Improvements in Magnetic Resonance Imaging Excitation Pulse Design

by

Adam Charles Zelinski

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2008

Author .....
Department of Electrical Engineering and Computer Science
August 8, 2008

Certified by .....
Elfar Adalsteinsson
Associate Professor of Electrical Engineering
Associate Professor of Health Sciences and Technology
Thesis Supervisor

Certified by .....
Vivek K Goyal
Associate Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Improvements in Magnetic Resonance Imaging

Excitation Pulse Design

by

Adam Charles Zelinski

Submitted to the Department of Electrical Engineering and Computer Science
on August 8, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis focuses on the design of magnetic resonance imaging (MRI) radio-frequency (RF) excitation pulses, and its primary contributions are made through connections with the novel multiple-system single-output (MSSO) simultaneous sparse approximation problem. The contributions are both conceptual and algorithmic and are validated with simulations, as well as anthropogenic-object-based and in vivo trials on MRI scanners.

Excitation pulses are essential to MRI: they excite nuclear spins within a subject that are detected by a resonant coil and then reconstructed into images. Pulses need to be as short as possible due to spin relaxation, tissue heating, and main field inhomogeneity limitations. When magnetic spins are tilted by only a small amount, pulse transmission may be interpreted as depositing energy in a continuous three-dimensional Fourier-like domain along a one-dimensional contour to form an excitation in the spatial domain. Pulse duration is proportional to the length of the contour and inversely proportional to the rate at which it is traversed, and the rate is limited by system gradient hardware restrictions. Joint design of the contour and a corresponding excitation pulse is a difficult and central problem, while determining near-optimal energy deposition once the contour is fixed is significantly easier.

We first pose the NP-Hard MSSO problem and formulate greedy and convex relaxation-based algorithms with which to approximately solve it. We find that second-order-cone programming and iteratively-reweighted least squares approaches are practical techniques for solving the relaxed problem and prove that single-vector sparse approximation of a complex-valued vector is an MSSO problem.

We then focus on pulse design, first comparing three algorithms for solving linear systems of multi-channel excitation design equations, presenting experimental results from a 3 Tesla scanner with eight excitation channels.

Our aim then turns toward the joint design of pulses and trajectories. We take joint design in a novel direction by utilizing MSSO theory and algorithms to design short-duration sparsity-enforced pulses. These pulses are used to mitigate transmit field inhomogeneity in the human brain at 7 Tesla, a significant step towards the clinical use of high-field imaging in the study of cancer, Alzheimer's disease, and Multiple Sclerosis. Pulses generated by the sparsity-enforced method outperform those created via conventional Fourier-based techniques, e.g., when attempting to produce a uniform magnetization in the presence of severe RF inhomogeneity, a 5.7-ms 15-spoke pulse generated by the sparsity-enforced method produces an excitation with 1.28 times lower root-mean-square error than conventionally-designed 15-spoke pulses. To achieve this same level of uniformity, conventional methods

must use 29-spoke pulses that are 1.4 times longer.

We then confront a subset selection problem that arises when a parallel excitation system has more transmit modes available than hardware transmit channels with which to drive them. MSSO theory and algorithms are again applicable and determine surprising target-specific mixtures of light and dark modes that yield high-quality excitations.

Finally, we study the critical patient safety issue of specific absorption rate (SAR) of multi-channel excitation pulses at high field. We develop a fast SAR calculation algorithm and propose optimizing an individual pulse and time-multiplexing a set of pulses as ways to reduce SAR; the latter is capable of reducing maximum local SAR by 11% with no impact on pulse duration.

Thesis Supervisor: Elfar Adalsteinsson
Title: Associate Professor of Electrical Engineering
Associate Professor of Health Sciences and Technology

Thesis Supervisor: Vivek K Goyal
Title: Associate Professor of Electrical Engineering

Acknowledgments

First, I thank my parents, Linda and Edward Zelinski, for their unending encouragement and support; my mother is always willing to listen and my father has taught me to be persistent and always remain determined to succeed at whatever task is at hand.

Throughout my tenure here at MIT I have had the privilege to have two great thesis advisors, Elfar Adalsteinsson and Vivek Goyal. Their advice and guidance have been invaluable and helped me push forward with my research. Their down-to-earth, collaborative leadership styles have profoundly impacted how I myself lead and act as part of a team, and also how I approach and solve both quantitative and qualitative problems.

I am thankful to have been able to work alongside Lawrence Wald of Harvard Medical School's Radiology department. His advice allowed me to port some of my theoretical work to have direct impacts on upcoming medical imaging technology. He also was kind enough to serve as a reader of my thesis. Michael Elad of the Technion–Israel Institute of Technology was also great to work with; he spent a great deal of time with me and provided insight regarding the iterated shrinkage algorithms presented in this thesis.

Interacting with the members of the Magnetic Resonance Imaging Group has been a positive experience. Kawin Setsompop and Borjan Gagoski in particular have been great colleagues to work with—the former taught me the basics of MRI and helped me with my initial foray into the field, while the latter served more than once as a willing subject for my experiments; I truly value their friendship. Arlene Wint has always done a great job booking all of my travel. I've also enjoyed being part of the Signal Transformation and Information Representation Group, whose members—Lav Varshney, Julius Kusuma, and Daniel Weller in particular—have always been willing to give me useful feedback on my research and presentations. Eric Strattman has always provided great support as well.

Charles Swannack, Sourav Dey, Vijay Divi, and others here on the 6th floor of Building 36 have also helped make my time here at MIT an enjoyable experience. When I was a first-year graduate student, Dmitry Malioutov of the Stochastic Systems Group introduced me to the concept of sparse approximation, one of the themes of this thesis. Ivan Nausieda was my good friend during my undergraduate days at Carnegie Mellon University and has continued being one since coming to MIT to pursue his Ph.D.

Various researchers at Massachusetts General Hospital's A. A. Martinos Center for Biomedical Imaging have been instrumental in helping me obtain clinical results. Jonathan Polimeni, Thomas Witzel, and Vijay Alagappan selflessly spent many late nights helping me use the center's experimental system. Leonardo Angelone conducted valuable simulations which served as inputs for my work on patient safety.

Encouragement from our industry partner Franz Schmitt of Siemens Medical Solutions was invaluable. Being able to travel to Siemens' headquarters in Erlangen, Germany and contribute technology toward their upcoming high-field MRI systems was a rewarding experience.

Finally, I was fortunate enough to land a wonderful girlfriend while living here in Cambridge. Kirsten Learson has supported me unwaveringly over the course of my studies and is truly my best friend. I also am thankful we have so many great pets—Miss Whiskers, Wally, and Henry here in Cambridge—along with my Bombay, Watson, who lives with my parents in Annapolis, Maryland.

My research has been supported by National Institutes of Health grants 1P41RR14075, 1R-01EB000790, 1R01EB006847, and 1R01EB007942, National Science Foundation CA-

REER grant 643836, United States Department of Defense National Defense Science and Engineering Graduate Fellowship grant F49620-02-C-0041, the MIND Institute, the A. A. Martinos Center for Biomedical Imaging, and an R. J. Shillman Career Development Award.

Adam C. Zelinski
Cambridge, Massachusetts

Contents

List of Algorithms

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Motivation

Magnetic resonance imaging (MRI) is an extremely useful tool for producing high-resolution images of soft tissues in the human body. Unlike computed tomography (CT), MRI does not require the use of ionizing radiation and is generally a safe, practical procedure [22].

MRI may be viewed as a two-stage experiment that non-invasively yields a spatial mapping of hydrogen nuclei in living subjects. Nuclear spins within a subject are first excited using a radio-frequency (RF) excitation pulse and proportions of excited spins are then detected using a resonant coil; images are then reconstructed from this data. Excitation pulses need to be tailored to a user's specific needs and in most applications need to be as *short in duration* as possible due to spin relaxation, tissue heating, and main field inhomogeneity limitations. Further, in almost all imaging scenarios, excitations must not only be fast, but *spatially-selective* as well. That is, an excitation pulse should manipulate nuclear spins within only a particular two-dimensional (2-D) or three-dimensional (3-D) region of space (e.g., a 5-mm-thick slice of tissue), leaving spins outside of this region untouched and thus undetectable by the readout coil. Such spatially-selective excitations reduce the duration and complexity of both the data readout and image reconstruction processes [12]. Finally, *spatially-tailored* excitation pulses are often desired: these pulses are able to vary the extent to which nuclear spins are excited across space. In short, the design and analysis of such short-duration, spatially-selective, spatially-tailored excitation pulses, their application to open problems, and the study of a mathematical sparse approximation problem inspired by MRI pulse design are important topics and the focus of this thesis.

One may show that RF excitation pulse design, under a linear approximation, involves choosing to deposit energy in a continuous, 3-D, Fourier-like domain in order to form some desired excitation in the spatial domain [102]; this domain is referred to as *excitation k-space*. Energy may only be deposited along a 1-D contour in k-space, and there are limitations on where and how it may be placed: the velocity and acceleration with which one traverses the contour are subject to MRI system gradient hardware threshold values and switching times. The most important fact is that excitation pulse duration corresponds directly to the length of the chosen contour and the rate it is traversed. A key problem then is to find a short "trajectory" through this k-space (and a corresponding energy deposition along this trajectory) such that a high-fidelity version of the desired excitation forms in the spatial domain. Other problems of interest include analyzing the effect excitation pulses have on tissue heating, the design of pulses to mitigate such safety concerns, and the creation of new algorithms that will produce pulses suitable for use on newly-developed parallel transmission hardware.

1.2 Applications

Overall, in broad terms, each contribution to MRI excitation put forth in this thesis generally focuses on one, two, or all three of the following application areas:

1.2.1 High-Field Transmit Profile B_1^+ Inhomogeneity Mitigation

High-magnetic-field MRI systems exhibit great promise because they significantly increase tissue contrast and signal-to-noise ratio (SNR) [65], but *in vivo* human imaging at high field is impeded by the presence of severe B_1^+ transmit profile inhomogeneity [16], a phenomenon that arises due to wavelength interference effects [133, 136] and tissue-conductive RF amplitude attenuation [29]. When standard slice-selective RF excitation waveforms that work well on low-field systems are used to conduct high-field imaging experiments, B_1^+ inhomogeneity causes the resulting images to exhibit undesirable center brightening, spatial contrast variation, and SNR non-uniformity, despite the use of homogeneous volume RF excitation coils [26, 67, 73, 133, 136]. Solving this problem is a critical and necessary step towards the use of high-field imaging in the study of cancer, Alzheimer's disease, and Multiple Sclerosis [2]. Note that inhomogeneity is also a concern at low field when imaging

structures such as the spine [119] and body [15].

1.2.2 Parallel Excitation Multi-Channel Pulse Design

A recent MRI excitation concept, termed "parallel transmission" (and sometimes referred to as "parallel excitation" or "transmit SENSE"), involves equipping an MRI scanner with an RF excitation coil that differs substantially from those of standard systems in that it is comprised of multiple elements that are each capable of independent, simultaneous transmission [58, 79, 113, 130, 168, 169]. The presence of multiple elements allows one to *undersample* a given excitation k-space trajectory and yet in many cases still form a high-fidelity version of the desired excitation; undersampling the trajectory is greatly beneficial because it reduces the distance one travels in k-space, in turn reducing the duration of the corresponding pulse [58, 79, 113, 130, 168, 169]. This ability to "accelerate" in the Fourier-like k-space domain and reduce pulse duration arises due to the extra spatial degrees of freedom provided by the system's multiple transmit elements; significant acceleration is not possible on a conventional system equipped with a single-channel transmission coil unless sophisticated joint trajectory-pulse design techniques are employed. Parallel transmission systems also offer a flexible means for spatially-tailoring excitation patterns for inner-volume excitation [47] and addressing increased main field (B_0) and B_1^+ inhomogeneity observed at high field strengths [133, 136]. The design of parallel transmission pulses is an ongoing open problem; a key issue of interest involves understanding and exploring to what extent a given parallel transmission system may undersample excitation k-space and shorten pulse duration and yet still produce a high-quality, patient-safe excitation.

1.2.3 Specific Absorption Rate Analysis and Reduction

Specific absorption rate (SAR)—defined as the average energy deposition in an N-gram (Ng) region of tissue over a period of time due to the application of one or more radio-frequency (RF) excitation pulses—is an important safety concern when conducting MRI experiments on human subjects. Avoiding dangerously-high SAR is especially a concern for the parallel transmission of spatially-tailored multi-dimensional excitation pulses through a multi-channel transmission system [58, 79, 113, 130, 168, 169]. This is because when multiple transmit channels are simultaneously employed, the local electric fields generated by each channel undergo superposition and local extremes in electric field magnitude may arise [168],

leading to spikes in SAR within a local 1g or 10g region that are of major concern to regulatory bodies in both the United States [21] and Europe [70,71]. A recent study has confirmed the frequent occurrence of such "hot spots" and found that parallel excitation pulses produce relatively high ratios of local SAR to whole-head mean SAR [111], making *local* rather than global SAR the limiting factor of parallel transmission. Furthermore, it is unfortunate that the greatest strength of parallel transmission systems—their ability to undersample excitation k-space and reduce the length of the corresponding excitation pulse—creates an additional SAR concern beyond the aforementioned one. Namely, accelerating a k-space trajectory significantly increases peak pulse power [56, 78, 81, 141], which to the first order has a *quadratic* impact on SAR. For example, to maintain the same nuclear spin flip angle and excitation quality, a conventional "hard" (i.e., constant amplitude) excitation pulse [12] requires a peak power increase by a factor of C as its duration is shortened by the same factor, causing global SAR to increase by a factor of C^2. Even when the repetition time of the pulse is kept constant such that total RF duty cycle decreases by a factor of C, SAR still increases linearly with C [78]. Understanding the SAR behavior of accelerated parallel excitation pulses at both standard and high field strengths is an open problem, as is the optimization of parallel transmission pulses (and pulse sequences) for the purpose of local and global SAR reduction.

1.3 Algorithmic Focus: Multi-System Single-Output Simultaneous Sparse Approximation

Most of the advances in the first two MRI applications presented in this thesis are based on forging connections with a novel sparse approximation problem and developing algorithms for this problem. We will show that viewing pulse design from a sparsity approximation perspective [23, 24, 31, 34, 37, 41, 43, 44, 52, 93, 96, 107] gives rise to an NP-Hard linear inverse problem where simultaneously sparse vectors are required to solve a set of equations involving multiple system matrices and a single known observation vector that is by itself a focus worth pursuing from an applied mathematical standpoint. After studying multi-system, single-output (MSSO) systems and designing MSSO algorithms independently of excitation pulse design, we apply our work to the first two application areas listed above.

1.4 Thesis Outline and Contributions

The chapter-by-chapter structure of the dissertation is given below. Publications and intellectual property stemming from the work of each chapter are also listed.

Chapter 2 presents background information on MRI excitation theory, conventional single-channel and multi-channel pulse design, and sparse approximation. The most important segment of this chapter is our summary of how to linearize the continuous-time system of differential equations governing MRI RF excitation under the low-flip-angle assumption to arrive at the linear Fourier k-space relation between RF energy deposited by the system and the resulting excitation that forms in the spatial domain [102].

Chapter 3 proposes the multiple-system, single-output (MSSO) simultaneous sparsity problem, an NP-Hard linear inverse problem that requires the determination of multiple unknown signal vectors. Each unknown vector passes through a different system matrix and the results are added to yield a single observation vector. Given the matrices and lone observation, the objective is to find a *simultaneously sparse* set of unknown vectors that solves the system. Seven algorithms are formulated to approximately solve this problem. Three greedy techniques are developed (matching pursuit [93], orthogonal matching pursuit [23, 31, 96], and least squares matching pursuit [31]) along with four methods based on a convex relaxation (iteratively reweighted least squares [77], two forms of iterative shrinkage [34, 41, 43, 44], and formulation as a second-order cone program [17, 92]). While deriving the algorithms, we prove that *seeking a single sparse complex-valued vector is equivalent to seeking two* **simultaneously sparse** *real-valued vectors*, increasing the relevance and applicability of MSSO theory, and then proceed to evaluate how well the techniques perform during sparsity profile recovery and MRI multi-channel pulse design scenarios. Overall, each algorithm is found to have its own particular weaknesses and merits, e.g., the iterative shrinkage techniques converge slowly, but because they update only a subset of the overall solution per iteration rather than all unknowns at once, they are useful in cases where attempting the latter is prohibitive in terms of system memory. This work has been submitted as an article to the journal indicated below:

- A. C. Zelinski, V. K. Goyal, and E. Adalsteinsson. Simultaneously Sparse Solutions to Linear Inverse Problems with Multiple System Matrices and a Single Observation Vector. *Siam J Sci Comp*, In Review, 2008.

Chapter 4 studies three algorithms for solving linearized systems of RF waveform design equations for calculating accelerated spatially-tailored excitations on parallel excitation MRI systems. Their artifact levels, computational speed, and RF peak and root-mean-square voltages are analyzed. A singular value decomposition (SVD) [50, 51, 118] inversion method is compared to Conjugate Gradient Least Squares (CGLS) [62] and Least Squares QR (LSQR) [100, 101], two iterative algorithms designed to solve large linear systems. The excitation pulses calculated using these methods are used in both Bloch simulations and imaging experiments on an actual eight-channel parallel excitation array implemented on a 3T human scanner. Overall, experiments show that waveforms designed using LSQR and CGLS have lower peak and RMS waveform voltages and produce excitations with fewer artifacts than those generated by the SVD-based method. This work resulted in a journal article, conference paper, and patent:

- A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, V. K. Goyal, F. Hebrank, U. Fontius, F. Schmitt, and E. Adalsteinsson. Comparison of Three Algorithms for Solving Linearized Systems of Parallel Excitation RF Waveform Design Equations: Experiments on an Eight-Channel System at 3 Tesla *Concepts Magn Reson, Part B: Magn Reson Eng*, 31B(3):176–190, Aug. 2007.

- A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, F. Hebrank, U. Fontius, F. Schmitt, and E. Adalsteinsson. RF Pulse Design Methods for Reduction of Image Artifacts in Parallel RF Excitation: Comparison of 3 Techniques on a 3T Parallel Excitation System with 8 Channels. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1686, Berlin, Germany, 2007.

- A. C. Zelinski, E. Adalsteinsson, K. Setsompop, L. L. Wald, and U. Fontius. *Method for designing RF excitation pulses in magnetic resonance tomography.* US Patent 7336145, issued February 26, 2008.

Chapter 5 uses MSSO theory to develop a sparsity-enforcement algorithm that jointly determines quickly-traversable excitation k-space trajectories along with corresponding excitation pulses. The proposed method lets users specify a desired 3-D spatially-tailored and spatially-selective excitation and then generates a pulse and trajectory explicitly optimized for the task at hand. The algorithm functions by applying an ℓ_1-norm penalty while

searching over a large number of possible trajectory segments (and corresponding RF pulse segments); this process ultimately reveals a small, sparse subset of trajectory and pulse segments that alone form a high-fidelity version of the desired target excitation. The method is general: it applies to conventional single-channel as well as emerging multi-channel parallel transmission systems, and also to a variety of 2-D and 3-D k-space trajectories. After mathematically developing the joint pulse-trajectory design technique, we use it to design single-channel slice-selective pulses that mitigate B_1^+ inhomogeneity in the human brain at 7 Tesla and eight-channel pulses that produce highly-structured excitations at 3 Tesla, and single-channel 2-D spiral-trajectory pulses. Overall, this body of work led to two journal articles, a pending patent, and four conference abstracts:

- A. C. Zelinski, L. L. Wald, K. Setsompop, V. K. Goyal, and E. Adalsteinsson. Sparsity-Enforced Slice-Selective MRI RF Excitation Pulse Design. *IEEE Trans Med Imag*, In Press, 2008.

- A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, V. K. Goyal, and E. Adalsteinsson. Fast Slice-Selective RF Excitation Pulses for Mitigating B_1^+ Inhomogeneity in the Human Brain at 7T. *Magn Reson Med*, 59(6):1355-1364, June 2008.

- A. C. Zelinski, E. Adalsteinsson, V. K. Goyal, and L. L. Wald. *Sparsity-Enforced Joint Trajectory and RF Excitation Pulse Design*. US Patent Pending, Internal Case No. MGH 3726 / MIT 13074, 2008.

- A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald. Sparsity in MRI RF Excitation Pulse Design. In *Proc Conf Information Sciences and Systems*, pages 252–257, Princeton, NJ, March 2008.

- A. C. Zelinski, K. Setsompop, V. Alagappan, V. K. Goyal, L. L. Wald, and E. Adalsteinsson. *In Vivo* B_1^+ Inhomogeneity Mitigation at 7 Tesla Using Sparsity-Enforced Spatially-Tailored Slice-Selective Excitation Pulses. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 620, Toronto, Canada, 2008.

- A. C. Zelinski, V. K. Goyal, L. L. Wald, and E. Adalsteinsson. Sparsity-Enforced Joint Spiral Trajectory & RF Excitation Pulse Design. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1303, Toronto, Canada, 2008.

- A. C. Zelinski, K. Setsompop, V. K. Goyal, V. Alagappan, U. Fontius, F. Schmitt, L. L. Wald, and E. Adalsteinsson. Designing Fast 3-D RF Excitations by Optimizing the Number, Placement, and Weighting of Spokes in k-Space via a Sparsity-Enforcement Algorithm. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1691, Berlin, Germany, 2007.

Chapter 6 confronts a subset selection problem that arises when a parallel excitation system has more transmit modes available than hardware transmit channels with which to drive them. Here we show the applicability of MSSO theory and propose a fast target-dependent sparsity-enforced subset selection algorithm that explicitly accounts for the desired excitation pattern when choosing the mode subset, in contrast with principal component and covariance analysis methods that only use the spatial profiles of the transmit modes and thus determine only a single mode subset for all excitations. In one simulated experiment, the proposed fast algorithm actually finds the optimal solution to the underlying NP-Hard combinatoric subset selection problem. This work appeared at a conference and is also undergoing patent processing:

- A. C. Zelinski, V. Alagappan, V. K. Goyal, E. Adalsteinsson, and L. L. Wald. Sparsity-Enforced Coil Array Mode Compression for Parallel Transmission. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1302, Toronto, Canada, 2008

- A. C. Zelinski, L. L. Wald, V. Alagappan, V. K. Goyal, and E. Adalsteinsson. *Sparsity-Enforced Coil Array Mode Compression for Parallel Transmission.* US Patent Pending, Internal Case No. MGH 3673 / MIT 13014, 2008.

Chapter 7 investigates the behavior of whole-head and local SAR as a function of trajectory acceleration factor and target excitation pattern due to the parallel transmission of spatially-tailored excitations at a high field strength of 7 Tesla. Finite-difference time-domain simulations in a multi-tissue head model are used to obtain B_1^+ and electric field maps of an eight-channel transmit head array. Local and average SAR produced by 2-D spiral-trajectory excitations are examined as a function of trajectory acceleration factor and a variety of target excitation parameters when pulses are designed for constant root-mean-square excitation pattern error. To rapidly calculate local SAR, we develop a fast algorithm with a small memory footprint. Mean and local SAR are shown to vary by orders

of magnitude depending on acceleration factor and excitation pattern, often exhibiting complex, non-intuitive behavior. Furthermore, we find that the ratio of local to global SAR is often large, which implies that local SAR, rather than global SAR, is the limiting factor of eight-channel parallel transmission at 7T. These results suggest that the validation of individual target patterns and corresponding pulses is necessary to enable routine clinical use of parallel transmission systems. This project culminated in a journal article and two conference papers:

- A. C. Zelinski, L. M. Angelone, V. K. Goyal, G. Bonmassar, E. Adalsteinsson, and L. L. Wald. Specific Absorption Rate Studies of the Parallel Transmission of Inner-Volume Excitations at 7 Tesla. *J Magn Reson Imag*, In Press, 2008.

- A. C. Zelinski, L. M. Angelone, V. K. Goyal, G. Bonmassar, E. Adalsteinsson, and L. L. Wald. Specific Absorption Rate Studies of the Parallel Transmission of Inner-Volume Selective Excitations at 7 Tesla. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1315, Toronto, Canada, 2008.

- A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, L. L. Wald. Fast, Accurate Calculation of Maximum Local N-Gram Specific Absorption Rate. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1188, Berlin, Germany, 2008.

Chapter 8 proposes several ways to reduce the maximum local SAR produced by parallel excitation pulses, since Chapter 7 shows that local SAR is a major safety concern of such pulses. We first pose a linear-algebraic formulation to evaluate whole-head or local Ng SAR, showing that local Ng SAR at any location may be computed using a highly-sparse, redundant block-diagonal matrix; this generalizes the mean-SAR matrix given in [168]. We then introduce a method to explore excitation fidelity, mean SAR and pulse duration tradeoffs, pose a constrained optimization problem that ensures local Ng SAR meets certain constraints, and discuss the computational implications of such an optimization. This work appeared as two conference articles:

- A. C. Zelinski, K. Setsompop, V. Alagappan, B. A. Gagoski, L. M. Angelone, G. Bonmassar, U. Fontius, F. Schmitt, E. Adalsteinsson, and L. L. Wald. Pulse Design Methods for Reduction of Specific Absorption Rate in Parallel RF Excitation. In

Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM), page 1698, Berlin, Germany, 2007.

- A. C. Zelinski, V. K. Goyal, L. M. Angelone, G. Bonmassar, L. L. Wald, and E. Adalsteinsson. Designing RF Pulses with Optimal Specific Absorption Rate (SAR) Characteristics and Exploring Excitation Fidelity, SAR, and Pulse Duration Tradeoffs. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1699, Berlin, Germany, 2007.

Chapter 9 also focuses on SAR-reduced parallel transmission, introducing the novel concept time-multiplexing a set of pulses that each produce approximately the same excitation pattern yields a lower maximum local SAR than does transmitting any individual pulse over many repetition times. We then present an algorithm for determining the optimal multiplexing scheme (in the lowest maximum local SAR sense) when given a set of candidate pulses, demonstrating the technique via simulations of a 7-Tesla eight-channel parallel transmission system. We plan to submit this work to the following journal:

- A. C. Zelinski, V. K. Goyal, and E. Adalsteinsson. Reduction of Maximum Local Specific Absorption Rate via Pulse Multiplexing. *J Magn Reson Imag*, In Preparation for Submission, 2008.

Chapter 10 summarizes the contributions this thesis makes to the MRI and sparse approximation communities and outlines directions for future research.

Manuscript structure. Applied mathematicians interested in the general MSSO problem should read the background on sparse approximation given in Chapter 2 and then turn their attention to Chapter 3. The MRI applications and methods that rely in part on MSSO theory and algorithms are located in Chapter 5 and Chapter 6. MRI excitation pulse designers will be most interested in Chapter 5 (and to some extent, Chapter 6), whereas RF safety researchers interested in the study and reduction of SAR due to parallel transmission should refer to Chapter 7, Chapter 8, and Chapter 9.

Chapter 2

Background

2.1 Magnetic Resonance Imaging

Here we provide an overview of magnetic resonance imaging (MRI). While comprehensive descriptions of MRI may be found in, e.g., [12, 89], what follows is a brief introduction to the modality to provide adequate context for discussing MRI excitation pulse design. The most important concept to take away from this chapter is how pulse design, under a linear approximation, involves choosing to deposit energy along a 1-D contour in a continuous, 3-D, Fourier-like "k-space" in order to form a desired excitation in the spatial domain.

2.1.1 System Overview

Fig. 2-1 depicts a prototypical MRI system that consists of:

- A strong *magnetic main field* that in most cases is generated by driving a large DC current through superconducting coils. This field, referred to as "B_0" (and of strength B_0 in units of Tesla), is directed from the feet to the head and defines the spatial z axis. When a subject is placed within the bore of the magnet, a small proportion of hydrogen atoms (H^+) within the subject transition into a steady state, aligning fully with B_0 and precessing at the Larmor frequency, $\omega_0 = \gamma B_0$ (rad/sec), where γ is the *gyromagnetic ratio*, a known physical constant (Hz/T). The gyromagnetic ratio of H^+ is 42.576 MHz/T.
- A set of *gradient coils* that are able to impart controlled spatially-linear variations (gradients) on the z-directed B_0 field as a function of x, y, or z when driven with

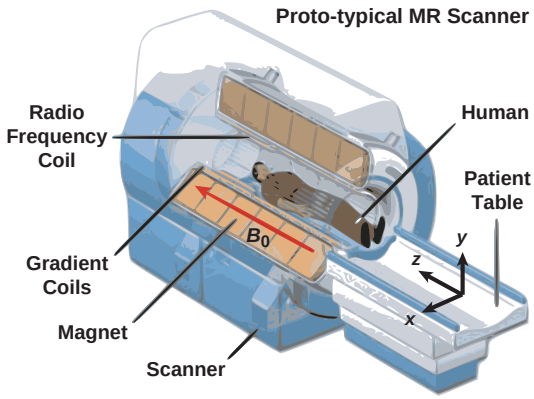

Figure 2-1: **Cutaway view of a prototypical MRI scanner**. The system's main field, B_0, is oriented along the spatial z axis (directed from the feet to the head). In the steady resting state, nuclear spins within the subject align with this main field and precess at the Larmor frequency, $\omega_0 = \gamma B_0$ (rad/sec). [Image ©2008 National Magnetic High Field Laboratory, Tallahassee, FL, USA. Used with permission.]

the proper currents. The continuously-valued gradient waveforms are denoted $\mathbf{G}(t) = [G_x(t), G_y(t), G_z(t)]^{\mathrm{T}}$, where $^{\mathrm{T}}$ is the vector transpose, and are typically in units of mT/m. The amplitude and slew rates of these waveforms are limited by hardware switching time and amplitude constraints. We will show shortly that these constraints limit the speed at which we may traverse a contour through excitation k-space and thus are an impediment to designing short-duration pulses.

- A *radio-frequency transmission coil* whose field is oriented perpendicularly to the static main field and tuned to the Larmor frequency. This coil is able to influence nuclear spins in the subject when driven by an RF voltage waveform, $b(t)$, or a current waveform, $a(t)$, placed on top of a carrier waveform. In almost all cases, a *quadrature coil* [126] is employed: one that is able to influence both the x and y components of the effective magnetic field simultaneously. We use complex variables to represent these two orthogonal components at each instant in time, which allows us to treat $b(t)$ and $a(t)$ as complex-valued time-varying signals.
- A *reception coil* that is able to detect, by induction, precessing *transverse components* of spins that have deviated from the steady state. Transverse components are the components of spins that lie in the (x, y) plane orthogonal to the spin component along the z axis. Reception hardware is often integrated directly with the transmission coil.

- A *digital interface* that is able to send trains of waveform samples to the gradient and RF coils, and that is also able to sample and store inductive signals sensed by the receive coil.

2.1.2 Larmor Precession and an Illustrative Thin-Slice-Imaging Example

Now let us consider an imaging scenario where our goal is to image the density of protons within a thin slice of the subject's brain. To begin, a subject lies motionless within the scanner; at all locations, proportions of magnetic spins align with z, yielding net magnetizations solely along the z axis. The gradient coils are dormant, and the reception coil, only capable of detecting transverse (non-z-directed) components of oscillating spins, senses only thermal noise. Suddenly, a specially-crafted RF pulse is sent through the transmission coil along with a judiciously-chosen set of gradient waveforms. The pulse excites only those spins within a thin slice of tissue within the subject's brain, most often only those spins within $(z_0 - \delta/2, z_0 + \delta/2)$, producing a *single-slice excitation.* Spins within this δ-mm-thick slice are tilted away from their steady-state z-directed position, while those outside the slice are left in perfect alignment with the main field. Assume here that spins within the slice are all tilted *uniformly* by a small angle; e.g., by 30 degrees at all $(x, y, z = z_0 \pm \delta/2)$ locations. These *small-tip-angle* spins precess at the Larmor frequency, rotating around the z axis at the angular rate w_0, all while returning to the z-directed steady state position.

We note here that spatial flip-angle uniformity during excitation is crucial: if spins are tilted non-uniformly across space, the intensity and contrast of the resulting image will no longer correspond to the actual density of underlying spins and the image will contain contrast and SNR non-uniformities that erode its diagnostic quality.

The transient behavior of a prototypical excited magnetic spin is illustrated in Fig. 2-2. The rate at which the spin's z-component recovers is modeled well by an exponential time constant, denoted T_1, and the rate at which its transverse components decay is modeled well by an exponential time constant denoted T_2. These constants depend on tissue type as well as field strength, e.g., white brain matter has $\mathrm{T}_1 \approx 780$ ms and $\mathrm{T}_2 \approx 80$ ms when $B_0 = 1.5$ Tesla [12].

As the transverse components decay, they induce a current in the receive coil; this is referred to as *free induction decay* (FID). During this brief time, the system enables its analog-to-digital converter and samples the FID signal, $s(t)$. The Fourier transform of

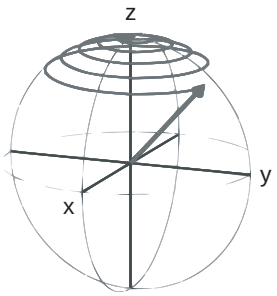

Figure 2-2: **Larmor precession and spin recovery**. A spin is excited by an RF pulse and tilted away from the z axis. After the transmission of the pulse, it then returns to its steady-state position, precessing at a rate $\omega_0 = \gamma B_0$ when in the presence of a main field of strength B_0. The rate at which the longitudinal z component recovers, and at which the transverse components decay, is modeled well by the tissue-dependent exponential time constants T_1 and T_2. Here, the spin has only been tilted by a small angle away from z; this is called *small-tip-angle excitation*.

this waveform is then computed, but because all spins within the thin slice of tissue have precessed at ω_0, no spatial information is obtained: the magnitude of the Fourier component corresponding to ω_0 simply indicates the density of excited spins within the entire slice. The receive coil has done little more than provide us with the integral of all detectable spins; only the "DC" component of the image we wish to generate has been determined.

What must be done to produce a useful spatial image of spin proportions within the thin slice? The late Paul Lauterbur conceived of the solution [87]: induce a gradient along the main field during the excitation process and make spins at different spatial locations precess at different rates; this lets the readout coil observe a signal with a spread of frequency components, each corresponding to spin proportions at a particular point in space. Consider, for example, applying a gradient of magnitude G_x such that the z-directed B_0 field varies with x, i.e., $B(x) = B_0 + xG_x$. Now a spin at x precesses at rate $w(x) = \gamma B(x)$ after applying the RF pulse, rather than simply at rate ω_0. The signal observed by the readout coil is now comprised of various significant frequency components whose magnitudes provide spatially-dependent spin density information. Extending this process to two dimensions (using both x and y gradients) permits the readout and reconstruction of a 2-D image whose intensity at (x, y) corresponds with the density of spins at $(x, y, z_0 \pm \delta/2)$.

The scenario outlined above—namely, exciting a thin slice of tissue—is referred to as "thin-slice excitation" and is the predominant type of MRI excitation conducted today. This type of excitation dominates the field because it simplifies the readout stage by permitting

the system to record and reconstruct an image from only two dimensions of data. One topic that has not yet been explained in this subsection is how to design an RF pulse that uniformly excites spins within only a thin slice of tissue. The creation of such *spatially-selective, spatially-tailored* pulses is the focus of our work. The design problem is essentially one of choosing $\mathbf{G}(t)$ and $b(t)$ such that spins across (x, y, z) undergo a user-defined excitation.

2.1.3 The Bloch Equation

Let us now back away from the illustrative example and simply consider a cluster of many magnetic spins within a tiny voxel centered about $\mathbf{r} = [x, y, z]^{\mathrm{T}}$ in space. Here we will employ vector notation to differentiate between field and spin components along the x, y, and z axes; to begin, let $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ be unit vectors along these respective axes, and let us neglect the effect of the time constants T_1 and T_2. The discussion here is largely based upon Ch. 5 and Ch. 6 of [98] and Sec. 1 of [102].

In the absence of a main field, the many spins within the voxel are oriented at random, and the net magnetizations along each axis, $M_x(\mathbf{r})$, $M_y(\mathbf{r})$, and $M_z(\mathbf{r})$, are essentially zero. We may represent these with a vector: $\mathbf{M}(\mathbf{r}) = [M_x(\mathbf{r}), M_y(\mathbf{r}), M_z(\mathbf{r})]^{\mathrm{T}}$. An inductive coil that is able to detect transverse components will thus observe no significant magnetization along the x and y axes.

However, in the presence of the z-directed main field $\mathbf{B}_0 = B_0\mathbf{z}$, a fraction of the magnetized spins aligns with this field, creating a non-zero magnetized equilibrium component along z, denoted M_0. But because the many spins in the tiny voxel that do not align with the main field still have random orientations, the net magnetization along x and y remains zero and spins within the tiny voxel remain undetected by the reception coil. In this situation, $\mathbf{M}(\mathbf{r}) = \mathrm{M}_0\mathbf{z}$.

Now let us apply a time-varying excitation pulse through the transmission coil, injecting the following circularly-polarized signal into the RF coil: $\mathbf{B}_1(t) = B_1(t)e^{-j\omega_0 t}$, where $B_1(t)$ may be viewed as a complex-valued modulation term (a waveform with both magnitude and phase) and ω_0 as a carrier frequency. This RF signal influences the x and y components of the overall magnetic field at $\mathbf{r}$ and impacts a small fraction of the magnetic spins in the region of interest. If we consider a *reference frame* that rotates about z at angular velocity $\omega = \gamma B$, the net magnetic spin components obey the following continuous-time

three-dimensional differential Bloch equation [14]:

$$\frac{d\mathbf{M}(\mathbf{r},t)}{dt} = \gamma\mathbf{M}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t) - \omega \times \mathbf{M}(\mathbf{r},t), \tag{2.1}$$

where $\times$ is the vector cross product, $\mathbf{B}(\mathbf{r},t)$ is the magnetic field at $\mathbf{r}$ (in this case equal to $\mathbf{B}_0 + \mathbf{B}_1(t)$), and the rightmost term accounts for the rotation about z. Essentially, (2.1) describes how the components of the magnetization vector $\mathbf{M}$ evolve over time due to the application of the RF pulse in the presence of the main field (hence $\mathbf{M}$'s new dependence on time) [98].

Now assume we drive each gradient coil with a time-varying waveform to modulate the z component of the magnetic field at $\mathbf{r}$. We express the gradient as the vector $\mathbf{G}(t) = G_x(t)\mathbf{x} + G_y(t)\mathbf{y} + G_z(t)\mathbf{z}$, so at location $\mathbf{r}$ and time t its impact on the magnetic field is $(\mathbf{G}(t) \cdot \mathbf{r})\mathbf{z} = [0, 0, G_x(t)x + G_y(t)y + G_z(t)z]^{\mathrm{T}}$, where $\cdot$ is the vector dot product. Thus the overall effective field at $\mathbf{r}$ and time t becomes $\mathbf{B}(\mathbf{r},t) = \mathbf{B}_0 + \mathbf{B}_1(t) + (\mathbf{G}(t) \cdot \mathbf{r})\,\mathbf{z}$, and (2.1) becomes:

$$\frac{d}{dt}\begin{pmatrix} M_x(\mathbf{r},t) \\ M_y(\mathbf{r},t) \\ M_z(\mathbf{r},t) \end{pmatrix} = \gamma \begin{pmatrix} 0 & \mathbf{G}(t)\cdot\mathbf{r} & -B_{1,y}(t) \\ -\mathbf{G}(t)\cdot\mathbf{r} & 0 & B_{1,x}(t) \\ B_{1,y}(t) & -B_{1,x}(t) & 0 \end{pmatrix} \begin{pmatrix} M_x(\mathbf{r},t) \\ M_y(\mathbf{r},t) \\ M_z(\mathbf{r},t) \end{pmatrix}, \tag{2.2}$$

where $B_{1,x}(t)$ and $B_{1,y}(t)$ are the demodulated orthogonal time-varying real-valued components of the applied RF field that are produced by the quadrature transmission coil [102].

2.1.4 Time Constant Equations

Before studying (2.2), here we simply note the behavior of $\mathbf{M}(\mathbf{r},t)$ when $\mathrm{T}_1(\mathbf{r})$ and $\mathrm{T}_2(\mathbf{r})$ are accounted for. Assuming that an excitation pulse is applied prior to time $t = 0$ and ends at $t = 0$, the longitudinal component returns to its equilibrium state based on $\mathrm{T}_1(\mathbf{r})$ as follows:

$$M_z(\mathbf{r},t) = M_z(\mathbf{r},0^+)e^{-t/\mathrm{T}_1(\mathbf{r})} + \mathrm{M}_0(\mathbf{r})(1 - e^{-t/\mathrm{T}_1(\mathbf{r})}), \tag{2.3}$$

where $M_z(\mathbf{r},0^+)$ is the value of the longitudinal component immediately after the excitation pulse has finished being transmitted. If we represent the transverse components $M_x(\mathbf{r},t)$ and $M_y(\mathbf{r},t)$ via the complex-valued variable $m(\mathbf{r},t) = M_x(\mathbf{r},t) + jM_y(\mathbf{r},t)$, then the rate

of decay of the transverse components is governed by $\mathrm{T}_2(\mathbf{r})$ as follows:

$$m(\mathbf{r},t) = m(\mathbf{r},0^+)e^{-t/\mathrm{T}_2(\mathbf{r})}e^{-j\omega(\mathbf{r})t}. \tag{2.4}$$

This is a decaying complex-valued sinusoid. Here, $\omega(\mathbf{r}) = \gamma B(\mathbf{r})$ and $B(\mathbf{r})$ is simply the z-directed magnetic field strength at location $\mathbf{r}$. The main point here is that after exciting a desired region, the transverse components decay away as modeled by T_2.

2.1.5 The Small-Tip-Angle Assumption and Excitation k-Space

We now continue our study of the expanded Bloch equation given in (2.2). Let us assume that the RF energy deposited by the system tilts the z component of the net magnetization by only a small angle at location $\mathbf{r}$ and thus $M_z(\mathbf{r},t) \approx \mathrm{M}_0$, the latter being the original constant longitudinal steady-state net magnetization when only the main field is active. Imposing this assumption *decouples* the third component of (2.2) from the other two.[1] If we define the transverse magnetization as a complex-valued variable $m(\mathbf{r},t) = M_x(\mathbf{r},t) + jM_y(\mathbf{r},t)$ and the applied RF pulse as the complex-valued time waveform $b(t) = B_{1,x}(t) + jB_{1,y}(t)$, then according to Pauly et al. [102], the x and y components of (2.2)—those components visible to the reception coil—may be structured as a single complex-valued differential equation:

$$m(\mathbf{r},t) = -j\gamma\left(\mathbf{G}(t)\cdot\mathbf{r}\right)m(\mathbf{r},t) + j\gamma b(t)\mathrm{M}_0. \tag{2.5}$$

Recalling that in the steady-state $\mathbf{M}(\mathbf{r}) = \mathrm{M}_0\mathbf{z}$, and imposing this as an initial condition (i.e., that $m(\mathbf{r},0) = 0$), Pauly et al. solve (2.5) to yield the final magnetization at time L [102]:

$$m(\mathbf{r},L) = m(\mathbf{r}) = j\gamma\mathrm{M}_0\int_0^L b(t)\exp\left(-j\gamma\mathbf{r}\cdot\int_t^L \mathbf{G}(s)ds\right)dt \tag{2.6}$$

This is the core equation that motivates the interpretation of RF pulse design as depositing energy in excitation k-space (a spatial frequency domain) to tailor the resulting magneti-

[1] It is worth noting that it has been empirically shown that this assumption holds extremely well for up to a 60-degree tip angle and approximately for up to a 90-degree angle, breaking down for flip angles nearing 180 degrees [102].

zation $m(\mathbf{r})$ in a desired way. To see this, define $\mathbf{k}(t)$ as follows [102]:

$$\mathbf{k}(t) = -\gamma \int_t^L \mathbf{G}(s)ds, \tag{2.7}$$

and then substitute (2.7) into (2.6) to obtain:

$$m(\mathbf{r}) = j\gamma \mathrm{M}_0 \int_0^L b(t) \exp\left(j\mathbf{r} \cdot \mathbf{k}(t)\right) dt. \tag{2.8}$$

This results in a striking change: (2.8) is now simply a 3-D Fourier transform to within a multiplicative constant.

One may interpret (2.8) as follows: $\mathbf{k}(t)$ is a 1-D contour that traces through a 3-D continuous spatial frequency domain over time; RF energy as represented by the complex-valued time-varying signal $b(t)$ is deposited along this contour at different frequencies, and a Fourier transform of this energy describes the approximate transverse magnetization pattern arising in the spatial domain. We see too the explicit dependence of the trajectory on the gradients: the *negated remaining running time integral* of $\mathbf{G}(t)$ from time t to L determines our location in k-space at time t. Because the gradient waveforms are driven by real hardware, and because such hardware has its own amplitude, slew rate, and switching time constraints, this means that the rate and extent to which we may traverse a contour through k-space is itself constrained.

We see now that our desire to obtain short-duration excitation pulses (as discussed in Ch. 1) is directly at odds with the gradient-related impediments on k-space traversal. This is why the design of efficient, short trajectories through k-space (and RF waveforms to accompany them) remains an open, high-impact problem even nearly 20 years after Pauly first proposed excitation k-space.

2.1.6 Data Readout and Image Reconstruction

We briefly alluded to post-excitation data readout and image reconstruction in the thought experiment of Sec. 2.1.2. From here onward we will simply assume that if one forms a high-fidelity version of the desired excitation, the readout stage is able to be accomplished. There is a great deal of literature discussing MRI readout theory that details many styles of data collection and image reconstruction, e.g., [12, 89, 98].

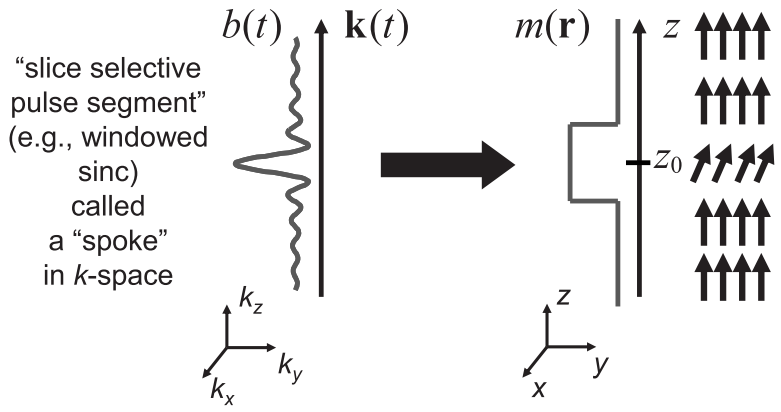

Figure 2-3: **Single-spoke slice-selective excitation example.** A sinc-like RF pulse segment is played over a period of time where the gradients are constructed such that the trajectory proceeds through a straight line along the k_z axis in excitation k-space. Due to the small-tip-angle Fourier-like relation given in (2.8), this sinc-like energy along k_z yields a box-shaped excitation profile along z, i.e., it primarily excites only those spins within the small region $z_0 \pm \delta$, such that $m(x, y, z)$ is a constant positive value for $z \in (z_0 - \delta/2, z_0 + \delta/2)$ and zero otherwise.

2.1.7 Spoke-Trajectory Pulses for Thin-Slice Excitation

We now briefly revisit the thin-slice excitation discussion of Sec. 2.1.2 and focus on a class of slice-selective pulses comprised of k-space trajectory segments that resemble sine cardinals (sincs). We refer to these pulse segments as "spokes" [108, 113, 132] because as each is played, its trajectory in k-space is a straight line. In the small-tip-angle regime, based on the relation given in (2.8), a rectangle-like slice profile along the z axis is achieved by placing a sinc-like RF pulse segment (a spoke) in the k_z direction of excitation k-space. In practice, a true sinc along k_z is replaced by a finite-length, sinc-like waveform [113]. The time-bandwidth product and k_z-extent of the segment influence the thickness and transition edges of the slice; such Fourier properties as well as others hold in this situation due to (2.8).

Single-spoke pulses. Fig. 2-3 provides a simple illustrative example of a one-spoke pulse, the accompanying RF waveform, and the resulting transverse magnetization pattern. The reason spoke-based pulses are effective at exciting thin slices is because they spend the majority of their time (and deposit energy) throughout moderate-to-high k_z frequencies, which allows them to strongly influence the magnetization $m(\mathbf{r})$ along the spatial z axis.

Multi-spoke pulses. Playing a single sinc-like spoke at $(k_x, k_y) = (0, 0)$ (DC) in k-space is one way to excite a thin slice, and it is the traditional method of slice-selective excitation. Depositing energy at a single frequency in (k_x, k_y), however, does not provide

any ability to tailor the resulting *in-plane* (x, y) spatial excitation pattern when using a single-channel excitation system; even with a multi-channel excitation system, significant tailoring of the in-plane excitation is not possible with this approach. Thus in order to strongly influence the in-plane excitation flip angle while simultaneously maintaining slice selectivity, one must place a number of spokes at various locations in (k_x, k_y), modulating the amplitude and phase of each to form a desired in-plane transverse magnetization; here, the complex weightings in (k_x, k_y) form the in-plane excitation, while the sinc-encoded k_z traversals provide slice selectivity in z. Simultaneous slice selection and in-plane tailoring are possible due to the separability of the 3-D Fourier transform in (2.8) that relates the spatial excitation to the deposition of energy in k-space. Unfortunately, using multiple spokes has the negative consequence of increasing pulse duration. Thus an ideal spoke-based pulse is one that not only achieves the user-specified in-plane excitation, but does so using as few spokes as possible. (We will confront this problem in Ch. 5 by using a sparse-approximation-inspired technique to optimize spoke placements in the (k_x, k_y) plane.)

2.1.8 Spiral-Trajectory Pulses for Structured In-Plane Excitation without Slice Selectivity

Now let us consider a different trajectory, one that spends all of its time spiraling within the (k_x, k_y) plane and never travels away from $k_z = 0$. No matter how much RF energy we place along this trajectory, its energy deposition is simply a Dirac delta with respect to k_z frequencies, which means that its excitation along the spatial z axis will be nonzero along z everywhere $m(x, y) \neq 0$. In other words, a 2-D spiral trajectory—as well as any other trajectory that fails to deposit energy at nonzero k_z frequencies—is incapable of exciting spins within only a thin slice along z. The upside of spiral trajectory pulses (and other 2-D trajectories) is that they are able to highly-structure the in-plane magnetization $m(x, y)$ because they travel out to and may deposit energy at moderate-to-high k_x and k_y frequencies. Fourier properties hold here: for example, the radial spacing of the spiral rings determines the extent to which aliasing may occur within the field of view, and the maximum frequencies traversed by the spiral inherently limit the resolution of the excitation we may form. When designing a spiral trajectory, one must account for these properties to avoid aliased excitations or patterns too low in resolution.

Figure 2-4 illustrates a spiral trajectory, the gradients that implement the spiral, and

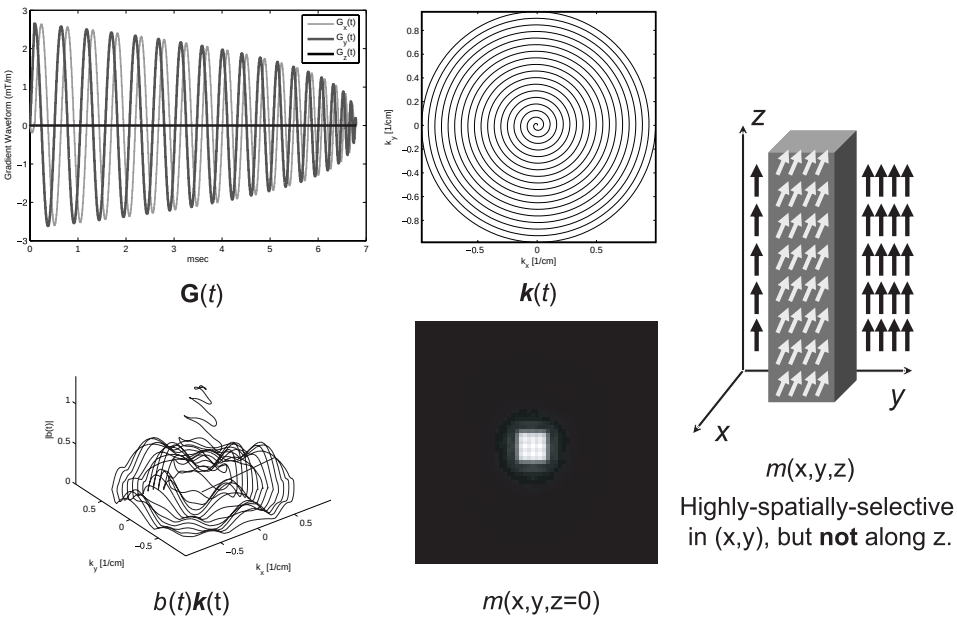

Figure 2-4: **Spiral-trajectory excitation example**. The gradients and (k_x, k_y)-space trajectory are given, along with the RF waveform, $b(t)$, deposited by the system as the trajectory is traversed. Within the (x, y) plane, a highly-structured box forms (due to energy placed at high k_x and k_y frequencies), but when viewing the resulting magnetization along z, the pulse fails to excite spins within only a thin slice.

the pulse magnitude $b(t)$ one may play along the trajectory to yield a highly-structured box-shaped excitation in the (x, y) spatial plane, but due to the fact that no energy is placed away from $k_z = 0$, we see that there is no spatial selection in z.

2.1.9 Non-Idealities of Excitation: B_1^+ and Main Field Inhomogeneity

When the small-tip-angle approximation is applied to a realistic single-transmit-channel system, we arrive at the following equation [102, 146] rather than at (2.8):

$$m(\mathbf{r}) = j\gamma \mathrm{M}_0 S(\mathbf{r}) \int_0^L b(t) e^{j\Delta B_0(\mathbf{r})(t-L)} e^{j\mathbf{r}\cdot\mathbf{k}(t)}\, dt = S(\mathbf{r})p(\mathbf{r}). \tag{2.9}$$

Notice that (2.9) is identical to (2.8) except for the presence of $S(\mathbf{r})$ and $e^{j\Delta B_0(\mathbf{r})(t-L)}$. Here, $m(\mathbf{r})$ is in radians, t in seconds, and γ in rad/T/s. The implicitly defined term, $p(\mathbf{r})$, will be discussed shortly.

Spatial transmit B_1^+ profile. The first new variable, $S(\mathbf{r})$, is complex-valued, varies across space, and is referred to as the *spatial transmit profile* of the transmission channel. In other words, $S(\mathbf{r})$ is essentially a *spatial basis function* that conveys how capable the coil is

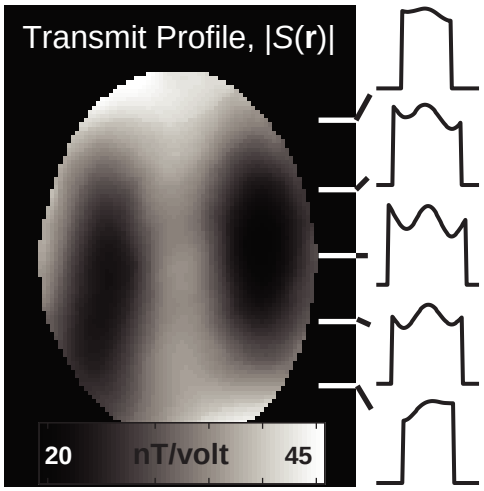

Figure 2-5: **Single-channel system spatial B_1^+ transmit profile in a water phantom at 7T**. $|S(x, y, z_0)|$ is depicted in units of nT/V, varying significantly with $\mathbf{r}$ due to the B_1^+ inhomogeneity phenomenon. It will be difficult to produce a uniform flip angle across space within a thin slice without using a sophisticated pulse (e.g., a multi-spoke pulse).

at exciting spins across space. For example, if $S(\mathbf{r}) = 0$, then the transmit coil is incapable of tilting spins at $\mathbf{r}$. Here, $S(\mathbf{r})$ is in units of T/V.[2] The true term for $S(\mathbf{r})$ is "B_1^+ map", "B_1^+ profile", or "$B_1^+(\mathbf{r})$", but we use $S(\mathbf{r})$ to avoid confusion with the system's reception profile, denoted B_1^-, and to simplify notation.

B_0 **main field inhomogeneity**. The second new variable, $\Delta B_0(\mathbf{r})$, is a *field map* of B_0 inhomogeneity in radians/second. Essentially, the system's main field is never perfectly equal to B_0 everywhere. When considering this non-ideality from a k-space perspective, it causes a phase accrual over time at each location $\mathbf{r}$, which is handled by the $e^{j\Delta B_0(\mathbf{r})(t-L)}$ term in (2.9). A main field inhomogeneity map is easy to acquire on a modern system and its phase accrual effect may be mitigated by simply incorporating it into the design process, as will be detailed shortly. B_1^+ inhomogeneity is a significantly worse problem.

Nominal excitation. The implicitly defined term of (2.9), $p(\mathbf{r})$, is the *nominal excitation* that forms before accounting for $S(\mathbf{r})$. In other words, $p(\mathbf{r})$ is the excitation that *would* form across space if $S(\mathbf{r}) = 1$ for all $\mathbf{r}$.

Detrimental effect of B_1^+ inhomogeneity at high field. At high magnetic field strengths, where B_1^+ inhomogeneity [16] is severe, $S(\mathbf{r})$ will vary significantly with $\mathbf{r}$ [29, 133, 136]. For example, Fig. 2-5 shows the in-plane $|S(x, y, z_0)|$ that arises when a water

[2] If a current waveform $a(t)$ were substituted for the voltage waveform $b(t)$ in (2.9), $S(\mathbf{r})$ would be in units of T/A.

phantom (an anthropogenic object used for test imaging) is placed in a 7 Tesla scanner. When a low-flip-angle pulse is transmitted on a high-field system, its nominal excitation, $p(\mathbf{r})$, is multiplied (pointwise) by the inhomogeneous $S(\mathbf{r})$ to yield the actual magnetization, as given in (2.9). Applying a standard one-spoke slice-selective excitation, where $|p(x, y, z_0)| = 1$, thus results in a non-uniform in-plane flip angle magnitude, equal to $|S(x, y, z_0)p(x, y, z_0)| = |S(x, y, z_0)|$. This is extremely detrimental because a non-uniform flip angle across space causes all resulting images to have unwanted contrast and SNR variations across space, limiting their diagonistic quality.

In contrast, an *ideal mitigation pulse* produces a nominal excitation $p(\mathbf{r})$ such that $|S(\mathbf{r})p(\mathbf{r})|$ is constant for all $\mathbf{r}$ in some region of interest, i.e., the ideal $|p(\mathbf{r})|$ equals the pointwise inverse of $S(\mathbf{r})$, denoted $|S(\mathbf{r})|^{-1}$, to within a multiplicative constant. This pulse is ideal in the sense that it mitigates the magnitude of the inhomogeneity; it does not impose phase uniformity because the latter is not stringently required in most clinical imaging applications.

Recalling that thin-slice uniform excitation is fundamental for most imaging schemes, we see that in order to achieve such an excitation in the presence of inhomogeneity, one must not only place energy throughout moderate-to-high k_z frequencies to be selective about $z = z_0$, one must also judiciously place energy throughout (k_x, k_y) to tailor the nominal excitation to mitigate the inhomogeneous in-plane $S(x, y, z = z_0)$. In Ch. 5, we will enable high-field imaging by implementing sparse approximation techniques that design short-duration thin-slice B_1^+-mitigation pulses.

2.1.10 Single-Channel Excitation Pulse Design

We now outline a spatial-domain-based method to design an excitation pulse to achieve a desired excitation $d(\mathbf{r})$ across space. This method accounts for $S(\mathbf{r})$ and thus may be able to mitigate inhomogeneity in certain spatial regions if the trajectory covers the proper spatial frequencies. The approach here is largely due to [146].

Assumptions. First we assume that $S(\mathbf{r})$ and $\Delta B_0(\mathbf{r})$ are known at all points $\mathbf{r}$ within a particular region of interest, called the *field of excitation* (FOX). Further, let the gradient waveforms, $\mathbf{G}(t)$, be fixed, which in turn fixes the trajectory through k-space, $\mathbf{k}(t)$, due to the latter's linear dependence on the former as given in (2.7).

Discretization. We now discretize (2.9) by sampling space at locations $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$ within the FOX and sampling time at $t_1, \ldots, t_{N_t} \in [0, L]$; the t_i are spaced uniformly by Δ_t. This yields

$$\mathbf{m} = \mathbf{SFb} = \mathbf{Ab} \tag{2.10}$$

Here, $\mathbf{S} \in \mathbb{C}^{N_s \times N_s}$ is a diagonal matrix comprised of N_s samples of the spatial profile $S(\mathbf{r})$ taken within the user-defined FOX. The next matrix, $\mathbf{F} \in \mathbb{C}^{N_s \times N_t}$, brings energy placed in k-space into the spatial domain at the N_s locations where the coil profile is sampled within the FOX. Formally, $\mathbf{F}(m, n) = j\gamma \mathrm{M}_0 \Delta_t e^{j\Delta B_0(\mathbf{r}_m)(t_n - L)} e^{j\mathbf{r}_m \cdot \mathbf{k}(t_n)}$. Finally, $\mathbf{b} \in \mathbb{C}^{N_t}$ is comprised of samples of the RF waveform, $b(t)$, spaced by Δ_t [146].

To within aliasing constraints, users are free to choose how finely to sample the spatial and temporal variables in (2.9) and thus control whether $\mathbf{m} = \mathbf{Ab}$ is underdetermined or overdetermined [80,146]. Furthermore, by sampling $\mathbf{r}$ only within the FOX, (2.10) becomes blind to the excitation that forms outside of the desired FOX and is thus freed from needless spatial constraints.

Conventional pulse generation. To excite a pattern $d(\mathbf{r})$, we must determine an RF pulse to play through the transmit coil. To do this, one may form N_s samples of the desired pattern into the vector $\mathbf{d} \in \mathbb{C}^{N_s}$ and create a pulse that (approximately) achieves the desired excitation by solving

$$\mathbf{d} = \mathbf{SFb} \tag{2.11}$$

for $\mathbf{b}$. We might solve (2.11) via direct inversion (or pseudoinversion) of $\mathbf{SF}$, or perhaps instead solve a Tikhonov-regularized problem [123,124] that penalizes the ℓ_2 energy of $\mathbf{b}$:

$$\min_{\mathbf{b}} \ \{\|\mathbf{d} - \mathbf{SFb}\|_2^2 + \delta_{\text{tik}} \|\mathbf{b}\|_2^2\}, \tag{2.12}$$

where δ_{tik} is a small nonnegative regularization term. After solving (2.11) in some manner, we may extract the time samples of the pulse from $\mathbf{b}$ and play the pulse through a Bloch-equation simulator or the actual system. If (2.11) is solved accurately, the small-tip angle approximation is valid, and the SNR is sufficient, then the excitation that forms will resemble the target pattern, $d(\mathbf{r})$.

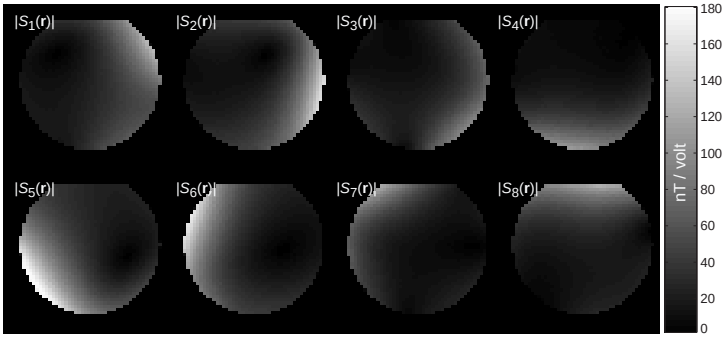

Figure 2-6: **Profile magnitudes of an eight-channel 3-Tesla parallel excitation MRI system**. Here the magnitudes of the $S_p(\mathbf{r})$s are depicted for $p = 1, \ldots, 8$; samples of each $S_p(\mathbf{r})$ may be taken within the nonzero region of influence (the chosen FOX) and stacked into the diagonal matrix $\mathbf{S}_p$. Across space, the $S_p(\mathbf{r})$s are not orthogonal—their regions of influence overlap each other to some extent.

2.1.11 Multi-Channel Parallel Excitation Pulse Design

Recently, systems have emerged whose transmit coils are comprised of $P > 1$ elements, each capable of independent, simultaneous transmission, e.g., [58, 79, 113, 130, 168, 169]. Because each of the P transmission elements has its own corresponding spatial transmit profile, denoted $S_p(\mathbf{r})$, and assuming the small-tip-angle assumption still holds, (2.9) extends as follows:

$$m(\mathbf{r}) = j\gamma \mathrm{M}_0 \sum_{p=1}^{P} S_p(\mathbf{r}) \int_0^L b_p(t) e^{j\Delta B_0(\mathbf{r})(t-L)} e^{j\mathbf{r}\cdot\mathbf{k}(t)}\, dt, \tag{2.13}$$

where $b_p(t)$ is the RF pulse played along the transmission coil's pth channel (V) [59]. If the gradients here are the same as in (2.9), the trajectory $\mathbf{k}(t)$ remains the same, but now there are P energy weightings being deposited along it; the pth weighting produces an excitation in the spatial domain (impacted by $S_p(\mathbf{r})$) and the *superposition* of the P individual excitations yields the overall excitation.

Multi-channel system transmit profiles. Fig. 2-6 depicts the coil profile magnitudes of an eight-channel system within an oil phantom at 3 Tesla. Spatially, the $S_p(\mathbf{r})$s are not orthogonal—their regions of influence overlap each other to some extent. If the strength of the main field is increased, the $S_p(\mathbf{r})$s will exhibit increased variation.

Conventional parallel pulse generation. Here we summarize the design method of [59]. Other pulse design approaches that rely on k-space but not the exact spatial-domain discretization steps given here are those of Katscher et al. [79] and Zhu et al. [168, 169].

To begin, suppose we have a P-channel system and want to produce a desired excitation $d(\mathbf{r})$. Formally, let us seek $b_p(t)$s in (2.13) such that the resulting $m(\mathbf{r})$ is close in the ℓ_2 sense to $d(\mathbf{r})$ for all $\mathbf{r}$ in a user-defined FOX. As in Sec. 2.1.10, a set of gradients is chosen to traverse a contour in k-space that is considered useful. We then discretize (2.13) in space at locations $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_s}$, and uniformly sampled in time at $t_1, t_2, \ldots, t_{N_f}$ (samples are spaced by Δ_t). This yields

$$\begin{aligned} \mathbf{m} &= \mathbf{S}_1\mathbf{F}\mathbf{b}_1 + \cdots + \mathbf{S}_P\mathbf{F}\mathbf{b}_P \\ &= [\mathbf{S}_1\mathbf{F}\cdots\mathbf{S}_P\mathbf{F}] \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_P \end{bmatrix} = \mathbf{A}_{\text{tot}}\mathbf{b}_{\text{tot}}, \end{aligned} \tag{2.14}$$

where $\mathbf{S}_p \in \mathbb{C}^{N_s \times N_s}$ is a diagonal matrix comprised of N_s samples of $S_p(\mathbf{r})$ taken within the FOX. The next matrix, $\mathbf{F} \in \mathbb{C}^{N_s \times N_t}$, is an operator that brings energy placed in k-space into the spatial domain at the N_s locations where each coil profile is sampled and is identical to the one given in Sec. 2.1.10. Finally, each $\mathbf{b}_p \in \mathbb{C}^{N_t}$ contains samples of $b_p(t)$ [59].

A set of P RF pulses that (approximately) achieves the desired excitation may now be generated by solving

$$\mathbf{d} = \mathbf{A}_{\text{tot}}\mathbf{b}_{\text{tot}} \tag{2.15}$$

for $\mathbf{b}_{\text{tot}}$, where $\mathbf{d} \in \mathbb{C}^{N_s}$ is constructed by sampling $d(\mathbf{r})$. Solving (2.15) via Tikhonov regularization is one way to find a well-conditioned solution $\mathbf{b}_{\text{tot}}$, i.e.,

$$\min_{\mathbf{b}_{\text{tot}}} \{\|\mathbf{d} - \mathbf{A}_{\text{tot}}\mathbf{b}_{\text{tot}}\|_2^2 + \delta\|\mathbf{b}_{\text{tot}}\|_2^2\}. \tag{2.16}$$

After solving the problem, samples of each of the P RF pulses may now be extracted from $\mathbf{b}_{\text{tot}}$, played through the P-channel system, and used to produce an excitation close to $d(\mathbf{r})$.

Undersampling. The presence of multiple excitation elements allows one to *undersample* an excitation trajectory relative to a conventional Nyquist-sampled trajectory and yet often still form a high-fidelity version of the desired excitation. Undersampling is greatly beneficial: it reduces the distance traveled in k-space and thus the duration of the corresponding pulse.

Unaccelerated vs. accelerated spiral-trajectory eight-channel design exam-

ple. Fig. 2-7 illustrates details of the design process when generating parallel excitation pulses to produce a 15-degree, 28-mm $\times$ 28-mm target box pattern with fixed in-box and overall fidelity constraints. Spirals are "accelerated" by factors of R, which means that the radial separation between spiral samples is increased R-fold relative to a Nyquist-sampled $R = 1$ design. The columns, from left to right, show design details for $R = 1$, 4, and 7 trajectories. The first and second rows depict the progressively-undersampled spiral trajectories fed to the design algorithm in each case and the gradients of each trajectory. The third row depicts the magnitude of the eighth channel's RF pulse shape, $|a_8(t)|$. (In the interest of space, only one of the eight RF pulse shapes in each case is shown). Finally, the bottom row shows the resulting excitations after Bloch-simulating the gradients and pulses; each pattern has the same fidelity. As R increases, the pulses grow shorter in duration as intended, and the eighth channel's RF pulse changes greatly in both magnitude and shape. The $R = 7$ pulse shape has a 100-times greater peak magnitude than the $R = 1$ pulse shape, a natural consequence of the fact that with only an $R = 7$ spiral, there are very few degrees of freedom remaining in k-space with which to form the excitation relative to the Nyquist-sampled $R = 1$ spiral, which forces the system to drive the channel profiles intensely with high-amplitude RF pulse shapes in order to form the desired pattern. This makes intuitive sense: the lack of k-space freedom forces the system to rely heavily upon its degrees of freedom in the spatial domain.

Relation to parallel readout. Parallel transmission is the excitation counterpart to the concept of readout-side acceleration where the use of multiple *reception* coils permits one to undersample readout-side data and substantially reduce readout time [106, 115]; the development of parallel transmission arrays has lagged behind that of multi-channel reception arrays because unlike readout-side scenarios where adding an additional channel is simplistic and low in cost, adding a transmit channel requires an expensive RF power amplifier and adds greatly to the complexity of the MRI system's online, real-time safety monitoring hardware.

2.1.12 Signal Intensity Equations

Here we outline the intensity (i.e., magnitude) one observes in a reconstructed image at location $\mathbf{r}$ that arises when one plays particular types of pulses through the scanner. These equations will play a crucial role in the next subsection.

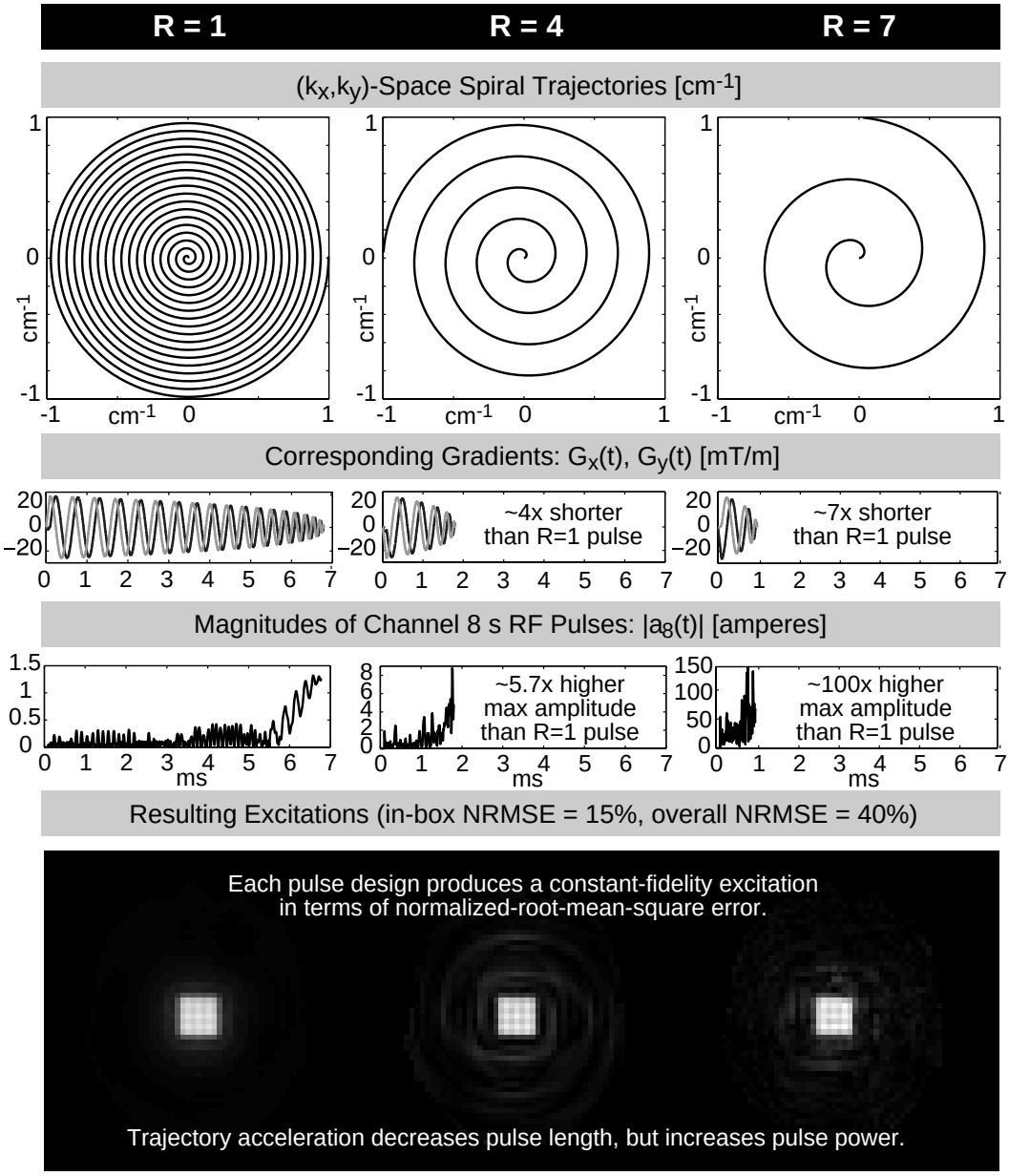

Figure 2-7: **Unaccelerated vs. accelerated spiral-trajectory parallel excitation eight-channel pulse design example**. Target pattern: 28-mm × 28-mm centered 15-degree square pattern with fixed in-box and overall normalized root-mean-square error constraints. Left, middle, and right columns: design details for $R = 1$, 4, and 7 trajectories, where R is the acceleration factor. Top row: progressively-undersampled spiral trajectories. Second row: gradient waveforms. Third row: magnitude of the eighth channel's RF pulse shape, $|a_8(t)|$. Bottom row: resulting excitations. As R increases, the pulses grow shorter in duration as intended. The eighth channel's RF pulse (along with the other seven pulses that are not shown) change greatly in both magnitude and shape with increasing R due to the loss of degrees of freedom in k-space.

First, assume a *standard one-spoke slice-selective pulse* is repeatedly transmitted with peak transmit voltage V through a transmit coil's pth channel with a fixed repetition time (TR). During this process, we form an intensity image SI via a readout procedure referred to as a *gradient-recalled echo* (GRE) [12, 89, 98]. In this scenario, the following holds:

$$\mathrm{SI}(\mathbf{r}, V) = c \cdot \rho(\mathbf{r}) \cdot |B_1^-(\mathbf{r})| \cdot \sin(V\alpha(\mathbf{r})) \cdot \frac{1 - E_1(\mathbf{r}, \mathrm{TR})}{1 - E_1(\mathbf{r}, \mathrm{TR}) \cos(V\alpha(\mathbf{r}))}, \tag{2.17}$$

where c is a gain constant, $\rho(\mathbf{r})$ is proton (H^+) density, $E_1(\mathbf{r}, \mathrm{TR}) = \exp(-\mathrm{TR}/\mathrm{T}_1(\mathbf{r}))$, and $\alpha(\mathbf{r})$ is the flip angle achieved at location $\mathbf{r}$ in rad/V [27, 137]. We thus see how the tissue relaxation parameter T_1, peak voltage, and choice of repetition time impact image intensity. The effect of T_2 has been ignored, which is a reasonable assumption [98, 137]. For a standard pulse of duration τ, $\alpha(\mathbf{r}) = \gamma\tau|S_p(\mathbf{r})|$ [27, 137].

The final parameter of (2.17), $|B_1^-(\mathbf{r})|$, is the dual of the $|S_p(\mathbf{r})|$s: it is the magnitude of the system's *reception* or *receive profile*.[3] Analogously to $|S_p(\mathbf{r})|$, the receive profile will exhibit significant spatial variation (inhomogeneity) on high-field systems. From (2.17), we see that an inhomogenous $|B_1^-(\mathbf{r})|$ indeed impacts SNR across space, but not image contrast. In practice, its image shading effect is removed simply via pointwise division from the resulting image SI.

Now assume we play a special-purpose *magnetization reset pulse* after the standard pulse as detailed in [33]. This eliminates the denominator of (2.17), yielding

$$\mathrm{SI}(\mathbf{r}, V) = c \cdot R(\mathbf{r}) \cdot (1 - E_1(\mathbf{r}, \mathrm{TR})) \cdot \sin(V\alpha(\mathbf{r})) = q(\mathbf{r}, \mathrm{TR}) \cdot \sin(V\alpha(\mathbf{r})), \tag{2.18}$$

where $R(\mathbf{r}) = \rho(\mathbf{r})|B_1^-(\mathbf{r})|$ is the *proton-density-weighted receive profile* and $q(\mathbf{r}, \mathrm{TR})$ is implicitly defined; this expression holds even for $\mathrm{TR} < \mathrm{T}_1$ [33].

Finally, consider a case where V is small enough such that $V\alpha(\mathbf{r})$ is small everywhere and a reset pulse is not used; here, $\cos(V\alpha(\mathbf{r})) \approx 1$ and $\sin(V\alpha(\mathbf{r})) \approx V\alpha(\mathbf{r})$, causing the $(1 - E_1(\mathbf{r}, \mathrm{TR}))$ terms of (2.17) to cancel, resulting in the following image intensity:

$$\mathrm{SI}(\mathbf{r}, V) = c \cdot V \cdot R(\mathbf{r}) \cdot \alpha(\mathbf{r}). \tag{2.19}$$

[3] For the purposes of this thesis, the systems we use for experimentation have only one reception profile regardless of how many transmit channels they may have.

2.1.13 Transmit Profile, Receive Profile, and Flip Angle Map Estimation

Throughout the chapter we assumed that the single-channel system transmit profile, $S(\mathbf{r})$, and the multi-channel system profiles, the $S_p(\mathbf{r})$s, were known. In reality, these profiles must be estimated. Here we first describe how to estimate $S_p(\mathbf{r})$ by collecting a series of images and exploiting (2.18). Other transmit profile magnitude estimation methods include [18,33,69,82,95,137]. We then describe a single-channel method to estimate $R(\mathbf{r}) = \rho(\mathbf{r})|B_1^-(\mathbf{r})|$ and $\alpha(\mathbf{r})$ due to any excitation pulse,[4] a seemingly novel approach that we introduced in [112, 161, 164].

Estimating the transmit profile magnitude $|S_p(\mathbf{r})|$. If B_1^+ inhomogeneity is not severe, one may exploit (2.18) to obtain $\alpha(\mathbf{r})$ (and subsequently $|S_p(\mathbf{r})|$) simply by collecting two short-TR images using standard slice-select pulses with peak voltages V_1 and V_2, where $V_2 = 2V_1$ (using the reset pulse each time), dividing the magnitude of the second image by the first (pointwise), and taking the inverse cosine [33, 69]. This method relies on the voltages being large enough such that the flip angle across the FOX is no longer in the linear regime [i.e., such that $\sin(V\alpha(\mathbf{r})) \neq V\alpha(\mathbf{r})$]. Unfortunately, when inhomogeneity is severe (e.g., in the human brain at $B_0 = 7$ Tesla), the voltages V_1 and V_2 fail to produce flip angles that fall outside of the linear regime across the entire FOX, and as a result the double-angle procedure fails to produce a stable $|S_p(\mathbf{r})|$ estimate at all spatial locations of interest. Therefore we adopt a different approach: using a standard pulse followed by a reset pulse each time, we vary V over a wide enough range to ensure that both low-flip and high-flip angles are achieved at each spatial location $\mathbf{r}$ and collect N short-TR images. For each $\mathbf{r}$, we then fit the N corresponding intensity samples to (2.18) in the least-squares sense using the Powell method [103]; this obtains $|S_p(\mathbf{r})|$ in T/V as well as $q(\mathbf{r}, \mathrm{TR})$. The Vs are chosen such that, for each $\mathbf{r}$, at least several of the N samples are in the large-tip-angle regime [82, 137].

Estimating the transmit profile phase $\angle(S_p(\mathbf{r}))$. To estimate the phase of $S_p(\mathbf{r})$, we keep V constant and collect one low-flip image and set $\angle(S_p(\mathbf{r}))$ equal to the phase of this image. This yields phase relative to the system's receive coil, which is sufficient [113, 165].

Estimating the proton-density-weighted receive profile $R(\mathbf{r})$ **on a one-channel system**. Fitting the transmit profile as described yields not only $|S(\mathbf{r})|$ but $q(\mathbf{r}, \mathrm{TR})$ as well.

[4] Estimating $\alpha(\mathbf{r})$ is critical in order to determine whether a non-standard mitigation pulse succeeds at producing a uniform flip angle magnitude across space.

However, obtaining the desired $R(\mathbf{r})$ from $q(\mathbf{r}, \mathrm{TR})$ is non-trivial because the latter depends on T_1 as seen in (2.18). Instead, we collect a low-flip image using a standard one-spoke pulse (without using the reset pulse), averaging multiple times such that (2.19) holds and SNR is large. We divide this image (pointwise) by $\alpha(\mathbf{r}) = \gamma\tau|S(\mathbf{r})|$ to obtain $R(\mathbf{r})$ (to within a multiplicative constant).

Estimating the flip angle map $\alpha(\mathbf{r})$ of any low-flip-angle pulse on a one-channel system. The weighted receive profile, $R(\mathbf{r})$, does not depend on the excitation pulse. Exploiting this, we may estimate the flip angle map achieved by any pulse, even a non-standard one such as a spoke-based mitigation waveform. First, we collect a low-flip image using the pulse of interest (without using the reset pulse); the intensity of the resulting image thus obeys (2.19). We then divide this image by the $R(\mathbf{r})$ estimate to obtain an estimate of the actual magnetization that arises when the pulse is played on the scanner (to within a multiplicative constant).

2.1.14 Safety Concern: Specific Absorption Rate

Specific absorption rate (SAR) is defined as the average energy deposition in an Ng region over an extended period of time due to the application of one or more radio-frequency (RF) excitation pulses and is a major safety concern as discussed in Ch. 1. We briefly state the equations that govern the "point SAR" that occurs at a single voxel $\mathbf{r}$ in space due to the parallel transmission of ampere-valued pulse shapes $a_1(t), \ldots, a_P(t)$ through a P-channel system[5,6] and the definition of local Ng SAR. Assume we know N_t time samples of each pulse shape spaced uniformly in time by Δ_t, i.e., for $p = 1, \ldots, P$, we know $a_p(n\Delta_t)$ for $n \in \{0, 1, \ldots, N_t - 1\}$. Recall from (2.13) that L is the duration of each pulse shape and thus $L = N_t\Delta_t$.

Electric fields. For $p = 1, \ldots, 8$, assume we know $\mathbf{E}_p(\mathbf{r}) = [E_{p,x}(\mathbf{r}), E_{p,y}(\mathbf{r}), E_{p,z}(\mathbf{r})]^{\mathrm{T}}$ (V/m/A). Formally, $\mathbf{E}_p(\mathbf{r})$ is the three-dimensional electric field (V/m) that arises at $\mathbf{r}$ when a unit ampere waveform tuned to the Larmor frequency is driven through channel p.

Point SAR at location $\mathbf{r}$. We calculate SAR (W/kg) at $\mathbf{r}$ by superimposing the electric field produced by each transmit channel due to each time sample in the RF pulse shape and then time averaging the net field's squared magnitude over the repetition time

[5]Here we have used $a_p(t)$s rather than $b_p(t)$s, i.e., current waveforms rather than voltage waveforms; the derivations here still hold if voltage waveforms are used with only a slight change in units.

[6]If one is concerned about single-channel system SAR, simply set $P = 1$ in this subsection.

(TR) interval in which the pulse is played and weighting by the known conductivity and density of that location [21, 70, 71]. The pulse shapes are transmitted from time 0 to L and nothing is transmitted during the time interval (L, TR). Formally,

$$\mathrm{SAR}(\mathbf{r}) = \frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})} \frac{1}{\mathrm{TR}} \int_0^{\mathrm{TR}} \|\mathbf{E}(\mathbf{r}, t)\|_2^2 \, dt = D \frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})} \frac{1}{L} \int_0^{L} \|\mathbf{E}(\mathbf{r}, t)\|_2^2 \, dt, \tag{2.20}$$

where $D = \frac{L}{\mathrm{TR}}$ is the duty cycle, $\rho(\mathbf{r})$ and $\sigma(\mathbf{r})$ are the density (kg/m^3) and conductivity (S/m) of the tissue at location $\mathbf{r}$, and $\mathbf{E}(\mathbf{r}, t)$ is the superposition of the electric fields generated by each of the channels scaled by the waveform samples transmitted at each time instant [168], i.e.,

$$\mathbf{E}(\mathbf{r}, t) = \sum_{p=1}^{P} a_p(t) \mathbf{E}_p(\mathbf{r}). \tag{2.21}$$

Since we know discrete samples of the $a_p(t)$s, we may approximately solve (2.20) via numeric integration:

$$\mathrm{SAR}(\mathbf{r}) \approx \frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})} \frac{\Delta_t}{\mathrm{TR}} \sum_{n=0}^{N_t - 1} \|\mathbf{E}(\mathbf{r}, n\Delta_t)\|_2^2 = \frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})} \frac{\Delta_t}{\mathrm{TR}} \sum_{n=0}^{N_t - 1} \left\| \sum_{p=1}^{P} a_p(n\Delta_t) \mathbf{E}_p(\mathbf{r}) \right\|_2^2 \tag{2.22}$$

Whole-head and local N-gram SAR. After obtaining $\mathrm{SAR}(\mathbf{r})$ for all $\mathbf{r}$ of interest (e.g., all locations in a human head model), global SAR (also referred to as mean SAR) is simply obtained by averaging the $\mathrm{SAR}(\mathbf{r})$ values. Likewise, Ng SAR at each $\mathbf{r}$ is obtained by finding an N-gram cube around each $\mathbf{r}$ and then averaging $\mathrm{SAR}(\mathbf{r})$ over all $\mathbf{r}$ within the cube, in line with [1]. (The FDA [21] and IEC [70, 71] mandate averaging over cubes rather than spheres.)

In Ch. 7, we will focus on the study of SAR produced by parallel transmission pulses. In Ch. 8 and Ch. 9, we will propose and study methods to design SAR-reduced pulses.

2.1.15 SAR-Reduced Parallel Transmission

To conclude the MRI section of this chapter, we briefly summarize two recent branches of work that have arisen that attempt to address and mitigate the SAR concerns of parallel transmission by improving upon the pulse design process.

- *ℓ_2 and ℓ_∞ constraints on pulse samples.* Because the vast majority of low-flip pulse design algorithms generate a parallel excitation pulse by solving a linear system of equations (e.g.,

[59, 102, 146, 168]), a simple way to (indirectly) reduce SAR is to impose regularizations while solving the linear system, constraining or reducing the root-mean-square or peak amperages of the resulting pTX pulse [59, 165]. This approach is simpler than those below—no knowledge of the local electric field generated by each transmit array element is required—but it does not guarantee a SAR decrease because pTX SAR does not scale directly as a function of a pulse's ℓ_2 and ℓ_∞ energies. (Only on a single-channel system is SAR guaranteed to scale directly with ℓ_2 and ℓ_∞ pulse energies.)

- *Explicit constraints on global and local SAR.* An alternative, more extensive approach is to explicitly build SAR constraints into the pulse design process. Because both whole-head mean SAR and local Ng SAR at any location may be expressed *quadratically* in terms of pulse sample values [56, 57, 158, 168], constraints on both whole-head and local SAR may be incorporated simply by adding quadratic constraints to the design algorithm [57, 158]. For example, the algorithm of [57] explicitly accounts for global SAR as well as local SAR at several spatial locations by incorporating several quadratic constraints into the design. It requires knowledge of the local electric field generated by each transmit array element per unit input, typically requiring FDTD simulations of a human body model. Examples of work in this area include [55, 57, 68, 134, 139, 142, 158].

Other pTX SAR research. In addition to the pulse design improvements above, other work involves the monitoring and tracking of SAR arising during a scan in real time [28, 54] and hardware-based SAR reduction by improving the efficiency and spatial encoding capabilities of pTX arrays [5, 6].

In Ch. 8, we will pose several SAR-reduced design approaches that were proposed alongside [56, 168], whereas in Ch. 9, we will introduce a novel approach to SAR reduction: *time multiplexing* a set of similar pulses to reduce maximum local SAR.

2.2 Sparse Approximation Theory and Algorithms

2.2.1 Overview

Moving away from MRI, we now provide background about sparse approximation theory. Over the past decade and a half, much work has been done on this topic; useful references include [23, 24, 31, 34, 37, 41, 43, 44, 52, 93, 96, 107]. Fundamentally, the goal of sparse ap-

proximation is to find a vector (or vectors) of unknowns *with a small number of nonzero elements* such that a (linear) system of equations (approximately) holds. The base case sparse approximation problem involves one known observation vector, a known system matrix, and a single sparse unknown vector [24,52,107]; we will refer to this as the *single-system single-output* (SSSO) problem. A generalization of the SSSO problem is the *single-system multiple-output* (SSMO) simultaneous sparsity problem, where there are multiple unknown vectors, a single system matrix, and a host of observation vectors [32, 92, 127, 129]. The background provided here on both SSSO sparsity and SSMO simultaneous sparsity will be useful when we study *multiple-system, single-output* (MSSO) simultaneous sparsity in Ch. 3; this novel linear inverse problem involves multiple unknown simultaneously sparse vectors, but unlike SSMO, each such vector passes through a different system matrix and the outputs of the various system matrices undergo a linear combination, yielding only one observation vector.

2.2.2 Single-System Single-Output (SSSO) Sparse Approximation

Consider a linear system of equations $\mathbf{d} = \mathbf{F}\mathbf{g}$, where $\mathbf{d} \in \mathbb{C}^M$, $\mathbf{F} \in \mathbb{C}^{M \times N}$, $\mathbf{g} \in \mathbb{C}^N$, and $\mathbf{d}$ and $\mathbf{F}$ are known. It is common to use the Moore-Penrose pseudoinverse of $\mathbf{F}$, denoted $\mathbf{F}^\dagger$, to determine $\hat{\mathbf{g}} = \mathbf{F}^\dagger \mathbf{d}$ as an (approximate) solution to the system of equations. When $\mathbf{d}$ is in the range of $\mathbf{F}$, $\hat{\mathbf{g}}$ is the solution that minimizes $\|\hat{\mathbf{g}}\|_2$, the Euclidean or ℓ_2 norm of $\hat{\mathbf{g}}$. When $\mathbf{d}$ is not in the range of $\mathbf{F}$, no solution exists; $\hat{\mathbf{g}}$ minimizes $\|\hat{\mathbf{g}}\|_2$ among the vectors that minimize $\|\mathbf{d} - \mathbf{F}\hat{\mathbf{g}}\|_2$.

When a sparse solution is desired, it is necessary for the analogue to $\hat{\mathbf{g}}$ to have only a small fraction of its entries differ from zero. We are faced with a sparse approximation problem of the form

$$\min_{\mathbf{g}} \|\mathbf{g}\|_0 \text{ subject to } \|\mathbf{d} - \mathbf{F}\mathbf{g}\|_2 \leq \epsilon, \tag{2.23}$$

where $\|\cdot\|_0$ denotes the number of nonzero elements of a vector. The subset of $\{1, 2, \ldots, N\}$ where there are nonzero entries in $\mathbf{g}$ is called the *sparsity profile*. For general $\mathbf{F}$, solving (2.23) essentially requires a search over up to $2^N - 1$ nonempty sparsity profiles. The problem is thus computationally infeasible except for very small systems of equations (e.g., even for $N = 30$, one may need to search 1,073,741,823 subsets). Formally, the problem is NP-Hard [36, 96].

For problems where (2.23) is intractable, a large body of work supports a greedy search over the columns of $\mathbf{F}$ to seek out a small subset of columns that, when weighted and linearly combined, yields a result that is close to $\mathbf{d}$ in the ℓ_2 sense along with a sparse $\mathbf{g}$ [23, 31, 37, 93, 96].

A second body of research supports the *relaxation* of (2.23) to find a sparse $\mathbf{g}$ [24]:

$$\min_{\mathbf{g}} \|\mathbf{g}\|_1 \text{ s.t. } \|\mathbf{d} - \mathbf{F}\mathbf{g}\|_2 \leq \epsilon. \tag{2.24}$$

This is a convex optimization and thus may be solved efficiently [17]. The solution of (2.24) does not always match the solution of (2.23)—if it did, the intractability of (2.23) would be contradicted—but certain conditions on $\mathbf{F}$ guarantee a proximity of their solutions [39, 40, 128]. Note that (2.24) applies an ℓ_1 norm to $\mathbf{g}$, but an ℓ_p norm (where $p < 1$) may also be used to promote sparsity [24, 52]; this leads to a non-convex problem and will not considered in this disseration.

The optimization

$$\min_{\mathbf{g}} \ \left\{ \tfrac{1}{2} \|\mathbf{d} - \mathbf{F}\mathbf{g}\|_2^2 + \lambda \|\mathbf{g}\|_1 \right\} \tag{2.25}$$

has the same set of solutions as (2.24). The first term of (2.25) keeps residual error down, whereas the second promotes sparsity of $\mathbf{g}$ [24,122]. As the control parameter, λ, is increased from zero to infinity, the algorithm yields sparser solutions and the residual error increases; sparsity is traded off with residual error. Various methods may be used to solve (2.25), such as iteratively reweighted least squares (IRLS) [77] (e.g., FOCUSS [52]), iterative shrinkage [34, 41, 43, 44], and second-order cone programming (SOCP) [17, 92]. Throughout the rest of this thesis, we will focus on formulations resembling (2.25) and its analogues rather than (2.24).

It is important to understand that a problem of the form (2.25) may arise as a proxy for (2.23) or as the inherent problem of interest. For example, in a Bayesian estimation setting, (2.25) yields the maximum a posteriori probability estimate of $\mathbf{g}$ from $\mathbf{d}$ when the observation model involves $\mathbf{F}$ and Gaussian noise and the prior on $\mathbf{g}$ is Laplacian. Similar statements can be made about the relaxations of the simultaneous sparse approximation problems posed in the following sections.

2.2.3 Single-System Multiple-Output (SSMO) Simultaneous Sparse Approximation

Let us now consider a generalization of the base-case SSSO problem where there are multiple unknown vectors, a single system matrix, and a host of observation vectors; the pth observation vector arises by multiplying the single system matrix with the pth unknown vector. The caveat here is that the system requires a *simultaneously sparse* set of vectors as the solution, i.e., a set of vectors where only a small number of each vector's entries are nonzero, and where the vectors' *sparsity profiles* (the locations of the nonzero entries) are equivalent (or are promoted to be equivalent with an increasing penalty given otherwise) [32, 92, 127, 129]. We refer to this as the *single-system multiple-output* problem.

Formally, the single-system multiple-output (SSMO) problem consists of P observation vectors ("snapshots"), all of which arise from the same system matrix:

$$\mathbf{d}_p = \mathbf{F}\mathbf{g}_p, \text{ for } p = 1, \ldots, P, \tag{2.26}$$

where $\mathbf{d}_p \in \mathbb{C}^M$ is known for $p = 1, \ldots, P$ along with $\mathbf{F} \in \mathbb{C}^{M \times N}$. In this scenario, we want to constrain the number of positions at which *any* of the $\mathbf{g}_p$s are nonzero. Thus we seek approximate solutions in which the $\mathbf{g}_p$s are not only sparse, but the union of their sparsity patterns is small; that is, a *simultaneously sparse* set of vectors is desired [92, 127]. Unfortunately, optimal approximation with a simultaneous sparsity constraint is even harder than (2.23).

Extending single-vector sparse approximation greedy techniques is one way to find an approximate solution [32, 129]. Another approach is to extend the relaxation (2.25) as follows:[7]

$$\min_{\mathbf{G}} \left\{ \tfrac{1}{2} \|\mathbf{D} - \mathbf{F}\mathbf{G}\|_{\mathrm{F}}^2 + \lambda \|\mathbf{G}\|_{\mathrm{S}} \right\}, \tag{2.27}$$

where $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_P] \in \mathbb{C}^{M \times P}$, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_P] \in \mathbb{C}^{N \times P}$, $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm, and

$$\|\mathbf{G}\|_{\mathrm{S}} = \sum_{n=1}^{N} \sqrt{\sum_{p=1}^{P} |\mathbf{G}(n,p)|^2} = \sum_{n=1}^{N} \sqrt{\sum_{p=1}^{P} |\mathbf{g}_p[n]|^2}, \tag{2.28}$$

i.e., $\|\mathbf{G}\|_{\mathrm{S}}$ is the ℓ_1 norm of the ℓ_2 norms of the rows of the $\mathbf{G}$ matrix.[8] This latter operator

[7] For $P = 1$, (2.27) collapses to the base case of (2.25).

[8] Although here we have applied an ℓ_1 norm to the ℓ_2 row energies of $\mathbf{G}$, an ℓ_p norm (where $p < 1$) could

is a *simultaneous sparsity norm*: it penalizes the program (produces large values) when the columns of $\mathbf{G}$ have dissimilar sparsity profiles [92]. Fixing λ to a sufficiently large value and solving this optimization yields simultaneously sparse $\mathbf{g}_p$s. Given the convex objective function in (2.27), one may then attempt to find a solution that minimizes the objective using, for example, IRLS-based [32] or SOCP-based [92] approaches. We conclude by noting that a formal analysis of the minimization of the convex objective (2.27) may be found in [127].

be used in place of the ℓ_1 norm if one is willing to deal with a non-convex objective function. Further, an ℓ_q norm (where $q > 2$) rather than an ℓ_2 norm could be applied to each row of $\mathbf{G}$ because the purpose of the row operation is to collapse the elements of the row into a scalar value without introducing a sparsifying effect.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Multiple-System Single-Output (MSSO) Simultaneous Sparse Approximation

3.1 Introduction

Here we propose a new type of linear inverse problem (independently of MRI) that requires a *simultaneously sparse* set of vectors as the solution. As discussed in Sec. 2.2, prior work on simultaneously sparse solutions to linear inverse problems focuses on the *single-system multiple-output* (SSMO) simultaneous sparsity problem, where there are multiple unknown vectors, a single system matrix, and a host of observation vectors [32, 92, 127, 129].

Here we study a problem different from the aforementioned one. This *multiple-system single-output* (MSSO) simultaneous sparsity problem still consists of multiple unknown vectors, but now each such vector is passed through a *different* system matrix and the outputs of the various system matrices undergo a linear combination, yielding only *one* observation vector. Given the matrices and lone observation, one must find a simultaneously sparse set of vectors that (approximately) solves the system. This problem has been explored in an MRI RF low-flip excitation pulse design context [157, 162, 167], and may also have applications to source localization using sensor arrays [74, 84] and signal denoising [24, 38, 44, 49].

Like SSMO, MSSO arises as a generalization of the *single-system single-output* (SSSO) sparse approximation problem, where there is one known observation vector, a known system matrix, and the solution one seeks is a single sparse unknown vector [24, 52, 107].

In this chapter, we propose three forward-looking greedy techniques—matching pursuit (MP) [93], orthogonal matching pursuit (OMP) [23, 31, 96], and least squares matching pursuit (LSMP) [31]—and also develop IRLS-based, shrinkage-based, and SOCP-based algorithms to approximately solve the NP-Hard MSSO simultaneous sparsity problem. We then evaluate the performance of the algorithms across three experiments: the first and second involve sparsity profile recovery in noiseless and noisy scenarios, respectively, while the third deals with linear MRI RF excitation pulse design from the perspective of an applied mathematician rather than an MRI pulse designer. We also prove that single-vector sparse approximation of a complex vector readily maps to the MSSO problem.

We first formulate the MSSO problem in Sec. 3.2 and then propose seven algorithms for solving the problem in Sec. 3.3. Details and results of the numerical experiments are given in Sec. 3.4, while the strengths and weaknesses of the algorithms are discussed in Sec. 3.5. Concluding remarks appear in Sec. 3.6. A version of the work in this chapter is currently under review [155].

3.2 MSSO Problem Formulation

We outline the MSSO problem in a style analogous to that of SSMO systems in (2.26, 2.27) and then pose a second formulation that is useful for deriving several algorithms in Sec. 3.3.

3.2.1 Standard Formulation

Consider the following system:

$$\begin{aligned} \mathbf{d} &= \mathbf{F}_1\mathbf{g}_1 + \cdots + \mathbf{F}_P\mathbf{g}_P \\ &= [\mathbf{F}_1 \cdots \mathbf{F}_P] \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_P \end{bmatrix} = \mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}, \end{aligned} \tag{3.1}$$

where $\mathbf{d} \in \mathbb{C}^M$ and the $\mathbf{F}_p \in \mathbb{C}^{M\times N}$ are known. Unlike the SSMO problem in (2.26), there is now only one observation and P different system matrices. Here we again desire an

approximate solution where the $\mathbf{g}_p$s are simultaneously sparse; formally,

$$\min_{\mathbf{g}_1,\ldots,\mathbf{g}_P} \|\mathbf{d} - \mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}\|_2 \text{ s.t. the simultaneous } K\text{-sparsity of the } \mathbf{g}_p\text{s.} \tag{3.2}$$

This is, of course, harder than the base case SSSO problem in (2.23) and thus NP-Hard. To keep the problem as general as possible, there are no constraints on the values of M, N, or P, i.e., there is no explicit requirement that the system be overdetermined or underdetermined. Further, we have used complex-valued rather than real-valued variables.

In the first half of Sec. 3.3, we will pose three approaches that attempt to solve the MSSO problem (3.2) in a greedy fashion. Another approach to solve the problem is to apply a relaxation similar to (2.25, 2.27):

$$\min_{\mathbf{G}} \left\{ \tfrac{1}{2} \|\mathbf{d} - \mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}\|_2^2 + \lambda \|\mathbf{G}\|_{\text{S}} \right\}, \tag{3.3}$$

where $\mathbf{G}$ and $\|\mathbf{G}\|_{\text{S}}$ are the same as in (2.27) and (2.28), respectively. In the second half of Sec. 3.3, we will outline four algorithms for solving this relaxed problem. By setting $P = 1$, (3.3) collapses to the base case of (2.25).

3.2.2 Alternate Formulation

In several upcoming derivations, it will be useful to view the MSSO problem from a different standpoint. To begin, we construct several new variables that are simply permutations of the $\mathbf{F}_p$s and $\mathbf{g}_p$s. First we define N new matrices:

$$\mathbf{C}_n = [\mathbf{f}_{1,n}, \ldots, \mathbf{f}_{P,n}] \in \mathbb{C}^{M \times P}, \text{ for } n = 1, \ldots, N, \tag{3.4}$$

where $\mathbf{f}_{p,n}$ is the nth column of $\mathbf{F}_p$. Next we construct N new vectors:

$$\mathbf{h}_n = [\mathbf{g}_1[n], \ldots, \mathbf{g}_P[n]]^{\text{T}} \in \mathbb{C}^P, \text{ for } n = 1, \ldots, N, \tag{3.5}$$

where $\mathbf{g}_p[n]$ is the nth element of $\mathbf{g}_p$ and $^{\mathrm{T}}$ is the transpose operation. Given the $\mathbf{C}_n$s and $\mathbf{h}_n$s, we now pose the following system:

$$\begin{aligned}\mathbf{d} &= \mathbf{C}_1\mathbf{h}_1 + \cdots + \mathbf{C}_N\mathbf{h}_N \\ &= [\mathbf{C}_1 \cdots \mathbf{C}_N] \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_N \end{bmatrix} = \mathbf{C}_{\mathrm{tot}}\mathbf{h}_{\mathrm{tot}}.\end{aligned} \tag{3.6}$$

Due to (3.4, 3.5), the system in (3.6) is *equivalent* to the one in (3.1). The relationship between the $\mathbf{g}_p$s and $\mathbf{h}_n$s implies that if we desire to find a set of simultaneously sparse $\mathbf{g}_p$s to solve (3.1, 3.2), we should seek out a set of $\mathbf{h}_n$s where many of the $\mathbf{h}_n$s equal an all-zeros vector, $\mathbf{0}$, but a few $\mathbf{h}_n$s are high in ℓ_2 energy (typically with all elements being nonzero). This claim is apparent if we consider the fact that $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ is equal to the transpose of $\mathbf{G}$, and that the $\mathbf{g}_p$s are only simultaneously sparse when $\|\mathbf{G}\|_{\mathrm{S}}$ is sufficiently small.

Given this setup, the NP-Hard formulation here equivalent to that of (3.2) is as follows:

$$\min_{\mathbf{h}_1, \ldots, \mathbf{h}_N} \|\mathbf{d} - \mathbf{C}_{\mathrm{tot}}\mathbf{h}_{\mathrm{tot}}\|_2 \;\; \text{s.t. the usage of only } K \text{ of the } \mathbf{h}_n\text{s}, \tag{3.7}$$

which, similarly to (3.2), might be (approximately) solved from a greedy standpoint.

Continuing with this alternate formulation, and given our desire to find a solution where most of the $\mathbf{h}_n$s are all-zero vectors and a few are dense, we relax the problem as follows:

$$\min_{\mathbf{h}_{\mathrm{tot}}} \left\{ \tfrac{1}{2} \|\mathbf{d} - \mathbf{C}_{\mathrm{tot}}\mathbf{h}_{\mathrm{tot}}\|_2^2 + \lambda \sum_{n=1}^{N} \|\mathbf{h}_n\|_2 \right\}. \tag{3.8}$$

Fixing λ to a sufficiently large value and solving this optimization yields many low-energy $\mathbf{h}_n$s (each close to $\mathbf{0}$), along with several dense high-energy $\mathbf{h}_n$s. Further, because $\sum_{n=1}^{N} \|\mathbf{h}_n\|_2$ is *equivalent* to $\|\mathbf{G}\|_{\mathrm{S}}$, this means (3.8) is *equivalent* to (3.3), and thus just like (3.3), the approach in (3.8) finds a set of simultaneously sparse $\mathbf{g}_p$s.

3.2.3 Differences between the SSMO and MSSO Problems

In the SSMO problem, we see from (2.26) that there are many different $\mathbf{d}$s and a single $\mathbf{F}$. The ratio of unknowns to knowns always equals N/M regardless of the number of observations, P. A large P when solving SSMO is actually beneficial because the simultaneous

sparsity of the underlying $\mathbf{g}_p$s becomes more exploitable; empirical evidence of improved sparsity profile recovery with increasing P may be found in both [32] and [92].

In contrast, we see from (3.1) that in the MSSO problem there is a single $\mathbf{d}$ and many different $\mathbf{F}$s. Here the ratio of unknowns to knowns is no longer constant with respect to P; rather it is equal to PN/M. We will show in Sec. 3.4 that as P increases, the underlying simultaneous sparsity of the $\mathbf{g}_p$s is not enough to combat the increasing number of unknowns, and that for large P, correctly identifying the sparsity profile of the underlying unknown $\mathbf{g}_p$s is a daunting task.

3.3 Proposed Algorithms

We now derive algorithms to (approximately) solve the MSSO problem defined in Sec. 3.2.

3.3.1 Matching Pursuit (MP)

To begin, we extend the single-vector case of matching pursuit [93] to an MSSO context. The classic MP technique first finds the column of the system matrix that best matches with the observed vector and then removes from the observation vector the projection of this chosen column. It proceeds to select a second column of the system matrix that best matches with the *residual observation*, and continues doing so until either K columns have been chosen (as specified by the user) or the residual observation ends up as a vector of all zeros. This algorithm is fast and computationally-efficient because the best-matching column vector during each iteration is determined simply by calculating the inner product of each column vector with the residual observation and ranking the resulting inner product magnitudes.

Now let us consider the MSSO system as posed in (3.6). In the first iteration of standard MP, we seek out the single column of the system matrix that can best represent $\mathbf{d}$. But in the MSSO context, we need to seek out which of the N $\mathbf{C}_n$ matrices can be best used to represent $\mathbf{d}$ when the columns of $\mathbf{C}_n$ undergo an arbitrarily-weighted linear combination. The key difference here is that on an iteration-by-iteration basis, we are no longer deciding which column vector best represents the observation, but which *matrix* does so. The intuition behind this approach is that ideal solutions should consist of only a few dense $\mathbf{h}_n$s and many all-zeros $\mathbf{h}_n$s. For the kth iteration of the algorithm, we need to select the proper

index $n \in \{1, \ldots, N\}$ by solving:

$$q_k = \underset{n}{\operatorname{argmin}} \ \min_{\mathbf{h}_n} \ \|\mathbf{r}_{k-1} - \mathbf{C}_n \mathbf{h}_n\|_2^2, \tag{3.9}$$

where q_k is the index that will be selected and $\mathbf{r}_{k-1}$ is the current residual observation. For fixed n, the solution to the inner minimization is obtained via the pseudoinverse, $\mathbf{h}_n^{\text{opt}} = \mathbf{C}_n^{\dagger} \mathbf{r}_{k-1}$, yielding

$$q_k = \underset{n}{\operatorname{argmin}} \ \min_{\mathbf{h}_n} \ \|\mathbf{r}_{k-1} - \mathbf{C}_n (\mathbf{C}_n^{\dagger} \mathbf{r}_{k-1})\|_2^2 = \underset{n}{\operatorname{argmax}} \ \mathbf{r}_{k-1}^{\text{H}} \mathbf{C}_n \mathbf{C}_n^{\dagger} \mathbf{r}_{k-1}, \tag{3.10}$$

where $^{\text{H}}$ is the Hermitian transpose. From (3.10) we see that, analogously to standard MP, choosing the best index for iteration k involves computing and ranking a series of inner-product-like quadratic terms.

Algorithm 3.1 outlines the entire procedure. After K iterations, one obtains $I_K \subset \{1, \ldots, N\}$ (of cardinality $T \leq K$), a set indicating the chosen $\mathbf{C}_n$ matrices. The weights to apply to each chosen matrix (i.e., the corresponding $\mathbf{h}_n$s) are obtained via a finalization step; they are the best weightings in the ℓ_2-residual-error sense with which to linearly combine the columns of the chosen $\mathbf{C}_n$ matrices to best match the observation $\mathbf{d}$. Since T total matrices end up being chosen by the process, there is no penalty in retuning the T associated $\mathbf{h}_n$ vectors because they are allowed to be dense. The $N - T$ other $\mathbf{C}_n$s (and corresponding $\mathbf{h}_n$s) are not used.[1]

One property of note is that if $M \leq P$, *Algorithm 3.1* stops after one iteration. This is because $\mathbf{C}_n \mathbf{C}_n^{\dagger}$ in this case is simply an $M \times M$ identity matrix for all $n \in \{1, \ldots, N\}$, and thus any one of the $\mathbf{C}_n$s is enough to represent $\mathbf{d}$ exactly when its columns are properly weighted and linearly combined.

3.3.2 Orthogonal Matching Pursuit (OMP)

In single-vector MP, the residual $\mathbf{r}_k$ always ends up orthogonal to the kth column of the system matrix, but as the algorithm continues iterating, there is no guarantee the residual remains orthogonal to column k or is minimized in the least-squares sense with respect to

[1] From the perspective of $\mathbf{F}_p$s and $\mathbf{g}_p$s in (3.1), *Algorithm 3.1* determines weights to place along only T rows of $\mathbf{G}$ (leaving the other $N - T$ rows zeroed out) that still yields a good approximation of $\mathbf{d}$ in the ℓ_2 error sense. It is seeking out the best rows of $\mathbf{G}$ which, when densely filled, yield a sound approximation of $\mathbf{d}$.

Algorithm 3.1 MSSO Matching Pursuit

Task: greedily choose up to K of the $\mathbf{C}_n$s to best represent $\mathbf{d}$ via $\mathbf{C}_1\mathbf{h}_1 + \cdots + \mathbf{C}_N\mathbf{h}_N$.

Data and Parameters: $\mathbf{d}$ and $\mathbf{C}_n, n \in \{1, \ldots, N\}$ are given. K iterations.

Precompute: $\mathbf{Q}_n = \mathbf{C}_n\mathbf{C}_n^\dagger$, for $n \in \{1, \ldots, N\}$.

Initialize: Set $k = 0$, $\mathbf{r}_0 = \mathbf{d}$, $I_0 = \emptyset$, $\mathbf{S}_0 = [\,]$.

Iterate: Set $k = 1$ and apply:

- $q_k = \text{argmax}_n \ \mathbf{r}_{k-1}^{\text{H}}\mathbf{Q}_n\mathbf{r}_{k-1}$.
- **if** $q_k \notin I_{k-1}$ **then**
  $I_k = I_{k-1} \cup \{q_k\}$
  $\mathbf{S}_k = [\mathbf{S}_{k-1}, \mathbf{C}_{q_k}]$
  else
  $I_k = I_{k-1}$
  $\mathbf{S}_k = \mathbf{S}_{k-1}$
  end if
- $\mathbf{r}_k = \mathbf{r}_{k-1} - \mathbf{Q}_{q_k}\mathbf{r}_{k-1}$.
- $k = k + 1$. Terminate loop if $k > K$ or $\mathbf{r}_k = \mathbf{0}$. I_K ends with $T \leq K$ elements.

Compute Weights: $\mathbf{x} = \mathbf{S}_K^\dagger\mathbf{d}$, unstack $\mathbf{x}$ into $\mathbf{h}_{q_1}, \ldots, \mathbf{h}_{q_T}$; set remaining $\mathbf{h}_n$s to $\mathbf{0}$.

the entire set of k chosen column vectors (indexed by $q_1, \ldots, q_k$). Furthermore, K iterations of single-vector MP do not guarantee K different columns will be selected. Single-vector OMP is an extension to MP that attempts to mitigate these problems by improving the calculation of the residual vector. During the kth iteration of single-vector OMP, column q_k is chosen exactly as in MP (by ranking the inner products of the residual vector $\mathbf{r}_{k-1}$ with the various column vectors), but the residual vector is updated by accounting for *all* columns chosen up through iteration k rather than simply the last one [31, 96].

To extend OMP to the MSSO problem, we choose matrix q_k during iteration k as in MSSO MP and then in the spirit of single-vector OMP compute the new residual as follows:

$$\mathbf{r}_k = \mathbf{d} - \mathbf{S}_k(\mathbf{S}_k^\dagger\mathbf{d}), \tag{3.11}$$

where $\mathbf{S}_k = [\mathbf{C}_{q_1}, \ldots, \mathbf{C}_{q_k}]$ and $\mathbf{S}_k^\dagger\mathbf{d}$ is the best choice of $\mathbf{x}$ that minimizes the residual error $\|\mathbf{d} - \mathbf{S}_k\mathbf{x}\|_2$. That is, to update the residual we now employ all chosen matrices, weighting and combining them to best represent $\mathbf{d}$ in the least-squares sense, yielding an $\mathbf{r}_k$ that is orthogonal to the columns of $\mathbf{S}_k$ (and thus orthogonal to $\mathbf{C}_{q_1}, \ldots, \mathbf{C}_{q_k}$), which has the benefit of ensuring that OMP will select a new $\mathbf{C}_n$ matrix at each step.

Algorithm 3.2 describes the OMP algorithm; the complexity here is moderately greater

Algorithm 3.2 MSSO Orthogonal Matching Pursuit

Task: greedily choose up to K of the $\mathbf{C}_n$s to best represent $\mathbf{d}$ via $\mathbf{C}_1\mathbf{h}_1 + \cdots + \mathbf{C}_N\mathbf{h}_N$.

Data and Parameters: $\mathbf{d}$ and $\mathbf{C}_n, n \in \{1, \ldots, N\}$ are given. K iterations.

Precompute: $\mathbf{Q}_n = \mathbf{C}_n\mathbf{C}_n^\dagger$, for $n \in \{1, \ldots, N\}$.

Initialize: Set $k = 0$, $\mathbf{r}_0 = \mathbf{d}$, $I_0 = \emptyset$, $\mathbf{S}_0 = [\,]$.

Iterate: Set $k = 1$ and apply:

- $q_k = \text{argmax}_{n \notin I_{k-1}} \ \mathbf{r}_{k-1}^{\text{H}}\mathbf{Q}_n\mathbf{r}_{k-1}$.
- $I_k = I_{k-1} \cup \{q_k\}$
- $\mathbf{S}_k = [\mathbf{S}_{k-1}, \mathbf{C}_{q_k}]$
- $\mathbf{r}_k = \mathbf{d} - \mathbf{S}_k\mathbf{S}_k^\dagger\mathbf{d}$.
- $k = k + 1$. Terminate loop if $k > K$ or $\mathbf{r}_k = \mathbf{0}$. I_K ends with $T \leq K$ elements.

Compute Weights: $\mathbf{x} = \mathbf{S}_K^\dagger\mathbf{d}$, unstack $\mathbf{x}$ into $\mathbf{h}_{q_1}, \ldots, \mathbf{h}_{q_T}$; set remaining $\mathbf{h}_n$s to $\mathbf{0}$.

than that of MP because the pseudoinversion of an $M \times Pk$ matrix is required during each iteration k.

3.3.3 Least Squares Matching Pursuit (LSMP)

Beyond OMP there exists a greedy algorithm with an even greater computational complexity known as LSMP. In single-vector LSMP, one solves $N - k + 1$ least squares minimizations during iteration k in order to determine which column of the system matrix may be used to best represent $\mathbf{d}$ [31].

Thus to extend LSMP to MSSO systems, we must ensure that during iteration k we account for the $k - 1$ previously chosen $\mathbf{C}_n$ matrices when choosing the kth one to best construct an approximation to $\mathbf{d}$. Specifically, index q_k is selected as follows:

$$q_k = \underset{n \in \{1, \ldots, N\}, n \notin I_{k-1}}{\text{argmin}} \ \min_{\mathbf{x}} \ \|\mathbf{S}_k^{(n)}\mathbf{x} - \mathbf{d}\|_2^2, \tag{3.12}$$

where I_{k-1} is the set of indices chosen up through iteration $k - 1$, $\mathbf{S}_k^{(n)} = [\mathbf{S}_{k-1}, \mathbf{C}_n]$, $\mathbf{S}_{k-1} = [\mathbf{C}_{q_1}, \ldots, \mathbf{C}_{q_{k-1}}]$, and $\mathbf{x} \in \mathbb{C}^{Pk}$. For fixed n, the solution of the inner iteration is $\mathbf{x}^{\text{opt}} = (\mathbf{S}_k^{(n)})^\dagger\mathbf{d}$; it is this step that ensures the residual observation error will be minimized by using *all* chosen matrices. Substituting $\mathbf{x}^{\text{opt}}$ into (3.12) and simplifying the expression

Algorithm 3.3 MSSO Least Squares Matching Pursuit

Task: greedily choose K of the $\mathbf{C}_n$s to best represent $\mathbf{d}$ via $\mathbf{C}_1\mathbf{h}_1 + \cdots + \mathbf{C}_N\mathbf{h}_N$.

Data and Parameters: $\mathbf{d}$ and $\mathbf{C}_n, n \in \{1, \ldots, N\}$ are given. K iterations.

Initialize: Set $k = 0$, $I_0 = \emptyset$, $\mathbf{S}_0 = [\,]$.

Iterate: Set $k = 1$ and apply:

- $q_k = \text{argmax}_{n \notin I_{k-1}} \ \mathbf{d}^{\text{H}}(\mathbf{S}_k^{(n)})(\mathbf{S}_k^{(n)})^{\dagger}\mathbf{d}$, where $\mathbf{S}_k^{(n)} = [\mathbf{S}_{k-1}, \mathbf{C}_n]$
- $I_k = I_{k-1} \cup \{q_k\}$
- $\mathbf{S}_k = [\mathbf{S}_{k-1}, \mathbf{C}_{q_k}]$
- $k = k + 1$. Terminate loop if $k > K$ or $\mathbf{r}_k = \mathbf{0}$. I_K ends with $T \leq K$ elements.

Compute Weights: $\mathbf{x} = \mathbf{S}_K^{\dagger}\mathbf{d}$, unstack $\mathbf{x}$ into $\mathbf{h}_{q_1}, \ldots, \mathbf{h}_{q_T}$; set remaining $\mathbf{h}_n$s to $\mathbf{0}$.

yields

$$q_k = \underset{n \notin I_{k-1}}{\text{argmax}} \ \mathbf{d}^{\text{H}}\mathbf{Q}_k^{(n)}\mathbf{d}, \tag{3.13}$$

where $\mathbf{Q}_k^{(n)} = (\mathbf{S}_k^{(n)})(\mathbf{S}_k^{(n)})^{\dagger}$.

Algorithm 3.3 describes the LSMP method. The complexity here is much greater than that of OMP because $N - k + 1$ pseudoinversions of an $M \times Pk$ matrix are required during each iteration k. Furthermore, the dependence of $\mathbf{Q}_k^{(n)}$ on both n and k makes precomputing all such matrices infeasible in most cases. One way to decrease computation and runtime might be to extend the projection-based recursive updating scheme of [31] to MSSO LSMP.

3.3.4 Iteratively Reweighted Least Squares (IRLS)

Having posed three greedy approaches for solving the MSSO problem, we now turn our attention toward minimizing (3.8), the relaxed objective function. Here, the regularization term λ is used to trade off simultaneous sparsity with residual observation error.

One way to minimize (3.8) is to use an IRLS-based approach [77]. To begin, consider

manipulating the right-hand term of (3.8) as follows:

$$\begin{aligned}
\lambda \sum_{n=1}^{N} \|\mathbf{h}_n\|_2 &= \lambda \sum_{n=1}^{N} \frac{\|\mathbf{h}_n\|_2^2}{\|\mathbf{h}_n\|_2} = \lambda \sum_{n=1}^{N} \frac{|\mathbf{h}_n[1]|^2 + \cdots + |\mathbf{h}_n[P]|^2}{\|\mathbf{h}_n\|_2} \\
&\approx \frac{\lambda}{2} \sum_{n=1}^{N} [\mathbf{h}_n^*[1], \ldots, \mathbf{h}_n^*[P]] \begin{bmatrix} \frac{2}{\|\mathbf{h}_n\|_2+\epsilon} & & \\ & \ddots & \\ & & \frac{2}{\|\mathbf{h}_n\|_2+\epsilon} \end{bmatrix} \begin{bmatrix} \mathbf{h}_n[1] \\ \vdots \\ \mathbf{h}_n[P] \end{bmatrix} \\
&= \frac{\lambda}{2} \sum_{n=1}^{N} \mathbf{h}_n^{\mathrm{H}} \mathbf{W}_n \mathbf{h}_n \\
&= \frac{\lambda}{2} \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}} \cdots \mathbf{h}_N^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \mathbf{W}_1 & & \\ & \ddots & \\ & & \mathbf{W}_N \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_N \end{bmatrix} = \frac{\lambda}{2} \mathbf{h}_{\mathrm{tot}}^{\mathrm{H}} \mathbf{W}_{\mathrm{tot}} \mathbf{h}_{\mathrm{tot}},
\end{aligned} \tag{3.14}$$

where * is the complex conjugate of a scalar, $\mathbf{W}_n$ is a $P \times P$ real-valued diagonal matrix whose diagonal elements each equal $2/(\|\mathbf{h}_n\|_2 + \epsilon)$, and ϵ is some small non-negative value introduced to mitigate poor conditioning of the $\mathbf{W}_n$s. If we fix $\mathbf{W}_{\mathrm{tot}} \in \mathbb{R}^{PN \times PN}$ by computing it using some prior estimate of $\mathbf{h}_{\mathrm{tot}}$, then the right-hand term of (3.8) becomes a quadratic function and (3.8) transforms into a Tikhonov optimization [123, 124]:

$$\min_{\mathbf{h}_{\mathrm{tot}}} \left\{ \tfrac{1}{2} \|\mathbf{d} - \mathbf{C}_{\mathrm{tot}} \mathbf{h}_{\mathrm{tot}}\|_2^2 + \tfrac{\lambda}{2} \mathbf{h}_{\mathrm{tot}}^{\mathrm{H}} \mathbf{W}_{\mathrm{tot}} \mathbf{h}_{\mathrm{tot}} \right\}. \tag{3.15}$$

Finally, by performing a change of variables and exploiting the properties of $\mathbf{W}_{\mathrm{tot}}$, we can convert (3.15) into an expression that may be minimized using the robust and reliable conjugate-gradient (CG) least-squares solver LSQR [100, 101], so named because it applies a QR decomposition [51] when solving the system in the least-squares sense. LSQR works better in practice than several other CG methods [13] because it restructures the input system via the Lanczos process [86] and applies a Golub-Kahan bidiagonalization [50] prior to solving it. (Aside: we will study LSQR in an MRI pulse design context in Ch. 4.)

To apply LSQR to (3.15), we first construct $\mathbf{W}_{\mathrm{tot}}^{1/2}$ as the element-by-element square-root of the diagonal matrix $\mathbf{W}_{\mathrm{tot}}$ and then take its inverse to obtain $\mathbf{W}_{\mathrm{tot}}^{-1/2}$. Defining $\mathbf{q} = \mathbf{W}_{\mathrm{tot}}^{1/2} \mathbf{h}_{\mathrm{tot}}$ and $\mathbf{A} = \mathbf{C}_{\mathrm{tot}} \mathbf{W}_{\mathrm{tot}}^{-1/2}$, (3.15) becomes:

$$\min_{\mathbf{q}} \left\{ \|\mathbf{d} - \mathbf{A}\mathbf{q}\|_2^2 + \lambda \|\mathbf{q}\|_2^2 \right\}. \tag{3.16}$$

Algorithm 3.4 MSSO Iteratively Reweighted Least Squares

Task: Minimize $\left\{\frac{1}{2}\left\|\mathbf{d}-\mathbf{C}_{\text{tot}}\mathbf{h}_{\text{tot}}\right\|_2^2+\lambda\sum_{n=1}^{N}\left\|\mathbf{h}_n\right\|_2\right\}$ using an iterative scheme.

Data and Parameters: λ, $\mathbf{d}$, $\mathbf{C}_{\text{tot}}$, δ, and K are given.

Initialize: Set $k=0$ and $\mathbf{h}_{\text{tot},k=0}=(\mathbf{C}_{\text{tot}})^{\dagger}\mathbf{d}$ (or e.g. $\mathbf{h}_{\text{tot},k=0}=\mathbf{1}$).

Iterate: Set $k=1$ and apply:

- Create $\mathbf{W}_{\text{tot}}$ from $\mathbf{h}_{\text{tot},k-1}$; construct $\mathbf{W}_{\text{tot}}^{1/2}$, $\mathbf{W}_{\text{tot}}^{-1/2}$, and let $\mathbf{A}=\mathbf{C}_{\text{tot}}\mathbf{W}_{\text{tot}}^{-1/2}$.
- Obtain $\mathbf{q}_{\text{tmp}}$ by using LSQR to solve $\min_{\mathbf{q}}\left\{\|\mathbf{d}-\mathbf{A}\mathbf{q}\|_2^2+\lambda\|\mathbf{q}\|_2^2\right\}$.
- Set $\mathbf{h}_{\text{tot,tmp}}=\mathbf{W}_{\text{tot}}^{-1/2}\mathbf{q}_{\text{tmp}}$.
- Line search: find $\mu_0\in[0,1]$ such that $(1-\mu)\mathbf{h}_{\text{tot},k-1}+\mu\mathbf{h}_{\text{tot,tmp}}$ minimizes (3.8).
- Set $\mathbf{h}_{\text{tot},k}=(1-\mu_0)\mathbf{h}_{\text{tot},k-1}+\mu_0\mathbf{h}_{\text{tot,tmp}}$.
- $k=k+1$. Terminate loop when $k>K$ or (3.8) decreases by less than δ.

Finalize: Unstack the last $\mathbf{h}_{\text{tot}}$ solution into $\mathbf{h}_1,\ldots,\mathbf{h}_N$.

This problem may be solved directly by simply providing $\mathbf{d}$, λ, and $\mathbf{A}$ to the LSQR package because LSQR is formulated to solve the exact problem in (3.16). Calling LSQR with these variables yields $\mathbf{q}^{\text{opt}}$, and the solution $\mathbf{h}_{\text{tot}}^{\text{opt}}$ is backed out via $\mathbf{W}_{\text{tot}}^{-1/2}\mathbf{q}^{\text{opt}}$.

Algorithm 3.4 outlines how one may iteratively apply (3.16) to attempt to find a solution that minimizes the original cost function, (3.8). The technique iterates until the objective function decreases by less than δ or the maximum number of iterations, K, is exceeded. The initial solution estimate is obtained via pseudoinversion of $\mathbf{C}_{\text{tot}}$ (an all-zeros initialization would cause $\mathbf{W}_{\text{tot}}$ to be poorly conditioned). A line search is used to step between the prior solution estimate and the upcoming one; this improves the rate of convergence and ensures the objective decreases at each step. This method is global in the sense that all PN unknowns are being estimated concurrently per iteration.

3.3.5 Row-by-Row Shrinkage (RBRS)

The proposed IRLS technique solves for all PN unknowns during each iteration, but this is cumbersome when PN is large. An alternative approach is to apply an inner loop that fixes n and then iteratively tunes $\mathbf{h}_n$ while holding the other $\mathbf{h}_m$s ($m\neq n$) constant; thus only P (rather than PN) unknowns need to be solved for during each inner iteration.

This idea inspires the RBRS algorithm. The term "row-by-row" is used because each $\mathbf{h}_n$ corresponds to row n of the $\mathbf{G}$ matrix in (3.3), and "shrinkage" is used because the ℓ_2

energy of most of the $\mathbf{h}_n$s will essentially be "shrunk" (to some extent) during each inner iteration: when λ is sufficiently large and many iterations are undertaken, many $\mathbf{h}_n$s will be close to all-zeros vectors and only several will be dense and high in energy.

RBRS for real-valued data

Assume $\mathbf{d}$ and the $\mathbf{C}_n$s of (3.8) are real-valued. We now seek to minimize the function by extending the single-vector sequential shrinkage technique of [43] and making modifications to (3.8). Assume we have prior estimates of $\mathbf{h}_1$ through $\mathbf{h}_N$, and that we now desire to update only the jth vector while keeping the other $N-1$ fixed. The shrinkage update of $\mathbf{h}_j$ is achieved via:

$$\min_{\mathbf{x}} \left\{ \tfrac{1}{2} \left\| \left[\Sigma_{n=1}^{N} \mathbf{C}_n \mathbf{h}_n - \mathbf{C}_j \mathbf{h}_j \right] + \mathbf{C}_j \mathbf{x} - \mathbf{d} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_2 \right\}, \tag{3.17}$$

where $\Sigma_{n=1}^{N} \mathbf{C}_n \mathbf{h}_n - \mathbf{C}_j \mathbf{h}_j$ forms an approximation of $\mathbf{d}$ using the prior solution coefficients, but discards the component contributed by the original jth vector, replacing the latter via an updated solution vector, $\mathbf{x}$. The left-hand term promotes a solution $\mathbf{x}$ that keeps residual error down, whereas the right-hand term penalizes $\mathbf{x}$s that contain nonzeros. It is crucial to note that the right-hand term does not promote the element-by-element sparsity of $\mathbf{x}$; rather, it penalizes the overall ℓ_2 energy of $\mathbf{x}$, and thus both sparse and dense $\mathbf{x}$s are penalized equally if their overall ℓ_2 energies are the same.

One way to solve (3.17) is to take its derivative with respect to $\mathbf{x}^{\mathrm{T}}$ and find $\mathbf{x}$ such that the derivative equals $\mathbf{0}$. By doing this and shuffling terms, and assuming we have an initial estimate of $\mathbf{x}$, we may solve for $\mathbf{x}$ iteratively:

$$\mathbf{x}_i = \left[\mathbf{C}_j^{\mathrm{T}} \mathbf{C}_j + \frac{\lambda}{\left\| \mathbf{x}_{i-1} \right\|_2 + \epsilon} \mathbf{I} \right]^{-1} \mathbf{C}_j^{\mathrm{T}} \mathbf{r}_j, \tag{3.18}$$

where $\mathbf{r}_j = \mathbf{d} + \mathbf{C}_j \mathbf{h}_j - \Sigma_{n=1}^{N} \mathbf{C}_n \mathbf{h}_n$, $\mathbf{I}$ is a $P \times P$ identity matrix, and ϵ is a small value that avoids ill-conditioned results.[2] By iterating on (3.18) until (3.17) changes by less than δ_{inner}, we arrive at a solution to (3.17), $\mathbf{x}^{\text{opt}}$, and this then replaces the prior solution vector, $\mathbf{h}_j$. Having completed the update of the jth vector, we proceed to update the rest of the vectors, looping this outer process K times or until the main objective function, (3.8), changes by

[2] Eq. (3.18) consists of a direct inversion of a $P \times P$ matrix, which is acceptable in this paper because all experiments involve $P \leq 10$. If P is large, (3.18) could be solved via a CG technique (e.g., LSQR).

Algorithm 3.5 MSSO Row-by-Row Sequential Iterative Shrinkage

Task: Minimize $\left\{ \frac{1}{2} \left\| \mathbf{d} - \mathbf{C}_{\text{tot}} \mathbf{h}_{\text{tot}} \right\|_2^2 + \lambda \sum_{n=1}^{N} \left\| \mathbf{h}_n \right\|_2 \right\}$ using an iterative scheme when all data is *real-valued*.

Data and Parameters: λ, $\mathbf{d}$, $\mathbf{C}_n$ ($n \in \{1, \ldots, N\}$), δ_{outer}, δ_{inner}, K, and I are given.

Initialize: Set $k = 0$ and $\mathbf{h}_{\text{tot}} = (\mathbf{C}_{\text{tot}})^{\dagger} \mathbf{d}$ (or e.g. $\mathbf{h}_{\text{tot}} = \mathbf{1}$), unstack into $\mathbf{h}_1, \ldots, \mathbf{h}_N$.

Iterate: Set $k = 1$ and apply:

- Sweep over row vectors: set $j = 1$ and apply:
  - Optimize a row vector: set $i = 1$ and $\mathbf{x}_0 = \mathbf{h}_j$ and then apply:
    - $\mathbf{x}_i = \left[\mathbf{C}_j^{\mathrm{T}} \mathbf{C}_j + \frac{\lambda}{\|\mathbf{x}_{i-1}\|_2 + \epsilon} \mathbf{I} \right]^{-1} \mathbf{C}_j^{\mathrm{T}} \mathbf{r}_j$, where $\mathbf{r}_j = \mathbf{d} + \mathbf{C}_j \mathbf{h}_j - \Sigma_{n=1}^{N} \mathbf{C}_n \mathbf{h}_n$.
    - $i = i + 1$. Terminate when $i > I$ or (3.17) decreases by less than δ_{inner}.
  - Finalize row vector update: set $\mathbf{h}_j$ to equal the final $\mathbf{x}$.
  - $j = j + 1$. Terminate loop when $j > N$.
- $k = k + 1$. Terminate loop when $k > K$ or (3.8) decreases by less than δ_{outer}.

Finalize: If λ was large enough, several $\mathbf{h}_n$s should be dense and others close to $\mathbf{0}$.

less than δ_{outer}. *Algorithm 3.5* details the entire procedure; unlike IRLS, here we essentially repeatedly invert $P \times P$ matrices to pursue a row-by-row solution, rather than $PN \times PN$ matrices to pursue a solution that updates *all* rows per iteration.

Extending RBRS to complex-valued data

If (3.8) contains complex-valued terms, we may structure the row-by-row updates as in (3.17), but because the derivative of the objective in (3.17) is complicated due to the presence of complex-valued terms, the simple update equation given in (3.18) is no longer applicable. One way to overcome this problem is to turn the complex-valued problem into a real-valued one.

To accomplish this conversion, let us first create several real-valued expanded vectors:

$$\widetilde{\mathbf{d}} = \begin{bmatrix} \mathrm{Re}(\mathbf{d}) \\ \mathrm{Im}(\mathbf{d}) \end{bmatrix} \in \mathbb{R}^{2M}, \quad \widetilde{\mathbf{h}}_n = \begin{bmatrix} \mathrm{Re}(\mathbf{h}_n) \\ \mathrm{Im}(\mathbf{h}_n) \end{bmatrix} \in \mathbb{R}^{2P}, \tag{3.19}$$

as well as real-valued expanded matrices:

$$\widetilde{\mathbf{C}}_n = \begin{bmatrix} \mathrm{Re}(\mathbf{C}_n) & -\mathrm{Im}(\mathbf{C}_n) \\ \mathrm{Im}(\mathbf{C}_n) & \mathrm{Re}(\mathbf{C}_n) \end{bmatrix} \in \mathbb{R}^{2M \times 2P}. \tag{3.20}$$

Due to the structure of (3.19, 3.20) and the fact that $\|\mathbf{h}_n\|_2$ equals $\|\widetilde{\mathbf{h}}_n\|_2$, the following optimization is *equivalent* to (3.8):

$$\min_{\widetilde{\mathbf{h}}_1,\ldots,\widetilde{\mathbf{h}}_N} \left\{ \tfrac{1}{2} \left\| \widetilde{\mathbf{d}} - \sum_{n=1}^{N} \widetilde{\mathbf{C}}_n \widetilde{\mathbf{h}}_n \right\|_2^2 + \lambda \sum_{n=1}^{N} \|\widetilde{\mathbf{h}}_n\|_2 \right\}. \tag{3.21}$$

This means we may apply RBRS to complex-valued scenarios by substituting the $\widetilde{\mathbf{h}}_n$s for the $\mathbf{h}_n$s and $\widetilde{\mathbf{C}}_n$s for the $\mathbf{C}_n$s in (3.17, 3.18) and *Algorithm 3.5.* [Eq. (3.18) becomes an applicable update equation because (3.17) will consist of only real-valued terms and the derivative calculated earlier is again applicable.] Finally, after running the algorithm to obtain finalized $\widetilde{\mathbf{h}}_n$s, we may simply restructure them into complex $\mathbf{h}_n$s.

3.3.6 Column-by-Column Shrinkage (CBCS)

Here we propose a dual of RBRS—a technique that sequentially updates the *columns* of $\mathbf{G}$ (i.e., the $\mathbf{g}_p$s) in (3.1, 3.3) rather than its rows (the $\mathbf{h}_n$s). Interestingly, we will show that this approach yields a *separable* optimization and reduces the overall problem to simply repeated *element-by-element* shrinkages of each $\mathbf{g}_p$.

CBCS for real-valued data

Assume the $\mathbf{g}_p$s, $\mathbf{F}_p$s, and $\mathbf{d}$ in (3.3) are real-valued and that we have prior estimates of the $\mathbf{g}_p$s. Let us consider updating the pth vector while keeping the other $P-1$ fixed. This reduces (3.3) to

$$\min_{\mathbf{x}} \left\{ \tfrac{1}{2} \|\mathbf{r} - \mathbf{F}_p \mathbf{x}\|_2^2 + \lambda \sum_{n=1}^{N} \sqrt{(\mathbf{x}[n])^2 + \mathbf{b}[n]} \right\}, \tag{3.22}$$

where $\mathbf{x}$ will be the update of $\mathbf{g}_p$, and $\mathbf{r}$ and $\mathbf{b}$ are as follows:

$$\mathbf{r} = \mathbf{d} + \mathbf{F}_p \mathbf{g}_p - \sum_{q=1}^{P} \mathbf{F}_q \mathbf{g}_q, \tag{3.23}$$

and

$$\mathbf{b}[n] = -(\mathbf{g}_p[n])^2 + \sum_{q=1}^{P} (\mathbf{g}_q[n])^2, \text{ for } n = 1,\ldots,N. \tag{3.24}$$

If the $\mathbf{b}[n]$s were not present, (3.22) would reduce to the standard problem iterated shrinkage is intended to solve [43, 44].

Now let us apply a proximal relaxation [30, 34, 48] to (3.22) and seek a solution $\mathbf{x} \in \mathbb{R}^N$ as a shrinkage update of $\mathbf{g}_p$:

$$\min_{\mathbf{x}} \left\{ \tfrac{1}{2} \left(\|\mathbf{r} - \mathbf{F}_p \mathbf{x}\|_2^2 + \alpha \left\|\mathbf{x} - \mathbf{g}_p\right\|_2^2 - \left\|\mathbf{F}_p(\mathbf{x} - \mathbf{g}_p)\right\|_2^2 \right) + \lambda \sum_{n=1}^{N} \sqrt{(\mathbf{x}[n])^2 + \mathbf{b}[n]} \right\}, \quad (3.25)$$

where α is chosen such that $\alpha \mathbf{I} - \mathbf{F}_p^{\mathrm{T}} \mathbf{F}_p$ is positive definite (e.g., α may be set to the maximum singular value of $\mathbf{F}_p$). The idea here is to replace $\mathbf{g}_p$ with the solution $\mathbf{x}$ and then iterate this procedure, repeatedly solving (3.25). This ultimately yields an updated $\mathbf{x}$ that globally minimizes (3.22) because the proximal method is guaranteed to arrive at a local minimum [34, 48] and (3.22) itself is convex. Having obtained $\mathbf{x}$, we perform an update, $\mathbf{g}_p = \mathbf{x}$, and then repeat the overall process for the next $\mathbf{g}_p$, and so forth. Additionally, we add a layer of iteration on top of this column-by-column sweep, optimizing each of the P vectors a total of K times.

The only obstacle that remains in order for us to implement the entire algorithm is an efficient way to solve (3.25). We pursue such an approach by first expanding the terms of (3.25):

$$\min_{\mathbf{x}} \left\{ \mathbf{c} + \mathbf{v}^{\mathrm{T}} \mathbf{x} + \frac{\alpha}{2} \mathbf{x}^{\mathrm{T}} \mathbf{x} + \lambda \sum_{n=1}^{N} \sqrt{(\mathbf{x}[n])^2 + \mathbf{b}[n]} \right\}, \quad (3.26)$$

where $\mathbf{c} = \frac{1}{2}\mathbf{r}^{\mathrm{T}}\mathbf{r} + \frac{\alpha}{2}\mathbf{g}_p^{\mathrm{T}}\mathbf{g}_p - \frac{1}{2}\mathbf{g}_p^{\mathrm{T}}\mathbf{F}_p^{\mathrm{T}}\mathbf{F}_p\mathbf{g}_p$ and $\mathbf{v} = \mathbf{F}_p^{\mathrm{T}}\mathbf{F}_p\mathbf{g}_p - \alpha\mathbf{g}_p - \mathbf{F}_p^{\mathrm{T}}\mathbf{r}$. Since $\mathbf{c}$ is constant, we may ignore it in the optimization. Upon closer inspection, we see that (3.26) is a *separable* problem and that the individual scalar elements of $\mathbf{x}$ may be optimized independently. For the nth element of $\mathbf{x}$, (3.26) simplifies to:

$$\min_{\mathbf{x}[n]} \left\{ \mathbf{v}[n]\mathbf{x}[n] + \frac{\alpha}{2}(\mathbf{x}[n])^2 + \lambda \sqrt{(\mathbf{x}[n])^2 + \mathbf{b}[n]} \right\}, \quad (3.27)$$

Having burrowed down to an element-by-element problem, all that remains is to efficiently solve (3.27). One approach is to compute the derivative of its objective with respect to $\mathbf{x}[n]$ and find $\mathbf{x}[n]$ such that the derivative equals zero. The derivative equals the following nonlinear scalar equation:

$$\mathbf{v}[n] + \mathbf{x}[n] \left(\alpha + \frac{\lambda}{\sqrt{(\mathbf{x}[n])^2 + \mathbf{b}[n]}} \right). \quad (3.28)$$

Setting the derivative in (3.28) to zero and assuming we have an initial estimate of $\mathbf{x}[n]$, we may solve for $\mathbf{x}[n]$ iteratively as follows:

$$(\mathbf{x}[n]_i) = -\mathbf{v}[n] \left(\alpha + \frac{\lambda}{\sqrt{(\mathbf{x}[n]_{i-1})^2 + \mathbf{b}[n] + \epsilon}} \right)^{-1}, \tag{3.29}$$

where ϵ is simply a small value that avoids ill-conditioned scenarios.

We may now formulate CBCS as *Algorithm 3.6*. As we seek to update a fixed $\mathbf{g}_1$, note how we iteratively tune its N elements, one at a time, via (3.29), but instead of moving on immediately to update $\mathbf{g}_2$, we update $\mathbf{g}_1$, $\mathbf{r}$, $\mathbf{v}$, and $\mathbf{b}$, and tune over the elements of $\mathbf{g}_1$ yet again, doing this repeatedly until the per-vector objective, (3.22), stops decreasing—only then moving on to $\mathbf{g}_2$. Empirically, we find this greatly speeds up the rate at which the $\mathbf{g}_p$s converge to a simultaneously sparse solution, but unfortunately, even with this extra loop, CBCS still requires excessive iterations for larger problems (see Sec. 3.4). Similarly to RBRS in *Algorithm 3.5*, note how the inner loops are cut off when the objective function stops decreasing to within some small value δ or some fixed number of iterations has been exceeded.

Extending CBCS to complex-valued data

If (3.3) contains complex-valued terms, we may structure the column-by-column updates as in (3.22, 3.25), but the expansion and derivative of the latter equation's objective function does not lend itself to the simple update equations given in (3.26, 3.27, 3.29). One way to overcome this problem is to turn the complex-valued problem into a real-valued one. This approach is not equivalent to the one used to extend RBRS to complex data.

First we stack the target vector, $\mathbf{d}$, into a real-valued vector:

$$\widetilde{\mathbf{d}} = \begin{bmatrix} \mathrm{Re}(\mathbf{d}) \\ \mathrm{Im}(\mathbf{d}) \end{bmatrix} \in \mathbb{R}^{2M}, \tag{3.30}$$

and then *split*, rather than stack, the unknown vectors into $2P$ new vectors:

$$\mathbf{g}_p^{(\mathrm{Re})} = \mathrm{Re}(\mathbf{g}_p) \in \mathbb{R}^N, \;\; \mathbf{g}_p^{(\mathrm{Im})} = \mathrm{Im}(\mathbf{g}_p) \in \mathbb{R}^N, \text{ for } p = 1, \ldots, P. \tag{3.31}$$

We then aggregate these vectors into $\widetilde{\mathbf{G}} = [\mathbf{g}_1^{(\mathrm{Re})}, \mathbf{g}_1^{(\mathrm{Im})}, \ldots, \mathbf{g}_P^{(\mathrm{Re})}, \mathbf{g}_P^{(\mathrm{Im})}]$. Next, we split each

Algorithm 3.6 MSSO Column-by-Column Sequential Iterative Shrinkage

Task: Minimize $\left\{ \frac{1}{2} \left\| \mathbf{d} - \mathbf{F}_{\text{tot}} \mathbf{g}_{\text{tot}} \right\|_2^2 + \lambda \left\| \mathbf{G} \right\|_{\text{S}} \right\}$ when all data is *real-valued*.

Data and Parameters: λ, $\mathbf{d}$, $\mathbf{F}_p, p \in \{1, \ldots, P\}$, δ_{tot}, δ_{vec}, δ_{elem}, K, J, I are given.

Initialize: $\mathbf{g}_{\text{tot}} = (\mathbf{F}_{\text{tot}})^{\dagger} \mathbf{d}$; split into $\mathbf{g}_1, \ldots, \mathbf{g}_P$; set $\alpha =$ max. sing. val. among $\mathbf{F}_p$s.

Iterate: Set $k = 1$ and apply:

- Sweep over column vectors: set $p = 1$ and apply:
  - Optimize a column vector: set $j = 1$ and apply:
    - Construct $\mathbf{r} = \mathbf{d} + \mathbf{F}_p \mathbf{g}_p - \sum_{q=1}^{P} \mathbf{F}_q \mathbf{g}_q$.
    - Construct $\mathbf{b}[l] = -(\mathbf{g}_p[l])^2 + \sum_{q=1}^{P} (\mathbf{g}_q[l])^2$, for $l = 1, \ldots, N$.
    - Construct $\mathbf{v} = \mathbf{F}_p^{\text{T}} \mathbf{F}_p \mathbf{g}_p - \alpha \mathbf{g}_p - \mathbf{F}_p^{\text{T}} \mathbf{r}$.
    - Set $\mathbf{x}_0 = \mathbf{g}_p$.
    - Sweep over column elements: set $n = 1$ and apply:
      - Optimize nth element of $\mathbf{x}$: set $i = 1$ and apply:
        - $(\mathbf{x}[n]_i) = -\mathbf{v}[n] \left(\alpha + \frac{\lambda}{\sqrt{(\mathbf{x}[n]_{i-1})^2 + \mathbf{b}[n] + \epsilon}} \right)^{-1}$.
        - $i = i + 1$. Stop if $i > I$ or (3.27) decreases by less than δ_{elem}.
      - $n = n + 1$. Terminate when $n > N$.
    - Update column vector: set $\mathbf{g}_p$ to equal the final $\mathbf{x}$.
    - $j = j + 1$. Terminate when $j > J$ or (3.22) decreases by less than δ_{vec}.
  - $p = p + 1$. Terminate when $p > P$.
- $k = k + 1$. Terminate loop when $k > K$ or (3.3) decreases by less than δ_{tot}.

Finalize: If λ was sufficiently large, $\mathbf{g}_1, \ldots, \mathbf{g}_P$ should be simultaneously sparse.

$\mathbf{F}_p$ into two separate matrices, for $p = 1, \ldots, P$:

$$\mathbf{F}_p^{(\text{A})} = \begin{bmatrix} \text{Re}(\mathbf{F}_p) \\ \text{Im}(\mathbf{F}_p) \end{bmatrix} \in \mathbb{R}^{2M \times N}, \; \mathbf{F}_p^{(\text{B})} = \begin{bmatrix} -\text{Im}(\mathbf{F}_p) \\ \text{Re}(\mathbf{F}_p) \end{bmatrix} \in \mathbb{R}^{2M \times N}, \tag{3.32}$$

yielding $2P$ new real-valued matrices.

Due to the structure of (3.30, 3.31, 3.32), the following optimization is *equivalent* to

(3.3):

$$\min_{\widetilde{\mathbf{G}}} \left\{ \frac{1}{2} \left\| \widetilde{\mathbf{d}} - \sum_{p=1}^{P} \mathbf{F}_p^{(\mathrm{A})} \mathbf{g}_p^{(\mathrm{Re})} - \sum_{p=1}^{P} \mathbf{F}_p^{(\mathrm{B})} \mathbf{g}_p^{(\mathrm{Im})} \right\|_2^2 + \lambda \sum_{n=1}^{N} \sqrt{\sum_{p=1}^{P} (\mathbf{g}_p^{(\mathrm{Re})}[n])^2 + \sum_{p=1}^{P} (\mathbf{g}_p^{(\mathrm{Im})}[n])^2} \right\}. \tag{3.33}$$

The equivalence arises because the first and second terms of (3.33) are equivalent to $\frac{1}{2}\|\mathbf{d} - \mathbf{F}_{\mathrm{tot}}\mathbf{g}_{\mathrm{tot}}\|_2^2$ and $\|\mathbf{G}\|_{\mathrm{S}}$ in (3.3), respectively.

This means we may apply CBCS to complex-valued problems by performing column-by-column optimization over the $2P$ real-valued unknown vectors. This works because CBCS will pursue solutions where the $\mathbf{g}_1^{(\mathrm{Re})}, \mathbf{g}_1^{(\mathrm{Im})}, \ldots, \mathbf{g}_P^{(\mathrm{Re})}, \mathbf{g}_P^{(\mathrm{Im})}$ vectors are simultaneously sparse, which is equivalent to pursuing simultaneously sparse $\mathbf{g}_1, \ldots, \mathbf{g}_P$s. After running CBCS on the $2P$ vectors, we simply restructure them into P complex-valued $\mathbf{g}_p$s.

Finally, let us set $P = 1$ and thus consider the case of single-vector sparse approximation. The above derivations show that *seeking a single sparse complex-valued vector is equivalent to seeking two* **simultaneously sparse** *real-valued vectors*. In other words, single-vector sparse approximation of a complex vector readily maps to the MSSO problem, increasing the applicability of algorithms that solve the latter.

3.3.7 Second-Order Cone Programming (SOCP)

We now propose a seventh and final algorithm to solve the MSSO problem as given in (3.3). We branch away from the shrinkage approaches that operate on individual columns or rows of the $\mathbf{G}$ matrix and again seek to concurrently estimate all PN unknowns. Rather than using an IRLS technique, however, we pursue a second-order cone programming approach, motivated by the fact that second-order cone programs may be solved via efficient interior point algorithms [120, 125] and are able to encapsulate conic, convex-quadratic [97], and linear constraints. (Quadratic programming is not an option because the $\mathbf{g}_p$s, $\mathbf{F}_p$s, and $\mathbf{d}$ may be complex.)

Second-order conic constraints are of the form $\mathbf{a} = [a_1, \ldots, a_N]^{\mathrm{T}}$ such that

$$\|[a_1, \ldots, a_{N-1}]^{\mathrm{T}}\|_2 \leq a_N. \tag{3.34}$$

The generic format of an SOC program is

$$\min_{\mathbf{x}} \ \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \text{ and } \mathbf{x} \in \mathbf{K}, \tag{3.35}$$

where $\mathbf{K} = \mathbb{R}_+^N \times \mathbf{L}_1 \times \cdots \times \mathbf{L}_N$, $\mathbb{R}_+^N$ is the N-dimensional positive orthant cone, and the $\mathbf{L}_n$s are second-order cones [97]. To convert (3.3) into the SOC format, we first write

$$\begin{gathered} \min_{\mathbf{G}} \ \left\{\tfrac{1}{2}s + \lambda \mathbf{1}^{\mathrm{T}}\mathbf{t}\right\} \\ \text{s.t. } \mathbf{z} = \mathbf{d}_{\mathrm{tot}} - \mathbf{F}_{\mathrm{tot}}\mathbf{g}_{\mathrm{tot}} \text{ and } \|\mathbf{z}\|_2^2 \leq s \\ \text{and } \|[\mathrm{Re}(\mathbf{g}_1[n]), \mathrm{Im}(\mathbf{g}_1[n]), \ldots, \mathrm{Re}(\mathbf{g}_P[n]), \mathrm{Im}(\mathbf{g}_P[n])]^{\mathrm{T}}\|_2 \leq t_n \end{gathered} \tag{3.36}$$

where $n \in \{1, \ldots, N\}$ and $\mathbf{t} = [t_1, \ldots, t_N]^{\mathrm{T}}$. The splitting of the complex elements of the $\mathbf{g}_p$s mimics the approach used when extending CBCS to complex data, and (3.36) makes the objective function linear, as required. Finally, in order to represent the $\|\mathbf{z}\|_2^2 \leq s$ inequality in terms of second-order cones, an additional step is needed. Given that $s = \frac{1}{4}(s+1)^2 - \frac{1}{4}(s-1)^2$, the inequality may be rewritten as $\mathbf{z}^{\mathrm{H}}\mathbf{z} + \frac{1}{4}(s-1)^2 \leq \frac{1}{4}(s+1)^2$ and then expressed as a conic constraint: $\|[\mathbf{z}^{\mathrm{T}}, \frac{1}{2}(s-1)]^{\mathrm{T}}\|_2 \leq \frac{1}{2}(s+1)$ [91, 97]. Applying these changes yields

$$\begin{gathered} \min \ \left\{\tfrac{1}{2}s + \lambda \mathbf{1}^{\mathrm{T}}\mathbf{t}\right\} \\ \text{s.t. } \mathbf{z} = \mathbf{d}_{\mathrm{tot}} - \mathbf{F}_{\mathrm{tot}}\mathbf{g}_{\mathrm{tot}} \text{ and } \|[\mathbf{z}^{\mathrm{T}}, u]^{\mathrm{T}}\|_2 \leq v, \\ u = (s-1)/2, v = (s+1)/2, s \geq 0, \\ \text{and } \|[\mathrm{Re}(\mathbf{g}_1[n]), \mathrm{Im}(\mathbf{g}_1[n]), \ldots, \mathrm{Re}(\mathbf{g}_P[n]), \mathrm{Im}(\mathbf{g}_P[n])]^{\mathrm{T}}\|_2 \leq t_n, \end{gathered} \tag{3.37}$$

which is a fully-defined SOC program that may be implemented and solved numerically. There is no *Algorithm* pseudocode for this technique because having set up the variables in (3.37), one may simply plug them into an SOCP solver. In this paper we implement (3.37) in SeDuMi (Self-Dual-Minimization) [120], a free software package consisting of `MATLAB` and `C` routines.

3.4 Experiments and Results

Our motivation for solving MSSO sparse approximation problems comes from MRI RF excitation pulse design. Due to the NP-hardness of the problem (3.2), there is no reasonable

way to check the accuracy of approximate solutions to these problem instances obtained with the algorithms introduced here. Thus, before turning to the MRI RF excitation pulse design problem in Sec. 3.4.3, we present several synthetic experiments. These experiments allow comparisons among algorithms and also reveal some empirical properties of the relaxation (3.3). Theoretical exploration of this relaxation is also merited but is beyond the scope of this dissertation.

All experiments are performed on a Linux server with a 3.0-GHz Intel Pentium IV processor. The system has 16 gigabytes of random access memory, ample to ensure that none of the algorithms require the use of virtual memory; this completely avoids excessive hard drive paging. MP, LSMP, IRLS, RBRS, CBCS are implemented in `MATLAB`, whereas SOCP is implemented in SeDuMi. The runtime of any method could be reduced significantly by implementing it in a completely compiled format such as `C`. Note: OMP is not evaluated because its performance always falls in between that of MP and LSMP.

3.4.1 Sparsity Profile Estimation in a Noiseless Setting

Overview

We now evaluate how well the algorithms of Sec. 3.3 estimate sparsity profiles when the underlying $\mathbf{g}_p$s are each strictly and simultaneously K-sparse and the observation $\mathbf{d}$ of (3.1) is known exactly and not corrupted by noise. This corresponds to a high-SNR source localization scenario where the sparsity profile indicates locations of emitters and our goal is to find the locations of these emitters [74, 84, 91, 92]. Our goal is to get an initial grasp of the challenges of solving the MSSO problem.

We synthetically generate real-valued sets of $\mathbf{F}_p$s and $\mathbf{g}_p$s in (3.1), apply the algorithms, and record the fraction of correct sparsity profile entries recovered by each. We vary M in (3.1) to see how performance at solving the MSSO problem varies when the $\mathbf{F}_p$s are underdetermined vs. overdetermined and also vary P to see how rapidly performance degrades as more system matrices and vectors are employed.

Details

For all trials, we fix $N = 30$ in (3.1) and $K = 3$, which means each $\mathbf{g}_p$ vector consists of thirty elements, three of which are nonzero. We consider $P \in \{1, 2, \ldots, 8\}$, and $M \in$

$\{10, 15, \ldots, 40\}$. For each of the fifty-six fixed (M, P) pairs, we create 50 random instances of (3.1). Each of the 2,800 instances is constructed and evaluated as follows:

- Pick a K-element subset of $\{1, \ldots, N\}$ uniformly at random. This is the sparsity profile.
- Create P total N-element vectors, the $\mathbf{g}_p$s. The K elements of each that correspond to the sparsity profile are filled in with draws from a Gaussian $\sim \mathcal{N}(0, 1)$ distribution; all other elements are set to zero.
- Create P total $M \times N$ matrices, the $\mathbf{F}_p$s. Each element of each matrix is determined by drawing from $\mathcal{N}(0, 1)$; each column of each matrix is normalized to have unit ℓ_2 energy.
- Compute $\mathbf{d} = \sum_{p=1}^{P} \mathbf{F}_p \mathbf{g}_p$. Shuffle $\mathbf{F}_p$s and $\mathbf{g}_p$s into $\mathbf{C}_n$s and $\mathbf{h}_n$s via (3.4, 3.5).
- Apply the algorithms:
  - MP, LSMP: iterate until K elements are chosen or the residual approximation is $\mathbf{0}$. If less than K terms are chosen, this hurts the recovery score.
  - IRLS, RBRS, CBCS, SOCP: approximate a λ oracle: proxy for a good choice of λ by looping over roughly seventy λs in $[0, 2]$, running the given algorithm each time. This sweep over λ results in high-energy, dense solutions through negligible-energy, all-zeros solutions. For each of the estimated $\widehat{\mathbf{g}}_{\text{tot}}$s (that vary with λ), estimate a sparsity profile by noting the largest ℓ_2 energy rows of the associated $\widehat{\mathbf{G}}$ matrix.[3] Remember the highest fraction recovered across all λs.

After performing the above steps, we average the results of the 50 trials associated with each fixed (M, P) to yield the average fraction of recovered sparsity profile elements.

Results

Each subplot of Fig. 3-1 depicts the average fraction of recovered sparsity profile elements versus the number of knowns, M, for a fixed value of P, revealing how performance varies as the $\mathbf{F}_p \in \mathbb{R}^{M \times N}$ matrices go from being underdetermined to overdetermined.

Recovery Trends. As the number of knowns M increases, recovery rates improve substantially, which is sensible. For large M and small P, the six algorithms behave similarly, consistently achieving nearly 100% recovery. For large P and moderate M, however, spar-

[3] For example, if the true sparsity profile is $\{1, 2, 9\}$ and the largest ℓ_2 energy rows of $\widehat{\mathbf{G}}$ are $\{2, 7, 8\}$, then the fraction of recovered sparsity profile terms equals $1/3$. Now suppose only two rows of $\widehat{\mathbf{G}}$ have nonzero energy and the profile estimate is only $\{7, 8\}$. The fraction recovered is now zero.

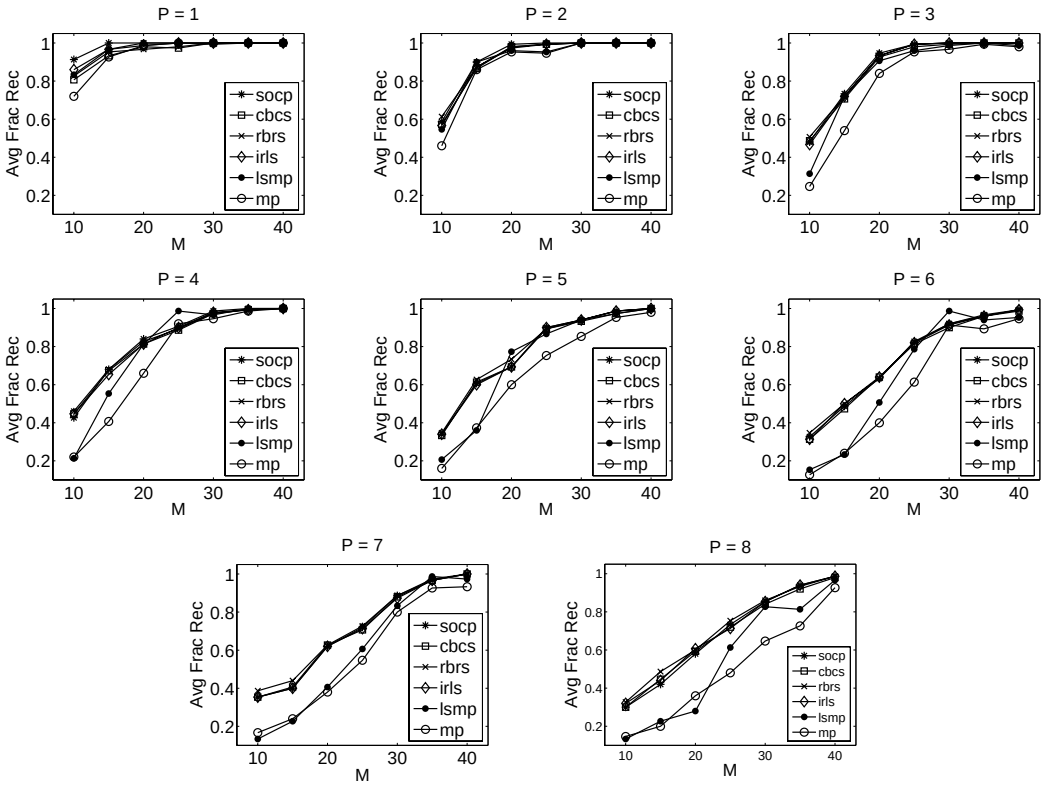

Figure 3-1: **Sparsity profile estimation in a noiseless setting**. Subplots depict average fraction of sparsity profile elements recovered over 50 trials of six algorithms as M is varied. P is fixed per subplot, and $N = 30$ and $K = 3$ for all trials. Data is generated as described in Sec. 3.4.1. Recovery scores for IRLS, RBRS, CBCS, and SOCP assume a good choice of λ is known. For large M, all algorithms exhibit high recovery rates; for large P, small M, or both, the algorithms that seek to minimize (3.3, 3.8) generally outperform those that greedily pursue a solution.

sity profile recovery rates are dismal—as P increases, the underlying simultaneous sparsity of the $\mathbf{g}_p$s is not enough to combat the increasing number of unknowns, PN. As M is decreased and especially when P is increased, the performance of the greedy techniques falls off relative to that of IRLS, RBRS, CBCS, and SOCP, showing that the convex relaxation approach itself is a sensible way to approximately solve the formal NP-Hard combinatorial MSSO simultaneous sparsity problem. Furthermore, the behavior of the convex algorithms relative to the greedy ones coincides with the studies of greedy vs. convex programming sparse approximation methods in single-vector [24,31] and SSMO contexts [32]. Essentially, in contrast with convex programming techniques, the greedy algorithms only look ahead by one term, cannot backtrack on sparsity profile element choices, and do not consider updating multiple rows of unknowns of the $\mathbf{G}$ matrix at the same time. LSMP tends to perform slightly better than MP because it solves a least squares minimization and explicitly considers earlier chosen rows whenever it seeks to choose another row of $\mathbf{G}$.

Convergence. Across most trials, IRLS, RBRS, CBCS, and SOCP converge rapidly and do not exceed the maximum limit of 500 outer iterations. The exception is CBCS when M is small and $P = 8$: here, the objective function frequently fails to decrease by less than the specified $\delta = 10^{-5}$.

Runtimes. For several fixed (M, P) pairs, Table 3.1 lists the average runtimes of each algorithm across the 50 trials associated with each pair.[4] For IRLS, RBRS, CBCS, and SOCP, runtimes are also averaged over the many λ runs. Among the convex minimization methods, SOCP seems superior given its fast runtimes in three out of four cases. Peak memory usage is not tracked here because it is difficult to do so when using MATLAB for such small problems; it will be tracked during the third experiment where the system matrices are vastly larger and differences in memory usage across the six algorithms are readily apparent.

Closer Look: Solution Vectors. We now observe how the algorithms that seek to minimize the convex objective behave during the 43rd trial when $K = 3$, $N = 30$, $M = 10$, and $P = 1$, corresponding to the base case problem of estimating one sparse real-valued vector, $\mathbf{g}_1$. Fig. 3-2 illustrates estimates obtained by SOCP, CBCS, RBRS, and IRLS when $\lambda = 0.03$; for each algorithm, a subplot shows elements of both the estimated and actual $\mathbf{g}_1$, and lists the estimated sparsity profile (ESP), number of profile terms recovered, and

[4] In the interest of space we do not list average runtimes for all fifty-six (M, P) pairs.

	(M, P)			
Algorithm	(10,8)	(20,1)	(30,5)	(40,8)
MP	5.4	1.8	2.6	4.0
LSMP	11.4	5.6	15.6	27.6
IRLS	92.6	10.1	73.2	175.0
RBRS	635.7	36.0	236.8	401.6
CBCS	609.8	7.1	191.4	396.3
SOCP	44.3	37.0	64.3	106.5

Table 3.1: **Average algorithm runtimes for noiseless sparsity profile estimation**. For several fixed (M, P) pairs, each algorithm's average runtime over the corresponding 50 trials is given in units of milliseconds; $N = 30$ and $K = 3$ for all trials (runtimes of the latter four algorithms are also averaged over the multiple λ runs per trial). MP is substantially faster than the other techniques, as expected. For larger problems, e.g. $(M, P) = (10, 8)$, the runtimes of both RBRS and CBCS are excessive relative to those of the other convex minimization techniques, IRLS and SOCP.

value of the objective function given in (3.3, 3.8). Although RBRS, CBCS, and SOCP yield slightly different solutions (among which SOCP yields the best profile estimate), they all yield an objective function equal to 0.028 ± 10^{-5}. Convex combinations of the three solutions continue to yield the same value, suggesting that the three algorithms have found solutions among a convex set that is the global solution to the objective posed in (3.3, 3.8). Given the fact that in this case SOCP outperforms RBRS and CBCS, we see that even the globally-optimal solution to the relaxed convex objective does not necessarily optimally solve the true K-sparse profile recovery problem. In contrast to the other methods, IRLS yields a slightly higher objective function value, 0.030, and its solution vector is not part of the convex set—it does however correctly determine 2 of the 3 terms of the true sparsity profile.

Closer Look: Objective Function Behavior. Concluding the experiment, Fig. 3-3 plots the objective vs. λ for the 25th trial when $M = 30$ and $P = 6$, studying how the objective (3.3, 3.8) varies with λ when applying SOCP, CBCS, RBRS, and IRLS. For all seventy values of $\lambda \in [0, 2]$, SOCP, CBCS, and RBRS generate solutions that yield the same objective function value. For $\lambda < \frac{1}{4}$, IRLS attains the same objective function value as the other methods, but as λ increases, IRLS is unable to minimize the objective function as well as SOCP, RBRS, and CBCS. The behavior in Fig. 3-3 occurs consistently across the fifty trials of the other (M, P) pairs.

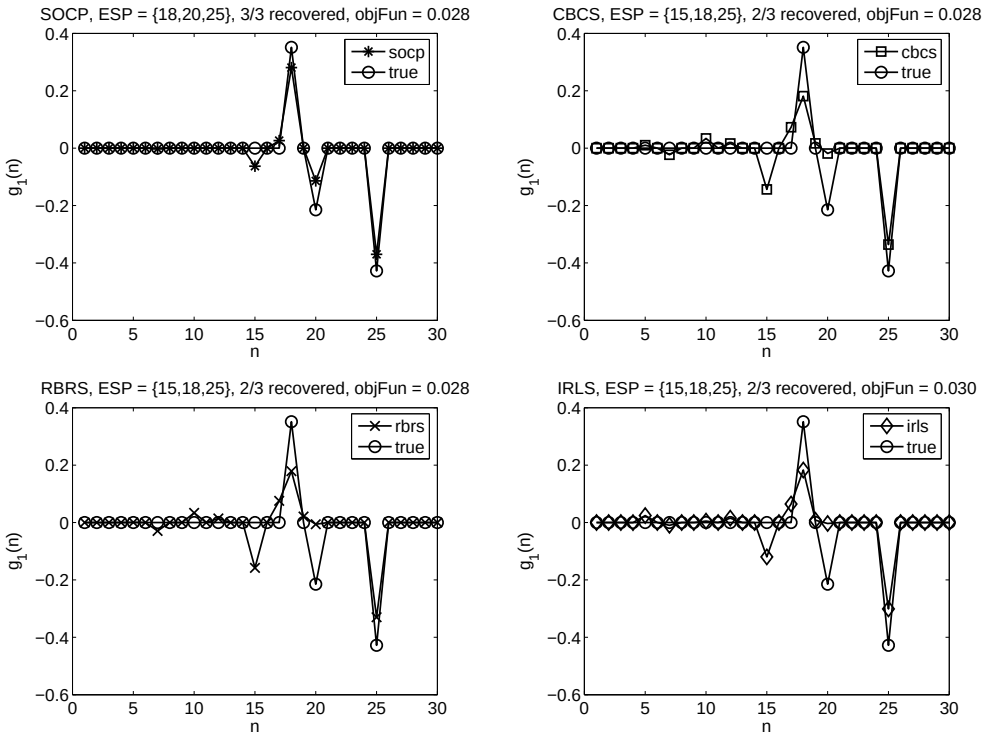

Figure 3-2: **Noiseless sparsity profile estimation with IRLS, RBRS, CBCS, SOCP**. Here $M = 10, N = 30, P = 1$, and $K = 3$. The algorithms are applied with λ fixed at 0.03 and attempt to estimate the single unknown vector, $\mathbf{g}_1$, along with the sparsity profile. Subplots depict the elements of both the estimated and actual $\mathbf{g}_1$, along with the estimated sparsity profile (ESP), number of profile terms recovered, and objective function value. SOCP leads to a superior sparsity profile estimate, and SOCP, RBRS, and CBCS seem to minimize the convex objective given in (3.3, 3.8). IRLS does not, but still manages to properly identify 2 out of 3 sparsity profile terms.

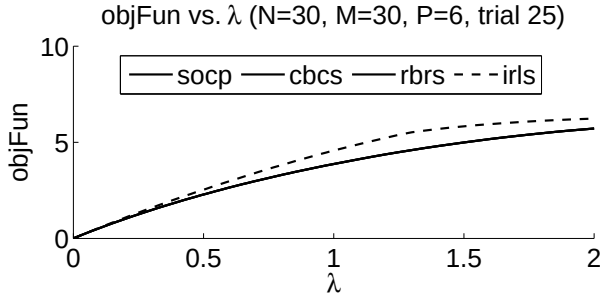

Figure 3-3: **Noiseless sparsity profile estimation: objective function behavior**. For the 25^{th} trial of the $(M, P) = (30, 6)$ series, SOCP, CBCS, RBRS, and IRLS are used to solve (3.3, 3.8) for 70 values of $\lambda \in [0, 2]$; the value of the objective function vs. λ is given above. For $\lambda > \frac{1}{4}$, IRLS's solutions do not minimize the objective as well as those produced by the three other methods.

3.4.2 Sparsity Profile Estimation in the Presence of Noise

Overview

We now evaluate how well the algorithms of Sec. 3.3 estimate sparsity profiles when the underlying $\mathbf{g}_p$s are each strictly and simultaneously K-sparse and the observation $\mathbf{d}$ of (3.1) is corrupted by additive white Gaussian noise. The signal-to-noise ratio (SNR) and K are varied across sets of Monte Carlo trials in order to gauge algorithm performance across many scenarios. For a given trial with a fixed SNR level in units of decibels (dB), the M elements of the true observation vector, $\mathbf{d}_{\text{true}}$, are corrupted with independent and identically distributed (i.i.d.) zero-mean Gaussian noise with variance σ^2, related to the SNR as follows:

$$\sigma^2 = \frac{1}{M}\|\mathbf{d}_{\text{true}}\|_2^2 \cdot 10^{-\text{SNR}/10} \tag{3.38}$$

This noise measure is analogous to that of [32].

Details

We fix $N = 30$, $M = 25$, and $P = 3$, and consider SNR $\in \{-10, -5, 0, \ldots, 25, 30\}$ and $K \in \{1, 3, 5, 7, 9\}$. For each fixed (SNR, K) pair, we generate 100 noisy observations and apply the algorithms as follows:

- Generate the sparsity profile, $\mathbf{g}_p$s, $\mathbf{F}_p$s, $\mathbf{h}_n$s, and $\mathbf{C}_n$s as in Sec. 3.4.1. The $\mathbf{g}_p$s are simultaneously K-sparse and all terms are real-valued.
- Compute $\mathbf{d}_{\text{true}} = \mathbf{F}_1\mathbf{g}_1 + \cdots + \mathbf{F}_P\mathbf{g}_P$.
- Construct $\mathbf{d}_{\text{noisy}} = \mathbf{d}_{\text{true}} + \mathbf{n}$ where $\mathbf{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ and σ^2 is given by (3.38).
- Apply the algorithms by providing them with $\mathbf{d}_{\text{noisy}}$ and the system matrices:
  - MP, LSMP: iterate until K elements are chosen or the residual approximation is $\mathbf{0}$. If less than K terms are chosen, this hurts the recovery score.
  - IRLS, RBRS, CBCS, SOCP: using a pre-determined *fixed* λ (see below), apply each algorithm to obtain estimates of the unknown vectors and sparsity profiles.

After performing the above steps, we average the results of the 100 trials associated with each fixed $(\text{SNR}, K, \text{alg})$ triplet to yield the average fraction of sparsity profile elements that each algorithm recovers.

Control Parameter Selection. The λ mentioned in the list above is determined as follows: before running the overall experiment, we generate three noisy observations for each (SNR, K) pair. We then apply IRLS, RBRS, CBCS, and SOCP, tuning the control parameter λ by hand until finding a single value that produces reasonable solutions. All algorithms then use this hand-tuned, fixed λ and are applied to the other 100 noisy observations associated with the (SNR, K) pair under consideration. Thus, in distinct contrast to the noiseless experiment, we no longer assume an ideal λ is known for each denoising trial.

Results

Each subplot of Fig. 3-4 depicts the average fraction of recovered sparsity profile elements versus SNR for a fixed K, revealing how well the six algorithms are able to recover the K elements of the sparsity profile amidst noise in the observation. Each data point is the average fraction recovered across 100 trials.

Recovery Trends. When $K = 1$, we see from the upper-left subplot of Fig. 3-4 that all algorithms have essentially equal performance for each SNR. Recovery rates improve substantially with increasing SNR, which is sensible. For each algorithm, we see across the subplots that performance generally decreases with increasing K; in other words, estimating a large number of sparsity profile terms is more difficult than estimating a small number of terms. This trend is evident even at high SNRs. For example, when SNR is 30 dB and $K = 7$, SOCP is only able to recover $\sim 70\%$ of sparsity profile terms. When $K = 9$, the recovery rate falls to $\sim 60\%$. For low SNRs, e.g., -5 dB, all algorithms tend to perform similarly, but the greedy algorithms perform increasingly worse than the others as K goes from moderate-to-large and SNR surpasses zero dB. In general, MP performs worse than LSMP, and LSMP in turn performs worse than IRLS, SOCP, RBRS, and CBCS, while the latter four methods exhibit essentially the same performance across all SNRs and Ks. For $K = 3$, MP's performance falls off relative to IRLS, SOCP, RBRS, and CBCS, but LSMP's does not. As K transitions from 3 to 5, however, LSMP performs as badly as MP at low SNRs, but its performance picks up as SNR increases. As K continues to increase beyond 5, LSMP's performance is unable to surpass that of MP, even when SNR is large. Overall, Fig. 3-4 shows that convex programming algorithms are superior to greedy methods when estimating sparsity profiles in noisy situations; this coincides with data collected in the noiseless experiment in Sec. 3.4.1, as well as the empirical findings of [31, 32].

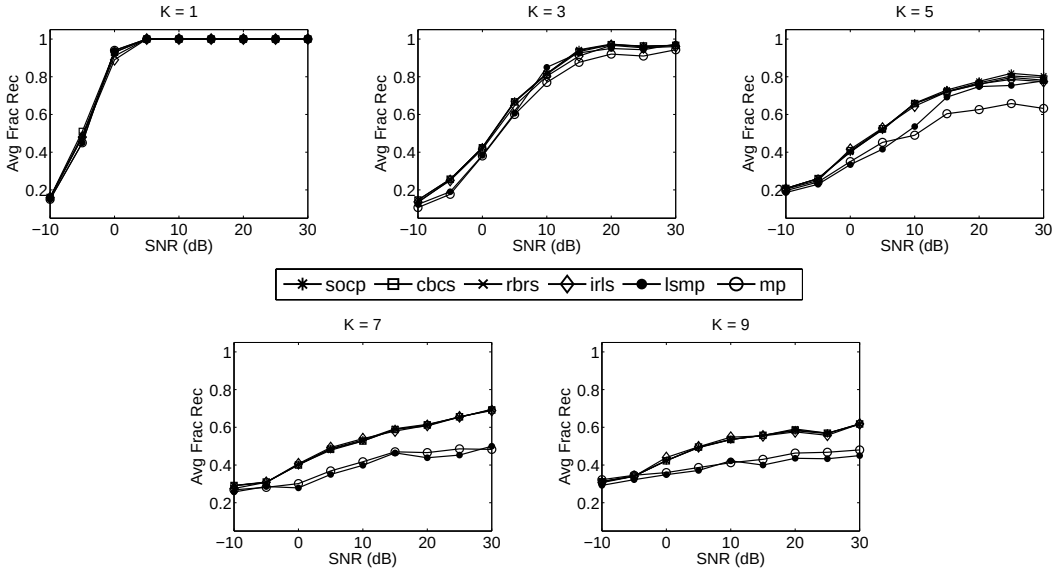

Figure 3-4: **Sparsity profile estimation in the presence of noise**. Each subplot depicts the average fraction of recovered sparsity profile elements versus SNR for a fixed K, revealing how well the algorithms recover the K elements of the sparsity profile amidst noise in the observation. Each data point is the average fraction recovered across 100 trials; data is randomly generated as described in Sec. 3.4.2. N, M and P are always fixed at 30, 25, and 3, respectively. For each (SNR, K) pair, a "good" lambda is chosen by denoising a few cases by hand and then using this fixed λ for 100 fresh denoising trials. Performance degrades with increasing K and decreasing SNR. For large K, the greedy algorithms perform worse than IRLS, SOCP, RBRS, and CBCS, whereas the latter four methods perform essentially identically across all (SNR, K) combinations.

Convergence. For many denoising trials, CBCS typically requires more iterations than the other techniques in order to converge. At times, it fails to converge to within the specified $\delta = 10^{-5}$, similarly to how it behaves during the noiseless experiment of Sec. 3.4.1.

Runtimes. Across all denoising trials, MP, LSMP, IRLS, RBRS, CBCS, SOCP have average runtimes of 3.1, 25.1, 57.2, 247.0, 148.5, and 49.2 milliseconds. It seems SOCP is best for denoising given that it is the fastest algorithm among the four methods that outperforms the greedy ones. IRLS is nearly as fast as SOCP and thus is a close second choice for sparsity profile estimation.

Closer Look: Mean Square Errors of Convex Minimization Methods Before and After Estimating the Sparsity Profile and Retuning the Solution. Now let us consider the 35th trial of the (SNR $= 0$ dB, $K = 3$) pair. We do away with the fixed λ assumption and now assume we care (to some extent) not only about estimating the sparsity profile, but the true solution $\mathbf{h}_{\text{tot}}$ as well. To proxy for this, we study how the mean square errors (MSEs) of solutions generated by IRLS, SOCP, RBRS, and CBCS behave across λ before and after identifying the sparsity profile and retuning the solution. Figure 3-5 depicts the results of this investigation.

Running each algorithm for a particular λ yields a solution $\widehat{\mathbf{h}}_{\text{tot}}(\text{alg}, \lambda)$. The left subplot simply illustrates the MSEs of the $\widehat{\mathbf{h}}_{\text{tot}}(\text{alg}, \lambda)$s with respect to the true solution. Among SOCP, RBRS, CBCS, and IRLS, only the last is able to determine solutions with MSEs less than unity (consider the IRLS error curve for $\lambda \geq 0.3$).

Consider now retuning each of the $\widehat{\mathbf{h}}_{\text{tot}}(\text{alg}, \lambda)$s as follows: unstack each into $\widehat{\mathbf{h}}_n(\text{alg}, \lambda)$ for $n \in \{1, \ldots, N\}$ and then remember the K vectors whose ℓ_2 energies are largest, yielding an estimate of the K-element sparsity profile. Let these estimated indices be $\{q_1, \ldots, q_K\}$. Now, generate a *retuned solution* by using the K matrices associated with the estimated sparsity profile and solving $\mathbf{d}_{\text{noisy}} = [\mathbf{C}_{q_1}, \ldots, \mathbf{C}_{q_K}]\mathbf{x}_{\text{tot}}$ for $\mathbf{x}_{\text{tot}} \in \mathbb{R}^{KP}$. This latter vector consists of KP elements and by unstacking it we obtain a retuned estimate of the $\widehat{\mathbf{h}}_n(\text{alg}, \lambda)$s, e.g., $\widehat{\mathbf{h}}_{q_1}(\text{alg}, \lambda)$ equals the first K elements of $\mathbf{x}_{\text{tot}}$, and so forth, while the other $\widehat{\mathbf{h}}_n(\text{alg}, \lambda)$s for $n \notin \{q_1, \ldots, q_K\}$ are now simply all-zeros vectors. Reshuffling the retuned $\widehat{\mathbf{h}}_n(\text{alg}, \lambda)$s yields $\widehat{\mathbf{g}}_p(\text{alg}, \lambda)$s that are strictly and simultaneously K sparse whose weightings yield the best match to the noisy observation in the ℓ_2 sense. Unlike the original solution estimate, which is not necessarily simultaneously K-sparse, here we have enforced true simultaneous K-sparsity. We may or may not have improved the MSE with respect to the true solution:

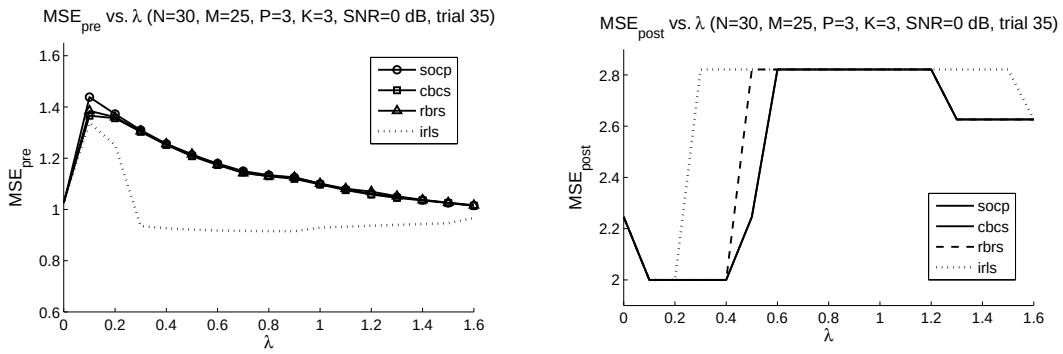

Figure 3-5: **MSEs of convex minimization methods before and after estimating the sparsity profile and retuning the solution**. Here MSE vs. λ is studied during the 35^{th} trial of the (SNR $= 0$ dB, $K = 3$) denoising series. Fixing λ and applying a given algorithm yields the solution $\widehat{\mathbf{h}}(\text{alg}, \lambda)$. Left plot: MSEs of the $\widehat{\mathbf{h}}(\text{alg}, \lambda)$s vs. the true solution $\mathbf{h}_{\text{tot}}$. Right plot: MSEs of the solution estimates after they undergo retuning to be strictly and simultaneously K-sparse. (Sec. 3.4.2 outlines the retuning process.) For all algorithms and λs, MSE increases substantially relative to the left plot. No method correctly estimates the true K-term sparsity profile and thus the retuning step causes every estimated solution to branch further away (in the MSE sense) from the actual one.

for example, if we have grossly miscalculated the sparsity profile, the MSE of the retuned solution is likely to increase substantially, but if we have estimated the true sparsity profile exactly, then the retuned solution will likely be quite close (in the ℓ_2 sense) to the true solution, and MSE will thus decrease.

The MSEs of these *retuned solutions* with respect to the true $\mathbf{h}_{\text{tot}}$ are plotted in the right subplot of Fig. 3-5. For all algorithms and λs, MSE has increased relative to the left subplot, which means that in every case our estimate of the true underlying solution has worsened. This occurs because across all algorithms and λs in Fig. 3-5, the true K-term sparsity profile is incorrectly estimated and thus the retuning step makes the estimated solution worse. The lesson here is that if one is interested in minimizing MSE in low-to-moderate SNR regimes it may be best to simply keep the original estimate of the solution rather than detect the sparsity profile and retune the result. If one is not certain that the sparsity profile estimate is accurate, retuning is likely to increase MSE by fitting the estimated solution weights to an incorrect set of generating matrices. On the other hand, if one is confident that the entire sparsity profile will be correctly identified with sufficiently high probability, retuning will be beneficial; see [46, 49, 53] for related ideas.

3.4.3 MRI RF Excitation Pulse Design from an Applied Mathematical Perspective

Overview

For the final experiment we study how well the six algorithms design MRI RF excitation pulses. Because the conversion of the physical problem into an MSSO format involves MRI physics and requires significant background, we only briefly outline how the system matrices arise and why simultaneously sparse solutions are necessary. A complete formulation of the problem will be given in Ch. 5 and is also described in [157, 167].

Formulation

For the purposes of this chapter design of an MRI RF excitation pulse reduces to the following problem: assume we are given M points in the 2-D spatial domain, $\mathbf{r}_1, \ldots, \mathbf{r}_M$, along with N points in a 2-D "Fourier-like" domain, $\mathbf{k}_1, \ldots, \mathbf{k}_N$. Each $\mathbf{r}_m$ equals $[x_m, y_m]^{\mathrm{T}}$, a point in space, while each $\mathbf{k}_n$ equals $[k_{x,n}, k_{y,n}]^{\mathrm{T}}$, a point in the Fourier-like domain, referred to as "k-space". The $\mathbf{r}_m$s and $\mathbf{k}_n$s are in units of centimeters (cm) and inverse centimeters (cm^{-1}), respectively. The $\mathbf{k}_n$s are Nyquist-spaced relative to the sampling of the $\mathbf{r}_m$s and may be visualized as a 2-D grid located at low k_x and k_y frequencies (where "k_x" denotes the frequency domain axis that corresponds to the spatial x axis). Under reasonable assumptions, energy placed at one or more points in k-space produces a pattern in the spatial domain; this pattern is related to the k-space energy via a Fourier-like transform [102]. Assume we place an arbitrary complex weight $g_n \in \mathbb{C}$ (i.e., both a magnitude and phase) at each of the N locations in k-space. Let us represent these weights using a vector $\mathbf{g} = [g_1, \ldots, g_N]^{\mathrm{T}} \in \mathbb{C}^N$. In an ideal (i.e., physically-unrealizable) setting, the following holds:

$$\mathbf{m} = \mathbf{A}\mathbf{g}, \tag{3.39}$$

where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is a known dense Fourier matrix[5] and the mth element of $\mathbf{m} \in \mathbb{C}^M$ is the image that arises at $\mathbf{r}_m$, denoted $m(\mathbf{r}_m)$, due to the energy deposition along the N points on the k-space grid as described by the weights in the $\mathbf{g}$ vector.

The goal now is to form a desired (possibly complex-valued) spatial-domain image $d(\mathbf{r})$ at the given set of spatial domain coordinates (the $\mathbf{r}_m$s) by placing energy at some of

[5] Formally, $\mathbf{A}(m,n) = j\gamma e^{i\mathbf{r}_m \cdot \mathbf{k}_n}$, where $j = \sqrt{-1}$ and γ is a known lumped gain constant.

the given $\mathbf{k}_n$ locations while obeying a special constraint on how the energy is deposited. To produce the spatial-domain image, we will use a "P-channel MRI parallel excitation system" [79, 113]—each of the system's P channels is able to deposit energy of varying magnitudes and phases at the k-space locations and is able to influence the resulting spatial-domain pattern $m(\mathbf{r})$ to some extent. Each channel p has a known "profile" across space, $S_p(\mathbf{r}) \in \mathbb{C}$, that describes how the channel is able to influence the magnitude and phase of the resulting image at different spatial locations. For example, if $S_3(\mathbf{r}_5) = 0$, then the 3rd channel is unable to influence the image that arises at location $\mathbf{r}_5$, regardless of how much energy it deposits along $\mathbf{k}_1, \ldots, \mathbf{k}_N$. The special constraint mentioned above is as follows: *the system's channels may only visit a small number of points in k-space—they may only deposit energy at $K \ll N$ locations.*

We now finalize the formulation of problem: first, we construct P diagonal matrices $\mathbf{S}_p \in \mathbb{C}^{M \times M}$ such that $\mathbf{S}_p(m,m) = S_p(\mathbf{r}_m), m = 1, \ldots, M$. Now we assume that each channel deposits arbitrary energies at each of the N points in k-space and describe the weighting of the k-space grid by the pth channel with the vector $\mathbf{g}_p \in \mathbb{C}^N$. Based on the physics of the P-channel parallel excitation system, the overall image $m(\mathbf{r})$ that forms is the *superposition* of the profile-scaled subimages produced by each channel:

$$\begin{aligned} \mathbf{m} &= \mathbf{S}_1\mathbf{A}\mathbf{g}_1 + \cdots + \mathbf{S}_P\mathbf{A}\mathbf{g}_P \\ &= \mathbf{F}_1\mathbf{g}_1 + \cdots + \mathbf{F}_P\mathbf{g}_P \\ &= \mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}, \end{aligned} \tag{3.40}$$

where $\mathbf{m} = [m(\mathbf{r}_1), \ldots, m(\mathbf{r}_M)]^{\mathrm{T}}$. Essentially, (3.40) is the real-world version of (3.39) for P-channel systems with profiles $S_p(\mathbf{r})$ that are not constant across $\mathbf{r}$.

Recalling that our overall goal is to deposit energy in k-space to produce the image $d(\mathbf{r})$, and given the special constraint that we may only deposit energy among a small subset of the N points in k-space, we arrive at the following problem:

$$\min_{\mathbf{g}_1, \ldots, \mathbf{g}_P} \|\mathbf{d} - \mathbf{m}\|_2 \ \text{ s.t. the simultaneous } K\text{-sparsity of the } \mathbf{g}_p\text{s}, \tag{3.41}$$

where $\mathbf{d} \in \mathbb{C}^M = [d(\mathbf{r}_1), \ldots, d(\mathbf{r}_M)]^{\mathrm{T}} \in \mathbb{C}^M$ and $\mathbf{m}$ is given by (3.40). That is, we seek out $K < N$ locations in k-space at which to deposit energy to produce an image $m(\mathbf{r})$ that is close in the ℓ_2 sense to the desired image $d(\mathbf{r})$. Strictly and simultaneously K-sparse $\mathbf{g}_p$s

are the only valid solutions to the problem.

One sees that (3.41) is precisely the MSSO system given in (3.2) and thus the algorithms posed in Sec. 3.3 are applicable to the pulse design problem. In order to apply the convex minimization techniques (IRLS, SOCP, RBRS, and CBCS) to this problem, the only additional step needed is to retune any given solution estimate $\widehat{\mathbf{g}}_{\text{tot}}(\text{alg}, \lambda)$ into a strictly and simultaneously K-sparse set of vectors; this retuning step is computationally tractable and was described in Sec. 3.4.2's "*Closer Look*" subsection.

Aside. An alternative approach to decide where to place energy at K locations in k-space is to compute the Fourier transform of $d(\mathbf{r})$ and decide to place energy at (k_x, k_y) frequencies where the Fourier coefficients are largest in magnitude [147]. This method does yield valid K-sparse energy placement patterns, but in Ch. 5 and [157, 162, 167] we empirically show that this technique is always outperformed by convex minimization approaches. Thus we do not delve into the Fourier-based method here.

Experimental Setup

Using an eight-channel system (i.e., $P = 8$) whose profile magnitudes (the $|S_p(\mathbf{r})|$s) are depicted in Fig. 2-6, we will design pulses to produce the desired complex-valued image shown in the left subplot of Fig. 3-6. We sample the spatial (x, y) domain at $M = 356$ locations within the region where at least one of the profiles in Fig. 2-6 is active—this region of interest is the field of excitation (FOX) discussed in Sec. 2.1.10.[6] The spatial samples are spaced by 0.8 cm along each axis and the FOX has a diameter of roughly 20 cm. Given our choice of $\mathbf{r}_1, \ldots, \mathbf{r}_{356}$, we sample the $S(\mathbf{r})$s and $d(\mathbf{r})$ and construct the $\mathbf{S}_p$s and $\mathbf{d}$. Next, we define a grid of $N = 225$ points in (k_x, k_y)-space that is 15×15 in size and extends outward from the k-space origin. The points are spaced by $\frac{1}{20}$ cm^{-1} along each k-space axis and the overall grid is shown in the right subplot of Fig. 3-6. Finally, because we know the 356 $\mathbf{r}_m$s and 225 $\mathbf{k}_n$s, we construct the 356×225 matrix $\mathbf{A}$ in (3.39, 3.40) along with the $\mathbf{F}_p$s in (3.40). We now have all the data we need to apply the algorithms and determine simultaneously K-sparse $\mathbf{g}_p$s that (approximately) solve (3.41).

We apply the algorithms and evaluate designs where the use of $K \in \{1, \ldots, 30\}$ candidate points in k-space is permitted (in practical MRI scenarios, K up to 30 is permissible).

[6] Sampling points outside of the FOX where no profile has influence is unnecessary because an image can never be formed at these points no matter how much energy any given channel places throughout k-space.

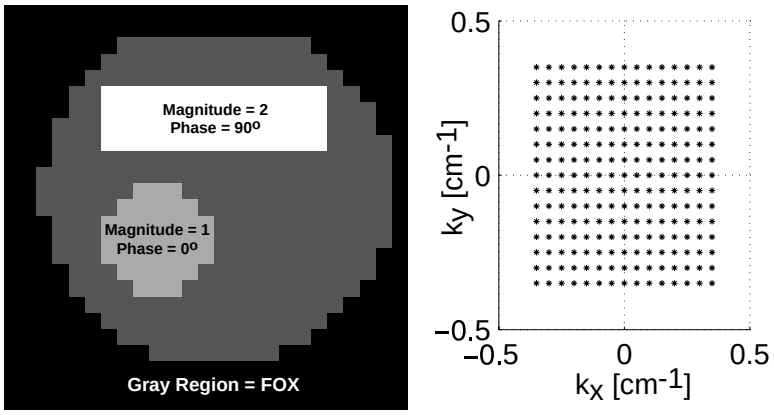

Figure 3-6: **Desired image and k-space grid**. Left image: desired complex-valued image, $d(\mathbf{r})$. Medium-gray region = FOX; other regions denote locations where we want image to be nonzero with the given magnitudes and phases. Sampling $d(\mathbf{r})$ at the 356 locations within the FOX allows us to construct $\mathbf{d}$ in (3.41). Right subplot: 15×15 grid of $N = 225$ candidate k-space locations, $\mathbf{k}_1, \ldots, \mathbf{k}_{225}$, at which the P channels may deposit energy and thus influence the resulting image. The physical constraints of the MRI excitation process force us to place energy at only a small number of grid locations.

Typically, the smallest K possible that produces a version of $d(\mathbf{r})$ to within some ℓ_2 fidelity is the design that the MRI pulse designer will use on a real system since this will correspond to the shortest pulse that accomplishes the desired task.

To obtain simultaneously K-sparse solutions with MP and LSMP, we set $K = 30$, run each algorithm once, remember the ordered list of chosen indices, and back out every solution for $K = 1$ through $K = 30$ via the retuning technique of Sec. 3.4.2.

For each convex minimization method (IRLS, SOCP, RBRS, and CBCS), we apply the following procedure: first, we run the algorithm for 14 values of $\lambda \in \left[0, \frac{1}{4}\right]$, storing each resulting solution, $\widehat{\mathbf{g}}_{\text{tot}}(\text{alg}, \lambda)$. Then for fixed K, to determine a simultaneously K-sparse deposition of energy on the k-space grid, we apply the retuning process of Sec. 3.4.2 to each of the 14 solutions, obtaining 14 strictly simultaneously K-sparse retuned sets of solution vectors, $\widehat{\mathbf{g}}_{\text{tot}}^{(K)}(\text{alg}, \lambda)$. The one solution among the 14 that best minimizes the ℓ_2 error between the desired and resulting images, $\|\mathbf{d} - \mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}^{(K)}(\text{alg}, \lambda)\|_2$, is chosen as the solution for the K under consideration. Essentially, we again assume we know a good value for λ when applying each of the convex minimization methods. To attempt to avoid convergence problems, RBRS and CBCS are permitted 5,000 and 10,000 maximum outer iterations, respectively (see below).

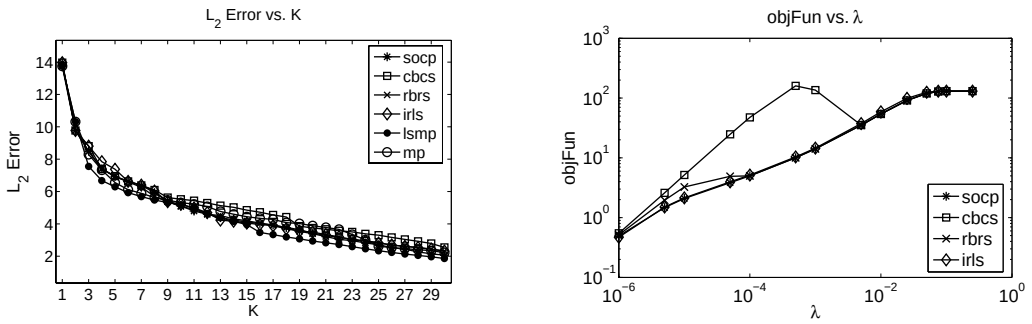

Figure 3-7: **MRI pulse design results**. Left subplot: ℓ_2 error vs. K is given for MP, LSMP, IRLS, RBRS, CBCS, and SOCP. For fixed K, LSMP consistently outperforms the other algorithms. Right subplot: objective function values vs. λ when SOCP, CBCS, RBRS, and IRLS attempt to minimize (3.3, 3.8). SOCP and IRLS converge and seem to minimize the objective; RBRS does so as well for most λs. CBCS routinely fails to converge even after 10,000 iterations and thus its solutions yield large objective function values.

Results

Image ℓ_2 Error vs. Number of Energy Depositions in k-Space. Figure 3-7's left subplot shows the ℓ_2 error versus K curves for each algorithm. As K is increased, each method produces images with lower ℓ_2 error, which is sensible: depositing energy at more locations in k-space gives each technique more degrees of freedom with which to form the image. In contrast to the sparsity profile estimation experiments in Sec. 3.4.1 and Sec. 3.4.2, however, here LSMP is the best algorithm: for each fixed K considered in Fig. 3-7, the LSMP technique yields a simultaneously K-sparse energy deposition that produces a higher-fidelity image than all other techniques. For example, when $K = 17$ LSMP yields a solution that leads to an image with ℓ_2 error of 3.3. In order to produce an image with equal or better fidelity, IRLS, RBRS, and SOCP need to deposit energy at $K = 21$ points in k-space, and thus produce less useful designs from an MRI perspective. CBCS fares the worst, needing to deposit energy at $K = 25$ grid points in order to compete with the fidelity of LSMP's $K = 17$ image.

Closer Look: Objective Function vs. λ. The right subplot of Fig 3-7 shows how well the four convex minimization algorithms minimize the objective function (3.3, 3.8) before retuning any solutions and enforcing strict simultaneous K-sparsity. For each fixed λ, SOCP and IRLS find solutions that yield the same objective function value. RBRS's solutions generally yield objective function values equal to those of SOCP and IRLS, but at times

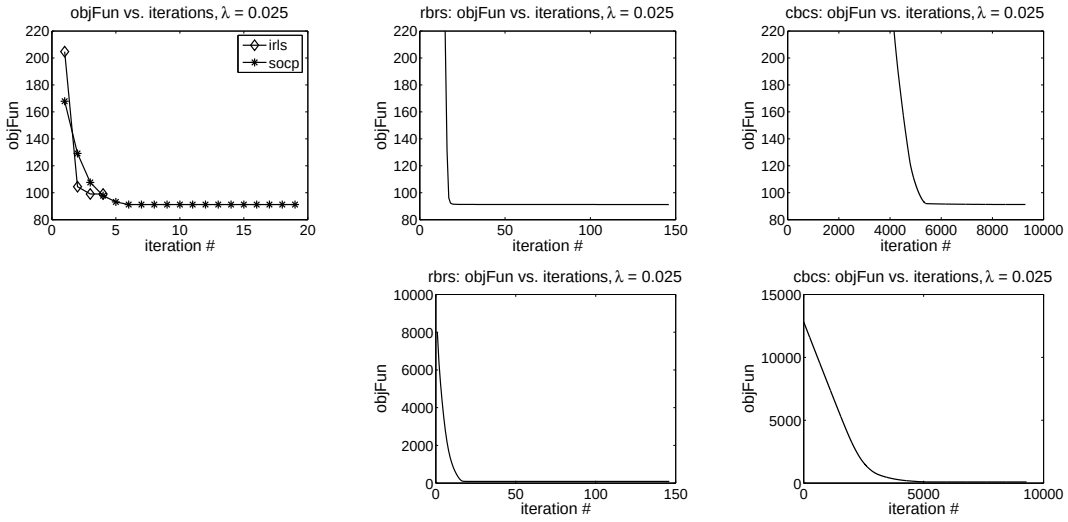

Figure 3-8: **Convergence behavior of IRLS, SOCP, RBRS, CBCS**. Each subplot illustrates the value of each algorithm's objective function (3.3, 3.8) as the algorithm iterates. Upper row subplots are scaled to have the same y axis, whereas the bottom row subplots are "zoomed out" to illustrate the overall behavior of RBRS and CBCS. IRLS and SOCP converge rapidly, within 4 and 19 iterations, respectively. RBRS and CBCS require roughly 150 and 10,000 iterations, respectively. The runtimes of IRLS, SOCP, RBRS, and CBCS in this case are 29, 121, 450, and 5,542 seconds.

lead to higher values: in these cases RBRS almost converges but does not do so completely. Finally, for most λs CBCS's solutions yield extremely large objective function values; in these cases CBCS completely fails to converge.

Closer Look: Objective Function Convergence for $\lambda = 0.025$. The right subplot of Fig 3-7 shows that for $\lambda = 0.025$, IRLS, SOCP, RBRS, and CBCS generate solutions that yield the same objective function value, suggesting that each method succeeds at minimizing the objective function. Figure 3-8 illustrates how the algorithms converge in this specific case: each subplot tracks the value of an algorithm's objective function as it iterates. Subplots along the top row all have the same y axis, giving a close look at how the algorithms behave. The two subplots along the bottom row "zoom out" along the y axis to show RBRS's and CBCS's total behavior. IRLS and SOCP converge rapidly, within 4 and 19 iterations, respectively. RBRS requires roughly 150 outer iterations, while CBCS requires nearly 10,000.

Runtimes and Peak Memory Usage. Setting $K = 30$, we run MP and LSMP and record the runtime of each. Across the 14 λs, IRLS, RBRS, CBCS, and SOCP's runtimes are recorded and averaged. The peak memory usage of each algorithm is also noted; these

Algorithm	Runtime	Peak Memory Usage (MB)
MP	11 sec	704
LSMP	46 min	304
IRLS	50 sec	320
RBRS	87 min	320
CBCS	3.3 hr	320
SOCP	96 sec	432

Table 3.2: **Algorithm runtimes and peak memory usage for MRI pulse design**. Each algorithm's runtime and peak memory usage is listed. The runtimes of the latter four algorithms are averaged over the fourteen λs per trial. MP is again faster than the other techniques, but consumes more memory because of its precomputation step (see *Algorithm 3.1*). IRLS and SOCP are quite similar performance-wise and minimize the convex objective function equally well (see Fig. 3-7), but we see here that IRLS is approximately 1.9 times faster and uses 1.4 times less peak memory than SOCP, making the former the superior technique among the convex methods.

statistics are presented in Table 3.2. In distinct contrast to the smaller-variable-size experiments in Sec. 3.4.1 and Sec. 3.4.2 where all four convex minimization methods have relatively short runtimes (under one second), here RBRS and CBCS are much slower, leaving IRLS and SOCP as the only feasible techniques among the four. Furthermore, while LSMP does indeed outperform IRLS and SOCP on an ℓ_2 error basis (as shown in Fig. 3-7), the runtime statistics here show that LSMP requires order-of-magnitude greater runtime to solve the problem—therefore, in some real-life scenarios where designing pulses in less than 5 minutes is a necessity, IRLS and SOCP are superior. Finally, in contrast to Sec. 3.4.1's runtimes given in Table 3.1, here IRLS is 1.9 times faster than SOCP and uses 1.4 times less peak memory, making it the superior technique for MRI pulse design since IRLS's ℓ_2 error performance and ability to minimize the objective function (3.3, 3.8) essentially equal that of SOCP.

Closer Look: Images and Chosen k-Space Locations for $K = 17$. To conclude the experiment, we fix $K = 17$ and observe the images produced by the algorithms along with the points at which each algorithm chooses to deposit energy along the grid of candidate points in (k_x, k_y)-space. Figure 3-9 illustrates the images (in both magnitude and phase) that arise due to each algorithm's simultaneously 17-sparse set of solution vectors,[7] while Fig. 3-10 depicts the placement pattern chosen by each method. From Fig. 3-9, we see

[7]Each image is generated by taking the corresponding solution $\mathbf{g}_{\text{tot}}$, computing $\mathbf{m}$ in (3.40), unstacking the elements of $\mathbf{m}$ into $m(\mathbf{r})$, and then displaying the magnitude and phase of $m(\mathbf{r})$.

that each algorithm forms a high-fidelity version of the desired image $d(\mathbf{r})$ given in the left subplot of Fig. 3-6, but among the images, LSMP's most accurately represents $d(\mathbf{r})$ (e.g., consider the sharp edges of the LSMP image's rectangular box). MP's and CBCS's images are noticeably fuzzy relative to the others. The placements in Fig. 3-10 give insight into these performance differences. Here, LSMP places energy at several higher frequencies along the k_y and k_x axes, which ensures the resulting rectangle is narrow with sharp edges along the spatial y and x axes. In contrast, CBCS fails to place energy at moderate-to-high (k_x, k_y)-space frequencies and thus cannot produce a rectangle with desirable sharp edges, while MP branches out to some extent but fails to utilize high k_y frequencies. IRLS, RBRS, and SOCP branch out to higher k_y frequencies but not to high k_x frequencies, and thus their associated rectangles in Fig. 3-9 are sharp along the y axis but exhibit less distinct transitions (more fuzziness) along the spatial x axis. In general, each algorithm has determined 17 locations at which to place energy that yield a fairly good image and each has avoided the computationally impossible scenario of searching over all N-choose-K (225-choose-17) possible placements.

3.5 Discussion

3.5.1 MRI Pulse Design vs. Denoising and Source Localization

The MRI pulse design problem in Sec. 3.4.3 differs substantially from the source localization problem in Sec. 3.4.1, the denoising experiment in Sec. 3.4.2, and other routine applications of sparse approximation (e.g. [24, 31, 32, 38, 44, 49, 92]). It differs not only in purpose but in numerical properties, the latter of which are summarized in Table 3.3. While this list will not always hold true on an application-by-application basis, it does highlight general differences between the two problem classes.

3.5.2 Merits of Row-by-Row and Column-by-Column Shrinkage

Even though LSMP, IRLS, and SOCP tend to exhibit superior performance across different experiments in this manuscript, RBRS and CBCS are worthwhile because unlike the former methods that update all PN unknowns concurrently, the shrinkage techniques update only a subset of the total variables during each iteration.

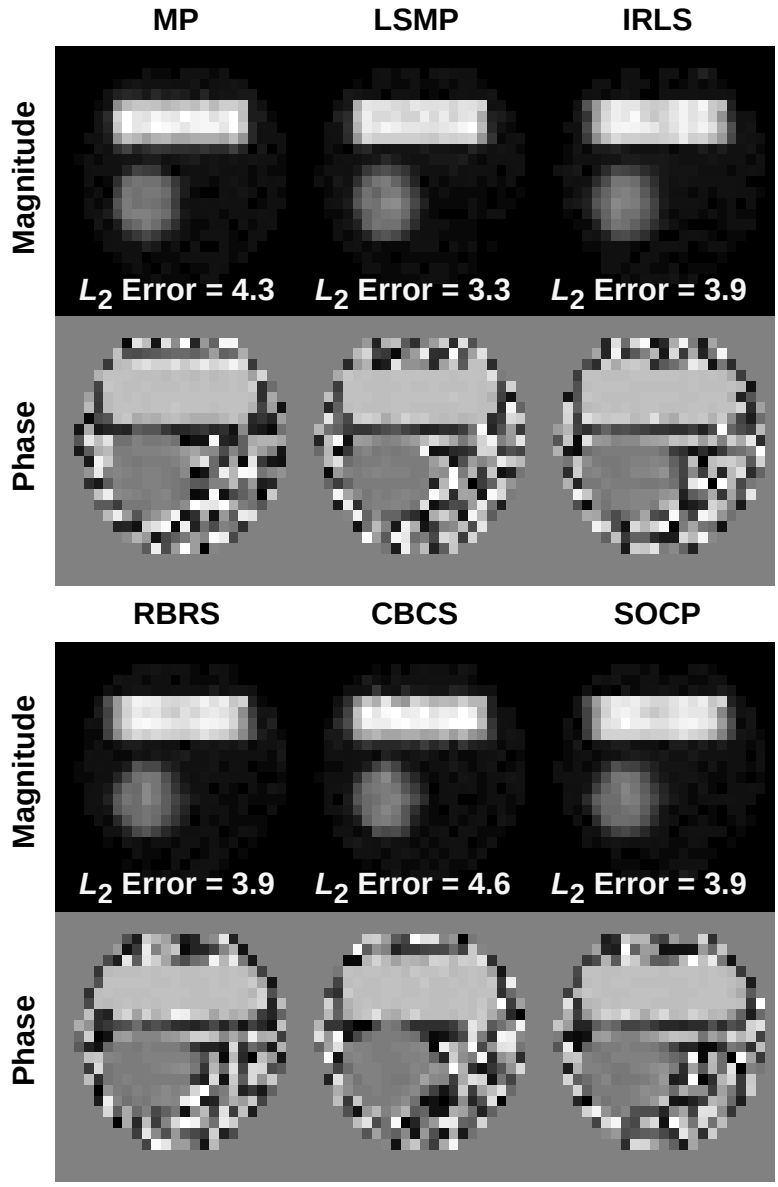

Figure 3-9: **MRI pulse design: images per algorithm for** $K = 17$. Each algorithm is used to solve the MRI pulse design problem using 17 energy depositions along the k-space grid, attempting to produce an image close to the desired one, $d(\mathbf{r})$, given in the left subplot of Fig. 3-6. From each set of simultaneously 17-sparse solution vectors, we calculate the resulting image via (3.40) and display both its magnitude and phase. LSMP's image best resembles the desired one; IRLS's, RBRS's, and SOCP's images are nearly as accurate; MP's and CBCS's images lack crisp edges, coinciding with their larger ℓ_2 errors.

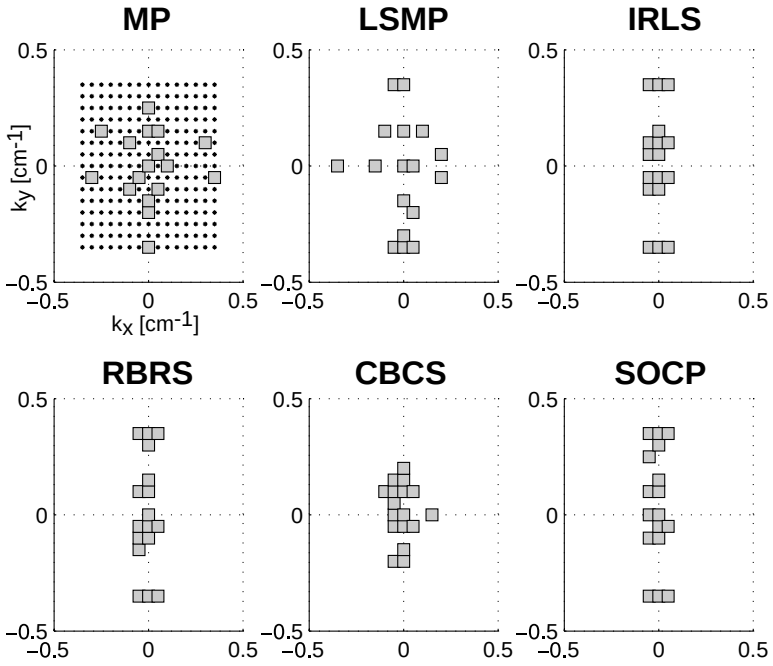

Figure 3-10: **MRI pulse design: energy deposition patterns per algorithm for** $K = 17$. Each algorithm's placement of energy in k-space is displayed. LSMP branches out to moderate k_x frequencies and high k_y frequencies, partly explaining the superiority of its image in Fig. 3-9. IRLS, RBRS, and SOCP succeed in branching out to higher k_y frequencies but do not place energy at $|k_x| \gg 0$. MP and CBCS fail to spread their energy to high spatial frequencies, and thus their images in Fig. 3-9 lack distinct edges and appear as "low-pass filtered" versions of $d(\mathbf{r})$.

MRI Pulse Design	**Denoising and Source Localization**
• $\mathbf{F}_p$s overdetermined	• $\mathbf{F}_p$s underdetermined
• No concept of noise: given $\mathbf{d}$ is $\mathbf{d}_{\text{true}}$	• Noisy: given $\mathbf{d}$ is not $\mathbf{d}_{\text{true}}$
• Sweep over λ useful	• Ideal λ unknown
• Metric: $\|\mathbf{d} - \mathbf{m}\|_2$	• Metrics: $\|\mathbf{g}_{\text{tot}} - \widehat{\mathbf{g}}_{\text{tot}}\|_2$, and/or fraction of rec. sparsity profile terms

Table 3.3: **Unique trends of the MRI pulse design problem**. This table highlights differences between the MRI problem and standard denoising and source localization applications. Items here will not always be true, instead providing general highlights about each problem class.

For example, RBRS as given in *Algorithm 3.5* updates only P unknowns at once, while CBCS as given in *Algorithm 3.6* updates but a single scalar at a time. RBRS and CBCS are thus applicable in scenarios where P and N are exceedingly large and tuning all PN parameters during each iteration is not possible. If storing and handling $M \times PN$ or $PN \times PN$ matrices exceeds a system's available memory and causes disk thrashing, RBRS and CBCS, though they require far more iterations, might still be better options than LSMP, IRLS, and SOCP in terms of runtime.

3.6 Conclusion

We defined the linear inverse multiple-system, single-output (MSSO) simultaneous sparsity problem where simultaneously sparse sets of unknown vectors are required as the solution. This problem differed from prior problems involving multiple unknown vectors because in this case, each unknown vector was passed through a different system matrix and the outputs of the various matrices underwent linear combination, yielding only one observation vector.

To solve the proposed MSSO problem, we formulated three greedy techniques, matching pursuit, orthogonal matching pursuit, and least squares matching pursuit, along with algorithms based on iteratively reweighted least squares, iterative shrinkage, and second-order cone programming methodologies. The MSSO algorithms were evaluated across noiseless and noisy sparsity profile estimation experiments as well as a magnetic resonance imaging pulse design experiment; for sparsity profile recovery, algorithms that minimized the relaxed convex objective function outperformed the greedy methods, whereas in the noiseless magnetic resonance imagine pulse design experiment, greedy LSMP exhibited superior performance.

Finally, when deriving CBCS for complex-valued data, we proved that seeking a single sparse complex-valued vector is equivalent to seeking two simultaneously sparse real-valued vectors—we mapped single-vector sparse approximation of a complex vector to the MSSO problem, increasing the applicability of algorithms that solve the latter.

Overall, while improvements upon these seven algorithms (and new algorithms altogether) surely do exist, we have laid the groundwork of the MSSO problem and conducted an initial exploration of algorithms with which to solve it.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Comparison of Three Algorithms for Solving Linearized Systems of Parallel Excitation Pulse Design Equations: Experiments on an Eight-Channel System at 3 Tesla

4.1 Introduction

In this chapter we investigate three algorithms for solving the linearized system of parallel excitation RF waveform design equations given in (2.14, 2.15) by conducting several experiments. The artifact levels and RF peak and root-mean-square (RMS) voltages of pulses generated by each method are analyzed along with algorithm runtime. The pulses calculated using these methods are used in both Bloch simulations and imaging experiments on an actual 8-channel parallel excitation coil array implemented on a 3T human scanner. Specifically, RF waveforms are designed for accelerated 2-D spiral k-space trajectories to produce a variety of 2-D target excitations and for a 3-D spokes trajectory to produce a uniform thin-slice excitation. The material of this chapter first appeared in [165, 166] and is patented [150].

In Experiment 1 (E1), pulses are designed using each algorithm to produce a square

target pattern for $R = 1$, 4, 6, and 8 spiral trajectories. For each algorithm and R value, the mean-square error (MSE) between the resulting Bloch-simulated excitation and target is calculated, which quantifies each method's excitation artifacts. Each waveform's peak voltage, V_{peak}, and root-mean-square (RMS) voltage, V_{RMS}, are also determined. In Experiment 2 (E2), the trajectories are again spirals and the target is a text logo. In addition to Bloch simulation analyses, the waveforms are played through a fully-implemented eight-channel parallel excitation system at 3T and actual excitations are analyzed. In Experiment 3 (E3), the trajectory is a fixed set of "spokes" in k_z that sample the (k_x, k_y)-plane to achieve slice selection in z [108, 113, 132] and the in-plane target is a uniform pattern. Finally, in Experiment 4 (E4), the trajectory is an $R = 8$ spiral and the target is again the text logo. Thousands of pulses are designed by looping over each method's primary control parameter, which provides extensive empirical data that shows how well each method trades off excitation quality with V_{peak} and V_{RMS}.

For each experiment, after fixing the target and trajectory, the Bloch equations relating the RF waveforms and target excitation are first linearized using the formalism of [59] as presented in Sec. 2.1.11, to which the reader may refer if the notation or formulation of the matrices here seems unfamiliar. After linearizing the system, each design method is used to generate a set of pulses. The methods have different regularizations and implementations that influence their optimization criteria and finite-precision arithmetic effects, which in turn strongly affect the resulting pulses, causing each method to produce a unique waveform and excitation.

The first design algorithm involves an approximate pseudoinverse generated via singular value decomposition (SVD), a popular approach for least-squares problems whose use is analytically justifiable [51, 118]. The other methods are Conjugate Gradient Least-Squares (CGLS) and Least-Squares QR (LSQR), iterative CG optimization algorithms for solving large linear systems [62, 100, 101]. An early use of an MRI-related CG method was the reconstruction of sensitivity encoded (SENSE) data by Pruessmann and Kannengießer [76, 104, 105]. More recently, CG methods have been used to design pulses for a single-channel system [146] and an emulated parallel excitation system [59].

4.2 Multi-Channel Parallel Excitation Pulse Design

Following the approach of [59] outlined in Sec. 2.1.11, we arrive at the system of equations given in (2.15) that allows us to design parallel excitation pulses to play through a P-channel system. We restate this system below for easy reference:

$$\mathbf{d} = \mathbf{A}\mathbf{b}, \tag{4.1}$$

where we have chosen to use $\mathbf{A}$ and $\mathbf{b}$ in place of the more cumbersome variables $\mathbf{A}_{\text{tot}}$ and $\mathbf{b}_{\text{tot}}$ and we assume that the coil profiles $S_1(\mathbf{r}), \ldots, S_P(\mathbf{r})$ are known. The vector $\mathbf{d}$ consists of N_s elements, while $\mathbf{A}$ and $\mathbf{b}$ are of size $N_s \times PN_t$ and $PN_t \times 1$, respectively; all three of these terms may be complex-valued. For all experiments considered in this chapter, $N_s = 1466$.

Recall that choosing a desired excitation pattern $d(\mathbf{r})$ and k-space trajectory $\mathbf{k}(t)$ implicitly determines $\mathbf{d}$ and $\mathbf{A}$, where $\mathbf{d}$ is $N_s \times 1$, formed by ordering the elements of $d(\mathbf{r})$ within the FOX. It is then necessary to find a candidate vector $\mathbf{b}$ that approximately solves $\mathbf{d} = \mathbf{A}\mathbf{b}$. Once $\mathbf{b}$ is found, voltage samples of the P waveforms $(b_1(t), \ldots, b_P(t))$ may be extracted, played through a Bloch simulation or actual system, and an excitation pattern may be recorded, the latter of which will resemble the target if the small-tip angle approximation [102] holds and the SNR is sufficient.

4.3 Three Algorithms for Solving a Linear System

4.3.1 SVD-Based Truncated Pseudoinversion

One may solve (4.1) via a truncated pseudoinverse generated by an SVD [51, 118], seeking a solution that minimizes $\|\mathbf{d} - \mathbf{A}\mathbf{b}\|_2$. This is accomplished with the Moore-Penrose pseudoinverse of $\mathbf{A}$, denoted as $\mathbf{A}^\dagger$, yielding $\mathbf{b}^{\text{opt}} = \mathbf{A}^\dagger\mathbf{d}$. To generate $\mathbf{A}^\dagger$, an SVD is used to decompose $\mathbf{A}$ into $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$, where $\mathbf{U}$ and $\mathbf{V}$ are $N_s \times N_s$ and $PN_t \times PN_t$ eigenvector matrices and $^{\text{H}}$ is the complex transpose. If $\mathbf{A}$ is of rank $J < \min(N_s, PN_t)$, then $\mathbf{\Sigma}$ is $N_s \times PN_t$ and diagonal, and its diagonal elements $\sigma_1 \geq \ldots \geq \sigma_J > 0$ are the nonzero singular values (SVs) of $\mathbf{A}$. Formally,

$$\mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}^{+}\mathbf{U}^{\text{H}} = \sum_{j=1}^{J} \sigma_j^{-1}\mathbf{v}_j\mathbf{u}_j^{\text{H}}, \tag{4.2}$$

where $\mathbf{u}_j$ and $\mathbf{v}_j$ are the jth columns of $\mathbf{U}$ and $\mathbf{V}$, and $\mathbf{\Sigma}^+$ is implicitly defined. When $\mathbf{A}$ is ill-conditioned, (4.2) yields a poor candidate for $\mathbf{b}$ because it uses all nonzero SVs, even those representing only the system's noise subspace. However, by retaining only the first $K < J$ singular values and avoiding the use of the smaller ones, a better-conditioned truncated pseudoinverse is obtained:

$$\mathbf{A}_K^\dagger = \sum_{k=1}^{K} \sigma_k^{-1} \mathbf{v}_k \mathbf{u}_k^{\mathrm{H}}, \tag{4.3}$$

allowing one to obtain a better-conditioned estimate $\mathbf{b}_K = \mathbf{A}_K^\dagger \mathbf{d}$. Therefore, K is this method's control parameter: as it is increased, the error $\|\mathbf{d} - \mathbf{A}\mathbf{b}_K\|_2$ decreases whereas the energy of the solution vector, $\|\mathbf{b}_K\|_2$, increases. One typically applies this method by retaining just enough SVs to yield a solution with acceptably low residual error while keeping $\|\mathbf{b}\|_2$ as small as possible. For large matrices this algorithm is slow because it computes an SVD, but for fixed K there exist fast methods to compute $\mathbf{A}_K^\dagger$ directly.

4.3.2 Conjugate Gradient Least-Squares (CGLS)

This algorithm solves the following optimization problem:

$$\min_{\mathbf{b}} \ \left\| \left(\mathbf{A}^{\mathrm{H}}\mathbf{A} + \lambda_{\mathrm{CGLS}}\mathbf{I} \right) \mathbf{b} - \mathbf{A}^{\mathrm{H}}\mathbf{d} \right\|_2 \tag{4.4}$$

where λ_{CGLS} is a regularization term. One sees from (4.4) that as λ_{CGLS} is increased, $\|\mathbf{b}\|_2$ decreases and residual error $\|\mathbf{r}\|_2 = \|\mathbf{d} - \mathbf{A}\mathbf{b}\|_2$ increases. CGLS does not perform an SVD and requires only $2N_s + 3PN_t$ complex multiplications per iteration i. When λ_{CGLS} is zero, CGLS is identical to Hestenes and Stiefel's iterative CG method for least-squares problems [62].

Pseudocode for CGLS is given in *Algorithm 4.1* and shows how the sequence of approximations $\mathbf{b}_i$ is generated. Analytically, the $\mathbf{b}_i$ are such that the residual error $\|\mathbf{r}_i\|_2$ decreases monotonically [62]. When solving (4.4), users may restrict the number of iterations or specify a threshold ϵ such that CGLS halts when $\|\mathbf{s}_i\|_2^2 / \|\mathbf{s}_0\|_2 < \epsilon$. CGLS may also incorporate preconditioning matrices, weighted norms, and initial conditions.

CGLS is similar to Sutton et al.'s CG method (SCG) [121] used in [146] and [59] for pulse design because both CGLS and SCG are based on the Hestenes-Stiefel method. Step

Algorithm 4.1 Conjugate Gradient Least Squares (CGLS)

1. Set $\mathbf{r}_0 = \mathbf{d}$, $\mathbf{s}_0 = \mathbf{A}^{\mathrm{H}}\mathbf{d}$, $\mathbf{p}_1 = \mathbf{s}_0$, $\gamma_0 = \|\mathbf{s}_0\|_2^2$, $\mathbf{b}_0 = \mathbf{0} = [0, \ldots, 0]^{\mathrm{T}}$
2. For $i = 1, 2, \ldots$ repeat the following:
   (a) $\mathbf{q} = \mathbf{A}\mathbf{p}_i$ (Hestenes and Stiefel's intermediate vector)
   (b) $\delta_i = \|\mathbf{q}_i\|_2^2 + \lambda_{\mathrm{CGLS}}\|\mathbf{p}_i\|_2^2$ (incorporate regularization term)
   (c) $\alpha_i = \gamma_{i-1}/\delta_i$ (calculate step size)
   (d) $\mathbf{b}_i = \mathbf{b}_{i-1} + \alpha_i\mathbf{p}_i$ (update set of RF waveforms)
   (e) $\mathbf{r}_i = \mathbf{r}_{i-1} - \alpha_i\mathbf{q}_i$ (update residual error vector)
   (f) $\mathbf{s}_i = \mathbf{A}^{\mathrm{H}}\mathbf{r}_i - \lambda_{\mathrm{CGLS}}\mathbf{b}_i$ (incorporate regularization term)
   (g) $\gamma_i = \|\mathbf{s}_i\|_2^2$
   (h) $\beta_i = \gamma_i/\gamma_{i-1}$
   (i) $\mathbf{p}_{i+1} = \mathbf{s}_i + \beta_i\mathbf{p}_i$
   (j) Test for convergence. Exit if a stopping criterion is met.

2(c) of CGLS given in *Algorithm 4.1* shows that the numerator of the step size α_i equals $\|\mathbf{s}_{i-1}\|_2$, and thus is guaranteed to be nonnegative real (zero if an exact solution is reached). The numerator of SCG's step size, however, is $\mathbf{p}_i^{\mathrm{H}}\mathbf{s}_{i-1} = (\mathbf{s}_{i-1} + \beta_i\mathbf{p}_{i-1})^{\mathrm{H}}\mathbf{s}_{i-1}$, and thus not guaranteed to be positive. We confirmed this numerically by providing SCG with randomly generated inputs and consistently observing complex-valued step sizes.

4.3.3 Least-Squares QR (LSQR)

This algorithm is an implementation of Tikhonov regularization and solves large linear least-squares problems in a numerically attractive manner [123, 124]. Its name comes from its use of the QR decomposition [51, 118]. The algorithm has one regularization parameter, λ_{LSQR}, and solves the following:

$$\min_{\mathbf{b}} \left(\|\mathbf{d} - \mathbf{A}\mathbf{b}\|_2^2 + \lambda_{\mathrm{LSQR}}^2 \|\mathbf{b}\|_2^2 \right). \tag{4.5}$$

As λ_{LSQR} is increased, more weight is placed on the energy of $\mathbf{b}$ than on the residual error, causing $\|\mathbf{b}\|_2$ to decrease and $\|\mathbf{d} - \mathbf{A}\mathbf{b}\|_2$ to increase. LSQR also avoids use of an SVD; it requires $3N_s + 5PN_t$ complex multiplications per iteration. Pseudocode for LSQR when $\lambda_{\mathrm{LSQR}} = 0$ is given in *Algorithm 4.2*, based on Sec. 4 of [101].

LSQR, like CGLS, generates $\mathbf{b}_i$ such that $\|\mathbf{r}_i\|_2$ decreases monotonically, but LSQR

Algorithm 4.2 Least-Squares QR (LSQR)

1. Initialize and begin the bidiagonalization.
   $\beta_1\mathbf{u}_1 = \mathbf{d}$, $\alpha_1\mathbf{v}_1 = \mathbf{A}^{\mathrm{H}}\mathbf{u}_1$, $\mathbf{w}_1 = \mathbf{v}_1$, $\mathbf{b}_0 = \mathbf{0}$, $\bar{\phi} = \beta_1$, $\bar{\rho} = \alpha_1$
2. For $i = 1, 2, \ldots$ repeat steps 3 through 6
3. Continue the bidiagonalization.
   (a) $\beta_{i+1}\mathbf{u}_{i+1} = \mathbf{A}\mathbf{v}_i - \alpha_i\mathbf{u}_i$
   (b) $\alpha_{i+1}\mathbf{v}_{i+1} = \mathbf{A}^{\mathrm{H}}\mathbf{u}_{i+1} - \beta_{i+1}\mathbf{v}_i$
4. Construct and apply orthogonal transformation.
   (a) $\rho_i = (\bar{\rho}_i^2 + \beta_{i+1}^2)^{1/2}$
   (b) $c_i = \bar{\rho}_i/\rho_i$
   (c) $s_i = \beta_{i+1}/\rho_i$
   (d) $\theta_{i+1} = s_i\alpha_{i+1}$
   (e) $\bar{\rho}_{i+1} = -c_i\alpha_{i+1}$
   (f) $\phi_i = c_i\bar{\phi}_i$
   (g) $\bar{\phi}_{i+1} = s_i\bar{\phi}_i$
5. Update $\mathbf{b}$, $\mathbf{w}$.
   (a) $\mathbf{b}_i = \mathbf{b}_{i-1} + (\phi_i/\rho_i)\mathbf{w}_i)$
   (b) $\mathbf{w}_{i+1} = \mathbf{v}_{i+1} - (\theta_{i+1}/\rho_i)\mathbf{w}_i$
6. Test for convergence. Exit if a stopping criterion has been met.

performs better in practice [13, 101] due to its unique restructuring of the input system (via the Lanczos process [86] and Golub-Kahan bidiagonalization [50]) prior to solving it. Empirical studies have shown that LSQR finds solutions with lower residual error than CGLS when $\mathbf{A}$ is ill-conditioned, and of similar fidelity when $\mathbf{A}$ is well-conditioned [13, 101]. Further description of how LSQR applies the Lanczos process and Golub-Kahan factorization, along with pseudocode when $\lambda_{\mathrm{LSQR}} \neq 0$, is located in [100, 101].

In addition to the above, LSQR's stopping rules are carefully designed to reflect the data's accuracy. Relative to CGLS's stopping rule, LSQR's ensures it always shuts down sooner and its corresponding b estimate is equally acceptable. This advantage becomes more pronounced as $\mathbf{A}$'s conditioning worsens [101].[1] Note that while LSQR indeed requires

[1]This was also confirmed directly by M. A. Saunders, one of the creators of LSQR.

$N_s + 2PN_t$ more complex multiplications per iteration than CGLS, this is mitigated by the fact that LSQR often requires fewer iterations to attain a similar-fidelity solution.

4.4 Quality Metrics

4.4.1 Image Quality Evaluation

For each experiment we conduct, we generate a 2-D image, $o(x, y)$, either via a Bloch simulation or by performing an excitation on the 8-channel system. We then evaluate the quality of $o(x, y)$ by using metrics that quantify how closely it matches the target pattern $d(x, y)$, each of which is explained below.

Mean-squared error (MSE). This measures how close $o(x, y)$ is to $d(x, y)$ over a chosen region of interest (ROI):

$$\mathrm{MSE}(o, d) = \frac{1}{\mathrm{card}(Z)} \sum_{(x,y)\in Z} |o(x, y) - d(x, y)|^2 \tag{4.6}$$

where Z is a set of coordinates that implicitly defines the spatial ROI over which the MSE is computed, and $\mathrm{card}(Z)$ is the *cardinality*, or number of elements, of Z.

Second-order statistics. Computing the mean μ and standard deviation σ of $o(x, y)$ in different ROIs quantifies the severity of artifacts and noise present within each, e.g., if the target is uniform in a particular ROI, a small σ implies that $o(x, y)$ closely matches $d(x, y)$ in that region.

Peak value. The maximum value in an ROI of $o(x, y)$ quantifies the worst-case artifact present, e.g., given two observations of the same target, larger peak values in one indicate it has more artifacts than the other.

Note on non-MSE metrics. Recent RF pulse design work uses MSE to evaluate the quality of an excitation, e.g., [59, 146]. As an extension of this methodology, we make use of non-MSE metrics in addition to MSE, because the latter is not always an ideal indicator of excitation quality. E.g., Wang et al. provide an example of six images with identical MSE, but three contain significant spike-like noise [138]. Using region-by-region peak value and second-order statistics analyses on these images causes the noise-ridden ones to exhibit worse scores, whereas MSE incorrectly indicates all images are of equal quality.

4.4.2 RF Waveform Characterization

Consider a set of waveforms represented by the PN_t-element vector, $\mathbf{b}$. The peak voltage, $V_{\max}$, equals $\|\mathbf{b}\|_\infty$, and the aggregated RMS voltage is defined as

$$V_{\text{RMS}} = \left(\frac{1}{PN_t} \sum_{n=1}^{PN_t} |\mathbf{b}[n]|^2 \right), \tag{4.7}$$

where $\mathbf{b}[n]$ is the nth element of $\mathbf{b}$. For each set of waveforms, knowing $V_{\max}$ allows us to compare pulses' relative peak powers. And since V_{RMS} is proportional to both integrated and average pulse power, comparing V_{RMS} values of different pulses tells how much power they dissipate relative to one another. Note that since $\mathbf{b}$ contains waveforms across all excitation channels, V_{peak} is the peak among all P waveforms.

4.5 Experiment Setup

System configuration. The parallel system is built around a Siemens 3T Tim Trio scanner (Siemens Medical Solutions, Erlangen, Germany). The transmit array is composed of eight circular, overlapped, 15-cm diameter, detunable surface coils arranged on a 28-cm diameter acrylic tube [7]. All scans are performed in a 17-cm low-dielectric oil phantom. For each RF design, the array's eight channels are driven, modulated in magnitude and phase by the pulses. Readouts are performed using a GRE sequence [12, 89, 98] with a repetition time (TR) of 30 ms, an echo time (TE) of 6 ms, and a bandwidth (BW) of 400 Hz/pixel.

Spatial profiles (B_1^+ maps). Spatial profiles $S_1(\mathbf{r}), \ldots, S_P(\mathbf{r})$ within the oil phantom are obtained using the approach outlined in Sec. 2.1.13. Specifically, low-flip-angle pulse is sent through each of the eight coil array elements, one at a time, and reception occurs on the system's body coil. B_1^+ maps are generated by recording a complex-valued image via a gradient-recalled echo (GRE) sequence with TR, TE, and BW equal to 20 ms, 6 ms, and 400 Hz/pixel, respectively, yielding 51 x 51 pixel, 4-mm resolution maps that capture the magnitude and relative phase of each array element. The magnitudes of these profiles are given in Fig. 2-6. Spatial variations in the body coil's reception profile are not removed because the profile is fairly uniform, exhibiting less than 5% variation. Before using the maps to generate the $\mathbf{A}$ matrix, each is scaled by a constant so the largest magnitude across all map pixels equals unity. This scaling makes the maps "qualitative" in the low flip angle

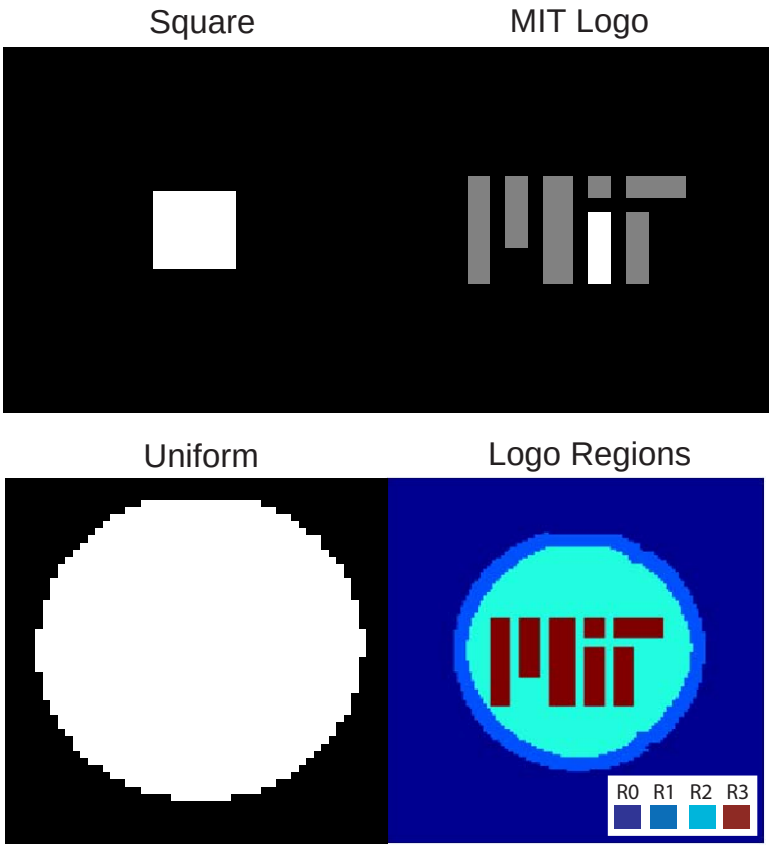

Figure 4-1: **Target excitation patterns and MIT logo regions**. The square, MIT logo, and uniformly-flat targets are shown. The MIT logo has two nonzero intensity levels-the lower part of the letter "i" is twice as intense as the others. The region-by-region breakdown of the logo is also shown. Region 0 is where statistical noise occurs; Region 1 is the ring-like edge region where the most glaring artifacts typically occur; Region 2 is the suppression region where the coil profiles interact and attempt to cancel out, but do not do so perfectly in practice; Region 3 is the letters of the target.

domain; they do not convey the exact flip angle achieved.

Target images. The targets are a square, a Massachusetts Institute of Technology (MIT) logo, and a uniformly-flat excitation. Fig. 4-1 depicts each 51 x 51 pixel, 4-mm resolution target. The logo has two nonzero intensity levels: the lower part of the "i" is twice the intensity of other letters. The intensities of the square, the uniform target, and the lower part of the "i" equal 0.01; this value is arbitrarily chosen since the B_1^+ maps are qualitative. This means that the V_{peak} and V_{RMS} of each $\mathbf{b}$ vector designed using these maps and targets are not actually in units of volts, but since 0.01 is used consistently through all experiments, it is possible to make relative comparisons between voltage statistics.

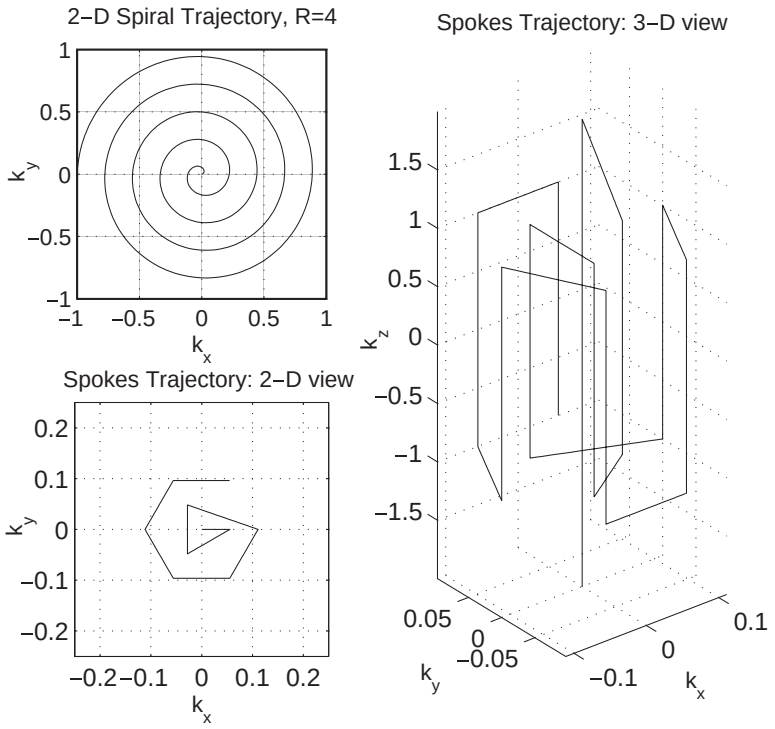

Figure 4-2: **An $R = 4$ spiral and 10-spoke slice-selective trajectory**.

Spiral trajectories. The 2-D k-space spirals are configured to have 4-mm resolution and an 18-cm field-of-view (FOV). Gradient amplitude and slew rate are 35 mT/m and 150 T/m/s, respectively. For $R = 1$, 4, 6, and 8, spirals are created that are 9.47, 2.42, 1.64, and 1.26 ms long, undergo 16, 4, 3, and 2 revolutions, and lead to $\mathbf{b}$ vectors of length 15152, 3872, 2642, and 2016, respectively. The $R = 4$ spiral is depicted in Fig. 4-2.

Spokes trajectory. The slice-selective trajectory consists of 10 spokes in k_z placed in the (k_x, k_y) plane to yield an 18-cm FOV, and is shown in Fig. 4-2. Slice thickness is 1 cm and the center spoke's time-bandwidth product is 4. Gradient amplitude and slew rate are 30 mT/m and 120 T/m/s, yielding a 5.84 ms pulse. To simplify the design process, we restrict the shape of each waveform to a Hanning-windowed sinc in k_z, which fixes the slice-selectivity of the trajectory and means that each design method only needs to calculate an amplitude and phase for each excitation channel to encode along each spoke. With 10 spokes and 8 coil elements, this means $\mathbf{A}$ has 80 columns and $\mathbf{b}$ has 80 elements.

Experiment summary. Table 4.1 summarizes the four experiments to be conducted.

Label	Target Pattern	Trajectory Type	Methodology
E1	Square	Spirals ($R = 1, 4, 6, 8$)	Simulations
E2	MIT Logo	Spirals ($R = 1, 4, 6, 8$)	Simulations + System Runs
E3	Uniformly-Flat	Spokes	Simulations
E4	MIT Logo	Spiral ($R = 8$)	Simulations + System Voltages

Table 4.1: **Experiment list**.

Experiments 1, 2, and 3 (Bloch simulations). In E1 and E2, the targets are the square and text logo, respectively, and the trajectories are $R = 1$, 4, 6, and 8 spirals. In E3, the in-plane target is the uniform pattern and the spokes trajectory is used. For every fixed target and trajectory, $\mathbf{d}$ and $\mathbf{A}$ are known, and $\mathbf{b}$ vectors are then calculated by solving $\mathbf{d} = \mathbf{A}\mathbf{b}$ using one of the algorithms. Once $\mathbf{b}$ is determined, pulses are Bloch-simulated as in [113].

For E1, E2, and E3, we first apply the SVD method, retaining enough SVs so the Bloch simulation of the resulting pulse yields an acceptable-looking excitation. Noting the MSE between this excitation and the target, we run LSQR and CGLS, tuning their parameters such that their Bloch-simulated pulses yield equal or lower MSE excitations. We attempt to make CGLS's MSE close to LSQR's, but this is difficult because CGLS is highly sensitive to λ_{CGLS} (see Discussion below). Voltage analyses are performed on each pulse once it is calculated.

Experiment 2 (8-channel system at 3T). For E2, each waveform designed during the simulation stage is played through the 8-channel excitation system by first scaling each "qualitative" $\mathbf{b}$ vector by a constant so its elements represent actual voltages. This scaling depends on both R and the design algorithm, and is chosen such that the flip angle of the MIT logo excited on the system is approximately constant across all experiments. Each excitation conducted on the system is then stored as a magnitude image, and MSE, voltage, and region-by-region second-order statistics and peak value analyses are conducted. The latter two metrics are calculated over the regions depicted in Fig. 4-1. Region 0 (R0) is where system noise is present, Region 1 (R1) the edge region where glaring artifacts tend to occur, Region 2 (R2) the suppression region where the profiles are interacting to cancel each other out, and Region 3 (R3) the letters.

Experiment 4. Here the trajectory is the $R = 8$ spiral and the target is the MIT logo. For each design algorithm, we loop over many choices of its control parameter, generating

	SVD			LSQR			CGLS		
	Kept SVs	Total SVs	RT	λ_{LSQR}	Iter.	RT	λ_{CGLS}	Iter.	RT
$R = 1$	976	1466	476	0.25	500	57	0.0125	1000	40
$R = 4$	861	1466	227	0.08	250	29	0.006	1000	19
$R = 6$	623	1466	187	0.09	500	16	0.0075	750	12
$R = 8$	391	1466	159	0.025	500	27	0.0005	750	37

Table 4.2: **Experiment 1 (square target, spiral trajectories): algorithm design parameters and runtimes**. Input parameters to each method used to compute each of the **b** vectors are listed. RT is runtime (sec).

thousands of pulses (**b**s). Then for each designed pulse, we compute its Bloch-simulated excitation's MSE with respect to the target, along with V_{peak} and V_{RMS}. Since we know from E2 how to properly scale each method's b vectors in order to play them on the eight-channel system, we scale them here as well, obtaining the actual voltage characteristics of each pulse. This essentially generates MSE vs. voltage tradeoff curves for each method. Note that extremely low-MSE pulses that yield completely unrealistic voltage values (e.g., $V_{\text{peak}} > 1000$ V) are disregarded.

4.6 Results

Experiments 1, 2, and 3 (Bloch simulations). For E1, the 12 resulting excitations are shown in Fig. 4-3. The rows and columns correspond to the $R = 1$, 4, 6 and 8 spiral trajectories and the three design algorithms, respectively. Each subplot depicts the Bloch-simulated excitation, the MSE between this simulated excitation and target, and (V_{max}, V_{RMS}). Fig. 4-4 shows the results of E2, and is formatted analogously to Fig. 4-3. For E3, Bloch-simulated images, MSEs, V_{max}, and V_{RMS} appear in Fig. 4-5. For E1, E2, and E3, each method's design parameters and runtime are listed in Table 4.2, Table 4.3, and Table 4.4, respectively. In the interest of space, second-order statistics and peak values are not shown for these experiments.

Experiment 2 (8-channel system at 3T). Fig. 4-6 shows the system images when the scaled **b** vectors are played through the 8-channel system. The region-by-region means, standard deviations, and peak values of each are shown in Fig. 4-7, where bar graphs of each statistic are shown for R0 through R3. The axes of each such graph are R and algorithm type.

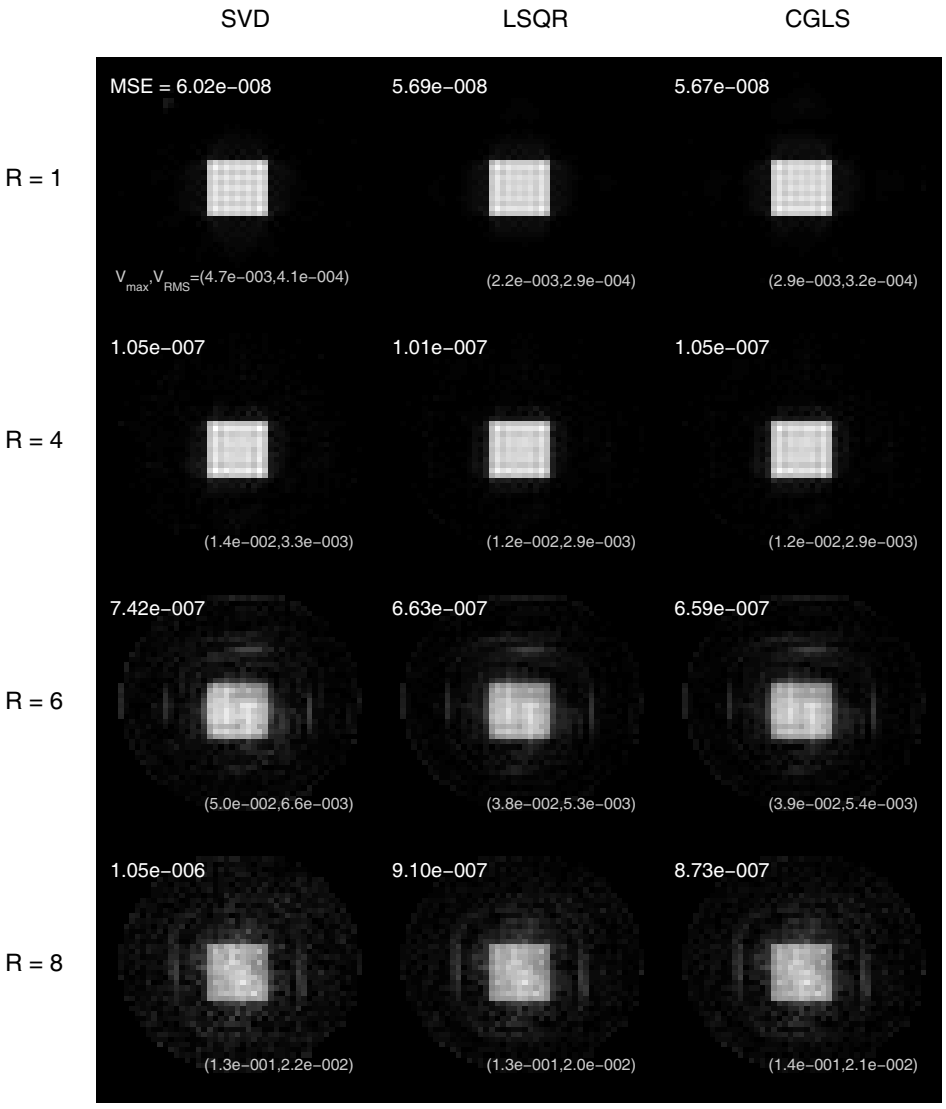

Figure 4-3: **Experiment 1's bloch-simulated excitations with MSE, $V_{\mathbf{max}}$, and $V_{\mathbf{RMS}}$ overlays**. The rows from top to bottom correspond to $R = 1$, 4, 6, and 8. From left to right, the columns correspond to RF waveforms designed using the SVD method, LSQR, and CGLS. The MSE between each excitation and the target is shown, along with $V_{\max}$ and V_{RMS} of each designed waveform.

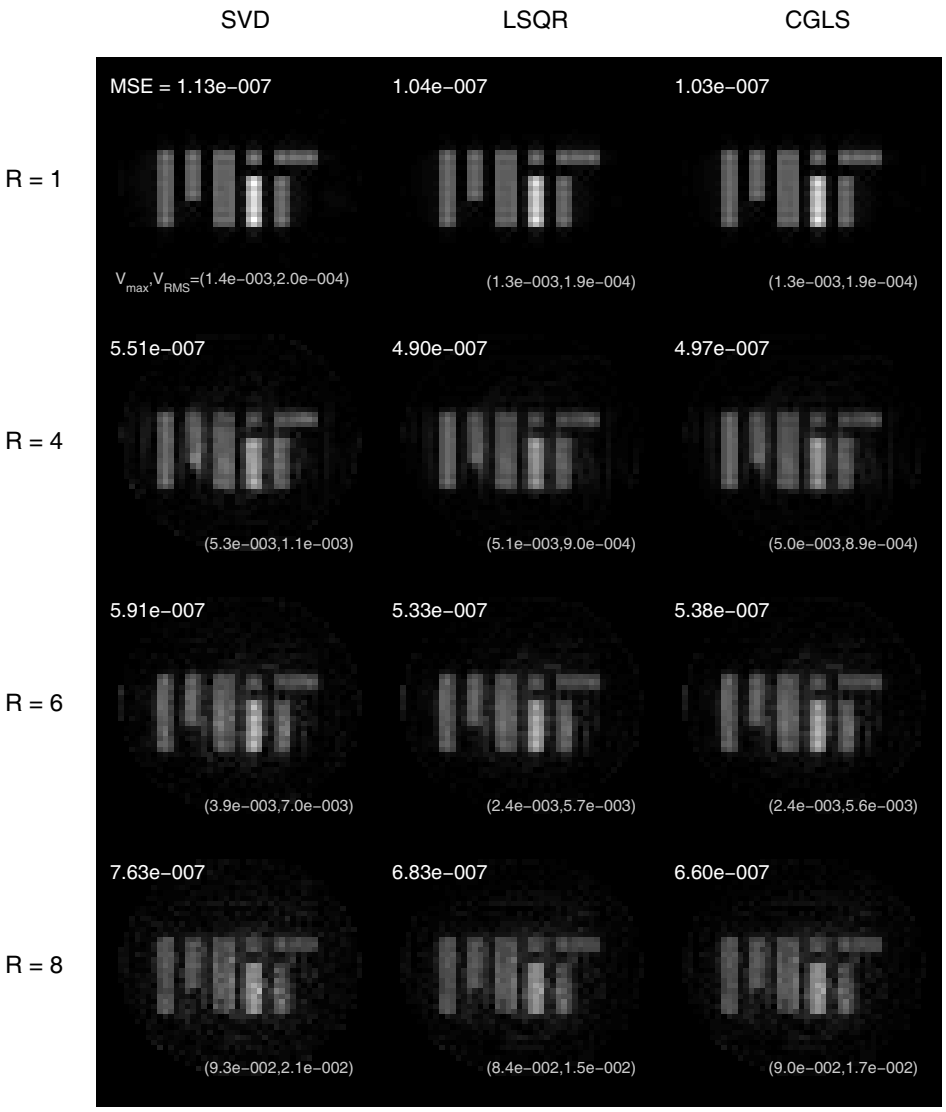

Figure 4-4: **Experiment 2's bloch-simulated excitation images with MSE, $V_{\mathbf{max}}$, and $V_{\mathbf{RMS}}$ overlays**. Formatting here is analogous to that of Fig. 4-3.

	SVD			LSQR			CGLS		
	Kept SVs	Total SVs	RT	λ_{LSQR}	Iter.	RT	λ_{CGLS}	Iter.	RT
$R = 1$	835	1466	663	0.75	3000	26	0.5	3000	10
$R = 4$	674	1466	220	0.46	250	7.1	0.22	250	5.3
$R = 6$	676	1466	187	0.06	500	22	0.0038	550	18
$R = 8$	767	1466	158	0.025	500	27	0.0005	1000	35

Table 4.3: **Experiment 2 (MIT logo target, spiral trajectories): algorithm design parameters and runtimes**. Formatting here is analogous to that of Table 4.1.

SVD			LSQR			CGLS		
Kept SVs	Total SVs	RT	λ_{LSQR}	Iter.	RT	λ_{CGLS}	Iter.	RT
39	80	0.8	3.0	200	0.15	0.01	100	0.1

Table 4.4: **Experiment 3 (uniformly-flat target, spokes trajectory): algorithm design parameters and runtimes**. Formatting here is analogous to that of Table 4.1.

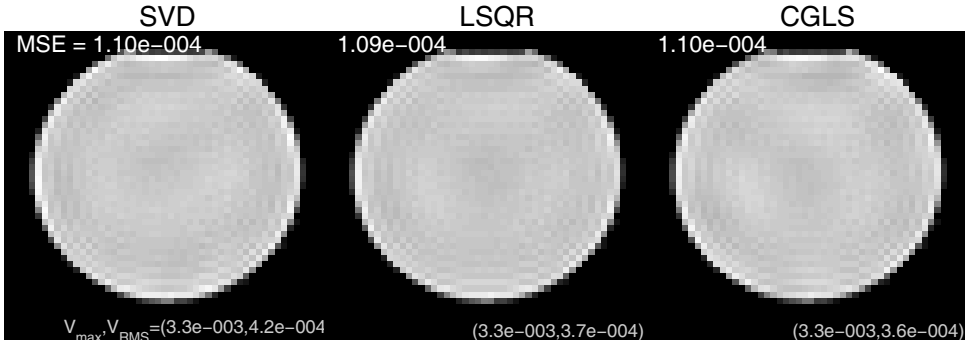

Figure 4-5: **Experiment 3's Bloch-simulated excitation images with MSE, $V_{\mathbf{max}}$, and $V_{\mathbf{RMS}}$ overlays**. Each image shows the in-slice excitation achieved by the RF waveform designed with each algorithm. From left to right, the images are due to the SVD method, LSQR, and CGLS.

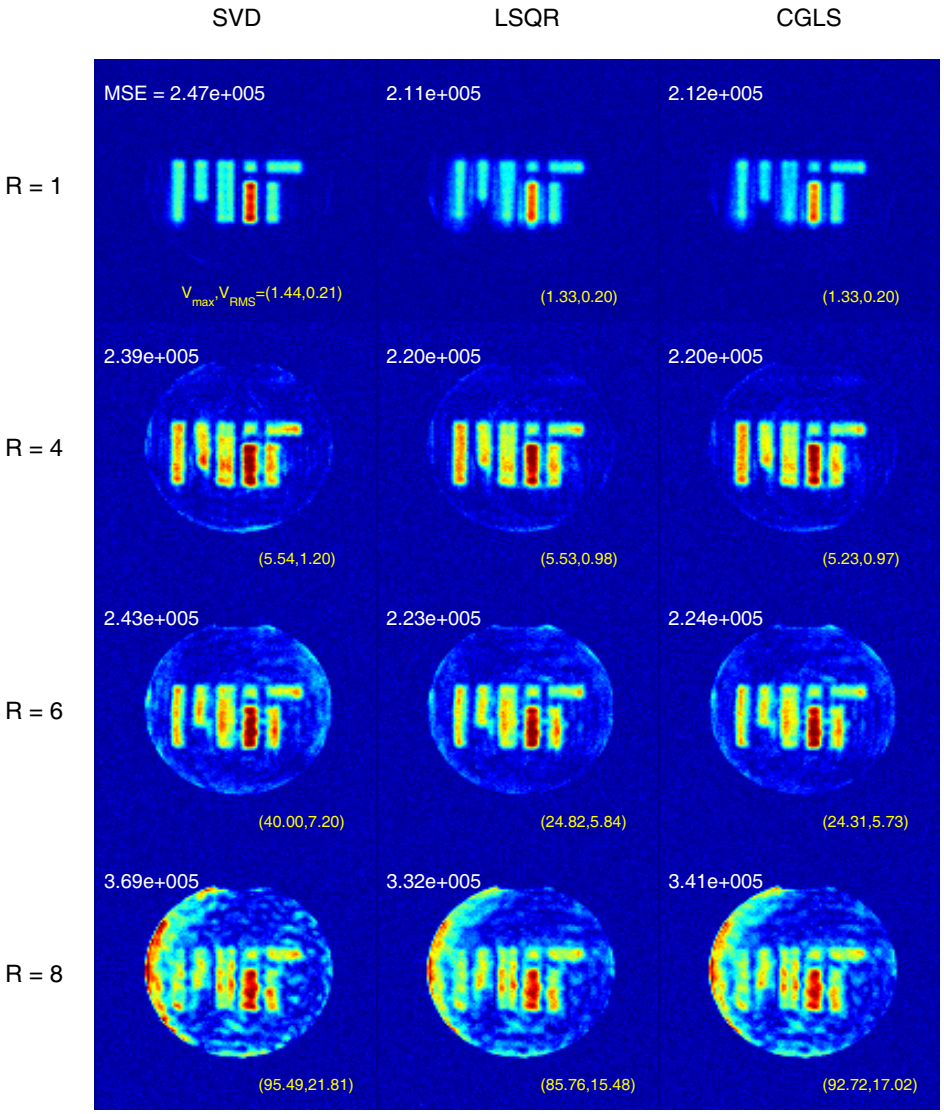

Figure 4-6: **Experiment 2's excitations conducted on the 8-channel system with MSE, $V_{\mathbf{max}}$, and $V_{\mathbf{RMS}}$ overlays**. Formatting here is analogous to that of Fig. 4-3.

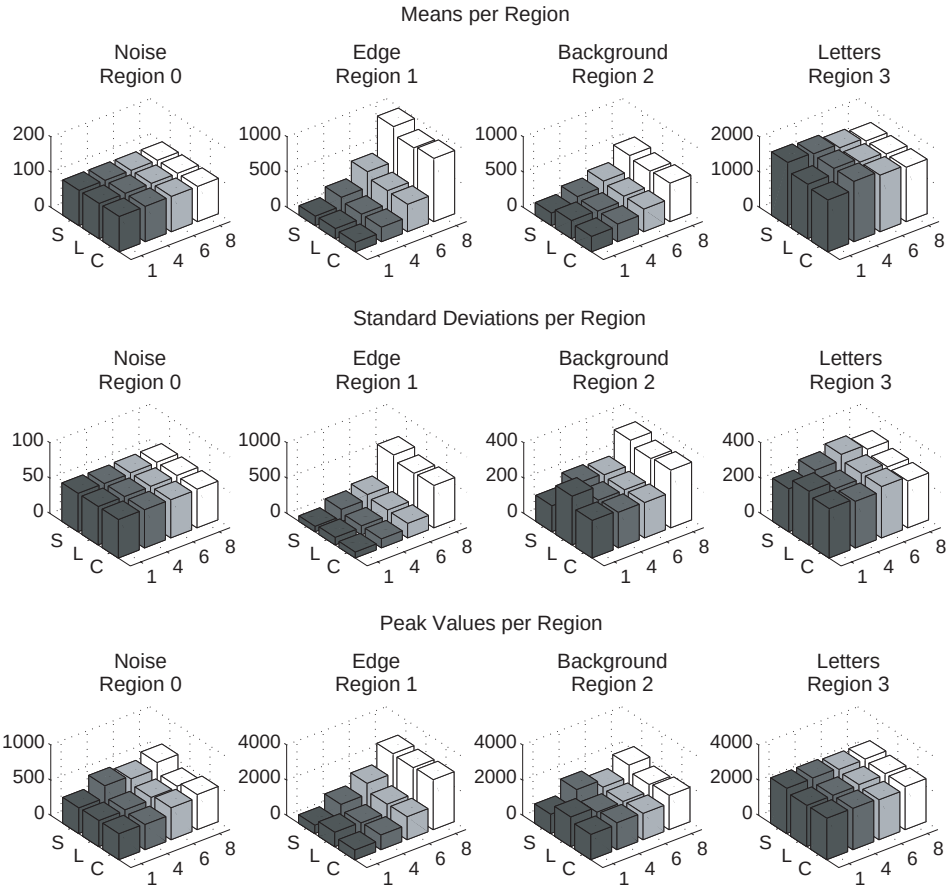

Figure 4-7: **Experiment 2's excitations conducted on the 8-channel system: means, standard deviations, and peak values per region**. For Regions 0 through 3, a bar graph of each statistic is shown. Each graph's x, y, and z axes denote the acceleration factor, algorithm type, and statistic value, respectively. "S", "L" and "C" stand for SVD method, LSQR, and CGLS.

Experiment 4. The upper, middle, and lower plots of Fig. 4-8 illustrate E4's results. The upper plot shows the iteration-by-iteration MSE performance of LSQR and CGLS when λ_{CGLS} and λ_{LSQR} equal zero. The middle plot shows the MSE vs. V_{peak} tradeoff of a variety of SVD pulses (by retaining different numbers of nonzero SVs) and the tradeoff curve for a large number of LSQR-designed pulses (generated by varying λ_{LSQR} over a wide range). The lower plot is analogous to the middle one, showing the MSE vs. V_{RMS} tradeoffs of the SVD and LSQR methods. CGLS data is not displayed in the middle and lower plots because, at the displayed scale, it is nearly identical to the LSQR data.

4.7 Discussion

Mean-square errors (MSEs). Referring to Fig. 4-3, Fig. 4-4, Fig. 4-5, Fig. 4-6, one sees that each LSQR and CGLS excitation has an equal or lower MSE than the corresponding SVD one, proving that we calibrated the designs properly. Further, as R increases, the MSEs of the SVD designs grow faster than those of LSQR and CGLS.

Experiments 1, 2, and 3 (Bloch simulations). For E1, Fig. 4-3 makes evident that excitations due to LSQR and CGLS are better than those due to SVD, i.e., for fixed R, LSQR and CGLS always result in lower MSE, V_{peak}, and V_{RMS}. Also, regardless of design technique, we see that artifacts always increase rapidly with R. In particular, when transitioning from $R = 6$ to 8, MSE increases by factors of 1.42, 1.37, and 1.32 for the SVD, LSQR, and CGLS designs, respectively. Analyzing Fig. 4-4, one sees E2 exhibits the same trends, e.g., for $R = 6$, the LSQR image has 1.11 times lower MSE and (1.63, 1.23) times lower voltages than the SVD method. For CGLS, (V_{peak}, V_{RMS}) are nearly identical to LSQR's, but CGLS's MSE is higher, so LSQR outperforms CGLS in excitation quality for the same amount of waveform energy.

In E3, $\mathbf{A}$ is 1466×80 in size and thus highly overdetermined, whereas in E1 and E2 it is highly underdetermined, which means E3 represents a substantially different design problem. Yet as Fig. 4-5 shows, LSQR and CGLS continue to outperform the SVD based method, e.g., the SVD and LSQR images have nearly the same MSE, but LSQR's V_{RMS} is 1.14 times lower than the SVD pulse's voltage. We do see that LSQR is not outperforming the SVD method as well as in E1 and E2, and thus conjecture that LSQR and CGLS provide a moderately better MSE vs. voltage tradeoffs than the SVD method when the

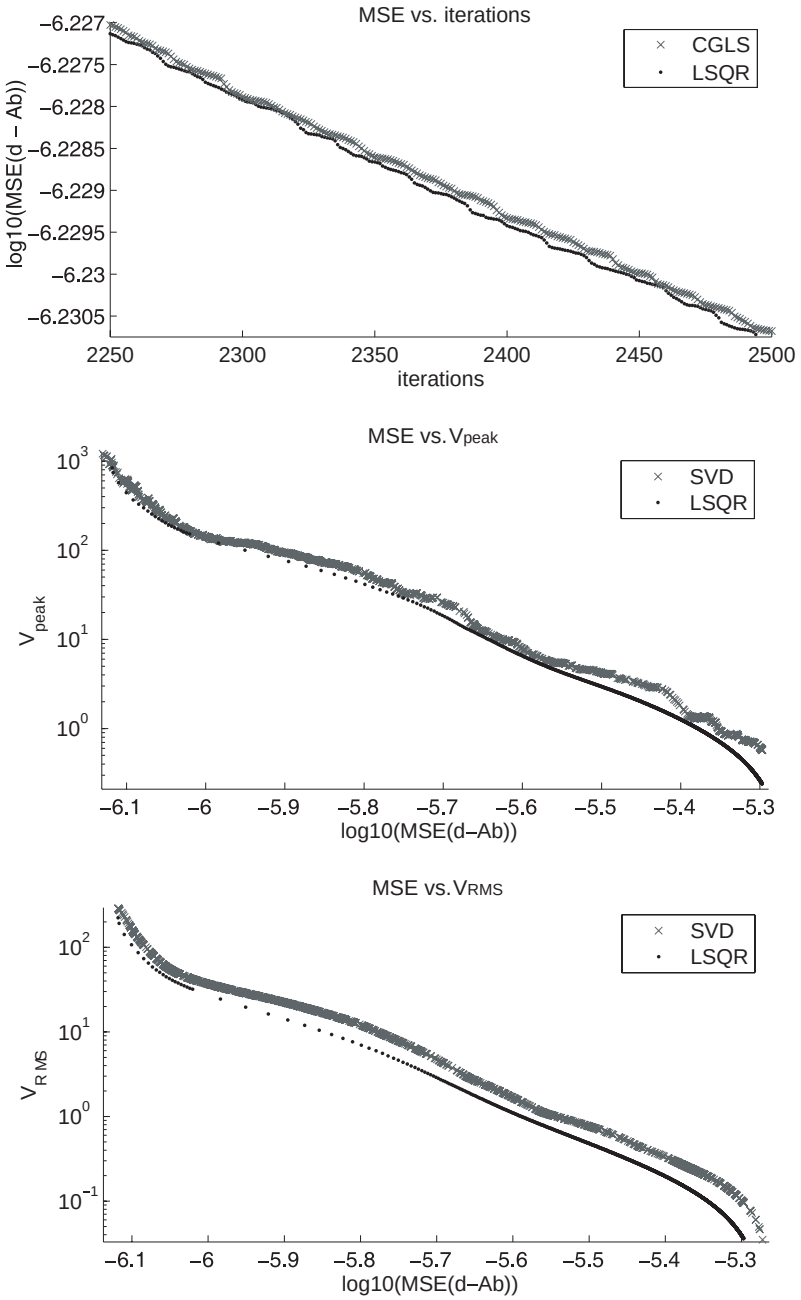

Figure 4-8: **Experiment 4's results**. The upper plot shows the iteration-by-iteration MSE performance of LSQR and CGLS when λ_{CGLS} and λ_{LSQR} equal zero. The middle plot shows the MSE vs. V_{peak} tradeoff of a variety of SVD pulses and the tradeoff curve for a large number of LSQR-designed pulses. The lower plot is analogous to the middle one, showing MSE vs. V_{RMS} tradeoffs of the SVD and LSQR methods.

system is small and overdetermined, and significantly better ones when it is large and underdetermined.

Experiment 2 (8-channel system at 3T). From Fig. 4-7, one sees that R0's μ and σ are equivalent across all images, confirming the noise floor is equivalent across all trials. For each R, the means of the images in R3 are nearly equal, proving each excitation achieved nearly the same flip angle. In R0 through R2, if no artifacts at all were present, the μ and σ values in Fig. 4-7 would equal those of the background noise, because the *ideal* target has zero intensity there. This means that smaller μ and σ in these regions imply fewer artifacts are present. In R1, the ring-like edge region that is particularly artifact-ridden, it is evident that for fixed R, LSQR and CGLS images exhibit superior second-order statistics than SVD images. E.g., in R1 for $R = 6$, (μ, σ) for the LSQR and CGLS images equal (206, 126) and (203, 127), respectively, whereas SVD's are significantly higher, equal to (260, 175). This same trend occurs in R2 and R3. The peak values also exhibit these trends, e.g., for $R = 4$ in R2, the SVD image has a peak of 1935 whereas the LSQR and CGLS images have much smaller peaks of 1393 and 1526.

Comparing the MIT logo simulation results in Fig. 4-4 with the eight-channel system results in Fig. 4-6, one sees that LSQR and CGLS's better MSE vs. voltage tradeoffs are present not just in the Bloch simulations, but in the system images as well. This proves that the advantage of LSQR and CGLS is due to the fundamental properties of these algorithms exhibited during the simulation stage, and is not due to system hardware effects such as transmission bandwidth.

Experiment 4. From the upper plot of Fig. 4-8, it is clear that for any of the given numbers of iterations, LSQR slightly outperforms CGLS in terms of MSE. From the middle and lower plots, it is evident that the LSQR method provides superior MSE vs. V_{peak} and MSE vs. V_{RMS} tradeoffs relative to the SVD method. Specifically, for *every* MSE evaluated, LSQR creates a pulse with lower peak and lower RMS voltage than the SVD method does. In many cases, the performance of LSQR relative to the SVD method is extreme, e.g., the SVD pulse that yields an image quality of log10(MSE) = -5.8 has $V_{\text{peak}} = 55.3$ V and $V_{\text{RMS}} = 12.3$ V, whereas the LSQR pulse achieves the exact same MSE with V_{peak} and V_{RMS} equal to merely 41.4 and 6.9 V. Note that these superior tradeoffs exhibited by LSQR are not limited to our chosen target and trajectory; they hold across many target-trajectory combinations, but due to space limitations we do not present these results. In conclusion,

E4 suggests that for the given target, trajectory, and fixed MSE, LSQR will always produce a pulse with better voltage characteristics than the SVD method.

Voltage characteristics. It is clear from E1 through E4 that LSQR and CGLS design pulses with better artifact vs. voltage tradeoffs than the SVD method. Because SAR is heavily influenced by V_{peak} and V_{RMS}, we speculate that the SAR values of LSQR and CGLS waveforms are significantly lower than those of SVD-based pulses.

One noticeable trend across all experiments is the rapid growth of V_{peak} and V_{RMS} with R. E.g., LSQR's V_{peak} jumps from 5.54 to 40.0 V when transitioning from $R = 4$ to 6, and jumps to 95.49 V when $R = 8$. These observations coincide with those of [81] and [131], extending the former's work from a strip-line coil to our circular array, and the latter's from a four-channel to an eight-channel system. This rapid voltage growth poses constraints on in vivo applications and implies the infeasibility of $R \gg 1$ designs due to algorithm development alone. However, since the maximum feasible R with moderate power requirements is strongly linked to the number of transmit elements and the design of the array [81], it may still be possible to achieve $R \gg 1$ pulses by designing arrays to handle higher voltages and have higher efficiencies.

Performance of LSQR and CGLS. LSQR and CGLS outperform the SVD method not only because of their numerical properties but because they directly penalize large $\mathbf{b}$ vectors, whereas the SVD algorithm does not do so. Across all experiments, LSQR and CGLS perform similarly in terms of each metric: this is because $\mathbf{A}$ is well-conditioned in all cases. Because of this, either algorithm may be used for waveform design instead of the SVD-based method. However, although they perform similarly in terms of MSE, voltage values and runtimes, LSQR should be used over CGLS, because of the various empirical studies showing its superior performance, especially in cases where A is ill-conditioned [13, 101].

Limitations. One limitation of this work is the qualitative nature of the spatial coil profiles. In future chapters, we will often use quantitative B_1^+ maps—those that describe the exact flip angle at each spatial location geenerated by each array element per input volt—to avoid this problem.

Another limitation is that during E2's eight-channel runs, several system images are saturated in a small sub-region of R3 because of the system imaging format's limited dynamic range relative to the acquisition parameters. To circumvent this issue, pixels within the more intense part of the letter "i" are discarded from R3's second-order statistic and peak

value computations, preventing error propagation. Note that in images where saturation was not an issue, the intensity level of the "i" was indeed twice that of the other letters.

4.8 Conclusion

Two iterative CG methods, LSQR and CGLS, have been shown to obtain better or equal quality excitations compared to an SVD-based truncated pseudoinversion method, while simultaneously producing pulses with significantly lower peak and RMS waveform voltages. This was shown to hold across a range of targets, k-space trajectories, and acceleration factors in both Bloch simulations and imaging experiments on an eight-channel system at 3T. Between LSQR and CGLS, the former had superior numerical properties.

To the best of our knowledge, this work has made the following contributions: (a) investigating different numerical methods of solving systems of parallel excitation equations that exhibit surprising performance differences, (b) validating these new methodologies across three types of targets and two different k-space trajectories, (c) introducing non-MSE metrics that avoid the possible pitfalls of judging excitation quality solely via MSE, (d) quantifying increases in V_{peak} and V_{RMS} voltage as a function of R for an 8-channel system with up to 8-fold trajectory accelerations, permitting exploration of the important issue of pulse energy, (e) validating our results by conducting actual accelerated parallel excitations on a realistic 8-channel system at 3T, and (f) showing for one target and trajectory, for all realistically-feasible system voltages, LSQR always generates a pulse with lower V_{peak} and V_{RMS} relative to the SVD method when excitation quality is fixed.

Chapter 5

Sparsity-Enforced Joint k-Space Trajectory and RF Excitation Pulse Design

5.1 Introduction

The purpose of this chapter is to improve the flexibility of the excitation process in MRI and generate optimized, target-excitation-specific sets of gradient waveforms and radio-frequency (RF) excitation pulses that outperform conventional non-optimized sets of gradient and excitation waveforms. We pose a design algorithm that yields short-duration, high-excitation-fidelity pulses and applies to both parallel transmission systems and conventional single-channel excitation systems. This algorithm is applicable to high-field MRI RF excitation pulse design where spatially-tailored excitation is useful for a variety of clinical imaging scenarios, but is not limited to high field systems. Our intent is to provide a means for the automatic generation of target-specific k-space trajectories (and corresponding sets of gradient waveforms and RF pulses). The contents of this chapter appear in two journal papers [164, 167], four conference abstracts [157, 159, 161, 162], and a pending patent [149].

Conventional non-joint trajectory and RF pulse design. In conventional pulse design, one decides on a general excitation format, e.g., a slice-selective excitation, and then simply chooses a standard set of gradients and an RF pulse to accomplish the task.

For example, a constant z gradient and sinc-shaped RF pulse may be chosen to cause the coil to excite a thin slice of tissue in the spatial z direction, as depicted in Fig. 2-3 and discussed in Sec. 2.1.7. A pulse designer puts thought into the task at hand: if a highly-structured 2-D in-plane pattern needs to be excited, the designer may choose a set of gradient waveforms that cause a 2-D echo-planar (EP) or spiral trajectory to be traversed in (k_x, k_y) in excitation k-space. If a parallel transmission system is available, a designer may "accelerate" the trajectory by undersampling it and using a shorter EP or spiral, as described in Sec. 2.1.11 and illustrated in Fig. 2-4. Finally, only after deciding upon the trajectory and gradients does the user proceed to design an RF pulse to accompany and deposit energy along the trajectory. There are many methods available to accomplish this task, e.g., [58, 59, 79, 108, 113, 146, 165, 168]. In each of these cases, pulse designers first decide on the trajectory and only then do they generate an RF waveform to accompany it; in other words, the RF pulse is automatically optimized based on the desired target excitation pattern, but the trajectory itself is not. It is worth noting an instance of work by Hargreaves et al. where k-space trajectories themselves undergo optimization by tuning gradient waveforms [61]. This approach, however, does not take the desired target excitation into account during the gradient optimization stage and RF pulses must thus be designed after the fact in the same fashion discussed above; the gradients are indeed "optimized", but not with respect to the desired excitation. Therefore this method, like those above, does not attempt to jointly solve the trajectory and pulse design problem.

Recent joint trajectory and RF pulse design work. Recently, researchers have begun to branch away from the conventional design approaches (e.g., [58,59,79,108,113,146, 165,168]) and have developed algorithms that attempt to jointly design both the trajectory and RF excitation pulse, optimizing both concurrently. These methods are similar in that they are provided a desired target excitation pattern and then (in some way) search over numerous trajectories (or types of trajectories), proceeding to find a "(trajectory, pulse)" pair that produces a version of the trajectory and resulting excitation that meet some predefined constraints, e.g., a particular trajectory duration (e.g., 5 milliseconds), a particular excitation fidelity (e.g., a normalized root-mean-square error of 20%), or both.

- **Levin et al.'s 2-D spiral trajectory optimization**. This approach focuses on optimizing 2-D spiral trajectories [88]. This method is limited in two ways: first, it is only able to optimize over a specific class of spiral trajectories that consists of concentric rings;

second, the algorithm and its underlying theory apply only to radially-symmetric 2-D desired excitation patterns; third, the algorithm is limited to optimizing over 2-D k-space and thus does not apply to routinely-desired 3-D excitations. The latter two limitations are hindering because they imply that the algorithm is incapable of being used to excite many commonly-desired spatially-tailored patterns, e.g., a 2-D square, a 3-D box, or a thin slice. There is one final limitation: from [88], it seems the method applies only to conventional systems and does not extend to parallel transmission coil arrays.

- **Yip et al.'s 2-D echo-planar trajectory optimization**. The second joint trajectory and pulse design method is due to Yip et al. and focuses solely on the optimization of 2-D echo-planar (EP) trajectories [148]. Unlike the method of Levin et al. [88], this method may be provided with a general 2-D excitation, i.e., radial symmetry of the desired excitation pattern is no longer a strict requirement. However, analogous to [88], this technique is limited in that it optimizes only 2-D EP trajectories. It is therefore sub-optimal for three reasons: first, it cannot be used to excite commonly-used 3-D patterns, such as thin slices; second, traversing EP trajectories at times have lengthy, impractical durations because EP traversals are limited by the MR system gradient's maximum amplitude and slew rates; finally, when 2-D EP trajectories and waveforms are applied in the presence of MR system non-idealities, they may cause the produced excitation to exhibit worse artifacts than an excitation produced by depositing energy along a 2-D spiral (i.e., spiral trajectories lead to less glaring imaging artifacts).

Moving beyond conventional and current joint design methods to *sparsity-enforced* joint trajectory and RF pulse design. In this chapter we overcome the limitations of prior approaches by proposing a general sparsity-enforcement algorithm that jointly determines sparse, quickly-traversable excitation k-space trajectories along with corresponding excitation pulses and gradient waveforms. The proposed method lets MRI system users specify a desired (3-D spatially-tailored and/or 3-D spatially-selective) excitation and then generates a customized set of gradients and RF excitation pulses explicitly optimized for the task at hand. The algorithm functions by applying an ℓ_1-norm penalty while searching over a large number of possible trajectory segments (and corresponding RF pulse segments); this process ultimately reveals a small, sparse subset of these segments (along with an overall RF pulse) that alone form a high-fidelity version of the user-specific target

excitation. The algorithm is general enough such that these segments may be of any shape or size, and thus applies to a wide variety of 2-D and 3-D k-space trajectories, so the algorithm may optimize any type of excitation pulse. Furthermore, the method provides users with intuitive control parameters that allow excitation quality to be traded off directly with pulse duration and is thus user-friendly.

Organization. In Sec. 5.2, we mathematically develop the sparsity-enforced joint trajectory-pulse design algorithm as generally as possible and show the relevance of MSSO simultaneous sparse approximation. In Sec. 5.3, we use it to design single-channel slice-selective pulses that mitigate B_1^+ inhomogeneity in the human brain at 7 Tesla. In Sec. 5.4, we study both sparsity-enforced spoke-based pulses from an empirical rather than *in vivo* standpoint, analyzing the performance of spoke-based sparsity-enforced pulses vs. conventionally designed pulses via simulations and an experiment within an oil phantom at 3T with an eight-channel parallel transmission system. In Sec. 5.5, we apply the joint design algorithm to a set of candidate rings in (k_x, k_y), optimizing spiral-like trajectories to produce a box-shaped excitation in the presence of single-channel B_1^+ inhomogeneity at 7T. Concluding remarks are given in Sec. 5.6.

5.2 General Sparsity-Enforced Joint Trajectory-Pulse Design

Here we will ultimately pose the algorithm as generally as possible, but to begin, let us consider only a single-transmit-channel system, as discussed in Sec. 2.1.9 and Sec. 2.1.10.

5.2.1 Derivation for Single-Channel Excitation Systems

Let $d(\mathbf{r})$ be a desired target excitation, where $\mathbf{r}$ is an index into the 3-D spatial domain, i.e., $\mathbf{r} = [x, y, z]^{\mathrm{T}}$. We now develop a method to jointly determine a target-specific, fast k-space trajectory and a corresponding RF pulse that achieve a high-fidelity version of this desired excitation. To begin, consider J contours in k-space; these may be spoke segments in k_z [108,113,132], rings in (k_x, k_y), or completely arbitrary curves tracing through (k_x, k_y, k_z); let the contours be denoted as $c_j(\mathbf{k})$ for $j = 1, \ldots, J$. Now assume that for $j = 1, \ldots, J$, we know N_j points in k-space along contour j, denoted $\mathbf{k}_{j,1}, \ldots, \mathbf{k}_{j,N_j}$. Thus, overall, $N_k = N_1 + \cdots + N_J$ discrete k-space points are known.

If we assume we are in the low-flip angle domain, the discussion of single-channel exci-

tation in Sec. 2.1.9 holds here, as does (2.9). Let us first eliminate the assumption of time from (2.9), i.e., we assume the traversal of a k-space contour, but do away with its time variable. Second, let us assume a homogeneous main field. This yields

$$m(\mathbf{r}) = j\gamma \mathrm{M}_0 S(\mathbf{r}) \int_{\mathbf{k}} g(\mathbf{k}) e^{j\mathbf{r}\cdot\mathbf{k}} d\mathbf{k}, \tag{5.1}$$

where $m(\mathbf{r})$, γ, M_0, and $S(\mathbf{r})$ are the same as in (2.9), and $g(\mathbf{k})$ is the complex-value energy weighting placed in k-space at a fixed point $\mathbf{k} = [k_x, k_y, k_z]^{\mathrm{T}}$. We now apply the linearized formalism discussed in Sec. 2.1.10 to (5.1), discretizing space at N_s locations $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$ within a chosen FOX and k-space at the N_k known k-space points along the J contours, i.e., the locations $\mathbf{k}_{j,1}, \ldots, \mathbf{k}_{j,N_j}$ for $j = 1, \ldots, J$. This lets us relate energy placed at each k-space contour to the resulting transverse magnetization that arises at the N_s points in space:

$$\begin{aligned} \mathbf{m} &= \mathbf{S}\mathbf{F}_1\mathbf{g}_1 + \cdots + \mathbf{S}\mathbf{F}_J\mathbf{g}_J \\ &= \mathbf{S}\left[\mathbf{F}_1 \cdots \mathbf{F}_J\right] \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_J \end{bmatrix} = \mathbf{S}\mathbf{F}_{\mathrm{tot}}\mathbf{g}_{\mathrm{tot}}, \end{aligned} \tag{5.2}$$

where $\mathbf{m}$ is an N_s-element vector of complex-valued samples of $m(\mathbf{r})$ at $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$, and $\mathbf{S}$ is an $N_s \times N_s$ complex-valued diagonal matrix of $S(\mathbf{r})$ within-FOX samples. Most importantly, $\mathbf{g}_j$ is an N_j-element vector of complex-valued weights the transmit channel places at the N_j k-space locations of contour j. Each $N_s \times N_j$ complex-valued matrix $\mathbf{F}_j$ relates how energy placed along contour j affects the resulting excitation. Formally,

$$\mathbf{F}_j(m, v) = j\gamma\Delta_t \mathrm{M}_0 \exp(j\mathbf{r}_m \cdot \mathbf{k}_{j,v}), \tag{5.3}$$

where Δ_t is the time-sample spacing of the RF waveform that will ultimately be designed. Each $\mathbf{F}_j$ in (5.3) is simply a specific instance of $\mathbf{F}$ in Sec. 2.1.10. To conclude the discretization stage, we sample the desired excitation $d(\mathbf{r})$ and stack the samples into $\mathbf{d}$.

The goal now is to find the *smallest subset of contours*, along with a corresponding pulse, such that when the pulse is played along the path defined by the small set of contours, a high-fidelity version of $d(\mathbf{r})$ is produced; the contour segments in the small subset reveal and

suggest a short, quickly-traversable k-space trajectory; the corresponding energy segments provide us with the RF pulse. Once this small contour subset is known, the individual contours within it may be connected using a gradient design algorithm (e.g., [61]), yielding a fully-determined trajectory and an accompanying RF pulse that have been obtained jointly and optimized with respect to the desired excitation pattern.

To find the truly optimal smallest contour subset, one must solve:

$$\min_{\mathbf{g}_1,\ldots,\mathbf{g}_J} \|\mathbf{d} - \mathbf{S}\mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}\|_2 \text{ s.t. the usage of only } K \text{ of the } \mathbf{g}_j\text{s}, \tag{5.4}$$

for the smallest K possible until the smallest-possible subset is found that achieves the excitation to within an ℓ_2 fidelity of ϵ with the proper choice of corresponding $\mathbf{g}_j$s. In other words, (5.4) essentially requires a search over all $2^J - 1$ nonempty subsets of the overall J-contour set; each subset of contours suggests a trajectory, from which a pulse may designed using one of the many conventional methods, e.g., [58,59,79,108,113,146,165,168]. Then the various (trajectory, pulse) pairs may be analyzed and the shortest (fastest) trajectory whose pulse generates a version of the target excitation that meets some desired fidelity constraint is the optimal solution to the problem. Unfortunately, this approach is computationally infeasible for even the smallest sets of contours, e.g., even for $J = 30$, one may need to search up to 1,073,741,823 possible trajectories. Formally, this problem is NP-Hard [36,96].

A brute-force search is clearly infeasible, but if we study (5.4) closely, we see that it strongly resembles the NP-Hard MSSO problem, (3.7), given in Sec. 3.2.2.[1] Note however that (5.4) is in fact more general than MSSO in (3.7)—each of the unknown vectors in the latter must be the same length, whereas here, the $\mathbf{g}_j$s may be quite different in size (each contains N_j elements). Despite the more general nature of the problem here, it still lends itself to the tractable algorithmic solutions posed in Ch. 3 because every technique given in Sec. 3.3 is already capable of solving this more general problem (proving this is beyond the scope of this dissertation). Given the strength of the convex relaxation techniques exhibited throughout the experiments of Sec. 3.4, we pose a relaxation here that is a generalization of (3.8):

$$\min_{\mathbf{g}_{\text{tot}}} \left\{ \tfrac{1}{2} \|\mathbf{d} - \mathbf{S}\mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}\|_2^2 + \lambda \sum_{j=1}^{J} \|\mathbf{g}_j\|_2 \right\}. \tag{5.5}$$

[1] The variable names in (5.4) differ slightly from those of (3.7) but the optimizations are indeed similar.

This optimization allows energy to be placed along any of the trajectory contour segments in k-space to form $d(\mathbf{r})$, but imposes a large penalty whenever one of the contours experiences a nonzero energy deposition (whenever $\|\mathbf{g}_j\|_2$ deviates from zero). This approach discourages the use of many contours while simultaneously encouraging those that remain to have $\mathbf{g}_j$s that (approximately) solve $\|\mathbf{d} - \mathbf{S}\mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot}}\|_2$. The ℓ_1-norm penalty on the $\|\mathbf{g}_j\|_2$ contour energies encourages sparsity of contour use. As λ is increased from 0 to 1, increasing numbers of contours have their energies driven to zero, residual error increases, and smaller contour subsets are revealed.

After solving (5.5) using a good choice of λ, an optimized Q-contour subset is formed, comprised of those $Q < J$ contours whose $\|\mathbf{g}_j\|_2$s are largest. The Q segments are then connected via a greedy method, producing a short, optimized k-space trajectory. Knowing the trajectory, we can generate a set of gradient waveforms, $\mathbf{G}(t)$, and subsequently determine the time-dependent k-space trajectory, $\mathbf{k}(t)$, discussed in Sec. 2.1.9. The time-dependent RF pulse, $g(t)$, has also been implicitly determined by (5.5) and is constructed directly from the complex-valued samples contained in the Q resulting $\mathbf{g}_j$ vectors.

Now that the time-dependent gradients and k-space trajectory are known, we may incorporate main field inhomogeneity as in (2.9) and retune (resolve for) the RF pulse by following the standard single-channel system pulse design steps given in Sec. 2.1.10 and solving (2.11).

5.2.2 Derivation for Multi-Channel Parallel Transmission Systems

Now let us assume a P-channel parallel transmission system is available. Assume that we have a sound estimate of each channel's transmission profile, i.e., we know $S_p(\mathbf{r})$ for $p = 1, \ldots, P$. This causes changes to (5.1, 5.2) because the resulting transverse magnetization, $m(\mathbf{r})$, is now a linear superposition of the excitations produced by each of the P channels, as discussed in Sec. 2.1.11. That is, $m(\mathbf{r}) = m_1(\mathbf{r}) + \cdots + m_P(\mathbf{r})$, or in discretized terms,

$$\mathbf{m} = \mathbf{m}_1 + \cdots + \mathbf{m}_P, \tag{5.6}$$

where $\mathbf{m}_p$ contains N_s samples of the excitation produced by the pth channel, $m_p(\mathbf{r})$, due to its deposition of energy along the J contours. In analogy to (5.2), the following holds for

channel p:

$$\begin{aligned}\mathbf{m}_p &= \mathbf{S}_p\mathbf{F}_1\mathbf{g}_{p,1} + \cdots + \mathbf{S}_p\mathbf{F}_J\mathbf{g}_{p,J} \\ &= \mathbf{S}_p\left[\mathbf{F}_1 \cdots \mathbf{F}_J\right]\begin{bmatrix}\mathbf{g}_{p,1}\\ \vdots \\ \mathbf{g}_{p,J}\end{bmatrix} = \mathbf{S}_p\mathbf{F}_{\text{tot}}\mathbf{g}_{\text{p,agg}},\end{aligned} \tag{5.7}$$

where $\mathbf{S}_p$ is an $N_s \times N_s$ diagonal matrix of samples of $S_p(\mathbf{r})$, $\mathbf{g}_{p,j}$ is an N_j element vector of energy weights placed by channel p along the N_j k-space locations of contour j, and $\mathbf{g}_{p,\text{agg}}$ is the aggregation of all J $\mathbf{g}_{p,j}$s. Notice that regardless of p, the $N_s \times N_k$ $\mathbf{F}_{\text{tot}}$ matrix is always present; this is because prior to the consideration of the transmission profile, there is always a linear Fourier relationship between energy deposited in k-space and the resulting magnetization. Now, substituting (5.7) into (5.6), we arrive at

$$\begin{aligned}\mathbf{m} &= \mathbf{S}_1\mathbf{F}_{\text{tot}}\mathbf{g}_{1,\text{agg}} + \cdots + \mathbf{S}_P\mathbf{F}_{\text{tot}}\mathbf{g}_{P,\text{agg}} \\ &= \left[\mathbf{S}_1\mathbf{F}_{\text{tot}} \cdots \mathbf{S}_P\mathbf{F}_{\text{tot}}\right]\begin{bmatrix}\mathbf{g}_{1,\text{agg}}\\ \vdots \\ \mathbf{g}_{P,\text{agg}}\end{bmatrix} = \mathbf{A}_{\text{tot}}\mathbf{g}_{\text{tot}},\end{aligned} \tag{5.8}$$

where $\mathbf{A}_{\text{tot}}$ is $N_s \times PN_k$ in size and $\mathbf{g}_{\text{tot}}$ is an PN_k-element vector representing the weights placed by all P channels at all N_k locations in k-space. Thus (5.8) expresses the final excitation achieved by transmission and deposition of energy by all channels at all contour segments. We now extend the MSSO-inspired single-channel sparsity enforcement algorithm of (5.5) to multi-channel systems as follows:

$$\min_{\mathbf{g}_{\text{tot}}} \left\{ \tfrac{1}{2}\left\|\mathbf{d} - \mathbf{A}_{\text{tot}}\mathbf{g}_{\text{tot}}\right\|_2^2 + \lambda \sum_{j=1}^{J} \left\| \left[\mathbf{g}_{1,j}^{\text{T}} \cdots \mathbf{g}_{P,j}^{\text{T}}\right]^{\text{T}} \right\|_2 \right\}, \tag{5.9}$$

where the regularization term, $\sum_j \left\| \left[\mathbf{g}_{1,j}^{\text{T}} \cdots \mathbf{g}_{P,j}^{\text{T}}\right]^{\text{T}} \right\|_2$, is the ℓ_1 norm of the overall ℓ_2 energies deposited by the system's P channels at each of the J contours. In other words, the sparsity-promoting ℓ_1-norm of *overall energies per contour* is being computed. We only promote sparsity along the "overall energy per contour" dimension because whether it is one channel or all channels that make large contributions to a contour does not matter if the use of

this contour greatly helps at forming the excitation. Fixing λ to a sufficiently large value and solving (5.9) will therefore yield a solution $\mathbf{g}_{\text{tot}}$ that is sparse in the contour dimension, revealing $Q < J$ trajectory contour segments that are useful in forming the excitation along with the RF energy that each channel should place upon each segment. From the resulting solution $\mathbf{g}_{\text{tot}}$, a high-energy $\|[\mathbf{g}_{1,j}^{\mathrm{T}} \cdots \mathbf{g}_{P,j}^{\mathrm{T}}]^{\mathrm{T}}\|_2$ term indicates that contour j is worth retaining and useful in forming the excitation, and the associated $\mathbf{g}_{p,j}$s indicate the RF pulse to transmit through the pth channel while traversing contour j. Similarly to (5.5), the optimization (5.9) may be implemented and solved numerically for practically-sized systems of equations. Finally, note that when $P = 1$, (5.9) reduces to the single-channel base case of (5.5).

Analogously to the single-channel case derived earlier, the time-dependent trajectory, $\mathbf{G}(t)$, and k-space trajectory, $\mathbf{k}(t)$, are readily obtained once (5.9) has been solved and Q contours selected, along with the RF pulses to play along each channel, $g_p(t)$, for $p = 1, \ldots, P$. If desired, these pulses may be retuned as in the single-channel case.

5.2.3 Algorithm Variants

It is possible to modify (5.5) and (5.9) such that an ℓ_∞ (maximum error) constraint is imposed on $\mathbf{d} - \mathbf{m}$ rather than an ℓ_2 penalty. Furthermore, if the conditioning of the $\mathbf{g}_{p,j}$s (and thus the overall RF pulse) is a concern, constraints may be placed on their ℓ_∞ and ℓ_2 energies. These variants and their combinations yield objective functions that are still convex.

5.3 Sparsity-Enforced Short-Duration Slice-Selective RF Excitation Pulses for Mitigating B_1^+ Inhomogeneity in the Human Brain at 7 Tesla

5.3.1 Overview

We now jointly optimize *spoke-based* trajectories [108, 113, 132] and RF excitation pulses to produce short-duration slice-selective excitations on a high-field single-channel system to mitigate severe B_1^+ inhomogeneity ($S(\mathbf{r})$ inhomogeneity) [16]. Mitigation pulses are designed and demonstrated at 7 Tesla in a head-shaped water phantom and the brain; in each case,

the pulses mitigate a significantly non-uniform transmit profile and produce nearly-uniform flip angles across the field of excitation.

At a conceptual level, this is accomplished by first fixing the overall shape and height of the spokes to be used, which reduces spoke-trajectory design to the problem of placing as few complex-valued weights as possible in (k_x, k_y) that still achieve a high-fidelity excitation. One then sets the J contour candidates discussed in Sec. 5.2.1 to be a grid of candidate points in 2-D (k_x, k_y) space. Providing the field profile of the in-plane B_1^+ inhomogeneity along with the desired excitation (in this case, a uniformly-flat flip angle across space), one may then solve the optimization (5.5) using one of the MSSO simultaneous sparse approximation techniques in Sec. 3.3 to reveal a small subset of good locations (and corresponding weights) at which to place spokes.

5.3.2 Motivation

The 3-D RF pulse designs proposed in [108,113,132] describe a class of slice-selective pulses capable of mitigating B_1^+ inhomogeneity that offer improvements over high-SAR adiabatic pulses [117] and image post-processing methods [25]. An overview of spoke-based pulses is given in Sec. 2.1.7; They consist of modulated sinc-like pulse segments ("spokes") in the k_z direction of excitation k-space positioned at locations in (k_x, k_y). Sinc-like RF depositions in k_z produce slice-selectivity in z and the amplitude and phase modulation of each spoke in (k_x, k_y) spatially tailors the excitation in (x, y) to mitigate the in-plane inhomogeneity. An ideal B_1^+ mitigation pulse excites the pointwise inverse of the inhomogeneity and yields a uniform magnetization; therefore in practice, spoke modulation terms are chosen such that they produce an in-plane excitation that closely resembles the ideal one. Recall that a "standard slice-selective" pulse was defined in Sec. 2.1.9 and Sec. 2.1.12 as a one-spoke pulse whose single spoke is placed at $(k_x = 0, k_y = 0)$.

In prior work, relatively few spokes have been used for inhomogeneity mitigation on single-channel [108] and multi-channel parallel transmission systems [113,132]. In all cases, work is performed at field strengths below 7T where B_1^+ inhomogeneity in the brain is less severe, resembling a quadratic function in space [108]. In contrast, B_1^+ inhomogeneity at 7T exhibits significant spatial variation and is not quadratic [133,136]. This means that spoke designs that utilize single-channel transmit systems and rely on quadratic assumptions about B_1^+ [108] are unlikely to mitigate brain inhomogeneity at 7T. Parallel excitation systems, on

the other hand, are indeed useful for B_1^+ mitigation at high field, but are expensive in terms of hardware and complexity: each transmission channel requires an RF power amplifier as well as a SAR monitor. Based on the above, it is evident that a method is needed to design fast, slice-selective, B_1^+ mitigation pulses for use on high-field single-channel systems.

Since B_1^+ is highly non-uniform at 7T [133,136], one approach to mitigating it would be to extend prior spoke-based designs by placing a large number of modulated spokes throughout (k_x, k_y)-space, covering both low and high spatial frequencies. Unfortunately, placing many spokes leads to impractically-long pulses. To avoid this problem, here we provide a B_1^+ map of the head to the sparsity-enforced joint trajectory-pulse algorithm to find a small number of spokes necessary to mitigate the inhomogeneity along with their placement in (k_x, k_y) and their proper modulations. The algorithm enforces sparsity on the number of spokes allowed while encouraging those that remain to be placed and modulated in a way that maximizes B_1^+ mitigation in the least-squares sense. In this work, we demonstrate the capabilities of sparsity-enforced pulse design by performing mitigation experiments at 7T in a head-shaped phantom and the human brain.

5.3.3 Methods

Transmit Profile, Receive Profile, and Flip Angle Map Estimation

In upcoming experiments, we estimate $R(\mathbf{r})$ and $|S(\mathbf{r})|$ (T/V) using a combination of standard one-spoke pulses, magnetization reset pulses [33], and a parametric fitting technique; readers are strongly encouraged to review the signal intensity equations of Sec. 2.1.12 and fitting overview in Sec. 2.1.13 before continuing further. At minimum, recall that one version of this technique allows us to compute the actual flip angle across space arising due to any low-flip pulse (to within a multiplicative constant), which we denote $\alpha(\mathbf{r})$. Finally, note that throughout this section $|B_1^+(\mathbf{r})|$ will be used synonymously with $|S(\mathbf{r})|$.

Sparsity-Enforced Spoke Placement and RF Pulse Design

Overview. Once an estimate of the transmit profile magnitude $|S(\mathbf{r})| = |B_1^+(\mathbf{r})|$ is known we may pursue our goal of exciting a thin uniform slice in the presence of the B_1^+ inhomogeneity. To achieve slice-selectivity, we will place sinc-like spokes in k_z. To ensure excitation uniformity, we will place and modulate the spokes such that the magnitude of the resulting

in-plane nominal excitation closely resembles the pointwise inverse of the transmit profile. Finally, to minimize pulse duration, we will use as few spokes as possible. Note that because we provide only the transmit profile *magnitude* to the algorithm, the pulse it designs will mitigate the magnitude of the inhomogeneity and not attempt to account for the profile's phase; this is acceptable from an imaging standpoint in most applications because the profile's phase is slowly varying.

Optimized spoke placement in 2-D k-space. Let us consider J candidate points in the (k_x, k_y) plane, $\mathbf{k}_1, \ldots, \mathbf{k}_J$, at which we may deposit energy; a nonzero deposition at a grid point commits us to playing a spoke at this location. Thus we must judiciously place energy on the grid to mitigate the inhomogeneity while using as few grid points as possible. This problem is clearly a special case of the NP-Hard general single-channel joint design problem of (5.4) in Sec. 5.2.1; further, because it involves the sparse approximation of a single complex-valued vector, this maps to the MSSO problem of Ch. 3 where the unknowns are two simultaneously sparse real-valued vectors. Therefore, we may solve the relaxed convex optimization given in (5.5) to find a tractable, approximate solution to the intractable NP-Hard spoke placement problem. With the proper choice of λ, solving (5.5) finds a sparse energy deposition across the grid of J candidate locations (the majority of energy depositions are zero or close to zero) that produces a relatively-flat magnetization: the few elements of the grid that are large in magnitude indicate good spoke locations, revealing a small set of points capable of producing the needed excitation. We now place T spokes at the $\mathbf{k}_j$s whose corresponding energy weightings are the largest in magnitude; T is thus another parameter trading off pulse duration with B_1^+ mitigation.

Gradient and RF pulse design. We now know T locations in (k_x, k_y) at which to place spokes, as well as the shape of the spoke to play along each. Fixing slice thickness and spoke shape ends up fixing the pulse shape. We are then able to produce a set of gradients and in turn determine the time-dependent k-space trajectory. All that remains is to calculate the complex weight to encode along each spoke. We may now apply the single-channel pulse design approach of Sec. 2.1.10 (and incorporate a $\Delta B_0(\mathbf{r})$ field map if desired), but because we only need to compute the in-plane weightings, (2.11) reduces to one where $\mathbf{SF}$ is $N_s \times T$ in size and $\mathbf{b}$ has T rather than N_t elements [113]; weights are computed simply via pseudoinversion of $\mathbf{SF}$. At this point, the gradients and pulse have been calculated.

Design parameters. For all T-spoke mitigation pulses presented in this section, λ equals 0.35, slice thickness is 20 mm, spokes are Hanning-windowed sincs, the DC spoke's time-bandwidth-product equals 4, the kz-lengths of the other spokes are half that of the DC spoke, and $\Delta B_0(\mathbf{r})$ is estimated from two images with a 1-ms echo time (TE) difference. The FOX in which N_s samples of $\mathbf{r}$ are taken is where the phantom or brain's inhomogeneity is nonzero. For all designs, the (k_x, k_y) grid is Nyquist-spaced corresponding to a 25.6-cm FOV and contains 289 candidate locations. Gradient amplitude and slew rate are constrained at 35 mT/m and 150 T/m/s. Given these parameters, the entire design process outlined in this section takes 3-5 minutes in MATLAB when (5.5) is implemented as a second-order-cone program as described in Sec. 3.3.7.

Choosing the number of spokes, T, is an essential part of the design process and is accomplished by solving the spoke placement problem once, designing a series of candidate pulses with increasing numbers of spokes, simulating the magnetization that arises due to each pulse, and recording the within-FOX standard deviation of each magnetization; T is then the smallest number of spokes needed to drive the standard deviation below some chosen value. This automated process takes several seconds.

Data Acquisition

Hardware. Experiments are conducted on a 7T Siemens scanner (Siemens Medical, Erlangen, Germany) with standard body gradients (40 mT/m maximum amplitude, 180 mT/m/s maximum slew rate). A quadrature bandpass birdcage coil [126,140] is used for transmission and reception.

Imaging parameters. When collecting intensity images to estimate $|B_1^+(\mathbf{r})|$ we use a standard slice-selective pulse followed by a 200-V 16-ms BIR4 adiabatic reset, collecting 128x128 GRE images with 25.6-cm FOV, 5-mm slice thickness, 2-mm in-plane resolution, 380 Hz/pixel bandwidth, 1-sec TR, and 8-ms TE. To obtain a low-flip reference image, L0(r), we apply a standard pulse without the reset and average 8 times; parameters are the same as above except TR = 100 ms and TE = 8 ms. Finally, when applying a mitigation pulse, we perform 3-D GRE readouts (without the reset) and collect 16 contiguous 5-mm slices, using the parameters above except here TR = 100 ms and TE = 8 ms. In-plane B_1^+ mitigation performance is judged by analyzing the magnitude of the center slice of the volume; slice-selectivity is judged by analyzing the through-plane intensity profile.

5.3.4 Experiments

Water phantom: $R|B_1^+|$ inhomogeneity mitigation. Here we design a pulse to mitigate the inhomogeneity presented by the combination of transmit and receive profiles in a uniform-T1 head-shaped water phantom. The motivation is as follows: because this pulse seeks to mitigate the combination of profiles rather than simply the transmit profile, it will produce a result that is easy to understand and evaluate, since it ideally will produce a uniform image. In contrast, a pulse that successfully mitigates only the transmit profile will produces an image that is still highly non-uniform, because in this latter case the non-uniform receive profile is not mitigated. (Note: a successful $R|B_1^+|$ mitigation pulse produces a non-uniform magnetization and is not practical for clinical scenarios; in clinical practice we want to mitigate only $|B_1^+|$.)

To begin, we collect a low-flip image $L_0(\mathbf{r})$ using a standard pulse; (2.19) implies that $L_0(\mathbf{r}) \propto R(\mathbf{r}) \cdot \alpha(\mathbf{r}) \propto R(\mathbf{r}) \cdot |B_1^+(\mathbf{r})|$. We then design a 23-spoke mitigation pulse by providing $L_0(\mathbf{r})$ as the transmit profile magnitude estimate and then running the algorithm. A stack of images is acquired using the pulse in conjunction with the 3-D readout. To quantify the degree to which the inhomogeneity is mitigated, we compare the standard deviation, σ, and worst-case variation (WV) of the original image $L_0(\mathbf{r})$ with those of the center slice of the post-mitigation readout volume and also observe five 1-D profiles. The WV of an image is the ratio of its brightest to its darkest pixel within the FOX. Unlike σ, WV is sensitive to the change of even a single pixel and thus reveals if the mitigation pulse causes the image to contain undesirable spikes or black holes.

Water phantom: $|B_1^+|$ inhomogeneity mitigation. We now transition to a practical scenario and design a pulse to mitigate solely the inhomogeneous transmit profile. We first estimate $|B_1^+(\mathbf{r})|$ by collecting 10 images with transmit voltages $V = 20, 60, 100, \ldots, 380$ V and performing the Powell fit [103]; collecting the images takes 17 minutes, while fitting takes under a minute. A low-flip image, $L_0(\mathbf{r})$, is also collected and $R(\mathbf{r})$ is then estimated. We then design a 19-spoke pulse by feeding $|B_1^+(\mathbf{r})|$ to the placement algorithm. The desired magnetization is a 10-degree uniform flip across the FOX. After playing the pulse with the 3-D readout, we extract the post-mitigation center slice, $L_m(\mathbf{r})$; taking $L_m(\mathbf{r})/R(\mathbf{r})$ (pointwise) yields $\alpha_m(\mathbf{r})$, the *post-mitigation* flip angle map. To calculate the performance of the pulse, five 1-D profiles of $|B_1^+(\mathbf{r})|$ and $\alpha_m(\mathbf{r})$ are considered along with the σ and WV

of each.

Human brain: $|B_1^+|$ inhomogeneity mitigation. Finally, we design a 19-spoke pulse to mitigate $|B_1^+|$ non-uniformity in an axial slice of the human brain at 7T. This experiment is conducted exactly like the water phantom $|B_1^+|$ mitigation trial. Experiments are conducted at the A. A. Martinos Center for Biomedical Imaging (Charlestown, MA) and obey all safety and Institutional Review Board (IRB) requirements.

5.3.5 Results and Discussion

Water phantom: $R|B_1^+|$ inhomogeneity mitigation. The original image, $L_0(\mathbf{r})$, along with the mitigated image due to a 200-V, 8.5-ms, 23-spoke pulse, $L_m(\mathbf{r})$, are presented in Panels A and B of Fig. 5-1, respectively. A through-plane profile of the mitigated readout volume (Panel C) proves the pulse achieves slice selection. Recall that here the in-plane goal of the pulse is to mitigate the combined transmit and receive profiles, so ideally $L_m(\mathbf{r})$ will be constant everywhere. From Panel B and the associated 1-D profiles, we see that the pulse has produced a more uniform image. Based on standard deviation, $L_m(\mathbf{r})$ (Panel B) is 2.6 times smoother than $L_0(\mathbf{r})$ (Panel A). Furthermore, worst-case variation has been reduced by a factor of 1.7. It seems that to some degree, the pulse mitigates the inhomogeneity presented by the combined profiles. Note that each image is scaled to display its entire dynamic range within the grayscale spectrum.

Fig. 5-2 shows pulse design details; sparsity-enforced spoke locations are shown in (k_x, k_y), along with the RF pulse magnitude, the gradients, and the 3-D k-space trajectory. The in-plane Bloch simulation of the RF closely resembles the inverse of $L_0(\mathbf{r})$ in Panel A of Fig. 5-1 as intended. (Note: because we are not mitigating solely the transmit profile, we do not know the flip angle per volt of this pulse, so the choice of a 200-V transmit voltage is completely arbitrary.)

Water phantom: $|B_1^+|$ inhomogeneity mitigation. Here a 19-spoke pulse attempts to produce a uniform magnetization. Panels A, B, and C of Fig. 5-3 depict the low-flip image $L_0(\mathbf{r})$, receive profile estimate $R(\mathbf{r})$, and transmit profile estimate $|B_1^+(\mathbf{r})|$; the latter is highly non-uniform with $\sigma = 0.16$ and WV $= 2.5$. The transmit and receive profiles are not equal; in fact, each seems to be the mirror image of the other (consider a reflection across the y axis). Note the smoothness and lack of noise in the transmit profile estimate (Panel C). This map is not smoothed; rather, it simply comes directly out of the fitting

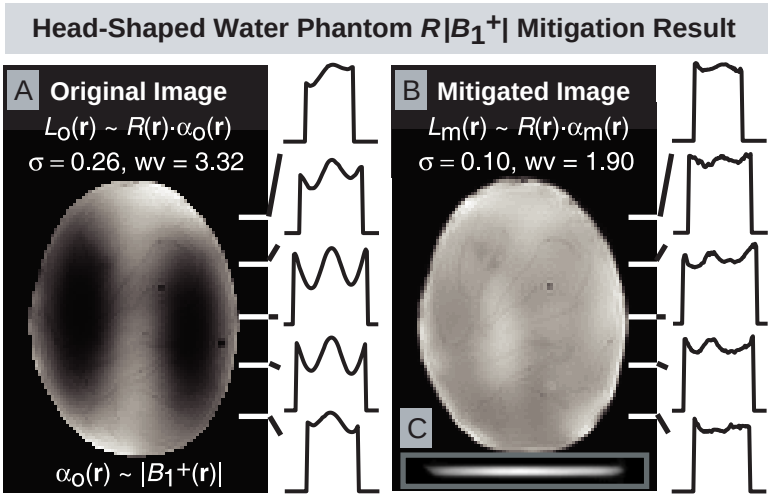

Figure 5-1: **Water phantom: $R|B_1^+|$ mitigation due to an 8.5-ms 23-spoke pulse**. Panel A: original image, $L_0(\mathbf{r})$, collected using standard pulse; highly non-uniform. Panel B: in-plane mitigated image, $L_m(\mathbf{r})$; standard deviation, σ, and worst-case variation (WV) reduced by 2.6x and 1.7x, respectively. Panel C: through-plane profile of mitigated readout volume showing successful slice selection. Given the 1-D profiles, σ, and WV metrics, and images themselves, it seems the pulse mitigates much of the inhomogeneity presented by $R(\mathbf{r})|B_1^+(\mathbf{r})|$.

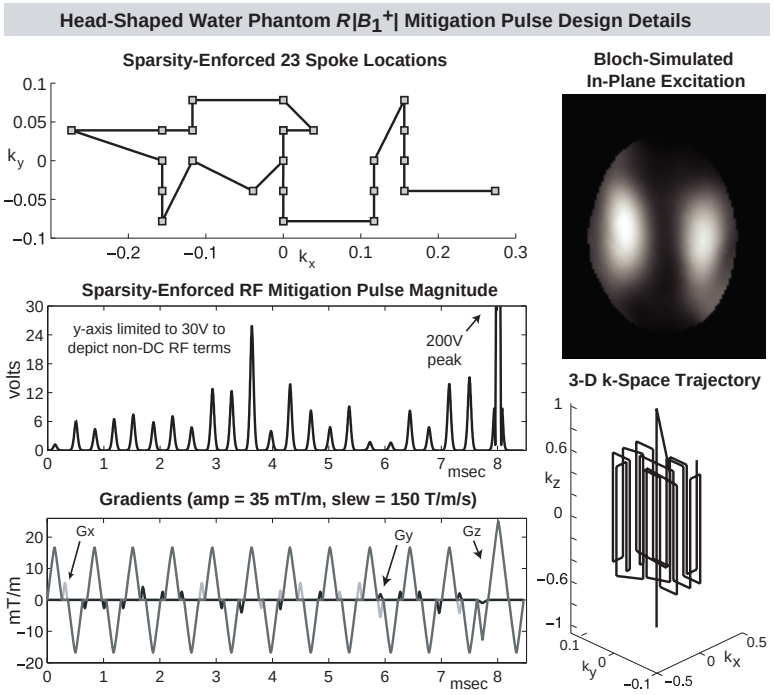

Figure 5-2: **Water phantom: 23-spoke $R|B_1^+|$ mitigation pulse design**. Upper-left: locations chosen by sparsity-enforced spoke placement method and their in-plane connections. Mid-left: 8.5-ms mitigation pulse magnitude (V). Bottom-left: gradients; G_z is dominant, smaller blips are G_x and G_y. Upper-right: in-plane excitation created by RF pulse (simulated). Lower-right: 3-D k-space trajectory.

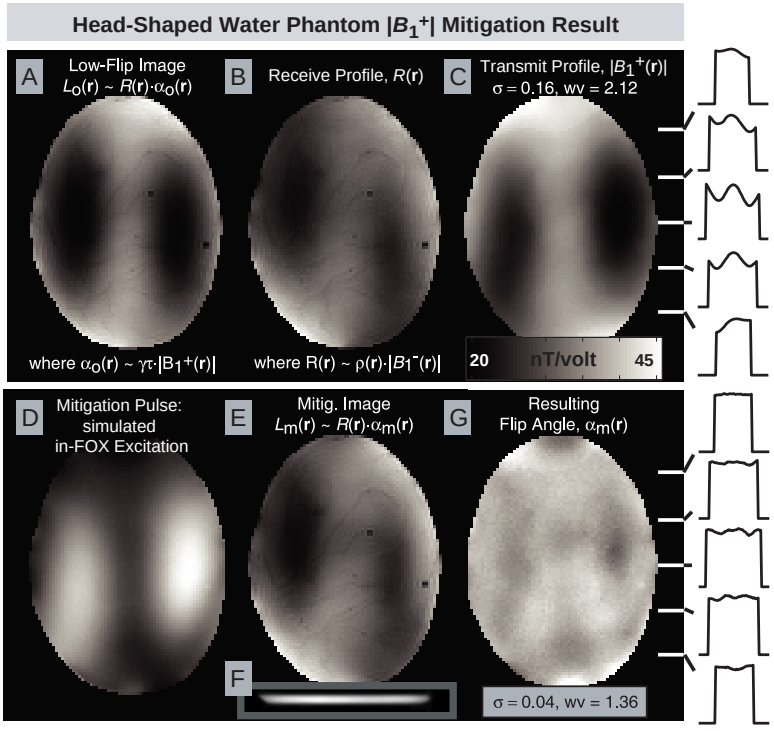

Figure 5-3: **Water phantom: $|B_1^+|$ mitigation due to a 7-ms 19-spoke pulse**. Panel A: original image, $L_0(\mathbf{r})$, collected using standard pulse. Panel B: receive profile estimate $R(\mathbf{r})$. Panel C: transmit profile estimate $|B_1^+(\mathbf{r})|$ in nT/V [proportional to unmitigated flip angle map $\alpha_0(\mathbf{r})$]. Panel D: in-plane excitation created by mitigation pulse (simulated); closely resembles $|B_1^+(\mathbf{r})|^{-1}$. Panel E: in-plane mitigated image, $L_m(\mathbf{r})$. Panel F: through-plane profile of mitigated readout volume showing successful slice selection. Panel G: post-mitigation flip angle map estimate, $\alpha_m(\mathbf{r})$; σ and WV reduced by 4x and 1.9x relative to $\alpha_0(\mathbf{r})$. It seems the pulse mitigates a large amount of $|B_1^+(\mathbf{r})|$ inhomogeneity.

algorithm whose inputs are non-smoothed raw images. Overall, this suggests that the fitted transmit profile is a realistic estimate.

The Bloch simulation of the in-plane excitation created by the 19-spoke pulse is given in Panel D; it closely resembles $|B_1^+(\mathbf{r})|^{-1}$, as intended. The similarity of the mitigated in-plane image, $L_m(\mathbf{r})$ (Panel E), to the receive profile, $R(\mathbf{r})$ (Panel B), suggests that the post-mitigation flip angle map may be fairly uniform, while the through-plane profile of the mitigated volume (Panel F) indicates the pulse succeeds at slice selection. The post-mitigation flip angle map, $\alpha_m(\mathbf{r})$ (Panel G), does indeed confirm that in-plane flip angle is fairly uniform across space. Quantitatively, σ and WV have been reduced by factors of 4 and 1.9, respectively.

The pulse itself is 7.5-ms long and transmitted at a peak value of 243 V. Pulse design details appear in Fig. 5-4. Note here that the spoke locations chosen by the sparsity-enforced

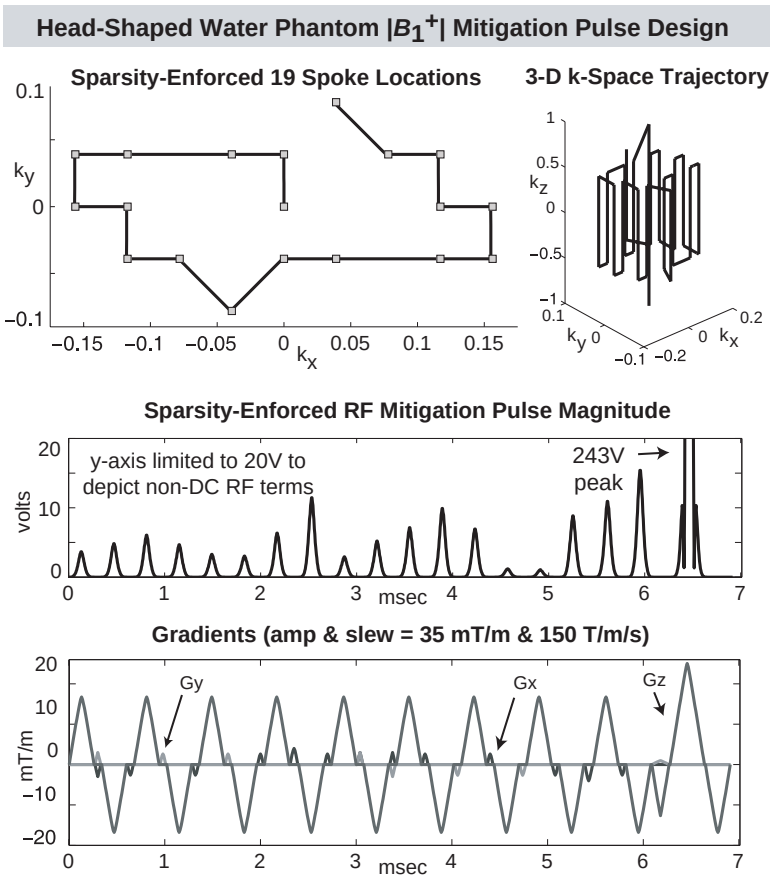

Figure 5-4: **Water phantom: 19-spoke $|B_1^+|$ mitigation pulse design**. Upper-left: locations chosen by sparsity-enforced spoke placement method. Upper-right: 3-D k-space trajectory. Middle row: 7-ms mitigation pulse magnitude (V). Bottom row: gradients.

method differ from those chosen in the earlier experiment (see Fig. 5-2) because the spoke patterns and pulses generated by the sparsity-enforced method depend on both the desired excitation and $|B_1^+|$ map.

***In vivo* human brain: $|B_1^+|$ inhomogeneity mitigation**. In this clinical scenario a 7.5-ms 19-spoke pulse attempts to produce a uniform magnetization in an axial slice of a healthy volunteer's brain. Fig. 5-5 depicts the low-flip angle image, receive profile, transmit profile, and other images; formatting here is identical to Fig. 4. As in prior experiments, the pulse's Bloch-predicted excitation (Panel D) closely resembles the inverse of $|B_1^+|$ (Panel C), and the through-plane profile of the mitigated volume (Panel F) confirms that the pulse excites only the intended region. The post-mitigation flip angle map, $\alpha_m(\mathbf{r})$ (Panel G), is more uniform that the original transmit profile (Panel C); this is apparent from the 1-D profiles as well as the fact $\alpha_m(\mathbf{r})$ has 3 times and 1.7 times lower σ and WV, respectively,

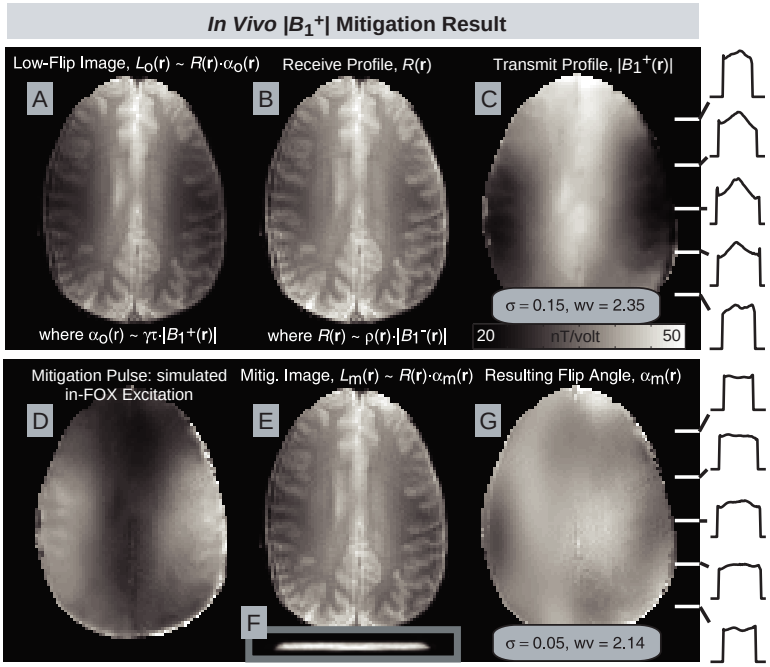

Figure 5-5: ***In vivo*** $|B_1^+|$ **mitigation due to a 7-ms 19-spoke pulse**. Panel A: original image, $L_0(\mathbf{r})$, collected using standard pulse. Panel B: receive profile estimate $R(\mathbf{r})$. Panel C: transmit profile estimate $|B_1^+(\mathbf{r})|$ in nT/V [proportional to unmitigated flip angle map $\alpha_0(\mathbf{r})$]. Panel D: in-plane excitation created by mitigation pulse (simulated). Panel E: in-plane mitigated image. Panel F: through-plane readout volume profile showing successful slice selection. Panel G: post-mitigation flip angle map, $\alpha_m(\mathbf{r})$; σ and WV reduced by 3x and 1.7x relative to non-mitigated $\alpha_0(\mathbf{r})$. It seems the pulse mitigates the $|B_1^+|$ inhomogeneity enough to produce a fairly-uniform magnetization.

than the original $|B_1^+|$ profile and flip angle map $\alpha_0(\mathbf{r})$.

Fig. 5-6 shows the design of the 19-spoke pulse. It is 7.5-ms long and transmitted at 203 V. We see from the (k_x, k_y) plot that the automated sparsity-enforced design method has chosen a placement pattern and pulse differing from those in earlier experiments (see Fig. 5-2, Fig. 5-4). The algorithm seems capable of determining good spoke locations in a variety of scenarios.

Robustness to λ. Empirically, we find that pulse designs are robust to the choice of λ in (5.5). That is, for various λs and fixed T, the algorithm often finds similar sets of spoke locations and produces magnetizations with similar degrees of uniformity.

Comparisons to prior work. The spoke-based B_1^+ mitigation pulse design method in [108] requires users to visually inspect and tune a control parameter while working on the scanner in order to produce a mitigated image, whereas the sparsity-enforced placement method is automated and seems robust to its λ parameter. Furthermore, prior work does

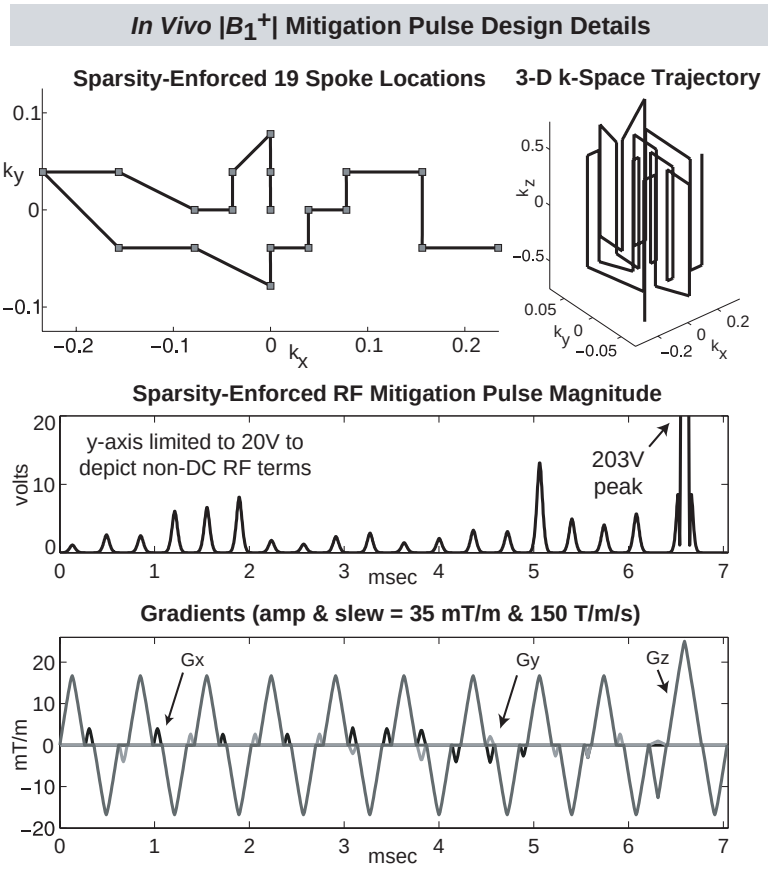

Figure 5-6: ***In vivo* 19-spoke $|B_1^+|$ mitigation pulse design**. Upper-left: locations chosen by sparsity-enforced spoke placement method; these differ from those chosen in the phantom experiment. Upper-right: 3-D k-space trajectory. Middle row: 7-ms mitigation pulse magnitude (V). Bottom row: gradients.

not provide a means to estimate the post-mitigation flip angle map in the presence of a non-uniform and possibly proton-weighted receive profile and is thus not able to truly characterize the extent to which B_1^+ inhomogeneity is mitigated by a pulse designed for that purpose.

Limitations. One limitation here is the lengthy time required to estimate $|B_1^+(\mathbf{r})|$. The other involves the slice thickness and duration of mitigation waveforms. In order to play 19 spokes in a feasible period of time given our gradient constraints, we chose to excite 20-mm slabs, but in many practical cases 5-mm slices are desired. Fortunately, this problem may be minimized by using commercially-available fast insert head gradients that are already in use at a number of sites. These gradients have amplitude and slew rate limits of 80 mT/m and 800 T/m/s, respectively; if these limits are conservatively constrained to 35 mT/m and 600 T/m/s, the 19-spoke patterns discussed earlier are able to be implemented to excite only 5 mm (10 mm) of tissue in less than 8 ms (5.5 ms).

5.3.6 Summary

Imaging experiments at 7T showed that applying sparsity-enforced joint trajectory-pulse design to spoke-based pulses yielded designs capable of mitigating B_1^+ inhomogeneity in both a head-shaped water phantom and the human brain, producing nearly uniform transverse magnetizations in each case. This seems to be a novel contribution to high field MRI RF excitation pulse design, B_1^+ inhomogeneity mitigation, and *in vivo* brain imaging at 7T.

5.4 Empirical Study of Single-Channel and Multi-Channel Sparsity-Enforced Spoke Placement and Comparison to Conventional Techniques

5.4.1 Overview

We continue to apply the general joint trajectory-pulse design algorithm of Sec. 5.2 to spoke-based pulses [108, 113, 132], but whereas in Sec. 5.3 we designed pulses solely for a single-channel system, here we design pulses for both single-channel and eight-channel systems. Our motivation is to compare sparsity-enforced design to conventional methods and explore the former's sensitivity to its control parameters. Specifically, the utility of

the sparsity-enforced spoke placement method is demonstrated by conducting a $|R||B_1^+|$ mitigation experiment in a head-shaped water phantom on a human scanner at 7T and a spatially-selective excitation experiment on an eight-channel parallel transmission system at 3T. In each experiment, the sparsity-enforced placement algorithm is compared to the conventional Fourier-based placement method and to an extension of the latter, which we call "inversion-based" placement. After demonstrating the superior performance of the sparsity-enforced placement algorithm, we proceed to characterize it in detail by exploring its sensitivity to its control parameters. Results for these various experiments are first obtained via Bloch-equation simulations and then validated with trials on the 7T and 3T systems; thus the sparsity-enforced method is investigated in both theory and practice.

Sec. 5.4.2 briefly summarizes spoke-based pulse design for multi-channel excitation systems along with transmit and receive profile estimation. Sec. 5.4.3 presents the conventional Fourier-based spoke placement method and the inversion-based technique, while a brief discussion of the sparsity-enforced algorithm in the context of conventional spoke-based pulse design) appears in Sec. 5.3.3. Sec. 5.4.5 describes the 7T and 3T systems and their associated experiments, while Sec. 5.4.6 presents and discusses the experimental results. A summary of this section's material is given in Sec. 5.4.7.

5.4.2 Methods

Spoke-based RF excitation pulse generation. For most spoke-based pulse designs, one fixes not only $\mathbf{k}(t)$ and $d(\mathbf{r})$, but desired slice thickness and spoke type as well. In this work, slice thickness is fixed at 10 mm and each spoke is a Hanning-windowed sinc. For all T-spoke pulses, the spoke at DC in (k_x, k_y) has a time-bandwidth product of 4, while the $T-1$ off-DC spokes have k_z-lengths half that of the former. (Using shorter off-DC spokes lets one reduce the duration of a pulse without noticeably impacting slice selection performance [113].) Overall, these constraints fix the shape of the P RF pulses and all properties of the slice profile (e.g., sidelobes and sharpness) [108, 113], while still letting the user retain control over the amplitude and phase each channel encodes along each spoke.

For a T-spoke pulse on a P-channel system, these constraints cause substantial changes to the design equations (2.14, 2.15) given in Sec. 2.1.11. First, all but T columns of $\mathbf{F}$ are discarded, since only T locations in (k_x, k_y)-space (those where the spokes are located) are available for weighting. Further, each $\mathbf{b}_p$ reduces from N_t to T elements. Finally, the

FOX reduces to a 2-D in-plane region, leading to a far smaller N_s than in a 3-D case. The system now has only PT unknowns, giving the user control over PT weights to influence the *in-plane* excitation pattern; these may be solved for via a pseudoinversion of $\mathbf{A}_{\text{tot}}$ in (2.15) or the Tikhonov regularization in (2.16). At this point, we are able to produce a set of gradients, determine the time-dependent k-space trajectory, and encode the proper amplitudes and phases along the P pulse shapes, concluding the design process.

Transmit and receive profile estimation. For all experiments in this section, we estimate the spatial transmit profiles, $S_1(\mathbf{r}), \ldots, S_P(\mathbf{r})$ (T/V), via the method outlined in Sec. 2.1.13. In our work, we only need coil profile estimates in the 2-D plane where the thin-slice excitation occurs, so here $\mathbf{r}$ indexes samples only within a 2-D FOX.

5.4.3 Conventional Spoke Placement

We now describe two conventional methods for determining the placement of spokes in (k_x, k_y) when designing a spoke-based RF waveform whose goal is to produce a user-defined in-plane target excitation, $d(x, y)$, using a P-channel system whose coil profiles are known. The second method will suggest spoke weights in addition to placement. Assume that the FOV is a rectangle, so when $d(x, y)$ is sampled, its N_{FOV} samples may be assembled into $\mathbf{D} \in \mathbb{C}^{M \times N}$. Furthermore, assume the FOX is smaller than the FOV and not necessarily rectangular. It is discretized and comprised of $N_{\text{FOX}} < N_{\text{FOV}}$ samples, which are a proper subset of the FOV samples. (The $N_{\text{FOV}} - N_{\text{FOX}}$ pixels in the FOV but not in the FOX comprise a "don't care" region.) Finally, assume the system's P coil profiles are sampled within the FOX and assembled into the diagonal matrices $\mathbf{S}_1$ through $\mathbf{S}_P \in \mathbb{C}^{N_{\text{FOX}} \times N_{\text{FOX}}}$. The goal now is to place spokes at T locations in (k_x, k_y) such that the resulting in-plane excitation is close (in the ℓ_2 sense) to $d(x, y)$ *within the FOX* (not the entire FOV).

Fourier-Based Spoke Placement

An intuitive way to determine where to place each spoke is to compute the Discrete Fourier Transform (DFT) of $\mathbf{D}$ and from the resulting discrete grid of (k_x, k_y) frequencies, choose the T whose Fourier coefficients are largest in magnitude [145, 147]. If the complex weights at these T locations are brought back into the spatial domain via an inverse transform, Plancherel's theorem implies that this image, $\widehat{\mathbf{D}}_T$, is the best ℓ_2 approximation of $\mathbf{D}$ *within the FOV* [90, 99]. Unfortunately, there are problems with this method: 1) The FOX is

typically smaller than the FOV, and the RF profile in the don't care region is not important. The Fourier method, however, has no concept of the FOX or the don't care region. This FOX-FOV mismatch means that the T chosen locations do not necessarily yield the best ℓ_2 representation of $d(x, y)$ *within the FOX*. 2) The influence of each spatial profile, $S_p(\mathbf{r})$, is not accounted for, even though each has a major impact on the in-plane excitation. 3) There is no concept of transmit channels, so the method cannot suggest a weight for each channel to place at each spoke location.

Inversion-Based Spoke Placement

Single-channel derivation. This method retains the logic of the Fourier technique while accounting for its shortcomings. For now, assume the system has a single channel and imagine a grid of N_f arbitrarily-spaced points in (k_x, k_y) located at $\mathbf{k}_1, \ldots, \mathbf{k}_{N_f}$. An arbitrary choice of complex weights at different points on this grid results in a nominal spatial domain excitation, which is related to the weighted grid by a Fourier transform. The nominal excitation is then multiplied (pointwise) by the coil profile to yield the actual excitation, $m(x, y)$. One may sample this excitation within the FOX at $\{\mathbf{r}_1, \ldots, \mathbf{r}_{N_{\text{FOX}}}\}$ and arrange the samples into $\mathbf{m} \in \mathbb{C}^{N_{\text{FOX}}}$. Likewise, the weights on the grid may be formed into $\mathbf{g} \in \mathbb{C}^{N_f}$. The relation between $\mathbf{m}$ and $\mathbf{g}$ is represented by $\mathbf{F} \in \mathbb{C}^{N_{\text{FOX}} \times N_f}$, where $\mathbf{F}(m, n) = i\gamma M_0 \Delta_t e^{i\mathbf{r}_m \cdot \mathbf{k}_n}$, where the nth column of $\mathbf{F}$ describes how the excitation at the $\mathbf{r}_m$s is influenced by the weight at $\mathbf{k}_n$. When samples of the coil profile are arranged into $\mathbf{S} \in \mathbb{C}^{N_{\text{FOX}} \times N_{\text{FOX}}}$, this leads to

$$\mathbf{m} = \mathbf{SFg}. \tag{5.10}$$

In short, (5.10) expresses the excitation that forms at a set of points in the FOX when an arbitrary set of complex weights is placed on an arbitrary k-space grid. For fixed $\mathbf{k}_n$, the column of $\mathbf{F}$ here is nearly identical to the column of its counterpart in (2.14, 2.15) of Sec. 2.1.11, except now there is no notion of time. For the most part, this is tolerable: $\mathbf{k}_n$ is known rather than $\mathbf{k}(t)$ at time t_n. The difference here is that time-dependent ΔB_0 phase accrual cannot be incorporated because $\mathbf{k}$'s dependence on t is not yet known. Fortunately, if B_0 is nearly homogeneous or pulse duration is short, then the column of $\mathbf{F}$ here (nearly) equals the column of its counterpart in (2.14, 2.15) of Sec. 2.1.11.

By sampling $d(x, y)$ analogously to $m(x, y)$, it is possible to form $\mathbf{d} \in \mathbb{C}^{N_{\text{FOX}}}$, and by

solving $\mathbf{d} = \mathbf{SFg}$ for $\mathbf{g}$, the weighting of the (k_x, k_y) grid that best achieves the target excitation in the ℓ_2 sense *within the FOX* may be obtained. This solution is generated via regularization rather than direct pseudoinversion:

$$\min_{\mathbf{g}} \{\|\mathbf{d} - \mathbf{SFg}\|_2^2 + \delta\|\mathbf{g}\|_2^2\}. \tag{5.11}$$

This yields a well-conditioned $\mathbf{g}$ that (approximately) solves the system and provides a set of grid weights that form the desired excitation. At this point, two pitfalls of the Fourier method have been mitigated: the coil profile has been explicitly incorporated and the unnecessary constraints on the excitation outside the FOX have been removed. Now, using the intuition that motivated the Fourier method, we decide to place spokes at the T points on the grid that correspond to the T largest-magnitude elements of $\mathbf{g}$.

Multi-channel derivation. For multi-channel systems, (5.10) generalizes to

$$\mathbf{m} = \mathbf{S}_1\mathbf{F}\mathbf{g}_1 + \cdots + \mathbf{S}_P\mathbf{F}\mathbf{g}_P = \mathbf{A}_{\text{tot}}\mathbf{g}_{\text{tot}}, \tag{5.12}$$

analogously to (2.14). A regularization analogous to (5.11) is then applied to solve $\mathbf{d} = \mathbf{A}_{\text{tot}}\mathbf{g}_{\text{tot}}$, yielding a well-conditioned solution, $\mathbf{g}_{\text{tot}}$, which is disassembled into $\{\mathbf{g}_p\}_{p=1}^P$, where $\mathbf{g}_p$ contains the weights for channel p. When an element of $\mathbf{g}_p$ is large in magnitude, channel p is using the corresponding grid point to strongly influence the resulting excitation, so placing a spoke here may be worthwhile. However, at this same location, imagine that the other $P - 1$ channels have weights of zero; should a spoke still be placed here, if only one channel suggests doing so, or should it be placed at an alternate location, one where each channel contributes toward the formation of $d(x, y)$?

The presence of multiple channels has led to a quandary about how to choose the T best spoke locations. A solution to this problem is to "compress" the energy of the resulting $\mathbf{g}_p$-weighted grids along the *channel dimension* using the ℓ_2 norm. This produces a single grid that represents the overall energy placed at each (k_x, k_y) location by all P channels, denoted $\mathbf{g}^{(\ell_2)}$, where $\mathbf{g}^{(\ell_2)}(i) = \|[\mathbf{g}_1(i), \ldots, \mathbf{g}_P(i)]^{\mathrm{T}}\|_2$. Using this norm to combine the energy deposited by all P channels at the ith grid location has no sparsifying effects, so $\mathbf{g}^{(\ell_2)}(i)$ will not radically differ if one channel deposits a large amount of energy at the ith grid point, or if this same amount of total energy is contributed by many channels. Forming $\mathbf{g}^{(\ell_2)}$ is sensible because when performing excitations using a multi-channel system, it doesn't

matter if it is one or many channels that place energy at a given point if such a deposition improves the excitation. Based on this reasoning, we reapply the intuition that inspired the Fourier-based method, and decide to place spokes at the T locations corresponding to the largest-magnitude elements of $\mathbf{g}^{(\ell_2)}$. A T-spoke pulse and corresponding gradients may then be generated using the approach given in Sec. 5.4.2.

Unfortunately, using Tikhonov regularization to find $\mathbf{g}_{\text{tot}}$ yields a dense solution, leading to a dense $\mathbf{g}^{(\ell_2)}$, one where *all* grid locations experience moderate amounts of energy deposition. Frequently, the energies at each of the T chosen locations are not much greater than the energies present at most other points. Up until this point, the magnitude of energy at a grid point has been used as proxy for whether that point is a good spoke location, but since the energies being deposited across all candidate locations are now similar, it is no longer clear if this intuition is useful. The ideal weights for each channel to place on the candidate grid have indeed been determined, but this process *has not restricted the number of points at which a channel may deposit energy.*

This is a serious drawback, because when T spokes are chosen, *only* T locations in (k_x, k_y) remain for the P channels to modulate and attempt to form $d(x, y)$. Thus in practice, when all weights are zeroed out except those PT weights associated with the chosen locations, the resulting excitation does not closely resemble the desired one. The achieved excitation is of low quality because the PT chosen weights relied on the weights placed at all other grid locations to form an accurate excitation. To mitigate this problem, the PT weights need to be retuned via the process given in Sec. 5.4.2). In short, this method does not encourage the channels to form the excitation by modulating only a small number of locations, despite the fact that only a small subset of candidate grid points are able to be retained for the final pulse design.

5.4.4 Multi-Channel Sparsity-Enforced Spoke Placement

To mitigate the inversion-based method's dense $\mathbf{g}^{(\ell_2)}$ grid problem, we rely on sparse approximation and pose an optimization that promotes the sparsity of $\mathbf{g}^{(\ell_2)}$:

$$\min_{\mathbf{g}_{\text{tot}}} \ \{\tfrac{1}{2}\,\|\mathbf{d} - \mathbf{A}_{\text{tot}}\mathbf{g}_{\text{tot}}\|_2^2 + \lambda\|\mathbf{g}^{(\ell_2)}\|_1\}. \tag{5.13}$$

This is equivalent to the MSSO relaxed convex optimization (3.3), the only difference being notational ($\|\mathbf{g}^{(\ell_2)}\|_1$ is used rather than $\|\mathbf{G}\|_{\mathrm{S}}$). With large enough λ, this method produces $\mathbf{g}_p$s that approximately solve the system while also yielding a sparse $\mathbf{g}^{(\ell_2)}$. Typically, solving (5.13) yields a $\mathbf{g}^{(\ell_2)}$ with many elements close to zero, with only several high-energy grid points. (The small grid weights are not exactly zero; this is because an ℓ_1 regularization is used to promote sparsity rather than an ideal but computationally intractable ℓ_0 penalty.)

In this way, the sparsity-enforced spoke placement reveals a small subset of the overall grid that alone is able to form the excitation. At this stage, T spoke locations are selected by choosing those points on the grid corresponding to the largest-magnitude elements of $\mathbf{g}^{(\ell_2)}$. In contrast with the inversion-based method, the PT weights associated with these T locations alone are typically able to generate a reasonable version of the desired excitation when other weights are zeroed out. Thus this method determines a small set of spoke locations and weightings that form an acceptable-quality excitation. Like the inversion-based method, the weights undergo retuning (discussed next), but in many cases they do not radically change.

Necessity of simultaneous sparsity. What if rather than promoting the sparsity of $\mathbf{g}^{(\ell_2)}$, we simply promote that of $\mathbf{g}_{\mathrm{tot}}$? This would produce a sparse $\mathbf{g}_{\mathrm{tot}}$, and thus individually-sparse $\mathbf{g}_p$. But this does not encourage *simultaneous sparsity* among the $\mathbf{g}_p$s, so their sparsity profiles differ, and thus when $\mathbf{g}^{(\ell_2)}$ is formed, it is not necessarily sparse, contradicting our goal. Further, $\|\mathbf{g}_{\mathrm{tot}}\|_1$ discourages solutions where many channels deposit energy at a single location, and thus penalizes solutions which are in fact useful. A T-spoke pulse and corresponding gradients may then be generated using the approach given in Sec. 5.4.2.

5.4.5 System and Experiment Setup

$|R||B_1^+|$ Mitigation in a Head-Shaped Water Phantom on a Single-Channel 7T Scanner

Overview. Each placement method will design pulses that mitigate $|R||B_1^+|$ inhomogeneity present in a head-shaped water phantom at 7T using different numbers of spokes. Here, the ideal mitigation pulse is one that excites the multiplicative inverse of the inhomogeneity [108] and produces a uniform magnetization across the FOX. Each method will attempt to

produce this ideal excitation.

Hardware. Work is conducted on a 7T whole-body human scanner (Siemens Medical Solutions, Erlangen, Germany). The scanner is built around a 90-cm diameter magnet (Magnex Scientific, Oxford, United Kingdom). The amplitude and slew rate of the system's head gradient coils are always constrained to 35 mT/m and 600 T/m/s, respectively. Potential gradient imperfections are minimized by operating at amplitude and slew rates below the hardware's maximum specifications and performing a one-time RF-gradient delay calibration. A single-channel, 28-cm diameter, quadrature bandpass birdcage coil is used for transmission and reception [11, 126].

Spatial profiles, target pattern, and spoke parameters. A GRE magnitude image of the head-shaped water phantom is obtained using a conventional low-flip slice-selective pulse with a repetition time (TR) of 20 ms, an echo time (TE) of 5 ms, a bandwidth (BW) of 390 Hz/pixel, and a FOV of 25.6 cm in each dimension. The dielectric properties of the water in the phantom cause its transmit and receive profiles to exhibit significant spatial variation, causing the resulting image intensity to be highly non-uniform. One sees from (2.17, 2.19) in Sec. 2.1.12 that this image is linearly related to both the transmit profile and the receive profile when in the small-tip angle regime, because $\sin(\alpha) \simeq \alpha$ for small α.

We use this image—the *combination* of the transmit and receive profiles—as proxy for the transmit profile. Naturally, when one implements a B_1^+ mitigation pulse for *in vivo* use, one should mitigate *only* the transmit profile, as this achieves optimal contrast and SNR, but here we will mitigate the combination of profiles, for two reasons: 1) The combination exhibits more spatial variation than the transmit profile alone, making the spoke selection problem more challenging. 2) A pulse designed to mitigate this combination, rather than simply the transmit profile, produces a result that is easy to understand and evaluate, since it ideally will produce a uniformly image.

The hypothetical transmit profile to be mitigated appears in Fig. 5-7 in nT/V, with a 25.6 cm $\times$ 25.6 cm FOV and 4-mm resolution. The FOX is where the phantom is present. The in-plane target magnetization magnitude is a uniform 10-degree flip and target phase is zero. (We have no desire to shape the nearly-constant phase that occurs naturally across the target.) Spoke parameters for all designs are given in Sec. 5.4.2. Based on these parameters and gradient constraints, the duration of a T-spoke pulse, $L(T)$, is close to $0.26T + 0.34$ ms, i.e., each spoke adds roughly 1/4 ms to pulse duration. The in-plane (k_x, k_y) traversals take

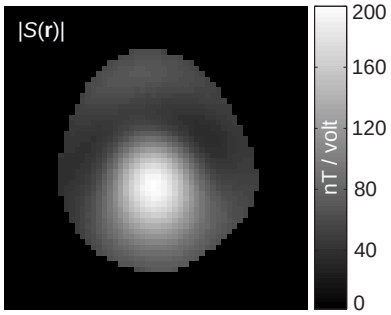

Figure 5-7: **Quantitative $|R||B_1^+|$ map of the head-shaped water phantom at 7T (nT/V)**. This hypothetical transmit profile is generated by collecting a GRE magnitude image and scaling the resulting pixels.

negligible time because most time is spent playing spokes in k_z. Finally, ΔB_0 is estimated from the phase of two GRE images (TE_1=5 ms, TE_2=6 ms).

Experiment details: simulations, quality metrics, and validation trial on real system. First, simulations are conducted as a function of T to demonstrate the superior performance of the sparsity-enforced technique; here, we assume we know a good value for the sparsity-enforced method's λ parameter. The quality of the in-plane excitation produced by each method is evaluated by computing the within-FOX, in-plane root-mean-square error (RMSE) and maximum error between each resulting excitation and the desired magnetization. Because the desired pattern is in degrees, each of these error terms are in degrees as well. Further, because the resulting pulses are in volts, we calculate the root-mean-square (V_{RMS}) and peak voltage (V_{peak}) of each. After the simulations, a sparsity-enforced pulse is played on the scanner to show that the experimental result closely matches the one predicted via simulation. We then return to simulations and characterize how the sparsity-enforced placement algorithm behaves when provided grids of oversampled candidate spoke locations. Finally, we investigate how RMSE varies as a function of λ and T. Results appear in Sec. 5.4.6.

Phase-Encoded, Spatially-Selective Excitation in an Oil Phantom on an Eight-Channel 3T Scanner

Overview. Here we evaluate how well the spoke placement methods design pulses that form a highly-structured excitation pattern in a 17-cm oil phantom, where the FOX is the phantom itself and transmission is conducted using an eight-channel parallel excitation

system.

Hardware. The parallel system is built around a 3T Magnetom Trio scanner (Siemens Medical Solutions, Erlangen, Germany). The coil array is composed of eight circular surface coils arranged on a 28-cm diameter acrylic tube [7]. After a set of pulses is designed for use on the scanner, the eight resulting waveforms are transmitted simultaneously through their respective coils and modulated in magnitude and phase as dictated by the pulse design. The system's body coil is used for reception. The amplitude and slew rate constraints of the gradients are 35 mT/m and 150 T/m/s, respectively. Gradient imperfections are mitigated by operating in a region well within the hardware's maximum specifications; RF-gradient mismatch is prevented via a one-time delay calibration.

Spatial profiles, target pattern, and spoke parameters. The system's 25.6 cm $\times$ 25.6 cm FOV, 4-mm resolution transmit profile magnitudes (nT/V) are illustrated in Fig. 2-6. The smooth variations exhibited by each profile occur without smoothing the fitted results, leading us to believe that the fitting method is robust and accurate. The receive profile (not shown) is smooth, varying less than 5% within the FOX. Spoke type and thickness are the same as in the single-channel case. The duration of a T-spoke pulse, $L(T)$, fits well to $0.47T + 0.65$ ms; ΔB_0 is again obtained from two phase maps with a 1-ms TE difference.

With the profiles of Fig. 2-6, it is possible to run the spoke placement algorithms and produce pulses. For all experiments, $d(x, y)$ is the phase-encoded bifurcation depicted in Fig. 5-8. This target has a high degree of spatial selectivity, experiencing a 10-degree flip within only two thin "veins" and no flip across the rest of the FOX. The left vein is 90 degrees out of phase with the right vein. Exciting this pattern is worthwhile because highly-structured excitations may have applications to clinical MRI. Furthermore, the strength of multi-channel excitation will be demonstrated. Finally, the phase-encoded nature of this excitation may have applications to phase-contrast magnetic resonance angiography (MRA) [42, 116], perhaps allowing MRA concepts to be ported to general MRI.

Experiment details: simulations, quality metrics, and validation trial on real system. Analogously to the single-channel system experiments, simulations are first conducted and RMSE and ℓ_∞ metrics are used to evaluate the performance of the different methods as a function of T. Voltages of the resulting pulses are also analyzed. Each design consists of eight RF pulses, so to succinctly present the data, the maximum peak and max-

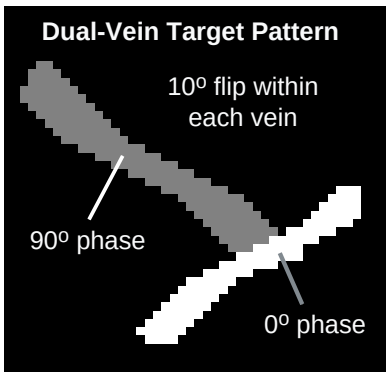

Figure 5-8: **Dual-phase bifurcation target excitation used for all 3T eight-channel system experiments**. Within the two veins the target commands a 10-degree flip and zero flip elsewhere. Each branch is phase encoded: the left vein is 90 degrees out of phase with the right. The ability to achieve this structured excitation pattern will show the strength of both multi-channel excitation and sparsity-enforced spoke placement.

imum RMS voltages observed across each set of eight waveforms are recorded. After the Bloch-simulated trials, a sparsity-enforced pulse is played on the scanner and the resulting excitation is compared to the one predicted via simulation. We then return to simulations, analyzing how the sparsity-enforced algorithm behaves when provided grids of candidate spoke locations with different extents in k-space, along with the method's sensitivity to its control parameter, λ.

5.4.6 Results and Discussion

Single-channel system: Bloch-simulated spoke placement analysis. Here, each method is used to place $T = 1, \ldots, 40$ spokes. Based on Sec. 5.4.5, we are able to generate all matrices and vectors needed to run the placement and pulse design routines, e.g., 1138 within-FOX samples of the transmit profile, $S(x, y)$, are used to form the spatial profile matrix, $\mathbf{S}$, all elements of $\mathbf{d}$ are set to 10 degrees based on the desired excitation $d(x, y)$, etc. The pointwise inverse of the inhomogeneity, $S^{-1}(x, y)$, is provided as the desired excitation to the Fourier placement method. The inversion-based method's Tikhonov parameter in (5.11) is set to 0.1.

The frequency grid of candidate spoke locations provided to the inversion-based and sparsity-enforced methods is centered at DC, Nyquist-spaced at $\frac{1}{25.6}\text{cm}^{-1}$, extends outward equally in both k_x and k_y, and is comprised of 17^2 points, leading to $\mathbf{F} \in \mathbb{C}^{1138\times 289}$. Providing this same grid to the inversion and sparsity-enforced methods promotes fair com-

parisons between the algorithms. Ideally, this grid would extend further out in k-space to reach the maximum and minimum frequencies suggested by the transmit profile's 4-mm sample spacing, but using such a large grid increases the sparsity-enforced method's runtime to 25 minutes (vs. 3 minutes for the 17^2-point grid). The smaller grid is acceptable, however, because the Fourier-based method—which selects spokes on the large, ideal grid—never chooses to place a spoke outside of this 17^2 region when $T \leq 40$, i.e., the majority of k-space energy lies on the small grid. (Later we will show how grid extent affects the sparsity-enforced method's performance.)

The final step is to run the sparsity-enforced routine for $R = 40$ values of λ, solving (5.13) each time and storing each resulting weighted grid, $\mathbf{g}^{(r)}$. When the sparsity-enforced method is used to place T spokes, R possible placements are evaluated—spoke locations being chosen based on the T maximum values of each $\mathbf{g}^{(r)}$—and the placement yielding the smallest residual error, $\|\mathbf{d} - \mathbf{S}\mathbf{F}\mathbf{g}^{(r)}_{\text{trunc},T}\|_2$, is retained. Thus when the sparsity-enforced method's results are presented, they implicitly assume that a good value for λ is known. (Later on, we will analyze sensitivity to λ.) We do not find it necessary to loop over the inversion-based method's Tikhonov term in (5.11) because we observe that once it is tuned past a certain threshold, all resulting solutions suggest essentially the same set of T spoke locations.

Each method designs pulses with $T = 1, \ldots, 40$ spokes, which are then simulated. The RMSE and maximum error of each resulting excitation pattern with respect to the uniform target are computed, along with the voltage characteristics of each corresponding pulse. Fig. 5-9 depicts these results, illustrating how each metric varies with T. For each method, RMSE decreases with T, because using more spokes allows for more spatial tailoring and inhomogeneity mitigation. For fixed T, the Fourier method is outperformed by the inversion-based method, which in turn is outperformed by the sparsity-enforced technique. For all T, the sparsity-enforced placement method produces the lowest RMSE.

When $T < 15$, the RMSE for each method is large, and the resulting magnetization still highly non-uniform. When $T \in [15, 30]$, the sparsity-enforced method, on average, produces an excitation with 1.32 times lower RMSE relative to the inversion method. The inversion-based method's excitations, in turn, have on average 1.06 times lower RMSE than those due to the Fourier method. For $T > 30$, the RMSE of the sparsity-enforced and inversion methods converges, but the resulting pulses are over 8-ms long and thus

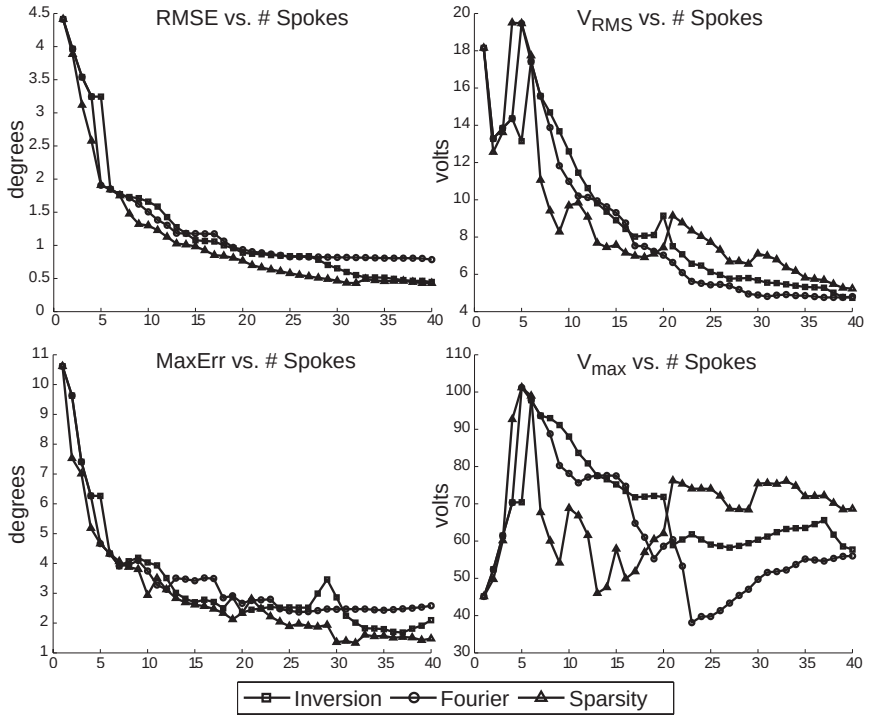

Figure 5-9: **Bloch-simulated spoke placement algorithm comparisons for mitigating $|R||B_1^+|$ inhomogeneity in the head-shaped water phantom at 7T**. Error and voltage statistics vs. number of spokes used (T) are shown for the Fourier, inversion, and sparsity-enforced spoke placement methods. Upper-left: RMSE vs. T. Lower-left: maximum error vs. T. Upper-right: V_{RMS} vs. T. Lower-right: $V_{\max}$ vs. T. For all three algorithms, the duration of a T-spoke pulse is close to $0.26T + 0.34$ ms.

impractical. Overall, it is clear that the sparsity-enforced method produces higher-quality excitations than the other methods for all practical pulse durations. To see the strength of the sparsity-enforced method, consider its 21-spoke 5.7-ms pulse with 0.68 RMSE. In order for the inversion method to produce the same quality excitation, a 29-spoke, 7.8-ms pulse is required. In other words, the sparsity-enforced pulse achieves the same quality excitation yet is 1.37 times shorter.

Unlike RMSE, maximum error does not vary smoothly with T, because the algorithms do not explicitly penalize ℓ_∞ error. In general, the maximum error curves of the inversion and sparsity-enforced methods are similar. The RMS voltages of pulses generated by each method exhibit a downward trend as T increases, because with more spokes, each individual spoke doesn't need to be modulated as intensely to form the excitation. Further, the peak voltages behave more erratically than the RMS voltages, analogously to how the ℓ_∞ error fluctuates more than RMSE. Finally, for $T \in [10, 20]$, the sparsity-enforced method often has the lowest voltages among all methods, whereas for $T > 20$, its voltages are higher.

Fig. 5-10 depicts the results of each method when $T = 21$. The left, center, and right columns show the Fourier, inversion, and sparsity-enforced method's results, respectively. The top, middle, and bottom rows depict the simulated excitation before accounting for transmit profile non-uniformity, the resulting magnetization after accounting for inhomogeneity, and a 2-D view of k-space showing where each method places its spokes and how they are traversed in-plane. Metrics are also given in correspondence with those in Fig. 5-9. From the top row, one sees that each method produces an excitation that approximates the inverse of the non-uniform profile in Fig. 5-7. Looking at the middle row, one appreciates the ability of each method to produce a relatively flat magnetization. The bottom row shows that the Fourier and inversion methods cluster spokes around DC. The sparsity-enforced method, however, places its spokes further out on the grid, leading to an excitation with 1.28 times lower RMSE. This improvement is achieved with no increase in pulse duration and only a moderate voltage increase. The spoke placement determined by the sparsity-enforced algorithm is not obvious, but by placing spokes at slightly higher spatial frequencies, the sparsity-enforced method produces a nominal excitation that is less symmetric, better resembling the pointwise inverse of the transmit profile than the excitations produced by the other methods.

Single-channel system validation. To validate our simulations, we play the 21-

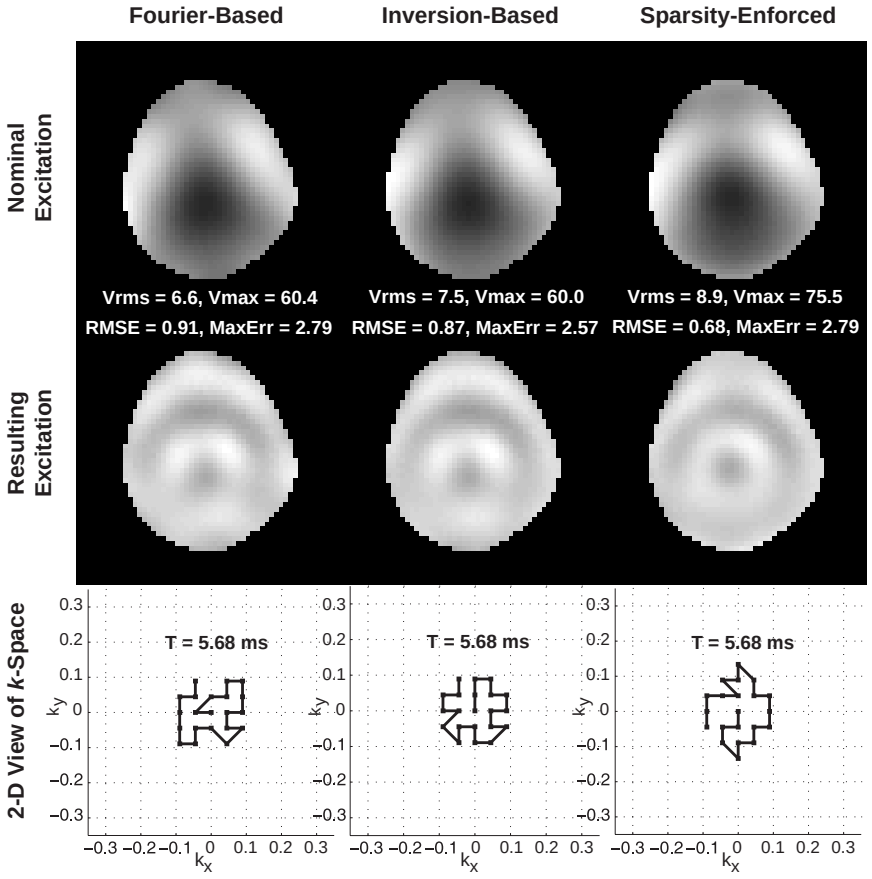

Figure 5-10: **Bloch-simulated 21-spoke pulses designed by the Fourier, inversion, and sparsity-enforced spoke placement algorithms for mitigating $|R||B_1^+|$ inhomogeneity in the head-shaped water phantom at 7T**. Columns, from left to right: results of the Fourier, inversion, and sparsity-enforced methods. Top row: excitations produced by each algorithm. Middle row: magnetizations after accounting for the inhomogeneity. Bottom row: 2-D view of k-space illustrating each spoke placement; each trajectory ends at the center of k-space. The sparsity-enforced pulse produces the lowest-RMSE excitation.

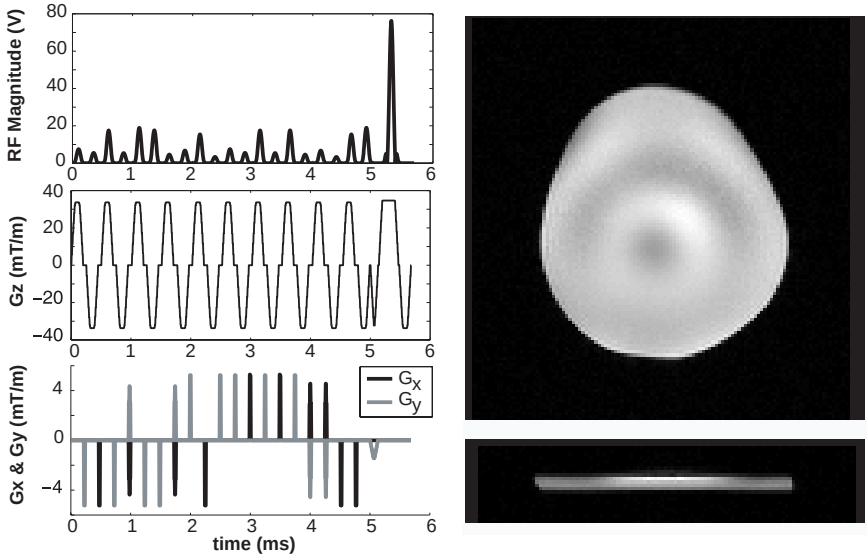

Figure 5-11: **Experimental result in the head-shaped water phantom on the single-channel 7T scanner**. Here, the 21-spoke sparsity-enforced RF pulse whose simulations appear in Fig. 5-10 is played on the actual system. Left: RF waveform magnitude and gradients. Upper-right: in-plane scanner result. Lower-right: through-plane scanner result. This experimental result closely resembles the Bloch-simulated image in Fig. 5-10, validating the simulation methodology and proving that the pulse design process is feasible for use on real systems.

spoke sparsity-enforced pulse from Fig. 5-10 on the scanner and perform a GRE readout. Since the pulse is designed to mitigate both the transmit and receive profiles, we validate our simulation by analyzing the magnitude of the readout image. Fig. 5-11 shows the RF magnitude and gradients of the 21-spoke pulse, along with the in-plane and through-plane result from the real system. The through-plane image demonstrates excellent slice selection. Further, there is a striking resemblance between the in-plane result and simulation (middle row, right column of Fig. 5-10). Note how Fig. 5-10's Fourier and inversion-based simulated patterns have two bright spots slightly north-west and north-east of center, whereas the simulated pattern and the scanner result in Fig. 5-11 have only a single bright spot north-east of center. This close match between the experimental and simulated result lends strong support to the simulation method and results.

Multi-channel system: Bloch-simulated spoke placement analysis. Here each method places 1 to 40 spokes, and RMSE, ℓ_∞ error, V_{RMS}, and V_{peak} are calculated for each of the resulting 120 pulses and excitations. The goal here is to produce the dual-vein target in Fig. 5-8. Since there are 8 channels, the matrices are 8 times larger, e.g., $\mathbf{A}_{\text{tot}}$

in (5.13) is now $1416 \times (289 * 8)$. For all trials, the inversion method's Tikhonov value in (5.11) is 0.01. The frequency grid is the same as in the single-channel experiment, so each $\mathbf{g}_p$ vector is comprised of 289 elements. With these variables, it takes about 25 minutes to solve (5.13) for one value of λ.

Analogously to the single-channel experiment, the sparsity-enforced routine is run for $R = 32$ different values of λ, so when the method is requested to place T spokes, it evaluates R possible placements and chooses the one yielding the smallest error. Here, in contrast with the single-channel experiment, $\lambda \in (10^{-9}, 10^{-1})$. Small values of λ are necessary because the application of $\|\cdot\|_{\mathrm{S}}$ to the $\mathbf{g}_p$s results in grids with barely any energy when $\lambda > 10^{-1}$.

After these steps, the algorithms place 1 to 40 spokes and design pulses for each. Fig. 5-12 shows the error and voltage metrics of each method as a function of T. The RMSE plot shows that the sparsity-enforced routine outperforms the other methods. For $T \in [10, 40]$, the sparsity-enforced technique, on average, has 1.18 times and 1.31 times lower RMSE than the Fourier and the inversion methods, respectively. Consider the sparsity-enforced method's 15-spoke, 7.5-ms pulse with 2.01 RMSE. For the Fourier method to attain this RMSE, a 26-spoke, 13-ms pulse is required, i.e., the sparsity-enforced pulse generates the same quality excitation yet is 1.73 times shorter in duration. In terms of voltages, all methods exhibit similar trends; this differs from the single-channel case where the sparsity-enforced method at times produced higher-voltage pulses. This means that the sparsity-enforced method produces pulses that yield higher-quality excitations without significantly increasing pulse duration, V_{RMS}, or V_{peak}.

These error plots also reveal a surprising result: for all T, the inversion method performs worse than the Fourier one, even though the former was designed to account for pitfalls of the latter. This differs from the single-channel case of Fig. 5-9, where the inversion method outperformed the Fourier one. The inversion method fails here because of the "grid compression" step discussed earlier. This problem does not occur in the single-channel experiment because in the latter case there is only a single $\mathbf{g}_p$ vector and the compacting of various grids to form a dense, problematic $\mathbf{g}^{(\ell_2)}$ never occurs.

Fig. 5-13 shows each method's excitation and spoke placement for $T = 15$. Its left, middle, and right columns show the results of the Fourier, inversion, and sparsity-enforced techniques. The rows, from top to bottom, depict the simulated excitation magnitude, excitation phase, and spoke placement in (k_x, k_y). Each method succeeds, to some extent,

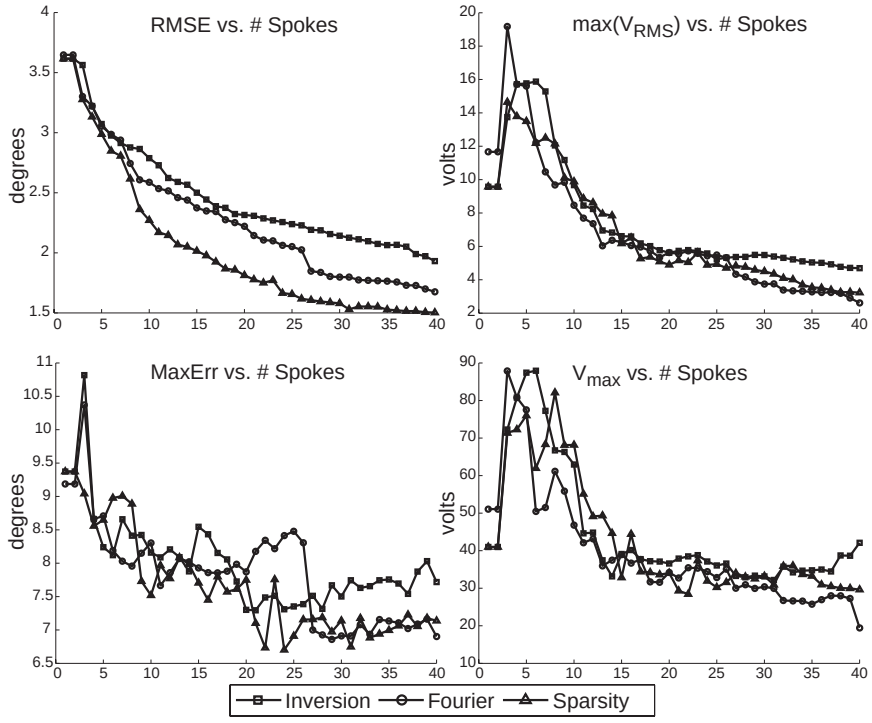

Figure 5-12: **Bloch-simulated spoke placement comparisons for forming the dual-vein target on the eight-channel parallel excitation system at 3T**. Error and voltage statistics vs. number of spokes (T) are shown for the Fourier, inversion, and sparsity-enforced placement methods. Upper-left: RMSE vs. T. Lower-left: maximum error vs. T. Upper-right: max V_{RMS} across all eight channels vs. T. Lower-right: max V_{peak} across all eight channels vs. T. For all algorithms, the duration of a T-spoke pulse is close to $0.47T + 0.65$ ms.

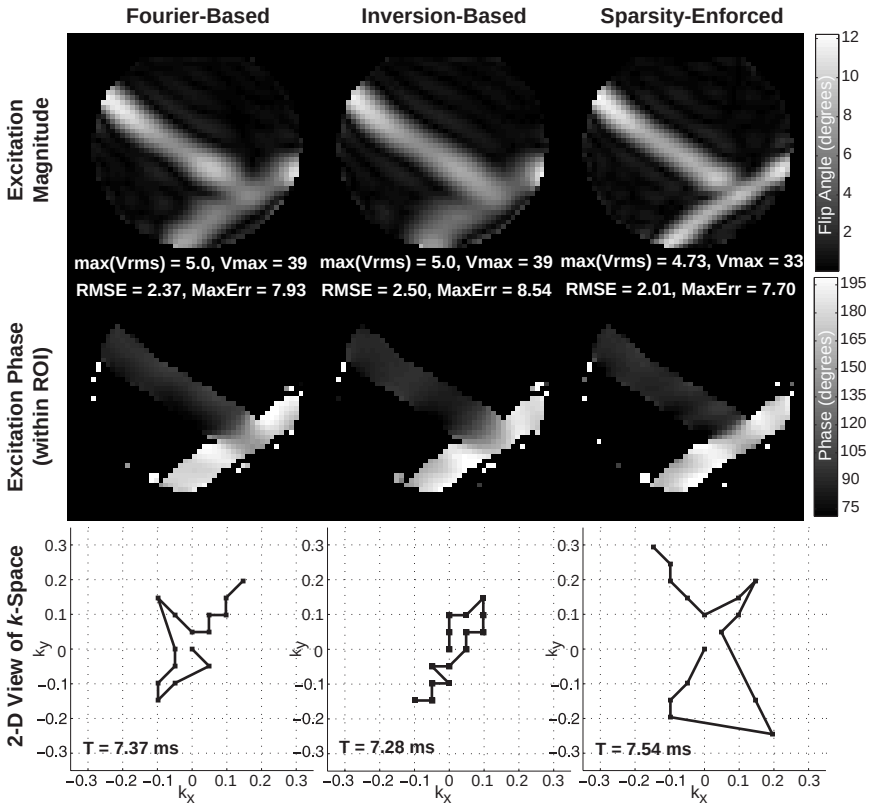

Figure 5-13: **Bloch-simulated 15-spoke pulses designed by the Fourier, inversion, and sparsity-enforced spoke placement algorithms for forming the dual-vein target on the eight-channel system at 3T**. Columns, from left to right: results of the Fourier, inversion, and sparsity-enforced spoke placement methods. Top row: excitation magnitudes. Middle row: excitation phases. Bottom row: 2-D view of k-space showing each method's spoke placement; each trajectory ends at the center of k-space. The sparsity-enforced excitation has the lowest RMSE, ℓ_∞ error, V_{RMS}, and V_{peak}.

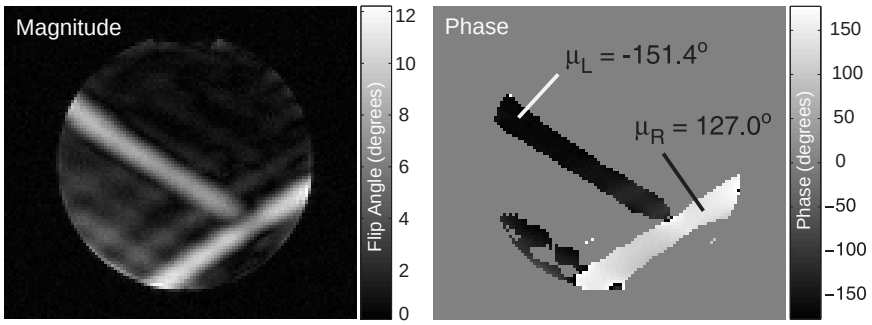

Figure 5-14: **Experimental result in an oil phantom on the eight-channel 3T system**. Here, the 15-spoke sparsity-enforced RF pulse and gradients whose simulation results appear in Fig. 5-13 are played on the actual system followed by a GRE readout. This experimental result closely resembles the Bloch-simulated magnitude and phase of this same waveform in Fig. 5-13, validating the simulation methodology and showing that the proposed pulse design process is applicable to real multi-channel systems.

in producing the desired excitation, but the sparsity-enforced method produces one that best resembles the target. Its excitation is better because it is not only less blurry than the others, its lower vein exhibits a degree of curvature not present in the veins of the other methods. This leads to the sparsity-enforced method's 2.01 RMSE, which is 1.18 and 1.24 times lower than those of the Fourier and inversion techniques. The sparsity-enforced method's spoke placement exhibits the same trend that it did in the single-channel case: it places spokes at slightly higher spatial frequencies than the Fourier method. The discussion about the poor performance of the inversion method (see Sec. 5.4.3) is bolstered by its placement shown here: the dense $\mathbf{g}^{(\ell_2)}$ grid causes the inversion-based technique to tightly cluster its spokes around DC, and because of this, its resulting excitation completely lacks distinct edges and is only a "lowpass" version of the target in Fig. 5-8.

Multi-channel system validation. Simulation results are validated by playing the 15-spoke sparsity-enforced pulse on the system. The magnitude and phase of the center slice are shown in Fig. 5-14, corresponding with the simulated images in the right column of Fig. 5-13. The experimental and simulated magnitude images are quite similar: there is a dark ridge in both images where the two veins intersect, the left vein in both images has barely any curvature, and the right vein of each image exhibits slight curvature. The lower-left of the experimental image contains ghosts of the left vein, an artifact not present in the simulation. We believe this occurs because gradient infidelities cause a spoke to be slightly misplaced in (k_x, k_y); this misplaced spoke ends up modulating spatial frequencies

different from those intended, and because the system's coil profiles are being driven and superposed to cancel each other out in the majority of the FOX, this slight deviation perturbs the intended cancellation of the profiles and creates noticeable artifacts. In other words, commanding no flip in most of the FOX while asking for a 10-degree flip within two thin veins is an ill-conditioned problem.

The phase of the experimental image in Fig. 5-14 resembles the phase map predicted by the simulation in Fig. 5-13. On average, the phase in the left vein differs from that in the right vein by 81.6 degrees, 8.4 degrees off from the ideal 90-degree separation. The majority of this 8.4-degree error occurs where the two veins intersect. In the ideal target in Fig. 5-8, we see there is a sharp 90-degree phase cutoff between the left and right veins, whereas in the experimental map, there is a gradual change in phase in this region. This behavior is reasonable: it is unrealistic to think that the real system is capable of exciting a target with a discontinuity. Besides this difference, the observed magnitude and phase maps match closely with the simulated ones, lending credence to the simulation results and sparsity-enforced pulse design.

Sparsity-enforced placement: grid oversampling analysis. We have seen that the sparsity-enforced placement algorithm is superior to the other methods, yielding improvements in excitation quality with negligible changes to pulse duration and, at most, moderate increases in voltage. We now investigate how oversampling the grid of candidate spoke locations affects the sparsity-enforced algorithm's RMSE performance. This is done via simulations in the context of the single-channel system's inhomogeneity mitigation scenario. Here, for $T = 1, 3, \ldots, 29$, we run the sparsity-enforced method using the 17×17 Nyquist grid discussed earlier, along with a 33×33 2x-oversampled grid and a 49×49 3x-oversampled grid. These oversampled grids extend out to nearly identical maximum and minimum spatial frequencies in k-space as does the Nyquist grid, so the only difference among the grids is their oversampling factor. With other parameters held constant, we run the sparsity-enforced method with various λs for each grid, design T-spoke pulses, and compute RMSE. The sparsity-enforced method's runtime increases when provided the oversampled grids, because they have roughly 4x and 8x as many candidate locations than does the Nyquist grid. Fig. 5-15.A shows the results of this Bloch-simulated experiment. Here, RMSE as a function of oversampling factor is relatively constant, e.g., for $T < 7, T > 17$, oversampling the grid has no RMSE benefit. For fixed $T \in [7, 15]$, RMSE decreases with

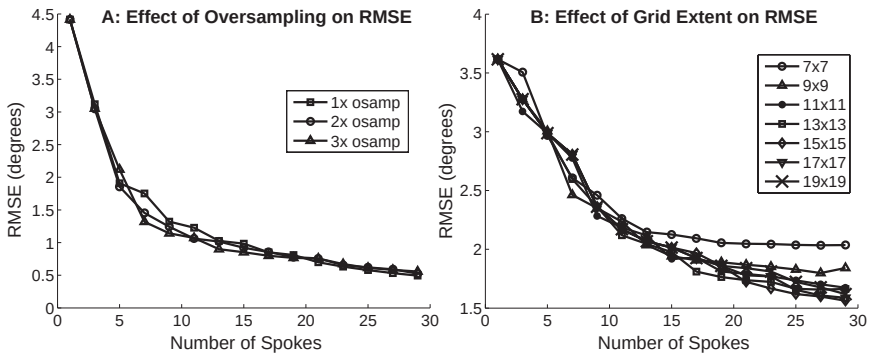

Figure 5-15: **Panel A, single-channel $|R||B_1^+|$ inhomogeneity mitigation experiment**: RMSE vs. number of spokes (T) when providing the sparsity-enforced method Nyquist-sampled, 2x-oversampled, and 3x-oversampled frequency grids of candidate locations. Grid oversampling seems to provide little RMSE benefit. **Panel B, eight-channel system, dual-vein target**: RMSE vs. T for Nyquist-sampled frequency grids of candidate spoke locations that are 7x7, 9x9, ..., 19x19 in size. Each grid is centered at DC and candidate points extent outward in k_x and k_y. The 11x11 grid yields reasonable RMSE results, suggesting that smaller grids may be used to decrease runtime.

increasing oversampling factor, but only slightly.

Sparsity-enforced placement: grid extent analysis. We now investigate whether increasing (reducing) the extent of the grid out to higher frequencies is of any benefit (detriment). Specifically, we alter the extent of the grid used in the multi-channel experiment and measure RMSE. We run the sparsity-enforced algorithm using the original Nyquist-spaced 17^2 grid, along with Nyquist-spaced grids that are $7^2, 9^2, \ldots, 15^2$, and 19^2 in size. In terms of runtime relative to the 17^2 grid, the sparsity-enforced algorithm runs 5.9, 3.6, 2.4, 1.7, and 1.3 times faster when the 7^2 through 15^2 grids are used, and 1.2 times slower when the 19^2 grid is used. For each grid, we sweep over λ (like in the earlier experiments) and then compute the RMSE of various T-spoke pulses for $T = 1, 3, \ldots, 29$. Fig. 5-15.B shows RMSE as a function of grid extent and T. Grids of size 11^2 and up yield relatively the same RMSE, which means that runtime may be reduced without a loss of performance by using smaller grids.

Based on the grid-extent results in Fig. 5-15.B and the oversampling results in Fig. 5-15.A, it is sufficient for our applications to simply pick a Nyquist-sampled grid that extends out a moderate rather than far distance into k-space. There is no need to extend the grid to high spatial frequencies or oversample it by even a factor of two. The strength of the sparsity-enforced algorithm does not come from placing spokes at high or finely-sampled

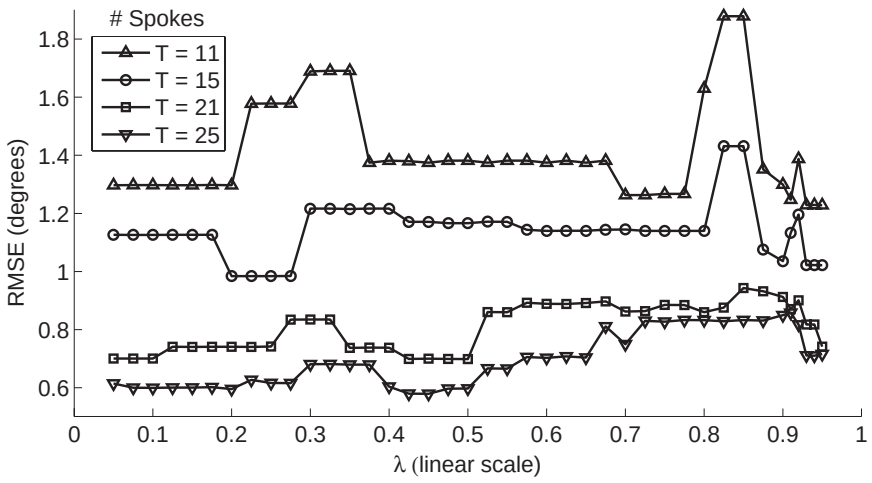

Figure 5-16: **Single-channel $|R||B_1^+|$ inhomogeneity mitigation experiment: λ sensitivity analysis**. The sparsity-enforced spoke placement algorithm's sensitivity to λ is analyzed when designing pulses comprised of 11, 15, 21, and 25 spokes. As T increases, sensitivity to λ decreases. Further, for fixed λ, RMSE does not exhibit a consistent trend across T.

frequencies, but in simply tuning placements outward from low frequencies and making slight—but certainly not obvious—alterations to the placements suggested by the Fourier method.

Sparsity-enforced placement: λ sensitivity analysis. The single-channel simulation results presented in Fig. 5-9 assumed that a good choice of λ was known. We now do away with this assumption, presenting RMSE as a function of λ and $T \in \{11, 15, 21, 25\}$ in Fig. 5-16. For small T, the choice of λ is crucial in order to best reduce RMSE, but as T increases, the algorithm's sensitivity to λ decreases significantly. Further, for fixed λ, RMSE does not exhibit a consistent trend across T.

Fig. 5-17 depicts the multi-channel experiment's sensitivity results, showing RMSE as a function of λ and $T \in \{11, 15, 21\}$. As T increases, the RMSE vs. λ curve moves smoothly downward. We are unsure why more erratic behavior occurs in the single-channel context (see Fig. 5-16).

Runtime. In the single-channel experiments where a 17^2 grid is used, it takes approximately 3 minutes to solve (5.13) using our MATLAB SeDuMi implementation on a Linux-based computer with a 3.0-GHz Pentium IV processor. In the eight-channel case with the 17^2 grid, runtime increases nearly linearly to 25 minutes. In general, the random-access memory footprint of the SOC program ranges from 200-700 megabytes.

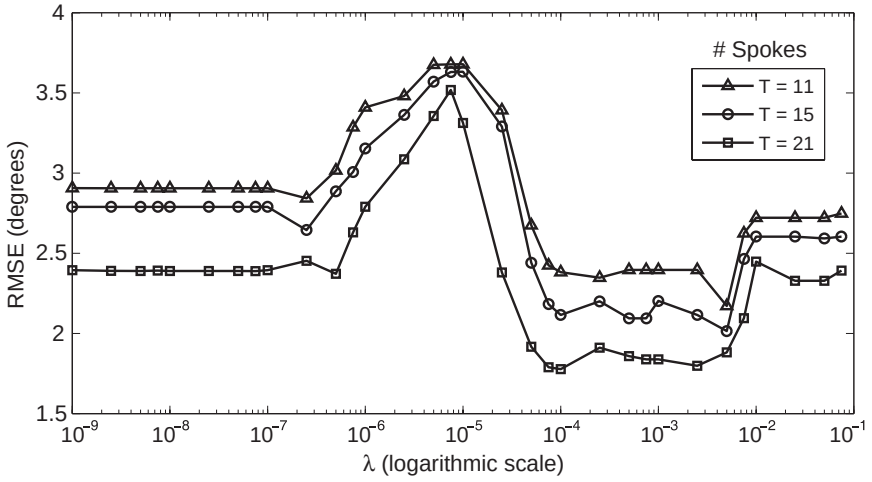

Figure 5-17: **Eight-channel system, dual-vein target: λ sensitivity analysis**. The sparsity-enforced spoke placement algorithm's sensitivity to λ is analyzed when designing pulses comprised of 11, 15, and 21 spokes. For fixed λ, RMSE decreases smoothly with T, in contrast with the behavior in Fig. 5-16.

5.4.7 Summary

The strength of sparsity-enforced spoke placement was demonstrated by designing fast, slice-selective RF pulses that mitigated $|R||B_1^+|$ inhomogeneity present in a head-shaped water phantom on a 7T single-channel system and that achieved a complex-valued target pattern using an eight-channel 3T parallel excitation system. In both cases, the sparsity-enforced method outperformed conventional methods, producing excitations with lower RMSE when pulse duration across the methods was fixed, and producing pulses with significantly shorter durations when excitation quality across the methods was fixed. The simulation results presented throughout this paper were validated by experiments on both the single-channel 7T system and the eight-channel 3T parallel excitation system and showed that sparsity-enforced pulses are applicable in real scenarios. Throughout both experiments, the sparsity-enforced algorithm automated the task of spoke placement and the design of the corresponding gradients and RF waveforms, yielding placement patterns that were not obvious and would be difficult or impossible to design by hand, freeing the designer from the task of slice-selective pulse design. The algorithm was shown to be highly robust to the choice of many of its input parameters, such as the extent and sample spacing of the grid of candidate spoke locations, and relatively robust to the choice of its sparsity-enforcing control parameter, λ, when larger numbers of spokes were placed during the single-channel

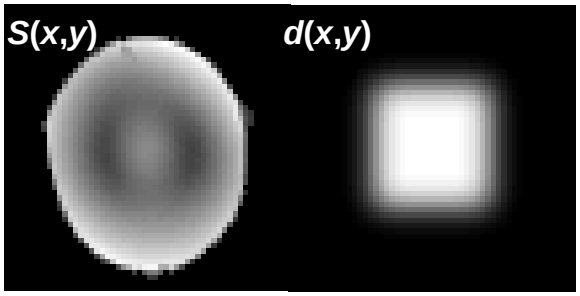

Figure 5-18: **Left: $|S(\mathbf{r})|$, map of in-plane $|B_1^+|$ inhomogeneity**. **Right: desired in-plane target excitation,** $d(\mathbf{r})$.

experiments and in general during the multi-channel experiments.

5.5 Joint Spiral-Trajectory RF Excitation Pulse Design

5.5.1 Overview

We now turn our attention towards the optimization of a 2-D spiral trajectory that attempts to form a smoothed box-shaped excitation in the presence of severe B_1^+ inhomogeneity present in a head-shaped water phantom at 7 Tesla on a single-channel system.

To jointly optimize spiral trajectories and pulses for use on single-channel systems, we define the J contours discussed in Sec. 5.2.1 as a set of *concentric rings* and then solve (5.5) to reveal a small subset of ellipses and corresponding energy weightings that form a high-fidelity version of the target excitation. The small, useful subset of rings out of the overall J-ring set is then connected into a spiral-like trajectory to yield a set of gradients, and the corresponding RF pulse is retuned as discussed in Sec. 5.2.1 by solving (2.11).

We focus on such an approach here and show that sparsity-enforced spirals lead to improvements in both excitation quality and pulse duration relative to conventional radially-undersampled ("accelerated") spirals.

5.5.2 Methods

Transmit profile and desired in-plane excitation pattern. Fig. 5-18 depicts the transmit profile, $S(\mathbf{r})$, of a birdcage coil in a head-shaped water phantom at 7 Tesla, along with the desired in-plane excitation pattern, $d(\mathbf{r})$. Forming $d(\mathbf{r})$ in the presence of the inhomogeneous $S(\mathbf{r})$ is non-trivial.

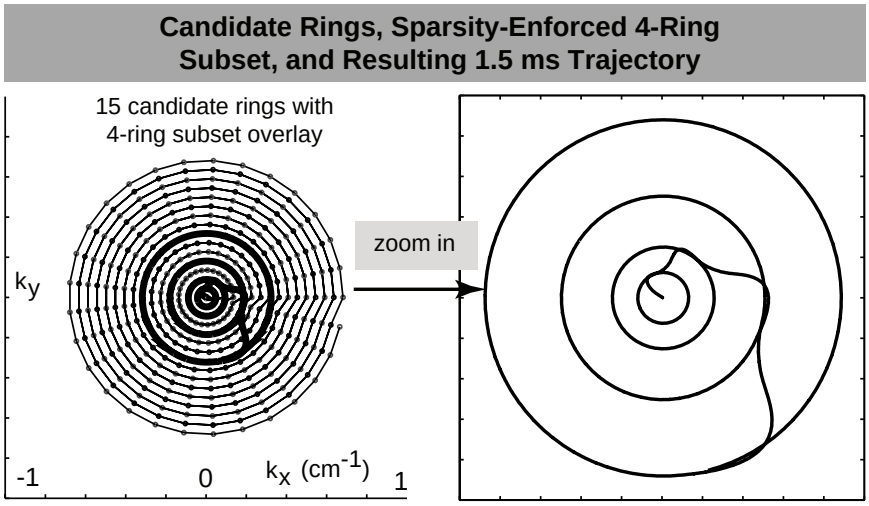

Figure 5-19: **Left: 15 concentric rings provided to the joint spiral-trajectory design algorithm and chosen 4-ring subset**. Fifteen rings, each comprised of 25 k-space points, along with a single point at the origin, are provided to the joint design algorithm as candidates. The four bold rings are those revealed by the algorithm. **Right: resulting 1.5-ms four-ring trajectory**. The trajectory is generated by connecting the four chosen rings using a greedy algorithm.

Conventional accelerated spirals. Consider a spiral whose radii are spaced at the Nyquist-limit (as defined by the desired FOV). Defining an unaccelerated Nyquist-sampled spiral an "$R = 1$" spiral, the conventional way to sparsify (accelerate) the trajectory is to radially undersample its rings by a factor of R. This process does indeed accelerate the trajectory and reduce pulse duration, but does not explicitly take the desired magnetization pattern into account.

Sparsity-enforced spiral design. By defining J concentric rings in 2-D k-space, the approach of Sec. 5.2.1 may be applied directly to yield a sparsity-enforced spiral-trajectory and accompanying RF excitationo pulse. Note that we could expand the algorithm's search space and ability to form excitations by providing the design algorithm additional ring-like contours in 2-D k-space such as asymmetrical ellipses.

5.5.3 Experimental Results

A sparsity-enforced spiral trajectory and RF excitation pulse are designed by first defining the $J = 16$ contours shown in the left panel of Fig. 5-19; there are 15 candidate rings, each comprised of 25 k-space points, along with a single point at DC (thus the parameter N_k discussed in Sec. 5.2.1 equals 376).

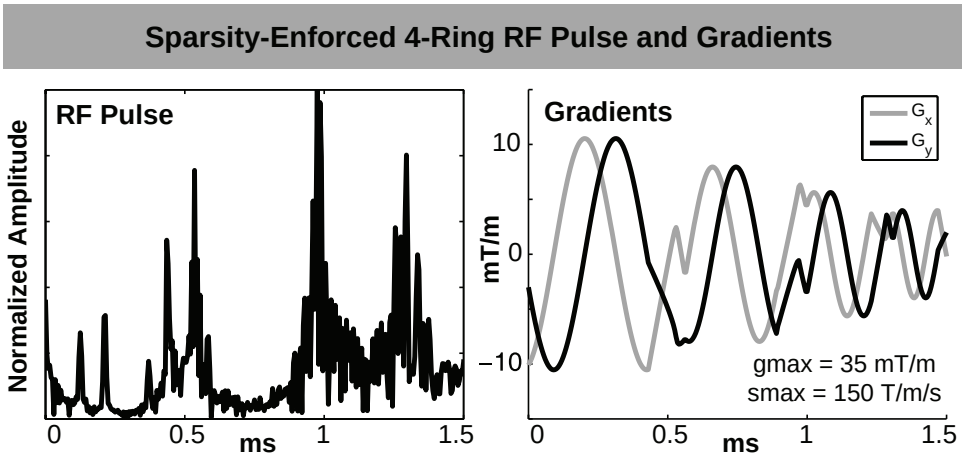

Figure 5-20: **Sparsity-enforced four-ring RF pulse and gradients**. The RF pulse plays along the trajectory given in the right panel of Fig. 5-19, and the gradients here trace the four-ring trajectory in 1.5 ms.

These contours, along with sampled versions of $S(\mathbf{r})$ and $d(\mathbf{r})$, are provided to the method, λ in (5.5) is set to 0.30, and (5.5) is then solved in under 2 minutes. Four rings are retained as the sparsity-enforced subset and appear as overlays on the left panel of Fig. 5-19. The right panel of Fig. 5-19 shows how the four-ring subset is connected into a 1.5-ms trajectory.

Figure 5-20 depicts the corresponding gradients and RF excitation pulse, the latter of which is generated via the steps outlined at the end of Sec. 5.2.1. Fig. 5-21 compares excitations due to the sparsity-enforced spiral and R-fold accelerated conventional spirals. The 1.5-ms sparsity-enforced spiral significantly outperforms the 4.9-ms, $R = 2$ spiral. The $R = 1$ spiral does indeed produce a near-perfect excitation, but is 6.4x longer than the optimized pulse. Note how the sparsity-enforced spiral traverses only a small segment of k-space yet is capable of forming a high-fidelity version of the box, in spite of the presence of the inhomogeneous $S(\mathbf{r})$. When $R = 7$, the conventionally-accelerated spiral has a duration on-par with that of the optimized spiral, but has four times larger error. Thus for fixed excitation quality, the sparsity-enforced spiral yields shorter pulses, and for fixed pulse duration, it yields lower-NRMSE excitations.

5.5.4 Summary

We have shown that the general single-channel joint trajectory-pulse design technique of Sec. 5.2.1 is indeed applicable to spiral trajectories. Here we showed that the approach

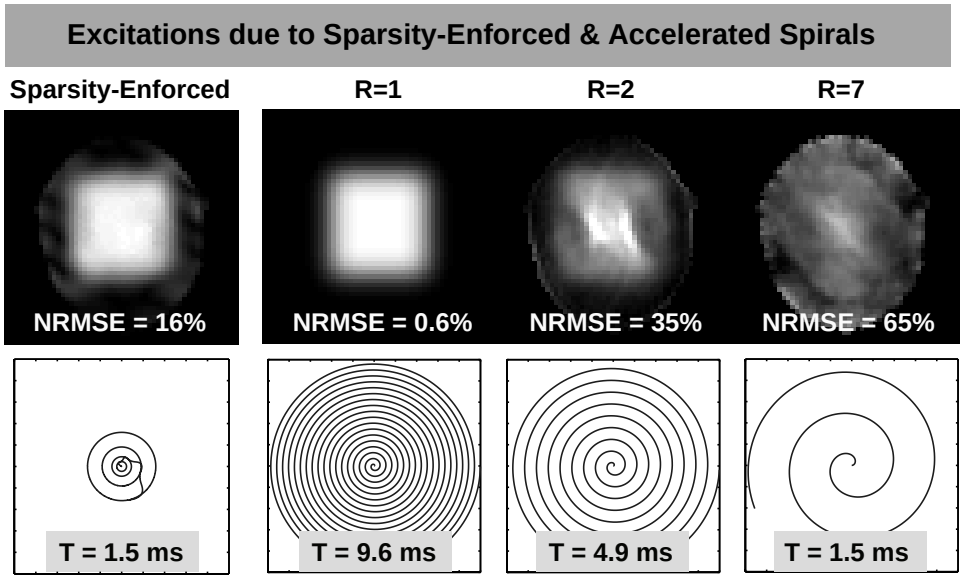

Figure 5-21: **Sparsity-enforced four-ring design vs. conventionally accelerated spiral designs**. Upper row: resulting in-plane excitations due to sparsity-enforced and $R = 1$, 2, and 7 accelerated spiral pulses. Lower row: corresponding trajectories with duration overlays. The sparsity-enforced trajectory is essentially a non-trivially-downsampled variable-density spiral.

was able to rapidly calculate fast, high-quality trajectories and corresponding RF pulses. Sparsity-enforced spiral trajectories significantly outperformed conventional spirals in simulated trials in a single-channel, 7T, non-uniform transmit profile setting.

5.6 Conclusion

Sparsity-enforced joint trajectory and pulse design was formulated generally and then applied to both spoke and spiral trajectories in single-channel and multi-channel contexts; in both B_1^+ mitigation and highly-structured excitation scenarios, the sparsity-enforced method yielded shorter pulses relative to conventional techniques when excitation quality was fixed, and higher-quality excitations than conventional techniques when pulse duration was fixed. To the best of our knowledge, our use of a sparsity-enforced optimization to jointly generate a target-pattern-specific, fast trajectory and corresponding RF waveform is a novel idea. The generality of the algorithm—in that it may apply to any set of trajectory segments and thus optimize over and generate a completely arbitrary trajectory—is, to the best of our knowledge, novel as well.

Chapter 6

Sparsity-Enforced Parallel Transmit Array Mode Subset Selection

6.1 Introduction

The purpose of this chapter is improve the flexibility of parallel excitation by optimizing the use of a multi-channel array's spatial profiles to improve the quality of accelerated spatially-tailored excitations. We confront an NP-Hard subset selection problem that arises when a parallel excitation system has more *transmit modes* available than hardware *transmit channels* with which to drive them. We will show the applicability of MSSO theory and propose a fast target-excitation-dependent sparsity-enforced mode subset selection (SEMSS) algorithm that explicitly accounts for the desired excitation pattern, $d(\mathbf{r})$, when choosing the mode subset, in contrast with conventional mode subset selection methods that only analyze the spatial profiles of the transmit modes and thus determine only a single mode subset for all desired excitations. Note that although we indeed make use of MSSO theory, the application here differs completely from Ch. 5's focus on joint pulse-trajectory design. The work of this chapter appeared at a conference [151] and also led to a pending patent [163].

Transmission ports. Let us consider a parallel transmission system, the basics of which were discussed in Sec. 2.1.11, whose physical hardware array consists of a vast number of input elements (ports). Attaching a port is of essentially no cost and a trivial task

for a hardware designer so this is a realistic scenario. In most cases, these elements are distributed radially about the array [3–6, 126, 140]. Let us assume P transmission channels available: each channel is simply a device that is capable of generating and transmitting a chosen waveform. In practice, one might simply connect each transmit channel to a single array element, spacing the channel-element connections equally around the coil array. This configuration style was used to generate the spatial profiles given in Fig. 2-6: the high-SNR regions of the eight spatial profiles are radially spaced by roughly 45 degrees because the transmit channels were connected in a radial fashion to ports spaced in 45 degree increments about the coil array.

Transmission modes. There is a more sophisticated approach to connecting a single transmit channel to the multi-port array, however. Specifically, rather than attaching a transmit channel directly to one array port, one might consider distributing the transmit channel waveform into many array ports simultaneously via the use of inexpensive splitter hardware, with the option of altering the magnitude and phase of the waveform entering each port via the use of inexpensive attenuators and phase shifters,[1] respectively. In other words, one may form *linear combinations* of the analog outputs of multiple ports using only one transmission channel, generating a single spatial profile—a *transmit mode*—that is not necessarily localized about a single array port as is each individual profile in Fig. 2-6. It is simple and inexpensive to configure an array to produce a transmit mode, which means that many different transmit modes may be made available. One such example of a multi-mode array is a *Butler matrix*. This device was first applied to electronically scanned antenna arrays by Butler and Lowe [19], but its use in an MRI excitation context has recently been validated; specifically, it has been confirmed that a Butler matrix is capable of producing a useful basis set of spatial transmit modes [5]. Notationally, let us define the N spatial transmit mode profiles of a given system as $S_1(\mathbf{r}), \ldots, S_N(\mathbf{r})$, where $\mathbf{r}$ is a spatial variable.

Transmission modes vs. transmission channels. Unfortunately, in direct contrast to the ease with which one may add an additional transmit mode is the cost and implementation difficulties one faces when attempting to add an additional transmit channel because each such channel (capable of driving only a single transmit mode profile) requires a separate digital RF generator, a high-power RF amplifier, and a SAR safety monitor. As a reflection of this, only up to eight-transmit-channel systems have been developed to

[1] Phase shifters are also called "delays".

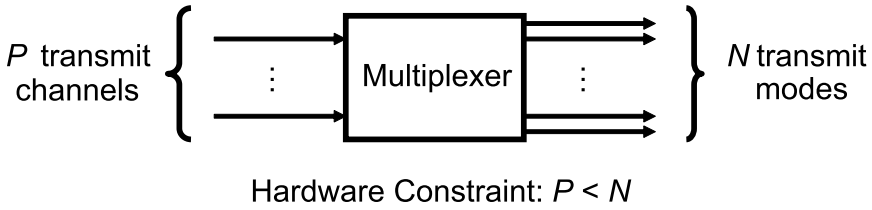

Figure 6-1: N**-mode,** P**-channel parallel excitation system**. Each transmit channel is capable of driving one transmit mode. Transmission modes are easy to construct and thus N is large. Transmission channels, however, are costly and complex to implement, and thus P is small. Given a user-desired target excitation, $d(\mathbf{r})$, which P-mode subset should one use to form the excitation? There are N-choose-P such subsets to choose from, but a brute-force search is infeasible for near-real-time pulse design.

date [7, 58, 113].

Our problem. In the interest of studying the above problem, we pose a case of practical interest: given a transmit array with $N = 16$ available transmit mode profiles[2] and a sytem with only $P = 8$ transmit channels, choose an advantageous eight-mode subset of these 16 modes to best produce a desired excitation pattern, $d(\mathbf{r})$.

Conventional mode subset selection techniques. One approach to choosing a P-mode subset is to choose the highest-energy modes among the N available (the "brightest" modes); another approach that is valid if $N = 2P$ and a Butler matrix is used is to select solely those P modes that are circularly-polarized (CP) (which are theoretically able to strongly tailor the effective magnetic field) and forego use of the P anti-circularly-polarized (ACP) modes (which theoretically produce no signal and are incapable of influencing the excitation). A full discussion of circular vs. anticircular polarization may be found in [72] and is beyond the scope of this dissertation. These are general approaches that rely on intuition and theory, respectively, but we see that they have no concept of the excitation the user wants to create and thus no ability to exploit this when choosing a subset of modes.

An alternative approach. Our approach differs substantially: we will take the user's desired excitation target pattern into account when searching for a subset of modes to best form the desired excitation. One might imagine that the best subset of modes that excites one desired pattern may differ from the best subset that excites a radically-different desired pattern. In short, different $d(\mathbf{r})$s may call for different subsets of modes. The

[2] In this chapter we do not focus on the *design* of the N transmit modes; rather, we simply assume a hardware array designer provides us an N-mode device. The design of such modes is itself an open and ongoing problem.

globally optimal way to solve this problem is to perform a brute-force search over all N-choose-P mode subsets of size P to find the best such subset; this technique, however, is incredibly expensive from a computational and runtime standpoint and infeasible for use in near-real-time pulse design scenarios. Thus, we instead propose a fast target-dependent sparsity-enforced subset selection (SEMSS) algorithm that explicitly accounts for the desired excitation pattern when choosing the mode subset.

Experiments. In this chapter, we use SEMSS to determine 8-mode subsets of a 16 mode array for use on an 8-channel parallel transmission system at 7 Tesla when forming slice-selective spatially-tailored excitations with uniform and ring-shaped target patterns in a water phantom. Excitations using the SEMSS subsets are compared to those due to bright-mode and circularly-polarized mode subsets. Brute-force search finds that in this experiment, SEMSS actually finds the best of all 12,871 possible subsets for the uniform excitation, and a near-optimal subset for the ring-shaped one. Our results confirm the hypothesis that the best mode subset may vary substantially depending on the desired excitation pattern.

Overview. In Sec. 6.2 we state the formal problem and outline pulse design for a fixed P-mode subset in Sec. 6.3. In Sec. 6.4 we discuss conventional mode subset selection methods and pose MSSO-inspired Sparsity-Enforced Mode Subset Selection (SEMSS) in Sec. 6.5. The $N = 16$-mode array and the two slice-selective excitation experiments are outlined in Sec. 6.6. Results are presented and discussed in Sec. 6.7. Concluding remarks appear in Sec. 6.8

6.2 Problem Statement and Assumptions

Given a uniform main field B_0, N available transmit mode spatial profiles $S_1(\mathbf{r}), \ldots, S_N(\mathbf{r})$, a desired excitation pattern $d(\mathbf{r})$, a given field-of-excitation (FOX) indexed by $\mathbf{r}$, a fixed set of gradients $\mathbf{G}(t)$ of duration L, and a fixed spoke-based k-space trajectory $\mathbf{k}(t)$ (for $t \in [0, L]$), which $P < N$ transmit mode spatial profiles should be chosen to best excite the pattern $d(\mathbf{r})$?

Discretization. Assume space has been sampled at N_s in-plane (x, y) locations, denoted $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$, and time has been sampled at N_t instants $t_1, \ldots, t_{N_t} \in [0, L]$ spaced uniformly by Δ_t.

Spokes trajectory. The slice-selective trajectory consists of T spokes in k_z placed in the (k_x, k_y) plane [108, 113, 132]. To simplify the design process, we fix the placement of T spokes, along with spoke shape and gradients. These constraints fix the shape of the P RF pulses and all properties of the slice profile (e.g., sidelobes and sharpness) [108, 113], while still letting the user retain control over the amplitude and phase each channel encodes along each spoke. For a T-spoke pulse and a given P-mode subset, these constraints cause major changes to (2.14, 2.15) given in Sec. 2.1.11. First, all but T columns of $\mathbf{F}$ are discarded, since only T locations in (k_x, k_y)-space (those where the spokes are located) are available for weighting. Further, each $\mathbf{b}_p$ reduces from N_t to T elements. Finally, the FOX reduces to a 2-D in-plane region, leading to a far smaller N_s than in a 3-D case. Users are then free to shape the in-plane (x, y) excitation pattern that arises by choosing the complex-valued weights that each of the P driven transmit modes deposits at each of the T locations (i.e., users may choose the amplitude and phase each of the P driven transmit modes will encode along each of the T spokes).

6.3 Parallel Excitation Pulse Design after Choosing a Mode Subset

Assume we have chosen a P-mode subset $S_{q_1}(\mathbf{r}), \ldots, S_{q_P}(\mathbf{r})$ among the N modes. (The q_ps uniquely index the overall set of available modes, $\{1, \ldots, N\}$.) To design the overall pulse we simply need to determine the PT weights the P driven modes will encode along the T in-plane k-space locations $\mathbf{k}_1, \ldots, \mathbf{k}_T$.

We first construct $\mathbf{S}_1, \ldots, \mathbf{S}_P$, where $\mathbf{S}_p$ is an $N_s \times N_s$ diagonal matrix comprised of samples of $S_{q_p}(\mathbf{r})$ taken within the in-plane 2-D FOX. We then construct the $N_s \times T$ matrix, $\mathbf{F}$, where $\mathbf{F}(m, t) = j\gamma \mathrm{M}_0 \Delta_t e^{j\mathbf{r}_m \cdot \mathbf{k}_t}$. Let $\mathbf{g}_p$ be a T-element vector representing the T weights the pth driven mode places at the corresponding T spoke locations, and $m(\mathbf{r}_m)$ be the resulting excitation that forms at $\mathbf{r}_m$ due to the P driven modes depositing energy at the T locations. Based on (2.13) and analogously to (2.14), we arrive at a system of linear equations:

$$\mathbf{m} = \mathbf{S}_1 \mathbf{F} \mathbf{g}_1 + \cdots + \mathbf{S}_P \mathbf{F} \mathbf{g}_P = \mathbf{A} \mathbf{g}_{\mathrm{tot}}, \tag{6.1}$$

where $\mathbf{m}$ is an N_s-element vector whose mth element equals $m(\mathbf{r}_m)$.

We then sample the desired excitation $d(\mathbf{r})$ at the same N_s $\mathbf{r}_m$s and stack these samples into the vector $\mathbf{d}$. Finally, analogously to (2.15), we solve

$$\mathbf{d} = \mathbf{A}\mathbf{g}_{\text{tot}} \tag{6.2}$$

for $\mathbf{g}_{\text{tot}}$, yielding the PT desired weights. This finalizes the pulse design and in the process yields the lowest ℓ_2-residual-error in-plane excitation possible given the P chosen modes and T fixed spoke locations. In this chapter, because $\mathbf{A}$ always consists of $PT \leq 90$ columns, (6.2) is well-conditioned and may be solved using any method (e.g., SVD-based inversion [51, 118], LSQR [100, 101], etc.)

6.4 Conventional Subset Selection Methods

Below we summarize two approaches with negligible computational cost for selecting P modes independently of $d(\mathbf{r})$, as well as the brute-force search method that indeed finds the optimal P-mode subset that produces $d(\mathbf{r})$, but whose use in real design scenarios is infeasible due to runtime constraints.

6.4.1 Circularly-Polarized Modes

These modes simply arise from the physical construction and use of the Butler matrix in the scenario where $N = 2P$. The modes are known based on the construction of the device [19].

6.4.2 Highest-Energy ("Brightest") Modes

Here we choose the P modes whose ℓ_2 energies within the specified FOX are largest in magnitude. Specifically, given the $S_n(\mathbf{r})$s, we first sample each at the N_s sampling locations $\mathbf{r}_1, \ldots, \mathbf{r}_M$ and then construct $\mathbf{s}_n = [S_n(\mathbf{r}_1), \ldots, S_n(\mathbf{r}_M)]^{\mathrm{T}}$ for $n = 1, \ldots, N$. The P-mode subset is then simply comprised of those P modes whose $\|\mathbf{s}_n\|_2$s are largest.

6.4.3 Brute-Force Search over N-Choose-P Modes

For each of the N-choose-P total possible P-mode subset choices, design a pulse according to the steps outlined in Sec. 6.3 and (6.1, 6.2), remembering the resulting residual error $\|\mathbf{d} - \mathbf{A}\mathbf{g}_{\text{tot}}\|_2$ for each case. The globally-optimal P-mode subset is then simply the subset

that yields the lowest ℓ_2 residual error. In case of a tie (i.e., two or more subsets producing equally-good excitations), any of the lowest-error subsets may be chosen.

6.5 MSSO-Inspired Target-Specific Transmit Mode Subset Selection

SEMSS is an optimization that designs a set of waveforms to drive the N (rather than P) modes to form the target pattern, but imposes a strong penalty whenever a mode waveform becomes nonzero, and in this way seeks out a P-mode subset among the N available modes. Essentially, SEMSS enforces sparsity on the usage of modes and thus reveals a small subset of good modes and corresponding waveforms with which to drive each mode. For the purposes of this chapter, here we derive SEMSS for T-spoke trajectories under the assumptions of Sec. 6.2, but SEMSS extends readily to other trajectories.

To begin, assume that all N modes may be driven. The equations that relate the pulses used to drive each mode to the resulting magnetization $m(\mathbf{r})$ are linearized using the approach of Sec. 6.3, except here we construct N rather than P matrices $\mathbf{S}_n, \ldots, \mathbf{S}_N$ and N rather than P vectors, $\mathbf{g}_1, \ldots, \mathbf{g}_N$. This yields

$$\mathbf{m} = \mathbf{S}_1\mathbf{F}\mathbf{g}_1 + \cdots + \mathbf{S}_N\mathbf{F}\mathbf{g}_N = \mathbf{A}\mathbf{g}_{\text{tot}}. \tag{6.3}$$

Recall that $\mathbf{g}_n$ contains the T weights the nth mode will encode at the T spoke locations to influence the samples of the resulting in-plane pattern contained in $\mathbf{m}$.

The desired excitation, denoted $d(\mathbf{r})$, is again vectorized into $\mathbf{d}$. To determine weightings for each mode to place at each spoke location, one may consider solving

$$\mathbf{d} = \mathbf{A}\mathbf{g}_{\text{tot}} \tag{6.4}$$

in the minimal-ℓ_2-error sense. Such an approach will indeed reveal weightings that form an excitation quite close to the desired one, but it results in a solution $\mathbf{g}_{\text{tot}}$ where all N modes deposit energy, i.e., all $\|\mathbf{g}_n\|_2$s will be nonzero. This fails to reveal a useful small mode subset.

Consider however (approximately) solving (6.4) while penalizing each nonzero mode energy (each $\|\mathbf{g}_n\|_2$), explicitly prohibiting the use of many modes, while encouraging those

that do remain in use to have $\mathbf{g}_n$s that still approximately solve (6.4). This could be accomplished by solving

$$\min_{\mathbf{g}_{\text{tot}}} \|\mathbf{d} - \mathbf{A}\mathbf{g}_{\text{tot}}\|_2^2 \text{ s.t. the use of only } P \text{ of the } \mathbf{g}_n\text{s.} \tag{6.5}$$

Upon close inspection, one sees that (6.5) is simply the brute-force method outlined in Sec. 6.4, except here it has been stated as a formal optimization problem. Upon even closer inspection, however, one sees that (6.5) is *identical* to the alternate formulation of the NP-Hard MSSO problem, (3.7), given in Sec. 3.2.2.[3]

Thus, just as in the alternate formulation of MSSO given in Sec. 3.2.2, we may relax (6.5) into the following tractable convex minimization problem:

$$\min_{\mathbf{g}_{\text{tot}}} \left\{ \frac{1}{2}\|\mathbf{d} - \mathbf{A}\mathbf{g}_{\text{tot}}\|_2^2 + \lambda \sum_{n=1}^{N} \|\mathbf{g}_n\|_2 \right\}. \tag{6.6}$$

For a fixed λ, such a norm encourages a sparse usage of the $\mathbf{g}_n$s. As λ is increased, increasing numbers of modes have their $\|\mathbf{g}_n\|_2$ energies driven to zero, residual excitation error increases, and smaller subsets of modes (and corresponding weightings) are revealed. SEMSS differs from the sparsity-enforced joint trajectory-pulse design methods posed in Ch. 5 because here we are enforcing sparsity on transmit mode usage rather than on the physical locations of spokes. In this chapter, we alter (6.6) into a second-order-cone program as outlined in Sec. 3.3.7 and implement the algorithm in SeDuMi [120].

After solving (6.6) using a sufficiently-large λ, the P modes whose $\|\mathbf{g}_n\|_2$ energies are largest are chosen as the optimized P-mode subset. The weights within the P $\mathbf{g}_n$s corresponding to the P chosen modes are retuned by truncating the $N - P$ unused $\mathbf{S}_n$s and $\mathbf{g}_n$s from $\mathbf{A}$ and $\mathbf{g}_{\text{tot}}$ in (6.3, 6.4) and finalizing the design as outlined in Sec. 6.3.

6.6 Experiments

We now study the performance of the four mode subset selection algorithms in the context of simulated slice-selective excitations within a water phantom at 7T where there are $N = 16$ transmit modes but only $P = 8$ transmit channels.

[3]The variable names in (6.5) differ slightly from those of (3.7) but the optimizations are indeed identical.

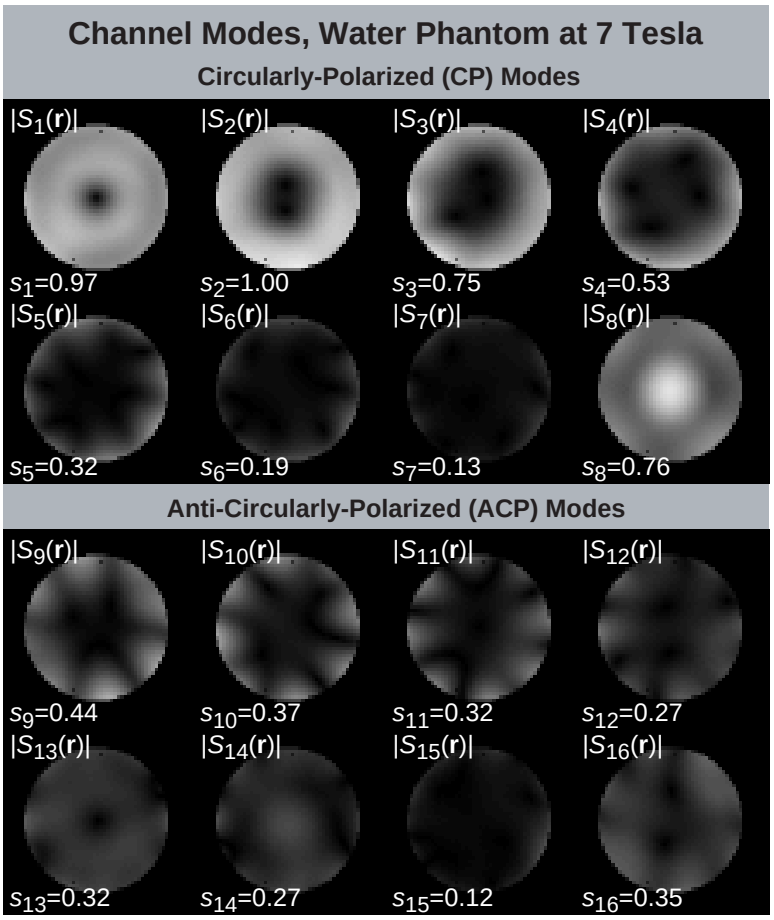

Figure 6-2: $N = 16$ **transmit modes of a Butler matrix in a water phantom at 7 Tesla**. For mode n, the in-plane mode magnitude, $|S_n(\mathbf{r})|$, is depicted and the normalized ℓ_2 mode energy, $s_n = \|\mathbf{s}_n\|_2$, is also provided.

6.6.1 $N = 16$ Butler Matrix Transmit Modes in a Water Phantom at 7T

Fig. 6-2 depicts the magnitudes of the $N = 16$ transmit modes, $S_n(\mathbf{r})$, of a stripline array in an orthogonal birdcage (BC) basis configuration, as obtained via a Butler matrix [19]. Modes include those with theoretically correct polarization for excitation (CP modes), as well as those with the opposite polarization (ACP modes). The in-plane, in-FOX ℓ_2 energy of each mode, $s_n = \|\mathbf{s}_n\|_2$, is calculated as described in Sec. 6.4 and overlaid onto the figure. The profiles are scaled by a constant so the largest magnitude across all map pixels equals unity. This scaling makes the mode maps "qualitative" in the low flip angle domain; they do not convey the exact flip angle achieved, but this is acceptable from a simulation standpoint.

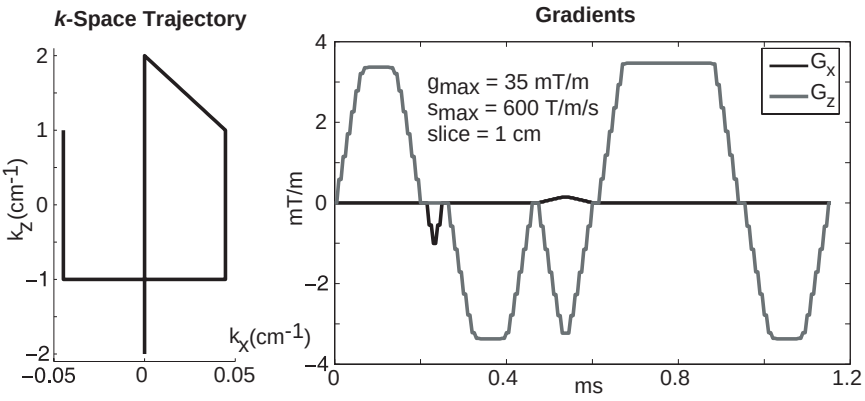

Figure 6-3: **Uniform excitation experiment: 3-spoke trajectory and gradients**. Both $\mathbf{k}(t)$ and $\mathbf{G}(t)$ are depicted. Each mode subset selection algorithm will pick $P = 8$ of the $N = 16$ modes given in Fig. 6-2 and then use each mode to place complex weights at each (k_x, k_y) spoke location to form uniformly slice-selective excitation.

6.6.2 Slice-Selective Uniform Excitation: Three-Spoke Trajectory, Eight Modes

For the first experiment, $d(\mathbf{r})$ is a uniform zero-phase excitation and a 1.2-ms three-spoke trajectory is used; the fixed $\mathbf{k}(t)$ and $\mathbf{G}(t)$ are illustrated in Fig. 6-3. The slice thickness of the trajectory is fixed at 10 mm and a Hanning-windowed sinc is used along each spoke. The spoke at $(k_x, k_y) = (0, 0)$ has a time-bandwidth product of 4, while the two off-DC spokes have k_z-lengths half that of the former. (Using shorter off-DC spokes lets one reduce the duration of a pulse without noticeably impacting slice selection performance [113].) Gradient amplitude and slew are constrained to 35 mT/m and 600 T/m/s, respectively.

The first three mode subset selection techniques are straightforward to implement. Regarding SEMSS, we sweep over twenty values of λ, solving (6.6) each time, and remembering the twenty P-mode subsets; the best subset of these twenty is retained. This sweep over λ is extremely fast and worthwhile: each individual run of (6.6) involves finding only $NT = 48$ complex-valued unknowns and takes only a matter of seconds. It allows us to avoid the intractable brute-force approach of evaluating all possible subsets and yet still conduct an exploration of the solution space in under a minute. The brute-force approach takes roughly an hour.

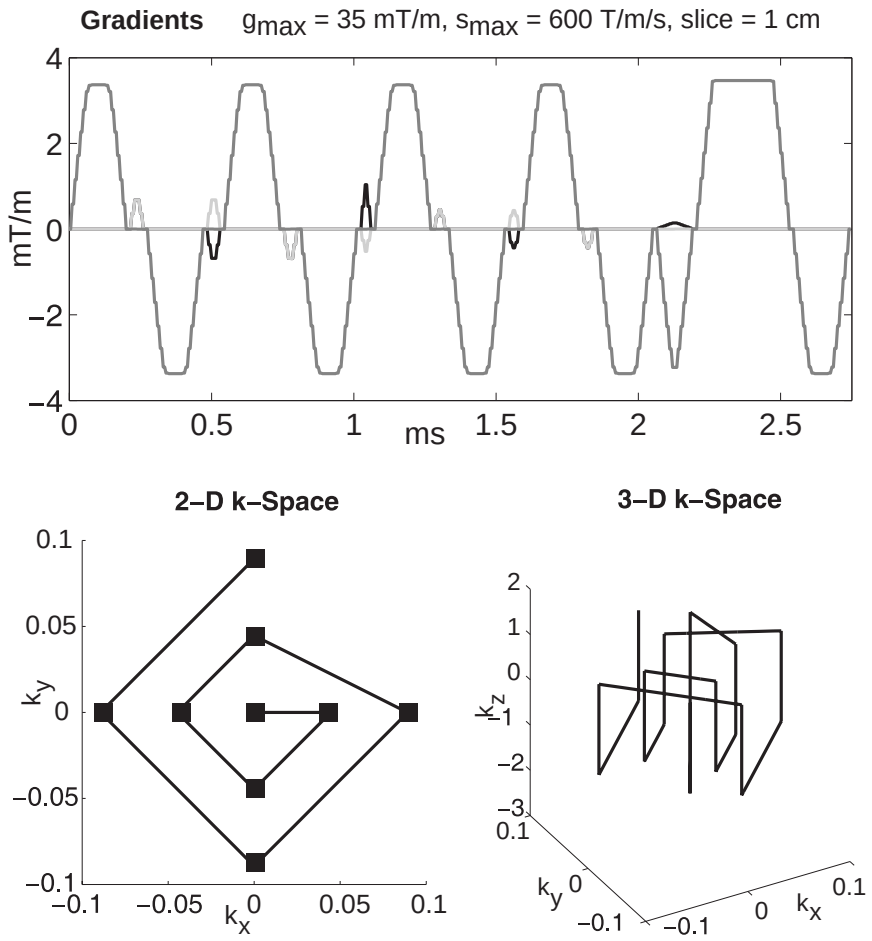

Figure 6-4: **Ring excitation experiment: 9-spoke trajectory and gradients**. Each mode subset selection algorithm will pick $P = 8$ of the $N = 16$ modes given in Fig. 6-2 and then use each mode to place complex weights at each (k_x, k_y) spoke location to form a slice-selective excitation whose in-plane pattern resembles a ring.

6.6.3 Slice-Selective Ring Excitation: Nine-Spoke Trajectory, Eight Modes

The second experiment involves exciting not only a thin slice, but an in-plane spatially-selective ring as well. Here we will use a nine-spoke trajectory to give the system more degrees of freedom in k-space with which to tailor the in-plane excitation. This trajectory and the accompanying gradients are given in Fig. 6-4; all spoke and SEMSS parameters equal those used in the uniform excitation experiment, except here $NT = 144$; The runtimes of SEMSS and brute-force search equal two minutes and three hours, respectively.

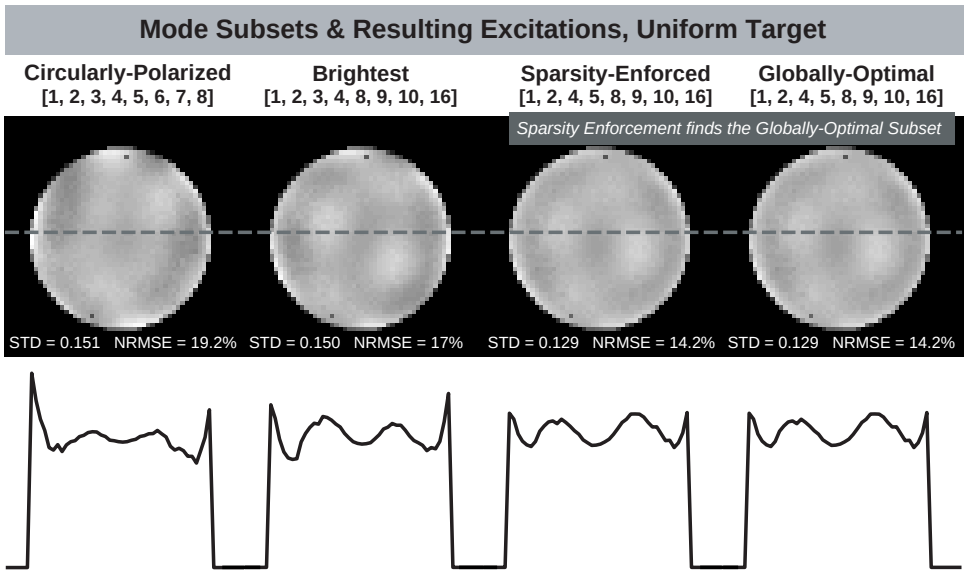

Figure 6-5: **Uniform excitation experiment: mode subset selection results**. The mode subsets selected by the CP, bright modes, globally-optimal brute-force, and sparsity-enforced SEMSS methods are listed and resulting in-plane excitations are displayed along with standard deviation (STD) and error overlays. CP produces the worst-quality excitation. In this specific case, SEMSS determines the globally-optimal subset in under a minute (vs. one hour via brute-force search).

6.7 Results and Discussion

6.7.1 Slice-Selective Uniform Excitation Results

The eight modes chosen by each of the four subset selection algorithms, along with images of the resulting Bloch-simulated in-plane excitations with in-plane standard deviation (STD) and error metric overlays, are provided in Fig. 6-5.[4]

In this case, the reasonable choice of CP modes produces the worst-quality excitation. Choosing the 8 bright modes improves NRMSE by a factor of 1.13, but fails to reduce STD. The SEMSS modes ($S_p(\mathbf{r}), p \in \{1, 2, 4, 5, 8, 9, 10, 16\}$) produce a noticeably more uniform excitation; STD and NRMSE improve by factors of 1.16 and 1.20 relative to the bright mode result, respectively. The bright subset makes use of $S_3(\mathbf{r})$ (of ℓ_2-energy 0.75), whereas SEMSS opts to use $S_5(\mathbf{r})$ (of ℓ_2-energy 0.32): thus we observe that using a low-energy "dark mode" in lieu of a bright one results in a better excitation. Brute-force search over all mode subsets shows that SEMSS determines the best subset among all 16-choose-8

[4]Normalized root-mean-square errors (NRMSEs) rather than ℓ_2 residual errors are provided for easier understanding; this is permissible because NRMSE always equals ℓ_2 error to within a known constant.

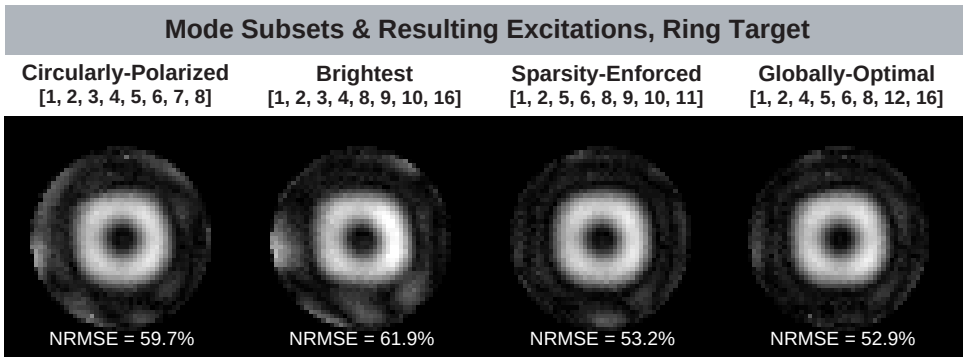

Figure 6-6: **Ring excitation experiment: mode subset selection results**. Formatting here is analogous to that of Fig. 6-5. Here, the bright-mode subset yields the worst-quality excitation. SEMSS finds a subset in two minutes that yields an excitation extremely close in quality to the globally-optimal one determined via a three-hour brute-force search.

(12,871) possible subsets, proving that it is able to rapidly determine target-specific subsets and outperform conventional techniques.

6.7.2 Slice-Selective Ring Excitation Results

The slice-selective ring excitation results are depicted in Fig. 6-6; formatting here is analogous to that of Fig. 6-5. In this case, choosing the bright modes yields the worst excitation. The SEMSS modes in this case are $S_p(\mathbf{r}), p \in \{1, 2, 5, 6, 8, 9, 10, 11\}$. Using these modes leads to an excitation that is 1.12 times better in quality than those produced by the CP-mode and bright-mode subsets. In this case, SEMSS's subset differs from the globally-optimal one determined via brute-force search, but leads to essentially the same result: the globally-optimal excitation has only 1.006 times lower NRMSE. As hypothesized, the optimal and SEMSS subsets are not equivalent to those of the uniform case given in Fig. 6-5: this proves that different targets call for the use of different modes. Furthermore, we see here that dark modes are present in both the SEMSS and globally-optimal solutions, e.g., each chooses $S_6(\mathbf{r})$, the second darkest of all 16 available modes.

6.8 Conclusion

By applying multiple-system single-output simultaneous sparsity theory, we derived a sparsity-enforced mode subset selection (SEMSS) algorithm that determines surprising target-specific mixtures of light and dark modes, providing increases in excitation quality relative to the

usage of CP and bright modes.

To the best of our knowledge, our use of a sparsity-enforced optimization procedure and incorporation of target profile information in order to determine a useful mode subset are novel ideas.

Chapter 7

Specific Absorption Rate Studies of the Parallel Transmission of Inner-Volume Excitations at 7 Tesla

7.1 Introduction

Our goal in this chapter is to investigate the behavior of whole-head and local N-gram (Ng) specific absorption rate (SAR) as a function of trajectory acceleration factor and target excitation pattern due to the parallel transmission (pTX) of spatially-tailored excitations at 7 Tesla (7T), especially the potential for a relatively high ratio of local SAR to average SAR. Along the way, we also propose a fast, low-memory algorithm for computing local SAR that enables us to analyze hundreds of candidate pTX pulses. The results presented here have been published in [152, 153, 156].

As outlined in Sec. 1.2.2 and Sec. 2.1.11, parallel transmission systems are promising devices because they enable one to accelerate through k-space and reduce the duration of an RF pulse even after one has exhausted the ability to do so by increasing the amplitude and slew rates of the system's gradient coils [58, 79, 113, 130, 168]. Unfortunately, as detailed in Sec. 1.2.3, parallel transmission poses two major SAR concerns: first, concurrently driving multiple transmit channels causes the electric fields generated by each channel to undergo

local superposition, which may produce local extremes in electric field magnitude [168] and thus spikes in local SAR; second, accelerating a k-space trajectory increases the peak power of the corresponding RF pulse and may significantly increase SAR [56, 78, 81, 141].

In this chapter, we investigate the above concerns and test if global and local SAR obey intuitive scaling rules as a function of excitation k-space trajectory acceleration factor and the parameters of an inner-volume excitation pattern excited by 2-D spatially-tailored RF excitation pulses. Our intent is to give pulse designers, coil array engineers, and RF safety researchers insight into the SAR characteristics of a high-field pTX system. Specifically, we study whole-head and maximum local 1 gram (1g) and 10 gram (10g) SAR in a multi-tissue head model during constant-fidelity excitation of 2-D boxes. Box-shaped excitations are useful because they allow one to exclude moving tissues from a volume undergoing imaging [47] and also permit reduced field-of-view (FOV) imaging, which improves temporal resolution without compromising spatial resolution [66]. Such excitations have applications to echo-volumnar imaging [135, 143] as well as perfusion territory mapping via arterial spin labeling [35].

The fidelity of each box-shaped excitation is kept constant as measured by normalized root-mean-square error (NRMSE) with respect to a target pattern. The simulations are based on an eight-channel parallel transmission system at 7T; this field strength is chosen because many researchers are focusing on using pTX to mitigate B_1^+ inhomogeneity [16] occurring at 7T. Local 1g and 10g SAR are studied in addition to whole-head mean SAR because these correspond to limits specified by the Food and Drug Administration (FDA) [21] and the International Electrotechnical Commission (IEC) [70, 71], and will permit observation of local SAR hot spots.

The study uses finite-difference time domain (FDTD) simulations [85] in a multi-tissue high-resolution head model as described by Angelone et al. [10], obtaining electric and B_1^+ fields generated by each array element when its corresponding channel is driven by a unit current, as described by Angelone et al. [8]. The B_1^+ maps are then used to design a variety of pTX box-shaped excitations, the latter of which are validated with Bloch-equation simulations. Finally, the electric field maps are used to calculate whole-head and maximum local 1g and 10g SAR due to each pTX pulse as a function of target flip angle, position, size, smoothness, and orientation, as well as spiral trajectory undersampling (acceleration) factor, R. The results demonstrate a wide range of SAR values that arise due to tailored

excitations that produce similar spatial patterns. This data, together with the non-intuitive behavior SAR exhibits as a function of the excitation parameters, suggests that explicit SAR calculations will likely be needed on a per-pulse basis, even for relatively minor variations in RF pulse properties, in order to ensure that local SAR values meet regulatory criteria.

7.2 Materials and Methods

7.2.1 Human Head and Eight-Channel Parallel Transmit Array Model

Head model. SAR characteristics of excitation pulses transmitted through an eight-channel head array are calculated in a high-resolution ($1 \times 1 \times 2$ mm^3) 29-tissue human head model, the latter of which is obtained via segmentation of anatomical MRI data [10]. Each of the tissues in the model is assigned a density, ρ (kg/m^3), and electrical conductivity, σ (S/m), using Federal Communications Commission data.[1] Overall, the model consists of roughly 2.5 million Yee cells [144].

Excitation array. The pTX system is modeled by placing eight copper circular loop elements at 45-degree increments along a 25-cm-diameter cylindrical surface, the latter of which is centered on the head, as described in [8]. Each loop element is overlapped to null mutual inductance with its neighbors and has a diameter of 15 cm, a trace width of 1 cm, and an input resistance of 50 Ω. Fig. 7-1's left subplot provides a snapshot of the head and transmit array, while the right subplot depicts a transverse slice through the head about $z = 0$, illustrating the variety of tissues and realism of the model.

Model resolution. The high-resolution model in this study meets or exceeds the requirements for accurate SAR calculation in volumes as small as 1g because each of its voxels contains no more than 0.0037 grams of tissue (every possible 1g region consists of at least 270 cells). Additionally, the model's fine resolution mitigates staircasing artifacts [20] that adversely impact local field and local SAR calculations because the model's ∼1000 cells per 300-MHz wavelength should reduce staircase error to less than one decibel [63].

7.2.2 Electromagnetic Field Simulations and B_1^+ Field Map Generation

Overview. Electric and B_1^+ fields produced in the head by each individual transmit channel are needed in order to design and calculate the SAR of pTX excitation pulses. We obtain

[1] This data is available at `www.fcc.gov/fcc-bin/dielec.sh`.

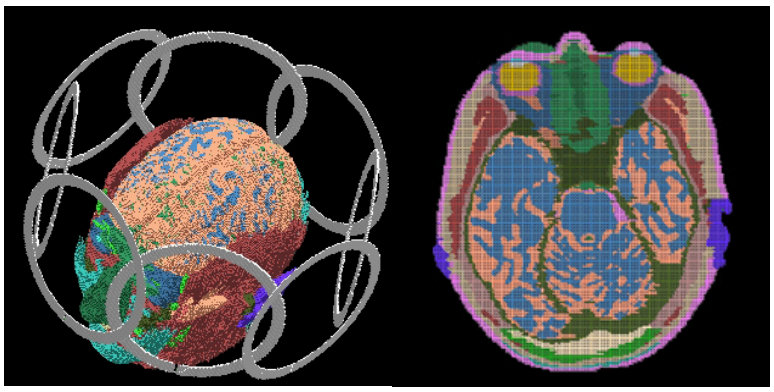

Figure 7-1: **Head and excitation array model**. Left subplot: head model centered within the eight-channel array of overlapped copper coils. Right subplot: transverse slice through the center of the head at $z = 0$; various finely-structured tissues are visible.

these via 300-MHz FDTD simulations using the XFDTD software package (REMCOM Inc., State College, PA) and then evaluate the SAR of any given excitation by superimposing modulated versions of these electric fields [78, 94, 111, 168]; this approach avoids the computationally-intractable scenario of FDTD-simulating each of the 162 excitations evaluated in this paper.

FDTD simulation. For $p = 1, \ldots, 8$, we drive the pth transmit channel of the eight-channel array with a 1-A peak-to-peak 300-MHz sinusoid, leave all other channels dormant, and use the FDTD method to obtain steady-state electric fields per ampere of input to the pth channel, $\mathbf{E}_p(\mathbf{r}) = [E_{p,x}(\mathbf{r}), E_{p,y}(\mathbf{r}), E_{p,z}(\mathbf{r})]^{\mathrm{T}}$ (V/m/A), and magnetic fields, $\mathbf{B}_p(\mathbf{r}) = [B_{p,x}(\mathbf{r}), B_{p,y}(\mathbf{r}), B_{p,z}(\mathbf{r})]^{\mathrm{T}}$ (T/A), generated at each of the 2.5 million spatial locations $\mathbf{r} = [x, y, z]^{\mathrm{T}}$ in the head. The use of current sources allows us to accurately approximate the simultaneous-drive behavior of the array by simply superimposing the field maps of individual channels [94].

Transmit profiles. The B_1^+ field that arises when channel p is driven by the unit ampere input, denoted $S_p(\mathbf{r})$, is derived at all locations $\mathbf{r}$ as follows [64]:

$$S_p(\mathbf{r}) = \tfrac{1}{2} \{B_{p,x}(\mathbf{r}) + iB_{p,y}(\mathbf{r})\} \quad \text{(T/A)} \tag{7.1}$$

Note: "S_p" is used (rather than "$B_{1,p}^+$") to keep upcoming formulations concise. Fig. 7-2 depicts the magnitude of the eight B_1^+ field maps obtained via (7.1) through the center transverse slice of the head about $z = 0$. Each is inhomogeneous.

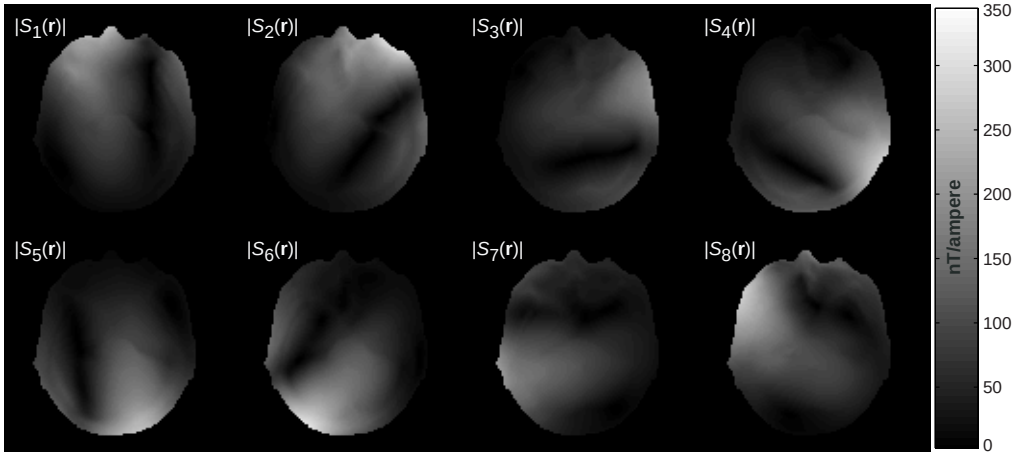

Figure 7-2: **Quantitative B_1^+ maps (nT/A) of the center transverse slice of the head derived from FDTD-simulated fields**. Each map exhibits severe inhomogeneity. $S_p(\mathbf{r})$ is the B_1^+ field profile that arises when channel p is driven with a 1-A peak-to-peak 300-MHz sinusoid.

7.2.3 Region-by-Region Error-Constrained Multi-Channel Pulse Design

Overview. We now describe one way to design a set of P RF pulse shapes, $a_1(t), \ldots, a_P(t)$ (A), each of duration L, to concurrently play through the P elements of a P-channel parallel transmission array in order to generate a user-defined target excitation, $d(\mathbf{r})$. Here, "pTX pulse" will be used as shorthand for "a set of P concurrently-transmitted RF pulse shapes that yields a box-shaped excitation". To begin, we will apply the linear formalism of [59] as outlined in Sec. 2.1.11 to reduce the design problem to that of solving a linear system and arrive essentially at (2.14), but instead of obtaining the pulses via (2.15, 2.16), we will pose and solve a different optimization. To begin, assume that the gradient waveforms, $\mathbf{G}(t)$, are fixed, the k-space trajectory, $\mathbf{k}(t)$, is predetermined, and that each of the P channel's B_1^+ profiles are known.

Design constants. For all upcoming simulations, we design pTX pulses that form approximations of a 2-D box-shaped inner-volume target in the center transverse slice of the head; the desired excitation has zero-degree flip angle at spatial locations outside of the box, a positive flip angle inside the box, and zero phase everywhere. The FOX is the center transverse slice of the head. The k-space trajectories are 2-D spirals that are radially undersampled (accelerated) by a factor of R relative to a 25.6 cm FOV, where $R = 1, 2, \ldots, 8$ (the $R = 1$ spiral is a conventional Nyquist-sampled spiral). The gradients are always constrained to amplitude and slew rates of 30 mT/m and 300 T/m/s, such that

the $R = 1, \ldots, 8$ spiral trajectories have durations of 6.8 ms, 3.47 ms, 2.36 ms, 1.81 ms, 1.48 ms, 1.26 ms, 1.11 ms, and 0.99 ms.

Small-Tip-Angle Approximation and Discretization. Imposing the small-tip-angle assumption [102] reduces the relation between the RF pulse shapes, gradients, and resulting magnetization to (2.13) as described in Sec. 2.1.11, except in this case we substitute $a_p(t)$s (current waveforms) for the $b_p(t)$s (voltage waveforms). We then discretize space at N_s locations $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$ within the user-defined FOX and sample time at N_t instants $t_1, \ldots, t_{N_t} \in [0, L]$; the t_i are spaced uniformly by Δ_t. This yields the system of equations given in (2.14), which we restate here using current vectors:

$$\begin{aligned} \mathbf{m} &= \mathbf{S}_1\mathbf{F}\mathbf{a}_1 + \cdots + \mathbf{S}_P\mathbf{F}\mathbf{a}_P \\ &= [\mathbf{S}_1\mathbf{F} \cdots \mathbf{S}_P\mathbf{F}] \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_P \end{bmatrix} = \mathbf{A}_{\text{tot}}\mathbf{a}_{\text{tot}}, \end{aligned} \tag{7.2}$$

where $\mathbf{m}$ is an N_s-element vector of samples of the resulting excitation $m(\mathbf{r})$, $\mathbf{S}_p$ is an $N_s \times N_s$ diagonal matrix containing samples of $S_p(\mathbf{r})$ taken within the FOX, and $\mathbf{F}$ is an $N_s \times N_t$ matrix that brings energy from k-space into the spatial domain. Finally, each $\mathbf{a}_p$ is an N_t-element vector of current samples of $a_p(t)$. Note that as a spiral trajectory is accelerated, fewer points in k-space are traversed and degrees of freedom in k-space are lost. The number of spatial-domain constraints N_s, however, remains constant, which means that the number of columns of $\mathbf{A}_{\text{tot}}$ decreases with R while the number of rows remains constant, causing the overall system in (7.2) to become increasingly overdetermined with R.

pTX pulse design. To excite a desired pattern, $d(\mathbf{r})$, P pulse shapes are needed. To generate these pulse shapes we first determine B_1^+ maps for each of the P channels and then decide on a k-space trajectory, which lets us generate the $\mathbf{S}_p$s, F, and $\mathbf{A}_{\text{tot}}$ in (7.2). We then sample $d(\mathbf{r})$ at $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$ and stack these samples into $\mathbf{d}$. At this point, one way to generate a set of pulse shapes that (approximately) produce the desired excitation is to solve (2.15) via pseudoinversion or the Tikhonov regularization [123, 124] given in (2.16) to obtain a solution $\mathbf{a}_{\text{tot}}$ In practice, setting δ in (2.16) to a small nonnegative value and solving (2.16) results in a reasonably-conditioned solution that produces an excitation close (in the ℓ_2 sense) to the one desired [165]. After solving (2.16), we extract samples of each $a_p(t)$

waveform from $\mathbf{a}_{\text{tot}}$ and play these waveforms on an actual system (or a Bloch-equation simulator) to produce an excitation that resembles the desired one [113].

Unfortunately, (2.16) does not let us allocate and fix the residual error between the resulting and desired excitation across different spatial regions. For example, for box-shaped excitations, designers are willing to tolerate larger errors outside of the box but require *within-box* error to be small. The approach of (2.16), however, gives us only a single variable, δ, with which to influence overall error, and we are thus unable to design pTX pulses that achieve both a chosen in-box error and a chosen overall (or out-of-box) error.

To circumvent this problem, we pose a novel algorithm that designs a pTX pulse that yields an excitation with both a desired in-box error and a desired overall error. First, we arrange N_{in} within-box samples of $d(\mathbf{r})$ into $\mathbf{d}_{\text{in}}$, and N_{out} out-of-box samples into $\mathbf{d}_{\text{out}}$ (where $N_{\text{in}} + N_{\text{out}} = N_s$), and then set $\mathbf{d} = [\mathbf{d}_{\text{in}}^{\text{T}}, \mathbf{d}_{\text{out}}^{\text{T}}]^{\text{T}}$. Likewise, we structure $\mathbf{A}_{\text{in}}$ and $\mathbf{A}_{\text{out}}$ such that $\mathbf{A}_{\text{tot}} = [\mathbf{A}_{\text{in}}^{\text{T}}, \mathbf{A}_{\text{out}}^{\text{T}}]^{\text{T}}$. These steps are analogous to those in Sec. 2.1.11 that lead to (2.16). With these new variables, we solve

$$\min_{\mathbf{a}_{\text{tot}}} \left\{ \|\mathbf{W}\left(\mathbf{d} - \mathbf{A}_{\text{tot}}\mathbf{a}_{\text{tot}}\right)\|_2^2 + \lambda \|\mathbf{a}_{\text{tot}}\|_2^2 \right\}, \tag{7.3}$$

where $\mathbf{W}$ is an $N_s \times N_s$ diagonal matrix such that $\mathbf{W}(n,n) = \alpha$ if $n \in \{1, \ldots, N_{\text{in}}\}$ and unity otherwise. Thus $\mathbf{W}$ weights within-box and out-of-box errors by α and unity, respectively. Eq. (7.3) has two variables, α and λ, which means we are now able to control both in-box error and overall error while still ensuring a well-conditioned solution.

In this paper, within-box and overall normalized root-mean-square error (NRMSE) are used as fidelity metrics and expressed as percentages. In-box NRMSE, ϵ_1, is defined as $100 \cdot \|\mathbf{d}_{\text{in}} - \mathbf{A}_{\text{in}}\mathbf{a}_{\text{tot}}\|_2 / \|\mathbf{d}_{\text{in}}\|_2$, whereas overall NRMSE, ϵ_{tot}, is defined as $100 \cdot \|\mathbf{d} - \mathbf{A}_{\text{tot}}\mathbf{a}_{\text{tot}}\|_2 / \|\mathbf{d}\|_2$. Whenever we design a pTX pulse, we first decide on a desired in-box error and overall error, denoted $\epsilon_{1,\text{des}}$ and $\epsilon_{\text{tot,des}}$. We then iteratively search over (α, λ), repeatedly solving (7.3) until a solution $\mathbf{a}_{\text{tot}}$ is found such that the resulting ϵ_1 and ϵ_{tot} are close to $\epsilon_{1,\text{des}}$ and $\epsilon_{\text{tot,des}}$. Finally, we simulate the waveform samples in $\mathbf{a}_{\text{tot}}$ and compare the simulated excitation to the desired one, ensuring that ϵ_1 and ϵ_{tot} remain close to $\epsilon_{1,\text{des}}$ and $\epsilon_{\text{tot,des}}$. Overall, this approach lets us design a variety of pTX pulses across different scenarios while guaranteeing that every excitation has essentially identical in-box and overall error.

7.2.4 Whole-Head and Local Specific Absorption Rate Calculations

Overview. After designing P pulse shapes to produce a desired excitation, we must determine the global and local SAR they produce in the head model. At this point, we know N_t time samples of each pulse shape spaced uniformly in time by Δ_t, i.e., for $p = 1, \ldots, P$, we have $[a_p(n\Delta_t)|n \in \{0, 1, N_t - 1\}]$. Recall that L is the overall duration of each pulse shape. Also assume we have decided upon the duty cycle, D, with which to play the pTX pulse (see end of subsection).

SAR per voxel. We first calculate SAR (W/kg) at each location $\mathbf{r}$, denoted SAR($\mathbf{r}$), by performing the numerical integration described in (2.22), which superimposes the electric field produced by each transmit channel due to each time sample in the RF pulse shape and then time averages the net field's squared magnitude over the pTX pulse duration and weights by the conductivity and density of that location; this procedure relies upon the FDTD-simulated $\mathbf{E}_p(\mathbf{r})$s and the σ and ρ properties of the head model depicted in Fig. 7-1.

Whole-head mean SAR. Having obtained SAR($\mathbf{r}$) for all 2.5 million locations $\mathbf{r}$ in the head, whole-head global SAR is obtained by averaging the SAR($\mathbf{r}$) values.

Fast Region Growth to Compute Local N-gram SAR. Local Ng SAR at $\mathbf{r}$ is obtained by finding an Ng cube around $\mathbf{r}$ and then averaging SAR($\mathbf{r}$) over all $\mathbf{r}$ within the cube, in line with [1]. To find an N-gram cube around each $\mathbf{r}$ we propose and implement a fast, low-memory algorithm rather than brute-force region growth because the latter is computationally infeasible given that there are 2.5 million locations $\mathbf{r}$ of interest and 162 pTX pulses to evaluate. We first form a list of spatial positions from a simple cubic lattice in order of distance from the origin and interpret this as a "universal" list of offsets from any given position $\mathbf{r}$. We then find the set of voxels that comprises N grams of tissue around $\mathbf{r}$ by choosing the shortest prefix of this list that yields sufficient total mass. Formally, given the mass per voxel around $\mathbf{r}$, $g(\mathbf{r})$, and this universal list, we form a voxel mass vector and search the cumulative sum of this voxel mass vector for the number of voxels needed to form the set; this rapidly determines the cluster of points that comprise an Ng cube around $\mathbf{r}$. These rapidly-determined points may be viewed as a set of indices, $I_N^{\mathbf{r}}$, such that

$$\sum_{i \in I_N^{\mathbf{r}}} g(\mathbf{r}_i) \approx N \text{ grams}, \tag{7.4}$$

and thus the Ng SAR at $\mathbf{r}$, denoted $\mathrm{SAR}^{N\mathrm{g}}$, is computed formally as:

$$\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}) = \frac{1}{|I_N^{\mathbf{r}}|} \sum_{i \in I_N^{\mathbf{r}}} \mathrm{SAR}(\mathbf{r}_i) \tag{7.5}$$

where $|I_N^{\mathbf{r}}|$ is the *cardinality* (number of elements) of $I_N^{\mathbf{r}}$. Readers should refer to [156] for more details about this algorithm and evidence of its superiority over conventional Ng SAR computation techniques. Finally, note that for a spatial location on the edge of the head bordering air (where $\sigma_{\mathrm{air}} = 0$), the associated $\mathbf{r}$ may not be at the center of the corresponding Ng averaging cube, but this is mitigated by the fact that this $\mathbf{r}$ is often part of the averaging cubes of many adjacent spatial locations [9].

Duty cycle. When computing SAR, the effect of the trajectory acceleration factor, R, is always accounted for to ensure that any SAR differences across R reflect only the extra power needed to maintain target fidelity. For example, $R = 1$ pTX pulses (6.8 ms long) have a 100% duty cycle, whereas $R = 4$ pTX pulses (1.81 ms long) have a 26% duty cycle. This is accomplished by fixing the repetition time (TR) to 6.8 ms for each effective sequence in which the pTX pulses are used, regardless of R.

7.3 Results

Birdcage mode. We begin by driving the array in a birdcage configuration, transmitting a 2.6-A, 3-ms, 100%-duty-cycle rectangular pulse shape through each channel in the absence of gradients and setting the phase of channel p's pulse shape to $45(p-1)$ degrees. This produces a 90-degree flip angle in the center transverse slice of the head, analogously to the "90-degree/3-ms" hard pulse of [27]. Fig. 7-3 depicts the inhomogeneous flip angle map that arises in the head when the set of pulse shapes that comprise one pTX pulse undergoes Bloch-equation simulation, along with the resulting SAR values. The image exhibits strong center brightening, resembling brain images collected on actual 7T systems equipped with homogeneous RF excitation volume birdcage coils [10, 133, 136]. In this case, there is approximately a 3-to-1 variation in peak-to-trough flip angle. The qualitative similarity of the simulated birdcage mode shown in Fig. 7-3 to the simulated and in vivo images presented in the lower half of the fifth figure in [136] suggests, to some extent, that the simulation methodology accurately captures the behavior of an eight-channel array at

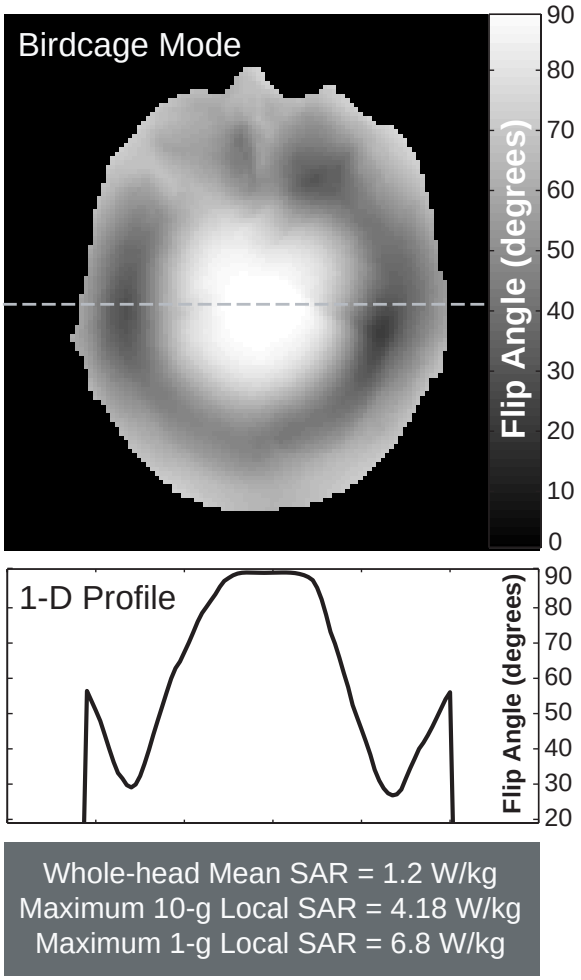

Figure 7-3: **Birdcage mode simulation**. For $p = 1, \ldots, 8$, transmission channel p is driven with a 3-ms, 2.6-A, 100%-duty-cycle rectangular pulse with phase equal to $45(p-1)$ degrees, producing a 90-degree flip in the center of the head (center transverse slice shown along with 1D profile and SAR statistics). Center brightening is evident, as well as signal loss left and right of center.

7T.

SAR as a function of R and θ for fixed excitation quality. The top row of Fig. 7-4 shows the simulated square inner-volume excitations for $R = 4$, along with ϵ_1 and ϵ_{tot} error with respect to the 28-mm $\times$ 28-mm target box pattern; for all cases, $\epsilon_1 = 15 \pm 2\%$ and $\epsilon_{\text{tot}} = 40 \pm 1\%$. Figure 7-4 also graphs global and maximum local 1g SAR for each acceleration factor ($R = 1, \ldots, 8$) as a function of flip angle ($\theta = 5, 15, \ldots, 45$degrees). For fixed R, global and local SAR scale quadratically with flip angle, whereas for fixed θ, mean and local SAR increase more than quadratically with R. Finally, global and local SAR vary strikingly by over five orders of magnitude across the various excitations.

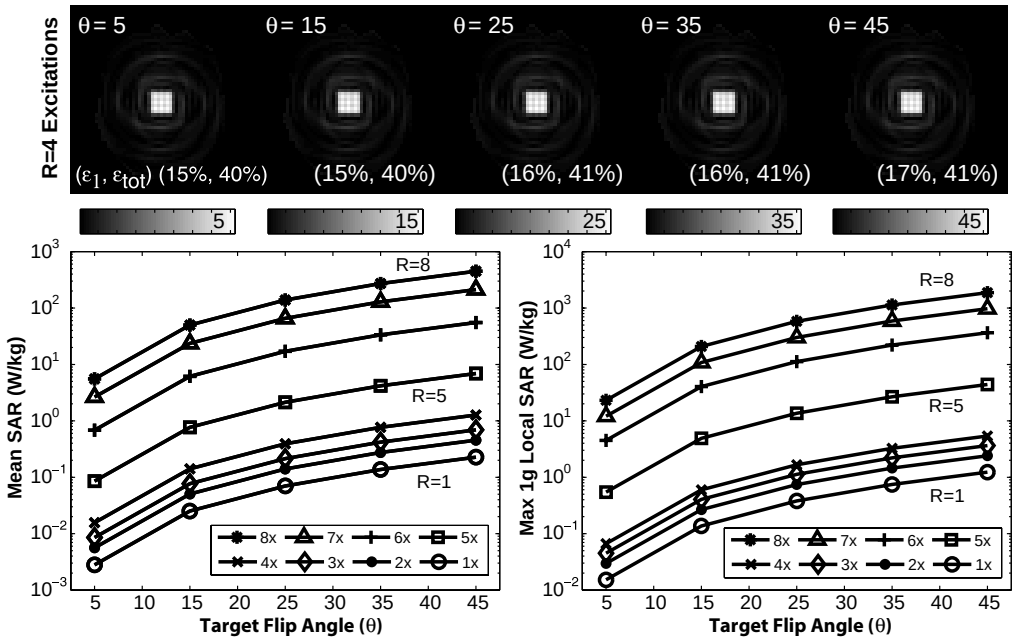

Figure 7-4: **SAR as a function of acceleration factor, R, and flip angle, θ (fixed excitation quality)**. Target: 28-mm × 28-mm centered square with in-box flip angle θ, $\epsilon_1 = 15 \pm 2\%$, $\epsilon_{\text{tot}} = 40 \pm 1\%$. Top row: $R = 4$ excitations. Bottom row: mean SAR and maximum 1g SAR as a function of (R, θ). For fixed θ, SAR grows rapidly with R; for fixed R, SAR grows quadratically with θ.

In general, the eight RF pulse shapes used to produce each of the excitations in Fig. 7-4 consistently exhibit low amperages during the time interval when the trajectory proceeds through high-frequency k-space regions, and progressively larger amperages as the trajectory spirals inward toward the origin of k-space. In other words, all RF pulse shapes deposit most of their energy at low spatial frequencies. Fig. 2-7 illustrates design details of the pTX pulses used to produce the 15-degree, 28-mm × 28-mm target box pattern of Fig. 7-4. We see here that peak RF magnitude across all eight pulse shapes increases from 1.5 A for $R = 1$, to 8 A for $R = 4$, to 150 A for $R = 7$. The $R = 7$ pulse shape has a much higher peak than the $R = 1$ pulse shape, a natural consequence of the fact that with only an $R = 7$ spiral, there are very few degrees of freedom in k-space with which to form the excitation (relative to the Nyquist-sampled $R = 1$ spiral), forcing the algorithm to drive the channel profiles intensely with high-amplitude RF pulse shapes in order to form the desired pattern. This lack of k-space freedom forces the system to rely heavily upon its degrees of freedom in the spatial domain and partly explains the high SAR values of the highly-accelerated pTX pulses.

Maximum local SAR as a function of R for fixed mean SAR and excitation quality. Since MR scanners typically monitor mean SAR (in the form of average forward power), it is informative to assess local SAR under the condition where the operator adjusts the sequence to achieve a fixed global average SAR. Figure 7-5 depicts such a scenario, showing iso-SAR operating contours in the (R, θ) parameter space that characterize how maximum local 1g and 10g SAR vary when mean SAR is held constant at 0.15 W/kg. Excitation patterns and residual errors here are the same as in Fig. 7-4. Figure 7-5 shows that maximum local 1g SAR is always higher than maximum local 10g SAR, which in turn is always higher than whole-head mean SAR. Across R, local 1g and 10g SAR vary from peak-to-trough by factors of 1.64 and 1.34, respectively. Furthermore, Fig. 7-5's left panel shows that in order to produce fixed-fidelity excitations with 0.15 W/kg mean SAR, in-box flip angle must decrease rapidly with R, e.g., only a 2-degree flip is achievable for $R = 6$, whereas a 36-degree flip is possible using an unaccelerated trajectory. Counter-intuitively, Fig. 7-5's right panel reveals that maximum local 1g and 10g SAR are not monotonic with R, e.g., the $R = 4$ excitation produces 1.3 times less maximum local SAR than the $R = 1$ excitation. Finally, the ratios of local 1g and 10g SAR to mean SAR are erratic: they are roughly constant for $R = 1$ to 3, decrease for $R = 4$, rise for $R = 5, 6$, and then decrease again for $R = 7, 8$.

SAR as a function of R and spatial position of the excitation. To assess the effect of excitation symmetry on SAR we design pTX pulses that excite 28-mm $\times$ 28-mm, 15-degree flip angle squares at different locations along the right-left (RL, or "x") axis and the anterior-posterior (AP, or "y") axis. Excitation quality is fixed across all designs. Figure 7-6 illustrates how global and maximum local 1g SAR behave when exciting boxes centered at different x locations. The upper row depicts $R = 4$ excitations, while the next two rows show local 1g SAR maps arising due to $R = 1$ and $R = 5$ excitations. For a given row of local SAR maps, all are displayed using the same dynamic range, permitting local SAR comparisons among the maps of each row. Furthermore, because a 1g local SAR data point exists for each of the voxels in the 3-D head model, the 2-D local SAR maps have been "collapsed" along the z-axis. Namely, for each location (x, y) in each given map, we have displayed $\max_z \text{SAR}_{\text{1-gram}}(x, y, z)$. Finally, the bottom row of Fig. 7-6 illustrates log-scaled plots of mean SAR and maximum local 1g SAR as a function of box position. Fig. 7-7 shows SAR behavior versus box shift along y and is formatted analogously to Fig. 7-6, except here

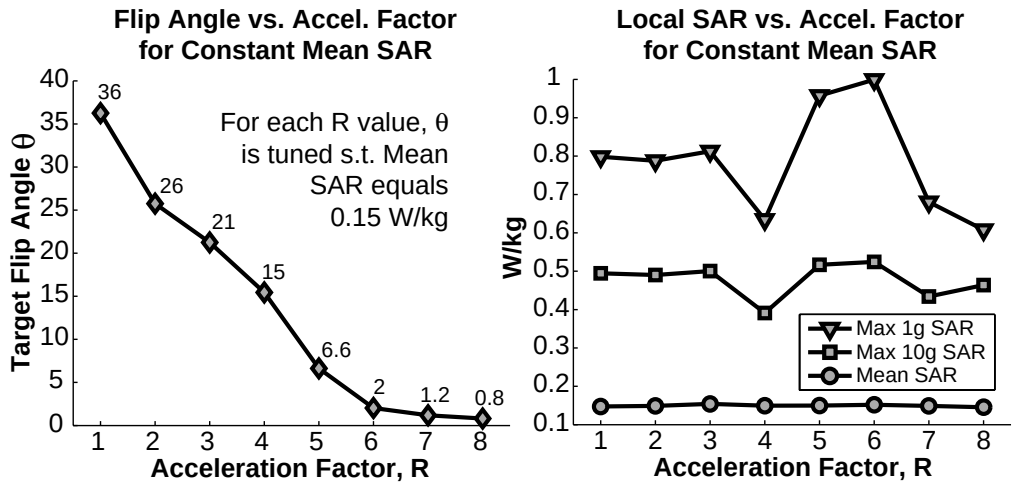

Figure 7-5: **Local SAR and θ as a function of R (fixed mean SAR and excitation quality)**. Target: 28-mm × 28-mm centered square with $\epsilon_1 = 15 \pm 2\%$ and $\epsilon_{\text{tot}} = 40 \pm 1\%$. For each R, target flip angle, θ, is varied until mean SAR equals 0.15 ± 0.01 W/kg. Left: to achieve fixed mean SAR, θ must decrease radically with R. Only a 2-degree flip is achievable for $R = 6$ whereas a 36-degree flip is achievable using an $R = 1$ unaccelerated trajectory. Right: maximum 1g and 10g SAR are not monotonic with R, e.g., the $R = 4$ excitation has 1.3 times lower local SAR than the $R = 1$ one.

local SAR maps are not displayed.

In both Fig. 7-6 and Fig. 7-7, SAR as a function of position is roughly convex for low trajectory acceleration factors ($R = 1$ to 4), i.e., centered excitation boxes yield the lowest SAR. At higher acceleration factors, however, centered excitations generally produce the highest SAR. Figure 7-6 also shows that mean and local SAR do not always behave similarly, e.g., for $R \leq 4$, the mean SAR vs. R curves are generally symmetric about $x_0 = 0$, whereas the local SAR curves are asymmetric. Figure 7-7 shows that qualitative SAR differences also arise when shifting the box along y. Furthermore, for fixed R, all of Fig. 7-7's local and mean SAR curves seem to exhibit the same shape, in contrast with the $R \leq 4$ curves of Fig. 7-6.

Regarding the local SAR maps of Fig. 7-6, the local 1g SAR patterns for $R = 1$ change across space in a way that is correlated with the position of the excited box, while the $R = 5$ local SAR maps, to the first order, seem to only scale by a multiplicative constant with box position. Finally, the peaks in both the $R = 1$ and $R = 5$ local SAR maps are not strongly correlated with the box position.

SAR as a function of R and in-box NRMSE, ϵ_1, for fixed overall NRMSE, ϵ_{tot}. Figure 7-8 shows the effect of excitation quality on SAR, illustrating global and maximum

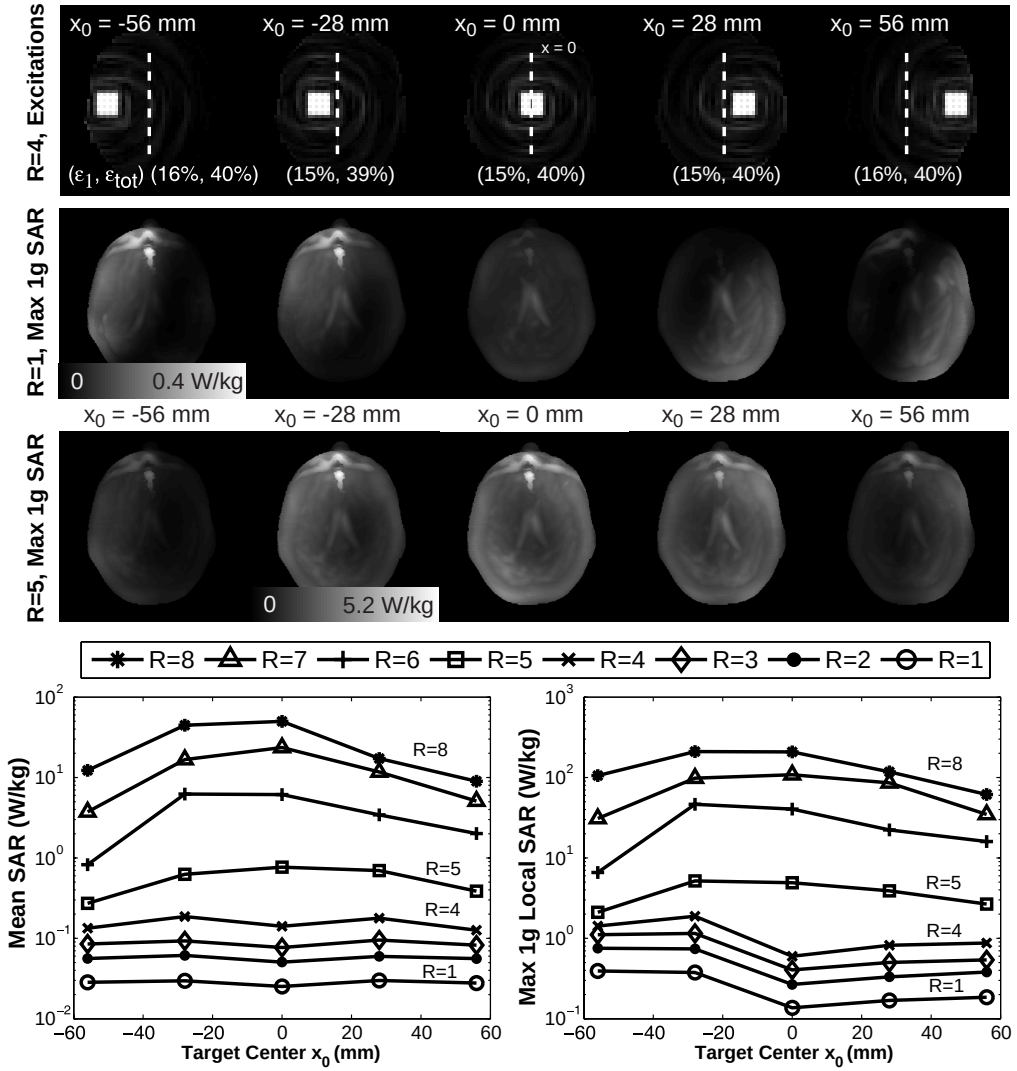

Figure 7-6: **SAR as a function of R and shift along x (fixed excitation quality).** Target: 15-degree, 28-mm × 28-mm square whose center x_0 varies along x with $\epsilon_1 = 15 \pm 1\%$ and $\epsilon_{\text{tot}} = 40 \pm 1\%$. Top row: $R = 4$ excitations. Second and third rows: local 1g SAR maps due to $R = 1$ and $R = 5$ pulses. Each spatial location (x, y) shown equals $\max_z \text{SAR}_{\text{1-gram}}(x, y, z)$, i.e., the maps have been collapsed along the z axis to efficiently display 3-D spatial SAR data. Bottom row: mean SAR and maximum 1g SAR as a function of (R, x_0). For $R \leq 4$, SAR increases as $|x_0|$ increases, whereas for $R > 4$, SAR decreases. For $R = 1$, local SAR seems to vary strongly across space with excitation position, whereas for $R = 5$, it seems to scale by only a multiplicative constant.

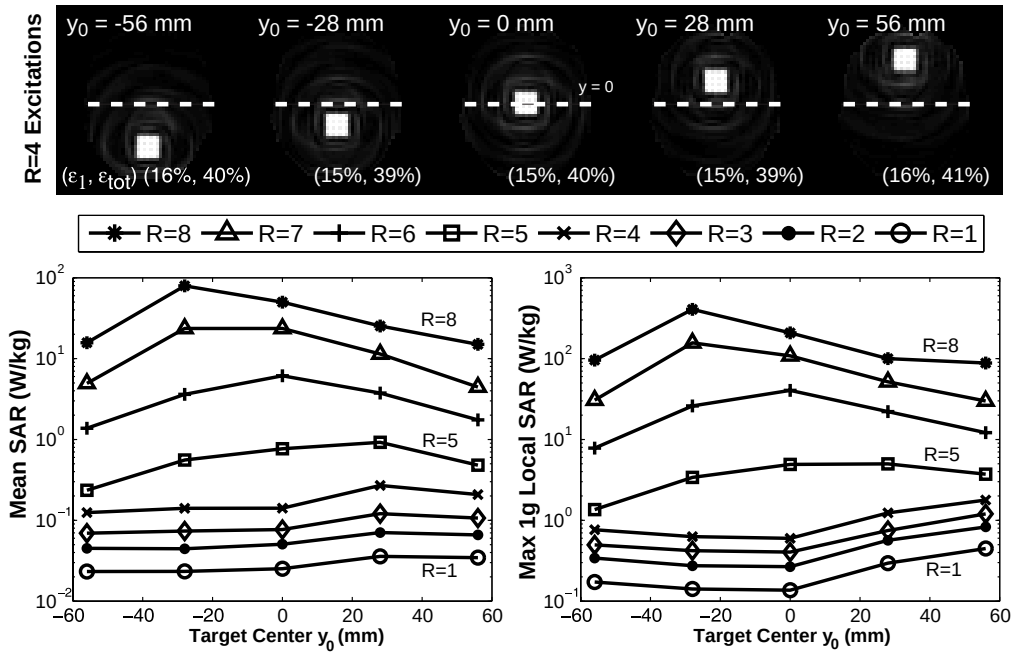

Figure 7-7: **SAR as a function of R and shift along y (fixed excitation quality)**. Target: 15-degree, 28-mm × 28-mm square whose center y_0 varies along y with $\epsilon_1 = 15 \pm 1\%$ and $\epsilon_{\text{tot}} = 40 \pm 1\%$. Top row: $R = 4$ excitations. Bottom row: mean SAR and maximum 1g SAR as a function of (R, y_0).

local 1g SAR for $R = 1, \ldots, 8$ due to exciting 28-mm × 28-mm centered boxes with different in-box fidelities; ϵ_1 varies from $10\%, 15\%, \ldots, 30\%$, while ϵ_{tot} is fixed at $40 \pm 2\%$. Figure 7-8's top row shows excitation patterns associated with $R = 4$, while the bottom row shows global and maximum local 1g SAR as a function of ϵ_1. We see from the top row of $R = 4$ excitations that the in-box flip angle decreases (the boxes grow darker) as more in-box error is permitted. Further, global SAR and maximum local 1g SAR both increase rapidly with R, ranging over three orders of magnitude for each fixed ϵ_1. Finally, SAR decreases fairly regularly with ϵ_1 for most accelerations, but this does not hold for $R = 5$ and $R = 6$, e.g., when $R = 6$, the $\epsilon_1 = 15\%$ excitation produces 1.3 and 1.5 times higher mean and local SAR than the $\epsilon_1 = 10\%$ excitation.

SAR as a function of R and ϵ_{tot} for fixed ϵ_1. Figure 7-9 illustrates the dual of Fig. 7-8's experiment, showing how global and maximum local 1g SAR behave across R when producing excitations with different overall fidelities ($\epsilon_{\text{tot}} = 20\%, 25\%, , 40\%$) while keeping $\epsilon_1 = 15 \pm 1\%$. Here, global SAR and local SAR vary by over four orders of magnitude across R and ϵ_{tot}. For $R \neq 5$, global and local SAR decrease monotonically with ϵ_{tot}, and even for $R = 5$, SAR is nearly monotonic: the lower-fidelity $\epsilon_{\text{tot}} = 35\%$ excitation produces

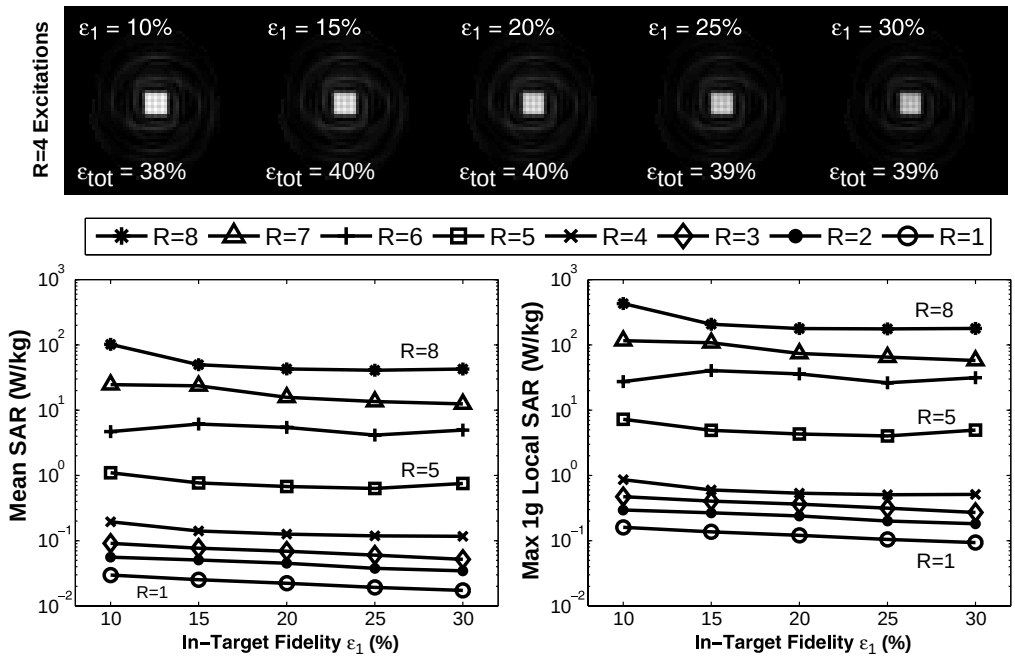

Figure 7-8: **SAR as a function of R and ϵ_1 (fixed $\epsilon_{\mathbf{tot}}$)**. Target: 15-degree, 28-mm × 28-mm square with varying ϵ_1 and $\epsilon_{\text{tot}} = 40 \pm 2\%$. Top row: $R = 4$ excitations; in-box flip angle decreases with ϵ_1. Bottom row: mean SAR and maximum 1g SAR as a function of (R, ϵ_1). For $R \neq 6$, SAR decreases fairly regularly with ϵ_1. For $R = 6$, SAR is actually higher when $\epsilon_1 = 15\%$ than when $\epsilon_1 = 10\%$.

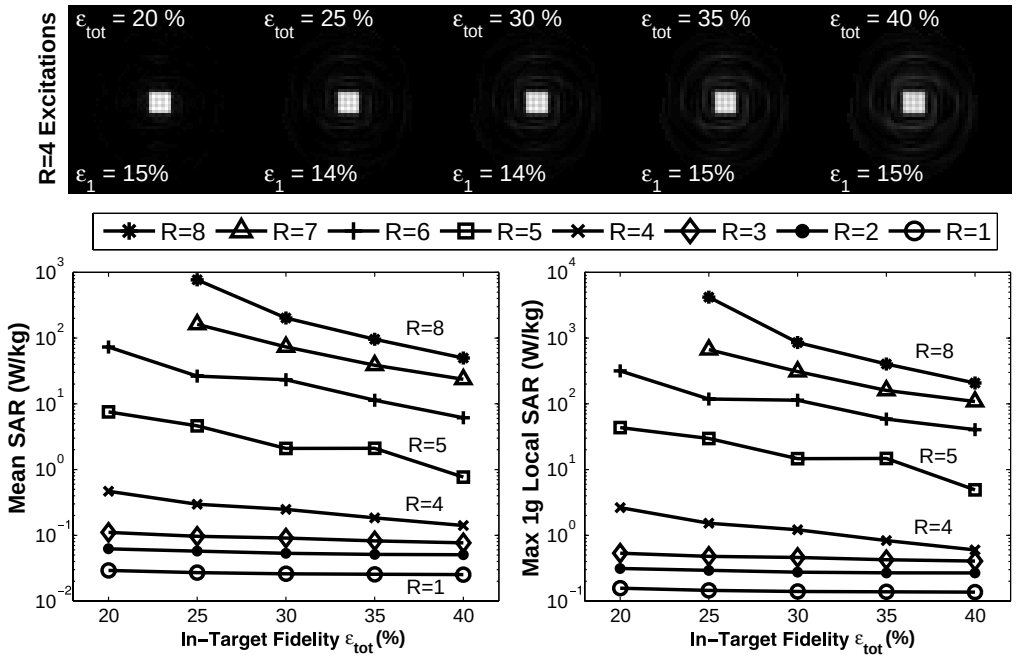

Figure 7-9: **SAR as a function of R and $\epsilon_{\mathbf{tot}}$ (fixed ϵ_1)**. Target: 15-degree, 28-mm × 28-mm centered square with varying ϵ_{tot} and $\epsilon_1 = 15 \pm 1\%$. Top row: $R = 4$ excitations. Outside of the square, artifacts increase with ϵ_{tot}. Bottom row: mean SAR and maximum 1g SAR as a function of $(R, \epsilon_{\text{tot}})$. For all R, SAR generally decreases smoothly with ϵ_{tot}.

only 1.007 times higher SAR than the $\epsilon_{\text{tot}} = 30\%$ case. (Note: there are no data points for $\epsilon_{\text{tot}} = 20\%$ when $R = 7$ and $R = 8$ because for these cases the pTX pulse design algorithm does not produce well-conditioned solutions.)

SAR vs. R and box orientation. Figure 7-10 illustrates the affect of excitation orientation on global and maximum local 1g SAR across the eight acceleration factors. Here we excite 44-mm × 28-mm rectangles in the center of the head that are rotated by ϕ degrees, characterizing SAR as a function of R and ϕ. Excitation fidelity is fixed as in prior experiments. For $R \leq 4$, SAR is relatively constant, whereas for $R > 4$, SAR exhibits spikes at particular rotations. For example, when $R = 5$, exciting a 90-degree-rotated box produces 3.3 W/kg of maximum local 1g SAR, whereas exciting a 135-degree-rotated box produces 9.54 W/kg, i.e., local SAR varies by a factor of 2.9 when simply rotating the excitation by 45 degrees, even with acceleration factor and excitation quality held constant. Finally, SAR is generally higher when boxes are highly asymmetric with respect to the AP-direction (y-axis) of the head (consider the $R = 6$, $\phi = 45$ degrees excitation).

SAR vs. R and box size. Figure 7-11 illustrates how the size of an excitation impacts global and maximum local 1g SAR. Here a series of increasingly-larger boxes of length and

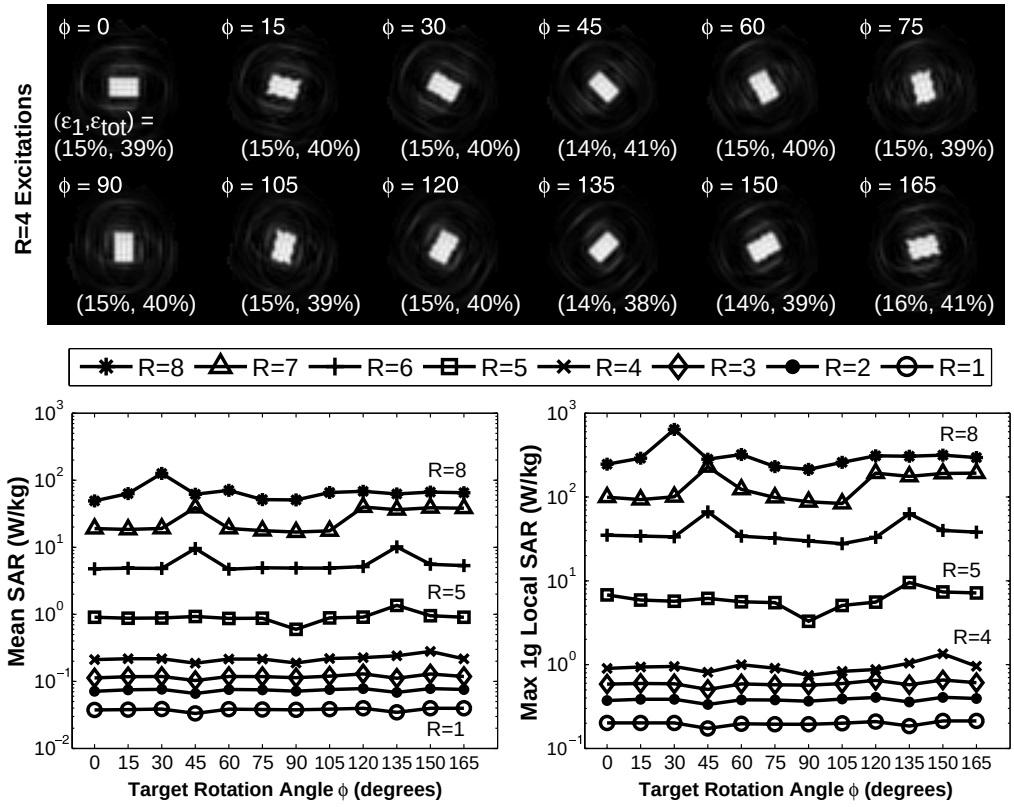

Figure 7-10: **SAR as a function of R and target rotation angle ϕ (fixed excitation quality)**. Target: 15-degree, 44-mm $\times$ 28-mm centered rectangle with varying ϕ, with $\epsilon_1 = 15 \pm 1\%$ and $\epsilon_{\text{tot}} = 40 \pm 2\%$. Top row: $R = 4$ excitations. Bottom row: mean SAR and maximum 1g SAR as a function of (R, ϕ). For $R \leq 4$, SAR is relatively constant, whereas for $R > 4$, SAR varies.

width N (mm) are excited across various acceleration factors. Excitation fidelity is fixed as in prior experiments, and Fig. 7-11 is formatted analogously to Figs. 7-7, 7-8, 7-9,7-10. Both the log-scaled and normalized data along the middle and bottom rows of Fig. 7-11 show that mean and local SAR behave quite differently depending on R: for $R \leq 4$, SAR grows rapidly with N, whereas for $R > 4$, SAR decreases with N. This means that for highly accelerated trajectories, exciting larger regions actually reduces energy deposition. For example, for $R = 5$, exciting a 52-mm box produces 3.2 times less mean SAR than exciting a 12-mm box, yet for $R = 1$, the opposite behavior occurs: exciting a 52-mm rather than 12-mm box leads to ten times higher mean SAR.

SAR vs. R and box smoothness. Figure 7-12 shows how sharp excitation pattern edges affect SAR. Here, increasingly-smooth 44-mm $\times$ 44-mm centered boxes are excited and the SAR of each is analyzed. The series of desired excitations is generated by applying successively larger M-mm $\times$ M-mm Gaussian smoothing kernels to the original sharp-edged target pattern and running the design algorithm each time to produce a pTX pulse. The top row of Fig. 7-12 shows that for $R = 4$, the excitations increase in smoothness with M as intended. The bottom row shows that mean and local SAR decrease significantly with target smoothness for all R. For example, when $R = 6$, exciting the smoothest box requires 2.8 times less global SAR and 2.7 times less maximum local 1g SAR than does exciting the sharpest-edged box.

Mean vs. maximum local 1g and 10g SAR. Across all experiments, maximum 1g SAR ranges from 3.8 to 13.8 times larger than corresponding mean SAR and on average is 5.6 times larger. Likewise, maximum 10g SAR is always 2.3 to 7.7 times larger than mean SAR and on average is 3.4 times larger. Finally, maximum 1g SAR ranges from 1.1 to 2.1 times larger than maximum 10g SAR and on average is 1.7 times larger.

7.4 Discussion

Comparisons across experiments. Considerable effort has been made to keep excitation fidelity constant across all experiments and thus permit SAR comparisons across experiments. One noticeable trend is that a consistent "jump" in mean and local SAR occurs as R transitions from 4 to 5, regardless of excitation shape, size, asymmetry, etc., which suggests the array is better conditioned in the $R \leq 4$ operating region.

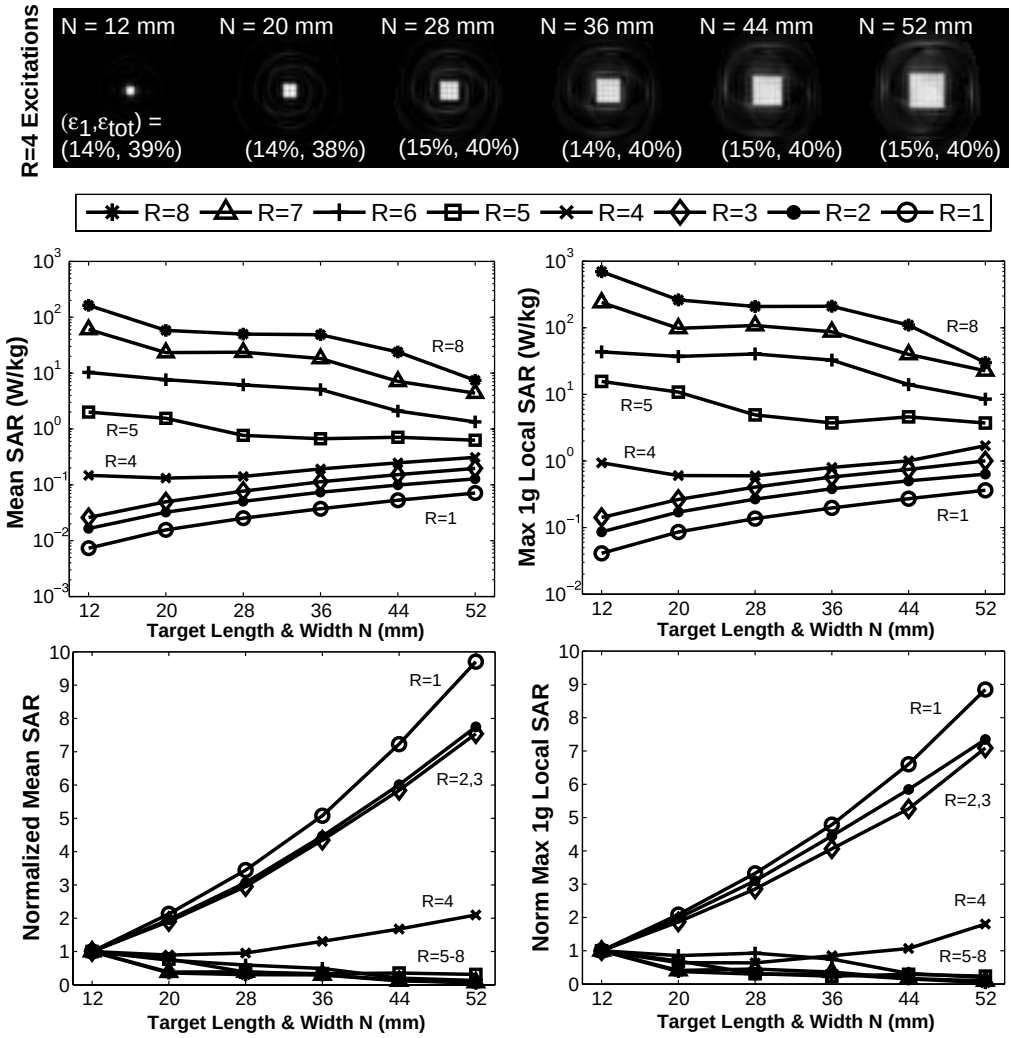

Figure 7-11: **SAR as a function of R and target size N (fixed excitation quality)**. Target: 15-degree centered square of varying size with $\epsilon_1 = 15 \pm 1\%$ and $\epsilon_{\text{tot}} = 40 \pm 2\%$. Top row: $R = 4$ excitations. Middle row: mean SAR and maximum 1g SAR as a function of (R, N). Bottom row: for each R, data from middle row is scaled so that SAR(R, $N = 12$ mm) is unity. For $R \leq 4$, SAR increases rapidly with N, whereas for $R > 4$, exciting larger regions reduces energy deposition.

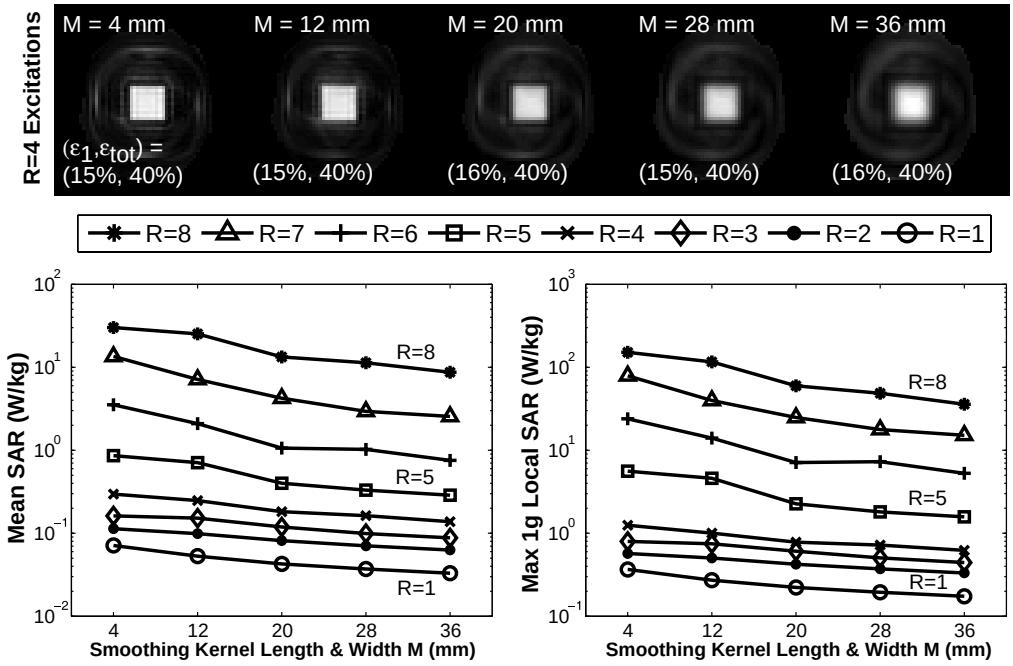

Figure 7-12: **SAR as a function of R and target smoothness (fixed excitation quality)**. Target: 15-degree, 44-mm $\times$ 44-mm square smoothed with various M-mm $\times$ M-mm Gaussian kernels with $\epsilon_1 = 15 \pm 1\%$ and $\epsilon_{\text{tot}} = 40 \pm 1\%$. The base case of no smoothing is when $M = 4$ mm (pixel resolution is 4 mm, so an $M = 4$ mm window is simply a one-pixel unit-energy kernel). Top row: $R = 4$ excitations. Bottom row: mean SAR and maximum 1g SAR as a function of (R, M). In general, SAR generally decreases smoothly with M.

The cost of trajectory acceleration and SAR-reduced design. Across all experiments, global and local SAR always increase dramatically with trajectory acceleration factor. Highly-accelerated pulses yield extremely high SAR values and may not be practical for *invivo* imaging. Furthermore, because even moderate trajectory accelerations may lead to order-of-magnitude SAR increases, it seems likely that further innovation beyond or in conjunction with conventional pTX pulse design is necessary in order to enable the use of highly accelerated trajectories. In Sec. 2.1.15, we provided an overview of all recent SAR-reduced pulse design research, and also cited real-time and hardware-based approaches to monitoring and reducing SAR. In Ch. 8, we will propose several SAR-reduced pTX design techniques, and in Ch. 9 we show that time multiplexing a set of similar pTX pulses is an effective low-computational-cost approach to reduce maximum local SAR.

SAR observations. In Fig. 7-4, mean and local SAR scale quadratically with flip angle. Fig. 7-5 shows that the ratio of local to mean SAR does not increase monotonically with acceleration factor. Figures 7-6 and 7-7 show that mean and local SAR behave non-intuitively when excitation spatial position is varied. For $R = 1$, local SAR varies significantly across space with excitation position, whereas for $R = 5$, to the first order, the local SAR maps simply undergo scalings by a multiplicative constant. Figure 7-8 shows that it is not always possible to reduce SAR by simply permitting more excitation error within a specific spatial region (e.g., within the box) and that one may in fact significantly increase SAR (e.g., by a factor of 1.5) by generating lower-quality excitations. Figure 7-9, on the other hand, shows that permitting more overall excitation pattern error generally decreases SAR. Figure 7-10 shows that for large acceleration factors, excitation asymmetry detrimentally impacts SAR. Figure 7-11 shows that SAR is sensitive to excitation size, suggesting that parallel transmission may complicate the ability to perform and flexibly scale reduced-FOV imaging [66]. Finally, Fig. 7-12 suggests that sharp edges are costly in terms of SAR, revealing that one reliable way to reduce SAR is to excite a smoother version of the desired pattern.

7.5 Conclusion

In conclusion, with the exception of dramatic SAR increases with trajectory acceleration factor, global and local SAR do not always exhibit intuitive trends. Nonetheless, it is clear

from the experiments that maximum local 1g and 10g SAR are always significantly higher than global SAR. Because both the United States and European Union safety standards impose limits on maximum local SAR [21, 70, 71], and because the ratio of local to global SAR is often considerably greater than the regulatory ratio required to maintain safety compliance for the human head by monitoring average power alone, it is evident that local SAR, rather than global SAR, is the limiting factor of eight-channel parallel transmission at 7T. Namely, it is likely that the safety limit imposed upon local SAR will preclude the user from utilizing the full limit of mean SAR. Seifert et al. have arrived at this identical conclusion after studying four-channel parallel transmission at 3T [111].

Although the range of excitation patterns studied here is not exhaustive, a sufficient number of variations of the size, shape, position, rotation and smoothness of the excitation pattern have been considered to suggest that mean and local SAR exhibit complex and at times non-intuitive behavior as a function of target excitation pattern and trajectory acceleration factor. In order to ensure patient safety, it seems that model-based validation of individual target patterns and corresponding sets of parallel transmission pulses will be required.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 8

Specific-Absorption-Rate-Reduced Multi-Channel Excitation Pulse Design

8.1 Introduction

In Ch. 7, we concluded that high local specific absorption rate (SAR) is a major safety concern of parallel excitation pulses. To address this problem, here we propose to account for electric-field interactions while designing a radio-frequency (RF) pulse and thus generate multi-channel excitation waveforms whose SAR characteristics are explicitly optimized. Versions of this work have appeared at a conference [158, 160].

We first pose a linear-algebraic formulation to evaluate whole-head or local N-gram (Ng) SAR when designing RF pulses on P-channel systems that exploits knowledge of the steady state electric fields generated per unit of current sent through each transmit channel, the tissue's electrical properties, and spatial sensitivity profiles (B_1^+ maps) of each coil; we show local Ng SAR at any location may be computed using a highly-sparse matrix. We then pose optimization problems that produce RF pulses with optimal SAR characteristics. In particular, we provide a closed-form solution for optimizing mean SAR, introduce a method to explore excitation fidelity, mean SAR, and pulse duration tradeoffs, and pose a constrained optimization problem that ensures local Ng SAR meets certain constraints. We then discuss the computational difficulties one ultimately faces when attempting to

implement such local-SAR-optimized pulse design methods.

Note that throughout this chapter we will rely on background material from Sec. 2.1.11 and Sec. 2.1.14, as well as the Ng SAR cubic region concept of Sec. 7.2.4. Readers interested in more information on SAR-reduced design should refer to Sec. 2.1.15 for a host of citations to recent work.

8.2 Assumptions

Let us consider a P-channel system whose transmission profiles $S_1(\mathbf{r}), \ldots, S_P(\mathbf{r})$ (T/A) are known for all M_{FOX} points of interest $\mathbf{r}$ within a user-defined field of excitation (FOX); these $\mathbf{r}$s that index within-FOX points are contained in the set C_{FOX}. Our goal is to determine a set of current waveforms $a_1(t), \ldots, a_P(t)$, each of duration L, to play through the system to form a high-fidelity version of the small-tip-angle excitation $d(\mathbf{r})$ within the FOX while constraining (or minimizing) whole-head mean SAR, local SAR, or both. The pulse will have a fixed and known repetition time, TR. Let us also assume that the gradient waveforms, $\mathbf{G}(t)$, are fixed, which means that the excitation k-space trajectory, $\mathbf{k}(t)$, is predetermined. Our specific goal now is to determine N_t discrete samples of each $a_p(t)$ (spaced by Δ_t in time) to transmit through each transmission channel to produce a low-SAR high-quality excitation.

To enable SAR computations, we have a model of the human body and full knowledge of the 3-D electric fields that arise at all M_{model} locations $\mathbf{r}$ within it due to the use of each transmission channel. That is, for $p = 1, \ldots, 8$, we know the steady-state electric fields per unit ampere of input to the pth channel, $\mathbf{E}_p(\mathbf{r}) = [E_{p,x}(\mathbf{r}), E_{p,y}(\mathbf{r}), E_{p,z}(\mathbf{r})]^{\text{T}}$ (V/m/A).[1] We also know the density, $\rho(\mathbf{r})$ (kg/m^3), and electrical conductivity, $\sigma(\mathbf{r})$ (S/m), at each location. Further, let all $\mathbf{r}$s that index the body be contained in the set C_{model}. This set may be quite large, e.g., in Ch. 7, the head model depicted in Fig. 7-1 consists of 2.5 million voxels. Finally, assume that for each $\mathbf{r}_m$, we know the cluster of points that comprise an Ng cube about $\mathbf{r}_m$; this cluster may be viewed as a set of R_m known indices, $I_N^{\mathbf{r}_m}$. [Ng region set notation and fast determination of such sets were discussed in Sec. 7.2.4; see (7.4).]

[1] Assuming the use of current sources allows us to accurately approximate the simultaneous-drive behavior of the array by simply superimposing the field maps of individual channels [94].

8.3 Regularized Multi-Channel RF Excitation Pulse Design

For a P-channel system, linearizing and discretizing the equations relating the RF pulse shapes played through each transmission channel to the resulting excitation as in Sec. 2.1.11 yields the system given in (2.14), which we restate:

$$\mathbf{m} = \mathbf{S}_1\mathbf{F}\mathbf{a}_1 + \cdots \mathbf{S}_P\mathbf{F}\mathbf{a}_P = \mathbf{A}\mathbf{a}_{\text{tot}}, \tag{8.1}$$

where here we have specified $\mathbf{a}_p$s rather than $\mathbf{b}_p$s (i.e., vectors of current samples rather than vectors of voltage samples). The vector $\mathbf{m}$ contains samples of the excitation $m(\mathbf{r})$ that forms within the FOX due to the transmission of the pulse samples in the $\mathbf{a}_p$s. The $\mathbf{S}_p$s, $\mathbf{F}$, and $\mathbf{A}$ are constructed only with respect to the FOX, so they are of sizes $M_{\text{FOX}} \times M_{\text{FOX}}$, $M_{\text{FOX}} \times N_t$, and $M_{\text{FOX}} \times PN_t$, respectively.

At this point, we might form M_{FOX} within-FOX samples of $d(\mathbf{r})$ into the vector $\mathbf{d}$ and design a set of pulses to accomplish the excitation via the Tikhonov regularization given in (2.16), but this would not explicitly account for whole-head or local SAR. We see that the linear expression $\mathbf{A}\mathbf{a}_{\text{tot}}$ only explains how a given set of pulses influences the excitation across space, not how a pulse set influences SAR. Thus, to capture SAR explicitly in the design process, we must first construct new linear-algebraic variables that allow us to express SAR as a function of $\mathbf{a}_{\text{tot}}$.

8.4 Linear-Algebraic Formulation of SAR using Sparse Block-Diagonal Matrices

We now derive a matrix-vector expression for $\text{SAR}(\mathbf{r}_m)$, the point SAR in W/kg at spatial location $\mathbf{r}_m$, as well as a SAR matrix that allows us to compute local Ng SAR at $\mathbf{r}_m$ given a set of pulse samples represented by $\mathbf{a}_{\text{tot}}$ in (8.1). Our work in this section ultimately obtains a matrix analogous to the mean-SAR matrix given in [168], but we arrive at our matrix using a different approach. Our resulting matrix is capable of computing local Ng SAR about any location rather than simply whole-head mean SAR. We hope our step-by-step formulation given here provides designers new insight into the design problem. We feel that the expressions below, given their sparse, structured natures, may be exploited from a computational standpoint to reduce the complexity of SAR calculations, but this is outside

the scope of the dissertation.

To begin, we know from (2.22) in Sec. 2.1.14 that we may accurately approximate $\mathrm{SAR}(\mathbf{r}_m)$ via numeric integration:

$$\mathrm{SAR}(\mathbf{r}_m) \approx w(\mathbf{r}_m) \sum_{n=0}^{N_t-1} \|\mathbf{E}(\mathbf{r}_m, n\Delta_t)\|_2^2, \tag{8.2}$$

where $w(\mathbf{r}_m) = \sigma(\mathbf{r}_m)\Delta_t/(2\rho(\mathbf{r}_m)\cdot \mathrm{TR})$ and $\mathbf{E}(\mathbf{r}, t)$ is given in (2.21). Now we expand (2.21) as follows:

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{E}_1(\mathbf{r})a_1(t) + \cdots + \mathbf{E}_P(\mathbf{r})a_P(t) \\ &= [\mathbf{E}_1(\mathbf{r}) \cdots \mathbf{E}_P(\mathbf{r})] \begin{bmatrix} a_1(t) \\ \vdots \\ a_P(t) \end{bmatrix} \\ &= \mathbf{F}(\mathbf{r})\mathbf{q}(t), \end{aligned} \tag{8.3}$$

where $\mathbf{F}(\mathbf{r})$ is a $3 \times P$ matrix containing all relevant electric field data at point $\mathbf{r}$ and $\mathbf{q}(t)$ is a vector of the P samples (in amperes) transmitted concurrently through the P channels at one time instant. Notice that if $P \geq 3$ and assuming $\mathbf{F}(\mathbf{r})$ is not singular (i.e., that the channels each influence the electric field at this location differently and that a substantial number of the $\mathbf{E}_p(\mathbf{r})$s are not collinear), then at any single instant in time we will be able to find a P-element solution vector $\mathbf{q}(t)$ that tailors $\mathbf{E}(\mathbf{r}, t)$ to any desired pattern. In the case of a $P = 2$-channel system, however, this is not possible: there are too few unknowns to guarantee an exact solution to the $3 \times P$ system of equations. This seems to imply that small-P systems—and especially $P = 2$ systems—will be unable to minimize or control SAR as well as high-P systems when both types of systems are faced with high-fidelity excitation constraints.

Returning to our derivation, we substitute (8.3) into (8.2), yielding

$$\mathrm{SAR}(\mathbf{r}_m) \approx w(\mathbf{r}_m) \sum_{n=0}^{N_t-1} \mathbf{q}^{\mathrm{H}}(n\Delta_t)\mathbf{F}^{\mathrm{H}}(\mathbf{r}_m)\mathbf{F}(\mathbf{r}_m)\mathbf{q}(n\Delta_t) = \mathbf{q}_{\mathrm{tot}}\mathbf{G}(\mathbf{r}_m)\mathbf{q}_{\mathrm{tot}}, \tag{8.4}$$

where $\mathbf{q}_{\mathrm{tot}} = [\mathbf{q}^{\mathrm{T}}(0), \ldots, \mathbf{q}^{\mathrm{T}}((N_t-1)\Delta_t)]^{\mathrm{T}} \in \mathbb{C}^{PN_t}$ and $\mathbf{G}(\mathbf{r}_m) = \mathbf{I}_{N_t} \otimes (w(\mathbf{r}_m)\mathbf{F}^{\mathrm{H}}(\mathbf{r}_m)\mathbf{F}(\mathbf{r}_m))$ is $PN_t \times PN_t$ elements in size, block-diagonal (with N_t identical $P \times P$ blocks), and very

sparse (at most, only P^2N_t of its $P^2N_t^2$ total elements are nonzero). Here, $\mathbf{I}_{N_t}$ is an $N_t \times N_t$ identity matrix and $\otimes$ indicates the *Kronecker product*, sometimes referred to as the *matrix direct product* [110].

Equation (8.4) shows that $\mathrm{SAR}(\mathbf{r}_m)$ is a purely quadratic function of the pulse samples. Further, the block-diagonal sparse $\mathbf{G}(\mathbf{r}_m)$ may be precomputed ahead of time for each $\mathbf{r}_m$ of interest. Finally, $\mathbf{q}_{\mathrm{tot}}$ is simply a *permutation* of $\mathbf{a}_{\mathrm{tot}}$ in (8.1), and thus $\mathbf{a}_{\mathrm{tot}} = \mathbf{Z}\mathbf{q}_{\mathrm{tot}}$ holds for a constant known permutation matrix $\mathbf{Z}$.

Now let us consider the ultimate goal of calculating local Ng SAR about location $\mathbf{r}_m$, denoted $\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}_m)$. Recalling that the cluster of R_m points that comprise an Ng cube about $\mathbf{r}_m$ are indexed by the set $I_N^{\mathbf{r}_m}$, we may write

$$\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}_m) = \frac{1}{R_m} \sum_{i \in I_N^{\mathbf{r}_m}} \mathrm{SAR}(\mathbf{r}_i). \tag{8.5}$$

Now, by substituting (8.4) into (8.5), we arrive at

$$\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}_m) = \frac{1}{R_m} \sum_{i \in I_N^{\mathbf{r}_m}} \mathbf{q}_{\mathrm{tot}}^{\mathrm{H}} \mathbf{G}(\mathbf{r}_i) \mathbf{q}_{\mathrm{tot}} = \mathbf{q}_{\mathrm{tot}}^{\mathrm{H}} \left(\frac{1}{R_m} \sum_{i \in I_N^{\mathbf{r}_m}} \mathbf{G}(\mathbf{r}_i) \right) \mathbf{q}_{\mathrm{tot}} = \mathbf{q}_{\mathrm{tot}}^{\mathrm{H}} \mathbf{Q}(\mathbf{r}_m) \mathbf{q}_{\mathrm{tot}}. \tag{8.6}$$

We have arrived at the single matrix that one needs, $\mathbf{Q}(\mathbf{r}_m)$, in order to compute local Ng SAR at location $\mathbf{r}_m$ due to the transmission of a set of pulses represented by $\mathbf{q}_{\mathrm{tot}}$. The matrix itself, being an average of the $\mathbf{G}(\mathbf{r}_i)$s, has the same highly-sparse, low-rank, redundant block-diagonal structure. With this formulation, we may now solve any optimization that explicitly constrains or regularizes Ng SAR. Note that since this is derived for general N, it is possible to construct a single matrix that computes whole-head mean SAR directly.

8.5 Closed-Form Solution for Mean-SAR Optimization

We first propose optimally minimizing the residual excitation error (i.e., maximizing the excitation fidelity) while limiting the whole-head mean SAR. Using the methodology of Sec. 8.4 and (8.5, 8.6), we construct $\mathbf{Q}$ that allows us to express *whole-head mean SAR* quadratically in $\mathbf{q}_{\mathrm{tot}}$. We then use the fact that $\mathbf{a}_{\mathrm{tot}} = \mathbf{Z}\mathbf{q}_{\mathrm{tot}}$ (where $\mathbf{Z}$ is the known

permutation matrix) to rewrite (8.1) as follows:

$$\mathbf{m} = \mathbf{A}\mathbf{a}_{\text{tot}} = \mathbf{A}\mathbf{Z}\mathbf{q}_{\text{tot}}. \tag{8.7}$$

Now, whereas Zhu solves the following optimization [168]:

$$\min_{\mathbf{q}_{\text{tot}}} \ \mathbf{q}_{\text{tot}}^{\text{H}}\mathbf{Q}\mathbf{q}_{\text{tot}} \quad \text{s.t. } \mathbf{d} = \mathbf{A}\mathbf{Z}\mathbf{q}_{\text{tot}}, \tag{8.8}$$

where $\mathbf{d}$ contains samples of the desired excitation, we instead propose a *relaxation* of (8.8)'s equality constraint

$$\min_{\mathbf{q}_{\text{tot}}} \ \|\mathbf{d} - \mathbf{A}\mathbf{Z}\mathbf{q}_{\text{tot}}\|_2^2 + \delta\mathbf{q}_{\text{tot}}^{\text{H}}\mathbf{Q}\mathbf{q}_{\text{tot}}, \tag{8.9}$$

which is a Tikhonov regularization [123, 124] that may be solved by a conjugate-gradient (CG) technique such as LSQR [100, 101] for a fixed $\delta \geq 0$. An optimal closed-form solution to (8.9) also exists for fixed $\delta \geq 0$:

$$\mathbf{q}_{\text{tot}}^{\text{opt}} = [(\mathbf{A}\mathbf{Z})^{\text{H}}(\mathbf{A}\mathbf{Z}) + \delta\mathbf{Q}]^{-1}(\mathbf{A}\mathbf{Z})^{\text{H}}\mathbf{d}. \tag{8.10}$$

Regardless of whether one solves for $\mathbf{q}_{\text{tot}}$ directly as in (8.10) or via a CG method, the optimal solution vector may then be substituted into both the residual error and mean SAR expressions ($\|\mathbf{d} - \mathbf{A}\mathbf{Z}\mathbf{q}_{\text{tot}}\|_2$ and $\mathbf{q}_{\text{tot}}^{\text{H}}\mathbf{Q}\mathbf{q}_{\text{tot}}$, respectively), which simply reduces them to functions of the scalar δ.

8.6 Exploring Excitation Fidelity, SAR and Pulse Duration Tradeoffs

Solving (8.9) for many values of $\delta \in [0, \infty)$ generates a curve of the lowest residual error $\|\mathbf{d} - \mathbf{A}\mathbf{Z}\mathbf{q}_{\text{tot}}\|_2$ achievable vs. mean SAR. This allows one to explore the entire range of fidelity-SAR tradeoffs. One may extend this concept by allowing pulse duration, L, to vary as well, i.e., turning overall pulse duration into a free variable. Then solving the system of equations for many (L, δ) pairs generates a contour of the best residual error achievable for a given mean SAR and pulse duration, illuminating fidelity vs. mean SAR vs. pulse duration tradeoffs.

8.7 Local N-Gram SAR Optimization

We now pose a constrained optimization that provides the highest-fidelity excitation while ensuring that the local Ng SARs at locations $\mathbf{r}_1, \ldots, \mathbf{r}_H$ are at most W_h W/kg:

$$\min_{\mathbf{q}_{\text{tot}}} \ \|\mathbf{d} - \mathbf{A}\mathbf{Z}\mathbf{q}_{\text{tot}}\|_2^2 \ \ \text{s.t.} \ \mathbf{q}_{\text{tot}}^{\text{H}}\mathbf{Q}(\mathbf{r}_1)\mathbf{q}_{\text{tot}} \leq W_1, \ldots, \mathbf{q}_{\text{tot}}^{\text{H}}\mathbf{Q}(\mathbf{r}_H)\mathbf{q}_{\text{tot}} \leq W_H, \tag{8.11}$$

where $\mathbf{Q}(\mathbf{r})$ is the matrix that computes local Ng SAR at location $\mathbf{r}$ as defined in (8.6). We may also easily include the whole-head mean SAR $\mathbf{Q}$ matrix and thus constrain mean SAR as well. To conclude this section we note that a form of (8.11) was recently implemented [57], but only for $H = 3$.

8.8 Implementation and Computational Discussion

The issue of most concern to regulatory bodies in both the United States [21] and Europe [70, 71] is *maximum* local Ng SAR. Therefore, one limitation that seems unavoidable from a design standpoint is that in order to design truly SAR-optimal pulses—ones where local SAR is guaranteed at *all* N_{model} spatial locations in the model—one is faced with the computationally-intractable problem of solving a system of equations with hundreds of thousands (millions) of quadratic constraints for moderate (high) resolution models; parallel computing might be one way to calculate SAR-optimal pulses in reasonable amounts of time [114].

8.9 Conclusion

Using novel derivations, we showed that local Ng SAR at any spatial point of interest $\mathbf{r}$ may be calculated using the highly-sparse, redundant-block-diagonal matrix $\mathbf{Q}(\mathbf{r})$ given in (8.6). We then posed a variety of optimization problems using such matrices, but unfortunately concluded that truly "SAR-optimal" pulse design remains, for the time being, a computationally intractable problem.

In Ch. 9, we will explore an approach that avoids computationally-intensive SAR-reduced design of a single pulse and instead time multiplexes a sequence of similar pulses to reduce maximum local SAR.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 9

Reduction of Maximum Local Specific Absorption Rate via Pulse Multiplexing

9.1 Introduction

In Ch. 7, we concluded that high local specific absorption rate (SAR) is a major safety concern of parallel excitation pulses, whereas in Ch. 8, we saw that reducing the SAR of a single pulse may be at times be computationally intractable from a design standpoint. Here we branch away from the optimization of a single pulse. The content of this chapter will soon be submitted to a journal [154].

A different SAR reduction approach: time multiplexing a set of pulses. The SAR-reduced pulse design techniques outlined in Sec. 2.1.15 and Ch. 8 all focus on designing a *single* pulse with favorable SAR characteristics; this pulse is then transmitted over many TRs to accomplish the imaging task for which it is intended. The distribution of SAR across space produced by this pulse may be determined exactly (along with mean and maximum local Ng SAR) by performing calculations involving the local electric field data and characteristics of the human model (see Sec. 7.2.4). But what if there were one or more pulses that accomplished essentially the same task (e.g., produced uniform thin-slice excitations), yet produced two significantly different distributions of SAR across space? For example, two pulses might each successfully excite a thin slice of tissue to within a

certain fidelity, yet the first might produce 1 and 3 W/kg of local SAR at locations A and B, whereas the second might produce 3 and 1 W/kg at these respective locations. If we transmit the first pulse repeatedly over many TRs, maximum local SAR equals 3 W/kg. This also holds true if we transmit the second pulse over many TRs. Consider, however, transmitting each pulse 50% of the time (e.g., playing the first pulse during odd TRs and the second pulse during even TRs): this produces a local SAR of 2 W/kg at both location A and location B.[1] *Time multiplexing* the two pulses has reduced maximum local SAR by a factor of 1.5; excitation quality differs from that of transmitting either pulse individually, but is not necessarily worse than each individual excitation because each pulse produces essentially the same excitation pattern. Based on a review of current literature, it seems multiplexing pulses to reduce SAR is a novel concept.

Purpose. The aim of this chapter is to develop an algorithm for determining the optimal multiplexing scheme (in the lowest maximum local SAR sense) when given a set of candidate pulses and develop the above thought experiment into a practical application: reducing the maximum local SAR of fast thin-slice excitation pulses transmitted through an eight-channel ($P = 8$) parallel transmission system. Although we focus on slice-selective pTX pulses, the concept of time multiplexing and the proposed algorithm are applicable to conventional single-channel systems and any type of pulse.

Overview. FDTD simulations [85] in a $1 \times 1 \times 2$ mm^3-resolution, 29-tissue head model [10] are conducted to obtain B_1^+ and electric field maps of an eight-channel transmit head array at a field strength of 7T as described in [8]. A field strength of 7 Tesla is chosen because many researchers are focusing on using pTX to mitigate B_1^+ inhomogeneity occurring at this strength.

The resulting B_1^+ maps are fed to a magnitude-least-squares (MLS) pulse design algorithm [112] to generate 162 candidate pulses, each comprised of two "spokes" (k_z traversals that ensure slice selection) [108, 113, 132] and capable of exciting a 1-cm-thick transverse slice of the brain located at $z = 0$ with a normalized root-mean-square error (NRMSE) of only 3.9 to 5.9% with respect to the desired target flip angle magnitude of 0.1 radians across the field of excitation (FOX). Each pulse's two spokes are placed at slightly different locations in k-space, parameterized by their distance from the k-space origin and their

[1] In this case, the overall local SAR that arises at point X is the *average* of the local SAR produced by each pulse at this point (see Sec. 9.2.5).

degree of rotation. SAR calculations are then conducted using the electric field data and whole-head, local 10g, and local 1g SAR due to each pTX pulse are calculated at each of the head model's 2.5 million spatial locations (voxels) and stored as real-valued data vectors.

While Bloch simulations confirm that the pulses all accomplish essentially the same excitation, the SAR calculation process reveals that the pulses at times create substantially different distributions of SAR across space. This occurs because by slightly perturbing the spoke locations in k-space, the design algorithm generates pulses with different magnitudes and phases, and these magnitudes and phases impact the superposition of the electric fields over time and thus the spatial SAR distribution.

Optimal pulse multiplexing in the lowest maximum local 1g SAR sense is then posed as a convex optimization [17, 97]—essentially, this algorithm efficiently searches over different pulse multiplexing schemes by evaluating linear combinations of the local SAR vectors and determines one that yields the lowest-possible maximum local 1g SAR. The outputs of the algorithm in this case are 162 weights that indicate the proportion of TRs to play each pulse; during the search, each weight is constrained to equal a nonnegative value between zero and one and the sum of the weights is enforced to always equal one—this ensures the resulting solution is a physically meaningful multiplexing scheme. The optimal multiplexing scheme in this case is shown to decrease maximum local SAR by a moderate amount while producing an excitation whose NRMSE fidelity falls within the fidelity range of the individual pulses' excitations.

Structure. Sec. 9.2 details the FDTD simulations, B_1^+ map generation, MLS pulse design, SAR calculation, and the multiplexing optimization algorithm. Sec. 9.3 depicts and discusses the results when applying the multiplexing technique to slice-selective eight-channel pTX pulses at 7T. Final remarks are given in Sec. 9.4.

9.2 Materials and Methods

9.2.1 Human Head and Eight-Channel Parallel Transmit Array Model

The model used here is described in Sec. 7.2.1 and depicted in Fig. 7-1.

9.2.2 Electromagnetic Field Simulations and B_1^+ Field Map Generation

Using the approach outlined in Sec. 7.2.2, we obtain the steady-state electric field per ampere of input produced by each of the $P = 8$ transmit channels, i.e., for $p = 1, \ldots, P$, we obtain $\mathbf{E}_p(\mathbf{r}) = [E_{p,x}(\mathbf{r}), E_{p,y}(\mathbf{r}), E_{p,z}(\mathbf{r})]^{\mathrm{T}}$ (V/m/A). We also compute the B_1^+ field profiles $S_1(\mathbf{r}), \ldots, S_P(\mathbf{r})$ (T/A) whose inhomogeneous in-plane magnitudes are depicted in Fig. 7-2. The pulse design process described next will essentially find an efficient way to modulate these profiles in both magnitude and phase such that their combination over time yields a uniformly-flat excitation within a thin slice of the head through the $z = 0$ plane.

9.2.3 Magnitude Least Squares (MLS) Multi-Channel Pulse Design

Overview. We now summarize the key points of [112] and describe one way to design a set of P RF pulse shapes to concurrently play through the P elements of a P-channel parallel transmission array in order to generate a user-defined target excitation, $d(\mathbf{r})$. Note that "pTX pulse" will be used as shorthand for "a set of P concurrently-transmitted RF pulse shapes that yields a thin-slice excitation".

Spoke-based pulses. We will use small-tip-angle slice-selective spoke-based pulses [108, 113, 132], descriptions of which are provided in Sec. 2.1.7.

Design parameters. For all upcoming simulations, we design pTX pulses that excite a 1-cm-thick transverse slice of the head; the FOX in each case is wherever the head is present within the slice. The desired excitation, $d(\mathbf{r})$, has a 0.1 radian flip angle at spatial locations in the FOX and a phase that varies slowly enough across space such that intravoxel dephasing is not a concern (slowly-varying phase is permissible for the majority of standard imaging scenarios where one is only interested in image magnitude). The MLS algorithm will produce pulses that achieve excitations close to the desired one. The k-space trajectory of each pulse consists of two spokes symmetrically spaced by a fixed distance away from the origin and rotated about the origin within the (k_x, k_y) plane by a specific amount. Twenty-seven radii choices for each pair are evaluated, along with six possible rotations, leading to 162 total two-spoke pulses.[2] Figure 9-1 provides an overhead (k_x, k_y)-space view of the 162 spoke pairs—the pair of gray boxes indicates the positions of the 52nd pulse's two spokes. The gradients are always constrained to amplitude and slew rates of 15 mT/m and 150

[2]Because eight transmit channels are present, only two spokes are needed to form uniform excitations, illustrating the strength of pTX systems.

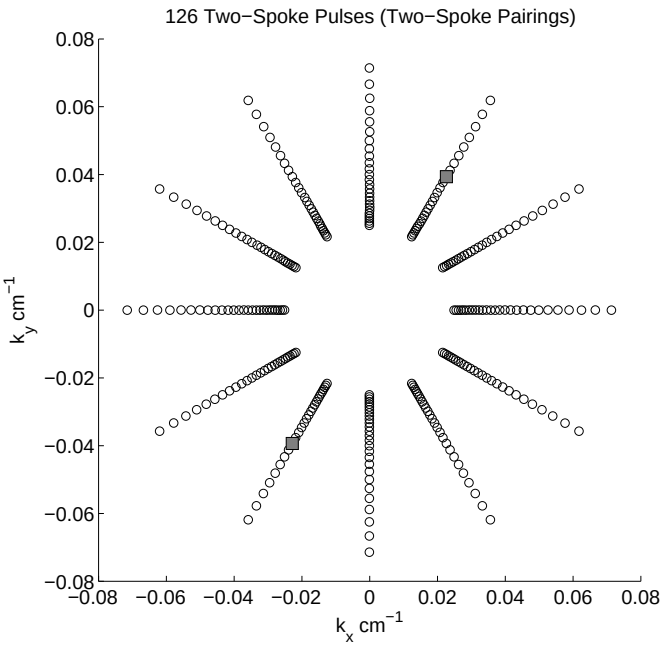

Figure 9-1: **Overhead view of spoke pairs placed parallel to the k_z axis**. The placements of spokes in (k_x, k_y) for each two-spoke pulse is displayed. Each pulse's trajectory consists of two spokes spaced by a fixed distance away from the origin and rotated within the (k_x, k_y) plane by a specific amount. Twenty-seven radii choices for each pair are evaluated, along with six possible rotations, leading to 162 total two-spoke pulses. The pair of gray boxes indicates the positions of the 52nd pulse's two spokes; details about this specific pulse are given in Fig. 9-2.

T/m/s, such that all pulses have durations of 2.87 to 2.91 ms.

Algorithm setup. To begin, assume we have decided where to place the pair of spokes. This means that the k-space trajectory, $\mathbf{k}(t) = [k_x(t), k_y(t), k_z(t)]^{\mathrm{T}}$, is fixed, along with the gradient waveforms, $\mathbf{G}(t) = [G_x(t), G_y(t), G_z(t)]^{\mathrm{T}}$. Also assume that each channel's B_1^+ profile, $S_p(\mathbf{r})$, is also known. Now, by assuming a small-tip angle excitation [102], we may apply the linear formalism of [59] to arrive at (2.13) as described in Sec. 2.1.11. We then discretize (2.13) by sampling space at locations $\mathbf{r}_1, \ldots, \mathbf{r}_{N_s}$ within the FOX and sampling time at $t_1, \ldots, t_{N_t} \in [0, L]$; the t_i are spaced uniformly by $\Delta_t = 10\mu$s, yielding (2.14), restated here:

$$\mathbf{m} = \mathbf{S}_1\mathbf{F}\mathbf{a}_1 + \cdots + \mathbf{S}_P\mathbf{F}\mathbf{a}_P = \mathbf{A}_{\mathrm{tot}}\mathbf{a}_{\mathrm{tot}}, \tag{9.1}$$

where we have used $\mathbf{a}_p$s (vectors of current samples) in place of $\mathbf{b}_p$s (vectors of voltage samples).

MLS algorithm. We now apply MLS to generate P RF pulse shapes that will excite a pattern close to the desired $d(\mathbf{r})$. We form N_s samples of the desired pattern into the vector $\mathbf{d} \in \mathbb{C}^{N_s}$. In this case, because we desire a uniform excitation, every element of $\mathbf{d}$ is fixed to equal 0.1 radians, but no phase is specified per element because a slowly-varying phase will be inherently determined by the MLS algorithm. On a conceptual level, MLS then solves the following optimization:

$$\min_{\mathbf{a}_{\text{tot}}} \{\|\mathbf{d} - |\mathbf{A}_{\text{tot}}\mathbf{a}_{\text{tot}}|\|_2^2 + \delta\|\mathbf{a}_{\text{tot}}\|_2^2\}, \tag{9.2}$$

where δ is a small nonnegative value that penalizes pulse candidates with excessive ℓ_2 pulse energies. It is crucial to note the absolute value operator surrounding $\mathbf{A}_{\text{tot}}\mathbf{a}_{\text{tot}}$: this means that the resulting phase of the excitation is given no heed during the optimization procedure and that (9.2) seeks only to find a set of pulse shapes that produce an excitation whose magnitude at all spatial locations is close to the desired magnitude. Therefore, in order to prevent rapidly-varying phase in the resulting excitation, $m(\mathbf{r})$, and avoid detrimental intravoxel dephrasing, the above optimization is solved a number of times, modifying the initial input to the optimization each time as described in [112]. Ultimately, this iterative process yields a solution $\mathbf{a}_{\text{tot}}$ whose pulse samples, when evaluated with a Bloch equation [14] simulator, produce a high-fidelity version of the desired magnitude pattern and whose phase variation is always less than 8 degrees per millimeter. For all 162 pulse designs, the NRMSE between the resulting and desired flip angle magnitude, defined as

$$\text{NRMSE} = 100 \cdot \frac{\|\mathbf{d} - |\mathbf{A}_{\text{tot}}\mathbf{a}_{\text{tot}}|\|_2}{\|\mathbf{d}\|_2} \ (\%), \tag{9.3}$$

always falls between 3.9 and 5.9%; that is, MLS successfully ensures that every two-spoke pulse it designs produces an essentially uniform excitation (even though the RF pulse shapes of each design may at times vary significantly in magnitude and phase due to their spoke position differences in (k_x, k_y)-space).

MLS pulse example. Details about the 52nd pulse designed by the MLS algorithm are given in Fig. 9-2. The Bloch-simulated magnitude, $|m(\mathbf{r})|$, and Bloch-simulated phase, angle$\{m(\mathbf{r})\}$, are displayed along the upper row. The flip angle is highly uniform across space and the phase is smoothly varying, as intended. The lower-left subplot depicts the two-spoke trajectory in 3D, while the lower-right subplots show the magnitude of the first

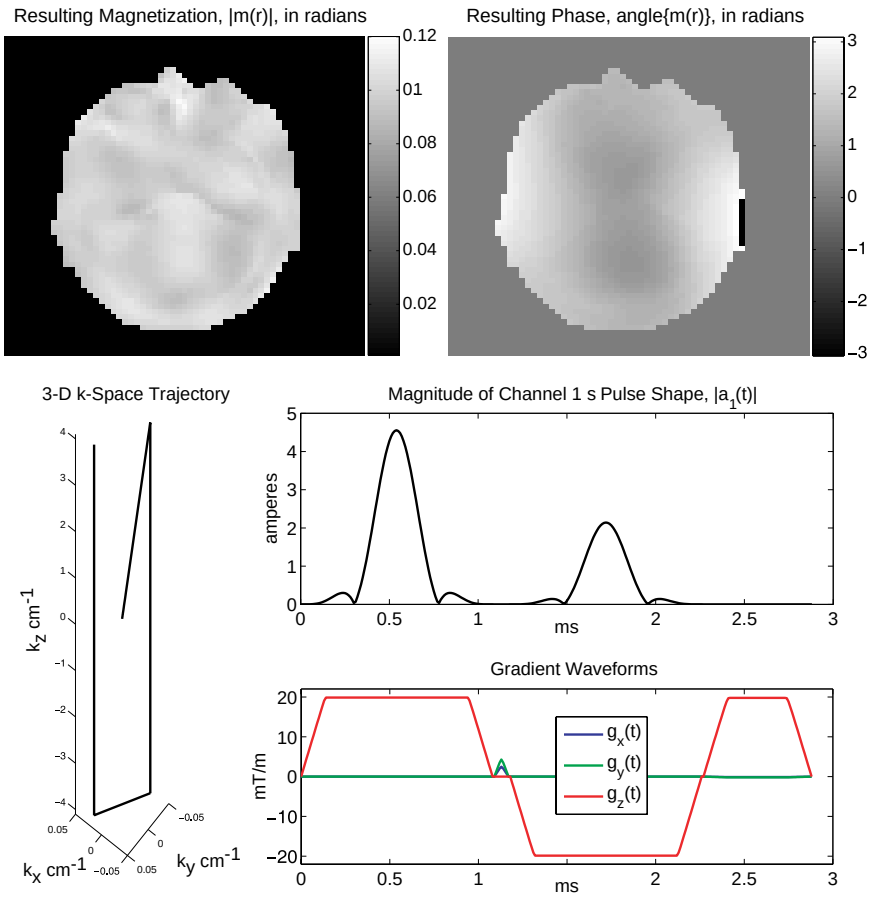

Figure 9-2: **Pulse #52 designed using MLS algorithm**. Upper row: Bloch-simulated magnitude and phase of the excitation that arises due to the application of the pulse; the flip angle is uniform across space and phase varies slowly, as intended. Lower-left subplot: 3D k-space trajectory. Lower-right subplots: magnitude of the first RF pulse shape, $|a_1(t)|$, and the gradient waveforms, $\mathbf{G}(t)$. Fig. 9-1's gray boxes indicate the locations of this pulse's two spokes in (k_x, k_y) space.

RF pulse shape, $|a_1(t)|$, along with the gradient waveforms, $\mathbf{G}(t)$. The locations of this pulse's two spokes in (k_x, k_y)-space are indicated by the gray boxes in Fig. 9-1.

9.2.4 Whole-Head and Local Specific Absorption Rate Calculations

Overview. After designing a pTX pulse (i.e., P RF pulse shapes) to produce a desired excitation, we must determine the global and local SAR the pulse produces in the head model. At this point, we know N_t time samples of each pulse shape spaced uniformly in time by Δ_t, i.e., for $p = 1, \ldots, P$, we know $a_p(n\Delta_t)$ for $n \in \{0, 1, \ldots, N_t - 1\}$. Recall from (2.13) that L is the duration of each pulse shape and thus $L = N_t\Delta_t$.

SAR per voxel. We first calculate SAR (W/kg) at each location $\mathbf{r}$, denoted SAR($\mathbf{r}$), via numerical integration as given in (2.22) of Sec. 2.1.14, where we assume that the P pulse shapes are concurrently transmitted from time 0 to L and nothing is transmitted during the time interval (L, TR). (See end of subsection regarding choice of TR and duty cycle.)

Whole-head and local N-gram SAR. Having obtained SAR($\mathbf{r}$) for all $\mathbf{r}$ in the head, whole-head global SAR is obtained by averaging the SAR($\mathbf{r}$) values. Likewise, Ng SAR at each $\mathbf{r}$ is obtained by finding an N-gram cube around each $\mathbf{r}$ and then averaging SAR($\mathbf{r}$) over all $\mathbf{r}$ within the cube, in line with [1]. We accomplish this rapidly by using the Ng region growth technique described in Sec. 7.2.4. Recall that for each location $\mathbf{r}$, this algorithm rapidly determines the cluster of points that comprise an N-gram cube around $\mathbf{r}$, and that this cluster may be perceived as a set of indices, $I_N^{\mathbf{r}}$, such that the mass equation (7.4) holds. Local Ng SAR at $\mathbf{r}$, denoted $\mathrm{SAR}^{N\mathrm{g}}$, is computed formally as:

$$\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}) = \frac{1}{|I_N^{\mathbf{r}}|} \sum_{i \in I_N^{\mathbf{r}}} \mathrm{SAR}(\mathbf{r}_i) \tag{9.4}$$

where $|I_N^{\mathbf{r}}|$ is the *cardinality* (number of elements) of $I_N^{\mathbf{r}}$.

Fixed repetition time per pulse. The effect of repetition time (TR) is accounted for when calculating the SAR of each pulse by setting TR equal to 20 ms for all pulses. Because all pulses are essentially $L = 2.9$ ms long, this corresponds to a constant duty cycle $D = L/\mathrm{TR}$ of 14.5%. This choice of TR is arbitrary in the sense that the optimal multiplexing scheme is invariant to the choice of TR.[3] Further, because the SAR of each pulse decreases linearly with TR [see (2.20, 2.22)], all *relative* SAR improvements stated in Sec. 9.3 hold for any physically-valid TR.

9.2.5 Pulse Time-Multiplexing for SAR Reduction (PTMSR)

Overview. Here we derive a convex optimization algorithm that determines how to time-multiplex a set of J candidate pulses such that the resulting maximum local Ng SAR is minimized as best as possible. Assume each pulse j produces approximately the same excitation, but none of the pulses are exactly equal to any of the others (i.e., there are at least slight magnitude and phase differences among the pulses). Because of this, the

[3] Of course, the choice of TR must be realistic—equal to or longer than the duration of the longest pulse among those being optimized over.

spatial distribution of SAR across space due to pulse j will not exactly equal the spatial distribution of SAR across space due to the other $J-1$ pulses. Thus, analogously to the thought experiment of Sec. 9.1 that involved two pulses, we may be able to multiplex these J pulses to reduce maximum local SAR.

Objective. The goal is to compute J weights, $w_j \in [0,1], j \in \{1,\ldots,J\}$, subject to the physically-realistic constraint that the weights sum to unity. Each w_j indicates the proportion of TRs to play pulse j assuming an infinite number of TRs are used. For example, if $w_{\mathrm{A}} = 0.837$, $w_{\mathrm{B}} = 0.163$, and all other w_js equal zero, then the optimal multiplexing scheme involves playing pulse A 83.7% of the time and pulse B 16.3% of the time. Since the number of TRs one employs to conduct an imaging task is always finite, these weights may simply be rounded off to obtain a nearly-optimal realistic solution, e.g., if 100 TRs are necessary, one would play pulse A eighty-four times and pulse B sixteen times. It is up to the user to choose the exact order in which to play the pulses.

Initialization. For $j = 1,\ldots,J$, we use the method of Sec. 9.2.4 to calculate the local Ng SAR produced by the jth pulse at spatial locations $\mathbf{r}_1,\ldots,\mathbf{r}_M$ in the head model, yielding $\mathrm{SAR}_j^{N\mathrm{g}}(\mathbf{r}_m)$ in W/kg at M locations. We then construct the following M-element vector:

$$\mathbf{s}_j = [\mathrm{SAR}_j^{N\mathrm{g}}(\mathbf{r}_1),\ldots,\mathrm{SAR}_j^{N\mathrm{g}}(\mathbf{r}_M)]^{\mathrm{T}}. \tag{9.5}$$

These vectors are useful representations, e.g., $\|\mathbf{s}_j\|_\infty$ is the maximum local Ng SAR produced by pulse j.

SAR due to the transmission of H pulses. We now derive the SAR that arises at location $\mathbf{r}$ when H pulses are played over H TRs. Note that we are not necessarily playing H *different* pulses. During the hth TR, pulse h, whose duration equals L_h, is played from time $(h-1)\mathrm{TR}$ through $(h-1)\mathrm{TR}+L_h$ and nothing is transmitted during the time interval beginning at $(h-1)\mathrm{TR}+L_h$ and ending at $h\mathrm{TR}$. By applying (2.20), we know the "standalone" SAR for this pulse, which we denote $\mathrm{SAR}_h(\mathbf{r})$. Moving forward, the SAR equation for this scenario is:

$$\mathrm{SAR}(\mathbf{r}) = \frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})}\frac{1}{H\cdot\mathrm{TR}}\int_0^{H\cdot\mathrm{TR}} \|\mathbf{E}(\mathbf{r},t)\|_2^2\,dt. \tag{9.6}$$

We now expand (9.6) and then apply (2.20) to each subinterval:

$$\begin{aligned}\mathrm{SAR}(\mathbf{r}) &= \frac{1}{H}\left(\frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})}\frac{1}{\mathrm{TR}}\int_0^{\mathrm{TR}}\|\mathbf{E}(\mathbf{r},t)\|_2^2\,dt+\cdots+\frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})}\frac{1}{\mathrm{TR}}\int_{(H-1)\cdot\mathrm{TR}}^{H\cdot\mathrm{TR}}\|\mathbf{E}(\mathbf{r},t)\|_2^2\,dt\right)\\ &= \frac{1}{H}\left(\mathrm{SAR}_1(\mathbf{r})+\cdots+\mathrm{SAR}_H(\mathbf{r})\right)\\ &= \frac{1}{H}\sum_{h=1}^{H}\mathrm{SAR}_h(\mathbf{r}).\end{aligned} \tag{9.7}$$

We have obtained the overall *point* SAR at $\mathbf{r}$ due to the train of H pulses. Now we seek to determine the Ng SAR that will arise at $\mathbf{r}$ when we apply this same pulse train. Recalling from Sec. 9.2.4 that for each $\mathbf{r}$, we know of an index set $I_N^{\mathbf{r}}$ that properly indexes an Ng cube about $\mathbf{r}$, we substitute (9.7) into (9.4) and shuffle the two resulting summations:

$$\begin{aligned}\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}) &= \frac{1}{|I_N^{\mathbf{r}}|}\sum_{i\in I_N^{\mathbf{r}}}\mathrm{SAR}(\mathbf{r}_i)\\ &= \frac{1}{|I_N^{\mathbf{r}}|}\sum_{i\in I_N^{\mathbf{r}}}\left(\frac{1}{H}\sum_{h=1}^{H}\mathrm{SAR}_h(\mathbf{r}_i)\right)\\ &= \frac{1}{H}\sum_{h=1}^{H}\left(\frac{1}{|I_N^{\mathbf{r}}|}\sum_{i\in I_N^{\mathbf{r}}}\mathrm{SAR}_h(\mathbf{r}_i)\right)\\ &= \frac{1}{H}\sum_{h=1}^{H}\mathrm{SAR}_h^{N\mathrm{g}}(\mathbf{r}).\end{aligned} \tag{9.8}$$

Eq. (9.8) thus proves that the *local* Ng SAR arising at $\mathbf{r}$ due to the H-pulse train is the *average* of the local Ng SARs produced at this same location by each of the individual pulses.

Let us conclude our derivations by assuming we play a very long train of pulses, i.e., we let H tend toward ∞. Returning to our set of J candidate pulses, let us assume in the limit we play pulse j a fraction $w_j \in [0,1]$ of the total number of TRs, imposing the constraint $\sum_{j=1}^{J} w_j = 1$ to ensure the scheme is physically sensible, i.e., *we assume some fixed, valid multiplexing scheme* for the J pulses. Under this multiplexing scheme, (9.8) becomes:

$$\mathrm{SAR}^{N\mathrm{g}}(\mathbf{r}) = \sum_{j=1}^{J} w_j \cdot \mathrm{SAR}_j^{N\mathrm{g}}(\mathbf{r}). \tag{9.9}$$

Eq. (9.9) shows that the overall local Ng SAR arising at $\mathbf{r}$ due to a particular multiplexing scheme of J pulses over many TRs is simply the *weighted average* of the J pulses' local Ng SARs based on the fraction of total TRs each pulse is transmitted.

Finally, having derived the local Ng SAR that arises at a single location $\mathbf{r}$, we propose a linear-algebraic way to compute the local Ng SAR at *all* M spatial locations $\mathbf{r}_1, \ldots, \mathbf{r}_M$ in the head model that arises due to the multiplexing scheme of the J pulses represented by the w_js:

$$\begin{aligned}
\mathbf{s}_{\text{overall}}^{Ng} = \begin{bmatrix} \text{SAR}^{Ng}(\mathbf{r}_1) \\ \vdots \\ \text{SAR}^{Ng}(\mathbf{r}_M) \end{bmatrix} &= \begin{bmatrix} w_1 \cdot \text{SAR}_1^{Ng}(\mathbf{r}_1) + \cdots + w_J \cdot \text{SAR}_J^{Ng}(\mathbf{r}_1) \\ \vdots \\ w_1 \cdot \text{SAR}_1^{Ng}(\mathbf{r}_M) + \cdots + w_J \cdot \text{SAR}_J^{Ng}(\mathbf{r}_M) \end{bmatrix} \\
&= \begin{bmatrix} \text{SAR}_1^{Ng}(\mathbf{r}_1) \\ \vdots \\ \text{SAR}_1^{Ng}(\mathbf{r}_M) \end{bmatrix} w_1 + \cdots + \begin{bmatrix} \text{SAR}_1^{Ng}(\mathbf{r}_1) \\ \vdots \\ \text{SAR}_1^{Ng}(\mathbf{r}_M) \end{bmatrix} w_J \\
&= \mathbf{s}_1 w_1 + \cdots \mathbf{s}_J w_J \\
&= \begin{bmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_J \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_J \end{bmatrix} \\
&= \mathbf{S}\mathbf{w},
\end{aligned} \tag{9.10}$$

where $\mathbf{s}_j = [\text{SAR}_j^{Ng}(\mathbf{r}_1), \ldots, \text{SAR}_j^{Ng}(\mathbf{r}_M)]^{\text{T}}$ [as in (9.5)], $\mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_J]$, and $\mathbf{w} = [w_1, \ldots, w_J]^{\text{T}}$.

Optimal pulse multiplexing algorithm. We now arrive at the core problem: given the J candidate pulses, find the optimal choice of w_js that best minimizes the maximum local Ng SAR that arises when applying the pulses over many TRs. The following optimization solves this problem in a globally optimal sense:

$$\min_{\mathbf{w}} \ \|\mathbf{S}\mathbf{w}\|_\infty \ \text{ s.t. } w_j \in [0,1] \text{ for } j \in 1, \ldots, J \text{ and } \sum_{j=1}^{J} w_j = 1. \tag{9.11}$$

This is a convex optimization [17,97]. By solving (9.11) we obtain $\mathbf{w}$ that is the best-possible minimizer of the objective function. After running the algorithm, one extracts the w_js from $\mathbf{w}$ and knows the multiplexing scheme to use that best minimizes local Ng SAR.

Software implementation. We implement (9.11) in SeDuMi (Self-Dual-Minimization)

[120], a free software package consisting of `MATLAB` and `C` routines. Since the true M of the head model equals 2.5 million, we downsample the $\mathbf{s}_j$ vectors from a resolution of $1\times1\times2$ mm^2 to $4\times4\times4$ mm^2, retaining the largest SAR value per downsampled region. This reduces the $\mathbf{s}_j$s to roughly 75,000 elements each and allows the optimization to be solved in 8 minutes on a Linux server with a 3.0-GHz Intel Pentium IV processor; the random access memory footprint of this process is well under one gigabyte. After determining $\mathbf{w}$ that yields the optimal multiplexing scheme in this worst-case lower-resolution setting, we make use of the original *full-resolution* $\mathbf{s}_j$s for all local SAR computations and figures given in Sec. 9.3 to present results as fairly and realistically as possible.

9.3 Experimental Results

Overview. Having created the 162 candidate pulses in Sec. 9.2.3 and derived the multiplexing algorithm in Sec. 9.2.5, we now investigate the SAR properties of these pulses, examine their local SAR distributions across space, determine the best individual pulse (in the lowest maximum SAR sense), compute the optimal multiplexing scheme of the 162 pulses, verify that the multiplexing scheme produces a high-fidelity excitation, and compare how well the multiplexed pulse scheme reduces maximum local SAR relative to the maximum local SAR produced simply by playing the best individual pulse among the original 162.

Mean and maximum local SARs of the 162 individual candidate pulses. Assume we transmit pulse j (and no other pulse) over many TRs. The resulting mean, maximum local 10g, and maximum local 1g SARs that arise in this situation are plotted in Fig. 9-3. Although the mean SARs of the pulses are similar, their maximum local SARs exhibit significant variation: some pulses are clearly superior to others in the maximum local SAR sense, even though all pulses produce similar high-fidelity excitations with NRMSEs of only 3.9 to 5.9%. Among the pulses, maximum 1g SAR ranges from 3.8 to 10.2 times greater than corresponding mean SAR. Likewise, maximum 10g SAR is always 2.6 to 5.8 times larger than mean SAR. If we are restricted to using only one pulse, #52 is the optimum choice because it produces a maximum local 1g SAR of only 0.083 W/kg, the lowest among all pulses.

Correlations of the spatial distribution of local 1g SAR of the 162 candidate pulses. We now analyze the extent to which the local 1g SAR distributions of the pulses are

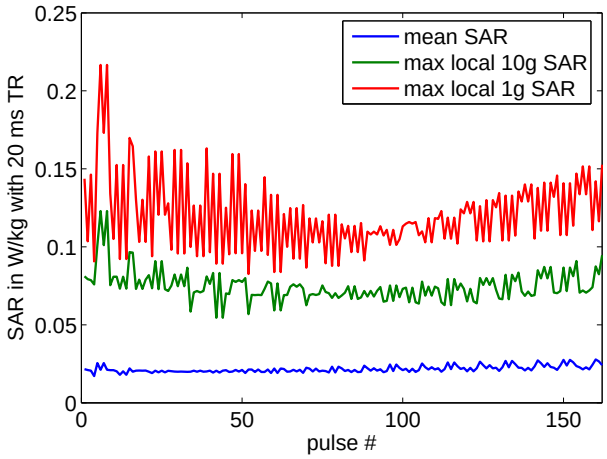

Figure 9-3: **Mean and maximum local SARs of each individual candidate pulse**. These SAR values arise when pulse j alone is transmitted over many TRs. Pulse 52 is the best individual pulse among the 162 evaluated in the maximum local 1g SAR sense, producing a maximum local 1g SAR of only 0.083 W/kg.

correlated across space by computing the Pearson product-moment correlation coefficient between the local 1g SAR vectors of each pair of pulses. Formally, for each pair (j, k) of M-element local 1g SAR vectors $(\mathbf{s}_j, \mathbf{s}_k)$, we compute

$$\rho_{j,k} = \frac{1}{M-1} \frac{(\mathbf{s}_j - \mu_j)^{\mathrm{T}} (\mathbf{s}_k - \mu_k)}{\sigma_j \sigma_k}, \tag{9.12}$$

where μ_j and σ_j indicate the mean and standard deviation of the M elements of $\mathbf{s}_j$. A value of $\rho_{j,k}$ close to unity means that the local 1g SAR patterns produced by pulses j and k across space are essentially identical, whereas $|\rho_{j,k}| \approx 0$ indicates the SAR distributions radically differ across space. Pulses that produce uncorrelated SAR patterns might be good candidates in the multiplexing sense. The overall correlation matrix (a display of all $\rho_{j,k}$s) is given in Fig. 9-4. Many pulse pairs produce highly-correlated distributions of local 1g SAR across space, while the lowest observed correlation is 70.6% and the average correlation is 87.0%.

Closer look: distributions of local 1g SAR across space arising due to the application of three pulses. The local 1g SAR distributions arising due to pulses 13, 118, and 160 are depicted in Fig. 9-5. All SAR maps are displayed using the same dynamic range, permitting direct comparisons among the maps. Furthermore, because a local 1g SAR data point exists for each of the voxels in the 3D head model, the 3D data has been displayed in

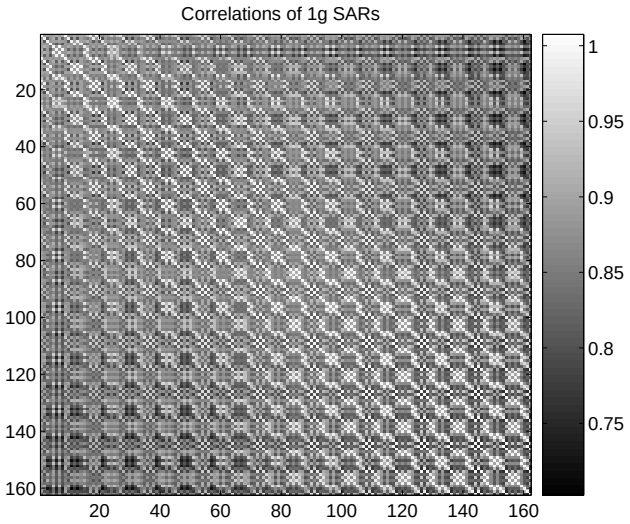

Figure 9-4: **Correlations of the spatial distributions of local 1g SAR produced by pairs of pulses**. For each pair of pulses (j, k), the extent of similarity between the distributions of local 1g SAR each pulse produces across space is quantified by computing the Pearson correlation coefficient between the corresponding $\mathbf{s}_j$ and $\mathbf{s}_k$ vectors. The pairwise vector correlation, on average, is 87.0%.

2D by "collapsing" SAR values along the spatial z axis. Namely, for each location (x, y) in pulse j's map, we have displayed $\max_z \text{SAR}_j^{1g}(x, y, z)$. Dashed boxes indicate the hotspot where local 1g SAR is largest. The distribution of local SAR indeed varies across space depending on which excitation pulse is used. The correlation coefficients $\rho_{13,118}$, $\rho_{13,160}$, and $\rho_{118,160}$ equal 81, 71, and 78%, respectively.

Optimal time multiplexing scheme. We now feed $\mathbf{s}_1, \ldots, \mathbf{s}_{162}$ to the multiplexing algorithm as described in Sec. 9.2.5 and obtain the optimized weights $w_1, \ldots, w_{162}$; these weights are displayed in Fig. 9-6. Only eight pulses have been utilized (less than 5% of available candidates) even though (9.11) does not penalize the use of many pulses in the multiplexing scheme. We convert the weights from percentiles into real values for use in a 61-TR scenario where we will collect 61 lines of k_y data in readout k-space. For each of the eight pulses whose weights are nonzero in Fig. 9-6, we use each one for 8, 8, 16, 15, 2, 2, 6, and 4 TRs, respectively. The overall resulting smooth excitation due to the multiplexing scheme is depicted in Fig. 9-7: its NRMSE of 5.5% is within the NRMSE range of the 8 excitations due to the individual pulses belonging to the multiplexing scheme, confirming its high degree of fidelity. Using (9.9, 9.10), we find that the maximum local 1g SAR that arises due to this multiplexing scheme is 0.075 W/kg; relative to the best individual pulse, we

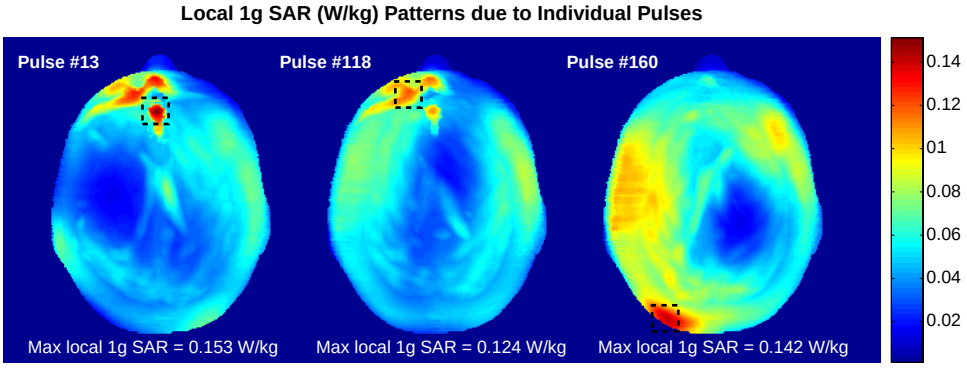

Figure 9-5: **Local 1g SAR distributions across space due to three pulses**. All maps have the same dynamic range. The 3D SAR data has been displayed in 2D by "collapsing" the spatial z axis, i.e., for each (x, y) coordinate in pulse j's map, $\max_z \text{SAR}_j^{1g}(x, y, z)$ is displayed. Dashed boxes indicate the (x, y) location where local 1g SAR is largest. The correlation coefficients $\rho_{13,118}$, $\rho_{13,160}$, and $\rho_{118,160}$ equal 0.81, 0.71, and 0.78, respectively.

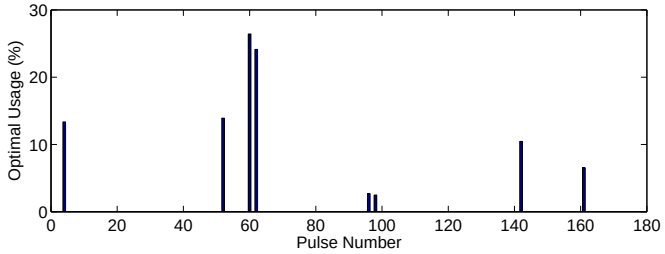

Figure 9-6: **Optimal time multiplexing scheme**. The optimized weights $w_j, j = 1, \ldots, 162$ as determined by the optimization (9.11) are displayed as percentages. Only eight of the 162 pulses are utilized in the multiplexing scheme. The resulting maximum local 1g SAR is 1.11 times lower than the maximum local 1g SAR of the best individual pulse. (See Table 9.1.)

have decreased maximum local 1g SAR by a factor of 1.11 with no loss in excitation quality relative to that of the candidate pulse set. The SAR statistics of the optimal multiplexing scheme vs. those of the best individual pulse are summarized in Table 9.1.

Local 1g SAR distribution of the time-multiplexed pulse scheme vs. that of the best individual pulse. Finally, we compare the local 1g SAR map of the best individual pulse to that which arises when the optimal multiplexing scheme is used. Fig. 9-8 displays these SAR maps side-by-side using the same dynamic range to allow for direct comparisons. The 3D SAR data has been collapsed along the z axis as in Fig. 9-5 and hatched boxes indicate the (x, y) location of the hotspot. The SAR maps are indeed similar, but when the multiplexing scheme is used the hotspot shifts its position and, most importantly,

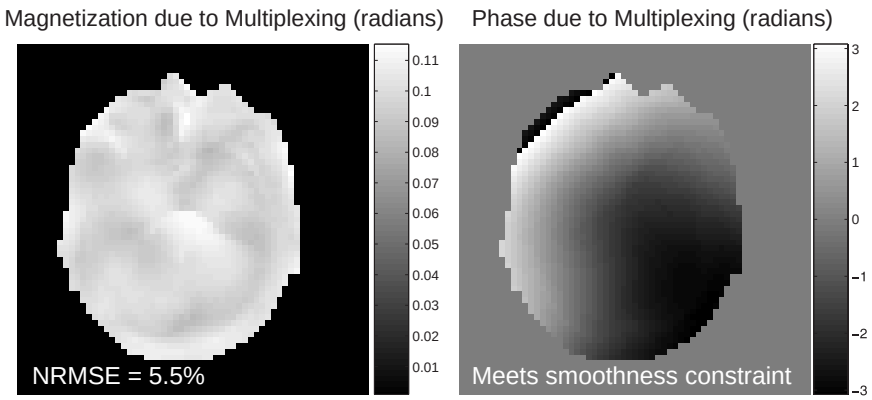

Figure 9-7: **Resulting excitation due to time multiplexing scheme**. We convert the eight optimized nonzero weights given in Fig. 9-6 from percentiles into real values for use in a 61-TR scenario where we will collect 61 lines of k_y data in readout k-space. For each of the eight pulses whose weights are nonzero in Fig. 9-6, we use each one for 8, 8, 16, 15, 2, 2, 6, and 4 TRs, respectively. The resulting excitation's magnitude, $|m(\mathbf{r})|$, is fairly uniform; its NRMSE is similar to those of the 162 individual candidates.

SAR Characteristics	**Best Indiv. Pulse**	**Opt. Multiplexing Scheme**
Mean SAR (W/kg)	0.021	0.022
Max. Local 10g SAR (W/kg)	0.057	0.064
Max. Local 1g SAR (W/kg)	0.083	0.075

Table 9.1: **SAR characteristics of the best individual pulse vs. those of the optimal multiplexing scheme**. The multiplexing scheme reduces maximum local 1g SAR by a factor of 1.11. Multiplexing indeed increases local 10g SAR by roughly the same factor, but this is permissible: the local 1g SAR, being the highest value in both cases, is the SAR characteristic of most concern and the limiting factor in this scenario.

is reduced in magnitude.

9.4 Conclusion

Using an optimization to determine the optimal time-multiplexing scheme of a set of candidate pulses reveals a small subset of pulses that, when played over a number of TRs, yields not only a high-fidelity excitation (5.5% NRMSE) but 1.11 times lower maximum local 1g SAR than does transmitting simply the best individual pulse. This concept was applied to slice-selective parallel transmission pulses, but readily applies to conventional single-channel transmission systems and other pulse types.

The concept of time multiplexing to reduce SAR, the optimal multiplexing convex optimization algorithm, and the linear-algebraic derivations of local SAR calculations all seem

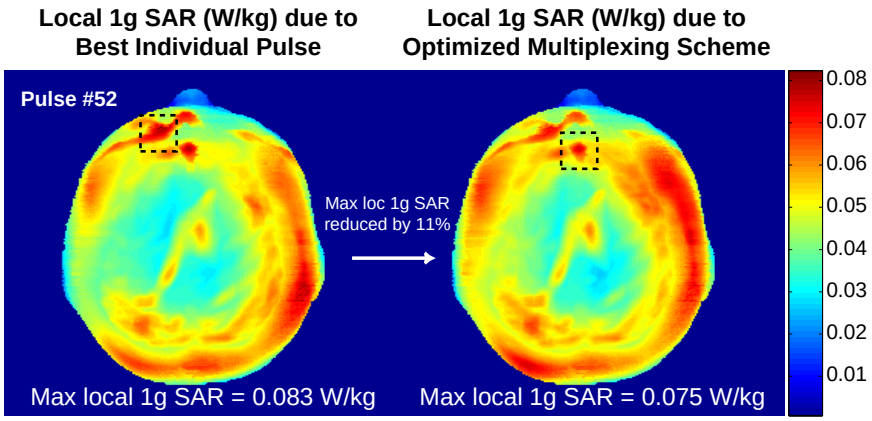

Figure 9-8: **Local 1g SAR distributions of best standalone pulse and time-multiplexed scheme**. The SAR maps due to the best individual pulse and the multiplexing scheme are displayed side-by-side. The 3D SAR data is collapsed along z axis as in Fig. 9-5. Hatched boxes indicate where local 1g SAR is highest. The multiplexing scheme shifts the hotspot and reduces it by a factor of 1.11.

to be novel contributions.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 10

Summary and Recommendations

10.1 Summary

Chapter 2 presented background information on MRI excitation theory, conventional single-channel and multi-channel pulse design, and sparse approximation.

Chapter 3 proposed the novel multiple-system, single-output (MSSO) simultaneous sparsity problem. Three greedy techniques were developed to approximately solve the problem (matching pursuit, orthogonal matching pursuit, and least squares matching pursuit) along with four methods based on a convex relaxation (iteratively reweighted least squares, two forms of iterative shrinkage, and formulation as a second-order cone program). While deriving the algorithms, we proved that *seeking a single sparse complex-valued vector is equivalent to seeking two* **simultaneously sparse** *real-valued vectors*, increasing the relevance and applicability of MSSO theory. The MSSO algorithms were then evaluated across noiseless and noisy sparsity profile estimation experiments as well as a magnetic resonance imaging pulse design experiment. For sparsity profile recovery, algorithms that minimized the relaxed convex objective function outperformed the greedy methods, whereas in the noiseless magnetic resonance imaging pulse design experiment, greedy least-squares matching pursuit exhibited superior performance.

Chapter 4 studied three algorithms for solving linearized systems of RF waveform design equations for calculating accelerated spatially-tailored excitations on parallel excitation MRI systems. Two iterative conjugate gradient (CG) methods, Least Squares QR (LSQR) and Conjugate Gradient Least Squares (CGLS), were shown to obtain better-quality or equal-quality excitations compared to a Singular Value Decomposition (SVD) truncated

pseudoinversion method, while simultaneously producing pulses with significantly lower peak and RMS waveform voltages. This was shown to hold across a range of targets, k-space trajectories, and acceleration factors. Between LSQR and CGLS, the former had superior numerical properties. Results were validated via Bloch-equation simulations and by conducting actual accelerated parallel excitations on an eight-channel system at 3 Tesla.

Chapter 5 used MSSO theory to develop a sparsity-enforcement algorithm that jointly determined quickly-traversable excitation k-space trajectories along with corresponding excitation pulses. The algorithm functioned by applying an ℓ_1-norm penalty while searching over a large number of possible trajectory segments (and corresponding RF pulse segments) that ultimately revealed a small subset of trajectory and pulse segments that alone formed a high-fidelity version of the desired target excitation. Imaging experiments at 7 Tesla showed that applying sparsity-enforced joint trajectory-pulse design to spoke-based pulses yielded designs capable of mitigating B_1^+ inhomogeneity in both a head-shaped water phantom and the human brain, a seemingly novel contribution to high field MRI RF excitation pulse design, B_1^+ inhomogeneity mitigation, and *in vivo* brain imaging at 7 Tesla. The strength of sparsity-enforced spoke placement was also demonstrated by designing fast, slice-selective RF pulses that achieved a complex-valued target pattern using an eight-channel 3-Tesla parallel excitation system. Finally, sparsity-enforced spiral trajectories were designed to mitigate non-uniform 7-Tesla transmit inhomogeneity using a single-channel system. In all cases, the sparsity-enforced method outperformed conventional methods, producing excitations with lower RMSE when pulse duration across the methods was fixed, and producing pulses with significantly shorter durations when excitation quality across the methods was fixed.

Chapter 6 confronted a subset selection problem that arose when a parallel excitation system had more transmit modes available than hardware transmit channels with which to drive them. By applying multiple-system single-output simultaneous sparsity theory, we derived a sparsity-enforced mode subset selection (SEMSS) algorithm that determined surprising target-specific mixtures of light and dark modes and provided increases in excitation quality relative to the usage of circularly-polarized and bright modes. In one simulated experiment, the proposed fast algorithm actually found the optimal solution to the underlying NP-Hard combinatoric subset selection problem.

Chapter 7 investigated the behavior of whole-head and local specific absorption rate

(SAR) as a function of trajectory acceleration factor and target excitation pattern due to the parallel transmission of spatially-tailored excitations at a high field strength of 7 Tesla. With the exception of dramatic SAR increases with trajectory acceleration factor, global and local SAR did not always exhibit intuitive trends. Nonetheless, it was clear from the experiments that maximum local 1g and 10g SAR were always significantly higher than global SAR. The results showed that local SAR, rather than global SAR, was the limiting factor of eight-channel parallel transmission at 7 Tesla. In order to ensure patient safety, it seems that model-based validation of individual target patterns and corresponding sets of parallel transmission pulses will be required.

Chapter 8 proposed several ways to reduce the maximum local SAR produced by parallel excitation pulses. Using novel derivations, it was shown that local Ng SAR at any spatial point of interest $\mathbf{r}$ might be calculated using a highly-sparse, redundant-block-diagonal matrix. A variety of optimization problems were then posed using such matrices.

Chapter 9 introduced the novel concept of time-multiplexing a set of similar-excitation-fidelity pulses to reduce maximum local SAR without impacting excitation quality. This approach led to safer excitations than did transmitting the best individual pulse (from a local SAR standpoint) over many repetition times. An optimization was formulated and used to determine the optimal time-multiplexing scheme of a set of candidate pulses. A small subset of pulses was revealed that, when played over a number of repetitions, yielded not only a high-fidelity excitation (5.5% NRMSE) but 1.11 times lower maximum local 1g SAR than did transmitting simply the best individual pulse in the lowest maximum local 1g SAR sense. This concept was applied to slice-selective parallel transmission pulses, but readily applies to conventional single-channel transmission systems and other pulse types. The concept of time multiplexing to reduce SAR, the optimal multiplexing convex optimization algorithm, and the linear-algebraic derivations of local SAR calculations all seem to be novel contributions.

10.2 Recommendations

We conclude the thesis simply by listing a variety of ideas for future research.

MSSO automated control parameter selection. A fast technique for finding ideal values of the regularization parameter, λ, in the MSSO optimization problems of Ch. 3,

Ch. 5 and Ch. 6 is an open problem. It might be worth investigating several approaches to automated parameter selection: the "L-curve" method [60], universal parameter selection [41], and min-max parameter selection [75].

MSSO algorithm runtime reduction. The runtimes of the seven MSSO algorithms given in Ch. 7 could be reduced significantly by implementing them in a completely compiled format such as `C`. The least-squares matching pursuit algorithm's computation and runtime could be improved upon by extending the projection based recursive updating schemes of [31, 32]. Runtime might also be reduced via a multi-resolution approach as in [92]. For example, in the context of sparsity-enforced spoke placement, one might first supply the algorithm with a coarse frequency grid, note which spoke locations are revealed, and then run the algorithm with a grid that is finely sampled around the locations suggested by the coarse result. This is faster than providing the algorithm a large, finely-sampled grid and attempting to solve the problem in one step. Further, because the desired solution to the spoke placement problem is sparse and the matrices involved are dense, solving this problem using iterative shrinkage techniques [34, 44], greedy-pursuit algorithms [24, 129], or special-purpose solvers (e.g., [83, 109]) may lead to major runtime improvements.

MSSO shrinkage algorithm convergence improvements. Both iterated shrinkage methods given in Ch. 3 (row-by-row and column-by-column shrinkage) required excessive iterations and hence exhibited lengthy runtimes. To mitigate these problems, one may consider extending parallel coordinate descent (PCD) shrinkage techniques used for single-system single-output sparse approximation (as in [43, 44]). Sequential subspace optimization (SESOP) [45] might also be employed to reduce runtime. Combining PCD with SESOP and adding a line search after each iteration would yield sophisticated versions of row-by-row and column-by-column shrinkage.

Study of the relaxed MSSO convex objective. Theoretical exploration of (3.3, 3.8) is merited, perhaps along the lines of [39, 40, 127, 128].

Multiple-system multiple-output (MSMO) simultaneous sparse approximation. In the future it may be useful to consider a combined problem where there are multiple observations as well as multiple system matrices. That is, assume we have a series of J observations, $\mathbf{d}_1, \ldots, \mathbf{d}_J$, each of which arises due to a set of P simultaneously K-sparse unknown vectors $\mathbf{g}_{1,j}, \ldots, \mathbf{g}_{P,j}$[1] passing through a set of P system matrices $\mathbf{F}_{1,j}, \ldots, \mathbf{F}_{P,j}$

[1]The K-term simultaneous sparsity profile of each set of $\mathbf{g}_{p,j}$s may or may not change with j.

and then undergoing linear combination, as follows:

$$\mathbf{d}_j = \mathbf{F}_{1,j}\mathbf{g}_{1,j} + \cdots + \mathbf{F}_{P,j}\mathbf{g}_{P,j} = \sum_{p=1}^{P} \mathbf{F}_{p,j}\mathbf{g}_{p,j} \text{ for } j = 1, \ldots, J. \tag{10.1}$$

If $\mathbf{F}_{p,j}$ is constant for all J observations then the problem reduces to

$$\mathbf{d}_j = \mathbf{F}_1\mathbf{g}_{1,j} + \cdots + \mathbf{F}_P\mathbf{g}_{P,j} = \mathbf{F}_{\text{tot}}\mathbf{g}_{\text{tot},j} \text{ for } j = 1, \ldots, J, \tag{10.2}$$

and we may stack the matrices and terms as follows:

$$[\mathbf{d}_1, \ldots, \mathbf{d}_J] = \mathbf{F}_{\text{tot}} \left[\mathbf{g}_{\text{tot},1}, \ldots, \mathbf{g}_{\text{tot},J}\right]. \tag{10.3}$$

Having posed (10.1, 10.2, 10.3), one may formulate optimization problems similar to (2.27, 3.3) to determine simultaneously sparse $\mathbf{g}_{p,j}$s that solve (10.3). Algorithms to solve such problems may arise by combining the concepts of SSMO algorithms [32, 92, 127, 129] with those of the MSSO algorithms posed in Ch. 3 of this thesis.

Transmission mode design. In Ch. 6 we focused solely on the *selection* of transmit modes, not on their *design*. The latter might be an interesting topic that blends optimization with transmission array hardware research.

Rapid, automated B_1^+ profile mapping. The sparsity-enforced single-channel spoke-based pulse design algorithm of Sec. 5.3 required a $|B_1^+|$ map to design a pulse to mitigate inhomogeneity. This requirement posed a challenge because $|B_1^+|$ varied per slice and per subject and estimating the map for a given slice and subject required 17 minutes (collecting 10 images at 1.7 minutes/image and then fitting). To enable clinical use of such techniques, mapping time must be reduced. First, it may not be necessary to collect ten high-resolution images for B_1^+ mapping; it seems 5-6 lower-resolution images may be sufficient, but at most this reduces mapping time to 4-5 minutes. Instead, or additionally, it may be possible to rapidly map $|B_1^+|$ in under a minute by exploiting some empirical trends we have observed: for example, $|B_1^+(\mathbf{r})|$ varies slowly with z, so a map estimate obtained at $z = z_0$ may be accurate within some range $z_0 \pm \delta$, allowing $|B_1^+|$ to be mapped once per slab rather than once per slice. It also seems that $|B_1^+|$ does not differ radically across subjects for a fixed axial slice. Thus it may be possible to develop a prototypical slice-by-slice $|B_1^+|$ model of the average brain and retune the slice maps of this model for a given subject by simply collecting

a small set of rapidly-acquired calibration scans to account for individual differences from the atlas.

SAR-reduced pulse design. The SAR expressions given in Ch. 8, given their sparse, structured natures, might be exploited from a computational standpoint to reduce the complexity of SAR calculations, perhaps via parallel computing.

SAR reduction via pulse optimization and multiplexing. The work of Ch. 8 and Ch. 9 might be extended by merging SAR-reduced parallel transmission pulse design technologies with the time multiplexing concept. It may also be worth investigating optimal time multiplexing schemes for spiral-trajectory-based (rather than spoke-based) parallel excitation pulses.

Bibliography

[1] *IEEE C95.3-2002. Recommended Practice for Measurements and Computations of Radio Frequency Electromagnetic Fields with Respect to Human Exposure to Such Fields, 100 kHz-300 GHz.* Institute of Electrical and Electronics Engineers, Inc., 2002.

[2] Ultra-high-field MRI Allows For Earlier Diagnosis Of Multiple Sclerosis. *Science Daily*, May 6, 2007. http://www.sciencedaily.com/releases/2007/05/070504121140.htm.

[3] V. Alagappan, E. Adalsteinsson, K. Setsompop, U. Fontius, A. C. Zelinski, G. C. Wiggins, F. Hebrank, F. Schmitt, and L. L. Wald. Comparison of Three Transmit Arrays for Parallel Transmit. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 165, 2007.

[4] V. Alagappan, J. Nistler, E. Adalsteinsson, K. Setsompop, U. Fontius, A. C. Zelinski, M. Vester, G. C. Wiggins, F. Hebrank, W. Renz, F. Schmitt, and L. L. Wald. A Degenerate Birdcage Coil for Parallel Excitation. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1028, 2007.

[5] V. Alagappan, J. Nistler, E. Adalsteinsson, K. Setsompop, U. Fontius, A. C. Zelinski, M. Vester, G. C. Wiggins, F. Hebrank, W. Renz, F. Schmitt, and L. L. Wald. Degenerate mode band-pass birdcage coil for accelerated parallel excitation. *Magn Reson Med*, 57(6):1148–1158, 2007.

[6] V. Alagappan, K. Setsompop, J. R. Polimeni, A. Potthast, A. C. Zelinski, G. C. Wiggins, U. Fontius, F. Schmitt, E. Adalsteinsson, and L. L. Wald. Mode Compression of Transmit and Receive Arrays for Parallel Imaging at 7T. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 619, Toronto, Canada, 2008.

[7] V. Alagappan, G. C. Wiggins, A. Potthast, K. Setsompop, E. Adalsteinsson, and L. L. Wald. An 8 Channel Transmit Coil for Transmit Sense at 3T. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 121, Seattle, Washington, USA, 2006.

[8] L. M. Angelone, N. Makris, C. E. Vasios, L. L. Wald, and G. Bonmassar. Effect of transmit array phase relationship on local Specific Absorption Rate (SAR). In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 2038, Seattle, WA, 2006.

[9] L. M. Angelone, A. Potthast, F. Segonne, S. Iwaki, J. W. Belliveau, and G. Bonmassar. Metallic electrodes and leads in simultaneous EEG-MRI: Specific absorption rate (SAR) simulation studies. *Bioelectromagnetics*, 25(4):285–295, 2004.

[10] L. M. Angelone, S. Tulloch, G. C. Wiggins, S. Iwaki, N. Makris, and G. Bonmassar. New high resolution head model for accurate electromagnetic field computation. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 881, Miami, FL, 2005.

[11] Enzo A. Barberi, Joseph S. Gati, Brian K. Rutt, and Ravi S. Menon. A Transmit-Only/Receive-Only (TORO) RF System for High-Field MRI/MRS Applications. *Magn Reson Med*, 43:284–289, 2000.

[12] Matt A. Bernstein, Kevin F. King, and Xiaohong J. Zhou. *Handbook of MRI Pulse Sequences*. Academic Press, sep 2004.

[13] A. Bjorck and T. Elfving. Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations. Technical Report Res Rep LiTH-MAT-R-1978-5, Department of Mathematics, Linkoping Univ., 1978.

[14] F. Bloch. Nuclear Induction. *Phys Rev A*, 70(7-8):460–474, 1946.

[15] H. Bomsdorf, T. Helzel, D. Kunz, P. Roschmann, O. Tschendel, and J. Wieland. Spectroscopy and Imaging with a 4 Tesla Whole-body MR System. *NMR in Biomedicine*, 1(3):151–158, 1988.

[16] P. Bottomley and E. Andrews. RF magnetic field penetration, phase shift and power dissipation in biological tissue: implications for NMR imaging. *Phys Med Biol*, 23:630–643, 1978.

[17] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, mar 2004.

[18] D. O. Brunner, S. Schweizer, and K. P. Pruessmann. Fast Mapping of Highly Inhomogenous RF Fields. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 353, Berlin, Germany, 2007.

[19] J. Butler and R. Lowe. Beamforming matrix simplifies design of electronically scanned antennas. *Electron Design*, 9:170–173, 1961.

[20] A. C. Cangellaris and D. B. Wright. Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in FDTD simulations of electromagnetic phenomena. *IEEE Trans Antennas Propagat*, 39:1518–1525, 1991.

[21] Center for Devices and Radiologic Health. *Guidance for the submission of premarket notifications for magnetic resonance diagnostic devices*. Food and Drug Administration, Rockville, MD, 1998.

[22] A. E. Chang, Y. L. Matory, A. J. Dwyer, S. C. Hill, M. E. Girton, S. M. Steinberg, R. H. Knop, J. A. Frank, D. Hyams, J. L. Doppman, and S. A. Rosenberg. Magnetic Resonance Imaging *Versus* Computed Tomography in the Evaluation of Soft Tissue Tumors of the Extremities. *Ann Surg*, 205(4):340–348, 1987.

[23] S. Chen and J. Wigger. Fast orthogonal least squares algorithm for efficient subset model selection. *IEEE Trans Signal Process*, 43(7):1713–1715, jul 1995.

[24] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM J. Scientific Computing*, 20(1):33–61, 1998.

[25] M. S. Cohen, R. M. DuBois, and M. M. Zeineh. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Hum Brain Mapp*, 10:204–211, 2000.

[26] C. M. Collins, S. Li, and M. B. Smith. SAR and B_1 distribution in a heterogeneous human head model within a birdcage coil. *Magn Reson Med*, 40:847–856, 1998.

[27] C. M. Collins and M. B. Smith. Signal-to-noise ratio and absorbed power as functions of main magnetic field strength, and definition of 90 degrees RF pulse for the head in the birdcage coil. *Magn Reson Med*, 45:684–691, 2001.

[28] C. M. Collins, Z. Wang, and M. B. Smith. A Conservative Method for Ensuring Safety within Transmit Arrays. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1092, Berlin, Germany, 2007.

[29] C. M. Collins, L. Wanzhan, W. Schreiber, Q. X. Yang, and M. B. Smith. Central brightening due to constructive interference with, without, and despite dielectric resonance. *J Magn Reson Imag*, 21(2):192–196, 2005.

[30] P. L. Combettes and J. C. Pesquet. Proximal Thresholding Algorithm for Minimization over Orthogonal Bases. *SIAM J Optim*, 18(4):1351–1376, nov 2007.

[31] S. F. Cotter, J. Adler, B. D. Rao, and K. Kreutz-Delgado. Forward sequential algorithms for best basis selection. In *Proc Inst Elect Eng Vision, Image, Signal Process*, volume 146, pages 235–244, oct 1999.

[32] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado. Sparse Solutions to Linear Inverse Problems with Multiple Measurement Vectors. *IEEE Trans Sig Proc*, 53(7):2477–2488, jul 2005.

[33] C. H. Cunningham, J. M. Pauly, and K. S. Nayak. Saturated Double-Angle Method for Rapid B1+ Mapping. *Magn Reson Med*, 55:1326–1333, 2006.

[34] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57(11):1413–1457, 2004.

[35] N. P. Davies and P. Jezzard. Selective arterial spin labeling (SASL): Perfusion territory mapping of selected feeding arteries tagged using two-dimensional radiofrequency pulses. *Magn Reson Med*, 49(6):1133–1142, 2003.

[36] G. Davis. *Adaptive Nonlinear Approximations*. PhD thesis, New York Univ., 1994.

[37] G. Davis, S. Mallat, and M. Avellaneda. Greedy adaptive approximation. *Constr Approx*, 13:57–98, 1997.

[38] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans Info Theory*, 41:613–627, may 1995.

[39] D. L. Donoho. For Most Large Underdetermined Systems of Equations, the Minimal ℓ_1-norm Near-Solution Approximates the Sparsest Near Solution. *Commun. Pure Appl. Math.*, 59(7):907–934, jul 2006.

[40] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. *IEEE Trans Info Theory*, 52(1):6–18, jan 2006.

[41] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, aug 1994.

[42] C. L. Dumoulin and H. R. Hart. Magnetic Resonance Angiography. *Radiology*, 161:717–720, 1986.

[43] M. Elad. Why Simple Shrinkage Is Still Relevant for Redundant Representations? *IEEE Trans Info Theory*, 52(12):5559–5569, dec 2006.

[44] M. Elad, B. Matalon, and M. Zibulevsky. Image Denoising with Shrinkage and Redundant Representations. In *Proc. IEEE CVPR*, volume 2, pages 1924–1931, 2006.

[45] M. Elad, B. Matalon, and M. Zibulevsky. Coordinate and Subspace Optimization Methods for Linear Least Squares with Non-Quadratic Regularization. *Appl Comp Harm Anal*, 23:346–367, nov 2007.

[46] M. Elad and I. Yavheh. A Weighted Average of Sparse Representations is Better than the Sparsest One Alone. *IEEE Trans Info Theory*, Submitted:1–35, 2008.

[47] D. A. Feinberg, J. C. Hoenninger, L. E. Crooks, L. Kaufman, J. C. Watts, and M. Arakawa. Inner volume MR imaging: technical concepts and their applicationInner volume MR imaging: technical concepts and their application. *Radiology*, 156:743–747, 1985.

[48] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE J Selected Topics in Signal Process*, 1(4):586–597, dec 2007.

[49] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran. Denoising by sparse approximation: Error bounds based on rate–distortion theory. *EURASIP J. Applied Signal Process.*, 2006:1–19, mar 2006.

[50] G. H. Golub and W. Kahan. Calculating the singular values and pseudoinverse of a matrix. *SIAM J. Numerical Analysis*, 2:205–224, 1965.

[51] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, 1983.

[52] I. F. Gorodnitsky and B. D. Rao. Sparse signal reconstruction from limited data using FOCUSS: a recursive weighted norm minimization algorithm. *IEEE Trans Sig Proc*, 45(3):600–616, 1997.

[53] V. K. Goyal, A. K. Fletcher, and S. Rangan. Compressive sampling and lossy compression. *IEEE Signal Process. Mag.*, 25(2):48–56, mar 2008.

[54] I. Graesslin, S. Biederer, K. Falaggis, P. Vernickel, H. Dingemans, G. Mens, P. Roeschmann, C. Leussler, Z. Zhai, M. Morich, and U. Katscher. Real-time SAR Monitoring to ensure Patient Safety for Parallel Transmission Systems. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1086, Berlin, Germany, 2007.

[55] I. Graesslin, S. Biederer, F. Schweser, K. H. Zimmermann, U. Katscher, and P. Bornert. SAR Reduction for Parallel Transmission using VERSE and k-space Filtering. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 674, Berlin, Germany, 2007.

[56] I. Graesslin, K. Falaggis, S. Biederer, D. Glaesel, P. Vernickel, P. Roeschmann, C. Leussler, A. Thran, Z. Zhai, M. Morich, and U. Katscher. SAR Simulations and Experiments for Parallel Transmission. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1090, Berlin, Germany, 2007.

[57] I. Graesslin, F. Schweser, B. Annighoefer, S. Biederer, U. Katscher, K. Nehrke, H. Stahl, H. Dingemans, G. Mens, and P. Bornert. A Minimum SAR RF Pulse Design Approach for Parallel Tx with Local Hot Spot Suppression and Exact Fidelity Constraint. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 621, Toronto, Canada, 2008.

[58] I. Graesslin, P. Vernickel, J. Schmidt, C. Findeklee, P. Roschmann, C. Leussler, P. Haaker, H. Laudan, K. M. Luedeke, J. Scholz, S. Buller, J. Keupp, P. Bornert, H. Dingemans, G. Mens, G. Vissers, K. Blom, N. Swennen, J. vd Heijden, L. Mollevanger, P. Harvey, and U. Katscher. Whole Body 3T MRI System with Eight Parallel RF Transmission Channels. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 129, Seattle, Washington, USA, 2006.

[59] W. A. Grissom, C. Y. Yip, Z. Zhang, V. A. Stenger, J. A. Fessler, and D. C. Noll. Spatial Domain Method for the Design of RF Pulses in Multicoil Parallel Excitation. *Magn Reson Med*, 56(3):620–629, 2006.

[60] P. C. Hansen. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6:1–35, 1994.

[61] B. A. Hargreaves, D. G. Nishimura, and S. M. Conolly. Minimum-Time Multi-Dimensional Gradient Waveform Design using Convex Optimization. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1001, Toronto, Canada, 2003.

[62] M. R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *J. Research of the National Bureau of Standards*, 49(6):409–436, 1952.

[63] R. Holland, V. P. Cable, and L. C. Wilson. Finite-volume time-domain (FVTD) techniques for EM scattering. *IEEE Trans Electromagn Compat*, 33:281–294, 1991.

[64] D. I. Hoult. The principle of reciprocity in signal strength calculations - A mathematical guide. *Concepts Magn Reson*, 12(4):173–187, 2000.

[65] D. I. Hoult and R. Richards. The signal to noise ratio of the nuclear magnetic resonance experiment. *J Magn Reson*, 24:71–85, 1976.

[66] X. Hu and T. Parrish. Reduction of field of view for dynamic imaging. *Magn Reson Med*, 31:691–694, 1994.

[67] T. S. Ibrahim, R. Lee, B. A. Baertlein, and P. M. Robitaille. B1 field inhomogeneity and SAR calculations for the birdcage coil. *Phys Med Biol*, 46:609–619, 2001.

[68] T. S. Ibrahim and L. Tang. A Whole-Body 7 Tesla RF Excitation Scheme with Much Improved B_1^+ Field Homogeneity and Local/Global SARs over Quadrature Excitation. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1013, Berlin, Germany, 2007.

[69] E. K. Insko and L. Bolinger. Mapping of the radiofrequency field. *J Magn Reson A*, 103:82–85, 1993.

[70] International Electrotechnical Commission. *International standard, medical equipment-part 2: particular requirements for the safety of magnetic resonance equipment for medical diagnosis, 2nd revision.* International Electrotechnical Commission, Geneva, 2002.

[71] International Electrotechnical Commission. *IEC Standard 60601-2-33. Particular requirements for the safety of magnetic resonance equipment for medical diagnosis.* International Electrotechnical Commission, Chicago, IL, 2004.

[72] J. D. Jackson. *Classical Electrodynamics.* Wiley, New York, NY, 1999.

[73] J. Jin and J. Chen. On the SAR and field inhomogeneity of birdcage coils loaded with the human head. *Magn Reson Med*, 21:192–196, 1997.

[74] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing–Concepts and Techniques.* Prentice Hall, 1993.

[75] I. M. Johnstone. On minimax estimation of a sparse normal mean vector. *Annals of Statistics*, 22(1):271–289, mar 1994.

[76] S. A. R. Kannengießer, A. R. Brenner, and T. G. Noll. Accelerated image reconstruction for sensitivity encoded imaging with arbitrary k-space trajectories. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 155, 2000.

[77] L. A. Karlovitz. Construction of nearest points in the ℓ^p, p even and ℓ^1 norms. *J Approx Theory*, 3:123–127, 1970.

[78] U. Katscher and P. Bornert. Parallel RF transmission in MRI. *NMR Biomed*, 19:393–400, 2006.

[79] U. Katscher, P. Bornert, C. Leussler, and J. S. van den Brink. Transmit SENSE. *Magn Reson Med*, 49(1):144–150, 2003.

[80] U. Katscher, I. Graesslin, H. Dingemans, and G. Mens. Experimental verification of over- and underdetermined Transmit SENSE. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 600, 2006.

[81] U. Katscher, P. Vernickel, and J. Overweg. Basics of RF power behaviour in parallel transmission. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 17, 2005.

[82] A. B. Kerr, C. H. Cunningham, J. M. Pauly, J. E. Piel, R. O. Giaquinto, R. D. Watkins, and Y. Zhu. Accelerated B1 Mapping for Parallel Excitation. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 352, Berlin, Germany, 2007.

[83] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A Method for Large-Scale ℓ_1-Regularized Least Squares Problems with Applications in Signal Processing and Statistics. Technical report, Stanford Univ., 2007.

[84] H. Krim and M. Viberg. Two decades of array signal processing research. The parametric approach. *IEEE Sig Proc Mag*, 13(4):67–94, jul 1996.

[85] K. S. Kunz and R. J. Luebbers. *The finite difference time domain method for electromagnetics*. CRC Press, Boca Raton, FL, 1993.

[86] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Research of the National Bureau of Standards*, 45:255–282, 1950.

[87] P.C. Lauterbur. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242:190–191, 1973.

[88] Y. S. Levin, L. J. Pisani, D. M. Spiclman, and J. M. Pauly. Trajectory optimization for variable-density spiral two-dimensional excitation. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 3012, Seattle, WA, USA, 2006.

[89] Z. P. Liang and P. C. Lauterbur. *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. IEEE Press on Biomedical Engineering, New York, NY, 2000.

[90] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, 1969.

[91] D. M. Malioutov. A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays. Master's thesis, Massachusetts Institute of Technology, 2003.

[92] D. M. Malioutov, M. Çetin, and A. S. Willsky. A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays. *IEEE Trans Sig Proc*, 53(8):3010–3022, 2005.

[93] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans Sig Proc*, 41(12):3397–3415, dec 1993.

[94] W. Mao, Z. Wang, M. B. Smith, and C. M. Collins. Calculation of SAR for transmit coil arrays. *Concepts Magn Reson, Part B: Magn Reson Eng*, 31B(2):127–131, 2007.

[95] J. P. Mugler and G. W. Miller. Rapid 3D Mapping of the B1 Field Using a Low-Flip Angle, Phase-Based Method with Improved Sensitivity. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 351, Berlin, Germany, 2007.

[96] B. K. Natarajan. Sparse Approximate Solutions to Linear Systems. *SIAM J. Scientific Computing*, 24(2):227–234, apr 1995.

[97] A. Nemirovski and A. Ben-Tal. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, Philadelphia, Pennsylvania, USA, 2001.

[98] D. G. Nishimura. *Principles of Magnetic Resonance Imaging*. Self-Published, 1996.

[99] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[100] C. C. Paige and M. A. Saunders. Algorithm 583; LSQR: Sparse linear equations and least-squares problems. *ACM Transactions on Mathematical Software*, 8(2):195–209, 1982.

[101] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, 1982.

[102] J. M. Pauly, D. Nishimura, and A. Macovski. A k-space analysis of small-tip-angle excitation. *J Magn Reson*, 81:43–56, 1989.

[103] M. J. D. Powell. A Fast Algorithm for Nonlinearly Constrained Optimization Calculations. *Lecture Notes in Mathematics*, 630:144–157, 1978.

[104] K. P. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger. A gridding approach for sensitivity encoding with arbitrary trajectories. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 276, 2000.

[105] K. P. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn Reson Med*, 46:638–651, 2001.

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

[107] B. D. Rao. Signal processing with the sparseness constraint. In *Proc. ICASSP*, volume 3, pages 1861–1864, Seattle, WA, may 1998.

[108] S. Saekho, C. Y. Yip, D. C. Noll, F. E. Boada, and V. A. Stenger. Fast-k_z three-dimensional tailored radiofrequency pulse for reduced B1 inhomogeneity. *Magn Reson Med*, 55(4):719–724, 2006.

[109] M. A. Saunders and B. Kim. PDCO: Primal-Dual interior method for Convex Objectives, 2007.

[110] R. D. Schafer. *An Introduction to Nonassociative Algebras*. Dover, New York, NY, 1995.

[111] F. Seifert, G. Wubbeler, S. Junge, B. Ittermann, and H. Rinneberg. Patient Safety Concept for Multichannel Transmit Coils. *J Magn Reson Imag*, 26:1315–1321, 2007.

[112] K. Setsompop, L. L. Wald, V. Alagappan, B. A. Gagoski, and E. Adalsteinsson. Magnitude least squares optimization for parallel radio frequency excitation design demonstrated at 7 Tesla with eight channels. *Magn Reson Med*, 59(4):908–915, apr 2008.

[113] K. Setsompop, L. L. Wald, V. Alagappan, B. A. Gagoski, F. Hebrank, U. Fontius, F. Schmitt, and E. Adalsteinsson. Parallel RF Transmission with 8 Channels at 3 Tesla. *Magn Reson Med*, 56(5):1163–1171, 2006.

[114] M. Smelyanskiy, S. Skedzielewski, and C. Dulong. Parallel Computing for Large-Scale Optimization Problems: Challenges and Solutions. *Intel Technology Journal*, 9(2):2–14, 2005.

[115] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magn Reson Med*, 38(4):591–603, 1997.

[116] S. P. Souza and C. L. Dumoulin. Phase-Contrast Magnetic Resonance Angiography. In *IEEE Engineering in Medicine & Biology Soc. 11th Annual Int. Conference*, 1989.

[117] R. Staewen, A. Johnson, B. Ross, T. Parrish, H. Merkle, and M. Garwood. 3-D FLASH imaging using a single surface coil and a new adiabatic pulse, BIR-4. *Invest Radiol*, 25:559–567, 1990.

[118] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 1993.

[119] P. W. Stroman. Magnetic Resonance Imaging of Neuronal Function in the Spinal Cord: Spinal fMRI. *Clinical Medicine and Research*, 3(3):146–156, 2005.

[120] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999. Special issue on Interior Point Methods.

[121] B. P. Sutton, D. C. Noll, and J. A. Fessler. Fast, Iterative Image Reconstruction for MRI in the Presence of Field Inhomogeneties. *IEEE Trans Med Imag*, 22(2):178–188, 2003.

[122] R. Tibshirani. Regression shrinkage and selection via the LASSO. *J. Royal Statistical Soc., Series B*, 58:267–288, nov 1996.

[123] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Doklady Akademii Nauk SSSR*, 151:501–504, 1963.

[124] A. N. Tikhonov and V. A. Arsenin. *Solution of Ill-posed Problems*. Winston & Sons, Washington, DC, USA, 1977.

[125] K. C. Toh, M. J. Todd, and R. H. Tutuncu. SDPT3 - a Matlab software package for semidefinite programming. *Optimization Methods and Software*, 11(12):545–581, 1999.

[126] J. Tropp. The hybrid bird cage resonator. In *Proc. Soc. for Magnetic Resonance in Medicine (SMRM)*, page 4009, Berlin, Germany, 1992.

[127] J. A. Tropp. Algorithms for simultaneous sparse approximation: part II: Convex relaxation. *Signal Process*, 86(3):589–602, 2006.

[128] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans Info Theory*, 52(3):1030–1051, mar 2006.

[129] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation: part I: Greedy pursuit. *Signal Process.*, 86(3):572–588, 2006.

[130] P. Ullmann, S. Junge, M. Wick, F. Seifert, W. Ruhm, and J. Hennig. Experimental analysis of parallel excitation using dedicated coil setups and simultaneous RF transmission on multiple channels. *Magn Reson Med*, 54(4):994–1001, 2005.

[131] P. Ullmann, G. Wuebbeler, S. Junge, F. Seifert, W. Ruhm, and J. Hennig. SAR-Analysis for Transmit SENSE with a 4-Channel Head Array at 3 T. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 601, Seattle, Washington, USA, 2006.

[132] J. Ulloa and J. V. Hajnal. Exploring 3D RF shimming for slice selective imaging. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 21, Miami Beach, Florida, USA, 2005.

[133] P. F. van de Moortele, C. Akgun, G. Adriany, S. Moeller, J. Ritter, C. M. Collins, M. B. Smith, J. T. Vaughan, and K. Ugurbil. B_1 Destructive Interference and Spatial Phase Patterns at 7T with a Head Transceiver Array Coil. *Magn Reson Med*, 54:1503–1518, 2005.

[134] B. Van den Berg, C. A. T. van den Bergen, J. B. Van de Kramer, B. W. Raaymakers, H. Kroeze, L. W. Bartels, and J. J. W. Lagendijk. Simultaneous B_1^+ Homogenization and Specific Absorption Rate Hotspot Suppression Using a Magnetic Resonance Phased Array Transmit Coil. *Magn Reson Med*, 57:577–586, 2007.

[135] W. van der Zwaag, S. Francis, and R. Bowtell. Improved echo volumar imaging (EVI) for functional MRI. *Magn Reson Med*, 56(6):1320–1327, 2006.

[136] J. T. Vaughan, M. Garwood, C. M. Collins, W. Liu, L. DelaBarre, G. Adriany, P. Anderson, H. Merkle, R. Goebel, M. B. Smith, and K. Ugurbil. 7T vs. 4T: RF power, homogeneity, and signal-to-noise comparison in head images. *Magn Reson Med*, 46(1):24–30, 2001.

[137] J Wang, M Qiu, Q. X. Yang, M. B. Smith, and R. T. Constable. Measurement and correction of transmitter and receivers induced nonuniformities in vivo. *Magn Reson Med*, 53:408–417, 2005.

[138] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[139] Z. Wang, S-H Oh, M. B. Smith, and C. M. Collins. RF Shimming Considering Both Excitation Homogeneity and SAR. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1022, Berlin, Germany, 2007.

[140] G. C. Wiggins, A. Potthast, C. Triantafyllou, C. J. Wiggins, and L. L. Wald. Eight-channel phased array coil and detunable TEM volume coil for 7 T brain imaging. *Magn Reson Med*, 54(1):235 – 240, 2005.

[141] X. Wu, C. Akgun, J. T. Vaughan, K. Ugurbil, and P. F. van de Moortele. SAR Analysis for Transmit SENSE at 7T with a Human Head Model. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 3350, Berlin, Germany, 2007.

[142] X. Wu, C. Akgun, J. T. Vaughan, K. Ugurbil, and P. F. van de Moortele. SAR Reduction in Transmit SENSE Using Adapted Excitation k-Space Trajectories. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 673, Berlin, Germany, 2007.

[143] Y. Yang, V. S. Mattay, D. R. Weinberger, J. A. Frank, and J. H. Duyn. Localized echo-volume imaging methods for functional MRI. *J Magn Reson Imag*, 7(2):371–375, 1997.

[144] K. S. Yee. Numerical solutions of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans Antennas Propagat*, AP-14:302–307, 1966.

[145] C. Y. Yip, J. A. Fessler, and D. C. Noll. A novel, fast and adaptive trajectory in three-dimensional excitation k-space. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 2350, Miami, FL, USA, 2005.

[146] C. Y. Yip, J. A. Fessler, and D. C. Noll. Iterative RF pulse design for multidimensional, small-tip-angle selective excitation. *Magn Reson Med*, 54(4):908–917, 2005.

[147] C. Y. Yip, J. A. Fessler, and D. C. Noll. Advanced 3-D tailored RF pulse for signal recovery in T_2^*-weighted functional magnetic resonance imaging. *Magn Reson Med*, 56(5):1050–1059, 2006.

[148] C. Y. Yip, W. A. Grissom, J. A. Fessler, and D. C. Noll. Joint Design of Trajectory and RF Pulses for Parallel Excitation. *Magn Reson Med*, 58(3):598–604, 2007.

[149] A. C. Zelinski, E. Adalsteinsson, V. K. Goyal, and L. L. Wald. *Sparsity-Enforced Joint Trajectory and RF Excitation Pulse Design*. US Patent Pending, Internal Case No. MGH 3726 / MIT 13074, 2008.

[150] A. C. Zelinski, E. Adalsteinsson, K. Setsompop, L. L. Wald, and U. Fontius. *Method for designing RF excitation pulses in magnetic resonance tomography*. US Patent 7336145, issued February 26, 2008.

[151] A. C. Zelinski, V. Alagappan, V. K. Goyal, E. Adalsteinsson, and L. L. Wald. Sparsity-Enforced Coil Array Mode Compression for Parallel Transmission. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1302, 2008.

[152] A. C. Zelinski, L. M. Angelone, V. K. Goyal, G. Bonmassar, E. Adalsteinsson, and L. L. Wald. Specific Absorption Rate Studies of the Parallel Transmission of Inner-Volume Selective Excitations at 7 Tesla. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1315, 2008.

[153] A. C. Zelinski, L. M. Angelone, V. K. Goyal, G. Bonmassar, E. Adalsteinsson, and L. L. Wald. Specific Absorption Rate Studies of the Parallel Transmission of Inner-Volume Excitations at 7 Tesla. *J Magn Reson Imag*, In Press, 2008.

[154] A. C. Zelinski, V. K. Goyal, and E. Adalsteinsson. Reduction of Maximum Local Specific Absorption Rate via Pulse Multiplexing. *J Magn Reson Imag*, In Preparation for Submission, 2008.

[155] A. C. Zelinski, V. K. Goyal, and E. Adalsteinsson. Simultaneously Sparse Solutions to Linear Inverse Problems with Multiple System Matrices and a Single Observation Vector. *SIAM J Sci Comp*, In Review, 2008.

[156] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald. Fast, Accurate Calculation of Maximum Local N-Gram Specific Absorption Rate. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1188, 2008.

[157] A. C. Zelinski, V. K. Goyal, E. Adalsteinsson, and L. L. Wald. Sparsity in MRI RF Excitation Pulse Design. In *Proc Conf Information Sciences and Systems*, pages 252–257, Princeton, NJ, mar 2008.

[158] A. C. Zelinski, V. K. Goyal, L. M. Angelone, G. Bonmassar, L. L. Wald, and E. Adalsteinsson. Designing RF Pulses with Optimal Specific Absorption Rate (SAR) Characteristics and Exploring Excitation Fidelity, SAR, and Pulse Duration Tradeoffs. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1699, Berlin, Germany, 2007.

[159] A. C. Zelinski, V. K. Goyal, L. L. Wald, and E. Adalsteinsson. Sparsity-Enforced Joint Spiral Trajectory & RF Excitation Pulse Design. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1303, 2008.

[160] A. C. Zelinski, K. Setsompop, V. Alagappan, B. A. Gagoski, L. M. Angelone, G. Bonmassar, U. Fontius, F. Schmitt, E. Adalsteinsson, and L. L. Wald. Pulse Design Methods for Reduction of Specific Absorption Rate in Parallel RF Excitation. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1698, 2007.

[161] A. C. Zelinski, K. Setsompop, V. Alagappan, V. K. Goyal, L. L. Wald, and E. Adalsteinsson. *In Vivo* B_1^+ Inhomogeneity Mitigation at 7 Tesla Using Sparsity-Enforced Spatially-Tailored Slice-Selective Excitation Pulses. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 620, 2008.

[162] A. C. Zelinski, K. Setsompop, V. K. Goyal, V. Alagappan, U. Fontius, F. Schmitt, L. L. Wald, and E. Adalsteinsson. Designing Fast 3-D RF Excitations by Optimizing the Number, Placement, and Weighting of Spokes in k-Space via a Sparsity-Enforcement Algorithm. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1691, Berlin, Germany, 2007.

[163] A. C. Zelinski, L. L. Wald, V. Alagappan, V. K. Goyal, and E. Adalsteinsson. *Sparsity-Enforced Coil Array Mode Compression for Parallel Transmission*. US Patent Pending, Internal Case No. MGH 3673 / MIT 13014, 2008.

[164] A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, V. K. Goyal, and E. Adalsteinsson. Fast Slice-Selective RF Excitation Pulses for Mitigating B_1^+ Inhomogeneity in the Human Brain at 7T. *Magn Reson Med*, 59(6):1355 – 1364, jun 2008.

[165] A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, V. K. Goyal, F. Hebrank, U. Fontius, F. Schmitt, and E. Adalsteinsson. Comparison of Three Algorithms for Solving Linearized Systems of Parallel Excitation RF Waveform Design Equations: Experiments on an Eight-Channel System at 3 Tesla. *Concepts Magn Reson, Part B: Magn Reson Eng*, 31B(3):176–190, aug 2007.

[166] A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, F. Hebrank, U. Fontius, F. Schmitt, and E. Adalsteinsson. RF Pulse Design Methods for Reduction of Image Artifacts in Parallel RF Excitation: Comparison of 3 Techniques on a 3T Parallel Excitation System with 8 Channels. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 1686, 2007.

[167] A. C. Zelinski, L. L. Wald, K. Setsompop, V. K. Goyal, and E. Adalsteinsson. Sparsity-Enforced Slice-Selective MRI RF Excitation Pulse Design. *IEEE Trans Med Imag*, In Press, 2008.

[168] Y. Zhu. Parallel excitation with an array of transmit coils. *Magn Reson Med*, 51(4):775–784, 2004.

[169] Y. Zhu, R. Watkins, R. Gianquinto, C. Hardy, G. Kenwood, S. Mathias, T. Valent, M. Denzin, J. Hopkins, W. Peterson, and B. Mock. Parallel excitation on an eight transmit channel MRI system. In *Proc. Int. Soc. for Magnetic Resonance in Medicine (ISMRM)*, page 14, Miami Beach, Florida, USA, 2005.