathbf{C}}_{\mathrm{ML}}$, resulting in the expressions:

$$\mathbf{F}^H \mathrm{diag} \left(\boldsymbol{\zeta}_m \right) \mathbf{F} \hat{\mathbf{c}}_m = \mathbf{F}^H \mathbf{s}_m \tag{5.8}$$

for $m = 1, \ldots, M$, where $\hat{\mathbf{c}}_m$ and $\mathbf{s}_m$ are the mth columns of $\hat{\mathbf{C}}_{\mathrm{ML}}$ and $\mathbf{S}$, respectively, $\{ \boldsymbol{\zeta}_m \}_{m=1}^M$ is the set of $Q \times 1$ vectors satisfying $\left[\boldsymbol{\zeta}_m \right]_q = 1$ when $\left(\bar{\mathbf{k}}_q, t_m \right) \in \{ (\mathbf{k}_p, t_p) \}_{p=1}^P$, and $\left[\boldsymbol{\zeta}_m \right]_q = 0$ otherwise, and $\mathrm{diag} \left(\boldsymbol{\zeta}_m \right)$ is a diagonal matrix with diagonal entries equal to the entries of the vector $\boldsymbol{\zeta}_m$. The necessary and sufficient condition for the uniqueness of a solution is that the normal matrices in Eq. (5.8) are invertible. Note that the $N \times N$ matrix $\mathbf{F}^H \mathrm{diag} \left(\boldsymbol{\zeta}_m \right) \mathbf{F}$ cannot be invertible unless $\{ \boldsymbol{\zeta}_m \}_{m=1}^M$ has at least N non-zero elements. As a result, each row of $\mathbf{FC}$ should be sampled at least N times, meaning that a total of at least NM samples are needed to have a well-posed reconstruction problem. As will be shown in the next section, the use of rank constraints can significantly relax these sampling requirements.

5.2 Low-Rank Matrix Recovery

5.2.1 Rank-Constrained Matrix Recovery with the PS Model

A consequence of using the Lth-order PS model, as introduced at the beginning of this chapter, is that for any sets of spatial locations $\{\mathbf{x}_n\}_{n=1}^N$ and time points $\{\bar{t}_m\}_{m=1}^M$, the Casorati matrix in Eq. (5.4) has at most rank L [399]. This can be easily understood by using the model for $\rho(\mathbf{x}, t)$ in Eq. (5.3) to decompose $\mathbf{C}$ as the sum of L rank-1 matrices:

$$\mathbf{C} = \sum_{\ell=1}^{L} \begin{bmatrix} u_\ell(\mathbf{x}_1) \\ \vdots \\ u_\ell(\mathbf{x}_N) \end{bmatrix} \begin{bmatrix} v_\ell(\bar{t}_1) & \cdots & v_\ell(\bar{t}_M) \end{bmatrix}. \tag{5.9}$$

This low-rank structure implies that use of the Lth-order PS model can considerably reduce sampling requirements. In particular, an $N \times M$ complex matrix of rank L has NM entries (specified by $2NM$ real numbers), but as the SVD illustrates (cf. Sec. 2.3), it has significantly fewer degrees of freedom. In particular, there are L real-valued degrees of freedom associated with the singular values themselves, there are at most $2NL - L^2$ degrees of freedom for the columns of $\mathbf{P}$ associated with non-zero singular values,[6] and there are at most $2ML - L^2 - L$ degrees of freedom for the columns of $\mathbf{Q}$ associated with non-zero singular values.[7] As a result, a rank-L matrix only has up to $2(N + M - L)L$ real-valued degrees of freedom. Thus, as few as $L(M + N - L)$ complex-valued samples could suffice for reconstruction, which would lead to

[6] Note that there are NL degrees of freedom for each of the real and imaginary parts of $\mathbf{P}$. This number is reduced because of constraints. In particular, there are L constraints on the magnitudes of the first L columns of $\mathbf{P}$, and $2\binom{L}{2}$ orthogonality constraints. As a result, there are at most $2NL - L(L-1) - L = 2NL - L^2$ degrees of freedom.

[7] The computation of the number of degrees of freedom of $\mathbf{Q}$ is similar to the computation for $\mathbf{P}$. However, unlike the case for $\mathbf{P}$, we apply L additional constraints to control the phase of the L different columns of $\mathbf{Q}$ (i.e., we force the first non-zero entry of each $\mathbf{q}_\ell$ to be real and nonnegative). This is necessary because for any set of scalars $\{a_\ell\}_{\ell=1}^L$ with $|a_\ell| = 1$, the sets of vectors $\left\{\frac{1}{a_\ell}\mathbf{p}_\ell\right\}_{\ell=1}^L$ and $\{a_\ell^*\mathbf{q}_\ell\}_{\ell=1}^L$ will be orthonormal if $\{\mathbf{p}_\ell\}_{\ell=1}^L$ and $\{\mathbf{q}_\ell\}_{\ell=1}^L$ are orthonormal, and $\frac{1}{a_\ell}\mathbf{p}_\ell a_\ell \mathbf{q}_\ell^H = \mathbf{p}_\ell \mathbf{q}_\ell^H$. Controlling the phase of the columns of $\mathbf{Q}$ eliminates this problem of having non-unique representation of $\mathbf{p}_\ell \mathbf{q}_\ell^H$. Combining terms, there are at most $2ML - L(L-1) - 2L = 2ML - L^2 - L$ degrees of freedom.

considerable savings over the NM samples required without the PS constraint, particularly when L is small relative to N and M. The constrained analog to Eq. (5.6) is

$$\hat{\mathbf{C}}_{\mathrm{PS}} = \arg \min_{\substack{\mathbf{C} \in \mathbb{C}^{N \times M} \\ \mathrm{rank}(\mathbf{C}) \leq L}} \left\| \Xi \left(\mathbf{F}\mathbf{C} \right) - \mathbf{s} \right\|_{\ell_2}^2 . \tag{5.10}$$

While this optimization problem looks very similar to Eq. (5.6), the nonlinearity of the rank constraint means that it is considerably more difficult to solve than a set of decoupled linear least-squares problems.

The problem formulation given in Eq. (5.10) will be the main focus of the rest of this chapter. However, before discussing this problem further, we will first review some other existing problem formulations for the recovery of low-rank matrices from limited data. While the PS-based reconstruction in Eq. (5.10) makes use of explicit rank constraints where L is assumed to be known, there also exist matrix recovery formulations where low-rank matrices are reconstructed from limited measurements using softer constraints. For the sake of generality, we will describe these alternative formulations with our previous MR-specific data measurement operator replaced by an arbitrary linear data-measurement operator $\mathcal{A} : \mathbb{C}^{N \times M} \rightarrow \mathbb{C}^{P}$. The data acquisition model from Eq. (5.5) can be obtained as the special case where $\mathcal{A}\left(\mathbf{C} \right) = \Xi \left(\mathbf{F}\mathbf{C} \right)$.

One alternative to Eq. (5.10) is the affine rank-minimization problem [201, 535]:

$$\hat{\mathbf{C}} = \arg \min_{\substack{\mathbf{C} \in \mathbb{C}^{N \times M} \\ \left\| \mathcal{A}(\mathbf{C}) - \mathbf{s} \right\|_{\ell_2} \leq \varepsilon}} \mathrm{rank} \left(\mathbf{C} \right) , \tag{5.11}$$

where ε represents a noise tolerance. This formulation is very similar to that of Eq. (5.10), with the main difference being that explicit constraints have shifted from the rank of the reconstructed matrix to the data consistency of the low-rank solution. As a result of this similarity, algorithms to solve one of these problems can also be used to solve the other. For example, noting that the optimal cost function value for Eq. (5.10) is monotonically decreasing in L, the solution to Eq. (5.11) can be achieved by solving Eq. (5.10) for increasing values of L until $\left\| \mathcal{A}\left(\mathbf{C} \right) - \mathbf{s} \right\|_{\ell_2} \leq$

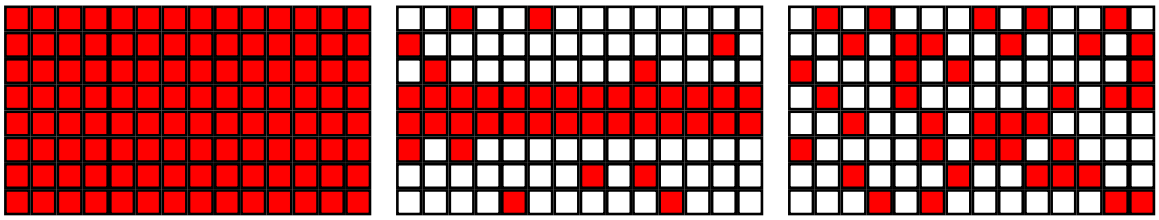
ε. Similarly, a solution to Eq. (5.10) can be solved by adjusting ε in Eq. (5.11) until the desired rank constraint is achieved. It should be noted that affine rank-minimization is known to be NP-hard [535].

Another alternative to Eq. (5.10) is based on regularization [1, 202, 203, 250, 427, 535]:

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C} \in \mathbb{C}^{N \times M}} \| \mathcal{A}(\mathbf{C}) - \mathbf{s} \|_{\ell_2}^2 + \lambda R(\mathbf{C}), \tag{5.12}$$

where $R(\cdot)$ is a regularization functional that favors matrices with low rank, and λ is a regularization parameter. Choices of $R(\cdot)$ that have been used previously include the nuclear norm (NN) ($R_{\text{NN}}(\mathbf{C}) = \sum_i \sigma_i$, where σ_i are the singular values of $\mathbf{C}$) [535], the log-determinant functional ($R_{\text{logdet}}(\mathbf{C}) = \sum_i \ln(\sigma_i + \delta)$, where δ is a small constant) [203], and the Schatten p-norm ($R_{\text{NN}}(\mathbf{C}) = \sum_i \sigma_i^p$) with $0 < p < 1$ [250, 427]. Other reasonable choices could include information-theoretic model selection criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) [97, 381, 591], which have been previously proposed in the context of sparsity-based reconstruction [28].

The NN-based regularization scheme is of particular interest [113, 535, 672]. The NN is the tightest convex relaxation of matrix rank, in just the same way that the ℓ_1-norm is the tightest convex relaxation of the ℓ_0-norm in the context of sparse-vector recovery [117]. Due to convexity, problems involving the NN can be solved globally using efficient algorithms [102, 247, 414, 423, 446, 535, 602]. And, similar to ℓ_1-minimization, the solution using NN minimization (NNM) can be proven to be equivalent to the solution to Eq. (5.11) under appropriate constraints on $\mathcal{A}$, $\mathbf{C}$, λ, and ε. The Schatten p-norm and the log-determinant functional are nonconvex relaxations of matrix rank. As a result, achieving global optimization with these functionals is nontrivial. However, similar to the discussion of sparse-vector reconstruction from the previous chapter, the use of nonconvex functionals can lead to meaningful improvements in sampling requirements and/or reconstruction quality.



(a) (b) (c)

Figure 5.3: Illustrations of the special sampling schemes considered in Sec. 5.2.2. Red matrix elements are sampled by $\Xi(\cdot)$. (a) Full sampling. (b) Specialized undersampling with a small number of fully-sampled "navigator"-rows. (c) General undersampling.

5.2.2 Sampling Considerations and Performance Guarantees

An important practical question is whether the various low-rank matrix recovery formulations lead to significant improvements in spatiotemporal data acquisition and image reconstruction. As we will see, the difficulty of numerical optimization and the strength of reconstruction performance guarantees are highly-related to the k-t space sampling pattern. For general linear measurement operators $\mathcal{A}$, theoretical results have been derived [112–114, 118, 359, 360, 391, 392, 392, 452, 535, 536] that provide sufficient conditions on $\mathcal{A}$ and $\mathbf{C}$ such that certain algorithms for solving Eqs. (5.10)-(5.12) and their variations are guaranteed to be successful. These sufficient conditions often require that $\mathcal{A}$ obeys a *rank-restricted isometry property* (rRIP), i.e., that $\|\mathcal{A}(\mathbf{C})\|_{\ell_2} \approx \|\mathbf{C}\|_F$ for any sufficiently low-rank $\mathbf{C}$.[8] While computing rRIPs is computationally intractable, it is known that certain classes of randomized sampling operators have good rRIP properties, as long as P is large enough [452, 535].

While the rRIP results with general $\mathcal{A}$ indicate that rank constraints can lead to meaningful reductions in sampling requirements, additional insight can be gained by analyzing the matrix recovery problem in the context of common MR-specific data-acquisition strategies. We illustrate three common sampling patterns in Fig. 5.3, which we will discuss in detail below.

[8]The definition of the rRIP is very similar to the definition of the RIP from the previous chapter. In addition, the performance guarantees for low-rank matrix recovery based on the rRIP are very similar to the guarantees for sparse vector recovery based on the RIP. This is not a coincidence, since there are many parallels between sparse vector recovery and low-rank matrix recovery. See Ref. [535] for more detail.

Full Sampling and $\mathbf{F}^H$ is a Tight Frame

Under full sampling conditions (i.e., $\Xi^*\left(\Xi\left(\cdot\right)\right)$ is an identity operator) and assuming that $\mathbf{F}^H$ is a tight frame (i.e., $\mathbf{F}^H\mathbf{F} = \alpha\mathbf{I}$, where $\alpha > 0$ and $\mathbf{I}$ is the $Q \times Q$ identity matrix),[9] the optimal solution to Eq. (5.10) is given by $\hat{\mathbf{C}}_{\text{PS}} = \tilde{\mathbf{H}}_L/\alpha$, where $\tilde{\mathbf{H}}_L$ is the rank-L Eckart-Young (EY) approximation of $\mathbf{H} = \mathbf{F}^H\mathbf{S}$. In this case, it can be shown (using Lemma 1.1 of [4]) that $\hat{\mathbf{C}}_{\text{PS}}$ satisfies

$$\begin{aligned}\|\mathbf{C} - \hat{\mathbf{C}}_{\text{PS}}\|_F \leq\ & \underbrace{\|\mathbf{C} - \tilde{\mathbf{C}}_L\|_F}_{\text{Approximation Error}} \\ & \underbrace{+ \frac{1}{\alpha}\|\tilde{\mathbf{N}}_L\|_F + \frac{2}{\sqrt{\alpha}}\sqrt{\|\tilde{\mathbf{N}}_L\|_F\|\tilde{\mathbf{C}}_L\|_F}}_{\text{Noise Error}},\end{aligned} \tag{5.13}$$

where $\tilde{\mathbf{C}}_L$ and $\tilde{\mathbf{N}}_L$ are the rank-L EY approximations of $\mathbf{C}$ and $\mathbf{F}^H\Xi^*\left(\boldsymbol{\eta}\right)$, respectively. Thus, when the low-rank approximation is good and the noise is not too large, we obtain good performance guarantees and a tractable reconstruction algorithm. Unfortunately, the requirements on $\Xi\left(\cdot\right)$ and $\mathbf{F}$ for this case can only be satisfied if $P \geq NM$. Thus, this setting is not useful for reconstruction from limited data, though it and its variations have proven useful for denoising, dimensionality-reduction, and analysis of spatiotemporal imaging data (e.g., [21, 75, 100, 222, 549, 599, 648]).

Specialized Undersampling with Navigator Data and F has a Trivial Null Space

Rather than fully sampling k-t space as in the previous section, we could instead manipulate the data acquisition physics to ensure that matrix recovery can be achieved with undersampled data using simple algorithms [399]; this approach has been utilized heavily in the existing PS literature, and will be referred to as the "basic PS" approach. We consider the case where the sampling operator $\Xi\left(\cdot\right)$ has been designed to fully measure at least L rows of the matrix $\boldsymbol{\Phi} = \mathbf{F}\mathbf{C}$. We will denote this fully-sampled submatrix of $\boldsymbol{\Phi}$ as $\boldsymbol{\Phi}_n$, with the corresponding submatrix of $\mathbf{F}$ denoted

[9] The assumption that $\mathbf{F}^H$ is a tight frame is reasonable for Cartesian-sampling of *k*-space, in which case $\mathbf{F}$ will often be proportional to the unitary DFT matrix.

as $\mathbf{F}_n$ such that $\mathbf{\Phi}_n = \mathbf{F}_n\mathbf{C}$. In the existing PS literature, $\mathbf{\Phi}_n$ is often called navigator or training data. We have the following result:

Theorem 5.1. *Let* $\mathbf{C}$ *have rank* L*, assume* $\boldsymbol{\eta} = \mathbf{0}$*, and let* $\mathbf{F}_n$ *and* $\mathbf{C}$ *be such that* $\mathbf{\Phi}_n$ *has rank* L*. Then the* L*-dimensional row-space of* $\mathbf{C}$ *is equal to the* L*-dimensional row-space of* $\mathbf{\Phi}_n$*.*

This follows from the fundamental theorem of linear algebra.

Corollary 5.1. *Under the conditions of the previous theorem,* $\mathbf{C}$ *can be represented as* $\mathbf{C} = \mathbf{R}\mathbf{Q}^H$ *for some matrix* $\mathbf{R} \in \mathbb{C}^{N\times L}$*, where* $\mathbf{Q}$ *is the* $M \times L$ *matrix of right singular vectors from the SVD of* $\mathbf{\Phi}_n$*.*

Assuming $\mathbf{Q}$ from the corollary is given, the nonlinear optimization problem in Eq. (5.10) can be recast in terms of the linear least-squares problem

$$\hat{\mathbf{R}} = \arg\min_{\mathbf{R}\in\mathbb{C}^{N\times L}} \left\| \Omega\left(\mathbf{F}\mathbf{R}\mathbf{Q}^H\right) - \mathbf{s} \right\|_{\ell_2}^2 . \tag{5.14}$$

The least-squares optimality condition for the solution to Eq. (5.14) is given by the normal equations

$$\mathbf{F}^H \Xi^* \left(\Xi \left(\mathbf{F}\hat{\mathbf{R}}\mathbf{Q}^H \right) \right) \mathbf{Q} - \mathbf{F}^H \Xi^* \left(\mathbf{s} \right) \mathbf{Q} = 0. \tag{5.15}$$

While this problem may seem complicated, the optimal $\hat{\mathbf{R}}$ can be computed efficiently using the same techniques that will be described for the IRPF algorithm in Sec. 5.3. Note also that when $\mathbf{F}$ is square and full rank, the solution to Eq. (5.14) can be equivalently obtained by solving for the matrix $\hat{\mathbf{Z}} = \mathbf{F}\hat{\mathbf{R}}$, and setting $\hat{\mathbf{R}} = \mathbf{F}^{-1}\hat{\mathbf{Z}}$. After this simplification, Eq. (5.14) becomes separable in the rows of $\mathbf{Z}$, leading to further improvements in computational efficiency.

In addition to fast computations, use of this sampling scheme has the advantage that only a very small number of k-t samples are required for reconstruction. In particular, at least LM samples are required to obtain $\mathbf{\Phi}_n$. In addition, at least $NL - L^2$ other samples of $\mathbf{\Phi}$ must be acquired to ensure that the total number of samples is at least as large as the number of degrees of freedom in $\mathbf{C}$. This minimum number of additional samples is achievable. For example, in the special case

where $\mathbf{F}$ is an identity matrix,[10] a unique reconstruction for $\mathbf{R}$ can only be obtained if each row of $\mathbf{\Phi} = \mathbf{C}$ is sampled at least L times. Though the LM samples for $\mathbf{\Phi}_n$ already cover L of the rows of $\mathbf{\Phi}$, there remain $N - L$ rows of $\mathbf{\Phi}$ that must each be sampled at least L times, leading to $NL - L^2$ additional required samples. As a result, the use of a specialized sampling scheme can potentially achieve the minimum $L(N + M - L)$ required samples.

However, one limitation of this approach is that the row-space of $\mathbf{C}$ is estimated only from a subset of the measured data. As a result, this procedure implicitly assumes that $\mathbf{\Phi}_n$ has the same dominant temporal subspace as $\mathbf{C}$, and neglects the effects of noise and modeling error.

General Undersampling and F is Square and Invertible

When $\mathbf{F}$ is square and invertible, the $\hat{\mathbf{C}}_{\text{PS}}$ obtained from solving Eq. (5.10) can equivalently be obtained by solving

$$\hat{\mathbf{\Phi}} = \arg \min_{\substack{\mathbf{\Phi} \in \mathbb{C}^{Q \times M} \\ \text{rank}(\mathbf{\Phi}) \leq L}} \|\Xi(\mathbf{\Phi}) - \mathbf{s}\|_{\ell_2}^2 , \tag{5.16}$$

and setting $\hat{\mathbf{C}}_{\text{PS}} = \mathbf{F}^{-1}\hat{\mathbf{\Phi}}$. The problem in Eq. (5.16) has recently been termed the low-rank *matrix completion* problem, and variations of this problem have recently been analyzed theoretically by a number of authors [113, 114, 118, 359, 360, 452]. The problem in Eq. (5.16) is non-convex, and the affine rank-minimization variation of Eq. (5.16) is NP-hard [535]. Nevertheless, there exist an ever-increasing number of relaxations and greedy heuristics to approximately solve this and related problems including the alternative formulations in Eqs. (5.11) and (5.12). For some completion algorithms, there also exist theoretical performance guarantees for these methods [113, 114, 118, 359, 360, 452]. Interestingly, it has been shown that for most low-rank matrices, if the operator $\Xi(\cdot)$ samples a uniformly random subset of the matrix entries and if the number of samples is large enough, then Eq. (5.16) is guaranteed to have a unique global minimizer that can be obtained using existing fast algorithms [113, 360]. In the presence of noise, this leads to performance guarantees that are weaker but have a similar form to Eq. (5.13). Importantly, high-quality guarantees can

[10] Note that whenever $\mathbf{F}$ is square and invertible, the previously described change of variables $\hat{\mathbf{Z}} = \mathbf{F}\hat{\mathbf{R}}$ can be used to formulate an equivalent optimization problem where $\mathbf{F}$ can be treated as an identity matrix.

be obtained when the number of measurements P is as small as $O\left(LN \ln N\right)$ [360], where we have assumed $N \leq M$. This result indicates that PS images can be recovered robustly from undersampled k-t space data without the use of specially-tailored sampling patterns, as long as appropriate algorithms are used for reconstruction.

Compared to the previous two special cases, the optimization problems in Eqs. (5.10) and (5.16) are flexible enough to accommodate very general sampling schemes. In addition, in contrast to the basic PS procedure, the general matrix completion/recovery approach makes use of all of the measured data to estimate the structure of the low-rank matrix $\mathbf{C}$. The price that is paid for this increased flexibility is that sampling requirements are no longer as easy to analyze, and that the optimization problem becomes significantly more complicated.

Algorithms for solving Eqs. (5.10) and (5.16) have existed in the literature for a long time. One of the earliest methods is based on alternating least-squares (ALS) . Early examples include the use of ALS techniques for nonlinear iterative partial least squares matrix factorizations (NIPALS) [654, 655], and low-rank decomposition of tensor-valued data with the CANDECOMP/PARAFAC model [89, 126, 293, 367], where a tensor is a higher-dimensional generalization of a matrix [367]. Also see Refs. [201, 224, 258, 268, 294, 467, 623] for descriptions of related ALS-based low-rank matrix recovery methods. Other approaches to solving Eqs. (5.10) and (5.16) include alternating projection algorithms [255], gradient descent and expectation-maximization algorithms [160, 467, 585], Newton methods [96, 483], optimization over Grassmann manifolds [153, 359, 360], and projected gradient algorithms [452]. Additionally, as described previously, many algorithms exist for solving the related problem formulations given in Eqs. (5.11) and (5.12). In this work, we will make use of the IRPF algorithm [268] to solve matrix completion/recovery problems. The IRPF algorithm, which falls into the class of ALS techniques, is described in detail in the next section.

5.3 The Incremented Rank PowerFactorization Algorithm

The incremented-rank PowerFactorization (IRPF) algorithm was designed to solve problems of the form

$$\hat{\mathbf{C}} = \arg \min_{\substack{\mathbf{C} \in \mathbb{C}^{N \times M} \\ \text{rank}(\mathbf{C}) \leq L}} \|\mathcal{A}(\mathbf{C}) - \mathbf{s}\|_{\ell_2}, \tag{5.17}$$

where $\mathcal{A} : \mathbb{C}^{M \times N} \rightarrow \mathbb{C}^{P}$ is an arbitrary linear operator.

IRPF is an ALS approach that makes use of the factorization $\mathbf{C} = \mathbf{UV}$, with $\mathbf{U} \in \mathbb{C}^{N \times L}$ and $\mathbf{V} \in \mathbb{C}^{L \times M}$, to enforce rank-$L$ structure implicitly. Subsequently, we seek a local minimum of

$$\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\} = \arg \min_{\substack{\mathbf{U} \in \mathbb{C}^{N \times L} \\ \mathbf{V} \in \mathbb{C}^{L \times M}}} \|\mathcal{A}(\mathbf{UV}) - \mathbf{s}\|_{\ell_2} \tag{5.18}$$

using an alternating minimization procedure over $\mathbf{U}$ and $\mathbf{V}$.

IRPF is similar to and inherits its name from the PowerFactorization algorithm introduced by Hartley and Schaffalitzky [294, 623]. Without being aware of PowerFactorization and many of the other ALS approaches for low-rank matrix recovery problems, Diego Hernando and I proposed and evaluated IRPF as part of a course project for a graduate class in the Fall 2008 semester at the University of Illinois at Urbana-Champaign. Our original name for IRPF was FARM (Fast Alternating-subsets descent algorithm for Rank-constrained Matrix recovery). After viewing our presentation on FARM, Professor Yi Ma informed us of the PowerFactorization algorithm, which inspired us to rename the algorithm. However, it should be noted that most of the previous ALS approaches (including PowerFactorization) assume the matrix completion problem, where $\mathcal{A}$ has the form of a sub-sampled identity operator. Unlike these methods, our formulation of IRPF works for arbitrary linear operators, and has the flexibility to solve both Eqs. (5.10) and (5.16). Empirical results indicate that IRPF is fast, works well for matrix recovery problems, and can give better solutions than alternatives like NNM and PowerFactorization [268] for solving general low-rank matrix recovery and completion problems.

5.3.1 Algorithm Description

Let the action of the linear operator $\mathcal{A}$ be expressed as

$$\begin{aligned}\left[\mathcal{A}\left(\mathbf{C}\right)\right]_p &= \sum_{n=1}^{N}\sum_{m=1}^{M} a_{nmp}\left[\mathbf{C}\right]_{nm} \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M} a_{nmp}\sum_{\ell=1}^{L}\left[\mathbf{U}\right]_{n\ell}\left[\mathbf{V}\right]_{\ell m}\end{aligned} \tag{5.19}$$

for appropriate constants a_{nmp}, and for $p = 1, 2, \ldots, P$. In this case, we can define

$$\mathcal{A}\left(\mathbf{UV}\right) \equiv \mathbf{A}_{\mathbf{U}}\mathrm{vec}\left(\mathbf{V}\right) \equiv \mathbf{A}_{\mathbf{V}}\mathrm{vec}\left(\mathbf{U}\right), \tag{5.20}$$

where $\mathrm{vec}\left(\cdot\right)$ stacks the columns of its matrix argument into a single column vector. The matrices $\mathbf{A}_{\mathbf{U}} \in \mathbb{C}^{P\times LN}$ and $\mathbf{A}_{\mathbf{V}} \in \mathbb{C}^{P\times ML}$ are defined as

$$\left[\mathbf{A}_{\mathbf{U}}\right]_{(p,\ell+L(m-1))} = \sum_{n=1}^{N} a_{nmp}\left[\mathbf{U}\right]_{n\ell} \tag{5.21}$$

and

$$\left[\mathbf{A}_{\mathbf{V}}\right]_{(p,n+M(\ell-1))} = \sum_{m=1}^{M} a_{nmp}\left[\mathbf{V}\right]_{\ell m}, \tag{5.22}$$

respectively, for $p \in \{1, \ldots, P\}$, $\ell \in \{1, \ldots, L\}$, $m \in \{1, \ldots, M\}$, and $n \in \{1, \ldots, N\}$.

PowerFactorization

The general PowerFactorization algorithm iterates by alternatingly optimizing $\mathbf{U}$ and $\mathbf{V}$ using a linear least-squares procedure. The earlier NIPALS algorithm [654, 655] can be considered as a special case of PowerFactorization, with $L = 1$. PowerFactorization (modified from Refs. [294, 623] to permit a general linear operator $\mathcal{A}$) runs as follows:

1. Initialize $\left(\mathbf{U}^{(0)} \in \mathbb{C}^{N\times L}, \mathbf{V}^{(0)} \in \mathbb{C}^{L\times M}\right)$. Set iteration number $q = 0$.

2. Holding $\mathbf{V}^{(q)}$ fixed, find $\mathbf{U}^{(q+1)}$ by solving

$$\mathbf{U}^{(q+1)} = \arg\min_{\mathbf{U}} \left\| \mathbf{A}_{\mathbf{V}^{(q)}} \mathrm{vec}\left(\mathbf{U}\right) - \mathbf{b} \right\|_2^2 . \tag{5.23}$$

3. Now fixing $\mathbf{U}^{(q+1)}$, find $\mathbf{V}^{(q+1)}$ by solving

$$\mathbf{V}^{(q+1)} = \arg\min_{\mathbf{V}} \left\| \mathbf{A}_{\mathbf{U}^{(q+1)}} \mathrm{vec}\left(\mathbf{V}\right) - \mathbf{b} \right\|_2^2 . \tag{5.24}$$

4. Increment q. If q exceeds a maximum number of iterations, if the iterations stagnate, or if the relative error $\left\| \mathcal{A}\left(\mathbf{U}^{(q)}\mathbf{V}^{(q)}\right) - \mathbf{b} \right\|_2 / \left\| \mathbf{b} \right\|_2$ is smaller than a desired threshold ε, then terminate the iterative procedure. Otherwise, repeat steps 2-4.

Various options exist for choosing the initialization $\left(\mathbf{U}^{(0)} \in \mathbb{C}^{N \times L}, \mathbf{V}^{(0)} \in \mathbb{C}^{L \times M}\right)$, and reasonable choices include random initialization or an initialization based on the EY approximation of $\mathcal{A}^*\left(\mathbf{s}\right)$. Note that for large-scale problems, the EY approximation can be obtained efficiently by using fast algorithms for computing a partial SVD (see, e.g., Ref. [102] and the associated source code for an example of one such partial SVD implementation).

When $\mathcal{A}^*\mathcal{A}$ is an identity operator, the PowerFactorization procedure is equivalent to the power method for computing singular value decompositions, which is known to converge quite rapidly to the rank-L EY approximation of $\mathbf{C}$ [248, 294]. Moreover, if the rank of $\mathbf{C}$ is not larger than the value of L used in the PowerFactorization procedure, then the algorithm will converge exactly to $\mathbf{C}$ in a single iteration.

The convergence speed and global optimality characteristics of PowerFactorization in the presence of missing data and/or more general $\mathcal{A}$ operators have not been rigorously characterized theoretically. However, in these cases, the PowerFactorization algorithm monotonically decreases the cost function in Eq. (5.18), and thus the value of the cost function is guaranteed to converge since it is bounded below by 0. In general, the estimates of $\mathbf{U}^{(q)}$ and $\mathbf{V}^{(q)}$ themselves are not guaranteed to converge, particularly in the case when there is sustained rank-deficiency in the linear least-squares

problems. In particular, it should be noted that the factorization $\mathbf{C} = \mathbf{UV}$ is not unique, since $\mathbf{C}$ could equivalently be factored as $\mathbf{C} = (\mathbf{UT})(\mathbf{T}^{-1}\mathbf{V})$ for any $L \times L$ invertible matrix $\mathbf{T}$. As a result, Eq. (5.18) will never have a unique solution for $\mathbf{U}$ and $\mathbf{V}$, though there can be a unique solution for the product matrix $\mathbf{C} = \mathbf{UV}$. Despite this non-uniqueness for the factored problem, our empirical results show that convergence of $\mathbf{U}^{(q)}$ and $\mathbf{V}^{(q)}$ is not generally an issue when the number of measurements P is large enough.

Incremented Rank PowerFactorization

For matrix recovery and completion problems, we have obtained better results and faster convergence by using a variation of the general PowerFactorization procedure that uses an incremented-rank strategy. IRPF starts with $L = 1$, and gradually increments L until the desired rank constraint is achieved. In this case, we initialize the new components of $\mathbf{U}$ and $\mathbf{V}$ using a rank-1 PowerFactorization fit to the current residual. Since the lower-rank fits tend to have better conditioning as a result of having fewer degrees of freedom, this incremented/continuation approach helps to prevent the algorithm from getting stuck in local optima. The IRPF algorithm is given below:

- Set $\hat{\mathbf{C}} \in \mathbb{C}^{N \times M} = 0$.
- For $\ell = 1, \ldots, L$
  1. Use PowerFactorization to find a rank-1 fit to the current residual:

$$\{\hat{\mathbf{u}}, \hat{\mathbf{v}}\} = \arg \min_{\substack{\mathbf{u} \in \mathbb{C}^{N \times 1} \\ \mathbf{v} \in \mathbb{C}^{1 \times M}}} \left\| \mathcal{A}(\mathbf{uv}) + \mathcal{A}\left(\hat{\mathbf{C}}\right) - \mathbf{s} \right\|_{\ell_2}. \tag{5.25}$$

  2. If $\ell = 1$, set $\hat{\mathbf{U}} = \hat{\mathbf{u}}$ and $\hat{\mathbf{V}} = \hat{\mathbf{v}}$. Otherwise, set $\hat{\mathbf{U}} = \begin{bmatrix} \hat{\mathbf{U}} & \hat{\mathbf{u}} \end{bmatrix}$ and $\hat{\mathbf{V}} = \begin{bmatrix} \hat{\mathbf{V}} \\ \hat{\mathbf{v}} \end{bmatrix}$.

3. Use PowerFactorization to solve

$$\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\} = \arg \min_{\substack{\mathbf{U}\in\mathbb{C}^{N\times\ell} \\ \mathbf{V}\in\mathbb{C}^{\ell\times M}}} \left\|\mathcal{A}\left(\mathbf{UV}\right) - \mathbf{s}\right\|_{\ell_2}, \tag{5.26}$$

initializing the algorithm with the $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ matrices from step 2.

4. Set $\hat{\mathbf{C}} = \hat{\mathbf{U}}\hat{\mathbf{V}}$.

Since IRPF uses the PowerFactorization procedure directly, it inherits all of its theoretical convergence properties.

The main computation in the IRPF procedure is solving the linear least-squares problems in Eqs. (5.23) and (5.24). However, these linear least-squares problems are quite classical (cf. Eq. (2.36)), and a number of efficient algorithms exist to directly compute solutions [248]. However, we do note that in some cases, the matrices $\mathbf{A}_{\mathbf{U}}$ and $\mathbf{A}_{\mathbf{V}}$ will not have full column rank, meaning that the least-squares solutions to Eqs. (5.23) and (5.24) can be non-unique; for example, if $\mathbf{V}$ is initialized to be identically zero, then $\mathbf{A}_{\mathbf{V}}$ is also identically zero. In these situations, it is beneficial to choose a least-squares solution that is distinct from the minimum-norm least-squares solution; in our implementation, we randomly choose a vector from the linear variety of least squares solutions.

Significant computational gains can be achieved when we apply IRPF to the MR-specific problems in Eqs. (5.10) and (5.16). In the context of Eq. (5.10), the least-squares optimality condition for $\mathbf{V}^{(q+1)}$ in Eq. (5.24) of the PowerFactorization procedure is that

$$\left[\mathbf{U}^{(q+1)}\right]^H \mathbf{F}^H \Xi^* \left(\Xi\left(\mathbf{F}\mathbf{U}^{(q+1)}\mathbf{V}^{(q+1)}\right)\right) - \left[\mathbf{U}^{(q+1)}\right]^H \mathbf{F}^H \Xi^*\left(\mathbf{s}\right) = 0. \tag{5.27}$$

Note that this expression has similar form to Eq. (5.7). As a result, similar to the simplification from Eq. (5.7) to Eq. (5.8), Eq. (5.27) is separable in the columns of $\mathbf{V}^{(q+1)}$. This separability implies that the system matrix $\mathbf{A}^H_{\mathbf{U}^{(q+1)}}\mathbf{A}_{\mathbf{U}^{(q+1)}}$ that appears in the normal equations for Eq. (5.24) can be permuted into decoupled block diagonal form, with $L \times L$ blocks. This leads to efficient

computation, either by direct inversion of each $L \times L$ block separately, or by using the iterative conjugate gradient algorithm to solve the sparse system. The iterative conjugate gradient (CG) algorithm for matrix inversion would find optimal solutions (assuming infinite numerical precision) after L iterations [305]. Similar decoupling exists for Eq. (5.23) of the PowerFactorization procedure when using the problem formulation in Eq. (5.16), or equivalently, the problem formulation in Eq. (5.10) when $\mathbf{F}$ is an identity matrix. Even when $\mathbf{F}$ in Eq. (5.10) is not an identity matrix, matrix-vector multiplications with $\mathbf{F}$ can often still be computed quickly (e.g., using the FFT for Fourier-domain observations), meaning that solving Eq. (5.23) using iterative methods like CG can still be very efficient.

Modified Incremented Rank PowerFactorization for Matrix Completion in the presence of Navigator Information

The previously described IRPF algorithm works well for many problems of interest. However, we have observed that a modified IRPF procedure can help to avoid local minima when solving matrix completion problems with specialized sampling schemes that include navigator information. As described previously, the fully sampled navigator rows of $\mathbf{\Phi}$ can provide strong information about the dominant temporal subspace for the fully-sampled $\mathbf{\Phi}$ matrix. However, it is often the case that the acquired navigator data samples constitute only a small percentage of the total P measured data samples. Since standard IRPF seeks to find a low-rank matrix that is as data-consistent as possible, it will frequently happen that the IRPF procedure neglects data consistency for the navigator samples in favor of improved data consistency with respect to the full dataset. Given sufficient measured data, this issue is not problematic, and IRPF still performs well; however, when P is very small, the IRPF procedure can more easily become trapped in local critical points of the cost function. Our proposed modification of IRPF leverages the strong subspace information that can be extracted from the navigator data to help IRPF avoid local critical points and obtain reconstructions that more closely match with the acquired data. Our proposed modification of IRPF is given below, formulated in the context of Eq. (5.16). It will be assumed that we have a total of J different

fully-sampled navigator rows.

- Set $\hat{\Phi} \in \mathbb{C}^{N \times M} = 0$.
- For $\ell = 1, \ldots, L$

  1. Use PowerFactorization to find a rank-1 fit to the current residual:

$$\{\hat{\mathbf{u}}, \hat{\mathbf{v}}\} = \arg \min_{\substack{\mathbf{u} \in \mathbb{C}^{N \times 1} \\ \mathbf{v} \in \mathbb{C}^{1 \times M}}} \left\| \Xi(\mathbf{uv}) + \Xi\left(\hat{\Phi}\right) - \mathbf{s} \right\|_{\ell_2}. \tag{5.28}$$

  2. If $\ell = 1$, set $\hat{\mathbf{U}} = \hat{\mathbf{u}}$ and $\hat{\mathbf{V}} = \hat{\mathbf{v}}$. Otherwise, set $\hat{\mathbf{U}} = \begin{bmatrix} \hat{\mathbf{U}} & \hat{\mathbf{u}} \end{bmatrix}$ and $\hat{\mathbf{V}} = \begin{bmatrix} \hat{\mathbf{V}} \\ \hat{\mathbf{v}} \end{bmatrix}$.
  3. Find the data-consistent full-rank matrix $\hat{\mathbf{B}}$ that is as close as possible (in the Frobenius norm) to the current estimate of the low-rank matrix:

$$\begin{aligned} \hat{\mathbf{B}} &= \arg \min_{\substack{\mathbf{B} \in \mathbb{C}^{N \times M} \\ \Xi(\mathbf{B}) = \mathbf{s}}} \left\| \mathbf{B} - \hat{\mathbf{U}}\hat{\mathbf{V}} \right\|_F \\ &= \Xi^*(\mathbf{s}) + \hat{\mathbf{U}}\hat{\mathbf{V}} - \Xi^*\left(\Xi\left(\hat{\mathbf{U}}\hat{\mathbf{V}}\right)\right). \end{aligned} \tag{5.29}$$

  4. Compute the rank-L EY approximation of the $(N + J) \times M$ matrix $\begin{bmatrix} \hat{\mathbf{B}} \\ \alpha \mathbf{\Phi}_n \end{bmatrix}$, where α is a positive scalar that is large enough to ensure that $\alpha \left\| \Phi_n \right\|_F >> \left\| \hat{\mathbf{B}} \right\|_F$. Let the EY approximation be written in terms of its SVD $\mathbf{P}\Sigma\mathbf{Q}^H$, where $\mathbf{P} \in \mathbb{C}^{(N+J) \times L}$ and $\mathbf{Q} \in \mathbb{C}^{M \times L}$ have orthonormal columns and $\Sigma \in \mathbb{R}^{L \times L}$ is a positive-definite diagonal matrix.
  5. Set $\hat{\mathbf{V}} = \mathbf{Q}^H$, and set $\hat{\mathbf{U}} = \mathbf{P}_J \Sigma$, where $\mathbf{P}_J$ is the $N \times L$ submatrix of $\mathbf{P}$ obtained by extracting its first N rows.

6. Use PowerFactorization to solve

$$\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\} = \arg \min_{\substack{\mathbf{U}\in\mathbb{C}^{N\times\ell} \\ \mathbf{V}\in\mathbb{C}^{\ell\times M}}} \left\|\Xi\left(\mathbf{UV}\right) - \mathbf{s}\right\|_{\ell_2}, \tag{5.30}$$

initializing the algorithm with the $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ matrices from step 5.

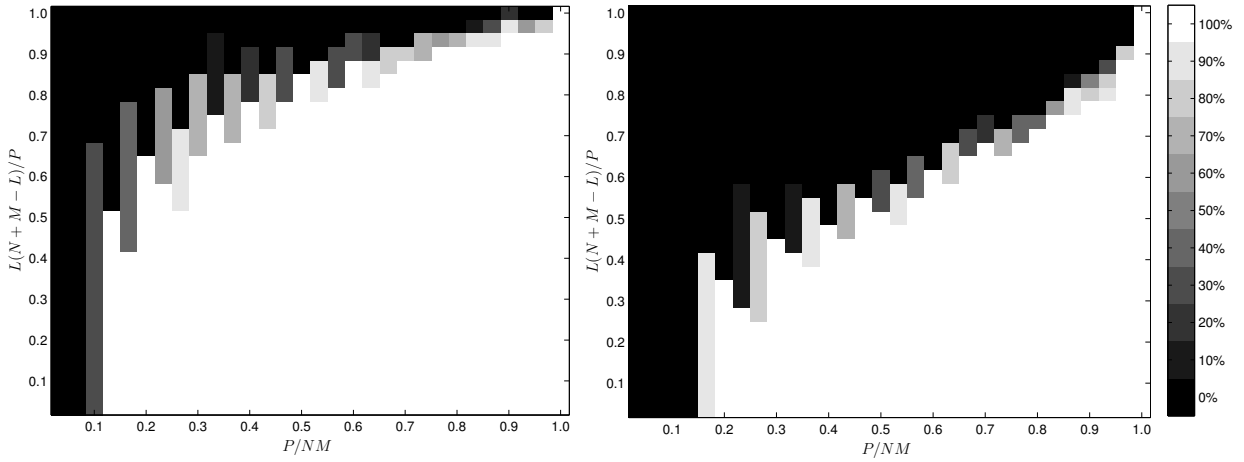
7. Set $\hat{\mathbf{\Phi}} = \hat{\mathbf{U}}\hat{\mathbf{V}}$.

The main difference between this algorithm and the previously described IRPF algorithm is the inclusion of steps 3-5. These steps are designed to ensure that the rank-L initialization for the PowerFactorization algorithm in step 6 is relatively data consistent with the navigator rows in Φ_n, while still leveraging estimation results from the previous iterations. Data consistency with the navigator rows will improve with increasing α; however, it is important to note that setting α extremely large could, in practice, lead to significant numerical errors in the computation of the SVD.

5.3.2 Comparison to Nuclear Norm Minimization

While IRPF has empirical speed advantages relative to NNM [268], we will focus here on the differences in empirical matrix recovery performance. The matrix recovery capabilities of IRPF were compared to those of NNM using two sets of experiments. In the first set of recovery experiments, the a_{mnp} coefficients defining linear operators $\mathcal{A}$ were generated at random from the Gaussian distribution, and random low-rank 30×30 matrices $\mathbf{C}$ were generated as $\mathbf{C} = \mathbf{M}_L\mathbf{M}_r$, where the entries of $\mathbf{M}_L \in \mathbb{C}^{30\times L}$ and $\mathbf{M}_R \in \mathbb{C}^{L\times 30}$ were also generated from the Gaussian distributed. Test cases were generated for many different combinations of the number of measurements P and $\mathrm{rank}\left(\mathbf{C}\right)$ (assumed to be known), and 10 realizations were computed for each (P, L) pair. Theoretical properties of NNM for Gaussian observations and matrix completion are discussed in [113, 114, 118, 391, 452, 535].

The second set of recovery experiments was identical to the first set, except that the linear



(a) Incremented-Rank PowerFactorization

(b) Nuclear norm minimization

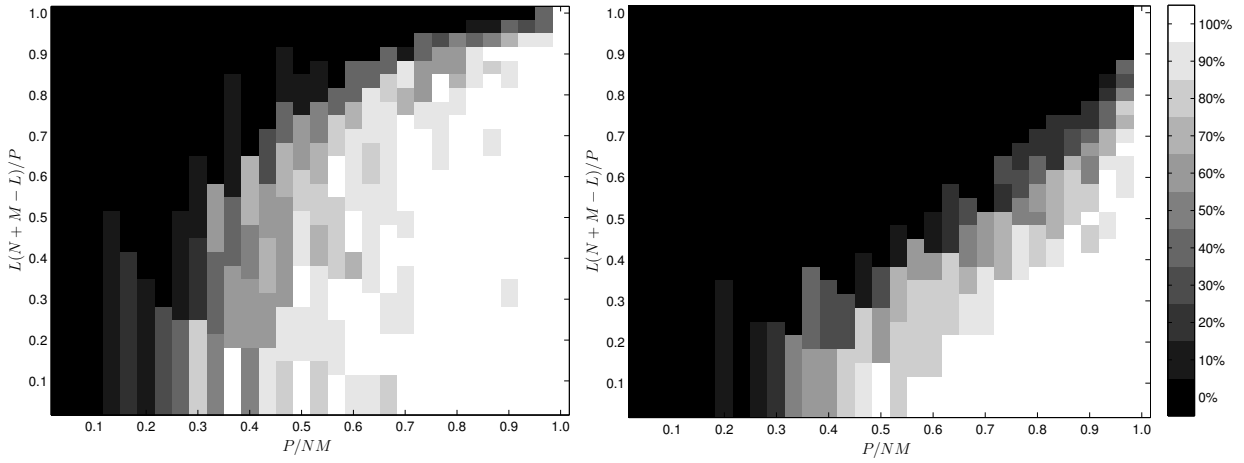
Figure 5.4: Matrix recovery results for (a) IRPF and (b) NNM with Gaussian observations. The color of each cell corresponds to the empirical recovery rate, with white denoting perfect recovery and black denoting failure in all 10 experiments. The vertical axis is $L(N+M-L)/P$, which is the ratio of the number of degrees of freedom for an $N \times M$ rank-L matrix to the number of measurements P.

operators $\mathcal{A}$ were chosen to correspond to the matrix completion problem, measuring P entries (selected uniformly at random) from $\mathbf{C}$.

Figure 5.4 shows the results of the experiment with Gaussian observations. While NNM is able to successfully recover a large fraction of the low-rank matrices, IRPF is able to recover a significant additional fraction that NNM is unable to recover. As in NNM [117, 535, 536], there appears to be phase-transition behavior for IRPF with Gaussian measurements, though the boundary of this phase transition appears in a different location.

Figure 5.5 shows the results of the matrix completion experiment. Again, there is a significant fraction of matrices that is successfully recovered by IRPF, but that is not recovered by NNM. The superiority of IRPF over NNM for completion problems is consistent with the results obtained by Dai and Milenkovic [153], in which IRPF also demonstrated advantages over several other recently-proposed matrix completion algorithms.

It should be noted for our completion experiment that we observed a small number of cases where NNM succeeded while IRPF failed, due to IRPF converging to a local stationary point of the cost function. These few cases are easily identified without knowing the true $\mathbf{C}$, due to a large



(a) Incremented-Rank PowerFactorization

(b) Nuclear norm minimization

Figure 5.5: Matrix completion results for (a) IRPF and (b) NNM. The color of each cell corresponds to the empirical recovery rate, as in Fig. 5.4.

residual data error. For moderate-size problems, this can be efficiently overcome by performing IRPF several times with randomly selected initializations. These results were obtained from relatively small matrices. Preliminary experiments indicate that an advantage of IRPF over NNM is maintained for larger matrices, although the asymptotic behavior is unknown.

While IRPF is more successful at matrix recovery and can converge faster than classical PowerFactorization, PowerFactorization alone can also perform surprisingly well given sufficient measurements and appropriately-chosen L. To illustrate this, we again generated Gaussian observation operators $\mathcal{A}$ for various P values, and random 40×40 matrices $\mathbf{C}$ of rank 8. We tested PF with this dataset, allowing L to range from 1 up to 20. The results of this experiment, averaged over 10 realizations, are shown in Fig. 5.6.

5.4 Application Examples

The following subsections present practical examples of applying the PS model and the IRPF algorithm to different spatiotemporal imaging scenarios.

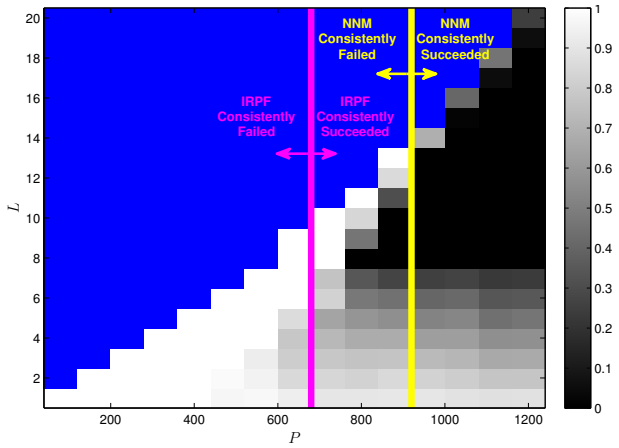
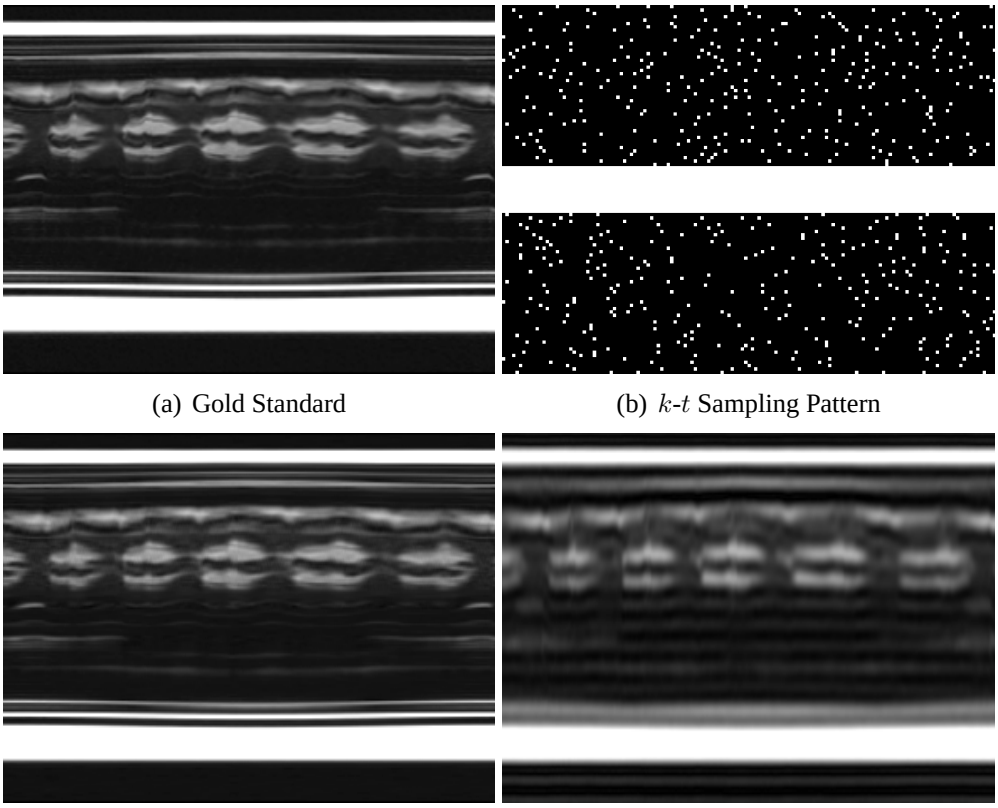


Figure 5.6: Mean relative reconstruction error using PowerFactorization for various L values. The true rank is 8. Blue indicates untested cases (the number of degrees-of-freedom exceeded P). The success/failure regimes for NNM and IRPF are indicated with yellow and pink lines, respectively.

5.4.1 Real-Time Cardiac Imaging Simulations

In this cardiac imaging example, simulations were performed using a version of the cardiac cine dataset shown in Fig. 5.1, which was spatiotemporally warped to emulate a 5 second free-breathing experiment in the presence of cardiac arrhythmia [675] with $M = 256$ different time points. Cartesian phase encoding was simulated, with full sampling along the readout dimension. Only 1/6th of the k-t locations were sampled; we acquired 27 fully-sampled navigator rows of $\mathbf{\Phi}$ at the center of *k*-space as training data, while the remaining rows were sampled uniformly at random. The use of Cartesian sampling with a square, invertible $\mathbf{F}$ matrix allowed us to use the matrix completion formulation of Eq. (5.16). For improved computational efficiency, Fourier inversion was first performed along the readout dimension, and the modified IRPF algorithm for matrix completion with navigator data was applied to reconstruct the undersampled phase-encoding dimension in a spatially-decoupled way. This led to a set of 256 different low-rank matrix reconstruction problems with $N = 200$ and $M = 256$. The reconstruction used $L = 18$, leading to a signal model with 10% of the degrees of freedom compared to a full-rank image model. The results of this simulation are shown in Fig. 5.7, illustrating the potential of low-rank constraints for accelerating this kind of imaging experiment. Other cardiac imaging results using the general matrix-recovery framework can be found in [675].



(a) Gold Standard (b) k-t Sampling Pattern

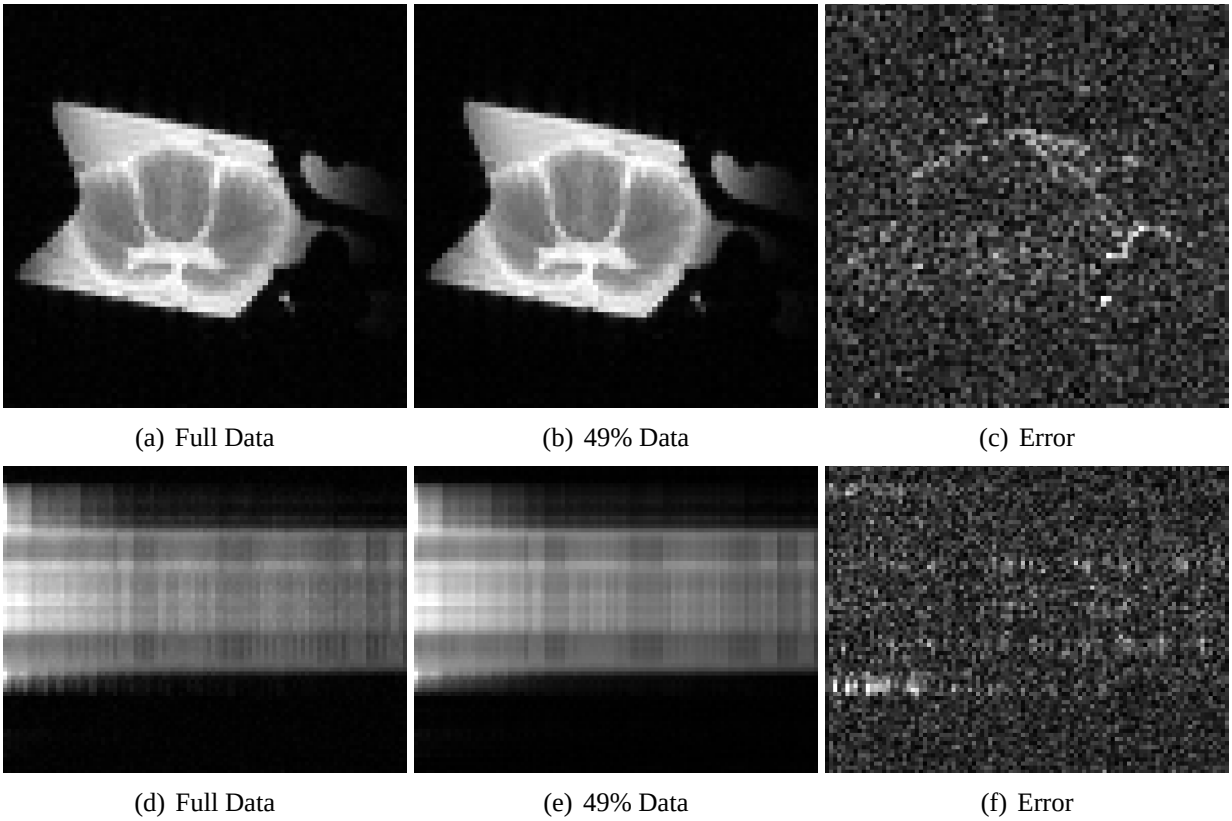
(c) Low-Rank Result (d) Low-Resolution Result

Figure 5.7: Low-rank matrix completion results from 1/6th of the entries of a simulated dynamic cardiac $\mathbf{\Phi}$ matrix. The horizontal axis for all images corresponds to t, while the vertical axis corresponds to the phase encoding dimension $\mathbf{x}$. (a) Gold standard image. (b) Sampling pattern in $\mathbf{k}$-t space. (c) Low-rank reconstruction using L=18. Low-rank modeling enables high-quality reconstruction from limited data. (d) Fourier reconstruction using a densely-sampled low-resolution acquisition with the same number P of data samples, illustrating the limitations of a more-conventional approach to data acquisition and reconstruction.

5.4.2 Diffusion MRI Experiments

In this diffusion MRI example, fully-sampled real data was acquired from a fixed *ex vivo* human spinal cord sample using a 4.7 T Varian scanner (data provided by Y. Wang and S.-K. Song at the Washington University in St. Louis). Data was acquired on a 96×96 Cartesian grid at the Nyquist rate for a 24 mm $\times$ 24 mm FOV. Acquisition used a standard spin-echo pulse sequence that acquires a single phase-encoding line per excitation, with an echo time of 35 ms and a repetition time of 2 s. The data was diffusion weighted to enable fitting with the Diffusion Basis Spectrum Imaging model [633]. In particular, a total of 99 different diffusion-weighted images were acquired, with b-values ranging between 0 and 3,200 s/mm^2 and diffusion times of $\delta = 8$ ms and $\Delta = 20$ ms. Total data acquisition time was approximately 5.3 hours.

Diffusion MRI experiments acquire a series of images, where each image is exposed to different diffusion-encoding gradients. Diffusion imaging can be treated as spatiotemporal imaging if the diffusion-encoding dimension is treated as a temporal dimension. The proposed low-rank matrix formulation was applied to enable high quality reconstruction from undersampled k-space data. In particular, we subsampled the dataset such that only 47 phase encoding lines were acquired per diffusion-encoded image, representing approximately 49% of the fully-sampled data. For a real experiment, this level of undersampling would reduce the total data acquisition time down to only 2.6 hours, representing a significant improvement in imaging efficiency. Of these 47 phase encoding lines, 32 lines were consistently acquired for each image at the center of *k*-space, while the remaining lines were randomly distributed to the remaining phase encoding locations. Results obtained by applying modified IRPF with $L = 7$ to solve Eq. (5.16) are shown in Fig. 5.8. As can be seen, the proposed low-rank matrix recovery framework enabled high quality reconstruction from significantly undersampled data, and could enable this type of experiment to be performed significantly faster.



(a) Full Data (b) 49% Data (c) Error

(d) Full Data (e) 49% Data (f) Error

Figure 5.8: Low-rank matrix completion results from 48% of the entries of a real diffusion-weighted MRI dataset. (a) One diffusion-weighted image from the fully-sampled dataset. (b) The same diffusion-weighted image after reconstruction using IRPF. This particular image was chosen for display because it was the image with the largest reconstruction error after reconstruction using the proposed method. (c) Error image (scaled by a factor of 15 for improved visualization) for the reconstruction shown in (b). (d) Spatiotemporal profile of the fully-sampled diffusion-weighted dataset. The horizontal axis corresponds to different diffusion-encoding parameters, while the vertical axis corresponds to the phase-encoding dimension. (e) Spatiotemporal profile of the IRPF-based reconstruction. (c) Error image (scaled by a factor of 15 for improved visualization) for the reconstruction shown in (e). These results indicate that the rank-7 PS model is able to accurately represent the diffusion contrast of the signal within the spine parenchyma, and that IRPF can be used to significantly accelerate data acquisition.

5.4.3 DCE Breast Imaging Simulations

In this DCE breast imaging example, an exogenous contrast agent is injected into the bloodstream, and the contrast kinetics are used to localize a tumor and provide information regarding its physiological and morphological properties. Fourier data was simulated from a series of magnitude images corresponding to 18 different time-points from a real DCE experiment. These 18 frames were subsequently interpolated onto a set of $M = 52$ equally-spaced time points. The simulated data acquisition assumed a 256×256 Cartesian *k*-space sampling grid (i.e., $N = 256^2$). The measurement operator Ξ sampled 25% of the entries in $\mathbf{\Phi}$ uniformly at random. This was augmented by an additional set of samples chosen to ensure that each row and each column of $\mathbf{C}$ was sampled at least 10 times,[11] resulting in a total of $\sim 27\%$ of the entries being sampled.

As in the previous examples, matrix recovery was performed using Eq. (5.16). Two temporal frames from a standard IRPF reconstruction with $L = 5$ are shown in Fig. 5.9. These results illustrate that high-quality PS-based reconstruction is possible from highly-undersampled data without the use of specialized sampling.

For comparison, reconstruction of this same dataset was also performed using a more traditional compressed sensing approach exploiting the known sparsity of

$$\hat{\rho}\left(\mathbf{x}, f\right) = \int \rho\left(\mathbf{x}, t\right) \exp\left(-\imath 2\pi f t\right) dt \tag{5.31}$$

in dynamic MR applications [228]. Figure 5.10 shows reconstructions performed by minimizing the ℓ_1 norm [117] of $\hat{\rho}\left(\mathbf{x}, f\right)$ subject to data-fidelity constraints, assuming a finite-dimensional spatiotemporal image model with Dirac delta voxel functions spaced uniformly on a Cartesian grid in x-f space. While both the IRPF and ℓ_1 reconstructions accurately recover the structure of the image, the ℓ_1 reconstruction shows more significant spatially-localized errors.

[11] A necessary condition for matrix completion to be well-posed in this context is that each row and each column is sampled at least L times. Sampling more than this will improve the conditioning of the problem.

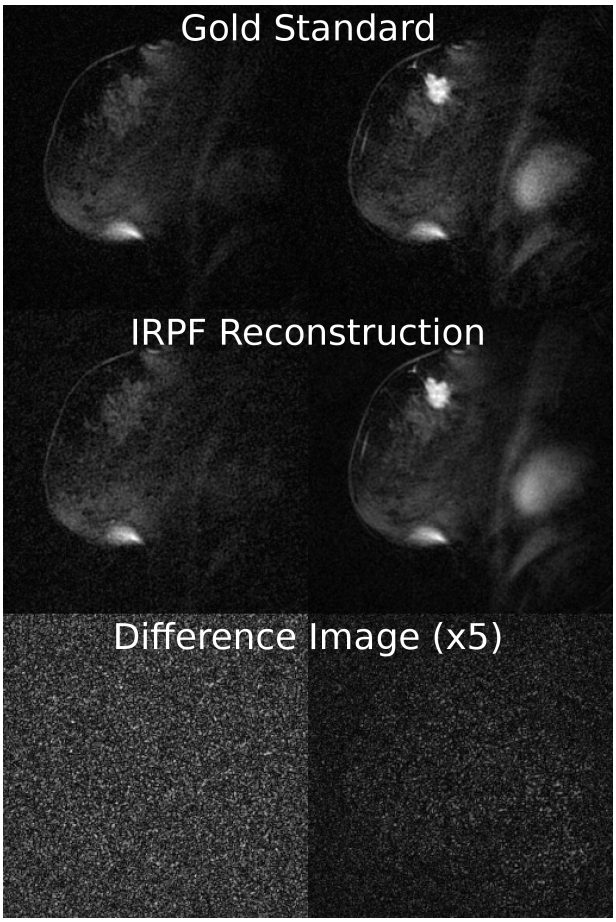


Figure 5.9: Simulated PS-based spatiotemporal reconstruction (L=5) using IRPF with $\sim 27\%$ of the full data. The left and right sides of the figure show different temporal frames.

5.5 Discussion

5.5.1 Specialized Sampling Versus Random Sampling

The DCE breast imaging example illustrated the potential of using IRPF for randomly-sampled spatiotemporal images with PS structure, while the cardiac and diffusion MRI examples used the specialized sampling scheme where a number of fully-sampled rows of the $\mathbf{\Phi}$ matrix were acquired. These results show that the proposed low-rank matrix recovery framework is very flexible, and can be used successfully with a range of different sampling patterns. However, it is also of interest to compare the random-sampling strategy with the special navigated sampling strategy. Simulation-based comparisons between these two different sampling strategies are shown in

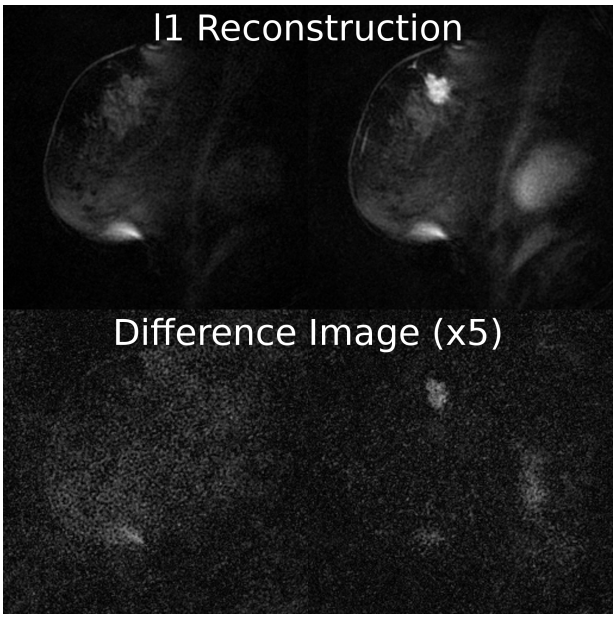


Figure 5.10: Compressed sensing reconstruction from $\sim 27\%$ of the full data. In contrast to IRPF reconstruction, the reconstruction errors have more apparent anatomically-correlated structure.

Figs. 5.11 and 5.12. These results illustrate that it can be beneficial to acquire several fully-sampled rows of $\mathbf{\Phi}$. Fully sampling low-frequency $\mathbf{k}$-space can be especially beneficial for realistic data, due to high signal energy concentration in this area, though it should also be noted that high-frequency *k*-space data can have higher temporal variability [59]. However, acquiring too many fully-sampled rows can degrade reconstruction performance, because there are fewer samples to distribute outside $\mathbf{\Phi}_n$. These figures also illustrate that the proposed IRPF method can be significantly more flexible than the basic PS method, enabling high-quality reconstructions across a range of different acquisition strategies.

5.5.2 Selection of L

One practical issue for Eq. (5.17) is the selection of L. If L is chosen too small, the PS model is not capable of accurately representing the true structure of the original spatiotemporal image, leading to significant bias in the final reconstructed image. On the other hand, if L is large, the PS model will have a large number of degrees of freedom. As a result, while large-L models can

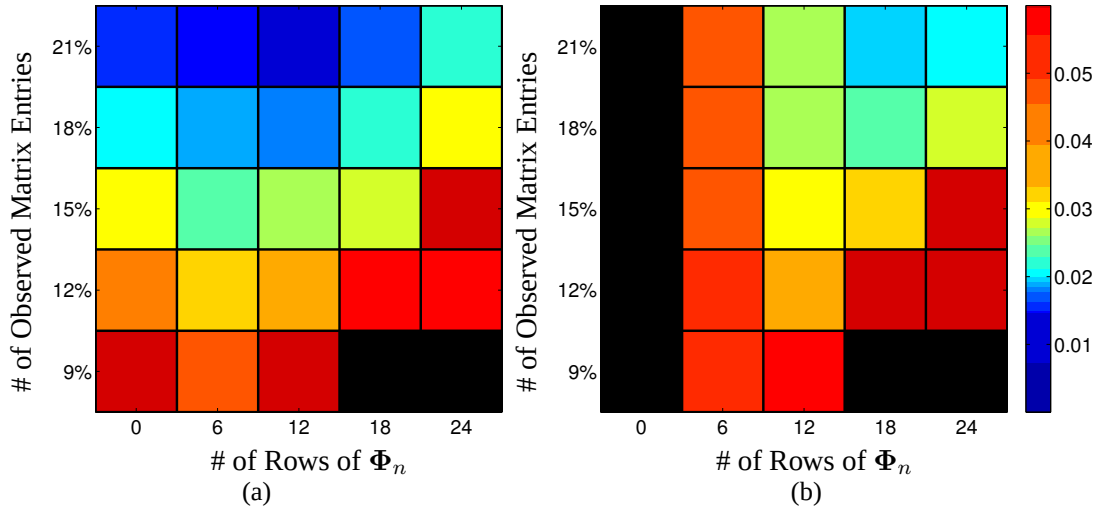


Figure 5.11: Relative reconstruction errors for the cardiac dataset described in Sec. 5.4.1 as a function of the number of observations P and the number of fully-sampled navigator rows. All navigator rows were located in the low-frequency region of *k*-space, while the remaining samples were randomly distributed. Relative reconstruction errors were measured in the Frobenius norm. For each case, L was set to the value that yielded the smallest reconstruction error. Black squares correspond to infeasible acquisition parameter combinations that were not tested. (a) Results using the modified IRPF algorithm. Due to its flexibility, IRPF is able to generate high-quality reconstructions across a range of different sampling strategies. However, acquiring a small number of fully-sampled navigator rows is observed to improve reconstruction performance, though it can be detrimental to acquire too many navigator rows. (b) Results using the basic PS procedure. In contrast to the IRPF method, the basic PS procedure is considerably less flexible. In particular, it is not possible to use the basic PS procedure when no navigator information is available, and the maximum rank L that can be used for reconstruction is limited by the total number of acquired navigator rows.

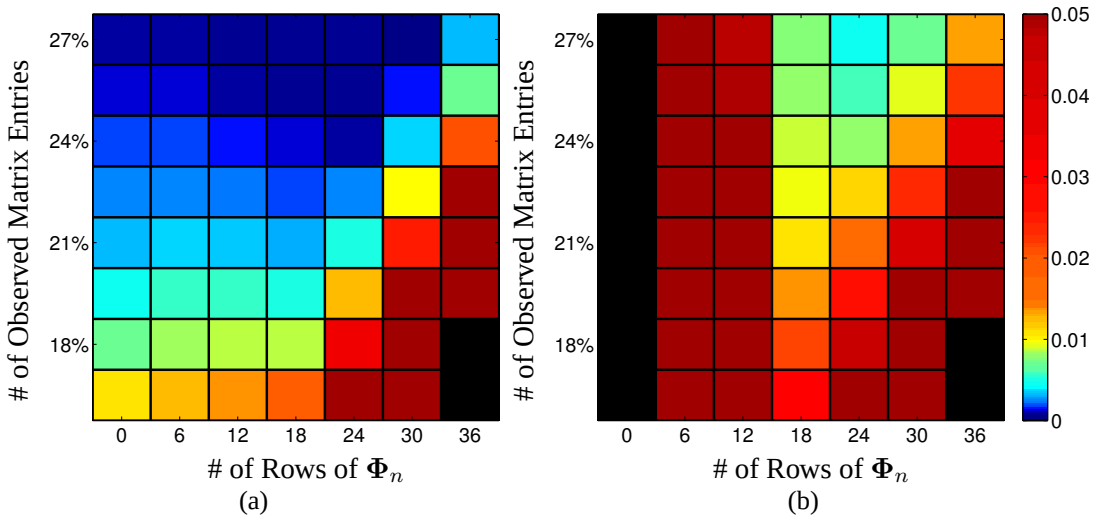


Figure 5.12: Relative reconstruction errors for random approximately low-rank matrices as a function of the number of observations P and the number of fully-sampled navigator rows. Except for the choice of matrices, simulation parameters were the same as those used to generate Fig. 5.11. Approximately low-rank matrices were generated using the following procedure: first, a random 200×256 was generated from the Gaussian distribution. Next, SVD was performed on this matrix, and its singular values were replaced with an exponentially-decaying set of values. In particular, we set $\sigma_k = \exp(-0.35k)$ for $k = 1, \ldots, 200$. This exponentially-decaying singular-value spectrum ensures that each of the random matrices can be well-approximated as a low-rank matrix. (a) Results using the modified IRPF algorithm. As with the cardiac data, the IRPF algorithm is flexible enough to accommodate a wide range of different sampling strategies. In some regimes, acquiring a number of fully-sampled navigators is beneficial. However, unlike the case shown in Fig. 5.11, there are also cases where pure random sampling outperforms the specialized navigator-based sampling. One explanation for this is that, unlike the case for real data, the energy-distribution for randomly-generated low-rank matrices is not concentrated near the center of *k*-space. (b) Results using the basic PS procedure. As before, these results indicate that the basic PS procedure is considerably less flexible than the IRPF-based low-rank matrix recovery approach.

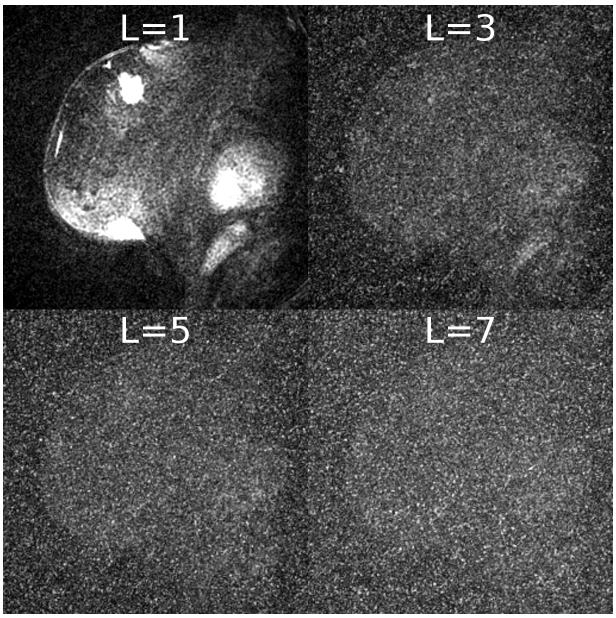


Figure 5.13: Mean-squared reconstruction error for different model orders. As L increases, the signal model has better capability to represent the true signal, leading to fewer signal features being apparent in the error map. However, this additional flexibility comes at the expense of conditioning, which is reflected by the increasing "noise" component of the error maps with increasing L.

always fit the measured data better than smaller-L models, the model-fitting problem with large L can be sensitive to noise and be prone to overfitting (i.e., interpreting noise perturbations as actual signal structure). In general model selection problems, this situation is frequently called the "bias/variance dilemma." This issue is illustrated in Fig. 5.13 for the DCE breast imaging dataset. A good choice of L requires a careful balance between the expressive power of the signal model and the robustness of the fitting procedure, and our choice of L in the previous examples was based on a qualitative assessment of visual reconstruction quality.

Estimation of the rank of a noisy fully-sampled matrix has been previously investigated by a number of authors (e.g., Refs. [106,150,218,288,315,368,463,638,657] and their references), who frequently made use of techniques from information-theoretic model selection [97, 259, 381, 383, 462,591]. The choice of L for PS model-fitting with limited data can also be considered within this information-theoretic framework. The literature on information-theoretic model selection provides a variety of different quantitative metrics for selecting an appropriate model to explain a given

dataset. In general, these methods favor models that fit the data accurately, while simultaneously penalizing models with a large number of degrees of freedom. Well-known model selection criteria include the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Final Prediction Error (FPE), and the Minimum Description Length (MDL). The assumptions and objectives leading to the derivation of these different criteria are outside the scope of this thesis, and the interested reader is referred to Refs. [97, 259, 381, 383, 462, 591] for more detail. While these criteria are all different, it is known that, as the number of measurements P grows to infinity, AIC and FPE are asymptotically equivalent [474] and that BIC and MDL are asymptotically equivalent in many cases of interest [259,462]. While various authors prefer different model selection criteria, it should be noted that the best-performing model selection criteria can vary from application to application [591].

We performed simulations to evaluate the performance of AIC, BIC, and FPE for model selection in matrix completion problems. Assuming observations according to Eq. (5.5) under white complex Gaussian noise with unknown variance, the $\mathrm{AIC_c}$ (the AIC with a correction to handle small sample sizes [97, 381]) for a rank-L image model is given by

$$\mathrm{AIC_c}\left(L\right) = 2P \ln\left(\left\|\Xi\left(\mathbf{F}\hat{\mathbf{C}}\right) - \mathbf{s}\right\|_{\ell_2}^2\right) + 2K\left(\frac{2P}{2P-K-1}\right), \tag{5.32}$$

where $K = 2L(N+M-L)+1$, and $\hat{\mathbf{C}}$ is the rank-L optimal solution to Eq. (5.10). Under the same assumptions and variable definitions, the BIC is given by

$$\mathrm{BIC}\left(L\right) = 2P \ln\left(\left\|\Xi\left(\mathbf{F}\hat{\mathbf{C}}\right) - \mathbf{s}\right\|_{\ell_2}^2\right) + K\ln\left(2P\right), \tag{5.33}$$

and the FPE is given by

$$\mathrm{FPE}\left(L\right) = \left\|\Xi\left(\mathbf{F}\hat{\mathbf{C}}\right) - \mathbf{s}\right\|_{\ell_2}^2 \left(\frac{1+K/(2P)}{1-K/(2P)}\right). \tag{5.34}$$

In all cases, the optimal choice of L will minimize the respective model selection criterion.

MDL was not evaluated in the simulations, because the computation of MDL requires the selection of a scheme to measure the joint complexity of the parametric model and the measured data. While there exist several different approaches to describing the stochastic complexity of a model [462], most of these would be very difficult to compute for the low-rank matrix recovery problem.[12] The literature we have found on the selection of matrix rank using MDL [218, 463, 638, 657] generally chooses a simple measure of complexity that leads to an equivalence between MDL and BIC.

Our simulations were performed using randomly-generated full-rank 128×128 complex Gaussian matrices, whose singular values were replaced with deterministic singular value distributions. In particular, the singular value distributions were chosen such that $\sigma_k = k^{-\beta}$ for $k = 1, \ldots, 128$. Matrix completion experiments were performed for matrices designed with $\beta = 1$, 2, 3, and 4. Note that higher values of β are associated with faster singular value decays, and therefore would be associated with better approximation using a low-rank matrix model. The set of observed matrix entries was chosen uniformly at random, and the number of observations P was varied between 20% and 80% of the total number of matrix entries. Simulations were performed 10 times (with different random matrices and sampling patterns) for each combination of parameters, and cumulative results are shown in Figs. 5.14-5.17.

Since the matrices used in this experiment were only approximately low rank, we would ideally want the different model selection criteria to choose a rank for which the relative reconstruction error for IRPF was minimized. In practice, this was never achieved by any of the different model selection criteria. In addition, it was frequently observed that there were large differences between the ranks selected by the different criteria, and that the reconstruction quality using the models selected by the different criteria depended heavily on β. In particular, model selection using FPE tended to result in more accurate matrix reconstruction when β was large, though the FPE results

[12] For example, one common approach to describing the complexity of the model requires the computation of the Fisher information matrix [354, 514] for the unknown parameters [259, 462]. Computation of the standard Fisher information matrix requires having an explicit parametric signal model. However, as far as we are aware, there is no model with exactly $2L(N + M - L)$ explicit real parameters that can represent an arbitrary rank-L complex matrix, meaning that the computation of the Fisher information matrix is seemingly nontrivial.

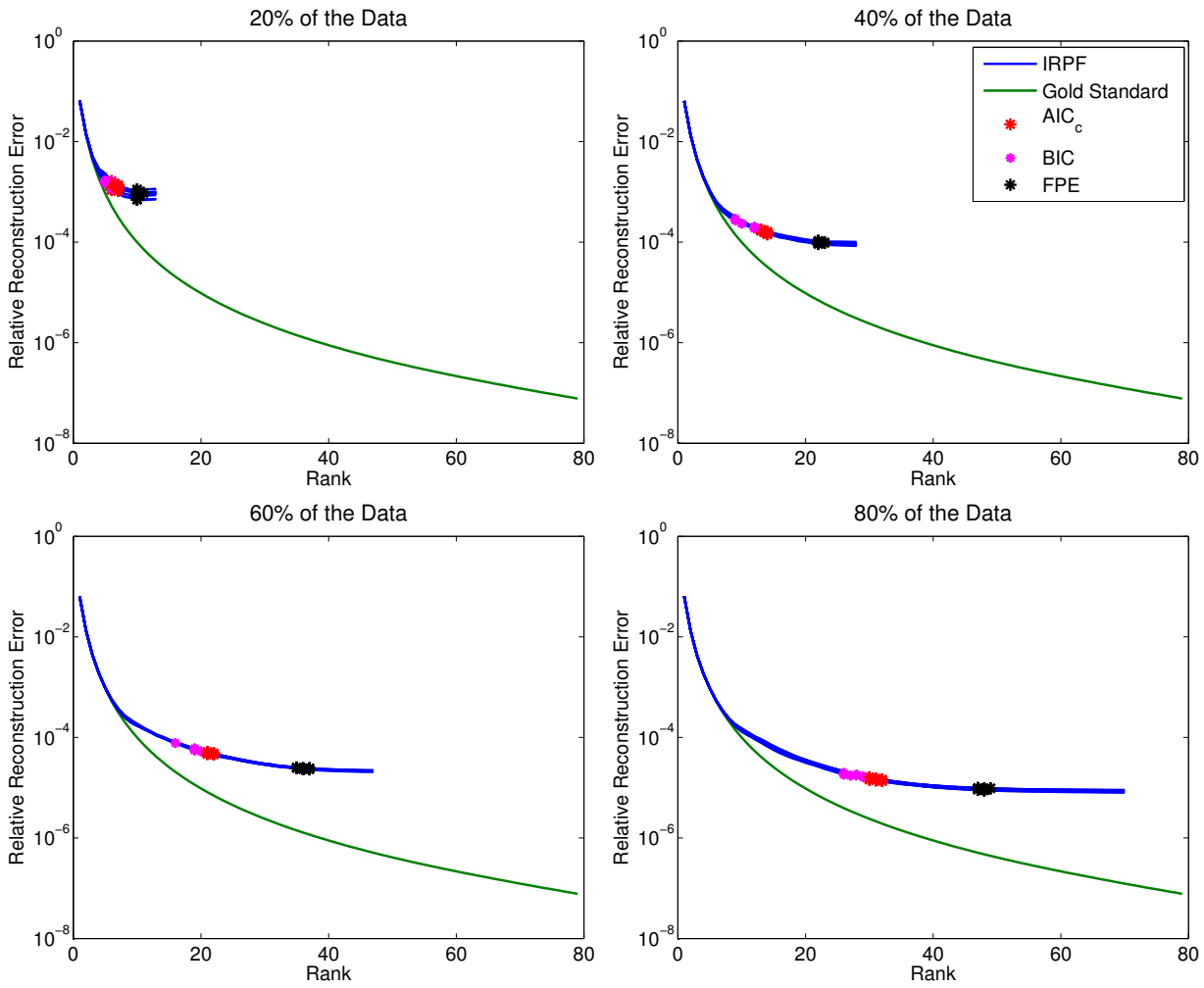


Figure 5.14: Comparison of information-theoretic model-selection criteria for matrix completion problems with different amounts of acquired data. The value of β specifying the singular value distribution was 4. The plots show the reconstruction error using IRPF as a function of rank for 10 different trials, in addition to the reconstruction error as a function of rank for the EY approximation of the fully-sampled original "gold-standard" matrix. Also indicated are the ranks selected by the AIC_c, BIC, and FPE model-selection criteria.

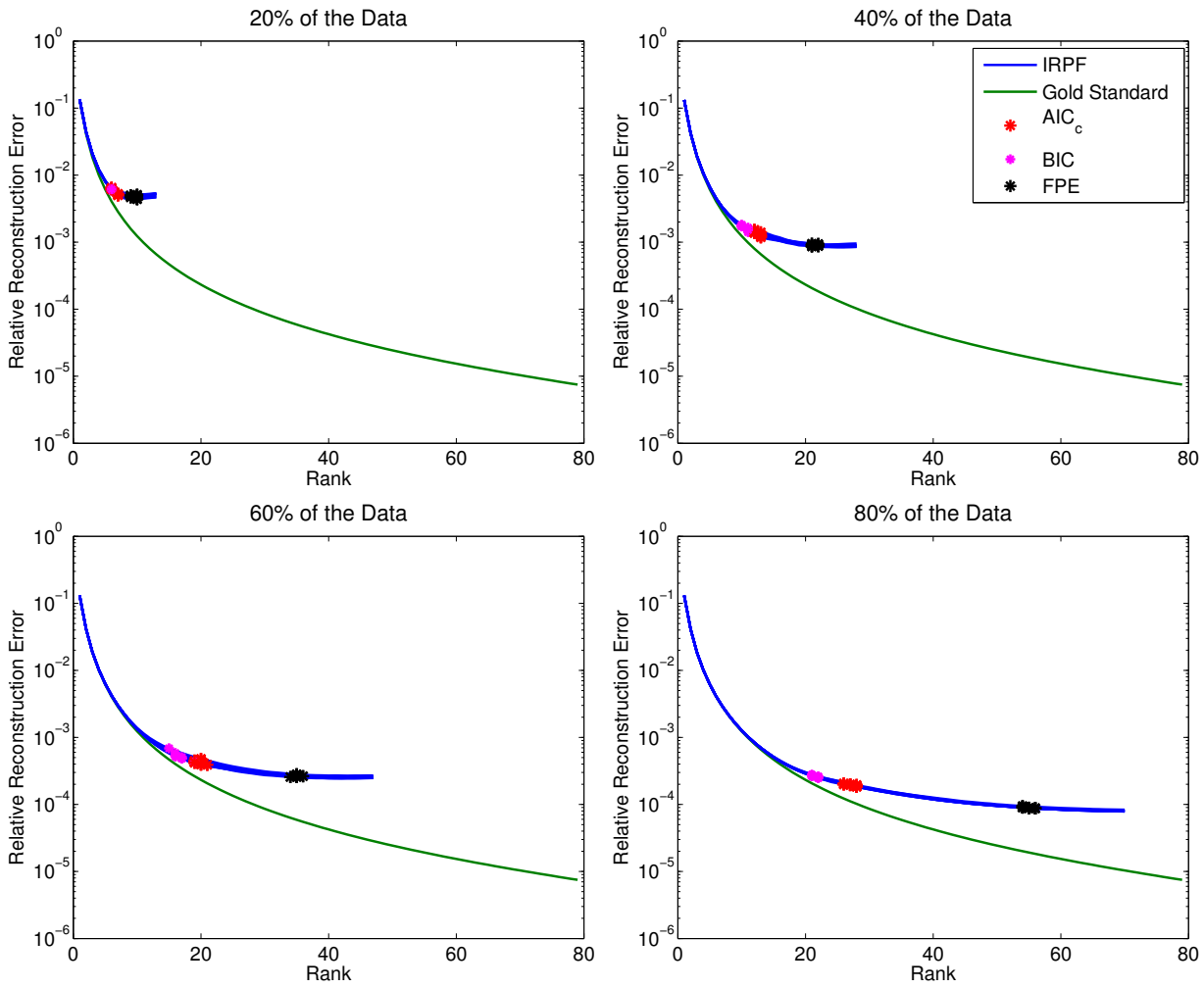


Figure 5.15: Comparison of information-theoretic model-selection criteria for matrix completion problems with different amounts of acquired data. The value of β specifying the singular value distribution was 3. The plots show the reconstruction error using IRPF as a function of rank for 10 different trials, in addition to the reconstruction error as a function of rank for the EY approximation of the fully-sampled original "gold-standard" matrix. Also indicated are the ranks selected by the AIC_c, BIC, and FPE model-selection criteria.

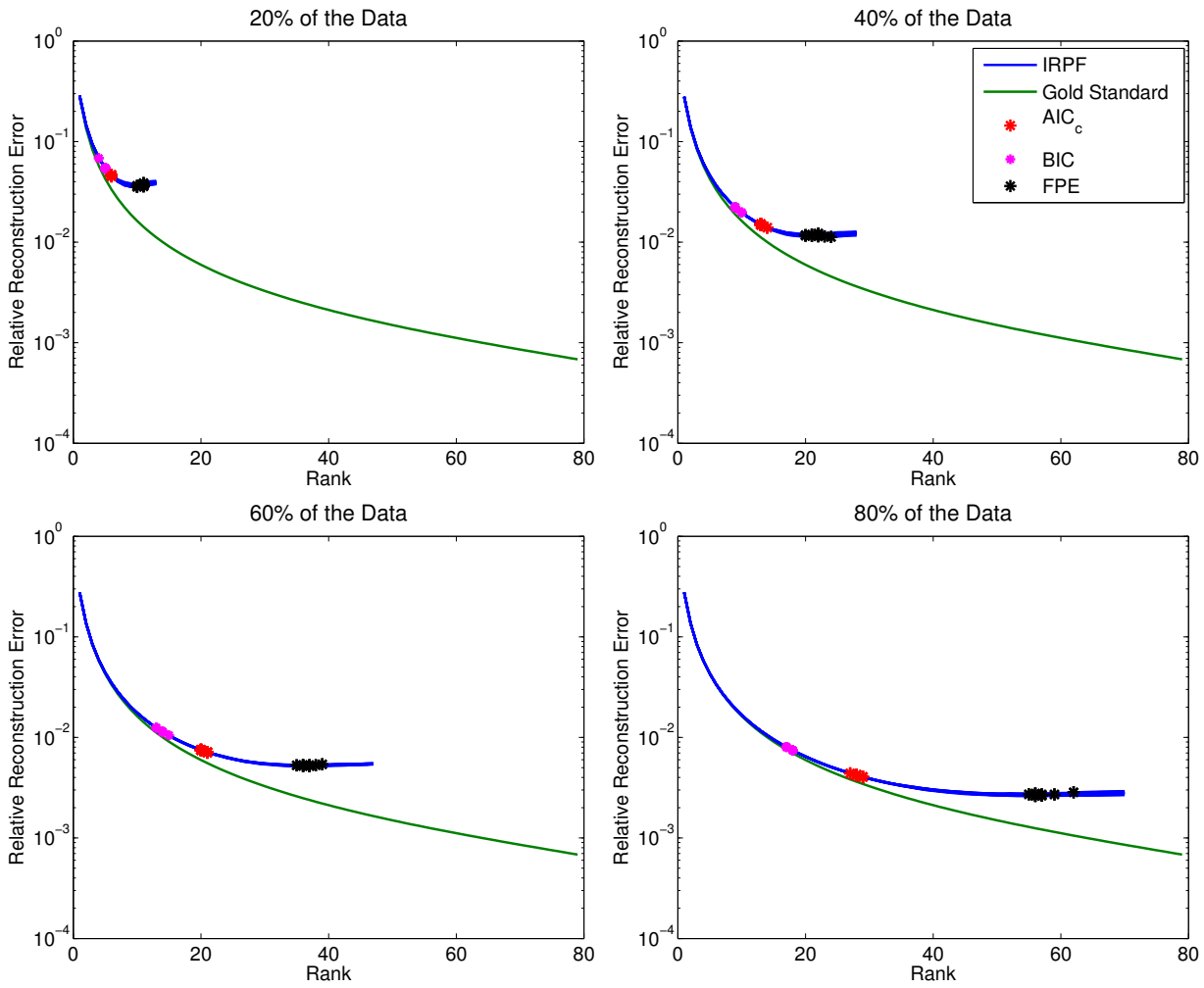


Figure 5.16: Comparison of information-theoretic model-selection criteria for matrix completion problems with different amounts of acquired data. The value of β specifying the singular value distribution was 2. The plots show the reconstruction error using IRPF as a function of rank for 10 different trials, in addition to the reconstruction error as a function of rank for the EY approximation of the fully-sampled original "gold-standard" matrix. Also indicated are the ranks selected by the AIC_c, BIC, and FPE model-selection criteria.

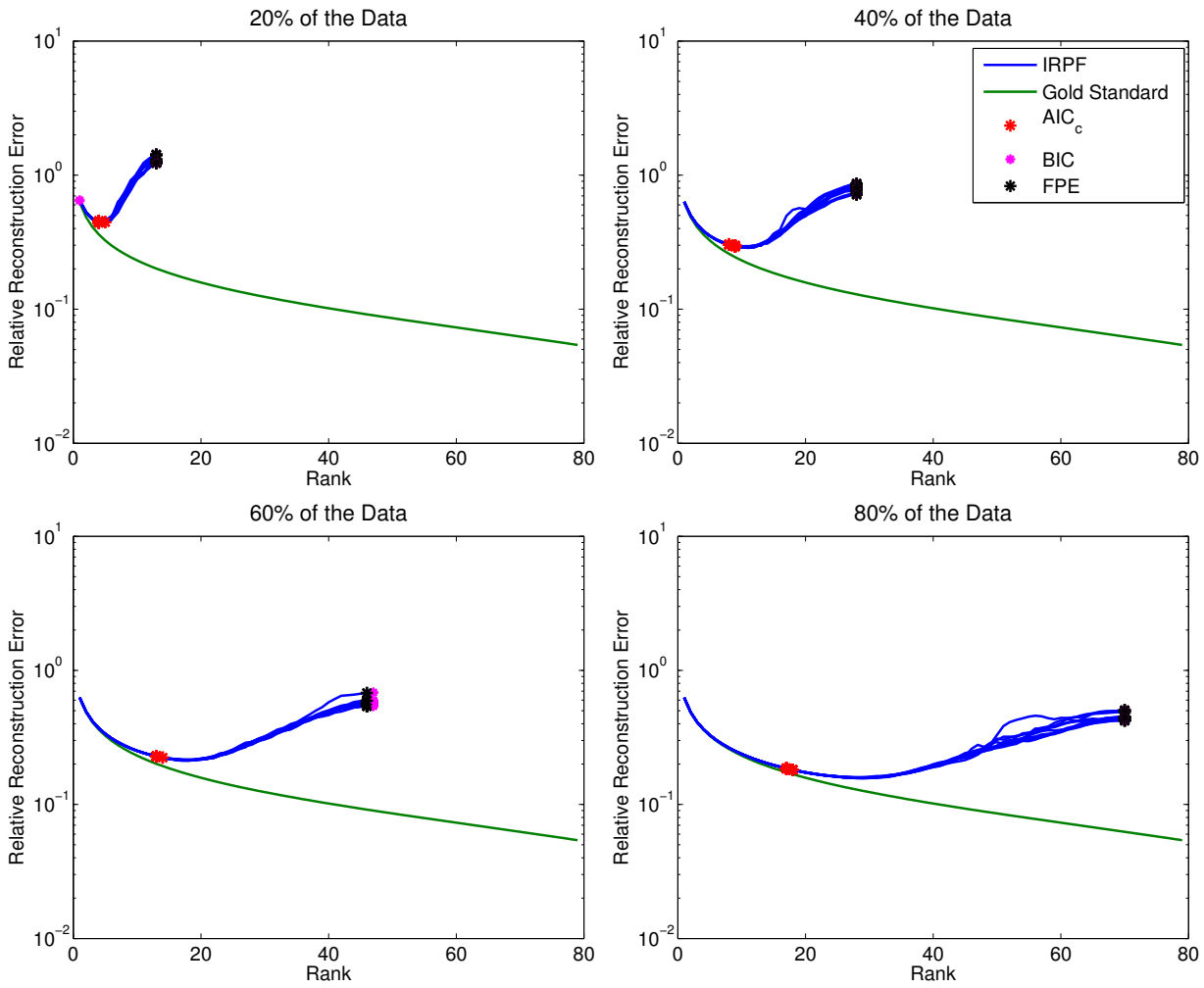


Figure 5.17: Comparison of information-theoretic model-selection criteria for matrix completion problems with different amounts of acquired data. The value of β specifying the singular value distribution was 1. The plots show the reconstruction error using IRPF as a function of rank for 10 different trials, in addition to the reconstruction error as a function of rank for the EY approximation of the fully-sampled original "gold-standard" matrix. Also indicated are the ranks selected by the AIC_c, BIC, and FPE model-selection criteria.

were very poor when $\beta = 1$. This illustrates that it is not necessarily a good idea to rely on these kinds of standard model-selection criteria without rigorous empirical testing in the specific problem context of interest.

Finally, it is also important to note that while the previous simulations judged performance based on reconstruction accuracy (with error measured in the Frobenius norm), this standard metric is not always very relevant for practical applications [636]. For example, revisiting the DCE breast imaging example in Fig. 5.13, it turns out that the minimum reconstruction error is achieved with $L = 2$, despite the fact that this choice fails to capture all of the signal dynamics of interest. We prefer to use $L = 5$, since this choice reconstructs local contrast kinetics more faithfully, despite higher total error. It is also interesting to note that AIC_c chooses $L = 3$, BIC chooses $L = 1$, and FPE chooses $L = 10$. We conclude that selection of L is complicated, and that the choices made by statistical model selection criteria are not necessarily consistent with each other or with our qualitative preferences. As a result, practical choice of L must be adapted based on the specific features and objectives of each imaging scenario.

5.5.3 Other Considerations

One important consideration for image reconstruction using the PS model is that the rank L needs to be small enough relative to both M and N that a constraint on $\text{rank}\,(\mathbf{C})$ significantly reduces the number of degrees of freedom. In practice, this can mean that more significant accelerations are possible for large-scale reconstruction problems where the number of reconstructed time points M is large. For example, a complex rank-5 matrix of size $256^2 \times 18$ has 6.6×10^5 real degrees of freedom and 1.2×10^6 complex entries, meaning that it would be necessary to sample at least 25% of the matrix to have any hope of successful reconstruction. In contrast, a complex rank-5 matrix of size $256^2 \times 52$ still has roughly 6.6×10^5 degrees of freedom, but has significantly more (i.e., 3.4×10^6) entries, so that accurate reconstruction is conceivable with only 10% of the data.

A further consideration is that, in some cases, we know more about a signal than merely the subspace in which it lives (e.g., in the case of parametric imaging experiments, where the observed

temporal contrast variations are known to obey a parametric model with a small number of unknowns). Spatiotemporal reconstruction from limited data in the presence of a parametric contrast model has been previously investigated in Ref. [272], and a procedure based on PS modeling alone would generally be relatively inefficient for this scenario.

It is also important to keep in mind that, similar to CS reconstructions, the spatial resolution and noise characteristics of nonlinear reconstructions based on matrix recovery have not yet been characterized rigorously. Thus, while the reconstructed images might appear visually pleasing, the reconstructions could potentially be missing important image features of interest. As a result, when choosing the PS model for a specific application, it is important to remember that use of the model is based on the assumption that it is reasonable to approximate the original spatiotemporal image by discarding the signal characteristics associated with small singular values. This suggests that the PS model might not be a good choice when the image features of interest would not contribute significantly to the Frobenius norm (e.g., subtle temporal contrast variations from highly-localized spatial regions). However, it may also be possible to alleviate this issue by performing a linear transform on the matrix to increase the transform-domain significance of the spatiotemporal features of interest.

A number of extensions to the proposed PS-based reconstruction scheme are also possible. For example, by invoking general linear sampling operators $\mathcal{A}$, it becomes possible to incorporate any prior information that might be available regarding the known spatial-spectral support of the spatiotemporal image [75, 77, 87]. A preliminary investigation of this is presented in [675], in the context of cardiac MRI. In addition, a more general choice of $\mathcal{A}$ makes it possible to model non-Fourier acquisition physics, which could be useful for a variety of imaging contexts. It is also relatively straightforward to include additional regularization in the formulation of the problem, if additional prior information is available. Note that use of rank-constraints is completely compatible with the use of other types of constraints (e.g., sparsity, support-limits, or anatomical prior information), and a combination of constraints can further improve the quality of reconstruction results over using a single constraint by itself (e.g., [84, 112, 143, 230, 231, 250, 251, 676, 679]).

Another interesting extension is the generalization to low-rank tensor recovery [229]. Low-rank tensor recovery can enable accelerated reconstruction of partially separable functions with higher-dimensional structure (e.g., see Eq. (4) of [399]). Methods designed for low-rank tensor recovery have already been successfully applied to address limited-data problems in computer vision [413], multidimensional NMR spectroscopy [339, 340, 487], and electroencephalography [3], with the potential to also be useful in many other high-dimensional imaging applications.

5.6 Conclusion and Summary

This chapter has presented a matrix recovery approach to estimating spatiotemporal images from sparsely sampled data, based on the assumption that the underlying function is partially separable. Spatiotemporal PS image modeling leads to the formulation of a low-rank matrix recovery problem, which can be solved to yield high-quality reconstructions from "arbitrarily"-sampled data. To solve matrix recovery problems, we proposed and investigated the IRPF algorithm, which was demonstrated to have good performance relative to existing algorithms. The proposed matrix-recovery framework was evaluated using simulated and experimental data. Results illustrated that the matrix-recovery framework was significantly more flexible than existing basic PS methods, and that the matrix-recovery formulation has the capability to enable image reconstruction with highly-undersampled spatiotemporal imaging data.

Chapter 6

Conclusion

As MR technology has developed over the past several decades, MRI has increasingly been used to study biological tissues and organs from a large number of different perspectives. Biological tissues are extremely complex, and a single voxel in typical MRI studies frequently contains signal contributions from multiple chemical species in a number of distinct environments (e.g., cytoplasm, extracellular fluid, and the surfaces and interiors of various organelles). Under different types of MR contrast encoding, each environment can influence the NMR signal in distinct ways. As a result, MRI can offer the ability to untangle the contributions from different environments, allowing MR experiments to probe the detailed anatomy, physiology, metabolism, and temporal behavior of normal and pathological tissues. However, despite this potential, MR technology has not matured enough to take full advantage of its capabilities. In particular, data acquisition for MR studies is relatively slow, which limits throughput and prevents many applications that would require high-dimensional encoding. In addition, experiments are frequently also limited by the low sensitivity of the NMR phenomenon.

This dissertation proposed novel approaches to address speed and sensitivity limitations in MRI. We developed fast data acquisition and image reconstruction methods that combine (1) the modeling and manipulation of physical imaging processes; (2) the use of a statistical modeling framework for image reconstruction; and (3) the use of novel constrained signal and image models. The utility of these methods has been demonstrated in a number of important MRI contexts, with significant potential for extension to other MR applications and other imaging modalities. The main contributions of this work include:

- The development and characterization of a novel method for enhancing the SNR of corre-

lated image sequences. This method makes use of a novel MRF model for correlated images, which uses explicit shared line-process variables to model correlated image structure and facilitates fast computation. The method has the capability to reduce noise contamination, while simultaneously preserving the high-resolution structure that is shared between the different frames of a sequence of images. We have shown that the resolution and SNR characteristics of the reconstruction scheme are easy to evaluate, giving users the ability to precisely control the trade-off between resolution and SNR. In addition, our analysis suggests new ways for designing MR experiments. In contrast to conventional wisdom, we have demonstrated that resolution/SNR efficiency can improve significantly if *k*-space data is acquired at frequencies beyond the nominal resolution of the experiment. This observation has many implications for the way that MR data acquisition should be performed.

- The development and characterization of a novel non-Fourier data acquisition method to accelerate encoding for images that are approximately sparse in a known transform domain. We have demonstrated that Fourier-domain undersampling can be inefficient for many CS-MRI applications, and that random encoding could overcome this limitation of Fourier encoding in high-SNR scenarios. In addition, we have demonstrated that the strong performance guarantees from CS theory are not applicable for many real CS-MRI problems of interest, underscoring the need for careful evaluation of CS-inspired MRI approaches on an application-by-application basis.

- The development and characterization of a novel framework for reconstructing spatiotemporal images based on the use of PS-based image modeling. We have placed PS-based spatiotemporal image reconstruction in the context of low-rank matrix recovery, and introduced and evaluated a new efficient algorithm for solving the resulting optimization problem. The proposed framework is significantly more flexible than previous PS-based reconstruction approaches, and was demonstrated to enable high-quality reconstruction from undersampled data in a number of spatiotemporal MRI contexts.

The use of novel imaging constraints will likely be essential for enabling the practical implementation of future generations of advanced MR experiments. This dissertation focused on constraints derived from two sources of prior information: (1) coregistered reference images from the same imaging subject, and (2) the fact that medical images can frequently be compressed using sparsity and/or low-dimensionality constraints. However, there are many other sources of prior information that could also be leveraged to improve MR imaging. Image databases, anatomical atlases, and historical medical images of the same patient contain a wealth of information that could be used to constrain image acquisition and reconstruction. In addition, it would also be possible to use other forms of context-based prior information to enhance MR imaging of specific body parts. For example, when reconstructing an MR brain image, it could be possible to leverage prior statistical information about the sizes, shapes, and spatial/geometrical relationships between different anatomical brain structures. These kinds of constraints are quite commonly used in image analysis applications [190, 583], but have largely not been used to influence the design of imaging experiments and image reconstruction methods.

Through centuries of careful observation and experimentation, humanity has accumulated a tremendous amount of knowledge about the characteristics of biological systems. However, to date, only a small fraction of this knowledge has been used to enhance data acquisition and image reconstruction in medical imaging applications. It will be a tremendous step forward for medical imaging when technology reaches the stage where data acquisition, image reconstruction, image interpretation, and image analysis are more tightly connected to each other, within a single integrated framework that can take advantage of all the prior information that is available.

References

[1] J. Abernethy, F. Bach, T. Evgeniou, and J.-P. Vert, "A new approach to collaborative filtering: Operator estimation with spectral regularization," *J. Mach. Learn. Res.*, vol. 10, pp. 803–826, 2009. 161, 169

[2] A. Abragam, *The Principles of Nuclear Magnetism*. London: Oxford University Press, 1961. 7

[3] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometr. Intell. Lab. Syst.*, vol. 106, pp. 41–56, 2011. 203

[4] D. Achlioptas and F. McSherry, "Fast computation of low rank matrix approximations," *J. ACM*, vol. 54, p. 9, 2007. 171

[5] J. J. H. Ackerman, "Surface (local) coils as NMR receivers," *Concepts Magn. Reson.*, vol. 2, pp. 33–42, 1990. 14

[6] N. Aggarwal and Y. Bresler, "Patient-adapted reconstruction and acquisition dynamic imaging method (PARADIGM) for MRI," *Inverse Probl.*, vol. 24, p. 045015, 2008. 115, 159, 160

[7] C. B. Ahn, J. H. Kim, and Z. H. Cho, "High-speed spiral-scan echo planar NMR imaging–I," *IEEE Trans. Med. Imag.*, vol. MI-5, pp. 2–7, 1986. 13

[8] S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET," *IEEE Trans. Med. Imag.*, vol. 27, pp. 413–424, 2008. 52, 154

[9] S. Aja-Fernández, C. Alberola-López, and C.-F. Westin, "Noise and signal estimation in magnitude MRI and Rician distributed images: a LMMSE approach," *IEEE Trans. Image Process.*, vol. 17, pp. 1383–1398, 2008. 16, 36

[10] S. Aja-Fernández, M. Niethammer, M. Kubicki, M. E. Shenton, and C.-F. Westin, "Restoration of DWI data using a Rician LMMSE estimator," *IEEE Trans. Med. Imag.*, vol. 27, pp. 1389–1403, 2008. 16, 36

[11] S. Ajraoui, K. J. Lee, M. H. Deppe, S. R. Parnell, J. Parra-Robles, and J. M. Wild, "Compressed sensing in hyperpolarized ^{3}He lung MRI," *Magn. Reson. Med.*, vol. 63, pp. 1059–1069, 2010. 116, 152

[12] A. Aldroubi and K. Gröchenig, “Nonuniform sampling and reconstruction in shift-invariant spaces,” *SIAM Rev.*, vol. 43, pp. 585–620, 2001. 18

[13] A. M. Alessio and P. E. Kinahan, “Improved quantitation for PET/CT image reconstruction with system modeling and anatomical priors,” *Med. Phys.*, vol. 33, pp. 4095–4103, 2006. 34, 38, 39

[14] M. E. Alexander, R. Baumgartner, A. R. Summers, C. Windischberger, M. Klarhoefer, E. Moser, and R. L. Somorjai, “A wavelet-based method for improving signal-to-noise ratio and contrast in MR images,” *Magn. Reson. Imag.*, vol. 18, pp. 169–180, 2000. 36

[15] M. Allain, J. Idier, and Y. Goussard, “On global and local convergence of half-quadratic algorithms,” *IEEE Trans. Image Process.*, vol. 15, pp. 1130–1142, 2006. 40, 47, 49, 78, 104

[16] A. H. Andersen, “On the Rician distribution of noisy MRI data,” *Magn. Reson. Med.*, vol. 36, pp. 331–333, 1996. 16

[17] A. H. Andersen, D. M. Gash, and M. J. Avison, “Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework,” *Magn. Reson. Imag.*, vol. 17, pp. 795–815, 1999. 162

[18] D. F. Andrews and C. L. Mallows, “Scale mixtures of normal distributions,” *J. R. Stat. Soc. Ser. B*, vol. 36, pp. 99–102, 1974. 42

[19] D. T. Andrews and P. D. Wentzell, “Applications of maximum likelihood principal component analysis: incomplete data sets and calibration transfer,” *Anal. Chim. Acta*, vol. 350, pp. 341–352, 1997. 161

[20] B. A. Ardekani, M. Braun, B. F. Hutton, I. Kanno, and H. Iida, “Minimum cross-entropy reconstruction of PET images using prior anatomical information,” *Phys. Med. Biol.*, vol. 41, pp. 2497–2517, 1996. 34

[21] K. Arfanakis, D. Cordes, V. M. Haughton, J. D. Carew, and M. E. Meyerand, “Independent component analysis applied to diffusion tensor MRI,” *Magn. Reson. Med.*, vol. 47, pp. 354–363, 2002. 37, 162, 171

[22] H.-E. Assemlal, D. Tschumperlé, and L. Brun, “Efficient and robust computation of PDF features from diffusion MR signal,” *Med. Image Anal.*, vol. 13, pp. 715–729, 2009. 37

[23] I. Atkinson, F. Kamalabadi, S. Mohan, and D. L. Jones, “Asymptotically optimal blind estimation of multichannel images,” *IEEE Trans. Image Process.*, vol. 15, pp. 992–1007, 2006. 37, 86

[24] I. C. Atkinson, A. Lu, J. P. Haldar, Z.-P. Liang, and K. R. Thulborn, “Human 17-oxygen imaging at 9.4T and enhanced reconstruction using 23-sodium,” in *Proc. Amer. Soc. Neuroradiology*, 2008, p. 214. 4, 84

[25] I. C. Atkinson, K. R. Thulborn, A. Lu, J. Haldar, X. J. Zhou, T. Claiborne, and Z.-P. Liang, "Quantitative 23-sodium and 17-oxygen MR imaging in human brain at 9.4 Tesla enhanced by constrained k-space reconstruction," in *Proc. Int. Soc. Magn. Reson. Med.*, 2008, p. 335. 4, 84

[26] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*. New York: Springer, 2006. 46

[27] W. P. Aue, S. Müller, T. A. Cross, and J. Seelig, "Volume-selective excitation. a novel approach to topical NMR," *J. Magn. Reson.*, vol. 56, pp. 350–354, 1984. 14

[28] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, "On the relation between sparse reconstruction and parameter estimation with model order selection," *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 560–570, 2010. 169

[29] S. P. Awate and R. T. Whitaker, "Feature-preserving MRI denoising: A nonparametric empirical Bayes approach," *IEEE Trans. Med. Imag.*, vol. 26, pp. 1242–1254, 2007. 16, 36

[30] K. Baete, J. Nuyts, K. Van Laere, W. Van Paesschen, S. Ceyssens, L. De Ceuninck, O. Gheysens, A. Kelles, J. Van den Eynden, P. Suetens, and P. Dupont, "Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET," *NeuroImage*, vol. 23, pp. 305–317, 2004. 34, 38, 39

[31] R. Bansal, L. H. Staib, D. Xu, A. F. Laine, J. Liu, and B. S. Peterson, "Using perturbation theory to reduce noise in diffusion tensor fields," *Med. Image Anal.*, vol. 13, pp. 580–597, 2009. 37

[32] Y. Bao and A. A. Maudsley, "Improved reconstruction for MRI spectroscopic imaging," *IEEE Trans. Med. Imag.*, vol. 26, pp. 686–695, 2007. 21, 34

[33] R. L. Barbour, H. L. Graber, J. Chang, S.-L. S. Barbour, P. C. Koo, and R. Aronson, "MRI-guided optical tomography: prospects and computation for a new imaging method," *IEEE Comp. Sci. Eng.*, vol. 2, pp. 63–77, 1995. 34

[34] A. V. Barger, W. F. Block, Y. Toropov, T. M. Grist, and C. A. Mistretta, "Time-resolved contrast-enhanced imaging with isotropic resolution and broad coverage using an undersampled 3D projection trajectory," *Magn. Reson. Med.*, vol. 48, pp. 297–305, 2002. 116, 121

[35] D. J. Bartholomew and M. Knott, *Latent variable models and factor analysis*, 2nd ed. London: Arnold, 1999. 161

[36] S. Basu, T. Fletcher, and R. Whitaker, "Rician noise removal in diffusion tensor MRI," in *Proc. MICCAI*, ser. Lecture Notes in Computer Science, M. Nielson and J. Sporring, Eds., vol. 4190. New York: Springer-Verlag, 2006, pp. 117–125. 16, 36, 37

[37] P. J. Beatty, D. G. Nishimura, and J. M. Pauly, "Rapid gridding reconstruction with a minimal oversampling ratio," *IEEE Trans. Med. Imag.*, vol. 24, pp. 799–808, 2005. 19

[38] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," California Institute of Technology, Tech. Rep., 2009. [Online]. Available: http://www.acm.caltech.edu/~nesta/ 107, 108, 131, 132

[39] R. Behin, J. Bishop, and R. M. Henkelman, "Dynamic range requirements for MRI," *Concepts Magn. Reson. B*, vol. 26B, pp. 28–35, 2005. 16

[40] R. M. Bell, Y. Koren, and C. Volinsky, "All together now: A perspective on the Netflix prize," *Chance*, vol. 23, pp. 24–29, 2010. 161

[41] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, pp. 5030–5043, 2010. 120, 150

[42] S. Benameur, M. Mignotte, J. Meunier, and J.-P. Soucy, "Image restoration using functional and anatomical information fusion with application to SPECT-MRI images," *Int. J. Biomed. Imaging*, vol. 2009, p. 843160, 2009. 34

[43] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process.*, vol. 14, pp. 1814–1830, 2005. 37

[44] J. J. Benedetto and P. J. S. G. Ferreira, Eds., *Modern Sampling Theory: Mathematics and Applications*. Boston: Birkhäuser, 2001. 18

[45] J. J. Benedetto and H. C. Wu, "Nonuniform sampling and spiral MRI reconstruction," *Proc. SPIE*, vol. 4119, pp. 130–141, 2000. 18

[46] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. Burlington: Elsevier Academic Press, 2004. 7, 13

[47] M. A. Bernstein, D. M. Thomasson, and W. H. Perman, "Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction," *Med. Phys.*, vol. 16, pp. 813–817, 1989. 16

[48] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. London: Institute of Physics Publishing, 1998. 17, 20, 118

[49] M. Bertero, C. De Mol, and E. R. Pike, "Linear inverse problems with discrete data. I: General formulation and singular system analysis," *Inverse Probl.*, vol. 1, pp. 301–330, 1985. 20

[50] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc. Ser. B*, vol. 48, pp. 259–302, 1986. 24, 38

[51] J. E. Besag, "Towards Bayesian image analysis," *J. Appl. Statist.*, vol. 16, pp. 395–407, 1989. 38

[52] G. Beylkin and M. J. Mohlenkamp, “Numerical operator calculus in higher dimensions,” *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 10 246–10 251, 2002. 160

[53] ——, “Algorithms for numerical analysis in high dimensions,” *SIAM J. Sci. Comput.*, vol. 26, pp. 2133–2159, 2005. 160

[54] M. J. Black and A. Rangarajan, “On the unification of line processes, outlier rejection, and robust statistics with applications in early vision,” *Int. J. Comput. Vis.*, vol. 19, pp. 57–91, 1996. 40, 43, 44, 45, 74

[55] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, “Robust anisotropic diffusion,” *IEEE Trans. Image Process.*, vol. 7, pp. 421–432, 1998. 40

[56] R. E. Blahut, *Theory of Remote Image Formation*. Cambridge: Cambridge University Press, 2004. 18

[57] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge: The MIT Press, 1987. 43

[58] R. Blanco-Sequeiros, R. Ojala, J. Kariniemi, J. Perälä, J. Niinimäki, H. Reinikainen, and O. Tervonen, “MR-guided interventional procedures: A review,” *Acta Radiol.*, vol. 46, pp. 576–586, 2005. 158

[59] M. S. Blasco, S. Krishnan, D. Moratal, S. Ramamurthy, and M. E. Brummer, “High spatial frequencies are more dynamic than low spatial frequencies in cardiac motion,” in *Proc. Int. Soc. Magn. Reson. Med.*, 2009, p. 4567. 191

[60] K. T. Block, M. Uecker, and J. Frahm, “Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint.” *Magn. Reson. Med.*, vol. 57, pp. 1086–1098, 2007. 116, 131, 152

[61] P. Blomgren and T. F. Chan, “Color TV: Total variation methods for restoration of vector-valued images,” *IEEE Trans. Image Process.*, vol. 7, pp. 304–309, 1998. 37

[62] B. Blümich, “Stochastic NMR imaging,” *J. Magn. Reson.*, vol. 60, pp. 37–45, 1984. 14

[63] ——, “Spatially resolved 2D spectroscopy with stochastic RF excitation,” *J. Magn. Reson.*, vol. 90, pp. 535–543, 1990. 14

[64] B. Blümich and H. W. Spiess, “NMR imaging with incommensurate sampling and gradient modulation rates,” *J. Magn. Reson.*, vol. 66, pp. 66–73, 1986. 14

[65] L. Bolinger and J. S. Leigh, “Hadamard spectroscopic imaging (HSI) for multivolume localization,” *J. Magn. Reson.*, vol. 80, pp. 162–167, 1988. 14

[66] J.-M. Bonny, O. Boespflug-Tanguly, M. Zanca, and J.-P. Renou, “Multi-exponential analysis of magnitude MR images using a quantitative multispectral edge-preserving filter,” *J. Magn. Reson.*, vol. 161, pp. 25–34, 2003. 16, 37

[67] P. A. Bottomley, “Localized NMR spectroscopy by the sensitive point method,” *J. Magn. Reson.*, vol. 50, pp. 335–338, 1982. 12

[68] ——, “Spatial localization in NMR spectroscopy *in vivo*,” *Ann. NY Acad. Sci.*, vol. 508, pp. 333–348, 1987. 14

[69] P. A. Bottomley, T. B. Foster, and R. D. Darrow, “Depth-resolved surface-coil spectroscopy (DRESS) for *in vivo* ^{1}H, ^{31}P, and ^{13}C NMR,” *J. Magn. Reson.*, vol. 59, pp. 338–342, 1984. 14

[70] N. Boussion, M. Hatt, F. Lamare, Y. Bizais, A. Turzo, C. Cheze-Le Rest, and D. Visvikis, “A multiresolution image based approach for correction of partial volume effects in emission tomography,” *Phys. Med. Biol.*, vol. 51, pp. 1857–1876, 2006. 34

[71] G. Boverman, E. L. Miller, A. Li, Q. Zhang, T. Chaves, D. H. Brooks, and D. A. Boas, “Quantitative spectroscopic diffuse optical tomography of the breast guided by imperfect *a priori* structural information,” *Phys. Med. Biol.*, vol. 50, pp. 3941–3956, 2005. 34

[72] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman, “Bayesian reconstruction and use of anatomical *a priori* information for emission tomography,” *IEEE Trans. Med. Imag.*, vol. 15, pp. 673–686, 1996. 34

[73] G. E. P. Box, J. S. Hunter, and W. G. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed. Hoboken: Wiley-Interscience, 2005. 118

[74] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004. 78, 80, 97

[75] J. G. Brankov, M. N. Wernick, M. A. King, Y. Yang, and M. V. Narayanan, “Spatially adaptive temporal smoothing for reconstruction of dynamic image sequences,” *IEEE Trans. Nuc. Sci.*, vol. 53, pp. 2769–2777, 2006. 37, 162, 171, 202

[76] J. G. Brankov, Y. Yang, and M. N. Wernick, “Tomographic imaging reconstruction based on a content-adaptive mesh model,” *IEEE Trans. Med. Imag.*, vol. 23, pp. 202–212, 2004. 21, 34

[77] Y. Bresler, N. Aggarwal, and B. Sharif, “Patient-adaptive spatio-temporal MRI: From PARADIGM to PARADISE and beyond,” in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2007, pp. 980–983. 202

[78] Y. Bresler and S. P. Litke, “A parametric technique for superresolution image reconstruction,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1990, pp. 1205–1208. 116

[79] X. Bresson and T. F. Chan, “Fast dual minimization of the vectorial total variation norm and applications to color image processing,” *Inverse Probl. Imaging*, vol. 2, pp. 455–484, 2008. 37, 78

[80] G. L. Bretthorst, “Automatic phasing of MR images. Part I: Linearly varying phase,” *J. Magn. Reson.*, vol. 191, pp. 184–192, 2008. 16

[81] ——, "Automatic phasing of MR images. Part II: Voxel-wise phase estimation," *J. Magn. Reson.*, vol. 191, pp. 193–201, 2008. 16

[82] ——, "Nonuniform sampling: Bandwidth and aliasing," *Concepts Magn. Reson. A*, vol. 32A, pp. 417–435, 2008. 116, 121

[83] F. A. Breuer, M. Blaimer, R. M. Heidemann, M. F. Mueller, M. A. Griswold, and P. M. Jakob, "Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging," *Magn. Reson. Med.*, vol. 53, pp. 684–691, 2005. 14, 124

[84] 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. 159, 160, 162, 202

[85] C. Brinegar, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang, "Real-time cardiac MRI without triggering, gating, or breath holding," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2008, pp. 3381–3384. 159, 160

[86] C. Brinegar, H. Zhang, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang, "First-pass perfusion cardiac MRI using the partially separable functions model with generalized support," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2010, pp. 2833–2836. 159, 160

[87] C. Brinegar, H. Zhang, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, D. Pocci, F. Lam, C. Ho, and Z.-P. Liang, "Real-time cardiac MRI using prior spatial-spectral information," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2009, pp. 4383–4386. 159, 160, 162, 202

[88] C. Brito-Loeza and K. Chen, "On high-order denoising models and fast algorithms for vector-valued images," *IEEE Trans. Image Process.*, vol. 19, pp. 1518–1527, 2010. 37

[89] R. Bro, "PARAFAC: Tutorial and applications," *Chemometr. Intell. Lab. Syst.*, vol. 38, pp. 149–171, 1997. 174

[90] H. R. Brooker, T. H. Mareci, and J. Mao, "Selective Fourier transform localization," *Magn. Reson. Med.*, vol. 5, pp. 417–433, 1987. 92

[91] B. A. Brooksby, H. Dehghani, B. W. Pogue, and K. D. Paulsen, "Near-infrared (NIR) tomography breast image reconstruction with *A Priori* structural information from MRI: Algorithm development for reconstructing heterogeneities," *IEEE J. Sel. Topics Quant. Electron.*, vol. 9, pp. 199–209, 2003. 34

[92] M. E. Brummer, D. Moratal-Pérez, C.-Y. Hong, R. I. Pettigrew, J. Millet-Roig, and W. T. Dixon, "Noquist: Reduced field-of-view imaging by direct Fourier inversion," *Magn. Reson. Med.*, vol. 51, pp. 331–342, 2004. 115, 159, 160

[93] P. P. Bruyant, H. C. Gifford, G. Gindi, and M. A. King, "Numerical observer study of MAP-OSEM regularization methods with anatomical priors for lesion detection in ^{67}Ga images," *IEEE Trans. Nuc. Sci.*, vol. 51, pp. 193–197, 2004. 34, 38, 39

[94] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *SIAM J. Multiscale Model. Simul.*, vol. 4, pp. 490–530, 2005. 36

[95] ——, "Image denoising methods. a new nonlocal principle," *SIAM Rev.*, vol. 52, pp. 113–147, 2010. 36

[96] A. M. Buchanan and A. W. Fitzgibbon, "Damped Newton algorithms for matrix factorization with missing data," in *IEEE Conf. Comp. Vision Patt. Recog.*, 2005, pp. 316–322. 161, 174

[97] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer-Verlag, 2002. 169, 194, 195

[98] R. B. Buxton, *Introduction to Functional Magnetic Resonance Imaging: Principles & Techniques*. Cambridge: Cambridge University Press, 2002. 15, 16, 158

[99] ——, "Spatial-smoothing to improve SNR," in *Introduction to Functional Magnetic Resonance Imaging: Principles & Techniques*. Cambridge: Cambridge University Press, 2002, pp. 280–283. 65, 68

[100] M. Bydder and J. Du, "Noise reduction in multiple-echo data sets using singular value decomposition," *Magn. Reson. Imag.*, vol. 24, pp. 849–856, 2006. 37, 88, 162, 171

[101] M. Bydder and M. D. Robson, "Partial Fourier partially parallel imaging," *Magn. Reson. Med.*, vol. 53, pp. 1393–1401, 2005. 21, 34, 89

[102] J.-F. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, pp. 1956–1982, 2010. 169, 177

[103] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4388–4394, 2010. 120, 121

[104] ——, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Inf. Theory*, vol. 56, pp. 3516–3522, 2010. 120, 150

[105] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 358–374, 2010. 121

[106] G. Camba-Mendez and G. Kapetanios, "Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling," *Economet. Rev.*, vol. 28, pp. 581–611, 2009. 194

[107] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Probl.*, vol. 23, pp. 969–985, 2007. 120, 150

[108] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, 2005. 115, 120, 121, 122, 152

[109] ——, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006. 115, 122, 124

[110] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008. 120, 121

[111] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, 2010, http://dx.doi.org/10.1016/j.acha.2010.10.002. 120, 122

[112] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Preprint*, 2009, http://arxiv.org/abs/0912.3599. 161, 170, 202

[113] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, pp. 925–936, 2010. 161, 169, 170, 173, 182

[114] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, pp. 717–772, 2009. 161, 170, 173, 182

[115] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, 2006. 115

[116] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006. 115, 120, 121

[117] E. J. Candès and T. Tao, "Reflections on compressed sensing," *IEEE Info. Theory Soc. News.*, vol. 58, pp. 14–17, 2008. 5, 24, 169, 183, 189

[118] ——, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2053–2080, 2010. 161, 170, 173, 182

[119] Y. Cao and D. N. Levin, "Feature-recognizing MRI," *Magn. Reson. Med.*, vol. 30, pp. 305–317, 1993. 21, 34, 116, 162

[120] ——, "On the relationship between feature-recognizing MRI and MRI encoded by singular value decomposition," *Magn. Reson. Med.*, vol. 33, pp. 140–142, 1995. 116, 162

[121] Y. Cao, D. N. Levin, and L. Yao, "Locally focused MRI," *Magn. Reson. Med.*, vol. 34, pp. 858–867, 1995. 21, 34, 116

[122] A. Cárdenas-Blanco, C. Tejos, P. Irarrazaval, and I. Cameron, "Noise in magnitude magnetic resonance images," *Concepts Magn. Reson. A*, vol. 32A, pp. 409–416, 2008. 16

[123] L. Carleson, P. Malliavin, J. Neuberger, and J. Wermer, Eds., *Collected Works of Arne Beurling*. Boston: Birkhäuser, 1989. 18

[124] J. W. Carlson, "An algorithm for NMR imaging reconstruction based on multiple RF receiver coils," *J. Magn. Reson.*, vol. 74, pp. 376–380, 1987. 14

[125] J. W. Carlson and T. Minemura, “Imaging time reduction through multiple receiver coil data acquisition and image reconstruction,” *Magn. Reson. Med.*, vol. 29, pp. 681–688, 1993. 14

[126] J. D. Carroll and J.-J. Chang, “Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition,” *Psychometrika*, vol. 35, pp. 283–319, 1970. 174

[127] Y. Censor, “Finite series-expansion reconstruction methods,” *Proc. IEEE*, vol. 71, pp. 409–419, 1983. 22

[128] C. Chan, R. Fulton, D. D. Feng, and S. Meikle, “Regularized image reconstruction with an anatomically adaptive prior for positron emission tomography,” *Phys. Med. Biol.*, vol. 54, pp. 7379–7400, 2009. 34, 38, 39

[129] T. F. Chan and W. L. Wan, “Analysis of projection methods for solving linear systems with multiple right-hand sides,” *SIAM J. Sci. Comput.*, vol. 18, pp. 1698–1721, 1997. 49

[130] S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Trans. Image Process.*, vol. 9, pp. 1532–1546, 2000. 88

[131] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, “Deterministic edge-preserving regularization in computed imaging,” *IEEE Trans. Image Process.*, vol. 6, pp. 298–310, 1997. 24, 40, 43, 44, 45, 47, 49, 78

[132] R. Chartrand, “Exact reconstruction of sparse signals via nonconvex minimization,” *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007. 120

[133] R. Chartrand and V. Staneva, “Restricted isometry properties and nonconvex compressive sensing,” *Inverse Probl.*, vol. 24, p. 035020, 2008. 120

[134] C. Chefd’hotel, D. Tschumperlé, R. Deriche, and O. Faugeras, “Regularizing flows for constrained matrix-valued images,” *J. Math. Imaging Vis.*, vol. 20, pp. 147–162, 2004. 37

[135] B. Chen and E. W. Hsu, “Noise removal in magnetic resonance diffusion tensor imaging,” *Magn. Reson. Med.*, vol. 54, pp. 393–407, 2005. 37

[136] C.-T. Chen, X. Ouyang, W. H. Wong, X. Hu, V. E. Johnson, C. Ordonez, and C. E. Metz, “Sensor fusion in image reconstruction,” *IEEE Trans. Nuc. Sci.*, vol. 38, pp. 687–692, 1991. 34, 38, 39

[137] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998. 24, 44

[138] Y. Chi, A. Pezeshki, L. L. Scharf, and R. Calderbank, “Sensitivity to basis mismatch of compressed sensing,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 3930–3933. 155

[139] P.-C. Chiao, W. L. Rogers, J. A. Fessler, N. H. Clinthorne, and A. O. Hero, "Model-based estimation with boundary side information or boundary regularization," *IEEE Trans. Med. Imag.*, vol. 13, pp. 227–234, 1994. 34

[140] D. Cho and T. D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Process.: Image Comm.*, vol. 20, pp. 77–89, 2005. 40

[141] P. L. Choyke, A. J. Dwyer, and M. V. Knopp, "Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging," *J. Magn. Reson. Imag.*, vol. 17, pp. 509–520, 2003. 158

[142] O. Christiansen, T.-M. Lee, J. Lie, U. Sinha, and T. F. Chan, "Total variation regularization of matrix-valued images," *Int. J. Biomed. Imaging*, vol. 2007, p. 27432, 2007. 37

[143] A. G. Christodoulou, C. Brinegar, J. P. Haldar, H. Zhang, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang, "High-resolution cardiac MRI using partially separable functions and weighted spatial smoothness regularization," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2010, pp. 883–886. 159, 160, 202

[144] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Am. Math. Soc.*, vol. 22, pp. 211–231, 2009. 120

[145] C. Comtat, P. E. Kinahan, J. A. Fessler, T. Beyer, D. W. Townsend, M. Defrise, and C. Michel, "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels," *Phys. Med. Biol.*, vol. 47, pp. 1–20, 2002. 34, 38, 39

[146] C. D. Constantinides, R. G. Weiss, R. Lee, D. Bolar, and P. A. Bottomley, "Restoration of low resolution metabolic images with a priori anatomic information: ^{23}Na MRI in myocardial infarction," *Magn. Reson. Imag.*, vol. 18, pp. 461–471, 2000. 34

[147] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, pp. 2477–2488, 2005. 40

[148] O. Coulon, D. C. Alexander, and S. Arridge, "Diffusion tensor magnetic resonance image regularization," *Med. Image Anal.*, vol. 8, pp. 47–67, 2004. 37

[149] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 27, pp. 425–441, 2008. 36

[150] J. G. Cragg and S. G. Donald, "Inferring the rank of a matrix," *J. Economet.*, vol. 76, pp. 223–250, 1997. 194

[151] C. H. Cunningham, G. A. Wright, and M. L. Wood, "High-order multiband encoding in the heart," *Magn. Reson. Med.*, vol. 48, pp. 689–698, 2002. 14, 124, 128

[152] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, pp. 2080–2095, 2007. 36

[153] W. Dai and O. Milenkovic, "SET: An algorithm for consistent matrix completion," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 3646–3649. 174, 183

[154] R. Damadian, L. Minkoff, M. Golsmith, M. Stanford, and J. Koutcher, "Field focusing nuclear magnetic resonance (FONAR): Visualization of a tumor in a live animal," *Science*, vol. 194, pp. 1430–1432, 1976. 12

[155] A. d'Aspremont, F. Bach, and L. El Ghaoui, "Optimal solutions for sparse principal component analysis," *J. Mach. Learn. Res.*, vol. 9, pp. 1269–1294, 2008. 150

[156] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Math. Program., Ser. B*, vol. 127, pp. 123–144, 2011. 150

[157] M. E. Davies and R. Gribonval, "Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2203–2214, 2009. 120

[158] S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue, and K. D. Paulsen, "Image-guided diffuse optical fluorescence tomography implemented with Laplacian-type regularization," *Opt. Express*, vol. 15, pp. 4066–4082, 2007. 34, 38, 39

[159] S. De Backer, A. Pižurica, B. Huysmans, W. Philips, and P. Scheunders, "Denoising of multicomponent images using wavelet least-squares estimators," *Image Vis. Comput.*, vol. 26, pp. 1038–1051, 2008. 37

[160] F. De la Torre and M. J. Black, "A framework for robust subspace learning," *Int. J. Comput. Vis.*, vol. 54, pp. 117–142, 2003. 161, 174

[161] F. de Pasquale, G. Sebastiani, E. Egger, L. Guidoni, A. M. Luciani, P. Marzola, R. Manfredi, M. Pacilio, A. Piermattei, V. Viti, and P. Barone, "Bayesian estimation of relaxation times T_1 in MR images of irradiated Fricke-agarose gels," *Magn. Reson. Imag.*, vol. 18, pp. 721–731, 2000. 37

[162] L. de Rochefort, T. Liu, B. Kressler, J. Liu, P. Pincemaille, V. Lebon, J. Wu, and Y. Wang, "Quantitative susceptibility map reconstruction from MR phase data using Bayesian regularization: Validation and application to brain imaging," *Magn. Reson. Med.*, vol. 63, pp. 194–206, 2010. 34, 38, 39

[163] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE Trans. Image Process.*, vol. 7, pp. 204–221, 1998. 24, 40, 44, 45, 47, 49, 78, 81

[164] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 2024–2036, 1989. 23

[165] A. J. den Dekker and A. van den Bos, "Resolution: a survey," *J. Opt. Soc. Am. A*, vol. 14, pp. 547–557, 1997. 67

[166] T. S. Denney and S. J. Reeves, "Bayesian image reconstruction from Fourier-domain samples using prior edge information," *J. Electron. Imaging*, vol. 14, p. 043009, 2005. 34, 38, 39

[167] R. Deriche, D. Tschumperlé, and C. Lenglet, "DT-MRI estimation, regularization and fiber tractography," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2004, pp. 9–12. 37

[168] B. Desplanques, J. Cornelis, E. Achten, R. Van de Walle, and I. Lemahieu, "Iterative reconstruction of magnetic resonance images from arbitrary samples in k-space," *IEEE Trans. Nuc. Sci.*, vol. 49, pp. 2268–2273, 2002. 22

[169] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complexity*, vol. 23, pp. 918–925, 2007. 121

[170] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inf. Theory*, vol. 38, pp. 719–746, 1992. 116

[171] O. Dietrich, S. Heiland, and K. Sartor, "Noise correction for the exact determination of apparent diffusion coefficients at low SNR," *Magn. Reson. Med.*, vol. 45, pp. 448–453, 2001. 16

[172] Z. Ding, J. C. Gore, and A. W. Anderson, "Reduction of noise in diffusion tensor images using anisotropic smoothing," *Magn. Reson. Med.*, vol. 53, pp. 485–490, 2005. 37, 88

[173] L. W. Dobrucki, E. D. de Muinck, J. R. Lindner, and A. J. Sinusas, "Approaches to multimodality imaging of angiogenesis," *J. Nucl. Med.*, vol. 51, pp. 66S–79S, 2010. 3

[174] I. Dologlou, D. van Ormondt, and G. Carayannis, "MRI scan time reduction through non-uniform sampling and SVD-based estimation," *Signal Process.*, vol. 55, pp. 207–219, 1996. 162

[175] M. Doneva, J. Sènègas, P. Börnert, H. Eggers, and A. Mertins, "Accelerated MR parameter mapping using compressed sensing with model-based sparsifying transform," in *Proc. Int. Soc. Magn. Reson. Med.*, 2009, p. 2812. 162

[176] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, pp. 613–627, 1995. 36

[177] ——, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006. 115

[178] ——, "For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," *Comm. Pure Appl. Math.*, vol. 59, pp. 907–934, 2006. 118

[179] ——, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math.*, vol. 59, pp. 797–829, 2006. 118

[180] ——, "Reflections on compressed sensing," *IEEE Info. Theory Soc. News.*, vol. 58, pp. 18–23, 2008. 5, 24

[181] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 2197–2202, 2003. 120, 150

[182] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, pp. 6–18, 2006. 120, 150, 151

[183] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2845–2862, 2001. 120

[184] D. L. Donoho and J. Tanner, "Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2002–2016, 2010. 121

[185] C. Dossal, G. Peyré, and J. Fadili, "A numerical exploration of compressed sampling recovery," *Linear Algebra Appl.*, vol. 432, pp. 1663–1679, 2010. 150

[186] A. Douiri, M. Schweiger, J. Riley, and S. R. Arridge, "Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information," *Meas. Sci. Techol.*, vol. 18, pp. 87–95, 2007. 34, 38, 39

[187] M. Doyle, E. G. Walsh, G. G. Blackwell, and G. M. Pohost, "Block regional interpolation scheme for k-space (BRISK): A rapid cardiac imaging technique," *Magn. Reson. Med.*, vol. 33, pp. 163–170, 1995. 159

[188] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sensing," *IEEE Signal Process. Mag.*, vol. 25, pp. 83–91, 2008. 124

[189] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Englewood Cliffs: Prentice-Hall, 1984. 18

[190] J. S. Duncan and N. Ayache, "Medical image analysis: Progress over two decades and the challenges ahead," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 22, pp. 85–106, 2000. 206

[191] A. Dutt and V. Rokhlin, "Fast Fourier transforms for nonequispaced data," *SIAM J. Sci. Comput.*, vol. 14, pp. 1368–1393, 1993. 19

[192] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211–218, 1936. 30

[193] A. Edelman, “Eigenvalues and condition numbers of random matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 9, pp. 543–560, 1988. 125, 145

[194] W. A. Edelstein, G. H. Glover, C. J. Hardy, and R. W. Redington, “The intrinsic signal-to-noise ratio in NMR imaging,” *Magn. Reson. Med.*, vol. 3, pp. 604–618, 1986. 65

[195] M. Elad, “Optimized projections for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 55, pp. 5695–5702, 2007. 150

[196] B. J. Erickson, A. Manduca, P. Palisson, K. R. Persons, F. Earnest, V. Savcenko, and N. J. Hangiandreou, “Wavelet compression of medical images,” *Radiology*, vol. 206, pp. 599–607, 1998. 116

[197] R. Eslami and M. Jacob, “Robust reconstruction of MRSI data using a sparse spectral model and high resolution MRI priors,” *IEEE Trans. Med. Imag.*, vol. 29, pp. 1297–1309, 2010. 34

[198] E. Eyal, D. Badikhi, E. Furman-Haran, F. Kelcz, K. J. Kirshenbaum, and H. Degani, “Principal component analysis of breast DCE-MRI adjusted with a model-based method,” *J. Magn. Reson. Imag.*, vol. 30, pp. 989–998, 2009. 162

[199] A. J. Fagan, G. R. Davies, J. M. S. Hutchison, F. P. Glasser, and D. J. Lurie, “Development of a 3-D, multi-nuclear continuous wave NMR imaging system,” *J. Magn. Reson.*, vol. 176, pp. 140–150, 2005. 14

[200] A. Fan, W. M. Wells, J. W. Fisher, M. Çetin, S. Haker, R. Mulkern, C. Tempany, and A. S. Willsky, “A unified variational approach to denoising and bias correction in MR,” *Lecture Notes Comp. Sci.*, vol. 2732, pp. 148–159, 2003. 36

[201] M. Fazel, H. Hindi, and S. Boyd, “Rank minimization and applications in system theory,” in *Proc. American Control Conf.*, 2004, pp. 3273–3278. 161, 168, 174

[202] M. Fazel, H. Hindi, and S. P. Boyd, “A rank minimization heuristic with application to minimum order system approximation,” in *Proc. American Control Conf.*, 2001, pp. 4734–4739. 169

[203] ——, “Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices,” in *Proc. American Control Conf.*, 2003, pp. 2156–2162. 169

[204] C. Feddern, J. Weickert, B. Burgeth, and M. Welk, “Curvature-driven PDE methods for matrix-valued images,” *Int. J. Comput. Vis.*, vol. 69, pp. 93–107, 2006. 37

[205] V. V. Federov, *Theory of Optimal Experiments*. New York: Academic Press, 1972. 97, 98, 100

[206] P. Feng and Y. Bresler, “Spectrum-blind minimum-rate sampling and reconstruction of multiband signals,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1996, pp. 1688–1691. 115, 159, 160

[207] J. A. Fessler, *Image reconstruction: Algorithms and analysis*, book draft. 19, 20, 22, 23, 47, 107, 118

[208] ——, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Process.*, vol. 5, pp. 493–506, 1996. 154

[209] ——, "Optimization transfer approach to joint registration/reconstruction for motion-compensated image reconstruction," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 596–599. 160

[210] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "Regularized emission image reconstruction using imperfect side information," *IEEE Trans. Nuc. Sci.*, vol. 39, pp. 1464–1471, 1992. 34, 38, 39

[211] J. A. Fessler, H. Erdoğan, and W. B. Wu, "Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise," *IEEE Trans. Image Process.*, vol. 9, pp. 1049–1055, 2000. 154

[212] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi, and D. C. Noll, "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity," *IEEE Trans. Signal Process.*, vol. 53, pp. 3393–3402, 2005. 13, 49

[213] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs," *IEEE Trans. Image Process.*, vol. 5, pp. 1346–1358, 1996. 52, 154

[214] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, pp. 560–574, 2003. 19

[215] J. A. Fessler, D. Yeo, and D. C. Noll, "Regularized fieldmap estimation in MRI," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2006, pp. 706–709. 36

[216] P. Fillard, X. Pennec, V. Arsigny, and N. Ayache, "Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics," *IEEE Trans. Med. Imag.*, vol. 26, pp. 1472–1482, 2007. 16, 37

[217] 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. 158

[218] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, pp. 1027–1036, 2002. 194, 196

[219] S. Foucart, "A note on guaranteed sparse recovery via ℓ_1-minimization," *Appl. Comput. Harmon. Anal.*, vol. 29, pp. 97–103, 2009. 120, 121

[220] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 395–407, 2009. 120

[221] J. Frahm, K.-D. Merboldt, and W. Hänicke, "Localized proton spectroscopy using stimulated echoes," *J. Magn. Reson.*, vol. 72, pp. 502–508, 1987. 14

[222] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Hum. Brain Mapp.*, vol. 2, pp. 56–78, 1994. 171

[223] R. E. Gabr, P. Aksit, P. A. Bottomley, A.-B. M. Youssef, and Y. M. Kadah, "Deconvolution-interpolation gridding (DING): Accurate reconstruction for arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 56, pp. 1182–1191, 2006. 21

[224] K. R. Gabriel and S. Zamir, "Lower rank approximation of matrices by least squares with any choice of weights," *Technometrics*, vol. 21, pp. 489–498, 1979. 174

[225] Y. Gal, A. J. H. Mehnert, A. P. Bradley, K. McMahon, D. Kennedy, and S. Crozier, "Denoising of dynamic contrast-enhanced MR images using dynamic nonlocal means," *IEEE Trans. Med. Imag.*, vol. 29, pp. 302–310, 2010. 37

[226] N. P. Galatsanos, A. K. Katsaggelos, R. T. Chin, and A. D. Hillery, "Least squares restoration of multichannel images," *IEEE Trans. Signal Process.*, vol. 39, pp. 2222–2236, 1991. 37, 38

[227] D. Gallichan, C. A. Cocosco, A. Dewdney, G. Schultz, A. Welz, J. Hennig, and M. Zaitsev, "Simultaneously driven linear and nonlinear spatial encoding fields in MRI," *Magn. Reson. Med.*, vol. 65, pp. 702–714, 2011. 14

[228] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, pp. 365–373, 2008. 116, 152, 159, 160, 189

[229] S. Gandy, B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Probl.*, vol. 27, p. 025010, 2011. 203

[230] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principle component analysis based four-dimensional computed tomography," *Preprint*, 2010, ftp://ftp.math.ucla.edu/pub/camreport/cam10-79.pdf. 160, 202

[231] H. Gao, H. Yu, and G. Wang, "True-color CT based on a prior rank, intensity and sparsity model (PRISM)," *Preprint*, 2011, ftp://ftp.math.ucla.edu/pub/camreport/cam11-01.pdf. 160, 202

[232] Q. Gao, J. P. Haldar, N. Rangawala, R. L. Magin, Z.-P. Liang, and X. J. Zhou, "Analysis of high b-value diffusion images using a fractional order diffusion model with denoising image reconstruction," in *Proc. Int. Soc. Magn. Reson. Med.*, 2009, p. 1418. 4, 84

[233] Y. Gao, S. M. Stakowski, S. J. Reeves, H. P. Hetherington, W.-J. Chu, and J.-H. Lee, "Fast spectroscopic imaging using online optimal sparse k-space acquisition and projections onto convex sets reconstruction," *Magn. Reson. Med.*, vol. 55, pp. 1265–1271, 2006. 34, 115

[234] A. N. Garroway, P. K. Grannell, and P. Mansfield, "Image formation in NMR by a selective irradiative process," *J. Phys. C: Solid State Phys.*, vol. 7, pp. L457–L462, 1974. 14

[235] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 14, pp. 367–383, 1992. 39, 43, 44, 46, 47

[236] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, pp. 932–946, 1995. 104

[237] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 6, pp. 721–741, 1984. 24, 38, 39, 40, 42, 43, 81

[238] J. J. Gerbrands, "On the relationships between SVD, KLT, and PCA," *Patt. Recog.*, vol. 14, pp. 375–381, 1981. 161

[239] G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Trans. Med. Imag.*, vol. 11, pp. 221–232, 1992. 37

[240] S.-M. Gho, Y. Nam, S.-Y. Zho, E. Y. Kim, and D.-H. Kim, "Three dimension double inversion recovery gray matter imaging using compressed sensing," *Magn. Reson. Imag.*, vol. 28, pp. 1395–1402, 2010. 116

[241] G. Gilbert, D. Simard, and G. Beaudoin, "Impact of an improved combination of signals from array coils in diffusion tensor imaging," *IEEE Trans. Med. Imag.*, vol. 26, pp. 1428–1436, 2007. 16

[242] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using registered anatomical images as priors," *Lecture Notes Comp. Sci.*, vol. 511, pp. 121–131, 1991. 34, 38, 39

[243] ——, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imag.*, vol. 12, pp. 670–680, 1993. 34, 38, 39

[244] M. E. Glidewell and K. T. Ng, "Anatomically constrained electrical impedance tomography for three-dimensional anisotropic bodies," *IEEE Trans. Med. Imag.*, vol. 16, pp. 572–580, 1997. 34

[245] G. H. Glover, "Phase-offset multiplanar (POMP) volume imaging: A new technique," *J. Magn. Reson. Imag.*, vol. 1, pp. 457–461, 1991. 14

[246] F. Godtliebsen, "Noise reduction using Markov random fields," *J. Magn. Reson.*, vol. 92, pp. 102–114, 1991. 36

[247] D. Goldfarb and S. Ma, "Convergence of fixed-point continuation algorithms for matrix rank minimization," *Found. Comput. Math.*, vol. 11, pp. 183–210, 2011. 169

[248] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed. London: The Johns Hopkins University Press, 1996. 21, 27, 110, 177, 179

[249] R. E. Gordon, P. E. Hanley, D. Shaw, D. G. Gadian, G. K. Radda, P. Styles, P. J. Bore, and L. Chan, "Localization of metabolites in animals using ^{31}P topical magnetic resonance," *Nature*, vol. 287, pp. 736–738, 1980. 12

[250] S. Goud, Y. Hu, E. Di Bella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imag.*, 2011, http://dx.doi.org/10.1109/TMI.2010.2100850. 159, 160, 169, 202

[251] S. Goud, Y. Hu, and M. Jacob, "Real-time cardiac MRI using low-rank and sparsity penalties," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 988–991. 159, 160, 202

[252] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast Fourier transform," *SIAM Rev.*, vol. 46, pp. 443–454, 2004. 19

[253] A. Greiser and M. von Kienlin, "Efficient k-space sampling by density-weighted phase-encoding," *Magn. Reson. Med.*, vol. 50, pp. 1266–1275, 2003. 92

[254] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, pp. 3320–3325, 2003. 120

[255] K. M. Grigoriadis and E. B. Beran, "Alternating projection algorithms for linear matrix inequalities problems with rank constraints," in *Advances in Linear Matrix Inequality Methods in Control*, L. El Ghaoui and S.-I. Niculescu, Eds. SIAM, 2000, pp. 251–267. 174

[256] 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. 14

[257] M. A. Griswold, P. M. Jakob, M. Nittka, J. W. Goldfarb, and A. Haase, "Partially parallel imaging with localized sensitivities (PILS)," *Magn. Reson. Med.*, vol. 44, pp. 602–609, 2000. 14

[258] B. Grung and R. Manne, "Missing values in principal component analysis," *Chemometr. Intell. Lab. Syst.*, vol. 42, pp. 125–139, 1998. 174

[259] P. D. Grünwald, *The minimum description length principle*. Cambridge: MIT Press, 2007. 194, 195, 196

[260] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, pp. 910–914, 1995. 16

[261] M. Guven, B. Yazici, X. Intes, and B. Chance, "Diffuse optical tomography with a priori anatomical information," *Phys. Med. Biol.*, vol. 50, pp. 2837–2858, 2005. 34

[262] E. M. Haacke, "Solving for non-ideal conditions in two-dimensional Fourier transform magnetic resonance imaging using a generalized inverse transform," *Inverse Probl.*, vol. 3, pp. 421–435, 1987. 21, 34, 89

[263] E. M. Haacke, R. F. Brown, M. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. New York: J. Wiley & Sons, 1999. 7

[264] E. M. Haacke, Z.-P. Liang, and S. H. Izen, "Superresolution reconstruction through object modeling and parameter estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 592–595, 1989. 116

[265] A. Haase, J. Frahm, D. Matthaei, W. Hänicke, and K.-D. Merboldt, "FLASH imaging. rapid NMR imaging using low flip-angle pulses," *J. Magn. Reson.*, vol. 67, pp. 258–266, 1986. 13

[266] K. R. Hahn, S. Prigarin, and K. Hasan, "A novel denoising technique for very noisy diffusion tensor imaging data," in *Proc. Int. Soc. Magn. Reson. Med.*, 2004, p. 1208. 37

[267] J. P. Haldar, Q. Gao, X. J. Zhou, and Z.-P. Liang, "Optimized measurement of anomalous diffusion," in *Proc. Int. Soc. Magn. Reson. Med.*, 2009, p. 3570. 4, 84

[268] 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. 5, 158, 174, 175, 182

[269] J. P. Haldar, D. Hernando, M. D. Budde, Q. Wang, S.-K. Song, and Z.-P. Liang, "High-resolution MR metabolic imaging," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2007, pp. 4324–4326. 4, 84

[270] ——, "High-resolution spectroscopic imaging with statistical reconstruction," in *Proc. Int. Soc. Magn. Reson. Med.*, 2007, p. 1231. 4, 84

[271] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Shaping spatial response functions for optimal estimation of compartmental signals from limited Fourier data," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2007, pp. 1364–1367. 34

[272] ——, "Super-resolution reconstruction of MR image sequences with contrast modeling," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 266–269. 202

[273] ——, "Compressed-sensing MRI with random encoding," *IEEE Trans. Med. Imag.*, 2010, http://dx.doi.org/10.1109/TMI.2010.2085084. 5, 115

[274] J. P. Haldar, D. Hernando, S.-K. Song, and Z.-P. Liang, "Anatomically constrained reconstruction from noisy data," *Magn. Reson. Med.*, vol. 59, pp. 810–818, 2008. 4, 38, 50, 84

[275] J. P. Haldar, D. Hernando, B. P. Sutton, and Z.-P. Liang, "Data acquisition considerations for compressed sensing in MRI," in *Proc. Int. Soc. Magn. Reson. Med.*, 2007, p. 829. 5, 124, 152

[276] J. P. Haldar, M. Jacob, A. Ebel, X. Zhu, N. Schuff, D. Hernando, B. Sutton, and Z.-P. Liang, "Constrained spectroscopic imaging with hard and soft anatomical boundary constraints," in *Proc. Int. Soc. Magn. Reson. Med.*, 2006, p. 3077. 4, 84

[277] ——, "Regularized inversion of noisy, incomplete MR spectroscopic imaging data with anatomical prior," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2006, pp. 718–721. 4, 84

[278] J. P. Haldar, J. H. Kim, S.-K. Song, and Z.-P. Liang, "Accelerated mouse spinal cord diffusion measurements with SNR-enhancing joint reconstruction," in *Proc. Int. Soc. Magn. Reson. Med.*, 2011. 4, 84

[279] J. P. Haldar and Z.-P. Liang, "High-resolution diffusion MRI," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2007, pp. 311–314. 4, 84

[280] ——, "Joint reconstruction of noisy high-resolution MR image sequences," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2008, pp. 752–755. 4, 38, 84

[281] ——, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 716–719. 5, 158, 159, 160, 165

[282] ——, "Low-rank approximations for dynamic imaging," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2011, pp. 1052–1055. 5, 158, 159, 160

[283] ——, "On MR experiment design with quadratic regularization," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2011, pp. 1676–1679. 4, 38, 66

[284] J. P. Haldar, K. Sakaie, and Z.-P. Liang, "Resolution and noise properties of linear phase-constrained partial Fourier reconstruction," in *Proc. Int. Soc. Magn. Reson. Med.*, 2009, p. 2862. 4, 84, 89

[285] J. P. Haldar, V. J. Wedeen, M. Nezamzadeh, G. Dai, N. Schuff, and Z.-P. Liang, "Improved SNR in diffusion spectrum imaging with statistical reconstruction," in *Proc. Int. Soc. Magn. Reson. Med.*, 2008, p. 141. 4, 84

[286] J. P. Haldar, T.-H. Wu, Q. Wang, C.-I. Chen, S.-K. Song, and Z.-P. Liang, "Further development in anatomically constrained MR image reconstruction: Application to multimodal imaging of mouse stroke," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2008, pp. 422–425. 4, 84

[287] G. Hamarneh and J. Hradsky, "Bilateral filtering of diffusion tensor magnetic resonance images," *IEEE Trans. Image Process.*, vol. 16, pp. 2463–2475, 2007. 37

[288] L. K. Hansen, J. Larsen, F. Å. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O. B. Paulson, "Generalizable patterns in neuroimaging: How many principal components?" *NeuroImage*, vol. 9, pp. 534–544, 1999. 194

[289] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*. Philadelphia: SIAM, 2010. 17, 20, 22, 118

[290] L. G. Hanson, "Is quantum mechanics necessary for understanding magnetic resonance imaging?" *Concepts Magn. Reson. A*, vol. 32A, pp. 329–340, 2008. 7

[291] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1996, pp. 1331–1334. 24, 118

[292] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, pp. 51–83, 1978. 68

[293] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970. 174

[294] R. Hartley and F. Schaffalitzky, "PowerFactorization: 3D reconstruction with missing or uncertain data," in *Australia-Japan Advanced Workshop on Computer Vision*, 2003. 174, 175, 176, 177

[295] L. He and I. R. Greenshields, "A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images," *IEEE Trans. Med. Imag.*, vol. 28, pp. 165–172, 2009. 16, 36

[296] X. He, L. Cheng, J. A. Fessler, and E. C. Frey, "Regularized image reconstruction algorithms for dual-isotope myocardial perfusion SPECT (MPS) imaging using a cross-tracer prior," *IEEE Trans. Med. Imag.*, 2010, http://dx.doi.org/10.1109/TMI.2010.2087031. 37, 38, 39, 40

[297] D. M. Healy and J. B. Weaver, "Two applications of wavelet transforms in magnetic resonance imaging," *IEEE Trans. Inf. Theory*, vol. 38, pp. 840–860, 1992. 14

[298] R. M. Heidemann, M. A. Griswold, A. Haase, and P. M. Jakob, "VD-AUTO-SMASH imaging," *Magn. Reson. Med.*, vol. 45, pp. 1066–1074, 2001. 14

[299] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Med. Phys.*, vol. 12, pp. 232–233, 1985. 16

[300] J. Hennig and M. Hodapp, "Burst imaging," *Mag. Reson. Mater. Phys. Bio. Med.*, vol. 1, pp. 39–48, 1993. 13

[301] J. Hennig, A. M. Welz, G. Schultz, J. Korvink, Z. Liu, O. Speck, and M. Zaitsev, "Parallel imaging in non-bijective, curvilinear magnetic field gradients: a concept study," *Magn. Reson. Mater. Phy.*, vol. 21, pp. 5–14, 2008. 14

[302] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 342–349, 2010. 136, 139

[303] D. Hernando, P. Kellman, J. P. Haldar, and Z.-P. Liang, "Robust water/fat separation in the presence of large field inhomogeneities using a graph cut algorithm," *Magn. Reson. Med.*, vol. 63, pp. 79–90, 2010. 36

[304] A. O. Hero, R. Piramuthu, J. A. Fessler, and S. R. Titus, "Minimax emission computed tomography using high-resolution anatomical side information and B-spline models," *IEEE Trans. Inf. Theory*, vol. 45, pp. 920–938, 1999. 34, 38, 39

[305] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Natl. Bur. Stand.*, vol. 49, pp. 409–436, 1952. 48, 180

[306] W. S. Hinshaw, "Image formation by nuclear magnetic resonance: The sensitive-point method," *J. Appl. Phys.*, vol. 47, pp. 3709–3721, 1976. 12

[307] W. S. Hoge, M. E. Kilmer, C. Zacarìas-Almarcha, and D. H. Brooks, "Fast regularized reconstruction of non-uniformly subsampled partial-Fourier parallel MRI data," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2007, pp. 1012–1015. 21, 34, 89

[308] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985. 27

[309] D. Hoult, "The origins and present status of the radio wave controversy in NMR," *Concepts Magn. Reson. A*, vol. 34A, pp. 193–216, 2009. 7

[310] D. I. Hoult, "Zeugmatography: A criticism of the concept of a selective pulse in the presence of a field gradient," *J. Magn. Reson.*, vol. 26, pp. 165–167, 1977. 14

[311] ——, "Rotating frame zeugmatography," *J. Magn. Reson.*, vol. 33, pp. 183–197, 1979. 14, 15

[312] ——, "The principle of reciprocity in signal strength calculations – a mathematical guide," *Concepts Magn. Reson.*, vol. 12, pp. 173–187, 2000. 10

[313] D. I. Hoult and R. E. Richards, "The signal-to-noise ratio of the nuclear magnetic resonance experiment," *J. Magn. Reson.*, vol. 24, pp. 71–85, 1976. 15

[314] W. V. Houston, "A compound interferometer for fine structure work," *Phys. Rev.*, vol. 29, pp. 478–484, 1927. 67

[315] D. C. Hoyle, "Automatic PCA dimension selection for high dimensional data and small sample sizes," *J. Mach. Learn. Res.*, vol. 9, pp. 2733–2759, 2008. 194

[316] C.-H. L. Hsu and R. M. Leahy, "PET image reconstruction incorporating anatomical information using segmented regression," *Proc. SPIE*, vol. 3034, pp. 381–392, 1997. 34

[317] S. Hu, M. Lustig, A. P. Chen, J. Crane, A. Kerr, D. A. C. Kelley, R. Hurd, J. Kurhanewicz, S. J. Nelson, J. M. Pauly, and D. B. Vigneron, "Compressed sensing for resolution enhancement of hyperpolarized ^{13}C flyback 3D-MRSI," *J. Magn. Reson.*, vol. 192, pp. 258–264, 2008. 116, 152

[318] X. Hu, V. Johnson, W. H. Wong, and C.-T. Chen, "Bayesian image processing in magnetic resonance imaging," *Magn. Reson. Imag.*, vol. 9, pp. 611–620, 1991. 36

[319] X. Hu, D. N. Levin, P. C. Lauterbur, and T. Spraggins, "SLIM: Spectral localization by imaging," *Magn. Reson. Med.*, vol. 8, pp. 314–322, 1988. 21, 34, 59

[320] X. Hu and T. Parrish, "Reduction of field of view for dynamic imaging," *Magn. Reson. Med.*, vol. 31, pp. 691–694, 1994. 115

[321] X. Hu and Z. Wu, "SLIM revisited," *IEEE Trans. Med. Imag.*, vol. 12, pp. 583–587, 1993. 55, 59

[322] P. J. Huber, *Robust Statistics*. New York: John Wiley & Sons, 1981. 74

[323] J. W. Hugg, A. A. Maudsley, M. W. Weiner, and G. B. Matson, "Comparison of k-space sampling schemes for multidimensional MR spectroscopic imaging," *Magn. Reson. Med.*, vol. 36, pp. 469–473, 1996. 92

[324] B. R. Hunt and O. Kübler, "Karhunen-Loeve multispectral image restoration, part I: Theory," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 592–600, 1984. 37

[325] M. A. Hurn, K. V. Mardia, T. J. Hainsworth, J. Kirkbride, and E. Berry, "Bayesian fused classification of medical images," *IEEE Trans. Med. Imag.*, vol. 15, pp. 850–858, 1996. 34

[326] M. Hutchinson and U. Raff, "Fast MRI data acquisition using multiple detectors," *Magn. Reson. Med.*, vol. 6, pp. 87–91, 1988. 14

[327] J. S. Hyde, A. Jesmanowicz, W. Froncisz, J. B. Kneeland, T. M. Grist, and N. F. Campagna, "Parallel image acquisition from noninteracting local coils," *J. Magn. Reson.*, vol. 70, pp. 512–517, 1986. 14

[328] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, pp. 1001–1009, 2001. 40, 43, 44, 78, 104

[329] D. Idiyatullin, C. Corum, J.-Y. Park, and M. Garwood, "Fast and quiet MRI using a swept radiofrequency," *J. Magn. Reson.*, vol. 181, pp. 342–349, 2006. 14

[330] X. Intes, C. Maloux, M. Guven, B. Yazici, and B. Chance, "Diffuse optical tomography with physiological and spatial a priori constraints," *Phys. Med. Biol.*, vol. 49, pp. N155–N163, 2004. 34

[331] S. Ito, S. Kawaharada, Y. Kamimura, and Y. Yamada, “Fast reconstruction of three-dimensional images in Fresnel transform imaging technique using optical imaging system,” *Sys. Comp. Japan*, vol. 33, pp. 21–29, 2002. 14

[332] S. Ito and Y. Yamada, “Alias-free image reconstruction using Fresnel transform in the phase-scrambling Fourier imaging technique,” *Magn. Reson. Med.*, vol. 60, pp. 422–430, 2008. 14

[333] J. I. Jackson, C. H. Meyer, D. G. Nishimura, and A. Macovski, “Selection of a convolution function for Fourier inversion using gridding,” *IEEE Trans. Med. Imag.*, vol. 10, pp. 473–478, 1991. 19

[334] M. Jacob, “Optimized least-square nonuniform fast Fourier transform,” *IEEE Trans. Signal Process.*, vol. 57, pp. 2165–2177, 2009. 19

[335] M. Jacob, X. Zhu, A. Ebel, N. Schuff, and Z.-P. Liang, “Improved model-based magnetic resonance spectroscopic imaging,” *IEEE Trans. Med. Imag.*, vol. 26, pp. 1305–1318, 2007. 34

[336] L. Jacques, “A short note on compressed sensing with partially known signal support,” *Signal Process.*, vol. 90, pp. 3308–3312, 2010. 116

[337] P. M. Jakob, M. A. Griswold, R. R. Edelman, and D. K. Sodickson, “AUTO-SMASH: A self-calibrating technique for SMASH imaging,” *Mag. Reson. Mater. Phys. Bio. Med.*, vol. 7, pp. 42–54, 1998. 14

[338] J. Jansen and B. Blümich, “Stochastic spectroscopic imaging,” *J. Magn. Reson.*, vol. 99, pp. 525–532, 1992. 14

[339] V. Jaravine, I. Ibraghimov, and V. Y. Orekhov, “Removal of a time barrier for high-resolution multidimensional NMR spectroscopy,” *Nat. Methods*, vol. 3, pp. 605–607, 2006. 203

[340] V. A. Jaravine, A. V. Zhuravleva, P. Permi, I. Ibraghimov, and V. Y. Orekhov, “Hyperdimensional NMR spectroscopy with nonlinear sampling,” *J. Am. Chem. Soc.*, vol. 130, pp. 3927–3936, 2008. 203

[341] J. Ji and Z.-P. Liang, “High resolution cardiac magnetic resonance imaging: A model-based approach,” in *Proc. IEEE Eng. Med. Bio. Conf.*, 2001, pp. 2268–2271. 159, 160

[342] G. Johnson, E. X. Wu, and S. K. Hilal, “Optimized phase scrambling for RF phase encoding,” *J. Magn. Reson. B*, vol. 103, pp. 59–63, 1994. 14, 124

[343] J. B. Johnson, “Thermal agitation of electricity in conductors,” *Phys. Rev.*, vol. 32, pp. 97–109, 1928. 15

[344] V. E. Johnson, “A framework for incorporating structural prior information into the estimation of medical images,” *Lecture Notes Comp. Sci.*, vol. 687, pp. 307–321, 1993. 34

[345] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002. 161

[346] A. W. Jones, J. Bland-Hawthorn, and P. L. Shopbell, "Towards a general definition for spectroscopic resolution," in *Astronomical Data Analysis Software and Systems IV*, ser. ASP Conference Series, R. A. Shaw, H. E. Payne, and J. J. E. Hayes, Eds., vol. 77, 1995. 65, 68

[347] A. Juditsky and A. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization," *Math. Program., Ser. B*, vol. 127, pp. 57–88, 2011. 150

[348] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med.*, vol. 61, pp. 103–116, 2009. 159, 160

[349] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques," *Int. J. Imag. Syst. Tech.*, vol. 20, pp. 81–98, 2010. 159, 160

[350] 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. 116, 152, 159, 160, 162

[351] Y. M. Kadah and X. Hu, "Pseudo-Fourier imaging (PFI): A technique for spatial encoding in MRI," *IEEE Trans. Med. Imag.*, vol. 16, pp. 893–902, 1997. 14

[352] J. Kaipio and E. Somersalo, "Statistical inverse problems: Discretization, model reduction and inverse crimes," *J. Comput. Appl. Math.*, vol. 198, pp. 493–504, 2007. 155

[353] Y. Katznelson, *An Introduction to Harmonic Analysis*, 2nd ed. New York: Dover Publications, Inc., 1976. 18

[354] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River: Prentice Hall, 1993. 22, 23, 24, 196

[355] K. Kazimierczuk, J. Stanek, A. Zawadzka-Kazimierczuk, and W. Koźmiński, "Random sampling in multidimensional NMR spectroscopy," *Prog. NMR Spect.*, vol. 57, pp. 420–434, 2010. 121

[356] S. F. Keevil, "Spatial localization in nuclear magnetic resonance spectroscopy," *Phys. Med. Biol.*, vol. 51, pp. R579–R636, 2006. 12

[357] P. Kellman, F. H. Epstein, and E. R. McVeigh, "Adaptive sensitivity encoding incorporating temporal filtering (TSENSE)," *Magn. Reson. Med.*, vol. 45, pp. 846–852, 2001. 159, 160

[358] J. R. Kelton, R. L. Magin, and S. M. Wright, "An algorithm for rapid image acquisition using multiple receiver coils," in *Proc. Int. Soc. Magn. Reson. Med.*, 1989, p. 1172. 14

[359] R. H. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2980–2998, 2010. 161, 170, 173, 174

[360] ——, “Matrix completion from noisy entries,” *J. Mach. Learn. Res.*, vol. 11, pp. 2057–2078, 2010. 161, 170, 173, 174

[361] Y. Kim, P. M. Thompson, and L. A. Vese, “HARDI data denoising using vectorial total variation and logarithmic barrier,” *Inverse Probl. Imaging*, vol. 4, pp. 273–310, 2010. 37, 38, 39, 40

[362] Y.-C. Kim, S. S. Narayanan, and K. S. Nayak, “Accelerated three-dimensional upper airway MRI using compressed sensing,” *Magn. Reson. Med.*, vol. 61, pp. 1434–1440, 2009. 116, 131, 152

[363] J. B. Kneeland and J. S. Hyde, “High-resolution MR imaging with local coils,” *Radiology*, vol. 17, pp. 1–7, 1989. 14

[364] T. Knopp, S. Kunis, and D. Potts, “A note on the iterative MRI reconstruction from nonuniform k-space data,” *Int. J. Biomed. Imaging*, vol. 2007, p. 24727, 2007. 22

[365] C. G. Koay, L.-C. Chang, J. D. Carew, C. Pierpaoli, and P. J. Basser, “A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging,” *J. Magn. Reson.*, vol. 182, pp. 115–125, 2006. 16

[366] C. G. Koay, E. Özarslan, and P. J. Basser, “A signal transformational framework for breaking the noise floor and its applications in MRI,” *J. Magn. Reson.*, vol. 197, pp. 108–119, 2009. 16

[367] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Rev.*, vol. 51, pp. 455–500, 2009. 161, 174

[368] K. Konstantinides and K. Yao, “Statistical analysis of effective singular values in matrix rank determination,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 757–763, 1988. 194

[369] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, pp. 30–37, 2009. 161

[370] J. Kornak and K. Young, “K-Bayes reconstruction for perfusion MRI II: Modeling and technical development,” *J. Digital Imaging*, vol. 23, pp. 374–385, 2010. 34, 38, 39

[371] J. Kornak, K. Young, N. Schuff, A. Du, A. A. Maudsley, and M. W. Weiner, “K-Bayes reconstruction for perfusion MRI I: Concepts and application,” *J. Digital Imaging*, vol. 23, pp. 277–286, 2010. 34

[372] J. Kornak, K. Young, B. J. Soher, and A. A. Maudsley, “Bayesian k-space-time reconstruction of MR spectroscopic imaging for enhanced resolution,” *IEEE Trans. Med. Imag.*, vol. 29, pp. 1333–1350, 2010. 34, 36, 38, 39

[373] F. R. Korosec, R. Frayne, T. M. Grist, and C. A. Mistretta, “Time-resolved contrast-enhanced 3D MR angiography,” *Magn. Reson. Med.*, vol. 36, pp. 345–351, 1996. 158, 159

[374] K. Krissian and S. Aja-Fernández, "Noise-driven anisotropic diffusion filtering of MRI," *IEEE Trans. Image Process.*, vol. 18, pp. 2265–2274, 2009. 16, 36

[375] S. Kulkarni, P. Khurd, I. Hsiao, L. Zhou, and G. Gindi, "A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors," *Phys. Med. Biol.*, vol. 52, pp. 3601–3617, 2007. 34, 38, 39

[376] A. Kumar, D. Welti, and R. R. Ernst, "NMR Fourier zeugmatography," *J. Magn. Reson.*, vol. 18, pp. 69–83, 1975. 12

[377] S. Kunis and D. Potts, "Time and memory requirements of the nonequispaced FFT," *Samp. Theory Signal Image Process.*, vol. 7, pp. 77–100, 2008. 19

[378] L.-W. Kuo, J. P. Haldar, Y.-C. Lo, C.-L. Liu, Z.-P. Liang, and W.-Y. I. Tseng, "Quantitative improvement of diffusion spectrum imaging tractography using statistical denoising," in *Proc. Int. Soc. Magn. Reson. Med.*, 2010, p. 1669. 4, 84

[379] D. Kwiat, S. Einav, and G. Navon, "A decoupled coil detector array for fast image acquisition in magnetic-resonance-imaging," *Med. Phys.*, vol. 18, pp. 251–265, 1991. 14

[380] W. E. Kyriakos, L. P. Panych, D. F. Kacher, C.-F. Westin, S. M. Bao, R. V. Mulkern, and F. A. Jolesz, "Sensitivity profiles from an array of coils for encoding and reconstruction in parallel (SPACE-RIP)," *Magn. Reson. Med.*, vol. 44, pp. 301–308, 2000. 14

[381] P. Lahiri, Ed., *Model Selection*. Beachwood: Institute of Mathematical Statistics, 2001. 169, 194, 195

[382] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, 1967. 18

[383] A. D. Lanterman, "Schwarz, Wallace, and Rissanen: Intertwining themes in theories of model selection," *Int. Stat. Rev.*, vol. 69, pp. 185–212, 2001. 194, 195

[384] M. Lassas and S. Siltanen, "Can one use total variation prior for edge-preserving Bayesian inversion?" *Inverse Probl.*, vol. 20, pp. 1537–1563, 2004. 155

[385] R. S. Laugesen, *Harmonic Analysis Lecture Notes*, University of Illinois at Urbana-Champaign, 2009, http://www.math.uiuc.edu/~laugesen/545/545Lectures.pdf. 18

[386] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, pp. 190–191, 1973. 3, 12

[387] P. C. Lauterbur, D. M. Kramer, W. V. House Jr., and C.-N. Chen, "Zeugmatographic high resolution nuclear magnetic resonance spectroscopy: Images of chemical inhomogeneity within macroscopic objects," *J. Am. Chem. Soc.*, vol. 97, pp. 6866–6868, 1975. 14

[388] D. Le Bihan, "Looking into the functional architecture of the brain with diffusion MRI," *Nat. Rev. Neurosci.*, vol. 4, pp. 469–480, 2003. 84, 86

[389] R. Leahy and X. Yan, "Incorporation of anatomical MR data for improved functional imaging with PET," *Lecture Notes Comp. Sci.*, vol. 511, pp. 105–120, 1991. 34, 38, 39

[390] K. Lee and Y. Bresler, "Computing performance guarantees for compressed sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 5129–5132. 150

[391] ——, "Guaranteed minimum rank approximation from linear observations by nuclear norm minimization with an ellipsoidal constraint," *Preprint*, 2009, http://arxiv.org/abs/0903.4742. 161, 170, 182

[392] ——, "ADMiRA: Atomic decomposition for minimum rank approximation," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4402–4416, 2010. 161, 170

[393] T. Lei, H. K. Song, F. W. Wehrli, and T. P. L. Roberts, "Statistical properties of spin noise in MRI," *Proc. SPIE*, vol. 6913, p. 69130H, 2008. 16

[394] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd ed. Chichester: John Wiley & Sons Ltd, 2008. 7

[395] A. Li, G. Boverman, Y. Zhang, D. Brooks, E. L. Miller, M. E. Kilmer, Q. Zhang, E. M. C. Hillman, and D. A. Boas, "Optimal linear inverse solution with multiple priors in diffuse optical tomography," *Applied Optics*, vol. 44, pp. 1948–1956, 2005. 34

[396] B. S. Y. Li, J. Regal, and O. Gonen, "SNR versus resolution in 3D ^{1}H MRS of the human brain at high magnetic fields," *Magn. Reson. Med.*, vol. 46, pp. 1049–1053, 2001. 66

[397] D. Liang, B. Liu, J. J. Wang, and L. Ying, "Accelerating SENSE using compressed sensing," *Magn. Reson. Med.*, vol. 62, pp. 1574–1584, 2009. 116, 152

[398] D. Liang, G. Xu, H. Wang, K. F. King, D. Xu, and L. Ying, "Toeplitz random encoding MR imaging using compressed sensing," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 270–273. 124

[399] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2007, pp. 988–991. 5, 159, 160, 161, 162, 165, 167, 171, 203

[400] Z.-P. Liang, F. Boada, 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. 17, 21, 34, 89

[401] Z.-P. Liang, E. M. Haacke, and C. W. Thomas, "High-resolution inversion of finite Fourier transform data through a localised polynomial approximation," *Inverse Probl.*, vol. 5, pp. 831–847, 1989. 116

[402] Z.-P. Liang, H. Jiang, C. P. Hess, and P. C. Lauterbur, "Dynamic imaging by model estimation," *Int. J. Imag. Syst. Tech.*, vol. 8, pp. 551–557, 1997. 159, 160, 161

[403] Z.-P. Liang and P. C. Lauterbur, "A generalized series approach to MR spectroscopic imaging," *IEEE Trans. Med. Imag.*, vol. 10, pp. 132–137, 1991. 21, 34

[404] ——, "A theoretical analysis of the SLIM technique," *J. Magn. Reson. B*, vol. 102, pp. 54–60, 1993. 55, 59

[405] ——, "An efficient method for dynamic magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 13, pp. 677–686, 1994. 21, 34, 159, 160

[406] ——, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. New York: IEEE Press, 2000. 3, 7

[407] J. M. Libove and J. R. Singer, "Resolution and signal-to-noise relationships in NMR imaging in the human body," *J. Phys. E: Sci. Instrum.*, vol. 13, pp. 38–44, 1980. 15

[408] F.-H. Lin, L. L. Wald, S. P. Ahlfors, M. S. Hämäläinen, K. K. Kwong, and J. W. Belliveau, "Dynamic magnetic resonance inverse imaging of human brain function," *Magn. Reson. Med.*, vol. 56, pp. 787–802, 2006. 14

[409] Y. Lin, W. C. Barber, J. S. Iwanczyk, W. Roeck, O. Nalcioglu, and G. Gulsen, "Quantitative fluorescence tomography using a combined tri-modality FT/DOT/XCT system," *Opt. Express*, vol. 18, pp. 7835–7850, 2010. 34

[410] B. Lipinski, H. Herzog, E. Rota Kops, W. Oberschelp, and H. W. Müller-Gärtner, "Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 129–136, 1997. 34, 38, 39

[411] C. Liu, R. Bammer, and M. E. Moseley, "Parallel imaging reconstructions for arbitrary trajectories using k-space sparse matrices (kSPA)," *Magn. Reson. Med.*, vol. 58, pp. 1171–1181, 2007. 14

[412] J. Liu and J. L. Koenig, "An automatic phase correction method in nuclear magnetic resonance imaging," *J. Magn. Reson.*, vol. 86, pp. 593–604, 1990. 16

[413] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, p. 2298. 203

[414] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM J. Matrix Anal. Appl.*, vol. 31, pp. 1235–1256, 2009. 169

[415] G. Lohmann, S. Bohn, K. Müller, R. Trampel, and R. Turner, "Image restoration and spatial resolution in 7-Tesla magnetic resonance imaging," *Magn. Reson. Med.*, vol. 64, pp. 15–22, 2010. 16, 36

[416] H. H.-S. Lu, C.-M. Chen, and I.-H. Yang, "Cross-reference weighted least square estimates for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 17, pp. 1–8, 1998. 34

[417] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: John Wiley & Sons, 1969. 19, 20, 21

[418] F. Luisier and T. Blu, "SURE-LET multichannel image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 17, pp. 482–492, 2008. 37

[419] R. Lukac, B. Smolka, K. Martin, K. N. Plataniotis, and A. N. Venetsanopoulos, "Vector filtering for color imaging," *IEEE Signal Process. Mag.*, vol. 22, pp. 74–86, 2005. 37

[420] 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. 24, 44, 116, 121, 128, 129, 131, 150, 152, 159, 160

[421] 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. Soc. Magn. Reson. Med.*, 2006, p. 2420. 116, 152, 159, 160

[422] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Trans. Image Process.*, vol. 12, pp. 1579–1590, 2003. 36

[423] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program., Ser. A*, 2010, http://dx.doi.org/10.1007/s10107-009-0306-5. 169

[424] A. Macovski, "Volumetric NMR imaging with time-varying gradients," *Magn. Reson. Med.*, vol. 2, pp. 29–40, 1985. 19

[425] ——, "Noise in MRI," *Magn. Reson. Med.*, vol. 36, pp. 494–497, 1996. 15

[426] 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. 115, 159, 160

[427] A. Majumdar and R. K. Ward, "An algorithm for sparse MRI reconstruction by Schatten p-norm minimization," *Magn. Reson. Imag.*, vol. 29, pp. 408–417, 2011. 162, 169

[428] M. Makowski, C. Jansen, I. Webb, A. Chiribiri, E. Nagel, R. Botnar, S. Kozerke, and S. Plein, "First-pass contrast-enhanced myocardial perfusion MRI in mice on a 3-T clinical MR scanner," *Magn. Reson. Med.*, vol. 64, pp. 1592–1598, 2010. 159, 160

[429] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic Press, 2008. 116, 122

[430] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. Garcia-Martí, L. Martí-Bonmatí, and M. Robles, "MRI denoising using non-local means," *Med. Image Anal.*, vol. 12, pp. 514–523, 2008. 36

[431] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J. Magn. Reson. Imag.*, vol. 31, pp. 192–203, 2010. 16, 36

[432] J. V. Manjón, N. A. Thacker, J. J. Lull, G. Garcia-Martí, L. Martí-Bonmatí, and M. Robles, "Multicomponent MR image denoising," *Int. J. Biomed. Imaging*, vol. 2009, p. 756897, 2009. 37

[433] P. Mansfield, "Multiplanar image formation using NMR spin echoes," *J. Phys. C: Solid State Phys.*, vol. 10, pp. L55–L58, 1977. 13

[434] P. Mansfield, A. A. Maudsley, and T. Baines, "Fast scan proton density imaging by NMR," *J. Phys. E: Sci. Instrum.*, vol. 9, pp. 271–278, 1976. 14

[435] R. F. Marcia, Z. T. Harmany, and R. M. Willett, "Compressive coded aperture imaging," *Proc. SPIE*, vol. 7246, p. 72460G, 2009. 124

[436] T. H. Mareci and H. R. Brooker, "Essential considerations for spectral localization using indirect gradient encoding of spatial information," *J. Magn. Reson.*, vol. 92, pp. 229–246, 1991. 92, 101

[437] R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*. New York: Springer-Verlag, 1991. 18

[438] J. Marot, C. Fossati, and S. Bourennane, "About advances in tensor data denoising methods," *EURASIP J. Adv. Signal Process.*, vol. 2008, p. 235357, 2008. 37

[439] G. J. Marseille, M. Fuderer, R. de Beer, A. F. Mehlkopf, and D. van Ormondt, "Reduction of MRI scan time through nonuniform sampling and edge-distribution modeling," *J. Magn. Reson. B*, vol. 103, pp. 292–295, 1994. 116, 121

[440] M. Martin-Fernandez, C. Alberola-López, J. Ruiz-Alzola, and C.-F. Westin, "Sequential anisotropic Wiener filtering applied to 3D MRI data," *Magn. Reson. Imag.*, vol. 25, pp. 278–292, 2007. 36

[441] M. Martin-Fernandez, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, and C. Alberola-López, "Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data," *Med. Image Anal.*, vol. 13, pp. 19–35, 2009. 16, 37

[442] T. F. Massoud and S. S. Gambhir, "Molecular imaging in living subjects: seeing fundamental biological processes in a new light," *Gene Dev.*, vol. 17, pp. 545–580, 2003. 3

[443] A. A. Maudsley, "Multiple-line-scanning spin density imaging," *J. Magn. Reson.*, vol. 41, pp. 112–126, 1980. 14

[444] ——, "Fourier imaging using rf phase encoding," *Magn. Reson. Med.*, vol. 3, pp. 768–777, 1986. 14

[445] ——, "Dynamic range improvement in NMR imaging using phase scrambling," *J. Magn. Reson.*, vol. 76, pp. 287–305, 1988. 14

[446] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral regularization algorithms for learning large incomplete matrices," *J. Mach. Learn. Res.*, vol. 11, pp. 2287–2322, 2010. 169

[447] G. McGibney and M. R. Smith, "An unbiased signal-to-noise ratio measure for magnetic resonance images," *Med. Phys.*, vol. 20, pp. 1077–1078, 1993. 16

[448] T. McGraw, B. Vemuri, E. Özarslan, Y. Chen, and T. Mareci, "Variational denoising of diffusion weighted MRI," *Inverse Probl. Imaging*, vol. 3, pp. 625–648, 2009. 37

[449] T. McGraw, B. C. Vemuri, Y. Chen, M. Rao, and T. Mareci, "DT-MRI denoising and neuronal fiber tracking," *Med. Image Anal.*, vol. 8, pp. 95–111, 2004. 37

[450] G. C. McKinnon, X. J. Zhou, and N. E. Leeds, "Phase corrected complex averaging for diffusion weighted spine imaging," in *Proc. Int. Soc. Magn. Reson. Med.*, 2000, p. 802. 16

[451] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskill, "Noise and filtration in magnetic resonance imaging," *Med. Phys.*, vol. 12, pp. 586–591, 1985. 15

[452] R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," *Preprint*, 2009, http://arxiv.org/abs/0909.5457. 161, 170, 173, 174, 182

[453] A. J. Miller and P. M. Joseph, "The use of power images to perform quantitative analysis on low SNR MR images," *Magn. Reson. Imag.*, vol. 11, pp. 1051–1056, 1993. 16

[454] 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. Imag.*, vol. 29, pp. 501–516, 2009. 17, 159, 160

[455] D. Mitsouras, G. P. Zientara, A. Edelman, and F. J. Rybicki, "Enhancing the acquisition efficiency of fast magnetic resonance imaging via broadband encoding of signal content," *Magn. Reson. Imag.*, vol. 24, pp. 1209–1227, 2006. 14, 124, 128

[456] M. J. Mohlenkamp and L. Monzón, "Trigonometric identities and sums of separable functions," *Math. Intell.*, vol. 27, pp. 65–69, 2005. 160

[457] R. Molina, J. Mateos, A. K. Katsaggelos, and M. Vega, "Bayesian multichannel image restoration using compound Gauss-Markov random fields," *IEEE Trans. Image Process.*, vol. 12, pp. 1642–1654, 2003. 37, 38, 39

[458] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River: Prentice Hall, 2000. 22, 23, 27

[459] H. Moriguchi and J. L. Duerk, "Iterative next-neighbor regridding (INNG): Improved reconstruction from nonuniformly sampled k-space data using rescaled matrices," *Magn. Reson. Med.*, vol. 51, pp. 343–352, 2004. 21

[460] P. Mukherjee, J. I. Berman, S. W. Chung, C. P. Hess, and R. G. Henry, "Diffusion tensor MR imaging and fiber tractography: Theoretic underpinnings," *Am. J. Neuroradiol.*, vol. 29, pp. 632–641, 2008. 29, 89, 91

[461] K. Murase, Y. Yamazaki, M. Shinohara, K. Kawakami, K. Kikuchi, H. Miki, T. Mochizuki, and J. Ikezoe, "An anisotropic diffusion method for denoising dynamic susceptibility contrast-enhanced magnetic resonance images," *Phys. Med. Biol.*, vol. 46, pp. 2713–2723, 2001. 36

[462] J. I. Myung, D. J. Navarro, and M. A. Pitt, "Model selection by normalized maximum likelihood," *J. Math. Psych.*, vol. 50, pp. 167–179, 2006. 194, 195, 196

[463] B. Nadler, "Nonparametric detection of signals by information theoretic criteria: performance analysis and an improved estimator," *IEEE Trans. Signal Process.*, vol. 58, pp. 2746–2756, 2010. 194, 196

[464] S. K. Nagle and D. N. Levin, "Multiple region MRI," *Magn. Reson. Med.*, vol. 41, pp. 774–786, 1999. 34, 115

[465] O. Nalcioglu and Z. H. Cho, "Limits to signal-to-noise improvement by FID averaging in NMR imaging," *Phys. Med. Biol.*, vol. 29, pp. 969–978, 1984. 15, 16

[466] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995. 118

[467] P. R. C. Nelson, P. A. Taylor, and J. F. MacGregor, "Missing data methods in PCA and PLS: Score calculations with incomplete observations," *Chemometr. Intell. Lab. Syst.*, vol. 35, pp. 45–65, 1996. 174

[468] M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM J. Appl. Math.*, vol. 61, pp. 633–658, 2000. 74

[469] ——, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM J. Multiscale Model. Simul.*, vol. 4, pp. 960–991, 2005. 45

[470] ——, "Model distortions in Bayesian MAP reconstruction," *Inverse Probl. Imaging*, vol. 1, pp. 399–422, 2007. 43

[471] M. Nikolova and R. H. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process.*, vol. 16, pp. 1623–1627, 2007. 40, 44, 47

[472] 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. 40, 44, 47, 49, 74, 78, 104, 107

[473] H. Nilgens, M. Thelen, J. Pagg, P. Blümler, and B. Blümich, "Hadamard NMR imaging with slice selection," *Magn. Reson. Imag.*, vol. 14, pp. 857–861, 1996. 14

[474] R. Nishii, "Asymptotic properties of criteria for selection of variables in multiple regression," *Ann. Stat.*, vol. 12, pp. 758–765, 1984. 195

[475] D. C. Noll, D. G. Nishimura, and A. Macovski, "Homodyne detection in magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 10, pp. 154–163, 1991. 16, 17, 34, 89

[476] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Process.*, vol. 8, pp. 1408–1419, 1999. 16, 36

[477] V. Ntziachristos, "Going deeper than microscopy: the optical imaging frontier in biology," *Nat. Methods*, vol. 7, pp. 603–614, 2010. 34

[478] J. Nuyts, "The use of mutual information and joint entropy for anatomical priors in emission tomography," in *Proc. IEEE Nuc. Sci. Symp.*, 2007, pp. 4149–4154. 34

[479] J. Nuyts, K. Baete, D. Bequé, and P. Dupont, "Comparison between MAP and postprocessed ML for image reconstruction in emission tomography when anatomical knowledge is available," *IEEE Trans. Med. Imag.*, vol. 24, pp. 667–675, 2005. 34, 38, 39

[480] H. Nyquist, "Thermal agitation of electric charge in conductors," *Phys. Rev.*, vol. 32, pp. 110–113, 1928. 15

[481] J. Oakley, J. Missimer, and G. Székely, "Magnetic resonance imaging based correction and reconstruction of positron emission tomography images," *Lecture Notes Comp. Sci.*, vol. 1654, pp. 301–316, 1999. 34

[482] F. Odille, S. Uribe, P. G. Batchelor, C. Prieto, T. Schaeffter, and D. Atkinson, "Model-based reconstruction for cardiac cine MRI without ECG or breath holding," *Magn. Reson. Med.*, vol. 63, pp. 1247–1257, 2010. 159, 160

[483] T. Okatani and K. Deguchi, "On the Wiberg algorithm for matrix factorization in the presence of missing components," *Int. J. Comput. Vis.*, vol. 72, pp. 329–337, 2007. 174

[484] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 281, pp. 607–609, 1996. 24, 116

[485] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Upper Saddle River: Prentice Hall, 1999. 19

[486] R. J. Ordidge, A. Connelly, and J. A. B. Lohman, "Image-selected *in vivo* spectroscopy (ISIS). a new technique for spatially selective NMR spectroscopy," *J. Magn. Reson.*, vol. 66, pp. 283–294, 1986. 14

[487] V. Y. Orekhov, I. Ibraghimov, and M. Billeter, "Optimizing resolution in multidimensional NMR by three-way decomposition," *J. Biomol. NMR*, vol. 27, pp. 165–173, 2003. 203

[488] J. D. O'Sullivan, "A fast sinc function gridding algorithm for Fourier inversion in computer tomography," *IEEE Trans. Med. Imag.*, vol. 4, pp. 200–207, 1985. 19

[489] X. Ouyang, W. H. Wong, V. E. Johnson, X. Hu, and C.-T. Chen, "Incorporation of correlated structural images in PET image reconstruction," *IEEE Trans. Med. Imag.*, vol. 13, pp. 627–640, 1994. 34, 38, 39

[490] V. M. Pai, S. Rapacchi, P. Kellman, P. Croisille, and H. Wen, "PCATMIP: Enhancing signal intensity in diffusion-weighted magnetic resonance imaging," *Magn. Reson. Med.*, 2011, http://dx.doi.org/10.1002/mrm.22748. 162

[491] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Soft.*, vol. 8, pp. 43–71, 1982. 48

[492] J. A. Palmer, D. P. Wipf, K. Kreutz-Delgado, and B. D. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Advances in Neural Information Processing Systems*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. Cambridge: MIT Press, 2006, vol. 18, pp. 1059–1066. 46

[493] S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 1262–1275, 1983. 156

[494] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Probl.*, vol. 25, p. 123009, 2009. 124, 150

[495] L. P. Panych and F. A. Jolesz, "A dynamically adaptive imaging algorithm for wavelet-encoded MRI," *Magn. Reson. Med.*, vol. 32, pp. 738–748, 1994. 116

[496] L. P. Panych, G. P. Zientara, and F. A. Jolesz, "MR image encoding by spatially selective RF excitation: an analysis using linear response models," *Int. J. Imag. Syst. Tech.*, vol. 10, pp. 143–150, 1999. 14, 124, 128

[497] D. L. Parker and G. T. Gullberg, "Signal-to-noise efficiency in magnetic resonance imaging," *Med. Phys.*, vol. 17, pp. 250–257, 1990. 65

[498] D. L. Parker, G. T. Gullberg, and P. R. Frederick, "Gibbs artifact removal in magnetic resonance imaging," *Med. Phys.*, vol. 14, pp. 640–645, 1987. 92, 101

[499] G. J. M. Parker, J. A. Schnabel, M. R. Symms, D. J. Werring, and G. J. Barker, "Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging," *J. Magn. Reson. Imag.*, vol. 11, pp. 702–710, 2000. 36

[500] T. Parrish and X. Hu, "Continuous update with random encoding (CURE): A new strategy for dynamic imaging," *Magn. Reson. Med.*, vol. 33, pp. 326–336, 1995. 116, 121, 129

[501] T. B. Parrish and X. Hu, "Hybrid technique for dynamic imaging," *Magn. Reson. Med.*, vol. 44, pp. 51–55, 2000. 159, 160

[502] J. Pauly, D. Nishimura, and A. Macovski, "A k-space analysis of small-tip-angle excitation," *J. Magn. Reson.*, vol. 81, pp. 43–56, 1989. 128

[503] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis," *Magn. Reson. Med.*, vol. 62, pp. 706–716, 2009. 159, 160

[504] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 12, pp. 629–639, 1990. 36

[505] D. C. Peters, F. R. Korosec, T. M. Grist, W. F. Block, J. E. Holden, K. K. Vigen, and C. A. Mistretta, "Undersampled projection reconstruction applied to MR angiography," *Magn. Reson. Med.*, vol. 43, pp. 91–101, 2000. 116, 121

[506] R. D. Peters and M. L. Wood, "Multilevel wavelet-transform encoding in MRI," *J. Magn. Reson. Imag.*, vol. 6, pp. 529–540, 1996. 128

[507] J. G. Pipe, "Spatial encoding and reconstruction in MRI with quadratic phase profiles," *Magn. Reson. Med.*, vol. 33, pp. 24–33, 1995. 14

[508] ——, "Reconstructing MR images from undersampled data: data-weighting considerations," *Magn. Reson. Med.*, vol. 43, pp. 867–875, 2000. 116, 121

[509] J. G. Pipe and P. Menon, "Sampling density compensation in MRI: Rationale and an iterative numerical solution," *Magn. Reson. Med.*, vol. 41, pp. 179–186, 1999. 19

[510] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, pp. 323–331, 2003. 16, 36

[511] S. K. Plevritis and A. Macovski, "MRS imaging using anatomically based k-space sampling and extrapolation," *Magn. Reson. Med.*, vol. 34, pp. 686–693, 1995. 115

[512] ——, "Spectral extrapolation of spatially bounded images," *IEEE Trans. Med. Imag.*, vol. 14, pp. 487–497, 1995. 22, 34

[513] S. L. Ponder and D. B. Twieg, "A novel sampling method for ^{31}P spectroscopic imaging with improved sensitivity, resolution, and sidelobe suppression," *J. Magn. Reson. B*, vol. 104, pp. 85–88, 1994. 92

[514] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 1994. 22, 23, 24, 196

[515] D. E. Prah, E. S. Paulson, A. S. Nencka, and K. M. Schmainda, "A simple method for rectified noise floor suppression: phase-corrected real data reconstruction with application to diffusion-weighted imaging," *Magn. Reson. Med.*, vol. 64, pp. 418–429, 2010. 16

[516] C. Prieto, P. G. Batchelor, D. L. G. Hill, J. V. Hajnal, M. Guarini, and P. Irarrazaval, "Reconstruction of undersampled dynamic images by modeling the motion of object elements," *Magn. Reson. Med.*, vol. 57, pp. 939–949, 2007. 159, 160

[517] K. P. Pruessmann, "Encoding and reconstruction in parallel MRI," *NMR Biomed.*, vol. 19, pp. 288–299, 2006. 14

[518] ——, "Sources of noise and limits of SNR," in *Proc. Int. Soc. Magn. Reson. Med. Weekend Educational Courses*, 2009. 15

[519] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 46, pp. 638–651, 2001. 14, 125, 132

[520] 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. 14

[521] F. Pukelsheim, *Optimal Design of Experiments*. New York: John Wiley & Sons, 1993. 23, 97, 98, 99, 100

[522] F. Pukelsheim and B. Torsney, "Optimal weights for experimental designs on linearly independent support points," *Ann. Stat.*, vol. 19, pp. 1614–1625, 1991. 101

[523] G. Puy, Y. Wiaux, R. Gruetter, J.-P. Thiran, D. V. de Ville, and P. Vandergheynst, "Spread spectrum for interferometric and magnetic resonance imaging," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 2802–2805. 124, 153

[524] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imag.*, vol. 19, pp. 493–506, 2000. 52, 154

[525] J. B. Ra and C. Y. Rim, "Fast imaging using subencoding data sets from multiple detectors," *Magn. Reson. Med.*, vol. 30, pp. 142–145, 1993. 14

[526] D. K. Ragan, S. Y. Lai, and J. A. Bankson, "Fast, reproducible measurement of the vascular input function in mice using constrained reconstruction and cardiac sampling," *NMR Biomed.*, 2010, http://dx.doi.org/10.1002/nbm.1601. 159, 160

[527] A. Raj, C. Hess, and P. Mukherjee, "Spatial HARDI: Improved visualization of complex white matter architecture with Bayesian spatial regularization," *NeuroImage*, vol. 54, pp. 396–409, 2011. 37, 38, 39

[528] S. Ramani and J. A. Fessler, "An accelerated iterative reweighted least squares algorithm for compressed sensing MRI," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 257–260. 110, 111

[529] ——, "Parallel MR image reconstruction using augmented Lagrangian methods," *IEEE Trans. Med. Imag.*, vol. 30, pp. 694–706, 2011. 114

[530] A. Rangarajan, I.-T. Hsiao, and G. Gindi, "A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction," *J. Math. Imaging Vis.*, vol. 12, pp. 199–217, 2000. 34

[531] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*. New York: John Wiley & Sons, Inc., 1971. 21, 23

[532] V. Rasche, R. Proska, R. Sinkus, P. Börnert, and H. Eggers, "Resampling of data between arbitrary grids using convolution interpolation," *IEEE Trans. Med. Imag.*, vol. 18, pp. 385–392, 1999. 19

[533] Lord Rayleigh, "Investigations in optics, with special reference to the spectroscope," *Philos. Mag.*, vol. 5, pp. 261–274, 1879. 67

[534] P. J. Ready and P. A. Wintz, "Information extraction, SNR improvement, and data compression in multispectral imagery," *IEEE Trans. Comm.*, vol. COM-21, pp. 1123–1131, 1973. 37

[535] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, pp. 471–501, 2010. 161, 168, 169, 170, 173, 182, 183

[536] B. Recht, W. Xu, and B. Hassibi, "Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization," in *Proc. IEEE Conf. Dec. Cont.*, 2008, pp. 3065–3070. 161, 170, 183

[537] S. J. Reeves and Z. Zhe, "Sequential algorithms for observation selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 123–132, 1999. 98

[538] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proc. Int. Conf. Mach. Learn.*, 2005. 161

[539] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Technol. J.*, vol. 23, pp. 282–332, 1944. 16

[540] S. J. Riederer, T. Tasciyan, F. Farzaneh, J. N. Lee, R. C. Wright, and R. J. Herfkens, "MR fluoroscopy: Technical feasibility," *Magn. Reson. Med.*, vol. 8, pp. 1–15, 1988. 159

[541] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *Amer. Stat.*, vol. 2, pp. 59–66, 1988. 36

[542] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, "The NMR phased array," *Magn. Reson. Med.*, vol. 16, pp. 192–225, 1990. 14

[543] M. S. Roos and S. T. S. Wong, "Spatial localization in stochastic NMR imaging with oscillating gradients:," *J. Magn. Reson.*, vol. 87, pp. 554–566, 1990. 14

[544] D. Rosenfeld, "New approach to gridding using regularization and estimation theory," *Magn. Reson. Med.*, vol. 48, pp. 193–202, 2002. 21

[545] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. Pure Appl. Math.*, vol. 61, pp. 1025–1045, 2008. 121

[546] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992. 24, 36, 44, 131

[547] W. Ruhm and P. Ullmann, "Multidimensional spatial encoding by parallel excitation," in *Proc. Int. Soc. Magn. Reson. Med.*, 2008, p. 1318. 14

[548] R. Saab and Ö. Yilmaz, "Sparse recovery by non-convex optimization – instance optimality," *Appl. Comput. Harmon. Anal.*, vol. 29, pp. 30–48, 2010. 120

[549] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, X. Mao, and L. C. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Trans. Med. Imag.*, vol. 23, pp. 1453–1465, 2004. 162, 171

[550] A. A. Samsonov and C. R. Johnson, "Noise-adaptive nonlinear diffusion filtering of MR images with spatially varying noise levels," *Magn. Reson. Med.*, vol. 52, pp. 798–806, 2004. 36

[551] S. Sastry and R. E. Carson, "Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: A tissue composition model," *IEEE Trans. Med. Imag.*, vol. 16, pp. 750–761, 1997. 34

[552] A. R. C. Sava, D. M. Sima, J.-B. Poullet, A. J. Wright, A. Heerschap, and S. Van Huffel, "Exploiting spatial information to estimate metabolite levels in two-dimensional MRSI of heterogeneous brain lesions," *NMR Biomed.*, 2011, http://dx.doi.org/10.1002/nbm.1628. 36

[553] K. Scheffler and J. Hennig, "Frequency resolved single-shot MR imaging using stochastic k-space trajectories," *Magn. Reson. Med.*, vol. 35, pp. 569–576, 1996. 116, 121

[554] ——, "Reduced circular field-of-view imaging," *Magn. Reson. Med.*, vol. 40, pp. 474–480, 1998. 115

[555] P. Scheunders, "Wavelet thresholding of multivalued images," *IEEE Trans. Image Process.*, vol. 13, pp. 475–483, 2004. 37

[556] P. Scheunders and S. De Backer, "Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors," *IEEE Trans. Image Process.*, vol. 16, pp. 1865–1872, 2007. 37

[557] C. O. Schirra, S. Weiss, S. Krueger, S. F. Pedersen, R. Razavi, T. Schaeffter, and S. Kozerke, "Toward true 3D visualization of active catheters using compressed sensing," *Magn. Reson. Med.*, vol. 62, pp. 341–347, 2009. 116, 152

[558] V. J. Schmid, B. Witcher, A. R. Padhani, N. J. Taylor, and G.-Z. Yang, "Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 25, pp. 1627–1636, 2006. 36

[559] S. Schmitter, N. N. Mistry, C. Brinegar, G. A. Johnson, and Z.-P. Liang, "High-resolution pulmonary perfusion imaging in rodents using a spatiotemporal model," in *Proc. Int. Soc. Magn. Reson. Med.*, 2008. 159, 160

[560] H. Schomberg and J. Timmer, "The gridding method for image reconstruction by Fourier transformation," *IEEE Trans. Med. Imag.*, vol. 14, pp. 596–607, 1995. 19

[561] G. Schultz, P. Ullmann, H. Lehr, A. M. Welz, J. Hennig, and M. Zaitsev, "Reconstruction of MRI data encoded with arbitrarily shaped, curvilinear, nonbijective magnetic fields," *Magn. Reson. Med.*, vol. 64, pp. 1390–1404, 2010. 14

[562] R. R. Schultz and R. L. Stevenson, "Stochastic modeling and estimation of multispectral image data," *IEEE Trans. Image Process.*, vol. 4, pp. 1109–1119, 1995. 37, 38, 39

[563] A. D. Scott, J. Keegan, and D. N. Firmin, "Motion in cardiovascular MR imaging," *Radiology*, vol. 250, pp. 331–351, 2009. 159

[564] G. C. Scott, "Theory of RF reciprocity," in *Proc. Int. Soc. Magn. Reson. Med. Weekend Educational Courses*, 2009. 10

[565] K. N. Scott, H. R. Brooker, J. R. Fitzsimmons, H. F. Bennett, and R. C. Mick, "Spatial localization of ^{31}P nuclear magnetic resonance signal by the sensitive point method," *J. Magn. Reson.*, vol. 50, pp. 339–344, 1982. 12

[566] F. M. Sebert, Y. Zou, B. Liu, and L. Ying, "Compressed sensing MRI with random B1 field," in *Proc. Int. Soc. Magn. Reson. Med.*, 2008, p. 1318. 124, 153

[567] H. Sedarat and D. G. Nishimura, "On the optimality of the gridding reconstruction algorithm," *IEEE Trans. Med. Imag.*, vol. 19, pp. 306–317, 2000. 19, 21

[568] M. Seeger, H. Nickisch, R. Pohmann, and B. Schölkopf, "Optimization of k-space trajectories for compressed sensing by Bayesian experimental design," *Magn. Reson. Med.*, vol. 63, pp. 116–126, 2010. 116, 152

[569] A. Sen Gupta and Z.-P. Liang, "Dynamic imaging by temporal modeling with principal component analysis," in *Proc. Int. Soc. Magn. Reson. Med.*, 2001, p. 10. 159, 160

[570] L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, pp. 2744–2756, 2002. 40

[571] 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. 115, 159, 160

[572] G. Shi, D. Gao, X. Xie, X. Chen, and D. Liu, "High-resolution imaging via moving random exposure and its simulation," *IEEE Trans. Image Process.*, vol. 20, pp. 276–282, 2011. 124

[573] H.-Y. Shum, K. Ikeuchi, and R. Reddy, "Principal component analysis with missing data and its application to polyhedral object modeling," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 17, pp. 854–867, 1995. 161

[574] J. Sijbers and A. J. den Dekker, "Maximum likelihood estimation of signal amplitude and noise variance from MR data," *Magn. Reson. Med.*, vol. 51, pp. 586–594, 2004. 16

[575] J. Sijbers, A. J. den Dekker, A. Van der Linden, M. Verhoye, and D. Van Dyck, "Adaptive anisotropic noise filtering for magnitude MR data," *Magn. Reson. Imag.*, vol. 17, pp. 1533–1539, 1999. 16, 36

[576] S. Silvey, *Optimal Design*. New York: Chapman and Hall, 1980. 97, 98, 100

[577] C. P. Slichter, *Principles of Magnetic Resonance*, 3rd ed. Berlin: Springer-Verlag, 1990. 7

[578] M. R. Smith, S. T. Nichols, R. M. Henkelman, and M. L. Wood, "Application of autoregressive moving average parametric modeling in magnetic resonance image reconstruction," *IEEE Trans. Med. Imag.*, vol. MI-5, pp. 132–139, 1986. 116

[579] D. L. Snyder, "Utilizing side information in emission tomography," *IEEE Trans. Nuc. Sci.*, vol. NS-31, pp. 533–537, 1984. 34

[580] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, pp. 591–603, 1997. 14

[581] S. Som, B. F. Hutton, and M. Braun, "Properties of minimum cross-entropy reconstruction of emission tomography with anatomically based prior," *IEEE Trans. Nuc. Sci.*, vol. 45, pp. 3014–3021, 1998. 34

[582] S. Somayajula, C. Panagiotou, A. Rangarajan, Q. Li, S. R. Arridge, and R. M. Leahy, "PET image reconstruction using information theoretic anatomical priors," *IEEE Trans. Med. Imag.*, vol. 30, pp. 537–549, 2011. 34

[583] M. Sonka and J. M. Fitzpatrick, Eds., *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*. Bellingham: SPIE Press, 2000. 206

[584] C. M. Sparrow, "On spectroscopic resolving power," *Astrophys. J.*, vol. 44, pp. 76–86, 1916. 67

[585] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 720–727. 174

[586] J. M. Star-Lack, "Optimal gradient waveform design for projection imaging and projection reconstruction echoplanar spectroscopic imaging," *Magn. Reson. Med.*, vol. 41, pp. 664–675, 1999. 92

[587] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, pp. 601–615, 2000. 154

[588] ——, "Efficient calculation of resolution and covariance for penalized-likelihood reconstruction in fully 3-D SPECT," *IEEE Trans. Med. Imag.*, vol. 23, pp. 1543–1555, 2004. 52, 154

[589] R. Stobbe and C. Beaulieu, "Advantage of sampling density weighted apodization over postacquisition filtering apodization for sodium MRI of the human brain," *Magn. Reson. Med.*, vol. 60, pp. 981–986, 2008. 92

[590] J. P. Stockmann, P. A. Ciris, G. Galiana, L. Tam, and R. T. Constable, "O-space imaging: Highly efficient parallel imaging using second-order nonlinear fields as encoding gradients with no phase encoding," *Magn. Reson. Med.*, vol. 64, pp. 447–456, 2010. 14

[591] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, pp. 36–47, 2004. 169, 194, 195

[592] E. M. Stokely and D. B. Twieg, "Reconstructing magnetic resonance spectroscopic images using spatial domain priors," in *Proc. IEEE Int. Conf. Image Process.*, 1994, pp. 6–10. 21, 22, 34

[593] S. S. Stone, J. P. Haldar, S. C. Tsao, W. m. W. Hwu, B. P. Sutton, and Z.-P. Liang, "Accelerating advanced MRI reconstructions on GPUs," *J. Parallel Distrib. Comput.*, vol. 68, pp. 1307–1318, 2008. 4, 33, 49

[594] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, p. 421425, 2009. 161

[595] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities," *IEEE Trans. Med. Imag.*, vol. 22, pp. 178–188, 2003. 13

[596] S. Tan and L. Jiao, "Multivariate statistical models for image denoising in the wavelet domain," *Int. J. Comput. Vis.*, vol. 75, pp. 209–230, 2007. 40

[597] J. Tang, B. M. W. Tsui, and A. Rahmim, "Bayesian PET image reconstruction incorporating anato-functional joint entropy," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2008, pp. 1043–1046. 34

[598] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Variational approach for edge-preserving regularization using coupled PDE's," *IEEE Trans. Image Process.*, vol. 7, pp. 387–397, 1998. 40

[599] C. G. Thomas, R. A. Harshman, and R. S. Menon, "Noise reduction in BOLD-based fMRI using component analysis," *NeuroImage*, vol. 17, pp. 1521–1537, 2002. 37, 162, 171

[600] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, pp. 267–288, 1996. 118

[601] A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*. Dordrecht: Kluwer Academic Publishers, 1995. 17

[602] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems," *Pacific J. Optim.*, vol. 6, pp. 615–640, 2010. 169

[603] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *Int. J. Comput. Vis.*, vol. 9, pp. 137–154, 1992. 161

[604] C. Triantafyllou, R. D. Hoge, and L. L. Wald, "Effect of spatial smoothing on physiological noise in high-resolution fMRI," *NeuroImage*, vol. 32, pp. 551–557, 2006. 66

[605] A. Tristán-Vega and S. Aja-Fernández, "DWI filtering using joint information for DTI and HARDI," *Med. Image Anal.*, vol. 14, pp. 205–218, 2010. 16, 37

[606] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, pp. 589–602, 2006. 40

[607] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, pp. 948–958, 2010. 118, 120

[608] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Med. Imag.*, vol. 28, pp. 106–121, 2009. 116, 131, 152

[609] ——, "Relaxed conditions for sparse signal recovery with general concave priors," *IEEE Trans. Signal Process.*, vol. 4, pp. 4347–4354, 2009. 120

[610] J. Tsao, "On the UNFOLD method," *Magn. Reson. Med.*, vol. 47, pp. 202–207, 2002. 115, 159, 160

[611] J. Tsao, B. Behnia, and A. G. Webb, "Unifying linear prior-information-driven methods for accelerated image acquisition," *Magn. Reson. Med.*, vol. 46, pp. 652–660, 2001. 21, 34

[612] 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. 159, 160

[613] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 27, pp. 506–517, 2005. 37

[614] D. B. Twieg, "The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods," *Med. Phys.*, vol. 10, pp. 610–621, 1983. 12

[615] M. Uecker, S. Zhang, D. Voit, A. Karaus, K.-D. Merboldt, and J. Frahm, "Real-time MRI at a resolution of 20 ms," *NMR Biomed.*, vol. 23, pp. 986–994, 2010. 159, 160

[616] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, pp. 22–38, 1999. 22

[617] J. R. Valenzuela and J. A. Fessler, "Joint reconstruction of Stokes images from polarimetric measurements," *J. Opt. Soc. Am. A*, vol. 26, pp. 962–968, 2009. 37, 38, 39, 40

[618] R. Van de Walle, H. H. Barrett, K. J. Myers, M. I. Altbach, B. Desplanques, A. F. Gmitro, J. Cornelis, and I. Lemahieu, "Reconstruction of MR images from data acquired on a general nonregular grid by pseudoinverse calculation," *IEEE Trans. Med. Imag.*, vol. 19, pp. 1160–1167, 2000. 20

[619] J. J. van Vaals, M. E. Brummer, W. T. Dixon, H. H. Tuithof, H. Engels, R. C. Nelson, B. M. Gerety, J. L. Chezmar, and J. A. den Boer, ""keyhole" method for accelerating imaging of contrast agent uptake," *J. Magn. Reson. Imag.*, vol. 3, pp. 671–675, 1993. 116, 159

[620] N. Vaswani, "LS-CS-residual (LS-CS): Compressive sensing on least squares residual," *IEEE Trans. Signal Process.*, vol. 58, pp. 4108–4120, 2010. 116

[621] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, pp. 4595–4607, 2010. 116

[622] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2D signals," in *Proc. IEEE Int. Conf. Image Process.*, 1998, pp. 752–756. 115, 120, 159, 160

[623] R. Vidal, R. Tron, and R. Hartley, "Multiframe motion segmentation with missing data using PowerFactorization and GPCA," *Int. J. Comput. Vis.*, vol. 79, pp. 85–105, 2008. 174, 175, 176

[624] V. Vitanis, R. Manka, D. Giese, H. Pedersen, S. Plein, P. Boesiger, and S. Kozerke, "High resolution three-dimensional cardiac perfusion imaging using compartment-based *k-t* principal component analysis," *Magn. Reson. Med.*, vol. 65, pp. 575–587, 2011. 159, 160

[625] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia: SIAM, 2002. 17, 20, 23, 47, 107, 118

[626] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Comput. Harmon. Anal.*, vol. 11, pp. 89–123, 2001. 40

[627] F. T. A. W. Wajer, "Non-Cartesian MRI scan time reduction through sparse sampling," Ph.D. dissertation, Technische Universiteit Delft, Delft, Netherlands, 2001. 49

[628] S. Walker-Samuel, M. Orton, J. K. R. Boult, and S. P. Robinson, "Improving apparent diffusion coefficient estimates and elucidating tumor heterogeneity using Bayesian adaptive smoothing," *Magn. Reson. Med.*, vol. 65, pp. 438–447, 2011. 37

[629] C.-H. Wang, J.-C. Chen, and R.-S. Liu, "Development and evaluation of MRI based Bayesian image reconstruction methods for PET," *Comput. Med. Imaging Graph.*, vol. 28, pp. 177–184, 2004. 34

[630] G. Wang and Y. Li, "Axiomatic approach for quantification of image resolution," *IEEE Signal Process. Lett.*, vol. 6, pp. 257–258, 1999. 68

[631] G. Wang and J. Qi, "Analysis of penalized likelihood image reconstruction for dynamic PET quantification," *IEEE Trans. Med. Imag.*, vol. 28, pp. 608–620, 2009. 52

[632] H. Wang, D. Liang, K. F. King, and L. Ying, "Toeplitz random encoding for reduced acquisition using compressed sensing," in *Proc. Int. Soc. Magn. Reson. Med.*, 2009, p. 2669. 124

[633] Y. Wang, Q. Wang, J. H. Kim, A. H. Cross, and S.-K. Song, "Noninvasive differentiation of co-existing inflammation, axonal injury, and demyelination in the mouse corpus callosum," in *Proc. Amer. Committee Treat. Res. Mult. Scler.*, 2010. 187

[634] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imaging Sci.*, vol. 1, pp. 248–272, 2008. 104, 105

[635] Z. Wang and G. R. Arce, "Variable density compressed image sampling," *IEEE Trans. Image Process.*, vol. 19, pp. 264–270, 2010. 129

[636] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, pp. 98–117, 2009. 153, 201

[637] Z. Wang, B. C. Vemuri, Y. Chen, and T. H. Mareci, "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI," *IEEE Trans. Med. Imag.*, vol. 23, pp. 930–939, 2004. 37, 38, 39

[638] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp. 387–392, 1985. 194, 196

[639] J. B. Weaver, Y. Xu, D. M. Healy, and L. D. Cromwell, "Filtering noise from images with wavelet transforms," *Magn. Reson. Med.*, vol. 21, pp. 288–295, 1991. 36

[640] J. B. Weaver, Y. Xu, D. M. Healy, and J. R. Driscoll, "Wavelet-encoded MR imaging," *Magn. Reson. Med.*, vol. 24, pp. 275–287, 1992. 14

[641] V. J. Wedeen, Y.-S. Chao, and J. L. Ackerman, "Dynamic range compression in MRI by means of a nonlinear gradient pulse," *Magn. Reson. Med.*, vol. 6, pp. 287–295, 1988. 14

[642] V. J. Wedeen, R. M. Weisskoff, T. G. Reese, G. M. Beache, B. P. Poncelet, B. R. Rosen, and R. E. Dinsmore, "Motionless movies of myocardial strain-rates using stimulated echoes," *Magn. Reson. Med.*, vol. 33, pp. 401–408, 1995. 162

[643] J. Weickert and H. Hagen, Eds., *Visualization and Processing of Tensor Fields*. Berlin: Springer, 2006. 37

[644] M. Weiger, F. Hennel, and K. P. Pruessmann, "Sweep MRI with algebraic reconstruction," *Magn. Reson. Med.*, vol. 64, pp. 1685–1695, 2010. 14

[645] R. A. Weisenseel, W. C. Karl, R. C. Chan, and T. J. Brady, "A variational approach to multi-modality subsurface data inversion and fusion based on shared image structure," *Subsurface Sensing Technol. Appl.*, vol. 4, pp. 375–394, 2003. 37, 38, 39, 40

[646] R. Weissleder and U. Mahmood, "Molecular imaging," *Radiology*, vol. 219, pp. 316–333, 2001. 3

[647] M. Welk, J. Weickert, F. Becker, C. Schnörr, C. Feddern, and B. Burgeth, "Median and related local filters for tensor-valued images," *Signal Process.*, vol. 87, pp. 291–308, 2006. 40

[648] J. A. Wells, D. L. Thomas, M. D. King, A. Connelly, M. F. Lythgoe, and F. Calamante, "Reduction of errors in ASL cerebral perfusion and arterial transit time maps using image de-noising," *Magn. Reson. Med.*, vol. 64, pp. 715–724, 2010. 37, 162, 171

[649] Y. Wiaux, G. Puy, R. Gruetter, J.-P. Thiran, D. V. de Ville, and P. Vandergheynst, "Spread spectrum for compressed sensing techniques in magnetic resonance imaging," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 756–759. 124, 153

[650] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot, "Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI," in *Proc. MICCAI*, 2007, pp. 344–351. 36, 37

[651] A. M. Wink and J. B. T. M. Roerdink, "Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing," *IEEE Trans. Med. Imag.*, vol. 23, pp. 374–387, 2004. 36

[652] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *Preprint*, 2011, http://dsp.ucsd.edu/~dwipf/wipf_draft2009.pdf. 120

[653] R. Wirestam, A. Bibic, J. Lätt, S. Brockstedt, and F. Ståhlberg, "Denoising of complex MRI data by wavelet-domain filtering: Application to high-*b*-value diffusion-weighted imaging," *Magn. Reson. Med.*, vol. 56, pp. 1114–1120, 2006. 36

[654] H. Wold, "Nonlinear estimation by iterative least squares procedures," in *Research Papers in Statistics*, F. David, Ed. New York: Wiley, 1966, pp. 411–444. 174, 176

[655] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometr. Intell. Lab. Syst.*, vol. 2, pp. 37–52, 1987. 161, 174, 176

[656] E. C. Wong, "Efficient randomly encoded data acquisition for compressed sensing," in *Proc. Int. Soc. Magn. Reson. Med.*, 2010, p. 4893. 124, 128

[657] K. M. Wong, Q.-T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 1959–1971, 1990. 194, 196

[658] S. Wong, L. Zaremba, D. Gooden, and H. K. Huang, "Radiologic image compression – a review," *Proc. IEEE*, vol. 83, pp. 194–219, 1995. 116

[659] S. T. S. Wong, M. S. Roos, R. D. Newmark, and T. F. Budinger, "Discrete analysis of stochastic NMR. I," *J. Magn. Reson.*, vol. 87, pp. 242–264, 1990. 128

[660] J. C. Wood and K. M. Johnson, "Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR," *Magn. Reson. Med.*, vol. 41, pp. 631–635, 1999. 16, 36

[661] X.-L. Wu, J. Gai, F. Lam, M. Fu, J. P. Haldar, Y. Zhuo, Z.-P. Liang, W.-M. Hwu, and B. P. Sutton, "IMPATIENT MRI: Illinois Massively Parallel Acceleration Toolkit for Image reconstruction with ENhanced Throughput in MRI," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2011, pp. 69–72. 49

[662] Q.-S. Xiang, "Accelerating MRI by skipped phase encoding and edge deghosting (SPEED)," *Magn. Reson. Med.*, vol. 53, pp. 1112–1117, 2005. 116

[663] Q.-S. Xiang and R. M. Henkelman, "*K*-space description for MR imaging of dynamic objects," *Magn. Reson. Med.*, vol. 29, pp. 422–428, 1993. 158

[664] D. Xu, L. Ying, and Z.-P. Liang, "Dynamic MRI using spatiotemporal modeling with phased array coils," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2006, pp. 722–725. 159, 160

[665] Y. Yamada, K. Tanaka, and Z. Abe, "NMR Fresnel transform imaging technique using a quadratic nonlinear field gradient," *Rev. Sci. Instrum.*, vol. 63, pp. 5348–5358, 1992. 14

[666] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Imaging Sci.*, vol. 2, pp. 569–592, 2009. 37, 38, 39, 40, 78

[667] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, pp. 250–278, 2011. 141

[668] J. C. Ye, S. Tak, Y. Han, and H. W. Park, "Projection reconstruction MR imaging using FOCUSS," *Magn. Reson. Med.*, vol. 57, pp. 764–775, 2007. 116, 152

[669] E. N. Yeh, C. A. McKenzie, M. A. Ohliger, and D. K. Sodickson, "Parallel magnetic resonance imaging with adaptive radius in k-space (PARS): Constrained image reconstruction using k-space locality in radiofrequency coil encoded data," *Magn. Reson. Med.*, vol. 53, pp. 1383–1392, 2005. 14

[670] S. Zaroubi and G. Goelman, "Complex denoising of MR data via wavelet analysis: application for functional MRI," *Magn. Reson. Imag.*, vol. 18, pp. 59–68, 2000. 36

[671] K. Zeng, H. Yu, S. Zhao, L. L. Fajardo, C. Ruth, Z. Jing, and G. Wang, "Digital tomosynthesis aided by low-resolution exact computed tomography," *J. Comput. Assist. Tomogr.*, vol. 31, pp. 976–983, 2007. 34

[672] H. Zhang, L. Cheng, and W. Zhu, "Nuclear norm regularization with a low-rank constraint for matrix completion," *Inverse Probl.*, vol. 26, p. 115009, 2010. 169

[673] X. Zhang, E. Yacoub, and X. Hu, "New strategy for reconstructing partial-Fourier imaging data in functional MRI," *Magn. Reson. Med.*, vol. 46, pp. 1045–1048, 2001. 21, 34, 89

[674] Y. Zhang, J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "Joint estimation for incorporating MRI anatomic images into SPECT reconstruction," in *Proc. IEEE Nuc. Sci. Symp.*, 1994, pp. 1256–1260. 34, 38, 39

[675] 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. Imag.*, 2010, pp. 996–999. 159, 160, 185, 202

[676] B. Zhao, J. P. Haldar, and Z.-P. Liang, "PSF model-based reconstruction with sparsity constraint: Algorithm and application to real-time cardiac MRI," in *Proc. IEEE Eng. Med. Bio. Conf.*, 2010, pp. pp. 3390–3393. 159, 160, 202

[677] X. J. Zhou, Q. Gao, O. Abdullah, and R. L. Magin, "Studies of anomalous diffusion in the human brain using fractional order calculus," *Magn. Reson. Med.*, vol. 63, pp. 562–569, 2010. 33

[678] G. P. Zientara, L. P. Panych, and F. A. Jolesz, "Near-optimal spatial encoding for dynamically adaptive MRI: Mathematical principles and computational methods," *Int. J. Imag. Syst. Tech.*, vol. 10, pp. 151–165, 1999. 14, 116, 162

[679] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Stat.*, vol. 15, pp. 265–286, 2006. 202

[680] G. Zyskind, "On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models," *Ann. Math. Stat.*, vol. 38, pp. 1092–1109, 1967. 101