Advances in Quantitative MRI: Acquisition, Estimation, and Application

by

Gopal Nataraj

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2018

Doctoral Committee:

Professor Jeffrey A. Fessler, Co-Chair
Associate Research Scientist Jon-Fredrik Nielsen, Co-Chair
Professor Douglas C. Noll
Associate Professor Clayton Scott
Associate Research Scientist Scott Swanson

Gopal Nataraj

gnataraj@umich.edu

ORCiD® ID: 0000-0002-2847-115X

---

to Amma and Appa, who have supported me since day one

ACKNOWLEDGMENTS

If reincarnation is in fact the way of the world, I must have done something good in my past life to have landed Jeffrey Fessler and Jon-Fredrik Nielsen as research co-advisors in this lifetime. Jeff saw research potential in me before I had even seriously considered graduate school: he took time to speak to me and encouraged me to apply when I visited the University of Michigan (UM) back in 2010 as an inexperienced sophomore. Jeff has since then helped me develop my intuition and yet also my rigor; my persistence and yet also my open-mindedness; my confidence and yet also my humility. I hope to one day lead a team with even half as much wisdom and grace as does Jeff. Asking Jon to serve as my co-advisor in 2014 (also per Jeff's suggestion) is perhaps the best decision I have made during graduate school. Since our earliest meetings, Jon has effortlessly balanced playing the roles of educator, collaborator, and more recently even tennis partner. Most practical things that I know about MRI today are due to Jon. Together, Jeff and Jon have provided a delicate combination of autonomy and direction that has been key to my development as an independent researcher.

The other three members of my thesis committee have each contributed to this work in unique and meaningful ways. Douglas Noll and I have shared many conversations to try and understand which of the many MR non-idealities dominated algorithmic performance loss in practice and to brainstorm how to compensate for these losses. Clayton Scott was a key contributor to the development of PERK (Ch. 5). Scott Swanson and I have shared several conversations about our new method for myelin water imaging (Ch. 6) and his commentary will be critical to its further validation and eventual clinical translation.

I have also enjoyed collaborating with very capable junior students. Undergraduate student Mingjie Gao helped adapt PERK to MR fingerprinting data [1] and ran myelin water imaging scan optimizations (used in Ch. 6). I was recently pleased to learn that he will proceed with graduate study working with Jeff. First-year graduate student Steven Whitaker has begun to work on exploiting off-resonance for myelin water imaging (discussed in Ch. 7). He plans to lead our group's myelin water imaging research thrust in the coming years.

I am grateful to UM for the resources it provided and for the relationships it enabled. UM funded a large portion of this early-stage and exploratory research through grants and fellowships. I met many wonderful people through the "lab of Jeff" (LOJ), the functional MRI lab, the graduate student council, teaching, tutoring, coursework, and probably other venues too. You have all made what is ultimately a solitary journey much more fun and fulfilling – thank you.

I next thank my roommates, Adam Mendrela and John "Trey" Ruppe. All three of us moved here together from Cornell back in 2012. None of us wore much facial hair back then (Trey still does not). I suppose Trey won the race to a doctorate; he now works in the greater Boston area. Adam has lived with me in the same apartment for the entire duration of our PhDs; I wish him the best of luck as he wraps things up in the coming months. These two have helped me work through every single one of the highs and lows of graduate school.

I next thank my fiancée, Manisha Rai. Simply put, she is the light of my life. Manisha, I cannot wait to get married and to start the next chapter of our lives together. I hope I can be even half as supportive during your residency training as you have been these past three years.

Finally, I thank my family. I doubt many people can say that three generations (9 members!) of family attended their PhD defense. I am humbled by your unconditional love and support. This thesis is at least as much yours as it is mine.

TABLE OF CONTENTS

Dedication . . . . **ii**

Acknowledgments . . . . **iii**

List of Figures . . . . **viii**

List of Tables . . . . **xii**

List of Appendices . . . . **xv**

Abstract . . . . **xvi**

Chapter

1 Introduction **1**

1.1 Overview 2
1.2 Organization 2

2 Background **4**

2.1 Relevant MR Physics 4
2.1.1 Bloch Equations 4
2.1.2 Steady-State Sequences 7
2.2 Optimization in QMRI 15
2.2.1 Iterative Local Optimization with Constraints 15
2.2.2 Partially Linear Models and the Variable Projection Method . . . 16

3 QMRI Parameter Estimation using Likelihood Models **18**

3.1 Introduction 18
3.2 Likelihood-Based Estimation in QMRI 18
3.2.1 The QMRI Scan Profile 18
3.2.2 Latent Object Parameter Estimation 20
3.3 Experimentation 23
3.3.1 T_1 estimation from two SPGR scans 23
3.3.2 T_2 estimation from one DESS scan 25
3.4 Discussion 27
3.5 Conclusion 28

4 QMRI Acquisition Design via Min-Max Optimization **30**

4.1 Introduction 30

4.2 A CRB-Inspired Scan Selection Method 32
4.2.1 The CRB and its Relevance to QMRI 32
4.2.2 Min-max Optimization Problem for Scan Design 33
4.3 Optimizing SS Sequences for Relaxometry in the Brain 34
4.3.1 Scan Design Details 35
4.3.2 Scan Profile Comparisons 36
4.4 Experimentation 39
4.4.1 Numerical Simulations 39
4.4.2 Phantom Experiments 42
4.4.3 *In vivo* Experiments 48
4.5 Sensitivity to Multiple-Compartment Relaxation 52
4.6 Discussion 53
4.7 Conclusion 55

5 QMRI Parameter Estimation via Regression with Kernels (PERK) 56

5.1 Introduction 56
5.2 Preliminaries 58
5.3 A Function Optimization Problem & Kernel Solution 59
5.4 Bias and Covariance Analysis 62
5.4.1 Conditional Bias 62
5.4.2 Conditional Covariance 63
5.5 Implementation Considerations 64
5.5.1 A Kernel Approximation 64
5.5.2 Tuning Parameter Selection 65
5.6 PERK Demonstration in a 1-D Toy Problem 67
5.7 Experimentation 69
5.7.1 Methods 69
5.7.2 Numerical Simulations 73
5.7.3 Phantom Experiments 74
5.7.4 *In vivo* Experiments 77
5.8 Robustness Studies 80
5.8.1 Mismatch in Training vs. Testing Noise Statistics 80
5.8.2 Mismatch in Scan Design vs. Sampling Distribution Support . . . 82
5.9 Discussion 82
5.10 Conclusion 87

6 Fast Myelin Water Imaging via Acquisition Design and PERK 89

6.1 Introduction 89
6.2 Multi-Compartmental Models for SS Sequences 91
6.2.1 A Two-Compartment SPGR Signal Model 91
6.2.2 A Two-Compartment DESS Signal Model 97
6.3 A Fast SS Acquisition for Precise Myelin Imaging 104
6.3.1 Scalable Acquisition Design 104
6.3.2 SPGR/DESS Scan Design Implementation Details 106
6.4 Experimentation 108

6.4.1 Methods 108
6.4.2 Numerical Simulations 110
6.4.3 *In vivo* Experiments 115
6.5 Discussion 119
6.6 Conclusion 120

7 Future Work 122

7.1 Combine PERK with Image Reconstruction 122
7.2 Exploit Off-Resonance for Myelin Water Imaging 123
7.3 Correlate with Other Myelin Biomarkers 123
7.4 Translate into a Clinical Setting 124

Appendices 125

Bibliography 138

LIST OF FIGURES

3.1 $\mathbf{T}_1$ MOM, ML, and RL estimates and corresponding error images, from two simulated SPGR scans. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Table 3.1 presents corresponding sample statistics. . . 25
3.2 $\mathbf{T}_2$ MOM, ML, and RL estimates and corresponding error images, from one simulated DESS scan. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Table 3.2 presents corresponding sample statistics. 27

4.1 Worst-case standard deviations $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1}$ (top), $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2}$ (middle), and cost $\widetilde{\Psi}^{\mathrm{t}}$ (bottom), versus pairs of nominal flip angles, holding other scan parameters fixed at selected profile $\widehat{\mathbf{P}}$. Subfigures (a)-(i), (j)-(l), and (m)-(o) correspond to scan profiles containing $(S_{\mathrm{SPGR}}, S_{\mathrm{DESS}}) = (2,1), (1,1)$, and $(0,2)$ SPGR and DESS scans, respectively. Selected scan parameters (starred) are within $\delta = 1\%$ of global minimizers and retain as much estimator precision as possible over a wide range of latent object parameters. All axes range from 5 to 90 degrees, in 5-degree increments. Colorbar ranges are $[0,100]$, $[0,10]$, and $[0,20]$ milliseconds for rows of $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1}$, $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2}$, and $\widetilde{\Psi}^{\mathrm{t}}$ subfigures, respectively. The optimized $(0,2)$ profile appears most robust to transmit field spatial variation. 38
4.2 Histograms of T_1 and T_2 estimates from noisy independent measurements of a *single* nominal WM or GM value. In each plot, two normal distributions are overlaid, each with latent means T_1 and T_2. In (a)-(b) and (c)-(d), the solid green curve is $\mathcal{N}\big(T_1, (\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1})^2\big)$ and $\mathcal{N}\big(T_2, (\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2})^2\big)$, respectively. In (a)-(d), the dashed maroon curves have variances computed from the Fisher information at *a priori* unknown T_1, T_2 values in WM or GM. These plots correspond to an optimized $(0,2)$ scan profile; analogous plots for other profiles are visually similar. At realistic noise levels, parameter estimates distribute with minimal bias and near-Gaussian shape. Thus, the CRB reliably approximates $\widehat{T}_1^{\mathrm{ML}}$ and $\widehat{T}_2^{\mathrm{ML}}$ errors. 41
4.3 Colorized $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from an HPD® quantitative phantom. Columns correspond to scan profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Fig. 4.4 provides identical grayscale images that enumerate vials. Colorbar ranges are in milliseconds. . 45

4.4 Grayscale $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from an HPD® quantitative phantom. Columns correspond to scan profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Vials are enumerated and color-coded to correspond with data points in Fig. 4.5. Fig. 4.3 provides identical colorized images. Colorbar ranges are in milliseconds. 46

4.5 Phantom within-ROI sample statistics of $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from optimized SPGR/DESS and reference IR/SE scan profiles, versus NIST NMR measurements [2]. Markers and error bars indicate ROI sample means and ROI sample standard deviations within the 14 labeled and color-coded vials in Fig. 4.4. Tight $\mathbb{X}^{\mathrm{t}}$ and broad $\mathbb{X}^{\mathrm{b}}$ latent parameter ranges are highlighted in orange and yellow, respectively. Table 4.3 replicates sample statistics within Vials 5-8. Our MR measurements are at 293K and NIST NMR measurements are at 293.00K. Within the designed parameter ranges, estimates from different acquisitions are in reasonable agreement with NIST measurements. 47

4.6 Colorized $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from the brain of a healthy volunteer. Columns correspond to profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Table 4.5 presents corresponding WM/GM within-ROI sample statistics. Colorbar ranges are in milliseconds. 49

4.7 Colorized monoexponential $\mathbf{T}_2$ ML estimates from pairs of (top) multi-exponential simulated and (bottom) *in vivo* SE datasets. Columns denote SE dataset echo times. Colorbar ranges are in milliseconds. 53

5.1 Illustrations of PERK for T_2 estimation from a single spin echo measurement. Subfigures vary the number N of PERK training points, marked with black circles. The orange and yellow curves plot PERK $\widehat{T}_2^{\mathrm{PERK}}$ and MOM $\widehat{T}_2^{\mathrm{MOM}}$ estimators evaluated at test points, marked with blue dots. Dashed black lines denote the sampling distribution support $\mathsf{supp}(\mathsf{p}_{T_2})$ over which each PERK estimator was trained. As N increases, $\widehat{T}_2^{\mathrm{PERK}}$ appears more similar to $\widehat{T}_2^{\mathrm{MOM}}$ within well-sampled regions of $\mathsf{supp}(\mathsf{p}_{T_2})$. 68

5.2 Holdout criterion $\Psi(\lambda, \rho)$ versus Gaussian kernel bandwidth scaling parameter λ and regularization parameter ρ. Each pixel is the weighted normalized root mean squared error of a candidate PERK estimator, where the empirical mean over 10^5 test points approximates an expectation with respect to training prior distribution $\mathsf{p}_{\mathsf{x},\boldsymbol{\nu}}$ and the weighting places emphasis on good T_1, T_2 estimation performance. A white star marks the minimizer $\left(\widehat{\lambda}, \widehat{\rho}\right) \leftarrow (2^{0.6}, 2^{-41})$. 71

5.3 m_0 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three m_0 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics. 74

5.4 T_1 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three T_1 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics. 75
5.5 T_2 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three T_1 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics. 76
5.6 VPM, PGPM, and PERK m_0, T_1, T_2 estimates in a quantitative phantom. Vials are enumerated and highlighted to correspond with markers and colored boxes in Fig. 5.7. PERK has only been trained to accurately estimate within vials 4-8; within these vials, VPM, PGPM, and PERK estimates appear visually similar. 77
5.7 Phantom sample statistics of VPM, PGPM, and PERK T_1, T_2 estimates and NIST NMR reference measurements [2]. Plot markers and error bars indicate sample means and sample standard deviations computed over ROIs within the 14 vials labeled and color-coded in Fig. 5.6. Yellow box boundaries indicate projections of the PERK sampling distribution's support $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. Missing markers lie outside axis limits. Corresponding tables replicate sample means $\pm$ sample standard deviations for vials within $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. 'V#' indicates vial numbers. All values are reported in milliseconds. Within $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$, VPM, PGPM, and PERK estimates agree excellently with each other and reasonably with NMR measurements. 78
5.8 VPM, PGPM and PERK estimates of m_0, T_1, T_2 in the brain of a healthy volunteer. Separate WM ROIs are distinguished by anterior/posterior (A/P) and right/left (R/L) directions. Four small anterior cortical GM polygons are pooled into a single GM ROI. Images are cropped in post-processing for display. 79
5.9 Performance criterion $\Psi(\sigma)$ versus PERK training noise standard deviation σ, over two different scales. Similar to Fig. 5.2, each point on the blue curve is the weighted normalized root mean squared error of a separately trained PERK estimator. In each subplot, a black star marks the performance criterion minimizer $\widehat{\sigma} \leftarrow 2^{-8.8}$ while a dashed red line marks the (latent) test data noise standard deviation $\sigma^\star \leftarrow 2^{-8.736}$. To within quantization error, PERK performs best when trained with training data whose noise statistics match test data noise statistics. Performance degradation as measured by Ψ does not exceed 10% for σ values within a factor of two of $\sigma^\star$. This result suggests that PERK is somewhat robust to moderate misspecification of the noise level. 81

5.10 Phantom sample statistics of VPM, PGPM, and PERK T_1, T_2 estimates and NIST NMR reference measurements [2] in the case of strong mismatch between acquisition design versus PERK training distribution support. Unlike analogous results in Fig. 5.7, here the PERK estimator was trained with a sampling distribution whose support extended well beyond the range of T_1, T_2 values for which the acquisition was optimized. Comparing to Fig. 5.7, we find that PERK estimator performance degrades within the highlighted T_1, T_2 range of interest. Plot markers and error bars indicate sample means and sample standard deviations computed over ROIs within the 14 vials labeled and color-coded in Fig. 5.11. Corresponding tables replicate sample means $\pm$ sample standard deviations for vials within the highlighted range. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. All values are in milliseconds. 83
5.11 More aggressively trained VPM, PGPM, and PERK m_0, T_1, T_2 estimates in a quantitative phantom. Here the PERK estimator was trained with a sampling distribution whose support extended over less well identified T_1, T_2 values. Comparing with analogous images in Fig. 5.6, PERK performance within vials 4-8 degrades, though in other vials performance clearly improves. Vials are enumerated and highlighted to correspond with markers and colored boxes in Fig. 5.10. 84

6.1 NNLS/RNNLS MESE f_M and ML/PERK DESS f_F estimates alongside corresponding magnitude error images, in a two-compartment simulation where none of the associated estimators incur bias due to model mismatch. Voxels not assigned WM- or GM-like compartmental fractions are masked out in post-processing for display. Table 6.2 presents corresponding sample statistics. 112
6.2 NNLS/RNNLS MESE f_M and ML/PERK DESS f_F estimates alongside corresponding magnitude error images, in a three-compartment simulation where any of the associated estimators could incur bias due to model mismatch. Voxels not assigned WM- or GM-like compartmental fractions are masked out in post-processing for display. Table 6.3 presents corresponding sample statistics. 114
6.3 Representative NNLS and RNNLS f_M estimates from a MESE acquisition alongside a PERK f_F estimate from a precision-optimized DESS acquisition, in the brain of a healthy volunteer. Using similar signal reception imaging parameters, MESE f_M estimates required 40m6s total scan time while DESS f_F estimates required 7m45s total scan time. PERK f_F estimates exhibit less WM variation and more clearly delineate cortical WM/GM boundaries than MESE f_M estimates. Table 6.4 presents corresponding sample statistics computed over manually selected WM and GM ROIs. 117

LIST OF TABLES

3.1 Sample means $\pm$ sample standard deviations of MOM, ML, and RL T_1, c_1 estimates from two simulated SPGR datasets, computed over 3001 WM-like and 1151 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. T_1 values are in milliseconds. c_1 values are unitless. Fig. 3.1 presents corresponding images. 25
3.2 Sample means $\pm$ sample standard deviations of MOM, ML, and RL T_2, c_2 estimates from one simulated DESS dataset, computed over 3001 WM-like and 1151 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. T_2 values are in milliseconds. c_2 values are unitless. Fig. 3.2 presents corresponding images. 27

4.1 Performance summary of different scan profiles, optimized by solving (4.9) subject to scan time constraint $T_{\max} = 41.9$ms. The first row defines each profile. The next four rows describe $\widehat{\mathbf{P}}$. The latter three pairs of rows show how worst-case values degrade from tight to broad ranges. Flip angles are in degrees; all other values are in milliseconds. 37
4.2 Sample means $\pm$ sample standard deviations of $\mathbf{T}_1$ and $\mathbf{T}_2$ ML estimates in WM and GM ROIs of simulated data, compared across different optimized scan profiles. Sample means exhibit insignificant bias, and sample standard deviations are consistent with worst-case standard deviations $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1}$ and $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2}$ reported in Table 4.1. All values are reported in milliseconds. 40
4.3 Phantom within-ROI sample means $\pm$ sample standard deviations of $\mathbf{T}_1$ and $\mathbf{T}_2$ estimates from optimized SPGR/DESS and reference IR/SE scan profiles, versus NIST NMR measurements (*cf.* slide 22 of e-poster corresponding to [2]). For sake of brevity, sample statistics corresponding only to phantom vials within (or nearly within) tight design range $\mathbb{X}^{\mathrm{t}}$ (color-coded orange in Fig. 4.4) are reported. Fig. 4.5 plots sample statistics for all vials. 'V#' abbreviates vial numbers. All values are reported in milliseconds. 48

4.4 Phantom pooled sample standard deviations $\pm$ pooled standard errors of sample standard deviations, from optimized SPGR/DESS scan profiles. Each entry is a measure of uncertainty of a typical voxel's T_1 or T_2 ML estimate, estimated over 10 repeated acquisitions. For sake of brevity, sample statistics corresponding only to phantom vials within (or nearly within) tight design range $\mathbb{X}^{\mathrm{t}}$ (color-coded orange in Fig. 4.4) are reported. 'V#' abbreviates vial numbers. All values are reported in milliseconds. 50
4.5 *Left*: WM/GM ROIs, overlaid on a representative anatomical (coil-combined, IR) image. Separate WM ROIs are distinguished by anterior-right (AR), anterior-left (AL), posterior-right (PR), and posterior-left (PL) directions. Four small anterior (A) cortical GM polygons are pooled into a single ROI. *Right*: Within-ROI sample means $\pm$ within-ROI sample standard deviations of $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from the brain of a healthy volunteer (Fig. 4.6 presents corresponding images). Sample statistics are computed within ROIs indicated in the anatomical image. All values are reported in milliseconds. 51
4.6 Within-ROI sample means $\pm$ within-ROI sample standard deviations of monoexponential $\mathbf{T}_2$ ML estimates, from pairs of *in vivo* SE datasets. Column headers indicate echo times $\mathbf{T}_{\mathrm{E}}$ (ms) of SE datasets. Sample statistics are computed within ROIs indicated in Table 4.5. Single-component $\widehat{T}_2^{\mathrm{ML}}$ estimates in WM depend on SE echo times. 54

5.1 Sample means $\pm$ sample standard deviations (RMSEs) of VPM, PGPM, and PERK m_0, T_1, T_2 estimates, computed in simulation over 7810 WM-like and 9162 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. m_0 values are unitless. T_1, T_2 values are in milliseconds. Figs. 5.3, 5.4, and 5.5 present corresponding images. 74
5.2 *In vivo* sample means $\pm$ sample standard deviations of VPM, PGPM, and PERK T_1, T_2 estimates, computed over color-coded ROIs indicated in Fig. 5.8. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. All values are in milliseconds. 80

6.1 SPGR/DESS flip angles and repetition times that comprise $\widehat{\mathbf{P}}$, a scan parameter matrix designed under total time budget $\sum_{d=1}^{D} T_{\mathrm{R},d} \leq 108$ms for precise f_{F} estimation in WM. For our noise variance measurements, this acquisition is expected to yield 42.5% relative standard deviation in asymptotically unbiased f_{F} estimates from two-compartment signal models. Interestingly, the optimized scan profile omits SPGR scans entirely. 107
6.2 Sample means $\pm$ sample standard deviations (RMSEs) of NNLS/RNNLS MESE f_{M} estimates and ML/PERK DESS f_{F} estimates in a two-compartment simulation where none of the associated estimators incur bias due to model mismatch. Sample statistics are computed over 7810 WM-like and 9162 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. Fig. 6.1 presents corresponding images. 113

6.3 Sample means $\pm$ sample standard deviations (RMSEs) of NNLS/RNNLS MESE f_{M} estimates and ML/PERK DESS f_{F} estimates in a three-compartment simulation where any of the associated estimators could incur bias due to model mismatch. Sample statistics are computed over 7810 WM-like and 9162 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. Fig. 6.2 presents corresponding images. 114

6.4 *Left*: WM/GM ROIs, overlaid on a representative anatomical MESE first-echo image. Separate lateral WM ROIs are distinguished by anterior-right (AR), anterior-left (AL), posterior-right (PR), and posterior-left (PL) directions and are respectively comprised of 90, 79, 182, and 201 voxels. Two internal capsule (IC) polygons are pooled into a single medial WM ROI comprised of 347 voxels. Three small anterior cortical (AC) GM polygons are pooled into a single GM ROI comprised of 78 voxels. *Right*: Sample means $\pm$ sample standard deviations of NNLS/RNNLS f_{M} estimates from a MESE acquisition as well as PERK f_{F} estimates from an optimized DESS acquisition, computed over WM/GM ROIs. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. Fig. 6.3 presents corresponding images. 118

LIST OF APPENDICES

A Multiple-Dataset Complex Coil Combination **125**

B SS-Informed RF Pulse Design **128**

ABSTRACT

Quantitative magnetic resonance imaging (QMRI) produces images of potential MR biomarkers: measurable tissue properties related to physiological processes that characterize the onset and progression of specific disorders. Though QMRI has potential to be more diagnostic than conventional qualitative MRI, QMRI poses challenges beyond those of conventional MRI that limit its feasibility for routine clinical use. This thesis first seeks to address two of these challenges. It then applies these solutions to develop a new method for myelin water imaging, a challenging application that may be specifically indicative of certain white matter disorders.

One challenge that presently precludes widespread clinical adoption of QMRI involves relatively long scan durations: to disentangle potential biomarkers from numerous nuisance MR contrast mechanisms, QMRI typically requires more data than conventional MRI and thus longer scans. Even allowing for long scans, it has previously been unclear how to systematically tune the "knobs" of highly flexible MR acquisitions so as to reliably enable precise biomarker estimation. Chapter 4 formalizes these challenges as a min-max optimal acquisition design problem that seeks scan parameter combinations that robustly enable precise object parameter estimation. It applies this technique to optimize combinations of spoiled gradient-recalled echo (SPGR) and dual-echo steady-state (DESS) scans for T_1, T_2 relaxometry in white matter (WM) and grey matter (GM) regions of the human brain at 3T field strength. Phantom accuracy experiments showed that SPGR/DESS scan combinations are in excellent agreement with reference measurements. Phantom precision experiments showed that trends in T_1, T_2 pooled sample standard deviations

reflect theoretical predictions. *In vivo* experiments showed that in WM and GM, T_1, T_2 estimates from a pair of optimized DESS scans exhibit precision comparable to that of optimized combinations of SPGR and DESS scans. To our knowledge, T_1 maps from DESS acquisitions are new. This example application illustrates that acquisition design can enable new biomarker estimation techniques from established MR pulse sequences, a fact that subsequent chapters exploit.

Another QMRI challenge involves the typically nonlinear dependence of MR signal models on the underlying potential biomarkers of interest: these nonlinearities cause conventional likelihood-based estimators to either scale very poorly with the number of unknowns or risk producing suboptimal estimates due to spurious local minima. Chapter 5 instead introduces a fast, general method for dictionary-free QMRI parameter estimation via regression with kernels (PERK). PERK first uses prior distributions and the nonlinear MR signal model to simulate many parameter-measurement pairs. Inspired by machine learning, PERK then takes these parameter-measurement pairs as labeled training points and learns from them a nonlinear regression function using kernel functions and convex optimization. PERK admits a simple implementation as per-voxel nonlinear lifting of MRI measurements followed by linear minimum mean-squared error regression. Chapter 5 demonstrates PERK for T_1, T_2 estimation using one of the SPGR/DESS acquisitions optimized in Chapter 4. Numerical simulations as well as single-slice phantom and *in vivo* experiments demonstrated that PERK and two well-suited maximum-likelihood (ML) estimators produce comparable T_1, T_2 estimates in WM and GM, but PERK is consistently at least $140\times$ faster. Similar comparisons to an ML estimator in a more challenging QMRI estimation problem (described in Chapter 6) suggest that this $140\times$ acceleration factor will increase by several orders of magnitude for full-volume QMRI estimation problems involving more latent parameters per voxel.

Chapter 6 applies ideas developed in previous chapters to design a new fast method for imaging myelin water content, a potential biomarker for healthy myelin. Since myelin degeneration characterizes certain white matter disorders (*e.g.*, multiple sclerosis), myelin water quantification could improve MRI specificity for monitoring the onset and progression of such demyelinating conditions. Specifically, Chapter 6 first develops a two-compartment DESS signal model and then uses a Bayesian variation of acquisition design (Chapter 4) to optimize a new DESS acquisition for precise myelin water imaging. The precision-optimized acquisition is as fast as conventional SS myelin water imaging acquisitions, but enables $\sim$2-3$\times$ better expected coefficients of variation in fast-relaxing fraction f_{F} estimates. Simulations without model mismatch demonstrate that PERK (Chapter 5) and ML f_{F} estimates from the proposed DESS acquisition exhibit comparable root mean-squared errors, but PERK is more than $500\times$ faster. Simulations with modest levels of model mismatch suggest that *in vivo* differences between DESS and conventional multi-echo spin-echo (MESE) myelin water images may be attributable to either unaccounted bulk-T_1 modeling errors and accounted flip angle variation in MESE estimates and/or the two-compartment assumption in DESS estimates. Nevertheless, *in vivo* experiments are to our knowledge the first to demonstrate lateral WM myelin water content estimates from a fast (3m15s) SS acquisition that are similar to conventional estimates from a slower (32m4s) MESE acquisition.

CHAPTER 1

Introduction

Magnetic resonance imaging (MRI) is a non-invasive tool that has earned widespread clinical adoption due (among other reasons) to its potential for excellent soft tissue contrast, its absence of ionizing radiation, and its flexibility to characterize a diversity of physical phenomena. Despite its numerous advantages, MRI requires highly specialized hardware, ongoing liquid-helium cooling of its superconducting main magnet, and long scan times. For these reasons, MRI is expensive relative to other medical imaging modalities. To better focus expenditures, one broad initiative advocated by the MR community is to increase the *value* of MRI examinations. One popular (and ambitious) measure of value is an MRI acquisition's specificity in distinguishing one disorder from a collection of candidates. The field of quantitative MRI (QMRI) seeks to estimate MR *biomarkers*, or measurable tissue properties that may be indicative of specific disorders of interest.

QMRI has potential to be more informative than conventional MRI. Conventional MRI is qualitative: it produces images comprised of voxels (*i.e.*, three-dimensional pixels) that are informative only relative to each other, not individually. Conventional MRI voxels are qualitative because they directly localize the MR signal, a typically complex function of not only biomarkers but also two types of confounds: *nuisance markers* that characterize undesired signal sources and/or MRI system imperfections; and *acquisition parameters* that characterize the MRI system's tunable "knobs". QMRI seeks to remove confound influence by instead imaging the biomarkers directly. Each QMR image voxel is thus a measurement of a given biomarker at a specific location. QMRI can therefore provide localized biomarker measurements (*e.g.*, myelin water content) related to a specific physiological process (*e.g.*, demyelination) that can through longitudinal study be used to monitor the onset and progression of disease (*e.g.*, multiple sclerosis).

QMRI poses several challenges beyond those of conventional MRI that currently limit its feasibility for routine clinical use. For example, accurate biomarker quantification traditionally requires multiple MR scans and thus long scan times. Furthermore, it has previously been unclear how to tune acquisition parameters of these multiple scans to ensure that biomarkers can be quantified precisely. Finally, MR biomarker quantification is a challenging estimation problem for which efficient algorithms have previously been unavailable. Addressing these challenges is essential for widespread clinical adoption of QMRI.

1.1 Overview

This thesis seeks to address the above challenges by developing an automated workflow for QMRI. We exploit tools from optimization, statistics, and machine learning to develop fast algorithms for quantifying biomarkers that characterize specific physiological processes. We apply this framework to challenging QMRI problems of clinical interest. Our goal is to introduce fast, automated tools that will increase the clinical value of QMRI.

Our solutions to two distinct subproblems in QMRI constitute two stages of our proposed QMRI workflow. Questions in *acquisition design* (Chapters 4, 6) ask how to assemble fast collections of scans that yield data rich in information about physical processes of interest. Questions in *parameter estimation* (Chapters 3, 5) ask how to quickly and reliably quantify biomarkers associated with these relevant physical processes. The overall workflow seeks to first design fast and informative scans based on the application, and to then accurately and precisely estimate clinically relevant biomarkers.

1.2 Organization

The main body of this thesis is organized as follows. Within this main body, Chapters 4-6 organize the key contributions to science of this thesis.

- Chapter 2 reviews relevant background material on MRI and optimization.
- Chapter 3 discusses methods for QMRI parameter estimation from likelihood models and applies these methods for model-based MR relaxometry, a simple and popular application. It partially derives content from conference papers [4, 5].
- Chapter 4 introduces a minimax optimization problem to aid robust and application-specific MR scan selection and optimization for precise latent parameter estimation. It optimizes several practical acquisitions and uses the likelihood-based estimation techniques introduced in Chapter 3 to assess the utility of scan optimization through simulations as well as phantom and *in vivo* experiments. It mainly derives content from published journal paper [6] that extends conference paper [7].
- Chapter 5 introduces a fast, general algorithm for dictionary-free QMRI parameter estimation via regression with kernels (PERK). It demonstrates orders-of-magnitude

acceleration over likelihood-based estimators through simulations as well as phantom and *in vivo* experiments. It also characterizes PERK performance through bias-covariance analysis and several robustness studies. It mainly derives content from accepted journal paper [8] that extends two conference papers [9, 1].

- Chapter 6 introduces a new method for imaging an MR biomarker of clinical interest. It applies ideas developed in earlier chapters to design a new fast QMRI workflow that may be specific to healthy myelin, whose degradation is associated with certain white matter disorders. It demonstrates this new method of potentially myelin-specific imaging in simulations and *in vivo* experiments. It partially derives content from in-preparation journal paper [10] that extends conference paper [11].

- Chapter 7 suggests several future research directions.

The appendices contain unpublished, less mature work and are organized as follows:

- Appendix A proposes an algorithm for simultaneously coil-combining a collection of MR coil image datasets without prior knowledge of coil sensitivity maps. Several chapters in the main body used a variation of this algorithm for coil data combination.

- Appendix B develops from first principles a new model for the influence of RF pulses on the steady-state (SS) transverse magnetization and then proposes two algorithms for SS-informed RF pulse design.

CHAPTER 2

Background

This chapter focuses only on background information pertinent to multiple subsequent chapters. We present further topic-specific information at the beginnings of corresponding chapters. Section 2.1 places emphasis on reviewing necessary MR fundamentals, and Section 2.2 proceeds to a shorter discussion regarding optimization as it pertains to QMRI.

2.1 Relevant MR Physics

This section begins with the fundamental Bloch equations and derives the signal models associated with two MR pulse sequences used extensively in this thesis. Our coverage of MRI is far from comprehensive, and omits fundamental but tangential topics such as signal localization. We refer the interested reader to books such as [12, 13, 14].

2.1.1 Bloch Equations

The Bloch equations [15] describe the macroscopic magnetization dynamics of *spin*, or (loosely) atomic nuclei with nonzero angular momentum and thus nonzero magnetic moment, *e.g.* ^{1}H. If the dominant source of magnetic flux arises (as is typical in MRI) from a main magnetic field that is oriented along the z-axis, the equations read

$$\frac{\partial}{\partial t} m_{xy}(\mathbf{r}, t) = i\gamma(m_z(\mathbf{r}, t) b_{xy}(\mathbf{r}, t) - m_{xy}(\mathbf{r}, t) b_z(\mathbf{r}, t)) - \frac{m_{xy}(\mathbf{r}, t)}{T_2(\mathbf{r})}; \tag{2.1}$$

$$\frac{\partial}{\partial t} m_z(\mathbf{r}, t) = \gamma(m_x(\mathbf{r}, t) b_y(\mathbf{r}, t) - m_y(\mathbf{r}, t) b_x(\mathbf{r}, t)) - \frac{m_z(\mathbf{r}, t) - m_0(\mathbf{r})}{T_1(\mathbf{r})}. \tag{2.2}$$

Here, $m_{xy}(\mathbf{r}, t) := m_x(\mathbf{r}, t) + im_y(\mathbf{r}, t) \in \mathbb{C}$ and $m_z(\mathbf{r}, t) \in \mathbb{R}$ are the transverse and longitudinal components of the magnetization vector at position $\mathbf{r} := [x, y, z]^\mathsf{T} \in \mathbb{R}^3$ and time $t \geq 0$; $b_{xy}(\mathbf{r}, t) := b_x(\mathbf{r}, t) + ib_y(\mathbf{r}, t) \in \mathbb{C}$ and $b_z(\mathbf{r}, t) \in \mathbb{R}$ are the transverse and longitudinal components (in an inertial reference frame) of the applied magnetic field; $T_1(\mathbf{r})$

and $T_2(\mathbf{r})$ are spin-lattice and spin-spin relaxation time constants; $m_0(\mathbf{r})$ is the equilibrium magnetization and is proportional to the density of spins per unit volume as well as the main field strength; γ is the gyromagnetic ratio; and $i := \sqrt{-1}$. As written, (2.1)-(2.2) specifically model the temporal dynamics of a single spin *isochromat*, or collection of macroscopically similar spins; later chapters consider higher-order effects such as multiple (possibly interacting) isochromat compartments (Chapter 6).

It is often convenient to study Bloch dynamics in a non-inertial reference frame rotating clockwise about the z-axis at Larmor frequency $\omega_0 := \gamma B_0$, where $B_0\hat{k}$ is the (nearly uniform) main magnetic field. In these coordinates, the apparent transverse magnetic field $b'_{xy}(\mathbf{r},t) = b'_x(\mathbf{r},t) + ib'_y(\mathbf{r},t) := b_{xy}(\mathbf{r},t)e^{i\omega_0 t}$ transforms only in phase, but the apparent longitudinal magnetic field $b'_z(\mathbf{r},t) := b_z(\mathbf{r},t) - B_0$ is greatly reduced in magnitude. The magnetization components transform more simply as $m'_{xy}(\mathbf{r},t) = m'_x(\mathbf{r},t) + im'_y(\mathbf{r},t) := m_{xy}(\mathbf{r},t)e^{i\omega_0 t}$ and $m'_z(\mathbf{r},t) := m_z(\mathbf{r},t)$. Remarkably, inserting these coordinate transformations into (2.1)-(2.2) does not change the form of the dynamical equations:

$$\frac{\partial}{\partial t}m'_{xy}(\mathbf{r},t) = i\gamma\big(m'_z(\mathbf{r},t)b'_{xy}(\mathbf{r},t) - m'_{xy}(\mathbf{r},t)b'_z(\mathbf{r},t)\big) - \frac{m'_{xy}(\mathbf{r},t)}{T_2(\mathbf{r})}; \tag{2.3}$$

$$\frac{\partial}{\partial t}m'_z(\mathbf{r},t) = \gamma\big(m'_x(\mathbf{r},t)b'_y(\mathbf{r},t) - m'_y(\mathbf{r},t)b'_x(\mathbf{r},t)\big) - \frac{m'_z(\mathbf{r},t) - m_0(\mathbf{r})}{T_1(\mathbf{r})}. \tag{2.4}$$

It thus suffices to consider how perturbations $\mathbf{b}'(\mathbf{r},t)$ to main field $B_0\hat{k}$ influence rotating-frame magnetization $\mathbf{m}'(\mathbf{r},t)$ via Eqs. (2.3)-(2.4). The inertial-frame magnetization $\mathbf{m}(\mathbf{r},t)$ is then easily constructed via $m_{xy}(\mathbf{r},t) = m'_{xy}(\mathbf{r},t)e^{-i\omega_0 t}$ and $m_z(\mathbf{r},t) = m'_z(\mathbf{r},t)$.

It is challenging to explicitly solve Eqs. (2.3)-(2.4) for arbitrary field perturbations $\mathbf{b}'(\mathbf{r},t)$. We discuss relevant special cases in the following.

2.1.1.1 Non-Selective Excitation

Here, we derive solutions to Eqs. (2.3)-(2.4) in the case of short, spatially non-selective excitations. We take the following common assumptions:

- We assume negligible spatial variation in the main magnetic field, so $b'_z(\mathbf{r},t) \approx 0$.
- We assume the transverse field separates in position and time; oscillates at the Larmor frequency (commonly in the radiofrequency (RF) range); and aligns at initial time $t \leftarrow t_0$ with the x-axis. Together, these assumptions restrict the so-called RF excitation to take form $b'_{xy}(\mathbf{r},t) \approx \kappa^{\mathrm{t}}(\mathbf{r})b'_{1,x}(t)\hat{\imath} + 0\hat{\jmath}$, where $\kappa^{\mathrm{t}}(\mathbf{r}) \in \mathbb{R}$ is the RF transmit coil spatial variation and $b'_{1,x}(t) \in \mathbb{R}$ is the RF excitation envelope.
- We assume that the duration T_{P} of RF excitation (often T_{P} ~1ms) is much shorter than relaxation time constants (typically T_1 ~1000ms and T_2 ~50ms in brain tissue)

and thus neglect relaxation effects during excitation.

Under these assumptions, Eqs. (2.3)-(2.4) reduce to the linear system

$$\frac{\partial}{\partial t}\begin{bmatrix} m'_x(\mathbf{r},t) \\ m'_y(\mathbf{r},t) \\ m'_z(\mathbf{r},t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \gamma\kappa^{\mathrm{t}}(\mathbf{r})b'_{1,x}(t) \\ 0 & -\gamma\kappa^{\mathrm{t}}(\mathbf{r})b'_{1,x}(t) & 0 \end{bmatrix} \begin{bmatrix} m'_x(\mathbf{r},t) \\ m'_y(\mathbf{r},t) \\ m'_z(\mathbf{r},t) \end{bmatrix}. \tag{2.5}$$

Eq. (2.5) admits the simple solution (for $t \geq t_0$)

$$\begin{bmatrix} m'_x(\mathbf{r},t) \\ m'_y(\mathbf{r},t) \\ m'_z(\mathbf{r},t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha(\mathbf{r},t;t_0)) & \sin(\alpha(\mathbf{r},t;t_0)) \\ 0 & -\sin(\alpha(\mathbf{r},t;t_0)) & \cos(\alpha(\mathbf{r},t;t_0)) \end{bmatrix} \begin{bmatrix} m'_x(\mathbf{r},t_0) \\ m'_y(\mathbf{r},t_0) \\ m'_z(\mathbf{r},t_0) \end{bmatrix}, \tag{2.6}$$

where $\mathbf{m}'(\mathbf{r},t_0) := \left[m'_x(\mathbf{r},t_0), m'_y(\mathbf{r},t_0), m'_z(\mathbf{r},t_0)\right]^{\mathsf{T}}$ is the initial magnetization and

$$\alpha(\mathbf{r},t;t_0) := \gamma\kappa^{\mathrm{t}}(\mathbf{r}) \int_{t_0}^{t} b'_{1,x}(\tau)\,\mathrm{d}\,\tau \tag{2.7}$$

is the nutation (or "flip") angle at time t. Eq. (2.6) reveals that on-resonance RF excitation causes the magnetization vector to rotate clockwise about an axis parallel to the direction of excitation. The nutation angle accumulated over an RF pulse of duration T_{P} is often decomposed as $\alpha(\mathbf{r},t_0+T_{\mathrm{P}};t_0) =: \alpha_0\kappa^{\mathrm{t}}(\mathbf{r})$, where α_0 is a prescribed nominal flip angle.

For deriving signal models in later sections, it is convenient and intuitive to define matrix operators that summarize relevant dynamics. Here, we rewrite Eq. (2.6) as

$$\mathbf{m}'(\mathbf{r},t) = \mathbf{R}_{x'}(\alpha(\mathbf{r},t;t_0))\mathbf{m}'(\mathbf{r},t_0), \tag{2.8}$$

where $\mathbf{R}_{x'}(\alpha(\mathbf{r},t;t_0))$ denotes a clockwise rotation of angle $\alpha(\mathbf{r},t;t_0)$ about the x'-axis.

2.1.1.2 Free Precession and Relaxation

Next, we derive solutions to the rotating-frame Bloch equations when no RF excitation is present, *i.e.* $b'_{xy}(\mathbf{r},t) \approx 0$. In this case, Eqs. (2.3)-(2.4) decouple, yielding separate dynamical equations for the transverse and longitudinal magnetization components:

$$\frac{\partial}{\partial t}m'_{xy}(\mathbf{r},t) = -i\gamma m'_{xy}(\mathbf{r},t)b'_z(\mathbf{r},t) - \frac{m'_{xy}(\mathbf{r},t)}{T_2(\mathbf{r})}; \tag{2.9}$$

$$\frac{\partial}{\partial t}m'_z(\mathbf{r},t) = -\frac{m'_z(\mathbf{r},t) - m_0(\mathbf{r})}{T_1(\mathbf{r})}. \tag{2.10}$$

Eqs. (2.9)-(2.10) admit simple solutions with no further assumptions:

$$m'_{xy}(\mathbf{r},t) = m'_{xy}(\mathbf{r},t_0)e^{-(t-t_0)/T_2(\mathbf{r})}e^{-i\phi'(\mathbf{r},t;t_0)}; \tag{2.11}$$
$$m'_{z}(\mathbf{r},t) = m'_{z}(\mathbf{r},t_0)e^{-(t-t_0)/T_1(\mathbf{r})} + m_0(\mathbf{r})\big(1-e^{-(t-t_0)/T_1(\mathbf{r})}\big), \tag{2.12}$$

where $m'_{xy}(\mathbf{r},t_0)$ and $m'_z(\mathbf{r},t_0)$ are the initial magnetization components and

$$\phi'(\mathbf{r},t;t_0) := \gamma \int_{t_0}^{t} b'_z(\mathbf{r},\tau)\,\mathrm{d}\,\tau \tag{2.13}$$

denotes the phase accumulation due to main field inhomogeneity (often called off-resonance effects). Eq. (2.11) reveals that without RF excitations, the transverse magnetization $m'_{xy}(\mathbf{r},t)$ relaxes to zero exponentially fast with time constant $T_2(\mathbf{r})$, while accruing phase due to off-resonance effects. Eq. (2.12) similarly reveals that without RF excitations, longitudinal magnetization $m'_z(\mathbf{r},t)$ recovers to $m_0(\mathbf{r})$ exponentially fast with time constant $T_1(\mathbf{r})$.

As in Section 2.1.1.2, we rewrite Eqs. (2.11)-(2.12) for $t \geq t_0$ using matrix operators:

$$\mathbf{m}'(\mathbf{r},t) = \mathbf{R}_{z'}(\phi'(\mathbf{r},t;t_0))\mathbf{E}(\mathbf{r},t;t_0)\mathbf{m}'(\mathbf{r},t_0) + \mathbf{m}_0(\mathbf{r},t;t_0) \tag{2.14}$$

where $\mathbf{m}_0(\mathbf{r},t;t_0) := m_0(\mathbf{r})\big(1-e^{-(t-t_0)/T_1(\mathbf{r})}\big)\hat{k}$;

$$\mathbf{R}_{z'}(\phi'(\mathbf{r},t;t_0)) := \begin{bmatrix} \cos(\phi'(\mathbf{r},t;t_0)) & \sin(\phi'(\mathbf{r},t;t_0)) & 0 \\ -\sin(\phi'(\mathbf{r},t;t_0)) & \cos(\phi'(\mathbf{r},t;t_0)) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.15}$$

denotes a clockwise rotation of angle $\phi'(\mathbf{r},t;t_0)$ about the z'-axis; and

$$\mathbf{E}(\mathbf{r},t;t_0) := \begin{bmatrix} e^{-(t-t_0)/T_2(\mathbf{r})} & 0 & 0 \\ 0 & e^{-(t-t_0)/T_2(\mathbf{r})} & 0 \\ 0 & 0 & e^{-(t-t_0)/T_1(\mathbf{r})} \end{bmatrix} \tag{2.16}$$

is an exponential relaxation operator. Section 2.1.2 (and later chapters) use matrix dynamical representations (2.8) and (2.14) to succinctly describe pulse sequence signal models.

2.1.2 Steady-State Sequences

MRI experiments typically involve repeated cycles of (pulsed) RF excitation; signal localization (not discussed here); and transverse T_2 relaxation and free precession, alongside (relatively slow) longitudinal T_1 recovery. We can build models of the received MR signal by considering the magnetization dynamics induced by specific pulse sequences.

Classical pulse sequences use relatively long cycle repetition times T_{R} to ensure near-complete T_1 recovery of the magnetization vector back to equilibrium state $m_0(\mathbf{r})\hat{k}$ prior to the start of each RF cycle. For such long-T_{R} sequences, it suffices to approximate the magnetization as fully recovered (*i.e.*, $\mathbf{m}'(\mathbf{r}, t_0 + rT_{\mathrm{R}}) \approx m_0(\mathbf{r})\hat{k}, \forall r \in \{0, 1, 2, \dots\}$) just prior to each RF excitation. This approximation yields a sequence of initial conditions and allows computation of the magnetization at corresponding times of data acquisition via direct application of Bloch dynamics (2.8) and (2.14). Resulting signal models are typically simple expressions of relaxation parameters $T_1(\mathbf{r})$ and $T_2(\mathbf{r})$; however, model accuracy often depends strongly on the long-T_{R} assumption, which requires long acquisitions.

Steady-state (SS) sequences [16] utilize short T_{R}, and can thus achieve much faster scan times. Due to short repetition times, SS sequences achieve only partial T_1 recovery in between RF excitations; thus, their magnetization responses do not obey the simple classical initial conditions (for the second RF cycle onwards). Although their transient magnetization dynamics can be complicated, SS sequences produce (under certain assumptions [17]) long-time magnetization responses that eventually[1] achieve a steady-state condition:

$$\lim_{t_0 \to \infty} \mathbf{m}'(\mathbf{r}, t_0 + rT_{\mathrm{R}}) = \mathbf{m}'(\mathbf{r}, t_0), \tag{2.17}$$

where repetition count $r \in \{1, 2, \dots\}$ for fixed RF excitations and off-resonance induced phase increments (as is assumed in the following). Subsections 2.1.2.1 and 2.1.2.2 use SS condition (2.17) and Bloch equation matrix operators introduced in (2.8) and (2.14) to derive long-time signal models for Spoiled Gradient-Recalled Echo (SPGR) and Dual-Echo Steady-State (DESS), two SS pulse sequences useful for quantitative MRI.

2.1.2.1 Spoiled Gradient-Recalled Echo (SPGR) Sequence

SPGR [19] is a fast pulse sequence that repeats cycles of fixed RF excitation (such that $b'_{1,x}(t + rT_{\mathrm{R}}) = b'_{1,x}(t), \forall t \in [t_0, t_0 + T_{\mathrm{P}}], r \in \{1, 2, \dots\}$); data acquisition; relaxation and recovery; and residual transverse magnetization "spoiling" (discussed later). Here we develop a simple and popular steady-state SPGR signal model.

Let $\mathbf{m}'(\mathbf{r}, t_0)$ denote the magnetization at an initial time t_0 selected well into the steady-state and just prior to excitation. The SPGR sequence first applies an RF excitation, which

[1]The progression to steady state takes on the order of $5T_2/T_{\mathrm{R}}$ RF cycles [17], typically a small but not insignificant period during which data acquisition is often foregone. This transition can (in some cases) be accelerated by prepending SS sequences with tailored "magnetization-catalyzing" modules [18].

rotates the initial magnetization as per (2.8):

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \mathbf{R}_{x'}(\alpha(\mathbf{r}, t_0 + T_\mathrm{P}; t_0))\mathbf{m}'(\mathbf{r}, t_0). \tag{2.18}$$

The excited magnetization then precesses and relaxes as per (2.14) until data acquisition, defined to occur at "echo time" $T_\mathrm{E} \in \left[\frac{T_\mathrm{P}}{2}, T_\mathrm{R}\right]$ after the (midpoint of) RF excitation:

$$\begin{aligned}\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) = \mathbf{R}_{z'}\left(\phi'\left(\mathbf{r}, \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; T_\mathrm{P}\right)\right)\mathbf{E}\left(\mathbf{r}, \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; T_\mathrm{P}\right)\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) \\ + \mathbf{m}_0\left(\mathbf{r}, \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; T_\mathrm{P}\right).\end{aligned} \tag{2.19}$$

The remaining transverse magnetization is spoiled[2] following signal reception[3] while the longitudinal component is unaffected. We model an ideal spoiling operation as

$$\mathbf{S}\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right), \text{ where } \qquad \mathbf{S} := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.20}$$

After spoiling, the longitudinal magnetization (partially) recovers until $t \leftarrow t_0 + T_\mathrm{R}$:

$$\begin{aligned}\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{R}) = \mathbf{R}_{z'}\left(\phi'\left(\mathbf{r}, T_\mathrm{R}; \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right)\right)\mathbf{E}\left(\mathbf{r}, T_\mathrm{R}; \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right)\mathbf{S}\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) \\ + \mathbf{m}_0\left(\mathbf{r}, T_\mathrm{R}; \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right).\end{aligned} \tag{2.21}$$

In steady-state, one cycle of excitation, acquisition, spoiling, and recovery returns the magnetization back to its initial state. We enforce this through the steady-state condition

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \mathbf{R}_{x'}(\alpha(\mathbf{r}, t_0 + T_\mathrm{P}; t_0))\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{R}) \tag{2.22}$$

[2]Transverse signal spoiling is often (nearly) achieved in practice using *gradient fields*, or strong induced field inhomogeneities that cause rapid transverse signal dephasing, in tandem with RF excitations that additionally impart nonlinear (often quadratically increasing) transverse magnetization phase [19]. Though the nonlinear RF phase used in so-called "RF-spoiling" prevents any one spin from reaching a true steady-state, the signal integrated over a typically-sized voxel achieves SS-like behavior [20].

[3]MRI signal reception also uses gradient fields to spatially localize signal; thus signal reception would in general impart additional magnetization phase. However, we assume hereafter that MR signal is received using "balanced" imaging gradients that impart zero net phase per repetition cycle; under this restriction, balanced imaging gradients have a negligible net affect on the magnetization vector.

which yields an algebraic system of equations. When it exists, the solution is

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \frac{1}{1 - e^{-(T_\mathrm{R}-T_\mathrm{P})/T_1(\mathbf{r})} \cos(\alpha(\mathbf{r}))} \begin{bmatrix} 0 \\ m_0(\mathbf{r}) \sin(\alpha(\mathbf{r}))\big(1 - e^{-(T_\mathrm{R}-T_\mathrm{P})/T_1(\mathbf{r})}\big) \\ m_0(\mathbf{r}) \cos(\alpha(\mathbf{r}))\big(1 - e^{-(T_\mathrm{R}-T_\mathrm{P})/T_1(\mathbf{r})}\big) \end{bmatrix}, \tag{2.23}$$

where $\alpha(\mathbf{r}) := \alpha(\mathbf{r}, t_0 + T_\mathrm{P}; t_0)$ is a slight abuse of notation. Remarkably, the SPGR steady-state magnetization immediately after excitation is approximately independent of both off-resonance effects and $T_2(\mathbf{r})$. Researchers more often cite the expression

$$\begin{aligned} m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{P}) &= m'_x(\mathbf{r}, t_0 + T_\mathrm{P}) + i m'_y(\mathbf{r}, t_0 + T_\mathrm{P}) \\ &= \frac{i m_0(\mathbf{r}) \sin(\alpha(\mathbf{r}))\big(1 - e^{-T_\mathrm{R}/T_1(\mathbf{r})}\big)}{1 - e^{-T_\mathrm{R}/T_1(\mathbf{r})} \cos(\alpha(\mathbf{r}))} \end{aligned} \tag{2.24}$$

for the complex transverse magnetization as it modifies (2.23) to include a simple first-order correction for unaccounted T_1 recovery during the RF pulse. Substituting (2.24) into (2.19) yields an expression for the transverse magnetization at the echo time:

$$\begin{aligned} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) &= m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{P}) e^{-(T_\mathrm{E}-T_\mathrm{P}/2)/T_2(\mathbf{r})} e^{-i\phi'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; t_0 + T_\mathrm{P}\right)} \\ &\approx m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{P}) e^{-T_\mathrm{E}/T_2(\mathbf{r})} e^{-i\phi'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; t_0 + \frac{T_\mathrm{P}}{2}\right)}, \end{aligned} \tag{2.25}$$

where the approximation again keeps in line with literature expressions.

The received signal is approximately proportional to the integrated transverse magnetization over a volume $\mathbb{V}$. To derive expressions, we take a few more usual assumptions:

- We assume that the signal is localized to a scale over which there is off-resonance phase variation, but minimal variation of $m_0(\mathbf{r})$, $T_1(\mathbf{r})$, $T_2(\mathbf{r})$, and $\alpha(\mathbf{r})$. This assumption is reasonable[4] when describing the signal arising from a typical voxel.
- We assume that (free-precession) off-resonance phase grows linearly with time, *i.e.* $\phi'\big(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; t_0 + \frac{T_\mathrm{P}}{2}\big) \approx \omega'(\mathbf{r}) T_\mathrm{E}$. We further assume that off-resonance frequency $\omega'(\mathbf{r})$ is distributed over the localized voxel as $\mathsf{p}_{\omega'} := \mathrm{Cauchy}(\bar{\omega}', R'_2)$, where $\bar{\omega}'(\mathbf{r})$ is the median off-resonance frequency and $R'_2(\mathbf{r})$ is the broadening bandwidth.

With these additional assumptions, the received steady-state SPGR (noiseless) signal model

[4] Model mismatch due to within-voxel spatial variation of relaxation parameters can be significant, especially for large voxels. Chapter 6 studies so-called partial volume effects and uses them for QMRI.

for a typically sized voxel centered at position $\mathbf{r}$ is (to within constants):

$$s_{\mathrm{S}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \,\mathrm{d}^3\,\mathbf{r} \tag{2.26}$$

$$\begin{aligned} &\approx m_{xy}(\mathbf{r}, t_0 + T_{\mathrm{P}}) e^{-T_{\mathrm{E}}/T_2(\mathbf{r})} \int_{\mathbb{R}} e^{-i\omega' T_{\mathrm{E}}} \mathsf{p}_{\omega'}(\omega') \,\mathrm{d}\,\omega' \\ &= m_{xy}(\mathbf{r}, t_0 + T_{\mathrm{P}}) e^{-T_{\mathrm{E}}/T_2(\mathbf{r})} e^{-R'_2(\mathbf{r})T_{\mathrm{E}} - i\bar{\omega}'(\mathbf{r})T_{\mathrm{E}}} \\ &= \frac{i m_0(\mathbf{r}) \sin(\alpha(\mathbf{r}))\left(1 - e^{-T_{\mathrm{R}}/T_1(\mathbf{r})}\right)}{1 - e^{-T_{\mathrm{R}}/T_1(\mathbf{r})} \cos(\alpha(\mathbf{r}))} e^{-T_{\mathrm{E}}/T_2^*(\mathbf{r})} e^{-i\bar{\omega}'(\mathbf{r})T_{\mathrm{E}}}, \end{aligned} \tag{2.27}$$

where $T_2^*(\mathbf{r}) := \left(\frac{1}{T_2} + R'_2\right)^{-1}$ is a modified spin-spin relaxation time that accounts for additional transverse magnetization decay due to off-resonance effects.

2.1.2.2 Dual-Echo Steady-State (DESS) Sequence

DESS [21, 22] is a fast pulse sequence that interlaces fixed, constant-phase RF excitations with fixed dephasing "gradients" (*i.e.*, induced main field inhomogeneities that vary nearly linearly with space) to produce two distinct signals per RF excitation. Here we develop simple steady-state DESS signal models.

As in Subsection 2.1.2.1, let $\mathbf{m}'(\mathbf{r}, t_0)$ denote the magnetization at an initial time t_0 selected well into the steady-state and just prior to excitation. The DESS sequence first applies a fixed RF rotation $\alpha(\mathbf{r}) := \alpha(\mathbf{r}, t_0 + rT_{\mathrm{R}} + T_{\mathrm{P}}; t_0 + rT_{\mathrm{R}}), \forall r \in \{0, 1, 2, \dots\}$:

$$\mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{P}}) = \mathbf{R}_{x'}(\alpha(\mathbf{r}))\mathbf{m}'(\mathbf{r}, t_0). \tag{2.28}$$

The excited transverse magnetization contributes to a first acquired signal; dephases (but does not spoil completely) due to gradient dephasing[5] and contributes again to a second (smaller, but nonzero) acquired signal. Since (with proper selection) dephasing gradients mainly contribute to off-resonance phase accrual, the net effect after data acquisition and gradient dephasing is well described simply by precession and relaxation:

$$\mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{R}}) = \mathbf{R}_{z'}(\phi'(\mathbf{r}))\mathbf{E}(\mathbf{r}, T_{\mathrm{R}}; T_{\mathrm{P}})\mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{P}}) + \mathbf{m}_0(\mathbf{r}, T_{\mathrm{R}}; T_{\mathrm{P}}), \tag{2.29}$$

where the abbreviation $\phi'(\mathbf{r}) := \phi'(\mathbf{r}, t_0 + (r+1)T_{\mathrm{R}}; t_0 + rT_{\mathrm{R}} + T_{\mathrm{P}}), \forall r \in \{0, 1, 2, \dots\}$ implies fixed phase accrual (due to gradient dephasing, field inhomogeneity, and other

[5] It is worth distinguishing gradient dephasing (commonly but somewhat misleadingly referred to as gradient spoiling) from RF spoiling. Gradient dephasing (used in DESS) primarily affects magnetization phase and is modeled simply as precession. RF spoiling (used in SPGR) combines gradient dephasing with nonlinear RF phase cycling and suppresses magnetization magnitude in steady-state.

unaccounted effects) over each repetition cycle.

In steady-state, one cycle of excitation, first acquisition, gradient spoiling, second acquisition, and (partial) recovery returns the magnetization back to its initial state. We enforce this through the steady-state condition

$$\mathbf{m}'(\mathbf{r}, t_0) = \mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{R}}) \tag{2.30}$$

which yields an algebraic system of equations. When it exists, the solution gives the steady-state magnetization just prior to RF excitation:

$$\mathbf{m}'(\mathbf{r}, t_0) = \begin{bmatrix} E_2(\mathbf{r}, T_{\mathrm{F}}) \sin\alpha(\mathbf{r}) \sin\phi'(\mathbf{r}) \\ -E_2(\mathbf{r}, T_{\mathrm{F}}) \sin\alpha(\mathbf{r}) (E_2(\mathbf{r}, T_{\mathrm{F}}) - \cos\phi'(\mathbf{r})) \\ 1 - E_2(\mathbf{r}, T_{\mathrm{F}}) \cos\phi'(\mathbf{r}) + E_2(\mathbf{r}, T_{\mathrm{F}}) \cos\alpha(\mathbf{r}) (E_2(\mathbf{r}, T_{\mathrm{F}}) - \cos\phi'(\mathbf{r})) \end{bmatrix} q(\mathbf{r}, T_{\mathrm{F}}), \tag{2.31}$$

where $T_{\mathrm{F}} := T_{\mathrm{R}} - T_{\mathrm{P}}$ is the free precession interval; $E_1(\mathbf{r}, t) := e^{-t/T_1(\mathbf{r})}$ and $E_2(\mathbf{r}, t) := e^{-t/T_2(\mathbf{r})}$ are relaxation operators; and $q(\mathbf{r}, t) :=$

$$\frac{m_0(\mathbf{r})(1 - E_1(\mathbf{r}, t))}{(1 - E_1(\mathbf{r}, t)\cos\alpha(\mathbf{r}))(1 - E_2(\mathbf{r}, t)\cos\phi'(\mathbf{r})) - E_2(\mathbf{r}, t)(E_1(\mathbf{r}, t) - \cos\alpha(\mathbf{r}))(E_2(\mathbf{r}, t) - \cos\phi'(\mathbf{r}))}.$$

Substituting (2.31) into (2.28) produces a similar expression for the steady-state magnetization immediately following RF excitation:

$$\mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{P}}) = \begin{bmatrix} E_2(\mathbf{r}, T_{\mathrm{F}}) \sin\alpha(\mathbf{r}) \sin\phi'(\mathbf{r}) \\ \sin\alpha(\mathbf{r}) (1 - E_2(\mathbf{r}, T_{\mathrm{F}}) \cos\phi'(\mathbf{r})) \\ \cos\alpha(\mathbf{r}) (1 - E_2(\mathbf{r}, T_{\mathrm{F}}) \cos\phi'(\mathbf{r})) + E_2(\mathbf{r}, T_{\mathrm{F}}) (E_2(\mathbf{r}, T_{\mathrm{F}}) - \cos\phi'(\mathbf{r})) \end{bmatrix} q(\mathbf{r}, T_{\mathrm{F}}). \tag{2.32}$$

The transverse magnetizations before and after RF excitation are then

$$m'_{xy}(\mathbf{r}, t_0) = -i \sin\alpha(\mathbf{r}) E_2(\mathbf{r}, T_{\mathrm{R}}) \Big(E_2(\mathbf{r}, T_{\mathrm{R}}) - e^{-i\phi'(\mathbf{r})} \Big) q(\mathbf{r}, T_{\mathrm{R}}); \tag{2.33}$$

$$m'_{xy}(\mathbf{r}, t_0 + T_{\mathrm{P}}) = +i \sin\alpha(\mathbf{r}) \Big(1 - E_2(\mathbf{r}, T_{\mathrm{R}}) e^{i\phi'(\mathbf{r})} \Big) q(\mathbf{r}, T_{\mathrm{R}}), \tag{2.34}$$

where (2.33)-(2.34) include simple first-order corrections for yet-unaccounted relaxation and recovery during excitation. Frequently, the DESS signals are acquired at symmetric echo times T_{E} before and after the center of each RF pulse. Substituting (2.34) into (2.9) gives the magnetization at the data acquisition time after RF excitation:

$$
\begin{aligned}
m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) &= m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{P}) e^{-(T_\mathrm{E} - T_\mathrm{P}/2)/T_2(\mathbf{r})} e^{-i\phi'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; t_0 + T_\mathrm{P}\right)} \\
&\approx m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{P}) e^{-T_\mathrm{E}/T_2(\mathbf{r})} e^{-i\phi'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; t_0 + \frac{T_\mathrm{P}}{2}\right)} \qquad (2.35) \\
&\approx m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{P}) e^{-T_\mathrm{E}/T_2(\mathbf{r})} e^{-i\omega'(\mathbf{r}) T_\mathrm{E}}, \qquad (2.36)
\end{aligned}
$$

where in (2.35) we again approximately correct for relaxation during excitation and in (2.36) we assume linear off-resonance phase accrual during free precession. To compute the magnetization at the acquisition time before excitation, we consider the free precession and relaxation that occurs between[6] signal reception and excitation:

$$
m'_{xy}(\mathbf{r}, t_0) = m'_{xy}\left(\mathbf{r}, t_0 - \left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\right) e^{-(T_\mathrm{E} - T_\mathrm{P}/2)/T_2(\mathbf{r})} e^{-i\phi'\left(\mathbf{r}, t_0; t_0 - \left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\right)}. \qquad (2.37)
$$

Rearranging (2.37) and applying approximations similar to those of (2.35)-(2.36),

$$
\begin{aligned}
m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} - T_\mathrm{E}\right) &= m'_{xy}(\mathbf{r}, t_0) e^{+(T_\mathrm{E} - T_\mathrm{P}/2)/T_2(\mathbf{r})} e^{+i\phi'\left(\mathbf{r}, t_0; t_0 - \left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\right)} \\
&\approx m'_{xy}(\mathbf{r}, t_0) e^{+T_\mathrm{E}/T_2(\mathbf{r})} e^{+i\phi'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2}; t_0 + \frac{T_\mathrm{P}}{2} - T_\mathrm{E}\right)} \qquad (2.38) \\
&\approx m'_{xy}(\mathbf{r}, t_0) e^{+T_\mathrm{E}/T_2(\mathbf{r})} e^{+i\omega'(\mathbf{r}) T_\mathrm{E}}. \qquad (2.39)
\end{aligned}
$$

The received signal is approximately proportional to the integrated transverse magnetization over a volume $\mathbb{V}$. To derive expressions, we retake assumptions used in Subsection 2.1.2.1 and append an additional assumption on the full-repetition phase accrual $\phi'(\mathbf{r})$:

- We assume that the signal is localized to a scale over which there is off-resonance phase variation, but minimal variation of $m_0(\mathbf{r})$, $T_1(\mathbf{r})$, $T_2(\mathbf{r})$, and $\alpha(\mathbf{r})$. This assumption is reasonable[7] when describing the signal arising from a typical voxel.
- We assume that free precession off resonance frequency $\omega'(\mathbf{r})$ is distributed over the localized voxel as $\mathsf{p}_{\omega'} \leftarrow \mathrm{Cauchy}(\bar{\omega}', R_2')$, where $\bar{\omega}'(\mathbf{r})$ is the median off-resonance frequency and $R_2'(\mathbf{r})$ is the broadening bandwidth.
- We assume that the dephasing gradient imparts an integral number n_cyc of across-

[6] Observe that we do not attempt to express the magnetization prior to (the next) RF excitation by simply operating on the magnetization after (the current) RF excitation with further precession and relaxation. The reason is due to the intermediate dephasing gradient, which causes phase accrual in excess of off-resonance effects and thus forbids an approximation akin to (2.36).

[7] Model mismatch due to within-voxel spatial variation of relaxation parameters can be significant, especially for large voxels. Chapter 6 studies so-called partial volume effects and uses them for QMRI.

voxel phase cycles[8] such that full-repetition phase accrual $\phi'(\mathbf{r})$ is distributed essentially uniformly as $\mathsf{p}_{\phi'} \leftarrow \mathrm{unif}(0, 2\pi n_{\mathrm{cyc}}), n_{\mathrm{cyc}} \in \{1, 2, 3, \dots\}$.

With these assumptions, the received steady-state DESS (noiseless) signal models for a typically sized voxel centered at position $\mathbf{r}$ are (to within constants):

$$
\begin{aligned}
s_{\mathrm{D}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) &\propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \,\mathrm{d}^3\,\mathbf{r} && (2.40)\\
&\approx \int_{\mathbb{R}}\int_{\mathbb{R}} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right)\mathsf{p}_{\phi'}(\phi')\mathsf{p}_{\omega'}(\omega') \,\mathrm{d}\,\phi' \,\mathrm{d}\,\omega' \\
&\approx e^{-T_{\mathrm{E}}/T_2(\mathbf{r})} \int_{\mathbb{R}} m'_{xy}(\mathbf{r}, t_0 + T_{\mathrm{P}})\mathsf{p}_{\phi'}(\phi') \,\mathrm{d}\,\phi' \int_{\mathbb{R}} e^{-i\omega' T_{\mathrm{E}}}\mathsf{p}_{\omega'}(\omega') \,\mathrm{d}\,\omega' \\
&= +im_0(\mathbf{r})E_2(\mathbf{r}, T_{\mathrm{E}})e^{-\left(R'_2(\mathbf{r})+i\bar{\omega}'(\mathbf{r})\right)T_{\mathrm{E}}} \tan\frac{\alpha(\mathbf{r})}{2}\left(1 - \frac{\eta(\mathbf{r}, T_{\mathrm{R}})}{\xi(\mathbf{r}, T_{\mathrm{R}})}\right); && (2.41)
\end{aligned}
$$

$$
\begin{aligned}
s_{\mathrm{D}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) &\propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) \,\mathrm{d}^3\,\mathbf{r} && (2.42)\\
&\approx \int_{\mathbb{R}}\int_{\mathbb{R}} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right)\mathsf{p}_{\phi'}(\phi')\mathsf{p}_{\omega'}(\omega') \,\mathrm{d}\,\phi' \,\mathrm{d}\,\omega' \\
&\approx e^{+T_{\mathrm{E}}/T_2(\mathbf{r})} \int_{\mathbb{R}} m'_{xy}(\mathbf{r}, t_0)\mathsf{p}_{\phi'}(\phi') \,\mathrm{d}\,\phi' \int_{\mathbb{R}} e^{+i\omega' T_{\mathrm{E}}}\mathsf{p}_{\omega'}(\omega') \,\mathrm{d}\,\omega' \\
&= -im_0(\mathbf{r})E_2^{-1}(\mathbf{r}, T_{\mathrm{E}})e^{-\left(R'_2(\mathbf{r})-i\bar{\omega}'(\mathbf{r})\right)T_{\mathrm{E}}} \tan\frac{\alpha(\mathbf{r})}{2}(1 - \eta(\mathbf{r}, T_{\mathrm{R}})), && (2.43)
\end{aligned}
$$

where (2.41) and (2.43) introduce intermediate variables

$$
\begin{aligned}
\eta(\mathbf{r}, t) &:= \sqrt{\frac{1 - E_2^2(\mathbf{r}, t)}{1 - E_2^2(\mathbf{r}, t)/\xi^2(\mathbf{r}, t)}};\\
\xi(\mathbf{r}, t) &:= \frac{1 - E_1(\mathbf{r}, t)\cos\alpha(\mathbf{r})}{E_1(\mathbf{r}, t) - \cos\alpha(\mathbf{r})}.
\end{aligned}
$$

In steady-state, the DESS signal is typically greatest immediately following excitation and defocuses with rate $\frac{1}{T_2} + R'_2$ until what we hereafter denote the *defocusing* echo time. After a low-signal period between RF pulses, the DESS signal then refocuses with rate $\frac{1}{T_2} - R'_2$ from what we hereafter denote the *refocusing* echo time until just prior the next

[8]In theory, it suffices to design dephasing gradients to impart as little as one cycle of net phase variation across a voxel. In practice, field inhomogeneities will induce spurious through-voxel field gradients that modify the effective dephasing gradient moment and thereby create partial phase cycles that distort the nominally uniform phase distribution. To reduce model mismatch due to so-called partial spoiling effects, dephasing gradients are usually designed to nominally impart multiple complete cycles of across-voxel phase variation. However, larger dephasing gradients cause greater DESS signal loss due to unaccounted diffusive effects.

excitation. Fortuitously, the defocusing (2.41) and refocusing (2.43) DESS signal models have significantly different dependence on relaxation parameters (especially T_2) and thus together are quite useful for relaxation parameter estimation.

2.2 Optimization in QMRI

This section overviews how optimization methods are leveraged in a substantial portion of this thesis to solve practical QMRI problems. For such problems, the central idea is to construct a suitable scalar cost function Ψ of some design variables $\mathbf{x}$, whose output $\Psi(\mathbf{x}) \in \mathbb{R}$ is designed to provide a measure of the undesirability of $\mathbf{x}$. We then employ tailored optimization algorithms to find an $\mathbf{x}$ that minimizes Ψ over a set $\mathbb{X}$, written as

$$\mathbf{x}^* \in \left\{ \arg\min_{\mathbf{x}\in\mathbb{X}} \Psi(\mathbf{x}) \right\}. \tag{2.44}$$

In either optimization-based parameter estimation (Chapter 3) or acquisition design (Chapter 4), we have reason to design Ψ to depend on corresponding design variables $\mathbf{x}$ through MR signal models. Because these models are often (strongly) nonlinear functions of design variables, corresponding cost functions are usually non-convex in $\mathbf{x}$ (though the search space $\mathbb{X}$ is almost always assumed convex in this thesis). Thus, most QMRI problems in the form of (2.44) are non-convex optimization problems.

In general, solving (2.44) is more challenging when Ψ is non-convex in $\mathbf{x}$ than otherwise, due in part to the possible presence of local extrema and/or saddle points. In the following, we discuss two strategies used in this thesis to cope with non-convex optimization. Subsection 2.2.1 relaxes (2.44) to instead seek a local minimizer via iterative methods. Subsection 2.2.2 restricts attention to signal models that are linear in a portion of $\mathbf{x}$ and discusses a specific problem for which (2.44) simplifies for such partially linear structures.

2.2.1 Iterative Local Optimization with Constraints

This subsection overviews a method for finding a local minimizer $\widehat{\mathbf{x}}$ of possibly non-convex cost function Ψ over convex constraint set $\mathbb{X}$. Such $\widehat{\mathbf{x}} \in \mathbb{X}$ must satisfy for some local neighborhood's radius $\delta > 0$

$$\Psi(\widehat{\mathbf{x}}) \leq \Psi(\mathbf{x}) \qquad \forall \mathbf{x} \in \mathbb{X} : \|\widehat{\mathbf{x}} - \mathbf{x}\|_2 < \delta. \tag{2.45}$$

Observe that a global optimizer $\mathbf{x}^*$ satisfies (2.45) for arbitrarily large δ; thus, any global minimizer is a local minimizer (but the converse is not necessarily true unless Ψ is convex).

As even locally optimal minimizers are often challenging to compute analytically, many algorithms construct $\widehat{\mathbf{x}}$ by iteratively updating an initial guess $\mathbf{x}^{(0)}$ until some convergence criterion is satisfied. For a differentiable cost and convex constraints, the gradient projection method [23] is one such iterative algorithm and repeats the following simple update:

$$\mathbf{x}^{(i)} \leftarrow \mathsf{P}_{\mathbb{X}}\left(\mathbf{x}^{(i-1)} - \mathbf{\Pi}\nabla_{\mathbf{x}}\Psi\left(\mathbf{x}^{(i-1)}\right)\right), \tag{2.46}$$

where $\mathsf{P}_{\mathbb{X}}$ denotes projection onto $\mathbb{X}$; $\nabla_{\mathbf{x}} := \left[\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_L}\right]$ denotes row gradient with respect to length-L vector $\mathbf{x}$; and $\mathbf{\Pi}$ is a diagonal preconditioning matrix that permits elements of $\mathbf{x}$ to take scale-informed step sizes along the negative gradient direction.

If Ψ is convex and sufficiently smooth, iterates produced via (2.46) converge to a limit point [24] that is a constrained global minimum (for appropriately selected $\mathbf{\Pi}$). If instead Ψ is non-convex (but $\mathbb{X}$ is still convex), statements regarding convergence[9] to a particular constrained local minimizer require additional (strong) assumptions regarding initialization and in general are still much weaker than in the convex case.

Since non-convex cost functions can have many local extrema (whose associated costs can vary dramatically), the utility of a locally optimal solution depends strongly on initialization quality. Accordingly, this thesis uses iterative local optimization for non-convex QMRI problems where a reasonable initialization is available and global optimization (to within quantization error) via exhaustive grid search is intractable.

2.2.2 Partially Linear Models and the Variable Projection Method

(Constrained, weighted) nonlinear least-squares is a specific non-convex optimization problem that is useful for many parameter estimation problems:

$$\mathbf{x}^* \in \left\{\arg\min_{\mathbf{x}\in\mathbb{X}} \|\mathbf{y} - \mathbf{f}(\mathbf{x})\|_{\mathbf{W}}^2\right\}, \tag{2.47}$$

where $\mathbf{f} : \mathbb{X} \mapsto \mathbb{C}^D$ is a nonlinear forward model that (barring noise) relates parameters $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{C}^L$ to data $\mathbf{y} \in \mathbb{C}^D$; weighted 2-norm $\|\boldsymbol{\iota}\|_{\mathbf{W}} := \sqrt{\boldsymbol{\iota}^{\mathsf{H}}\mathbf{W}\boldsymbol{\iota}}$ for a symmetric, positive-semidefinite weighting matrix $\mathbf{W} \in \mathbb{R}^{D\times D}$ and arbitrary vector $\boldsymbol{\iota} \in \mathbb{C}^D$; and $(\cdot)^{\mathsf{H}}$ denotes conjugate transpose. The variable projection method [25] reduces the complexity of (2.47) when the forward model takes the partially linear structure $\mathbf{f}(\mathbf{x}) \equiv \mathbf{A}(\mathbf{x}_{\mathrm{N}})\mathbf{x}_{\mathrm{L}}$ and

[9]For example, it suffices to assume that $\mathbf{x}^{(0)}$ lies in the *attraction basin* $\mathbb{B}_{\tilde{\mathbf{x}}}$ of a given unconstrained local minimum $\tilde{\mathbf{x}}$, where attraction basin is defined here as the largest convex set containing $\tilde{\mathbf{x}}$ over which Ψ is convex. If $\mathbb{B}_{\tilde{\mathbf{x}}} \cap \mathbb{X}$ is nonempty and step sizes within $\mathbf{\Pi}$ are small enough to contain iterates within $\mathbb{B}_{\tilde{\mathbf{x}}}$, then iterates converge to the limit point $\mathsf{P}_{\mathbb{X}}(\tilde{\mathbf{x}})$.

the feasible set takes the partially unconstrained form $\mathbb{X} \equiv \mathbb{C}^{L_\mathrm{L}} \times \mathbb{X}_\mathrm{N}$, where $\mathbf{x}_\mathrm{L} \in \mathbb{C}^{L_\mathrm{L}}$; $\mathbf{x}_\mathrm{N} \in \mathbb{X}_\mathrm{N}$; and $\mathbf{A} : \mathbb{X}_\mathrm{N} \mapsto \mathbb{C}^{D \times L_\mathrm{L}}$ is a matrix function. These restrictions on (2.47) define a so-called separable least-squares problem:

$$(\mathbf{x}_\mathrm{L}^*, \mathbf{x}_\mathrm{N}^*) \in \left\{ \arg \min_{\substack{\mathbf{x}_\mathrm{L} \in \mathbb{C}^{L_\mathrm{L}} \\ \mathbf{x}_\mathrm{N} \in \mathbb{X}_\mathrm{N}}} \|\mathbf{y} - \mathbf{A}(\mathbf{x}_\mathrm{N})\mathbf{x}_\mathrm{L}\|_\mathbf{W}^2 \right\}. \tag{2.48}$$

The variable projection method simplifies (2.48) by exploiting the partially linear structure of $\mathbf{f}$ to explicitly express the optimal $\mathbf{x}_\mathrm{L}^*$ as a function of any fixed $\mathbf{x}_\mathrm{N} \in \mathbb{X}_\mathrm{N}$:

$$\begin{aligned} \mathbf{x}_\mathrm{L}^*(\mathbf{x}_\mathrm{N}) &= \arg \min_{\mathbf{x}_\mathrm{L} \in \mathbb{C}^{L_\mathrm{L}}} \|\mathbf{y} - \mathbf{A}(\mathbf{x}_\mathrm{N})\mathbf{x}_\mathrm{L}\|_\mathbf{W}^2 \\ &= \left(\mathbf{W}^{1/2}\mathbf{A}(\mathbf{x}_\mathrm{N})\right)^\dagger \mathbf{W}^{1/2}\mathbf{y} \qquad (2.49) \\ &= \left(\mathbf{A}^\mathsf{H}(\mathbf{x}_\mathrm{N})\mathbf{W}\mathbf{A}(\mathbf{x}_\mathrm{N})\right)^{-1} \mathbf{A}^\mathsf{H}(\mathbf{x}_\mathrm{N})\mathbf{W}\mathbf{y}, \qquad (2.50) \end{aligned}$$

where $(\cdot)^\dagger$ denotes pseudoinverse; $\mathbf{W}^{1/2}$ denotes principal (matrix) square root; and (2.50) holds if the matrix inversion within exists. Substituting (2.50) into (2.48) yields a new non-convex optimization problem that contains L_L fewer unknowns than before:

$$\begin{aligned} \mathbf{x}_\mathrm{N}^* \in &\left\{ \arg \min_{\mathbf{x}_\mathrm{N} \in \mathbb{X}_\mathrm{N}} \left\| \mathbf{y} - \mathbf{A}(\mathbf{x}_\mathrm{N}) \left(\mathbf{A}^\mathsf{H}(\mathbf{x}_\mathrm{N})\mathbf{W}\mathbf{A}(\mathbf{x}_\mathrm{N})\right)^{-1} \mathbf{A}^\mathsf{H}(\mathbf{x}_\mathrm{N})\mathbf{W}\mathbf{y} \right\|_\mathbf{W}^2 \right\} \\ \equiv &\left\{ \arg \max_{\mathbf{x}_\mathrm{N} \in \mathbb{X}_\mathrm{N}} \mathbf{y}^\mathsf{H}\mathbf{W}\mathbf{A}(\mathbf{x}_\mathrm{N}) \left(\mathbf{A}^\mathsf{H}(\mathbf{x}_\mathrm{N})\mathbf{W}\mathbf{A}(\mathbf{x}_\mathrm{N})\right)^{-1} \mathbf{A}^\mathsf{H}(\mathbf{x}_\mathrm{N})\mathbf{W}\mathbf{y} \right\}, \qquad (2.51) \end{aligned}$$

where the equivalence leading to (2.51) omits terms independent of $\mathbf{x}_\mathrm{N}$.

In low-dimensional QMRI applications (*e.g.*, those discussed in Chapter 3), reduced problem (2.51) may be tractable via exhaustive grid search, in which case a global optimum $(\mathbf{x}_\mathrm{L}^*(\mathbf{x}_\mathrm{N}^*), \mathbf{x}_\mathrm{N}^*)$ is achievable to within quantization error. However, larger estimation problems involving more nonlinear parameters might only be tractable via iterative optimization (see Subsection 2.2.1) towards a local solution $(\widehat{\mathbf{x}}_\mathrm{L}(\widehat{\mathbf{x}}_\mathrm{N}), \widehat{\mathbf{x}}_\mathrm{N})$.

CHAPTER 3

QMRI Parameter Estimation using Likelihood Models

3.1 Introduction

This brief chapter[1] describes methods for QMRI parameter estimation from statistical likelihood models. The main purpose of this chapter is to serve as a bridge between the background information reviewed in Chapter 2 and more novel ideas introduced in later chapters. As such, we place emphasis here on development of notation and terminology over thorough validation. As instructional examples, we demonstrate likelihood-based parameter estimation on simple problems involving estimation of relaxation parameters T_1 and T_2, applications that Chapter 4 motivates and studies in much greater detail.

The remainder of this chapter is organized as follows. Section 3.2 introduces the notion of a QMRI scan profile, describes a signal model for parameter estimation, and formulates two likelihood-based estimators using this model. Section 3.3 demonstrates these likelihood-based estimators through simulation experiments in two simple applications where conventional estimators are available. Section 3.4 discusses advantages and drawbacks of these two likelihood-based estimators. Section 3.5 provides concluding remarks.

3.2 Likelihood-Based Estimation in QMRI

3.2.1 The QMRI Scan Profile

After image reconstruction, many MRI pulse sequences useful for parameter estimation produce at each voxel centered at position $\mathbf{r}$ a set of noisy voxel values $\{y_1(\mathbf{r}), \dots, y_D(\mathbf{r})\}$,

[1]This chapter partially derives content from conference papers [4, 5].

each of which can be described with the following general model:

$$y_d(\mathbf{r}) = s_d(\mathbf{x}(\mathbf{r}); \boldsymbol{\nu}(\mathbf{r}), \mathbf{p}_d) + \epsilon_d(\mathbf{r}), \tag{3.1}$$

where $d \in \{1, \dots, D\}$. Here, $\mathbf{x}(\mathbf{r}) \in \mathbb{C}^L$ collects L *latent* object parameters at $\mathbf{r}$; $\boldsymbol{\nu}(\mathbf{r}) \in \mathbb{C}^K$ collects K *known* object parameters at $\mathbf{r}$; $s_d : \mathbb{C}^L \times \mathbb{C}^K \times \mathbb{R}^A \mapsto \mathbb{C}$ is a (pulse-sequence dependent) function that models the noiseless signal obtained from the dth dataset using *acquisition* parameter $\mathbf{p}_d \in \mathbb{R}^A$; and $\epsilon_d \sim \mathbb{C}\mathcal{N}(0, \sigma_d^2)$ is assumed for simplicity[2] to be (circularly-symmetric) complex Gaussian noise [28, 29] with zero mean and variance σ_d^2. Semicolon positions in signal model (3.1) and similar expressions throughout this thesis distinguish unknown and known parameters. Concrete examples follow shortly.

For accurate, well-conditioned QMRI parameter estimation, it is typically necessary to acquire a collection of datasets, which we refer to hereafter as a *scan profile*. A scan profile consists of D datasets from up to D pulse sequences (some sequences yield more than one dataset, *e.g.*, DESS). Let $\mathbf{y}(\mathbf{r}) := [y_1(\mathbf{r}), \dots, y_D(\mathbf{r})]^\mathsf{T} \in \mathbb{C}^D$ collect noisy voxel values centered at $\mathbf{r}$ from a given scan profile. Then the vector signal model

$$\mathbf{y}(\mathbf{r}) = \mathbf{s}(\mathbf{x}(\mathbf{r}); \boldsymbol{\nu}(\mathbf{r}), \mathbf{P}) + \boldsymbol{\epsilon}(\mathbf{r}) \tag{3.2}$$

helps define the noiseless signal $\mathbf{s} := [s_1, \dots, s_D]^\mathsf{T} : \mathbb{C}^L \times \mathbb{C}^K \times \mathbb{R}^{A \times D} \mapsto \mathbb{C}^D$ and acquisition parameters $\mathbf{P} := [\mathbf{p}_1, \dots, \mathbf{p}_D] \in \mathbb{R}^{A \times D}$ associated with that scan profile. Here, noise $\boldsymbol{\epsilon}(\mathbf{r}) := [\epsilon_1(\mathbf{r}), \dots, \epsilon_D(\mathbf{r})]^\mathsf{T} \in \mathbb{C}^D$ typically has diagonal covariance structure $\boldsymbol{\Sigma} := \operatorname{diag}\left([\sigma_1, \dots, \sigma_D]^\mathsf{T}\right)$ due to independence across datasets, where $\operatorname{diag}(\cdot)$ assigns its argument to the diagonal entries of an otherwise zero (square) matrix.

The following subsections describe two concrete scan profiles whose signals can be modeled via (3.2) and that we study through experiments later in this chapter.

3.2.1.1 Example: An SPGR Scan Profile for T_1 estimation

We first consider the problem of $T_1(\mathbf{r})$ estimation at $\mathbf{r}$ from as few SPGR scans as possible, given a prior estimate of transmit field variation $\kappa^{\mathrm{t}}(\mathbf{r})$ (see (2.7)). Examining SPGR model (2.27) makes clear that by fixing echo time T_E across scans, SPGR signal dependence is

[2] Though the noise distribution of $\mathbf{k}$-space raw data is usually well-modeled as complex white Gaussian, the noise distribution of the dth reconstructed image y_d depends both on the acquisition and reconstruction. If single receive channel $\mathbf{k}$-space data is fully-sampled on a Cartesian grid, each dataset y_d is recoverable via separate Fourier transform, and is thus complex Gaussian and independent across datasets. However if $\mathbf{k}$-space data is multi-channel, undersampled, and/or non-Cartesian, it may be preferable that y_d be estimated by more sophisticated techniques, *e.g.* [26, 27]. In such cases, reconstructed image noise is unlikely to be Gaussian-distributed.

reduced to just two spatially varying latent parameters: desired parameter $T_1(\mathbf{r}) \in \mathbb{R}$ and nuisance parameter $c_1(\mathbf{r}) := im_0(\mathbf{r})e^{-T_E/T_2^*(\mathbf{r})}e^{-i\bar{\omega}'(\mathbf{r})T_E} \in \mathbb{C}$. We assign $\mathbf{x} \leftarrow [T_1, c_1]^\mathsf{T}$ and $\boldsymbol{\nu} \leftarrow \kappa^t$ for $L \leftarrow 2$ latent and $K \leftarrow 1$ known parameters, respectively.

With T_E fixed, prescribed flip angles α_0 and repetition times T_R are the only remaining $A \leftarrow 2$ acquisition parameters available to choose that appear explicitly in (2.27). Thus, an SPGR scan profile useful for T_1 estimation must vary $\mathbf{p}_d \leftarrow [\alpha_0, T_R]^\mathsf{T} \forall d \in \{1, \dots, D\}$ over S_{SPGR} scan repetitions to produce $D \geq L \leftarrow 2$ datasets for well-conditioned estimation.

3.2.1.2 Example: A DESS Scan Profile for T_2 estimation

We next consider the problem of $T_2(\mathbf{r})$ estimation at $\mathbf{r}$ from as few DESS scans as possible. Examining DESS models (2.41) and (2.43) makes clear that even with fixed T_E over possibly several acquisitions, there is signal dependence on five distinct object parameters: $\kappa^t(\mathbf{r}) \in \mathbb{R}$, $T_1(\mathbf{r}) \in \mathbb{R}$, $\bar{\omega}'(\mathbf{r}) \in \mathbb{R}$, $c_2(\mathbf{r}) := m_0(\mathbf{r})e^{-T_E/T_2^*(\mathbf{r})} \in \mathbb{C}$, and $T_2(\mathbf{r}) \in \mathbb{R}$. In this chapter, we take $\kappa^t(\mathbf{r}) \in \mathbb{R}$ and $T_1(\mathbf{r}) \in \mathbb{R}$ as known for simplicity. To avoid (separate or joint) $\bar{\omega}'(\mathbf{r})$ estimation, we choose to use magnitude DESS data, at the expense of slight model mismatch[3] due to Rician noise. These choices assign $\boldsymbol{\nu} \leftarrow [\kappa^t, T_1]^\mathsf{T}$ as $K \leftarrow 2$ known parameters and leave $L \leftarrow 2$ latent parameters $\mathbf{x} \leftarrow [T_2, c_2]^\mathsf{T}$ to be estimated.

With T_E again fixed, $\mathbf{p}_d \leftarrow [\alpha_0, T_R]^\mathsf{T} \forall d \in \{1, \dots, D\}$ collects the remaining $A \leftarrow 2$ tunable scan parameters that appear explicitly in (2.41) and (2.43). As in Example 3.2.1.1, $D \geq L \leftarrow 2$ datasets are necessary for well-conditioned estimation. Unlike before however, a minimum $D \leftarrow 2$ datasets need not require scan repetition, since S_{DESS} DESS scan repetitions produce $D \leftarrow 2S_{\text{DESS}}$ datasets.

3.2.2 Latent Object Parameter Estimation

3.2.2.1 Signal Model and Problem Statement

A scan profile's reconstructed images can be modeled to discretize the bulk MR signal into V localized voxels centered at positions $\mathbf{r}_1, \dots, \mathbf{r}_V$:

$$\mathbf{Y} = \mathbf{S}(\mathbf{X}; \mathbf{N}, \mathbf{P}) + \mathbf{E}. \tag{3.3}$$

[3] The assumption of complex Gaussian noise in noisy MRI images implies that corresponding magnitude MRI images are Rician-distributed. However, the statistical estimators we will develop in Subsection 3.2.2 are based on Gaussian data. Fortunately, this source of model mismatch is negligible (less than 1%) for signal-to-noise ratio (SNR) in excess of 10 [30], and the acquisitions we examine here are capable of producing SNR in tissue of at minimum 100 and usually more.

Here, signal model $\mathbf{S} : \mathbb{C}^{L \times V} \times \mathbb{C}^{K \times V} \times \mathbb{R}^{A \times D} \mapsto \mathbb{C}^{D \times V}$ is a matrix function that maps latent $\mathbf{X} := [\mathbf{x}(\mathbf{r}_1), \ldots, \mathbf{x}(\mathbf{r}_V)] \in \mathbb{C}^{L \times V}$ and known $\mathbf{N} := [\boldsymbol{\nu}(\mathbf{r}_1), \ldots, \boldsymbol{\nu}(\mathbf{r}_V)] \in \mathbb{C}^{K \times V}$ parameter images (with fixed acquisition parameter $\mathbf{P}$) to reconstructed image data $\mathbf{Y} := [\mathbf{y}(\mathbf{r}_1), \ldots, \mathbf{y}(\mathbf{r}_V)] \in \mathbb{C}^{D \times V}$, save for noise image $\mathbf{E} := [\boldsymbol{\epsilon}(\mathbf{r}_1), \ldots, \boldsymbol{\epsilon}(\mathbf{r}_V)] \in \mathbb{C}^{D \times V}$. The goal in QMRI parameter estimation is to estimate latent parameter images $\mathbf{X}$ from MR image data $\mathbf{Y}$, for a fixed scan profile defined by $\mathbf{S}$ and $\mathbf{P}$ and given (separately acquired, estimated, and here assumed) known parameter images $\mathbf{N}$.

3.2.2.2 Maximum Likelihood Methods

In maximum likelihood (ML) estimation, one seeks model parameters that maximize the likelihood of observing output data. We apply ML estimation to QMRI by first constructing a *likelihood function* that describes the probability of observing image data $\mathbf{Y}$ given latent parameters $\mathbf{X}$. We then formulate ML latent parameter estimator $\widehat{\mathbf{X}}_{\mathrm{ML}}$ by finding an estimate $\widehat{\mathbf{X}}_{\mathrm{ML}}(\mathbf{Y}; \mathbf{N}, \mathbf{P})$ of $\mathbf{X}$ that maximizes this likelihood function.

We first construct the likelihood function for the vth voxel's data $\mathbf{y}(\mathbf{r}_v)$ and latent parameter $\mathbf{x}(\mathbf{r}_v)$. For complex Gaussian noise, the likelihood function is

$$\mathsf{L}(\mathbf{x}(\mathbf{r}_v)) \propto \exp\left(-\|\mathbf{y}(\mathbf{r}_v) - \mathbf{s}(\mathbf{x}(\mathbf{r}_v); \boldsymbol{\nu}(\mathbf{r}_v), \mathbf{P})\|_{\boldsymbol{\Sigma}^{-1}}^2\right), \tag{3.4}$$

where (3.4) omits constants that are independent of $\mathbf{x}(\mathbf{r}_v)$ and are therefore irrelevant. Assuming noise independence across image voxels, we can next build a simple and practical likelihood function of the full image data as

$$\mathsf{L}(\mathbf{X}) = \prod_{v=1}^{V} \mathsf{L}(\mathbf{x}(\mathbf{r}_v)). \tag{3.5}$$

We form an ML parameter estimate by finding $\mathbf{X}$ that maximizes this likelihood function:

$$\begin{aligned}
\widehat{\mathbf{X}}_{\mathrm{ML}}(\mathbf{Y}; \mathbf{N}, \mathbf{P}) &\in \left\{\arg\max_{\mathbf{X} \in \mathbb{X}^V} \mathsf{L}(\mathbf{X})\right\} \\
&\equiv \left\{\arg\min_{\mathbf{X} \in \mathbb{X}^V} -\log \mathsf{L}(\mathbf{X})\right\} &\quad (3.6) \\
&= \left\{\arg\min_{\mathbf{X} \in \mathbb{X}^V} \sum_{v=1}^{V} \|\mathbf{y}(\mathbf{r}_v) - \mathbf{s}(\mathbf{x}(\mathbf{r}_v); \boldsymbol{\nu}(\mathbf{r}_v), \mathbf{P})\|_{\boldsymbol{\Sigma}^{-1}}^2\right\} \\
&= \left\{\arg\min_{\mathbf{X} \in \mathbb{X}^V} \left\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \mathbf{S}(\mathbf{X}; \mathbf{N}, \mathbf{P}))\right\|_{\mathrm{F}}^2\right\}, &\quad (3.7)
\end{aligned}$$

where $\mathbb{X}$ is a (typically convex) latent parameter search space; the set equivalence in (3.6) uses the monotonicity of the $\log$ function; and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius matrix norm.

Typically, QMR image model $\mathbf{S}$ is nonlinear in $\mathbf{X}$ and so ML estimation problem (3.7) involves non-convex optimization, which is challenging in general (see Section 2.2). Two properties of (3.7) guide our solution strategies. First, (3.7) is separable across voxels, so problem non-convexity is addressable on a voxel-by-voxel basis. Second, MR signal models are usually partially linear, in which case we may employ the variable projection method (described in Section 2.2.2) to further reduce problem complexity. For applications studied in this chapter, these properties allow for (3.7) to be solved via simple grid search.

3.2.2.3 Regularized Likelihood Methods

In regularized likelihood (RL) estimation, we modify ML estimation problem (3.6) to include additional information in the form of *regularization*:

$$\widehat{\mathbf{X}}_{\mathrm{RL}}(\mathbf{Y};\mathbf{N},\mathbf{P}) \in \left\{\arg\min_{\mathbf{X}\in\mathbb{X}^V} -\log\mathsf{L}(\mathbf{X}) + \mathsf{R}(\mathbf{X})\right\}. \tag{3.8}$$

Here, we have freedom to design regularizer $\mathsf{R} : \mathbb{C}^{L\times V} \mapsto \mathbb{R}$ to encourage desirable structure in estimates of $\mathbf{X}$. We observe that it is usually reasonable to assume that each latent object parameter map is *piecewise smooth* as a function of space: that is, each parameter is likely to vary smoothly in space, except for sharp discontinuities at tissue boundaries. To encourage piecewise-smoothness in parameter estimates, we use the regularizer

$$\mathsf{R}(\mathbf{X}) := \sum_{l=1}^{L} \beta_l \sum_{j=1}^{J} \phi_l\Big(\big[\mathbf{J}\mathbf{X}^{\mathsf{T}}\big]_{jl}\Big), \text{ where} \tag{3.9}$$

$$\phi_l(\cdot) := \gamma_l^2\left(\sqrt{1+|\cdot/\gamma_l|^2} - 1\right) \tag{3.10}$$

is a differentiable approximation of the absolute value function; $\mathbf{J} \in \mathbb{R}^{J\times V}$ evaluates J (multi-dimensional) finite-differencing operations; $[\cdot]_{jl}$ extracts the (j,l)th matrix element; and β_l is a regularization parameter that controls the relative importance of smoothing the lth latent object parameter image. Conceptually, this regularizer penalizes inconsistencies in adjacent latent parameter image voxels, but with a severity that depends on the degree of inconsistency. "Small" voxel-to-voxel differences are likely due to image data noise within a single tissue type and are penalized near-quadratically, while "large" differences are likely due to tissue boundaries and are penalized near-linearly. Useful notions of small versus large differences are governed by shape parameters $\gamma_l \,\forall l \in \{1,\ldots,L\}$, and vary for different latent parameter maps based on their units and relative scale.

In general, QMRI image signal model $\mathbf{S}$ is nonlinear in $\mathbf{X}$ and so RL estimation problem (3.8) requires non-convex optimization. Unlike in ML estimation, (3.8) is not separable across voxels due to regularization, precluding global optimization (via grid search or other methods). We instead take the corresponding ML estimate as initialization and solve (3.8) via iterative constrained local optimization (detailed in Section 2.2.1).

3.3 Experimentation

This section demonstrates likelihood-based estimation through two experiments in simulation that correspond to the simple problems defined in Section 3.2.1. Subsection 3.3.1 continues Example 3.2.1.1 in studying T_1 estimation from two SPGR scans. Subsection 3.3.2 continues Example 3.2.1.2 in studying T_2 estimation from one DESS scan.

3.3.1 T_1 estimation from two SPGR scans

We selected T_1 and T_2 WM and GM values based on previously reported measurements at 3T [31, 32] and extrapolated other nuisance latent object parameters m_0 and T_2^* from measurements at 1.5T [33]. For simplicity, we assumed no flip angle variation $\kappa^{\mathrm{t}} \leftarrow 1$ and no phase accrual due to off-resonance effects $\bar{\omega}' \leftarrow 0$. We assigned these parameter values to the 81st slice of the BrainWeb digital phantom [34, 33] to create ground truth $\mathbf{M}_0, \mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_2^* \in \mathbb{R}^V$ maps. We simulated 217×181 noiseless single-coil SPGR image datasets, varying nominal flip angles $\alpha_0 \leftarrow 5, 30°$ and fixing repetition times $T_{\mathrm{R}} \leftarrow 12.2$ms and echo times $T_{\mathrm{E}} \leftarrow 4.67$ms across $S_{\mathrm{SPGR}} \leftarrow 2$ scans. We corrupted noiseless datasets with additive complex Gaussian noise to yield $D \leftarrow 2$ noisy complex datasets with signal-to-noise ratio (SNR) ranging from 57-93, where SNR is defined here as

$$\mathsf{SNR}(\mathbf{S}, \mathbf{Y}) := \frac{\|\mathbf{S}\|_{\mathrm{F}}}{\|\mathbf{Y} - \mathbf{S}\|_{\mathrm{F}}}. \tag{3.11}$$

We estimated latent parameter maps $\mathbf{T}_1, \mathbf{C}_1$ using a conventional method-of-moments (MOM) estimator [35], the ML estimator (3.7), and the RL estimator (3.8). The MOM, ML, and RL estimators respectively took 0.11s, 0.75s, and 31s. The MOM estimator applies linear regression voxel-by-voxel to an appropriately transformed version of the noiseless magnitude SPGR signal model that is linear in T_1, c_1; see *e.g.* [35, 36] for details. We next describe our implementations of ML and RL estimation in turn.

The ML estimator applies the variable projection method (VPM; described in Subsection 2.2.2) to separate nonlinear $\mathbf{T}_1$ estimation from linear $\mathbf{C}_1$ estimation. Specifically, the algorithm first estimates $\mathbf{T}_1$ voxel-by-voxel via an exhaustive grid search (over 1000 T_1 values logarithmically spaced between $(10^{1.5}, 10^{3.5})$ms) for a maximizer of the separated least squares cost (2.51). It then estimates $\mathbf{C}_1$ via per-voxel linear regression.

The RL estimator applies a preconditioned variant of the classical gradient projection method (GPM; described in Subsection 2.2.1) to iteratively descend towards a local optimizer of the RL cost described in (3.8). We designed the preconditioner as the inverse of a positive definite diagonal majorizer of the RL cost function's Hessian matrix, updated for the first five iterations and fixed thereafter. We employed a diagonal preconditioner to retain the linear convergence guarantees of GPM [37] yet approach the practical performance of other unprojected second-order methods (*e.g.*, Newton's method). We employed a simple step-halving line search at each iteration to ensure monotone local convergence in cost. We initialized GPM with the ML estimates. We selected regularization parameters as described in Subsection 4.4.2.1. We used the Michigan image reconstruction toolbox [38] to construct the regularizer and rapidly evaluate its gradient and Hessian. We used the MATLAB® symbolic toolbox to generate analytical expressions for the gradient and Hessian of the SPGR signal model. At each iteration, we used these gradient and Hessian expressions to compute a preconditioned descent direction, updated the iterate (possibly after backtracking to ensure descent), and projected each voxel's T_1 iterate to within $[10, 3000]$ms. We continued iterations until the convergence criterion

$$\left\|\mathbf{\Omega}^{-1}\big(\mathbf{X}^{(i)} - \mathbf{X}^{(i-1)}\big)\right\|_{\mathrm{F}} < 10^{-7}\left\|\mathbf{\Omega}^{-1}\big(\mathbf{X}^{(i)}\big)\right\|_{\mathrm{F}} \tag{3.12}$$

was satisfied, where $(\cdot)^{(i)}$ denotes the ith iterate, $\mathbf{\Omega} := \mathrm{diag}\big(\mathrm{med}\big(\mathbf{X}^{(0)}\big)\big)$ is a weighting matrix, and $\mathrm{med}(\cdot)$ takes the median across the columns of its argument.

Fig. 3.1 compares MOM, ML, and RL $\mathbf{T}_1$ estimates alongside $10\times$ magnified absolute difference images with respect to the ground truth. Overall, all three estimators produce reasonable $\mathbf{T}_1$ maps. The MOM and ML $\mathbf{T}_1$ estimates are visually similar. The RL $\mathbf{T}_1$ estimates are smoother than the MOM and ML $\mathbf{T}_1$ estimates away from tissue interfaces, but the RL $\mathbf{T}_1$ estimate incurs systematically higher errors near tissue boundaries and provides reduced spatial resolution.

Table 3.1 presents T_1, c_1 samples statistics within WM-like and GM-like ROIs selected to contain voxels that are well away from tissue interfaces. In both WM and GM, MOM and ML $\mathbf{T}_1, \mathbf{C}_1$ estimates are comparable. RL estimates consistently exhibit the lowest variation, but the RL $\mathbf{T}_1$ estimate exhibits the greatest bias in WM. All RL bias values would be significantly greater if ROIs instead contained voxels at tissue interfaces.

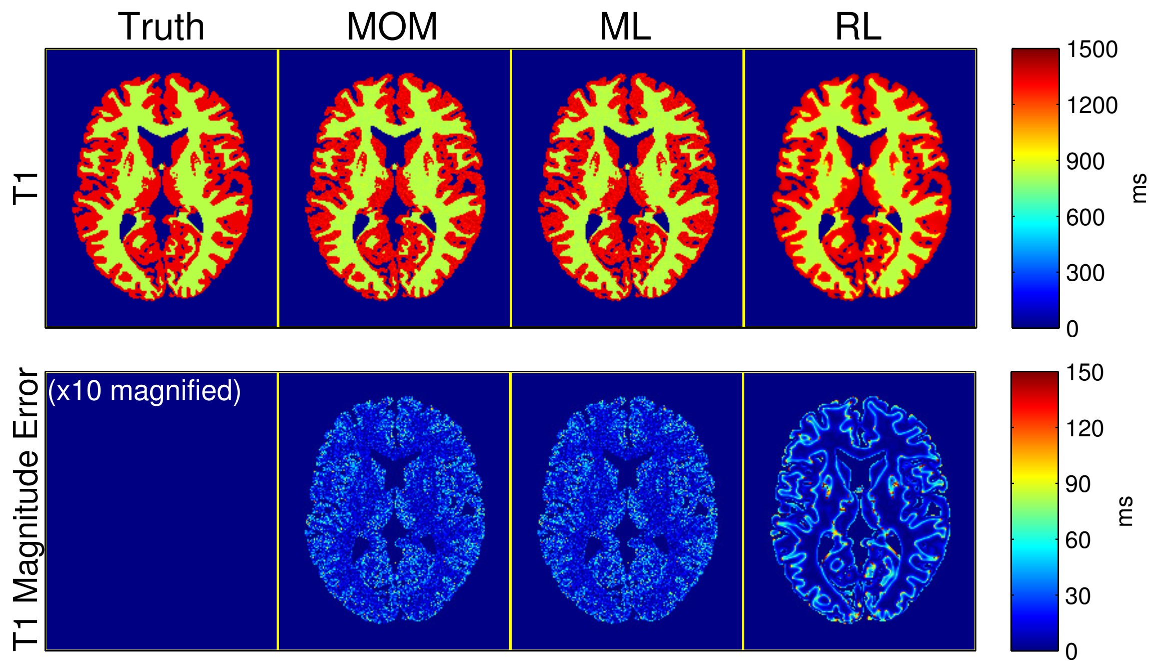


Figure 3.1: $\mathbf{T}_1$ MOM, ML, and RL estimates and corresponding error images, from two simulated SPGR scans. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Table 3.1 presents corresponding sample statistics.

	Truth	MOM	ML	RL
WM T_1	832	832.7 ± 15.6	832.7 ± 15.6	834.00 ± 2.77
GM T_1	1331	1332 ± 34.9	1332 ± 34.9	1332.2 ± 6.3
WM c_1	0.77	0.7266 ± 0.00744	0.7314 ± 0.00749	0.73184 ± 0.00475
GM c_1	0.86	0.8245 ± 0.0108	0.8301 ± 0.0109	0.8287 ± 0.0059

Table 3.1: Sample means $\pm$ sample standard deviations of MOM, ML, and RL T_1, c_1 estimates from two simulated SPGR datasets, computed over 3001 WM-like and 1151 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. T_1 values are in milliseconds. c_1 values are unitless. Fig. 3.1 presents corresponding images.

3.3.2 T_2 estimation from one DESS scan

Using the same ground truth parameters maps as in Subsection 3.3.1, we simulated noiseless single-coil DESS image datasets arising from $S_{\mathrm{DESS}} \leftarrow 1$ DESS scan with nominal flip angle $\alpha_0 \leftarrow 45°$, repetition time $T_{\mathrm{R}} \leftarrow 17.5$ms, and symmetric echo times $T_{\mathrm{E}} \leftarrow 4.67$ms. We corrupted noiseless datasets with additive complex Gaussian noise to yield $D \leftarrow 2$ noisy complex datasets with SNR ranging from 97-134, where SNR is defined as in (3.11).

We estimated latent parameter maps $\mathbf{T}_2, \mathbf{C}_2$ using a conventional MOM estimator [22], the ML estimator (3.7), and the RL estimator (3.8). The MOM, ML, and RL estimators respectively took 0.09s, 0.76s, and 26s; we describe their implementations next in turn. For symmetric echo times, the voxel-by-voxel MOM estimator assigns

$$\widehat{T}_2(y_1, y_2) \leftarrow -\frac{2(T_\mathrm{R} - T_\mathrm{E})}{\log |y_2/y_1|}, \tag{3.13}$$

where y_1 and y_2 are noisy measurements of the defocused (2.41) and refocused (2.43) DESS signals, respectively. With $\widehat{\mathbf{T}}_2$ fixed, the MOM method then estimates $\mathbf{C}_2$ via per-voxel linear regression. This MOM estimator incurs strong bias for flip angles that provide practical SNR levels, mainly because it neglects T_1 effects. This MOM estimator also amplifies noise due to the division operation.

Similar to Subsection 3.3.1, the ML estimator applies VPM to separate nonlinear $\mathbf{T}_2$ estimation from linear $\mathbf{C}_2$ estimation. The algorithm first estimates $\mathbf{T}_2$ voxel-by-voxel via an exhaustive grid search over 1000 T_2 values logarithmically spaced between $(10^{0.5}, 10^3)$ms. It then estimates $\mathbf{C}_2$ via per-voxel linear regression.

The RL estimator applies preconditioned GPM to iteratively descend towards a local optimizer of the RL cost described in (3.8). GPM implementation details remain largely unchanged from those described in Subsection 3.3.1. At each iteration, we computed a preconditioned descent direction (using MATLAB®-generated analytical expressions for the gradient and Hessian of the DESS signal models), updated the iterate (possibly after backtracking to ensure monotone descent), and projected each voxel's T_2 iterate to within $[1, 700]$ms. We continued iterations until convergence criterion (3.12) was satisfied.

Fig. 3.2 compares MOM, ML, and RL $\mathbf{T}_2$ estimates alongside $10\times$ magnified absolute difference images with respect to the ground truth. Overall, the ML and RL estimators produce more reasonable $\mathbf{T}_2$ maps than does the MOM estimator (but utilize additional κ^t, T_1 information). The RL $\mathbf{T}_2$ estimates are smoother than ML $\mathbf{T}_2$ estimates away from tissue interfaces, but the RL $\mathbf{T}_2$ estimate incurs systematically higher errors near tissue boundaries and provides reduced spatial resolution.

Table 3.2 presents T_2, c_2 sample statistics within the same well-isolated ROIs as was used in Table 3.1. MOM estimates are consistently most biased. The MOM and ML $\mathbf{T}_2$ estimates exhibit similar levels of variability. RL estimates consistently exhibit more bias and less variation than ML estimates.

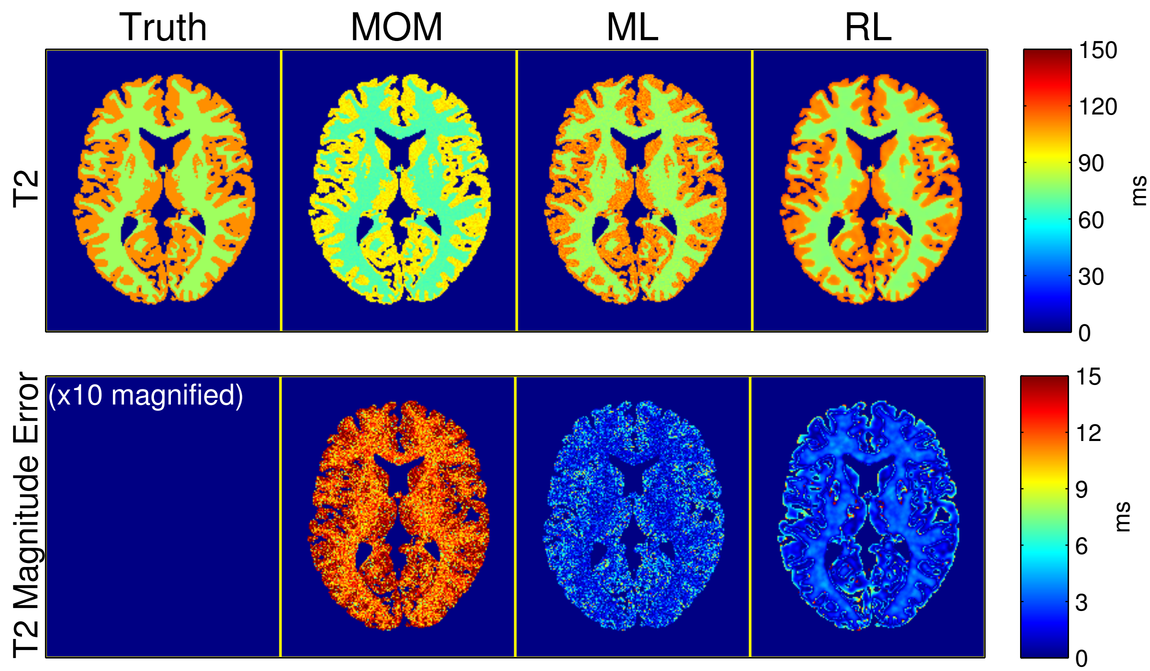


Figure 3.2: $\mathbf{T}_2$ MOM, ML, and RL estimates and corresponding error images, from one simulated DESS scan. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Table 3.2 presents corresponding sample statistics.

	Truth	MOM	ML	RL
WM T_2	79.6	68.13 ± 1.64	79.36 ± 2.18	76.402 ± 0.411
GM T_2	110.	95.86 ± 3.21	110.2 ± 4.19	111.57 ± 0.88
WM c_2	0.77	0.8578 ± 0.0148	0.7852 ± 0.0149	0.79290 ± 0.00457
GM c_2	0.86	0.9523 ± 0.0241	0.8545 ± 0.0240	0.8510 ± 0.0063

Table 3.2: Sample means $\pm$ sample standard deviations of MOM, ML, and RL T_2, c_2 estimates from one simulated DESS dataset, computed over 3001 WM-like and 1151 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. T_2 values are in milliseconds. c_2 values are unitless. Fig. 3.2 presents corresponding images.

3.4 Discussion

The simulated experiments in this chapter serve to illustrate that MRI parameter estimation from likelihood models can often offer greater accuracy than conventional MOM estimation, though usually at the expense of greater computation. Simulations corresponding to Ex. 3.2.1.1 demonstrated small but consistent ML and RL accuracy gains over MOM c_1 estimation, primarily because likelihood-based estimators here considered complex image noise statistics. Simulations corresponding to Ex. 3.2.1.2 demonstrated larger ML and RL accuracy gains over MOM T_2, c_2 estimation, primarily because likelihood-based estimators here required fewer bias-inducing signal model approximations. In general, such accuracy gains may be more substantial in more complicated QMRI estimation problems that for MOM estimation will require stronger model approximations.

Because likelihood-based estimators do not rely on (possibly intractable) algebraic manipulations of the application-specific signal model, they are also more general-purpose tools than are MOM estimators. Indeed, algorithms for implementing the ML (3.7) (*i.e.*, VPM with grid search) or RL (3.8) estimators (*i.e.*, GPM) are typically available even when the associated inverse problem is poorly conditioned. In such cases, ML estimates still need not necessarily be imprecise in all latent parameter entries, a behavior that Ch. 4 characterizes and then exploits. Because of their relative flexibility, we utilize likelihood-based estimators over MOM estimators for the more complex QMRI problems studied in Ch. 4.

Simulations herein demonstrate that ML versus RL estimation performance can be characterized by a bias-variance tradeoff: RL estimation reduces variation in regions well away from tissue interfaces, but increases bias near interfaces. Thus, a decision of whether to include regularization (and with what strength) should consider the degree to which regions of interest contain interfaces. Since the applications studied in later chapters take interest in resolving subtle WM/GM boundaries with high spatial resolution and associated experiments use fully-sampled $\mathbf{k}$-space data, we hereafter employ ML estimation by default (though Ch. 4 also provides comparisons with RL estimation). In other QMRI problems that utilize low spatial frequency or highly under-sampled data or involve poorly-conditioned parameter estimation, including regularization may instead be preferable.

Both simulated experiments used acquisition parameters similar to those used in earlier studies [36, 39]. While these studies provide intuitive reasoning for some acquisition parameter choices, it is unclear whether these choices are in any sense optimal for the respective tasks of T_1 or T_2 estimation. Motivated by this question, Ch. 4 defines one notion of acquisition parameter optimality and investigates how optimized acquisition parameters can improve T_1, T_2 ML estimation performance.

3.5 Conclusion

This transitional chapter has developed a formalism to describe a general QMRI scan profile and has described two likelihood-based estimators for QMRI parameter estimation. We have demonstrated these ML and RL estimators in two simple applications where conventional MOM estimators are available, namely T_1 estimation from two SPGR scans and T_2

estimation from one DESS scan. Simulations illustrate that ML and RL estimators can often offer greater accuracy than MOM estimators, though usually at the expense of greater computation. Because of their accuracy and generality, likelihood-based QMRI estimators will be used to validate a new method for scan design in Ch. 4 and for comparison with a new QMRI parameter estimation method in Ch. 5.

CHAPTER 4

QMRI Acquisition Design via Min-Max Optimization

4.1 Introduction

Fast, accurate *relaxometry*, or quantification of spin-lattice and spin-spin relaxation parameters T_1 and T_2 has been of longstanding interest in MRI. Many researchers have suggested that T_1, T_2 "maps" (*i.e.*, estimated parameter images) may serve as biomarkers for monitoring the progression of various disorders [40]. Neurological applications include: lesion classification in multiple sclerosis [41]; tumor characterization [42, 43]; and symptom onset prediction in stroke [44, 45]. In addition, T_1, T_2 have shown promise for detecting hip and knee cartilage degeneration [46, 47] and for assessing cardiac dysfunction due to iron overload [48] or edema [49]. Motivated by this broad interest in T_1, T_2 mapping, this chapter[1] describes a systematic method to guide QMRI scan design.

Classical pulse sequences such as inversion/saturation recovery (IR/SR) or (single) spin echo (SE) yield relatively simple methods for T_1 or T_2 estimation, respectively; however, these methods require several scans, each with long repetition time T_R, leading to undesirably long acquisitions. Numerous modifications such as the Look-Locker method [50], multi-SE trains [51], or fast $\mathbf{k}$-space trajectories [52, 53, 54] have been proposed to accelerate T_1 [55, 56, 57, 58] and T_2 [59, 60, 61, 62] relaxometry with these classical sequences. These techniques are more sensitive to model non-idealities [63, 64, 65], and are still speed-limited by the long T_R required for (near)-complete T_1 recovery.

Steady-state (SS) pulse sequences [16, 17] permit short T_R, and are thus inherently much faster than classical counterparts. SS techniques are well-suited for relaxometry because the signals produced are highly sensitive to T_1 and T_2 variation. However, short T_R times also cause SS signals to be complex functions of both desired and undesired (*nui-*

[1]This chapter mainly derives content from published journal paper [6] that extends conference paper [7].

sance) parameters, complicating quantification. Furthermore, some such methods [36, 66] still require scan repetition, though individual scans are now considerably shorter. Despite these difficulties, the potential for rapid scanning with high T_1, T_2 sensitivity has motivated numerous SS relaxometry studies [67, 36, 66, 68, 69, 70, 71, 39, 72, 73].

The dual-echo steady-state (DESS) sequence [22] was recently proposed as a promising SS imaging technique for T_2 estimation [71]. Because it produces two distinct signals per excitation, the DESS sequence can reduce scan repetition requirements by recording close to twice as much data per scan. As with other SS methods, the resulting signals [74, 75] are complicated functions of T_1, T_2, and other parameters (see Section 2.1.2.2 for derivations). Prior works have isolated T_2 dependencies using either algebraic manipulations of the first- and second-echo signals [71, 39] or separate scans to first estimate nuisance parameters [5]. Although DESS concurrently encodes rich T_1 and T_2 information, these methods have shied away from using DESS for T_1 estimation, either through bias-inducing approximations, or noise-propagating sequential estimation, respectively.

Whether it be with DESS, other sequences, or even combinations thereof, it is generally unclear how to best assemble a *scan profile* (*i.e.*, a collection of scans) for a fixed amount of scan time. Furthermore, for a given scan profile, it is typically not obvious how to best select acquisition parameters (*e.g.*, flip angles, repetition times, etc.) for relaxometry. In this and subsequent chapters, the term *scan design* refers to the related problems of scan profile selection and acquisition parameter optimization.

Historically, scan design for relaxometry has mainly been explored using figures of merit related to estimator precision. In particular, several studies have used the Cramér-Rao Bound (CRB), a statistical tool that bounds the minimum variance of an unbiased estimator. Earlier works have used the CRB and variations to select inversion times for recovery experiments [76, 77], flip angles for spoiled gradient-recalled echo (SPGR) sequences [78], and echo times for SE experiments [79]. More recent studies have considered additional scan design challenges, including scan time constraints [80], multiple latent parameters [81], multiple scan parameter types [82], and latent parameter spatial variation [83, 84, 85].

The aforementioned studies consider scan parameter optimization for profiles consisting of *only one* pulse sequence. In contrast, this chapter introduces a general framework for robust, application-specific scan design for parameter estimation from *combinations* of pulse sequences. The framework first finds multiple sets of scan parameters that achieve precise estimation within a tight, *application-specific* range of object parameters (*e.g.*, T_1, T_2, etc.). The framework then chooses the one scan parameter set most *robust* to estimator precision degradation over a broader range of object parameters. As a detailed example, we optimize three combinations of SPGR and DESS sequences for T_1, T_2 mapping.

For a fixed total scan time, we find that well-chosen DESS scans alone can be used to estimate both T_1 and T_2 with precision and robustness comparable to combinations of SPGR and DESS. This example illustrates that, with careful scan profile design, well-established pulse sequences can find use in new estimation problems.

This chapter is organized as follows. Section 4.2 describes a CRB-inspired min-max optimization problem for robust, application-specific scan design. Section 4.3 optimizes three practical DESS/SPGR combinations to show that, even in the presence of radiofrequency (RF) field inhomogeneity, DESS is a promising option for T_1, T_2 relaxometry. Section 4.4 describes simulation, phantom, and *in vivo* experiments and discusses corresponding results. Section 4.5 investigates the sensitivity of single-compartment T_2 estimates to model mismatch due to multi-compartment relaxation. Section 4.6 discusses advantages, challenges, and extensions. Section 4.7 summarizes key contributions.

4.2 A CRB-Inspired Scan Selection Method

4.2.1 The CRB and its Relevance to QMRI

Recall from Section 3.2.1 that after image reconstruction, we can model the single-voxel MR image domain data associated with a particular scan profile as

$$\mathbf{y} = \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) + \boldsymbol{\epsilon}, \tag{4.1}$$

where signal model $\mathbf{s} := [s_1, \ldots, s_D]^\mathsf{T} : \mathbb{C}^L \times \mathbb{C}^K \times \mathbb{R}^{A\times D} \mapsto \mathbb{C}^D$ relates latent $\mathbf{x} \in \mathbb{C}^L$, known $\boldsymbol{\nu} \in \mathbb{C}^K$, and acquisition $\mathbf{P} \in \mathbb{R}^{A\times D}$ parameters to noisy scan profile image data $\mathbf{y} \in \mathbb{C}^D$, barring noise $\boldsymbol{\epsilon} \in \mathbb{C}^D$. Assuming (as in Section 3.2.1) complex Gaussian noise $\boldsymbol{\epsilon} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, the likelihood function (3.4) is (to within constants independent of $\mathbf{x}$)

$$\mathsf{L}(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\|\mathbf{y} - \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})\|^2_{\boldsymbol{\Sigma}^{-1}}\right). \tag{4.2}$$

Under suitable[2] regularity conditions, the Fisher information matrix $\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) \in \mathbb{C}^{L\times L}$ [86] characterizes the imprecision of unbiased estimates of $\mathbf{x}$ from $\mathbf{y}$, given $\boldsymbol{\nu}$ and $\mathbf{P}$:

$$\begin{aligned}\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) &:= \mathsf{E}_\mathbf{y}\Big((\nabla_\mathbf{x} \log \mathsf{L}(\mathbf{x}|\mathbf{y}))^\mathsf{H} \nabla_\mathbf{x} \log \mathsf{L}(\mathbf{x}|\mathbf{y})\Big) \\ &= (\nabla_\mathbf{x}\mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}))^\mathsf{H} \boldsymbol{\Sigma}^{-1} \nabla_\mathbf{x}\mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}),\end{aligned} \tag{4.3}$$

[2]In particular, $\mathbf{s}$ must be analytic in complex components of $\mathbf{x}$.

where $\mathsf{E}_{\mathbf{y}}(\cdot)$ denotes element-wise expectation with respect to $\mathbf{y}$. In particular, the matrix CRB [87] ensures that any unbiased[3] estimator $\widehat{\mathbf{x}}$ satisfies

$$\operatorname{cov} \widehat{\mathbf{x}}; \boldsymbol{\nu}, \mathbf{P} \succeq \mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}), \tag{4.4}$$

where for arbitrary, equally-sized $\mathbf{C}_1$ and $\mathbf{C}_2$, matrix inequality $\mathbf{C}_1 \succeq \mathbf{C}_2$ means $\mathbf{C}_1 - \mathbf{C}_2$ is positive semi-definite. In the following, we design an optimization problem based on the CRB to guide QMRI scan design for relaxometry.

4.2.2 Min-max Optimization Problem for Scan Design

Following [88], we focus on minimizing a weighted average of the variances in each of the L latent object parameter estimates. A reasonable objective function for overall estimator precision is therefore given by

$$\Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) = \operatorname{tr}\left(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})\mathbf{W}^{\mathsf{T}}\right), \tag{4.5}$$

where $\mathbf{W} \in \mathbb{R}^{L \times L}$ is a diagonal, application-specific matrix of weights, preselected to control the relative importance of precisely estimating the L latent object parameters. For scan design, we would like to minimize (4.5) with respect to scan parameters $\mathbf{P}$.

The CRB depends not only on $\mathbf{P}$ but also on the spatially varying object parameters $\mathbf{x}$ and $\boldsymbol{\nu}$. Thus, one cannot perform scan design by "simply" minimizing Ψ with respect to scan parameters $\mathbf{P}$. Instead, we pose a *min-max* optimization problem for scan design: we seek candidate scan parameters $\breve{\mathbf{P}}$ over a search space $\mathbb{P}$ that *minimize* the worst-case (*i.e.*, *maximum*) cost $\widetilde{\Psi}^{\mathrm{t}}$, as viewed over "tight" object parameter ranges $\mathbb{X}^{\mathrm{t}}$ and $\mathbb{N}^{\mathrm{t}}$:

$$\breve{\mathbf{P}} \in \left\{ \arg \min_{\mathbf{P} \in \mathbb{P}} \widetilde{\Psi}^{\mathrm{t}}(\mathbf{P}) \right\}, \text{ where} \tag{4.6}$$

$$\widetilde{\Psi}^{\mathrm{t}}(\mathbf{P}) := \max_{\substack{\mathbf{x} \in \mathbb{X}^{\mathrm{t}} \\ \boldsymbol{\nu} \in \mathbb{N}^{\mathrm{t}}}} \Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \tag{4.7}$$

Here, we select *latent* parameter set $\mathbb{X}^{\mathrm{t}}$ based on the application and *known* parameter set $\mathbb{N}^{\mathrm{t}}$ based on the spatial variation typically observed in the known parameters $\boldsymbol{\nu}$. Min-max

[3] Provided that signal model $\mathbf{s}(\cdot)$ is injective and continuously differentiable in $\mathbf{x}$, maximum-likelihood (ML) estimates of $\mathbf{x}$ (discussed in Subsection 3.2.2.2) are asymptotically consistent and efficient, and thus asymptotically achieve the CRB. For Gaussian noise models, increasing sample size is statistically equivalent to increasing signal-to-noise ratio. Thus, in regions where the data provides sufficiently high SNR (and is thus approximately Gaussian-distributed even in magnitude [30]), ML estimates will exhibit minimal estimation bias, and the CRB can be used to reliably predict ML estimation error. Table 4.2 in Subsection 4.4.1 empirically explores the validity of this high-SNR assumption, through simulations at realistic noise levels.

approach (4.9) should ensure good estimation precision over a range of parameter values.

Since Ψ is in general non-convex with respect to $\mathbf{P}$, it may have multiple global minimizers as well as other scan parameters that are nearly global minimizers. To improve robustness to object parameter variations, we form an expanded set of candidate scan parameters by also including scan parameters that yield costs to within a tolerance $\delta \ll 1$ of the optimum. Mathematically, we define this expanded set of candidate scan parameter combinations (for a given scan profile) as

$$\breve{\mathbb{P}} := \left\{ \mathbf{P} : \widetilde{\Psi}^{\mathrm{t}}(\mathbf{P}) - \widetilde{\Psi}^{\mathrm{t}}\left(\breve{\mathbf{P}}\right) \leq \delta \widetilde{\Psi}^{\mathrm{t}}\left(\breve{\mathbf{P}}\right) \right\}. \tag{4.8}$$

To select amongst these candidate scan parameters, we employ a robustness criterion: we select the single scan parameter $\widehat{\mathbf{P}}$ that degrades the least when the worst-case cost is viewed over widened object parameter sets $\mathbb{X}^{\mathrm{b}} \supseteq \mathbb{X}^{\mathrm{t}}$ and $\mathbb{N}^{\mathrm{b}} \supseteq \mathbb{N}^{\mathrm{t}}$:

$$\widehat{\mathbf{P}} = \arg \min_{\mathbf{P} \in \breve{\mathbb{P}}} \widetilde{\Psi}^{\mathrm{b}}(\mathbf{P}), \text{ where} \tag{4.9}$$

$$\widetilde{\Psi}^{\mathrm{b}}(\mathbf{P}) := \max_{\substack{\mathbf{x} \in \mathbb{X}^{\mathrm{b}} \\ \boldsymbol{\nu} \in \mathbb{N}^{\mathrm{b}}}} \Psi(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}). \tag{4.10}$$

To compare different scan profiles, we select corresponding search spaces $\mathbb{P}$ to satisfy acquisition constraints (*e.g.*, total scan time), but otherwise hold optimization parameters $\mathbf{W}$, δ, $\mathbb{X}^{\mathrm{t}}$, $\mathbb{X}^{\mathrm{b}}$, $\mathbb{N}^{\mathrm{t}}$, $\mathbb{N}^{\mathrm{b}}$ fixed. Since Ψ is data-independent, we can solve (4.6) and (4.9) offline for each scan profile. The result of each profile's min-max optimization process (4.9) is a corresponding optimized scan parameter matrix $\widehat{\mathbf{P}}$ that is suitable for the range of latent $\mathbf{x}$ and known $\boldsymbol{\nu}$ object parameters specified in $\mathbb{X}^{\mathrm{t}}$ and $\mathbb{N}^{\mathrm{t}}$, and is robust to variations in those parameters over broader sets $\mathbb{X}^{\mathrm{b}}$ and $\mathbb{N}^{\mathrm{b}}$, respectively.

4.3 Optimizing SS Sequences for Relaxometry in the Brain

This section applies the methods of Section 4.2.2 to the problem of scan design for joint T_1, T_2 estimation from combinations of SS sequences. Section 4.3.1 details how we use optimization problems (4.6) and (4.9) to tailor three SPGR and DESS scan combinations for precise T_1, T_2 estimation in white matter (WM) and grey matter (GM) regions of the brain. Section 4.3.2 compares the predicted performance of the three optimized scan profiles.

4.3.1 Scan Design Details

There are numerous candidate scan profiles involving DESS and/or other pulse sequences that may be useful for fast, accurate T_1, T_2 mapping. In this chapter, we consider combinations of magnitude SPGR and DESS scans for estimating the $L \leftarrow 3$ latent parameters T_1, T_2, and proportionality constant c_2 (defined in Example 3.2.1.2), given knowledge of transmit field inhomogeneity κ^{t} as $K \leftarrow 1$ known parameter. With proper RF phase cycling and gradient spoiling, the SPGR signal s_{S} (as expressed in (2.27)) contains no explicit T_2 dependence. SPGR's reduced dependence on spatially varying unknowns is reason for its use in T_1 mapping [67, 66, 68] and subsequent T_2 mapping from other sequences [36, 5]. In a similar spirit, we examine scan profiles containing SPGR over other SS sequences because we predict that the SPGR sequence's T_2-independence may help estimators disentangle T_2 from other unknown sources of DESS signal contrast.

As respectively discussed in Examples 3.2.1.1-3.2.1.2, each SPGR and DESS scan leaves $\mathbf{p} \leftarrow [\alpha_0, T_{\mathrm{R}}]^{\mathsf{T}}$ as $A \leftarrow 2$ acquisition parameters available to optimize. A given scan profile consisting of S_{SPGR} SPGR and S_{DESS} DESS scans yields $D \leftarrow S_{\mathrm{SPGR}} + 2S_{\mathrm{DESS}}$ datasets. We optimize such a scan profile by solving (4.9) over a dimension-$AD \leftarrow 2(S_{\mathrm{SPGR}} + 2S_{\mathrm{DESS}})$ space of scan parameters.

We select constraints on search space $\mathbb{P}$ based on hardware limitations and desired scan profile properties. Since each pair of DESS signals must share the same $\mathbf{p}$, the search space $\mathbb{P}$ is reduced to $\mathbb{A}_{0,\mathrm{SPGR}}^{S_{\mathrm{SPGR}}} \times \mathbb{A}_{0,\mathrm{DESS}}^{S_{\mathrm{DESS}}} \times \mathbb{T}_{\mathrm{R,SPGR}}^{S_{\mathrm{SPGR}}} \times \mathbb{T}_{\mathrm{R,DESS}}^{S_{\mathrm{DESS}}}$ (superscripts denote Cartesian powers). We assign flip angle ranges $\mathbb{A}_{0,\mathrm{SPGR}} \leftarrow [5, 90]^{\circ}$ and $\mathbb{A}_{0,\mathrm{DESS}} \leftarrow [5, 90]^{\circ}$ to restrict RF energy deposition within the orifice of the main magnet. We set feasible T_{R} solution sets $\mathbb{T}_{\mathrm{R,SPGR}} \leftarrow [12.2, +\infty)$ms and $\mathbb{T}_{\mathrm{R,DESS}} \leftarrow [17.5, +\infty)$ms based on pulse sequence designs that control for other scan parameters. These control parameters are described in further detail in Section 4.4, and are held fixed in all subsequent SPGR and DESS experiments. To equitably compare optima from different scan profiles, we require

$$\mathbf{T}_{\mathrm{R}} := [T_{\mathrm{R},1}, \dots, T_{\mathrm{R},S_{\mathrm{SPGR}}}, T_{\mathrm{R},S_{\mathrm{SPGR}}+1}, \dots, T_{\mathrm{R},S_{\mathrm{SPGR}}+S_{\mathrm{DESS}}}]^{\mathsf{T}}$$

to satisfy a total time constraint, $\|\mathbf{T}_{\mathrm{R}}\|_1 \leq T_{\max}$. For a scan profile consisting of S_{SPGR} SPGR and S_{DESS} DESS scans, these constraints collectively reduce the search space dimension from AD to $2(S_{\mathrm{SPGR}} + S_{\mathrm{DESS}}) - 1$.

Prior works have considered T_1 or T_2 estimation from as few as 2 SPGR [78, 36] or 1 DESS [71] scan(s), respectively. We likewise elect to optimize the $(S_{\mathrm{SPGR}}, S_{\mathrm{DESS}}) \leftarrow (2, 1)$ scan profile as a benchmark. We choose $T_{\max} \leftarrow 2(12.2) + 1(17.5) = 41.9$ms and select other scan profiles capable of meeting this time constraint. Requiring that candidate profiles contain $S_{\text{DESS}} \geq 1$ DESS scans for T_2 contrast and satisfy $D \geq L(=3)$ for well-conditioned estimation, we note that $(1, 1)$ and $(0, 2)$ are the only other eligible profiles.

In the ensuing experiments, we focus on precise T_1, T_2 estimation in the brain. Noting that $T_1 \sim 10T_2$, we choose $\mathbf{W} \leftarrow \operatorname{diag}(0.1, 1, 0)$ to place approximately equal importance on precise T_1 versus T_2 estimation and zero weight on proportionality constant c_2 estimation (obviating the need for complex differentiation in (4.3)). Since Ψ then depends on c_2 through only a scale factor, it suffices to fix $c_2 \leftarrow 1$ and design the latent object parameter range as $\mathbb{X}^{\text{t}} \leftarrow \mathbb{T}_1^{\text{t}} \times \mathbb{T}_2^{\text{t}} \times 1$. Here, $\mathbb{T}_1^{\text{t}} \leftarrow [800, 1400]$ms and $\mathbb{T}_2^{\text{t}} \leftarrow [50, 120]$ms correspond to WM and GM regions of interest (ROIs) at 3T [31, 32]. We take $\mathbb{N}^{\text{t}} \leftarrow [0.9, 1.1]$ to account for 10% transmit field spatial variation. Broadened ranges $\mathbb{X}^{\text{b}} \leftarrow [400, 2000]\text{ms} \times [40, 200]\text{ms} \times 1$ and $\mathbb{N}^{\text{b}} \leftarrow [0.5, 2]$ are constructed to encourage solutions robust to a realistically wide range of object parameters. We assume constant noise variance $\sigma_1^2 = \cdots = \sigma_D^2 := \sigma^2$, where $\sigma^2 \leftarrow 1.49 \times 10^{-7}$ is selected to reflect measurements from normalized phantom datasets (*cf.* Section 4.4.2.1 for acquisition details). Lastly, we set $\delta \leftarrow 0.01$ to select a robust scan parameter $\widehat{\mathbf{P}}$ with associated worst-case cost $\widetilde{\Psi}^{\text{t}}\left(\widehat{\mathbf{P}}\right)$ within 1% of global optimum $\widetilde{\Psi}^{\text{t}}\left(\breve{\mathbf{P}}\right)$.

4.3.2 Scan Profile Comparisons

We solve (4.6) and (4.9) via grid search to allow illustration of $\widetilde{\Psi}^{\text{t}}(\mathbf{P})$ as well as worst-case T_1, T_2 standard deviations $\widetilde{\sigma}_{\text{T}_1}^{\text{t}}(\mathbf{P})$ and $\widetilde{\sigma}_{\text{T}_2}^{\text{t}}(\mathbf{P})$, each defined as

$$\widetilde{\sigma}_{\text{T}_1}^{\text{t}}(\mathbf{P}) := \max_{\substack{\mathbf{x} \in \mathbb{X}^{\text{t}} \\ \boldsymbol{\nu} \in \mathbb{N}^{\text{t}}}} \sigma_{T_1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}); \tag{4.11}$$

$$\widetilde{\sigma}_{\text{T}_2}^{\text{t}}(\mathbf{P}) := \max_{\substack{\mathbf{x} \in \mathbb{X}^{\text{t}} \\ \boldsymbol{\nu} \in \mathbb{N}^{\text{t}}}} \sigma_{T_2}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}), \tag{4.12}$$

where $\sigma_{T_1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})$ and $\sigma_{T_2}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})$ are corresponding diagonal elements of inverse Fisher matrix $\mathbf{F}^{-1}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P})$. Grid searches for the $(2, 1)$, $(1, 1)$, and $(0, 2)$ profiles each took about 4, 43, and 28 minutes, respectively. All experiments described hereafter were carried out using MATLAB® R2013a on a 3.5GHz desktop with 32GB RAM.

Table 4.1 compares optimized scan parameters for profiles consisting of $(2, 1)$, $(1, 1)$, and $(0, 2)$ SPGR and DESS scans, respectively. In addition to $\widetilde{\sigma}_{\text{T}_1}^{\text{t}}(\widehat{\mathbf{P}})$ and $\widetilde{\sigma}_{\text{T}_2}^{\text{t}}(\widehat{\mathbf{P}})$, Table 4.1 presents analogous worst-case standard deviations $\widetilde{\sigma}_{\text{T}_1}^{\text{b}}(\widehat{\mathbf{P}})$ and $\widetilde{\sigma}_{\text{T}_2}^{\text{b}}(\widehat{\mathbf{P}})$ over $\mathbb{X}^{\text{b}} \times \mathbb{N}^{\text{b}}$ to show how each estimator degrades over the broadened object parameter range. When viewed over tight range $\mathbb{X}^{\text{t}} \times \mathbb{N}^{\text{t}}$, the $(0, 2)$ profile provides a 11.5% reduction in worst-case cost over the other choices. Extending to broadened range $\mathbb{X}^{\text{b}} \times \mathbb{N}^{\text{b}}$, this reduction

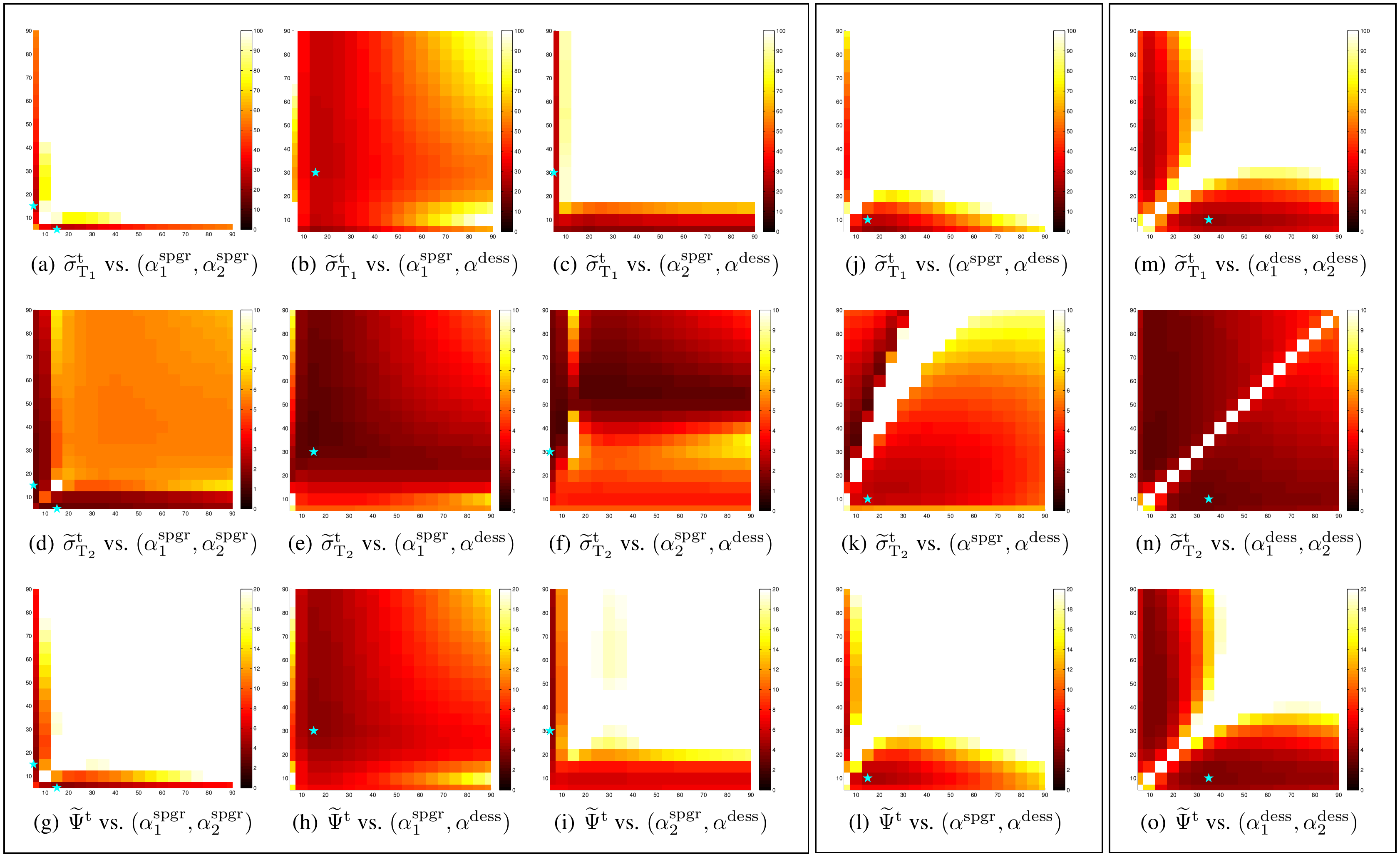
Scan	$(2,1)$	$(1,1)$	$(0,2)$
$\widehat{\alpha}_0^{\text{spgr}}$	(15,5)°	15°	–
$\widehat{\alpha}_0^{\text{dess}}$	30°	10°	(35,10)°
$\widehat{T}_R^{\text{spgr}}$	(12.2, 12.2)	13.9	–
$\widehat{T}_R^{\text{dess}}$	17.5	28.0	(24.4, 17.5)
$\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}}(\widehat{\mathbf{P}})$	28	27	**21**
$\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{b}}(\widehat{\mathbf{P}})$	154	169	**113**
$\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}}(\widehat{\mathbf{P}})$	**1.3**	2.8	1.5
$\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{b}}(\widehat{\mathbf{P}})$	9.1	8.8	**6.0**
$\widetilde{\Psi}^{\mathrm{t}}(\widehat{\mathbf{P}})$	4.0	4.9	**3.5**
$\widetilde{\Psi}^{\mathrm{b}}(\widehat{\mathbf{P}})$	17.7	17.9	**12.2**

Table 4.1: Performance summary of different scan profiles, optimized by solving (4.9) subject to scan time constraint $T_{\max} = 41.9$ms. The first row defines each profile. The next four rows describe $\widehat{\mathbf{P}}$. The latter three pairs of rows show how worst-case values degrade from tight to broad ranges. Flip angles are in degrees; all other values are in milliseconds.

grows dramatically to 31.4%. We thus observe that while the different optimized profiles afford similar estimator precision over a narrow range of interest, the $(0,2)$ profile may be preferable due to its robustness to a wide range of object parameters.

Fig. 4.1 displays heat maps of worst-case latent parameter standard deviations $\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}}$, $\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}}$ and worst-case cost $\widetilde{\Psi}^{\mathrm{t}}$ as pairs of flip angles are varied away from the optimized scan design $\widehat{\mathbf{P}}$. Boxes group subfigures corresponding to the same scan profile. Viewing the bottom row of subfigures, it is evident that $\widetilde{\Psi}^{\mathrm{t}}(\widehat{\mathbf{P}})$ takes similar values for the different scan profiles. However, it is apparent that the $(S_{\mathrm{SPGR}}, S_{\mathrm{DESS}}) = (0,2)$ profile is substantially more robust to transmit field variation than other tested profiles (namely, $(2,1)$ and $(1,1)$). Optimized worst-case cost over broadened latent parameter ranges $\widetilde{\Psi}^{\mathrm{b}}(\widehat{\mathbf{P}})$ captures this by expanding the range of possible flip angles from $\mathbb{N}^{\mathrm{t}} = [0.9, 1.1]$ to $\mathbb{N}^{\mathrm{b}} = [0.5, 2]$ to account for factor-of-two spatial variation in relative flip angle. As a result, we find that the properties of "broad" search criterion $\widetilde{\Psi}^{\mathrm{b}}$ provide a stronger reason to select the $(0,2)$ scan for joint T_1, T_2 estimation in the brain than the properties of "tight" search criterion $\widetilde{\Psi}^{\mathrm{t}}$.

As the DESS sequence has already found success for T_2 mapping from even one scan [71], it is reassuring but unsurprising that our analysis finds two DESS scans to yield the most precise T_2 estimates. More interestingly, our methods suggest that, with a minimum $S_{\mathrm{DESS}} = 2$ scans, DESS can be used to simultaneously estimate T_1 as well. In fact, for certain choices of parameter ranges, a second DESS scan is predicted to afford $\widehat{T}_1$ precision comparable to two SPGR scans.



(a) $\widetilde{\sigma}^{t}_{T_1}$ vs. $(\alpha_1^{\text{spgr}}, \alpha_2^{\text{spgr}})$ (b) $\widetilde{\sigma}^{t}_{T_1}$ vs. $(\alpha_1^{\text{spgr}}, \alpha^{\text{dess}})$ (c) $\widetilde{\sigma}^{t}_{T_1}$ vs. $(\alpha_2^{\text{spgr}}, \alpha^{\text{dess}})$ (j) $\widetilde{\sigma}^{t}_{T_1}$ vs. $(\alpha^{\text{spgr}}, \alpha^{\text{dess}})$ (m) $\widetilde{\sigma}^{t}_{T_1}$ vs. $(\alpha_1^{\text{dess}}, \alpha_2^{\text{dess}})$

(d) $\widetilde{\sigma}^{t}_{T_2}$ vs. $(\alpha_1^{\text{spgr}}, \alpha_2^{\text{spgr}})$ (e) $\widetilde{\sigma}^{t}_{T_2}$ vs. $(\alpha_1^{\text{spgr}}, \alpha^{\text{dess}})$ (f) $\widetilde{\sigma}^{t}_{T_2}$ vs. $(\alpha_2^{\text{spgr}}, \alpha^{\text{dess}})$ (k) $\widetilde{\sigma}^{t}_{T_2}$ vs. $(\alpha^{\text{spgr}}, \alpha^{\text{dess}})$ (n) $\widetilde{\sigma}^{t}_{T_2}$ vs. $(\alpha_1^{\text{dess}}, \alpha_2^{\text{dess}})$

(g) $\widetilde{\Psi}^{t}$ vs. $(\alpha_1^{\text{spgr}}, \alpha_2^{\text{spgr}})$ (h) $\widetilde{\Psi}^{t}$ vs. $(\alpha_1^{\text{spgr}}, \alpha^{\text{dess}})$ (i) $\widetilde{\Psi}^{t}$ vs. $(\alpha_2^{\text{spgr}}, \alpha^{\text{dess}})$ (l) $\widetilde{\Psi}^{t}$ vs. $(\alpha^{\text{spgr}}, \alpha^{\text{dess}})$ (o) $\widetilde{\Psi}^{t}$ vs. $(\alpha_1^{\text{dess}}, \alpha_2^{\text{dess}})$

Figure 4.1: Worst-case standard deviations $\widetilde{\sigma}^{t}_{T_1}$ (top), $\widetilde{\sigma}^{t}_{T_2}$ (middle), and cost $\widetilde{\Psi}^{t}$ (bottom), versus pairs of nominal flip angles, holding other scan parameters fixed at selected profile $\widehat{\mathbf{P}}$. Subfigures (a)-(i), (j)-(l), and (m)-(o) correspond to scan profiles containing $(S_{\text{SPGR}}, S_{\text{DESS}}) = (2, 1), (1, 1)$, and $(0, 2)$ SPGR and DESS scans, respectively. Selected scan parameters (starred) are within $\delta = 1\%$ of global minimizers and retain as much estimator precision as possible over a wide range of latent object parameters. All axes range from 5 to 90 degrees, in 5-degree increments. Colorbar ranges are $[0, 100]$, $[0, 10]$, and $[0, 20]$ milliseconds for rows of $\widetilde{\sigma}^{t}_{T_1}$, $\widetilde{\sigma}^{t}_{T_2}$, and $\widetilde{\Psi}^{t}$ subfigures, respectively. The optimized $(0, 2)$ profile appears most robust to transmit field spatial variation.

4.4 Experimentation

To test our approach to optimized scan design (described in Section 4.2.2), we estimate $\mathbf{T}_1$ and $\mathbf{T}_2$ maps (using maximum likelihood (ML) and regularized likelihood (RL) methods detailed in Section 3.2.2) from datasets collected using the scan profiles optimized in Section 4.3. In Section 4.4.1, we study estimator statistics from simulated data. In Sections 4.4.2-4.4.3, we progress to phantom and *in vivo* datasets to evaluate scan profile performance under increasingly complex settings. For the latter experiments, we use reference parameter maps from classical (long) pulse sequences, in lieu of ground truth maps.

4.4.1 Numerical Simulations

We select T_1 and T_2 WM and GM values based on previously reported measurements at 3T [31, 32] and extrapolate other nuisance latent object parameters m_0 and T_2^* from measurements at 1.5T [33]. We assign these parameter values to the discrete anatomy of the BrainWeb digital phantom [34, 33] to create ground truth $\mathbf{M}_0, \mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_2^* \in \mathbb{R}^V$ maps. We then choose acquisition parameters based on Table 4.1 (with fixed $T_\mathrm{E} = 4.67\text{ms}$) and apply models (2.27) and (2.41)-(2.43) to the 81st slices of these true maps to compute noiseless 217×181 SPGR and DESS image-domain data, respectively.

For each scan profile, we corrupt the corresponding (complex) noiseless dataset $\mathbf{S}$ with additive complex Gaussian noise, whose variance $\sigma^2 \leftarrow 1.49 \times 10^{-7}$ is set to match CRB calculations. This yields realistically noisy datasets $\mathbf{Y}$ ranging from 105-122 signal-to-noise ratio (SNR), where SNR is defined here as

$$\mathsf{SNR}(\mathbf{S}, \mathbf{Y}) := \frac{\|\mathbf{S}\|_\mathrm{F}}{\|\mathbf{Y} - \mathbf{S}\|_\mathrm{F}}. \tag{4.13}$$

We use each profile's noisy magnitude dataset $|\mathbf{Y}|$ to compute estimates $\widehat{\mathbf{T}}_1$ and $\widehat{\mathbf{T}}_2$ We then evaluate estimator bias and variance from latent ground truth $\mathbf{T}_1$ and $\mathbf{T}_2$ maps.

In these simulations, we intentionally neglect to model a number of physically realistic effects because their inclusion would complicate study of estimator statistics. First and foremost, we assume knowledge of a uniform transmit field, to avoid confounding κ^t and T_1, T_2 estimation errors. For a similar reason, spatial variation in the sensitivity of a single receive coil is also not considered. We omit modeling partial volume effects to ensure deterministic knowledge of WM and GM ROIs. We will explore the influence of these (and other) nuisance effects on scan design in later subsections and chapters.

To isolate bias due to estimator nonlinearity from regularization bias, we solve ML problem (3.7) only, and do not proceed to solve RL problem (3.8). This permits considera-

Scan	$(2, 1)$	$(1, 1)$	$(0, 2)$	Truth
WM $\widehat{T}_1^{\mathrm{ML}}$	830 ± 17	830 ± 15	830 ± 14	832
GM $\widehat{T}_1^{\mathrm{ML}}$	$1330 \pm 30.$	1330 ± 24	1330 ± 24	1331
WM $\widehat{T}_2^{\mathrm{ML}}$	$80. \pm 1.0$	$80. \pm 2.1$	79.6 ± 0.94	79.6
GM $\widehat{T}_2^{\mathrm{ML}}$	$110. \pm 1.4$	$110. \pm 3.0$	$110. \pm 1.6$	110

Table 4.2: Sample means $\pm$ sample standard deviations of $\mathbf{T}_1$ and $\mathbf{T}_2$ ML estimates in WM and GM ROIs of simulated data, compared across different optimized scan profiles. Sample means exhibit insignificant bias, and sample standard deviations are consistent with worst-case standard deviations $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1}$ and $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2}$ reported in Table 4.1. All values are reported in milliseconds.

tion of T_1, T_2 estimation from each of the 7733 WM or 9384 GM data points as voxel-wise independent realizations of the same estimation problem. To minimize quantization bias, we optimize (3.7) using a finely spaced dictionary of signal vectors from 1000 T_1 and T_2 values logarithmically spaced between $[10^2, 10^{3.5}]$ and $[10^1, 10^{2.5}]$, respectively. Using 10^6 dictionary elements, solving (3.7) took less than 7 minutes for each tested scan design $\widehat{\mathbf{P}}$.

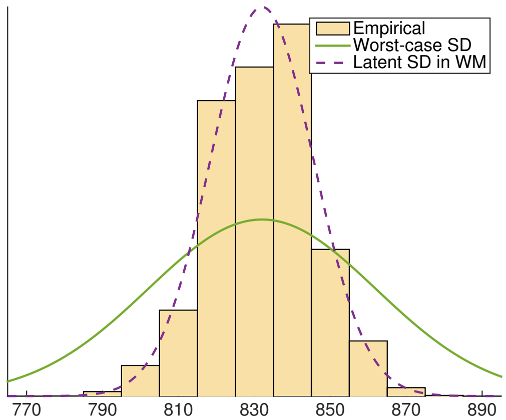
Table 4.2 verifies[4] that, despite model nonlinearity and Rician noise, estimation bias in WM- and GM-like voxels is negligible. Sample standard deviations are consistent with $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1}$ and $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2}$ (*cf.* Table 4.1). We observe that the $(1, 1)$ and $(0, 2)$ profiles afford high $\widehat{\mathbf{T}}_1^{\mathrm{ML}}$ precision, while the $(2, 1)$ and $(0, 2)$ scans afford high $\widehat{\mathbf{T}}_2^{\mathrm{ML}}$ precision. In agreement with the predictions of $\widetilde{\Psi}^{\mathrm{t}}$ and $\widetilde{\Psi}^{\mathrm{b}}$, these simulation studies suggest that at these SNR levels, an optimized profile containing 2 DESS scans can permit $\mathbf{T}_1$ and $\mathbf{T}_2$ estimation precision in WM and GM comparable to optimized profiles containing SPGR/DESS combinations.

Fig. 4.2 histograms (voxel-wise independent) ML estimates $\widehat{T}_1^{\mathrm{ML}}$ and $\widehat{T}_2^{\mathrm{ML}}$ from the $(0, 2)$ scan profile. Each histogram is over a WM or GM ROI, within which all voxels are assigned the same single-component true T_1 and T_2 nominal value, listed in Table 4.2.

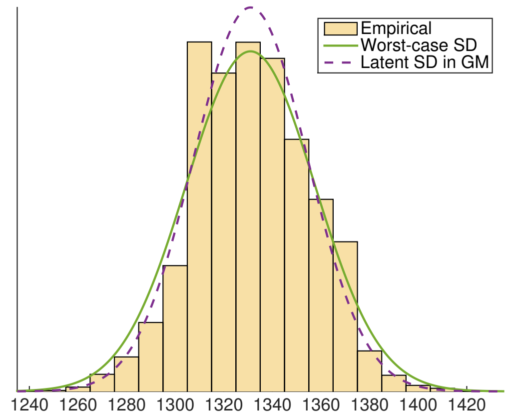
Overlaid in dashed maroon are normal distributions with latent means T_1 and T_2 and variances computed from the Fisher matrix at T_1, T_2 values in WM or GM. It is apparent that despite finite SNR and Rician noise, $\widehat{T}_1^{\mathrm{ML}}$ and $\widehat{T}_2^{\mathrm{ML}}$ exhibit negligible bias and near-Gaussian shape, suggesting locally linear behavior of the DESS signal model in T_1 and T_2 ($\widehat{T}_1^{\mathrm{ML}}$ and $\widehat{T}_2^{\mathrm{ML}}$ distributions from other profiles are similar).

The subfigures of Fig. 4.2 superimpose in solid green a second set of normal distributions, with the same means T_1 and T_2 as before, but worst-case standard deviations $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1}$ and $\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2}$. The separations between these distribution pairs visually depict how estimator

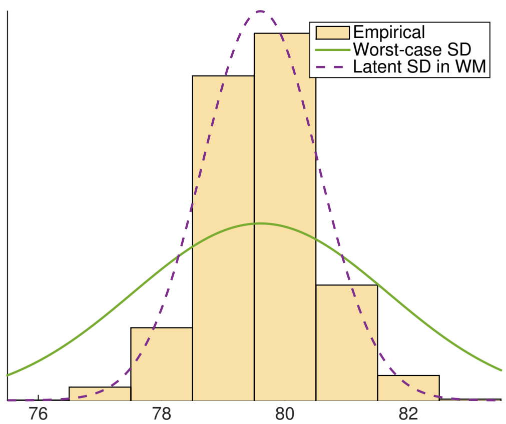
[4]Each sample statistic presented in this chapter is rounded off to the highest place value of its corresponding uncertainty measure. For simplicity, each uncertainty measure is itself endowed one extra significant figure. Decimal points indicate the significance of trailing zeros.



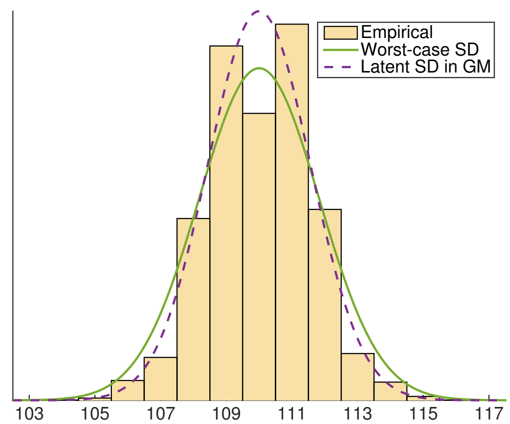
(a) $\widehat{T}_1^{\mathrm{ML}}$ in voxels with WM-like $T_1 \leftarrow 832$



(b) $\widehat{T}_1^{\mathrm{ML}}$ in voxels with GM-like $T_1 \leftarrow 1331$



(c) $\widehat{T}_2^{\mathrm{ML}}$ in voxels with WM-like $T_2 \leftarrow 79.6$



(d) $\widehat{T}_2^{\mathrm{ML}}$ in voxels with GM-like $T_2 \leftarrow 110$

Figure 4.2: Histograms of T_1 and T_2 estimates from noisy independent measurements of a *single* nominal WM or GM value. In each plot, two normal distributions are overlaid, each with latent means T_1 and T_2. In (a)-(b) and (c)-(d), the solid green curve is $\mathcal{N}\big(T_1, (\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_1})^2\big)$ and $\mathcal{N}\big(T_2, (\widetilde{\sigma}^{\mathrm{t}}_{\mathrm{T}_2})^2\big)$, respectively. In (a)-(d), the dashed maroon curves have variances computed from the Fisher information at *a priori* unknown T_1, T_2 values in WM or GM. These plots correspond to an optimized $(0, 2)$ scan profile; analogous plots for other profiles are visually similar. At realistic noise levels, parameter estimates distribute with minimal bias and near-Gaussian shape. Thus, the CRB reliably approximates $\widehat{T}_1^{\mathrm{ML}}$ and $\widehat{T}_2^{\mathrm{ML}}$ errors.

variances specific to WM or GM T_1 and T_2 values differ from worst-case variances. Using the fixed latent object parameters to optimize scan profiles can tailor scans for precise estimation in *either* WM *or* GM. In contrast, the proposed min-max formulation finds scan parameters that ensure precise estimation in *both* WM *and* GM.

4.4.2 Phantom Experiments

This subsection describes two experiments. In the first experiment, we compare SPGR/DESS scan profiles described in Table 4.1 (as well as a reference profile consisting of IR and SE scans) against nuclear magnetic resonance (NMR) measurements from the National Institute for Standards and Technology (NIST) [2]. These measurements provide information about *ROI sample means* and *ROI sample standard deviations* (Fig. 4.5), which we define as first- and second-order statistics computed across voxels within an ROI. In the second experiment, we repeat the SPGR/DESS scan profiles 10 times and compute *sample standard deviation maps* across repetitions. Taking ROI sample means of these maps gives *pooled sample standard deviations* (Table 4.4), which indicate relative scan profile precision.

4.4.2.1 Within-ROI Statistics

We acquire combinations of $(2, 1)$, $(1, 1)$, and $(0, 2)$ SPGR and DESS coronal scans of a High Precision Devices® MR system phantom T_2 array. For each scan profile, we prescribe the optimized flip angles $\widehat{\boldsymbol{\alpha}}_0$ and repetition times $\widehat{\mathbf{T}}_\mathrm{R}$ listed in Table 4.1, and hold all other scan parameters fixed. We achieve the desired nominal flip angles by scaling a 20mm slab-selective Shinnar-Le Roux excitation [89], of duration 1.28ms and time-bandwidth product 4. For each DESS (SPGR) scan, we apply 2 (10) spoiling phase cycles over a 5mm slice thickness. We acquire all steady-state phantom and *in vivo* datasets with a $256 \times 256 \times 8$ matrix over a $240 \times 240 \times 40$ mm^3 field of view (FOV). Using a 31.25kHz readout bandwidth, we acquire all data at minimum $T_\mathrm{E} \leftarrow 4.67$ms before or after RF excitations. To avoid slice-profile effects, we sample $\mathbf{k}$-space over a 3D Cartesian grid. After Fourier transform of the raw datasets, only one of the excited image slices is used for subsequent parameter mapping. Including time to reach steady-state, each steady-state scan profile requires 1m37s scan time.

To validate a reference scan profile for use in *in vivo* experiments, we also collect 4 IR and 4 SE scans. For (phase-sensitive, SE) IR, we hold $(T_\mathrm{R}, T_\mathrm{E}) \leftarrow (1400, 14)$ms fixed and vary (adiabatic) inversion time $T_\mathrm{I} \in \{50, 150, 450, 1350\}$ms across scans. For SE, we similarly hold $T_\mathrm{R} \leftarrow 1000$ms fixed and vary echo time $T_\mathrm{E} \in \{10, 30, 60, 150\}$ms across scans. We prescribe these scan parameters to acquire 256×256 datasets over the same

$240 \times 240 \times 5$ mm^3 slice processed from the SPGR/DESS datasets. Each IR and SE scan requires 5m58s and 4m16s, for a total 40m58s scan time.

We additionally collect a pair of Bloch-Siegert shifted 3D SPGR scans for separate transmit field estimation [90]. We insert a 9ms Fermi pulse at $\pm$8kHz off-resonance into an SPGR sequence immediately following on-resonant excitation. We estimate regularized transmit field maps [91] from the resulting pair of datasets. We normalize this transmit field map estimate by the 0.075G peak Fermi pulse amplitude to estimate transmit coil spatial variation map $\mathbf{s}_\mathrm{t}$. After calibration via separate measurements, we take $\mathbf{s}_\mathrm{t}$ as known. For consistency, we account for flip angle variation when estimating $\mathbf{T}_1$ and $\mathbf{T}_2$ from both candidate (SPGR/DESS) and reference (IR/SE) scan profiles. With a repetition time of 21.7ms, this transmit field mapping acquisition requires 1m40s total scan time.

We acquire all phantom datasets using a GE Discovery™ MR750 3.0T scanner with an 8-channel receive head array. We separately normalize and combine coil data from each scan profile using a natural extension of [92] to the case of multiple datasets. For each optimized SPGR/DESS scan profile $\widehat{\mathbf{P}}$, we pre-cluster known parameter maps $\mathbf{N}$ into 10 clusters using k-means++ [93] and use each of the 10 cluster means to compute a corresponding dictionary of signal vectors from 300 T_1 and T_2 values logarithmically spaced between $[10^{1.5}, 10^{3.5}]$ and $[10^{0.5}, 10^{3}]$, respectively. We then iterate over clusters and use each dictionary in conjunction with corresponding coil-combined magnitude image data to produce ML parameter estimates $\widehat{\mathbf{X}}_\mathrm{ML}\left(\mathbf{N}, \widehat{\mathbf{P}}\right)$. We subsequently solve RL problem (3.8) with initialization $\widehat{\mathbf{X}}_\mathrm{ML}\left(\mathbf{N}, \widehat{\mathbf{P}}\right)$ to obtain regularized estimates $\widehat{\mathbf{X}}_\mathrm{RL}\left(\mathbf{N}, \widehat{\mathbf{P}}\right)$ for each $\widehat{\mathbf{P}}$. We design regularizer (3.9) to encourage RL parameter estimates from different scan profiles to exhibit similar levels of smoothness. Letting $l \in \{1, 2, 3\}$ enumerate latent object parameters $\mathbf{T}_1$, $\mathbf{T}_2$, and the proportionality constant, we choose mild regularization parameters $(\beta_1, \beta_2, \beta_3) := D \times (2^{-21}, 2^{-23}, 2^{-26})$ to scale with the number of datasets and fix shape parameters $(\gamma_1, \gamma_2, \gamma_3) := (2^5 \text{ ms}, 2^2 \text{ ms}, 2^{-2})$ to values on the order of anticipated standard deviations. We iteratively update $\mathbf{X}$ until convergence criterion

$$\left\|\mathbf{X}^{(i)} - \mathbf{X}^{(i-1)}\right\|_\mathrm{F} < 10^{-7}\left\|\mathbf{X}^{(i)}\right\|_\mathrm{F} \tag{4.14}$$

is satisfied. For all steady-state profiles tested, ML initializations and RL reconstructions of phantom datasets require less than 3m30s and 9s, respectively.

We next describe sequential[5] $\mathbf{T}_1$, then $\mathbf{T}_2$ estimation from IR and SE reference scans.

[5]We initially attempted to circumvent sequential $\mathbf{T}_1$, then $\mathbf{T}_2$ estimation by instead jointly estimating $\mathbf{M}_0$, $\mathbf{T}_1$, $\mathbf{T}_2$, and inversion efficiency from the IR and SE datasets together. Even using magnitude data and signal models, this resulted in heavily biased parameter maps, possibly due to the dependence of adiabatic inversion efficiency on relaxation parameters [94].

We first jointly coil-combine all 8-channel IR and SE phantom datasets to produce complex images. We next estimate $\mathbf{T}_1$ along with a (nuisance parameter) inversion efficiency map via (3.7) and (3.8) from the 4 complex coil-combined IR images. By using the same flip angle scaling map $\mathbf{s}_\mathrm{t}$ as is used for SPGR/DESS profiles, we estimate $\mathbf{T}_1$ using a signal model similar to one proposed in [95], which accounts for imperfect excitation/refocusing and imperfect inversion. We then take both $\mathbf{T}_1$ and $\mathbf{s}_\mathrm{t}$ as known and estimate $\mathbf{T}_2$ along with nuisance parameter $\mathbf{M}_0$ (accounting for imperfect excitation/refocusing and incomplete recovery) via (3.7) and (3.8) from the 4 complex coil-combined SE images. We hold all other reconstruction details identical to those of SPGR/DESS reconstructions.

Figs. 4.3-4.4 compare in color and grayscale phantom $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from optimized scan profiles. Vials are enumerated in Fig. 4.5 in descending T_1 and T_2 order. Vials corresponding to tight $\mathbb{X}^\mathrm{t}$ and broad $\mathbb{X}^\mathrm{b}$ parameter ranges are highlighted with orange and yellow labels, respectively. Within these vials of interest, parameter maps from different scans appear visually similar.

In higher-T_1 vials (and the surrounding water), more bias is apparent in $\widehat{\mathbf{T}}_1$ ML and RL estimates from the $(0, 2)$ scan profile than from the $(2, 1)$ and $(1, 1)$ scan profiles. With the signal models used in this study, the images suggest that scan profiles consisting of at least one SPGR scan may offer increased protection against T_1 estimation bias.

Fig. 4.5 plots sample means and sample standard deviations computed within circular ROIs of phantom $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates. The highlighted orange and yellow parameter spaces correspond to design ranges $\mathbb{X}^\mathrm{t}$ and $\mathbb{X}^\mathrm{b}$. $\mathbf{T}_1$ estimates from both the candidate $(2, 1)$, $(1, 1)$, and $(0, 2)$ (SPGR, DESS) and reference $(4, 4)$ (IR, SE) profiles are in reasonable agreement with NIST estimates [2] across the vial range. $\mathbf{T}_2$ estimates from all profiles are also in good agreement with NIST for vials within $\mathbb{X}^\mathrm{b}$. SPGR/DESS profiles likely underestimate large T_2 values ($\geq$200ms) due to greater influence of diffusion in DESS [96, 97, 98]. SPGR/DESS profiles possibly overestimate and the IR/SE profile likely underestimates short ($\leq$30ms) and very short ($\leq$15ms) T_2 values, respectively, due to poorly conditioned estimation. Table 4.3 replicates sample statistics in Fig. 4.5 for vials 5-8. Compared to ML initializations, (weakly) regularized estimates reduce error bars without introducing substantial additional bias.

4.4.2.2 Across-Repetition Statistics

In a second study, we repeat the $(2, 1)$, $(1, 1)$, and $(0, 2)$ scan profiles 10 times each and separately compute $\mathbf{T}_1$ and $\mathbf{T}_2$ ML estimates for each repetition of each scan profile. We then estimate the standard deviation across repetitions on a per-voxel basis, to produce sample standard deviation maps for each profile. Each ROI voxel of the sample standard

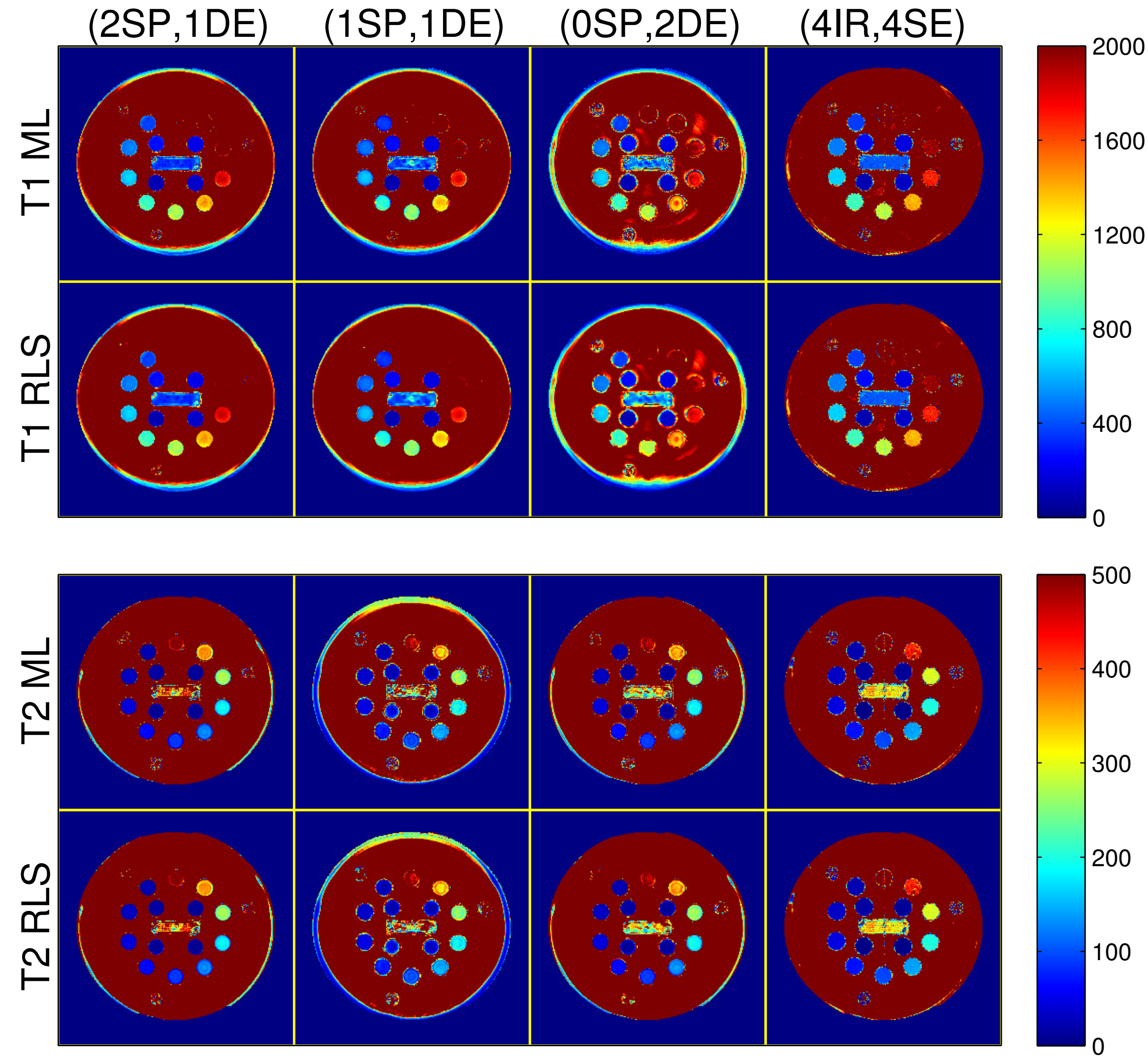


Figure 4.3: Colorized $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from an HPD® quantitative phantom. Columns correspond to scan profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Fig. 4.4 provides identical grayscale images that enumerate vials. Colorbar ranges are in milliseconds.

deviation map is a better estimate of the *population standard deviation* (which the CRB characterizes) than the ROI sample standard deviation from a single repetition, because the latter estimate is contaminated with slight spatial variation of voxel population means (due to imaging non-idealities such as Gibbs ringing due to $\mathbf{k}$-space truncation).

Table 4.4 reports pooled sample standard deviations and pooled standard errors of the sample standard deviations (computed via expressions in [3]) for phantom vials within (or nearly within) tight design range $\mathbb{X}^{\mathrm{t}}$ (marked orange in Fig. 4.4). Due to error propagation

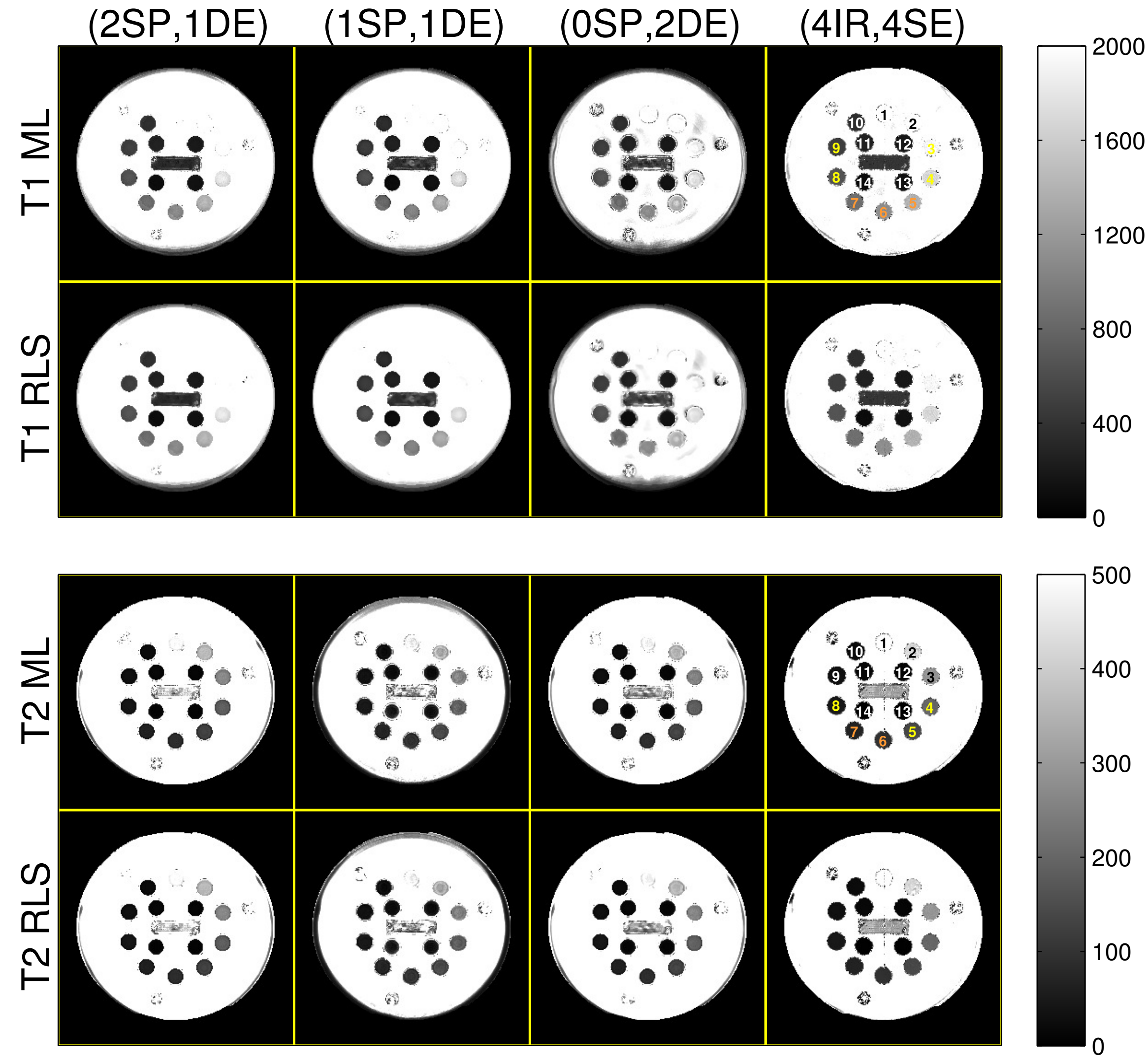
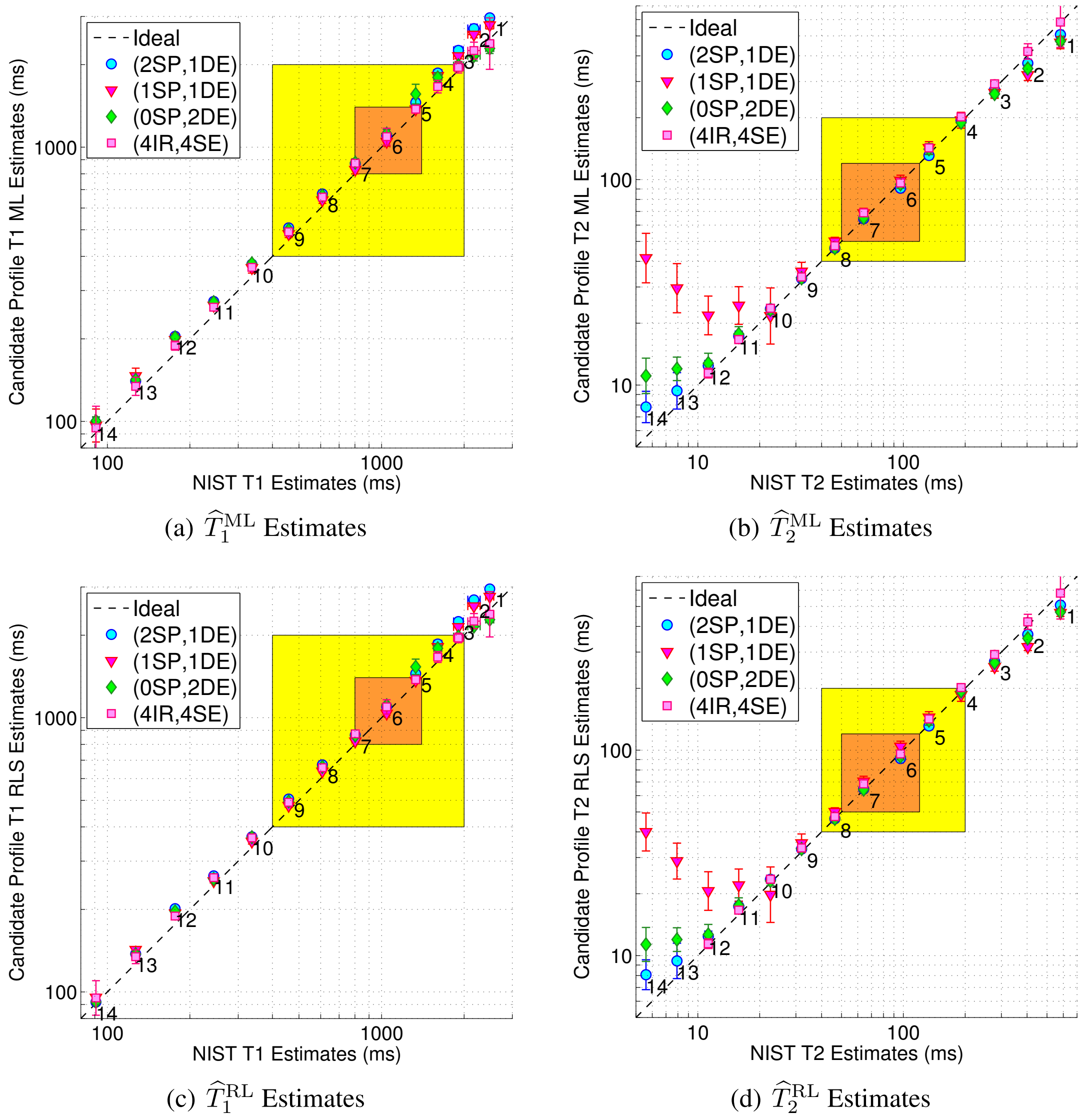


Figure 4.4: Grayscale $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from an HPD® quantitative phantom. Columns correspond to scan profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Vials are enumerated and color-coded to correspond with data points in Fig. 4.5. Fig. 4.3 provides identical colorized images. Colorbar ranges are in milliseconds.

from coil combination and $\mathbf{s}_\mathrm{t}$ estimation, pooled ML sample standard deviations cannot be compared *in magnitude* to worst-case predicted standard deviations (Table 4.1); however, *trends* of empirical and theoretical standard deviations are overall similar. In particular, the optimized $(0, 2)$ DESS-only scan profile affords T_1 ML estimation precision (in vials whose T_1, T_2 is similar to that of WM/GM) comparable to optimized $(2, 1)$ and $(1, 1)$ mixed (SPGR, DESS) profiles. Also in agreement with predictions, the optimized $(2, 1)$ and $(0, 2)$ profiles afford greater T_2 ML estimation precision than the optimized $(1, 1)$ profile.



(a) $\widehat{T}_1^{\mathrm{ML}}$ Estimates

(b) $\widehat{T}_2^{\mathrm{ML}}$ Estimates

(c) $\widehat{T}_1^{\mathrm{RL}}$ Estimates

(d) $\widehat{T}_2^{\mathrm{RL}}$ Estimates

Figure 4.5: Phantom within-ROI sample statistics of $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from optimized SPGR/DESS and reference IR/SE scan profiles, versus NIST NMR measurements [2]. Markers and error bars indicate ROI sample means and ROI sample standard deviations within the 14 labeled and color-coded vials in Fig. 4.4. Tight $\mathbb{X}^{\mathrm{t}}$ and broad $\mathbb{X}^{\mathrm{b}}$ latent parameter ranges are highlighted in orange and yellow, respectively. Table 4.3 replicates sample statistics within Vials 5-8. Our MR measurements are at 293K and NIST NMR measurements are at 293.00K. Within the designed parameter ranges, estimates from different acquisitions are in reasonable agreement with NIST measurements.

	(2SP,1DE)	(1SP,1DE)	(0SP,2DE)	(4IR,4SE)	NIST NMR
V5 $\widehat{T}_1^{\mathrm{ML}}$	$1450 \pm 50.$	1380 ± 41	1600 ± 130	1380 ± 44	1332 ± 0.8
V5 $\widehat{T}_1^{\mathrm{RL}}$	1450 ± 26	1370 ± 16	1540 ± 98	1380 ± 37	
V6 $\widehat{T}_1^{\mathrm{ML}}$	$1100 \pm 30.$	1050 ± 39	1120 ± 39	1100 ± 74	1044 ± 3.2
V6 $\widehat{T}_1^{\mathrm{RL}}$	1100 ± 15	1040 ± 14	1110 ± 16	1100 ± 64	
V7 $\widehat{T}_1^{\mathrm{ML}}$	870 ± 22	830 ± 29	880 ± 29	870 ± 25	801.7 ± 1.70
V7 $\widehat{T}_1^{\mathrm{RL}}$	865 ± 7.1	820 ± 11	860 ± 18	870 ± 21	
V8 $\widehat{T}_1^{\mathrm{ML}}$	680 ± 12	640 ± 18	670 ± 12	658 ± 8.8	608.6 ± 1.03
V8 $\widehat{T}_1^{\mathrm{RL}}$	674 ± 7.6	637 ± 7.4	662 ± 6.6	658 ± 7.1	
V5 $\widehat{T}_2^{\mathrm{ML}}$	131 ± 5.5	$140 \pm 10.$	141 ± 8.4	143 ± 4.9	133.27 ± 0.073
V5 $\widehat{T}_2^{\mathrm{RL}}$	131 ± 5.2	145 ± 9.1	139 ± 7.1	142 ± 4.8	
V6 $\widehat{T}_2^{\mathrm{ML}}$	91 ± 3.5	99 ± 6.0	95 ± 4.2	96 ± 2.7	96.89 ± 0.049
V6 $\widehat{T}_2^{\mathrm{RL}}$	91 ± 3.4	104 ± 6.2	93 ± 3.7	96 ± 2.6	
V7 $\widehat{T}_2^{\mathrm{ML}}$	64 ± 2.2	69 ± 3.9	65 ± 2.1	69 ± 1.2	64.07 ± 0.034
V7 $\widehat{T}_2^{\mathrm{RL}}$	65 ± 2.1	71 ± 4.3	64 ± 1.9	69 ± 1.2	
V8 $\widehat{T}_2^{\mathrm{ML}}$	46 ± 1.5	$50. \pm 2.3$	46 ± 1.1	47.6 ± 0.87	46.42 ± 0.014
V8 $\widehat{T}_2^{\mathrm{RL}}$	46 ± 1.5	$50. \pm 2.3$	46 ± 1.0	47.5 ± 0.85	

Table 4.3: Phantom within-ROI sample means $\pm$ sample standard deviations of $\mathbf{T}_1$ and $\mathbf{T}_2$ estimates from optimized SPGR/DESS and reference IR/SE scan profiles, versus NIST NMR measurements (*cf.* slide 22 of e-poster corresponding to [2]). For sake of brevity, sample statistics corresponding only to phantom vials within (or nearly within) tight design range $\mathbb{X}^{\mathrm{t}}$ (color-coded orange in Fig. 4.4) are reported. Fig. 4.5 plots sample statistics for all vials. 'V#' abbreviates vial numbers. All values are reported in milliseconds.

4.4.3 *In vivo* Experiments

In a single long study of a healthy volunteer, we acquire the same optimized scan profiles containing $(2,1)$, $(1,1)$, and $(0,2)$ SPGR/DESS scans (*cf.* Table 4.1), as well as the reference profile containing $(4,4)$ IR/SE scans. We obtain axial slices from a 32-channel Nova Medical® receive head array. To address bulk motion between acquisitions and to compare within-ROI statistics, we rigidly register[6] each coil-combined image to an IR image prior to parameter mapping. All acquisition and reconstruction details are otherwise the same as in phantom experiments (*cf.* Section 4.4.2.1). For all SS scan profiles tested, ML and RL reconstructions of brain datasets require less than 3m30s and 9s, respectively.

[6]For each coil-combined dataset, we compute a separate 2D rigid transformation (with respect to the $T_{\mathrm{I}} = 50$ms IR dataset) via the MATLAB® function `imregtform` and then apply the transformation via `imwarp`. We choose to use rigid transformations instead of affine distortions to avoid scaling; however in doing so we sacrifice compensating for small through-plane rotations. We do not find registration to substantially change subsequently estimated relaxation maps; however, this extra step substantially improves alignment of (especially cortical GM) ROIs in T_1 and T_2 estimates from different scan profiles.

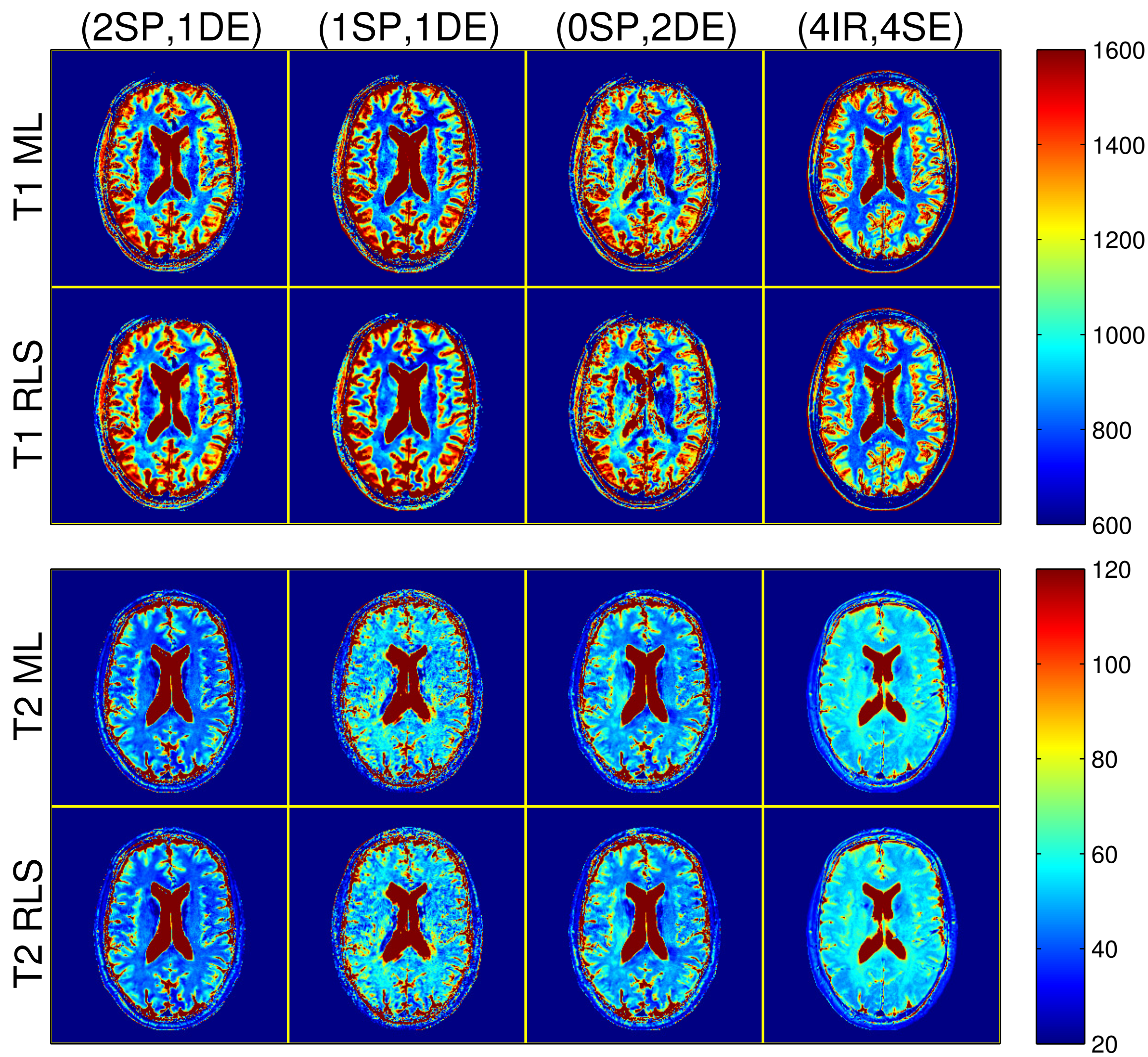


Figure 4.6: Colorized $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from the brain of a healthy volunteer. Columns correspond to profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Table 4.5 presents corresponding WM/GM within-ROI sample statistics. Colorbar ranges are in milliseconds.

	(2SP,1DE)	(1SP,1DE)	(0SP,2DE)
V5 $\widehat{\sigma}_{\widehat{T}_1^{\mathrm{ML}}}$	50 ± 12	$40 \pm 10.$	39 ± 9.4
V6 $\widehat{\sigma}_{\widehat{T}_1^{\mathrm{ML}}}$	70 ± 18	60 ± 15	70 ± 16
V7 $\widehat{\sigma}_{\widehat{T}_1^{\mathrm{ML}}}$	60 ± 13	50 ± 13	50 ± 13
V8 $\widehat{\sigma}_{\widehat{T}_1^{\mathrm{ML}}}$	23 ± 5.4	$20. \pm 4.7$	18 ± 4.3
V5 $\widehat{\sigma}_{\widehat{T}_2^{\mathrm{ML}}}$	2.6 ± 0.63	6 ± 1.4	3.5 ± 0.84
V6 $\widehat{\sigma}_{\widehat{T}_2^{\mathrm{ML}}}$	1.9 ± 0.46	5 ± 1.1	2.3 ± 0.54
V7 $\widehat{\sigma}_{\widehat{T}_2^{\mathrm{ML}}}$	1.4 ± 0.34	3.4 ± 0.80	1.5 ± 0.35
V8 $\widehat{\sigma}_{\widehat{T}_2^{\mathrm{ML}}}$	1.1 ± 0.26	3.5 ± 0.84	1.4 ± 0.33

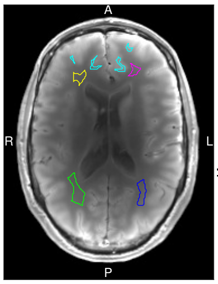
Table 4.4: Phantom pooled sample standard deviations $\pm$ pooled standard errors of sample standard deviations, from optimized SPGR/DESS scan profiles. Each entry is a measure of uncertainty of a typical voxel's T_1 or T_2 ML estimate, estimated over 10 repeated acquisitions. For sake of brevity, sample statistics corresponding only to phantom vials within (or nearly within) tight design range $\mathbb{X}^{\mathrm{t}}$ (color-coded orange in Fig. 4.4) are reported. 'V#' abbreviates vial numbers. All values are reported in milliseconds.

Fig. 4.6 compares brain $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from optimized scan profiles. Though in-plane motion is largely compensated via registration, through-plane motion and non-bulk motion likely persist, and will influence ROI statistics. Due to motion (and scan duration) considerations, we examine within-ROI statistics from a single repetition as in Section 4.4.2.1, and do not attempt across-repetition statistics as in Section 4.4.2.2.

Visually, $\widehat{\mathbf{T}}_1$ maps from steady-state profiles exhibit similar levels of contrast in WM/GM regions well away from cerebrospinal fluid (CSF) as that seen in the reference $\widehat{\mathbf{T}}_1$ estimate. Since we did not optimize any scan profiles for estimation in high-T_1 regions, it is expected that greater differences may emerge in voxels containing or nearby CSF. In particular, $\mathbf{T}_1$ is significantly underestimated within and near CSF by the $(0, 2)$ DESS-only profile. This suggests that with the signal models used in this work, including at least one SPGR scan in an optimized profile may offer greater protection against estimation bias in high-T_1 regions.

Table 4.5 summarizes within-ROI sample means and sample standard deviations compute[7] over four separate WM ROIs containing 96, 69, 224, and 148 voxels and one pooled cortical GM ROI containing 156 voxels. Within-ROI $\widehat{\mathbf{T}}_1$ sample standard deviations are comparable across SS profiles. In agreement with Table 4.1, $\mathbf{T}_2$ estimates from the optimized $(1, 1)$ scan profile exhibit higher within-ROI sample variation than corresponding $(2, 1)$ and $(0, 2)$ $\widehat{\mathbf{T}}_2$ maps. Compared to ML counterparts, RL estimates generally reduce within-ROI sample variation and do not significantly change within-ROI sample means.

[7] We have taken effort to select ROIs that reflect expected anatomy in all coil-combined and registered images, including adjacent slices in images from 3D acquisitions. However, we acknowledge the possibility of some contamination across tissue boundaries, especially WM and/or CSF contamination into cortical GM.



	ROI	(2SP,1DE)	(1SP,1DE)	(0SP,2DE)	(4IR,4SE)
$\widehat{T}_1^{\text{ML}}$	AR WM	840 ± 32	770 ± 31	840 ± 43	780 ± 22
	AL WM	740 ± 61	660 ± 45	740 ± 55	760 ± 24
	PR WM	890 ± 88	860 ± 72	960 ± 84	810 ± 26
	PL WM	$860 \pm 70.$	850 ± 61	880 ± 79	820 ± 37
	A GM	1200 ± 210	1200 ± 230	1300 ± 230	1300 ± 180
$\widehat{T}_1^{\text{RL}}$	AR WM	840 ± 24	$770 \pm 20.$	840 ± 43	$780 \pm 20.$
	AL WM	740 ± 51	670 ± 37	740 ± 54	760 ± 23
	PR WM	890 ± 79	860 ± 61	960 ± 82	810 ± 24
	PL WM	870 ± 62	$850 \pm 50.$	880 ± 78	820 ± 35
	A GM	1200 ± 200	1200 ± 220	1300 ± 230	1300 ± 180
$\widehat{T}_2^{\text{ML}}$	AR WM	$40. \pm 1.3$	54 ± 3.8	46 ± 1.5	55 ± 1.9
	AL WM	$40. \pm 1.7$	$50. \pm 4.5$	44 ± 1.7	53 ± 1.8
	PR WM	43 ± 2.7	$60. \pm 6.9$	51 ± 3.6	59 ± 2.1
	PL WM	43 ± 1.8	57 ± 4.9	49 ± 2.5	57 ± 1.8
	A GM	50 ± 12	60 ± 15	60 ± 11	59 ± 6.0
$\widehat{T}_2^{\text{RL}}$	AR WM	$40. \pm 1.3$	54 ± 3.4	46 ± 1.5	55 ± 1.9
	AL WM	$40. \pm 1.7$	$50. \pm 4.4$	43 ± 1.7	53 ± 1.8
	PR WM	43 ± 2.8	$60. \pm 6.7$	51 ± 3.7	58 ± 2.3
	PL WM	43 ± 1.7	57 ± 4.7	49 ± 2.5	57 ± 1.8
	A GM	50 ± 12	60 ± 15	60 ± 11	59 ± 6.4

Table 4.5: *Left*: WM/GM ROIs, overlaid on a representative anatomical (coil-combined, IR) image. Separate WM ROIs are distinguished by anterior-right (AR), anterior-left (AL), posterior-right (PR), and posterior-left (PL) directions. Four small anterior (A) cortical GM polygons are pooled into a single ROI. *Right*: Within-ROI sample means $\pm$ within-ROI sample standard deviations of $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from the brain of a healthy volunteer (Fig. 4.6 presents corresponding images). Sample statistics are computed within ROIs indicated in the anatomical image. All values are reported in milliseconds.

In most cases, $\widehat{\mathbf{T}}_1$ within-ROI sample means from optimized SPGR/DESS scan profiles do not deviate substantially from each other or from reference IR/SE measurements. Two notable exceptions are $\widehat{T}_1^{\text{ML}}$ in anterior left and posterior right WM from $(1, 1)$ and $(0, 2)$ profiles: these estimates are significantly lower and higher than analogous estimates from other profiles, respectively. Results thus suggest that the optimized $(2, 1)$ scan profile yields WM $\widehat{T}_1^{\text{ML}}$ estimates that are more consistently similar to IR WM $\widehat{T}_1^{\text{ML}}$ estimates than other optimized SPGR/DESS profiles.

Systematic differences in $\widehat{\mathbf{T}}_2$ sample means are evident across scan profiles, particularly within WM ROIs. Curiously, the $(1, 1)$ profile agrees most consistently (in WM/GM $\widehat{T}_2^{\text{ML}}$ within-ROI sample mean) with reference estimates, though with relatively high sample variation. The $(2, 1)$ and $(0, 2)$ SPGR/DESS profiles produce consistently lower WM $\widehat{T}_2^{\text{ML}}$ than the reference IR/SE profile, though the $(0, 2)$ profile is in reasonable agreement with other steady-state estimates [99]. These discrepancies may due to differences in sensitivity to multi-compartmental relaxation [100]. Specifically, different signal models with different scan parameter choices might be more or less sensitive to the model mismatch incurred by neglecting to distinguish the multiple T_2 components within each voxel. Section 4.5 studies $\mathbf{T}_2$ estimation bias due to multi-compartmental relaxation in more detail.

4.5 Sensitivity to Multiple-Compartment Relaxation

This section investigates through simulations and *in vivo* experiments the sensitivity of single-compartment $\mathbf{T}_2$ estimates to model mismatch due to multi-compartment relaxation. The exploratory study in SE data discussed here inspired other similar unreported studies in SPGR/DESS data that contributed to the development of Ch. 6.

We simulate multi-exponential data to arise from three non-exchanging pools of myelinated water $(T_1, T_2) \leftarrow (500, 20)$ms, intracellular and extracellular water $(T_1, T_2) \leftarrow (1000, 80)$ms, and free water $(T_1, T_2) \leftarrow (3500, 250)$ms [100, 101]. We assign pool fractions of $(0.15, 0.80, 0.05)$ in WM and $(0, 0.95, 0.05)$ in GM to the 81st slice of the BrainWeb digital phantom [33, 34] to create ground truth $\mathbf{M}_0$, $\mathbf{T}_1$, and $\mathbf{T}_2$ compartment-wise maps. We simulate component-wise IR signals (acquisition parameters in Subsection 4.4.2.1) and add them to yield noiseless multi-component IR data. We likewise simulate and then add component-wise SE signals to construct three scan profiles consisting of pairs of multi-component SE datasets with variable $\mathbf{T}_\mathrm{E} \in \left\{[10, 30]^\mathsf{T}, [10, 60]^\mathsf{T}, [10, 150]^\mathsf{T}\right\}$. To avoid confounding sources of bias, we assume knowledge of a uniform transmit field and a uniform sensitivity profile of a single-channel receive coil. We estimate a single-component $\widehat{\mathbf{T}}_1$ ML map from multi-component IR data, which we then use to estimate a single-component $\widehat{\mathbf{T}}_2$ ML map from each multi-exponential SE scan profile.

The upper row of Fig. 4.7 compares $\widehat{\mathbf{T}}_2$ maps from simulated multi-exponential SE data. The lower row of Fig. 4.7 compares *in vivo* $\widehat{\mathbf{T}}_2$ maps from corresponding subsets of the SE reference profile discussed in Section 4.4.3. As echo times are further separated, $\widehat{T}_2$ in WM approaches $\widehat{T}_2$ in GM, creating an apparent reduction in $\widehat{T}_2$ WM/GM contrast.

Table 4.6 summarizes $\widehat{T}_2^{\mathrm{ML}}$ sample means and sample standard deviations (computed within WM/GM ROIs depicted in Table 4.5) from *in vivo* SE scan profiles. Single-component $\widehat{T}_2^{\mathrm{ML}}$ estimates depend on SE echo times more significantly in WM than in GM. Comparing with Table 4.5, trends suggest that disagreement in $\widehat{T}_2^{\mathrm{ML}}$ estimates across scan profiles may in part be attributable to the substantial differences of acquisition parameters (*e.g.*, echo times) used in different pulse sequences.

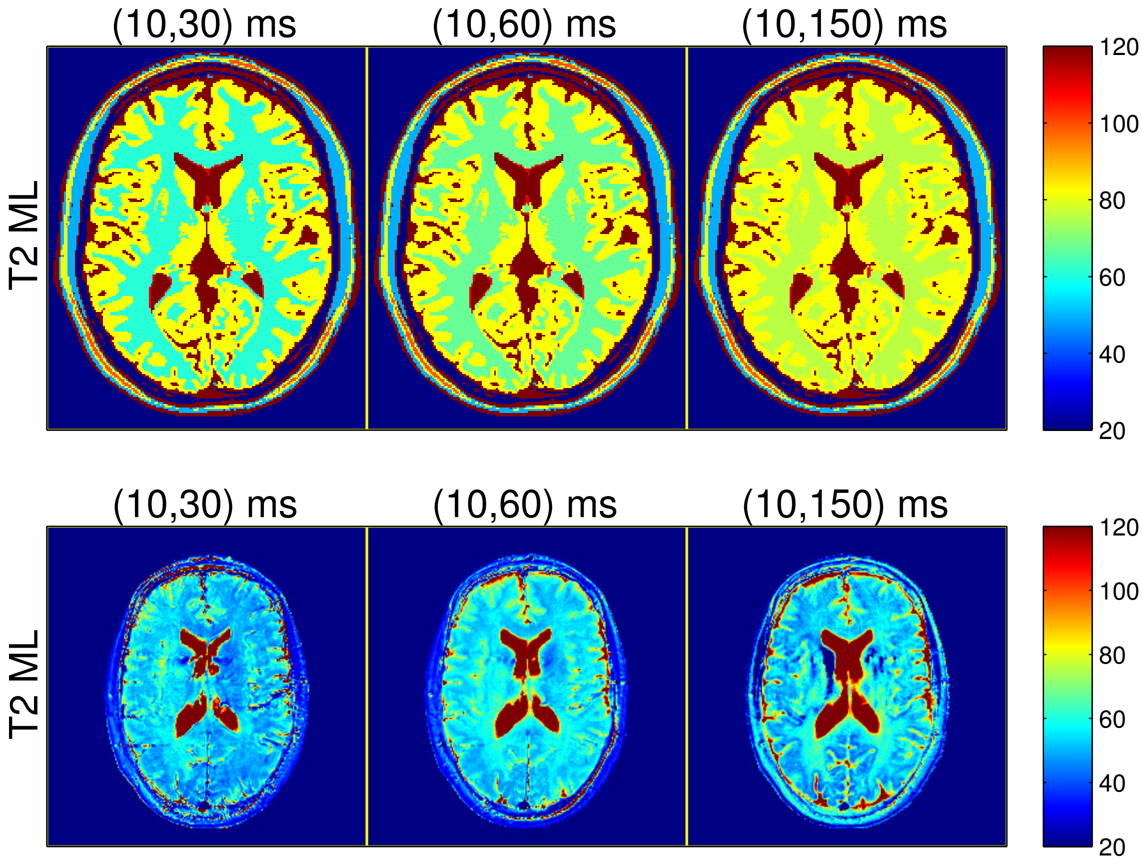


Figure 4.7: Colorized monoexponential $\mathbf{T}_2$ ML estimates from pairs of (top) multi-exponential simulated and (bottom) *in vivo* SE datasets. Columns denote SE dataset echo times. Colorbar ranges are in milliseconds.

4.6 Discussion

Phantom experiments show that optimized scan profiles consisting of $(2,1)$, $(1,1)$, and $(0,2)$ (SPGR, DESS) scans yield accurate WM/GM T_1, T_2 estimates, and that empirical precision trends across profiles agree reasonably with CRB-based predictions. However, *in vivo* experiments reveal that even with scan optimization, it may be challenging to achieve clinically viable levels of precision from the aforementioned SS profiles, at least at 3T. At the expense of greater scan time, it is of course possible that optimized profiles containing greater numbers of SPGR, DESS, and/or other SS scans can provide clinically acceptable precision levels. For these and other more complicated scan profiles, estimator dependence on scan parameters becomes even less intuitive, increasing the need for scan design.

The proposed scan design framework addresses spatial variation in object parameters through a min-max design criterion. The min-max criterion guarantees an upper bound on

	ROI	$(10,30)$	$(10,60)$	$(10,150)$
$\widehat{T}_2^{\mathrm{ML}}$	AR WM	54 ± 3.0	56 ± 1.9	54 ± 2.4
	AL WM	$50. \pm 2.2$	54 ± 1.8	54 ± 2.4
	PR WM	55 ± 2.6	58 ± 2.2	61 ± 2.6
	PL WM	50 ± 2.2	57 ± 2.0	61 ± 2.1
	A GM	58 ± 6.5	61 ± 6.8	57 ± 7.3

Table 4.6: Within-ROI sample means $\pm$ within-ROI sample standard deviations of mono-exponential $\mathbf{T}_2$ ML estimates, from pairs of *in vivo* SE datasets. Column headers indicate echo times $\mathbf{T}_\mathrm{E}$ (ms) of SE datasets. Sample statistics are computed within ROIs indicated in Table 4.5. Single-component $\widehat{T}_2^{\mathrm{ML}}$ estimates in WM depend on SE echo times.

a weighted sum of variances and assumes no prior knowledge of distributions. However, in general it is non-differentiable in $\mathbf{P}$, precluding gradient-based optimization. Furthermore, it is conservative by nature, and often selects scan parameters based on corner cases of the object parameter space. To reduce the influence of corner cases, it may be desirable to instead construct a cost function related to the coefficient of variation as in [79, 77, 80, 81], perhaps by setting parameter weights $\mathbf{W}^{-1} \leftarrow \mathrm{diag}(\mathbf{x})$ for $\mathbf{x} \neq 0$ in (4.5).

As a less conservative alternative to min-max design, other recent works [83, 84] have addressed object parameter spatial variation by instead constructing cost functions related to the Bayesian CRB [102], which characterizes the expected precision with respect to a prior distribution on object parameters. Bayesian cost functions are usually differentiable and can also, with appropriate priors, penalize object parameter coefficients of variation instead of variances, as in [83]. However, prior distributions are generally unknown, and may need to be estimated from data, as in [84].

Careful calibration of flip angle scaling κ^{t} is essential for accurate $\mathbf{T}_1, \mathbf{T}_2$ estimation from SPGR/DESS scan profiles. In this work, we estimate κ^{t} from *separate* acquisitions and adjust nominal flip angles prior to reconstruction, but acknowledge that non-idealities in those separate acquisitions may themselves cause resultant transmit field estimation errors to propagate into our $\mathbf{T}_1, \mathbf{T}_2$ estimates. To reduce error propagation, it may be desirable to instead design scan profiles to permit *joint* estimation of κ^{t}, in addition to other latent object parameters. Unfortunately, we find that optimizing the $(2,1)$ or $(0,2)$ profile to allow for four-parameter $\mathbf{x}(\mathbf{r}) \leftarrow [T_1(\mathbf{r}), T_2(\mathbf{r}), c_2(\mathbf{r}), \kappa^{\mathrm{t}}(\mathbf{r})]^\mathsf{T}$ estimation results in unacceptably high amplification of the worst-case T_1 standard deviation. (Incidentally however, precise $\mathbf{T}_2$ ML and RL estimation alone from the $(2,1)$ or $(0,2)$ profile is possible [5].) It remains an open scan design question as to whether time spent collecting Bloch-Siegert data for separate κ^{t} mapping could instead be better spent collecting additional SPGR, DESS, and/or other data for joint estimation.

By working with closed-form signal expressions, we neglect to model several higher-order effects (*e.g.*, interactions between multiple intravoxel water compartments). However, it is apparent that the nonlinear estimation procedures required for many mapping problems can amplify the influence of these secondary effects, often inducing significant estimation bias. Since the CRB (as described) applies only to unbiased estimators, it is thus desirable to use signal models that are as complete as possible for CRB-based scan design. In theory, scan optimization approach (4.9) is even compatible with acquisitions where a closed-form model relating data to latent and scan parameters is unknown, as in [61, 103]. In practice, difficulties arise in efficient computation of signal gradients required in (4.3), which may demand more specialized techniques, as in [104]. Designing scan profiles involving such complex signal models would likely necessitate optimization techniques more involved than the simple grid searches used in this work.

4.7 Conclusion

This chapter has introduced a CRB-inspired min-max optimization approach to guide MR scan design for precise parameter estimation. As a detailed example, we have optimized combinations of fast SPGR and DESS scans for T_1, T_2 relaxometry in WM and GM regions of the human brain at 3T. Numerical simulations show that at typical noise levels and with accurate flip angle prior knowledge, WM- and GM-like T_1, T_2 ML estimates from optimized scans are nearly unbiased, and so worst-case CRB predictions yield reliable bounds on ROI sample variances. Phantom accuracy experiments show that optimized combinations of $(2, 1)$, $(1, 1)$, or $(0, 2)$ (SPGR, DESS) scans are in excellent agreement with NIST and IR/SE measurements over the designed latent object parameter range of interest. Phantom precision experiments show that these SPGR/DESS combinations exhibit trends in pooled sample standard deviations that reasonably reflect CRB predictions.

In vivo experiments suggest that with optimization, the $(0, 2)$ profile can yield comparable $\widehat{\mathbf{T}}_1, \widehat{\mathbf{T}}_2$ precision to the more conventional $(2, 1)$ [5] scan profile in well-isolated WM/GM ROIs; however, the $(0, 2)$ $\mathbf{T}_1$ estimates are unreliable within and near the CSF and do not agree with IR measurements in WM as consistently as the $(2, 1)$ profile. This and other disagreements across profiles *in vivo* may be attributable to differences in signal model sensitivities to multi-compartmental relaxation. Nevertheless, the example application studied in this chapter illustrates that scan optimization can enable new parameter mapping techniques from established pulse sequences.

CHAPTER 5

QMRI Parameter Estimation via Regression with Kernels (PERK)

5.1 Introduction

This chapter[1] reconsiders the problem of QMRI parameter estimation (first introduced in Chapter 3), where one seeks to estimate latent parameter images from suitably informative data. Chapter 3 applied a common parameter estimation strategy to QMRI that involves minimizing a cost function related to a statistical likelihood function. Because MR signal models are typically nonlinear functions of the underlying latent parameters, such likelihood-based estimation usually requires non-convex optimization. To seek good solutions, Chapter 3 as well as many other works (*e.g.*, [105, 106, 95, 107, 103, 108, 109, 110, 61, 111, 112, 113, 6, 114, 115]) approach estimation with algorithms that employ exhaustive grid search, which requires either storing or computing on-the-fly a "dictionary" of signal vectors. These works estimate a small number (2-3) of nonlinear latent parameters, so grid search is practical. However, for moderate or large sized problems, the required number of dictionary elements renders grid search undesirable or even intractable, unless one imposes artificially restrictive latent parameter constraints. Though several recent works [109, 112, 114, 115] focus on reducing dictionary storage requirements, all of these methods ultimately rely on some form of dictionary-based grid search.

There are numerous QMRI applications that could benefit from an alternative parameter estimation method that scales well with the number of latent parameters. For example, vector (*e.g.*, flow [116]) and tensor (*e.g.*, diffusivity [117] or conductivity [118]) field mapping techniques require estimation of at minimum 4 and 7 latent parameters per voxel, respectively. Phase-based longitudinal [119] or transverse [120, 90] field mapping could avoid

[1] This chapter mainly derives content from accepted journal paper [8] that extends two conference papers [9, 1].

noise-amplifying algebraic manipulations on reconstructed image data that are conventionally used to reduce signal dependencies on nuisance latent parameters. Compartmental fraction mapping [100, 11] from steady-state pulse sequences requires estimation of at least 7 [121] and as many as 10 [101] latent parameters per voxel. In these and other applications, greater estimation accuracy requires more complete signal models that involve more latent parameters, increasing the need for scalable estimation methods.

The fundamental challenge of scalable MRI parameter estimation stems from MR signal model nonlinearity: standard linear estimators would be scalable but inaccurate. One natural solution strategy involves nonlinearly preprocessing reconstructed images such that the transformed images are at least approximately linear in the latent parameters. As an example, for simple T_2 estimation from measurements at multiple echo times, one could apply linear regression to the logarithm of the measurements (Subsection 5.6 builds further intuition using this simple application). However, such simple transformations are generally not evident for more complicated signal models. Without such problem-specific insight, sufficiently rich nonlinear transformations could dramatically increase problem dimensionality, hindering scalability. Fortunately, a celebrated result in approximation theory [122] showed that simple transformations involving *reproducing kernel* functions [123] can represent nonlinear estimators whose evaluation need not directly scale in computation with the (possibly very high) dimension of the associated transformed data. These kernel methods later found popularity in machine learning (initially for classification [124] and quickly thereafter for other applications, *e.g.*, regression [125]) because they provided simple, scalable nonlinear extensions to fast linear algorithms.

The general idea of using linearization to simplify a nonlinear estimation problem has been used before in QMRI. For example, orthogonal transforms have been used to linearly represent exponential [126] and extended phase graph [127] models for T_2 estimation. An unscented Kalman filter has been used to linearly represent nonlinear models for general multiple-parameter estimation up to third-order accuracy [128]. Whereas these prior works largely focus on parameter estimation accuracy gains in under-sampled acquisitions, this paper focuses on acceleration for general per-voxel MRI parameter estimation from reconstructed images.

This chapter introduces a fast dictionary-free method for MRI parameter estimation via regression with kernels (PERK). PERK first simulates many instances of latent parameter inputs and measurement outputs using prior distributions and a general nonlinear MR signal model. PERK takes such input-output pairs as simulated *training points* and then *learns* (using an appropriate nonlinear kernel function) a nonlinear *regression function* from the training points. PERK may scale considerably better with the number of latent parameters than likelihood-based estimation via grid search.

The remainder of this chapter is organized as follows. Section 5.2 reviews pertinent background information about kernels. Section 5.3 formulates a function optimization problem for MRI parameter estimation and efficiently solves this problem using kernels. Section 5.4 studies bias and covariance of the resulting PERK estimator. Section 5.5 addresses practical implementation issues such as computational complexity and model selection. Section 5.6 provides intuition into PERK through a simple toy problem. Section 5.7 demonstrates PERK in numerical simulations as well as phantom and *in vivo* experiments. Section 5.8 investigates PERK robustness to two types of non-idealities. Section 5.9 discusses advantages, challenges, and extensions. Section 5.10 summarizes key contributions.

5.2 Preliminaries

This brief section reviews relevant definitions and facts about kernels. A (real-valued) *kernel* $k : \mathbb{P}^2 \mapsto \mathbb{R}$ is a function that describes a measure of similarity between two pattern vectors $\mathbf{q}, \mathbf{q}' \in \mathbb{P}$. The matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$ associated with kernel k and $N \in \mathbb{N}$ patterns $\mathbf{q}_1, \ldots, \mathbf{q}_N \in \mathbb{P}$ consists of entries $k(\mathbf{q}_n, \mathbf{q}_{n'})$ for $n, n' \in \{1, \ldots, N\}$. A *positive definite kernel* is a kernel for which $\mathbf{K}$ is positive semidefinite (PSD) for any finite set of pattern vectors, in which case $\mathbf{K}$ is a *Gram matrix*. A *symmetric kernel* satisfies $k(\mathbf{q}, \mathbf{q}') = k(\mathbf{q}', \mathbf{q})$ $\forall \mathbf{q}, \mathbf{q}' \in \mathbb{P}$. We hereafter restrict attention to symmetric, positive definite (SPD) kernels.

An SPD kernel $k : \mathbb{P}^2 \mapsto \mathbb{R}$ defines an inner product in a particular Hilbert function space $\bar{\mathbb{H}}$ that we briefly describe here because it characterizes the class of candidate regression functions over which PERK operates. To envision $\bar{\mathbb{H}}$, first define a kernel's associated *(canonical) feature map* $\mathbf{z} : \mathbb{P} \mapsto \mathbb{R}^{\mathbb{P}}$ that assigns each $\mathbf{q} \in \mathbb{P}$ to a *(canonical) feature* $k(\cdot, \mathbf{q}) \in \mathbb{R}^{\mathbb{P}}$. Then $\bar{\mathbb{H}}$ is a completion of the space $\mathbb{H} := \left\{\sum_{n=1}^N a_n k(\cdot, \mathbf{q}_n)\right\}$ spanned by point evaluations of the feature map, where $N \in \mathbb{N}$, $a_1, \ldots, a_N \in \mathbb{R}$, and $\mathbf{q}_1, \ldots, \mathbf{q}_N \in \mathbb{P}$ are arbitrary. Let $\langle \cdot, \cdot \rangle : \bar{\mathbb{H}}^2 \mapsto \mathbb{R}$ denote the inner product on $\bar{\mathbb{H}}$. Then for any $h, h' \in \mathbb{H}$ that have finite-dimensional canonical representations $h := \sum_{n=1}^N a_n k(\cdot, \mathbf{q}_n)$ and $h' := \sum_{n'=1}^N b_{n'} k(\cdot, \mathbf{q}_{n'})$, the assignment

$$\langle h, h' \rangle_{\bar{\mathbb{H}}} = \sum_{n=1}^N \sum_{n'=1}^N a_n b_{n'} k(\mathbf{q}_{n'}, \mathbf{q}_n) \tag{5.1}$$

is consistent with the inner product on $\bar{\mathbb{H}}$. This inner product exhibits $\forall h \in \bar{\mathbb{H}}, \mathbf{q} \in \mathbb{P}$ an interesting *reproducing property*

$$\langle h, k(\cdot, \mathbf{q})\rangle_{\bar{\mathbb{H}}} = h(\mathbf{q}) \tag{5.2}$$

that can be seen to directly follow from (5.1) for $h \in \mathbb{H}$.

A *reproducing kernel* (RK) is a kernel that satisfies (5.2) for some real-valued Hilbert space $\bar{\mathbb{H}}$. A kernel is reproducing if and only if it is SPD. There is a bijection between RK k and $\bar{\mathbb{H}}$, and so $\bar{\mathbb{H}}$ is often called the *reproducing kernel Hilbert space* (RKHS) uniquely associated with RK k. This bijection is critical to practical function optimization over an RKHS in that it translates inner products in a (usually high-dimensional) RKHS $\bar{\mathbb{H}}$ into equivalent kernel operations in the (lower-dimensional) pattern vector space $\mathbb{P}$. The following sections exploit the bijection between an RKHS and its associated RK.

5.3 A Function Optimization Problem & Kernel Solution

Recall from Section 3.2.1 that after image reconstruction, many QMRI acquisitions produce at each voxel position a sequence of noisy measurements $\mathbf{y} \in \mathbb{C}^D$, modeled as

$$\mathbf{y} = \mathbf{s}(\mathbf{x}, \boldsymbol{\nu}) + \boldsymbol{\epsilon}, \tag{5.3}$$

where $\mathbf{x} \in \mathbb{R}^L$ denotes L *latent* parameters; $\boldsymbol{\nu} \in \mathbb{R}^K$ denotes K *known* parameters; $\mathbf{s} : \mathbb{R}^L \times \mathbb{R}^K \mapsto \mathbb{C}^D$ models D noiseless continuous signal functions; and $\boldsymbol{\epsilon} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_D, \boldsymbol{\Sigma})$ is complex Gaussian noise with zero mean $\mathbf{0}_D \in \mathbb{R}^D$ and known covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$. (As a concrete example, for T_2 estimation from spin echo measurements, $\mathbf{x}$ could collect spin density and T_2; $\boldsymbol{\nu}$ could collect known longitudinal and transverse field inhomogeneities; and $\mathbf{y}$ could collect measurements at D echo times.) We seek to estimate on a per-voxel basis each latent parameter $\mathbf{x}$ from measurement $\mathbf{y}$ and known parameter $\boldsymbol{\nu}$.

To develop an estimator $\widehat{\mathbf{x}}$, we simulate many instances of forward model (5.3) and use kernels to estimate a nonlinear inverse function. We sample part of $\mathbb{R}^L \times \mathbb{R}^K \times \mathbb{C}^D$ and evaluate (5.3) N times to produce sets of object parameter and noise realizations $\{(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \dots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)\}$ and corresponding measurements $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$. We seek a function $\widehat{\mathbf{h}} : \mathbb{R}^Q \mapsto \mathbb{R}^L$ and an offset $\widehat{\mathbf{b}} \in \mathbb{R}^L$ that together map each pure-real[2] regressor $\mathbf{q}_n := [|\mathbf{y}_n|^\mathsf{T}, \boldsymbol{\nu}_n^\mathsf{T}]^\mathsf{T}$ to an estimate $\widehat{\mathbf{x}}(\mathbf{q}_n) := \widehat{\mathbf{h}}(\mathbf{q}_n) + \widehat{\mathbf{b}}$ that is "close" to corresponding

[2]We present our methodology assuming pure-real patterns $\mathbf{q}$ and estimators $\widehat{\mathbf{x}}$ for simplicity and to maintain consistency with experiments, in which we choose to use magnitude images for unrelated reasons (see Subsection 5.7.1 for details). It is straightforward to generalize Theorem 1 for complex-valued kernels and thereby address the cases of complex patterns and/or estimators.

regressand $\mathbf{x}_n$, where $Q := D + K$, $n \in \{1, \ldots, N\}$, and $(\cdot)^\mathsf{T}$ denotes vector transpose. For any finite N, there are infinitely many candidate estimators that are consistent with training points in this manner. We use function regularization to choose one estimator that smoothly interpolates between training points:

$$\left(\widehat{\mathbf{h}}, \widehat{\mathbf{b}}\right) \in \arg \min_{\substack{\mathbf{h} \in \bar{\mathbb{H}}^L \\ \mathbf{b} \in \mathbb{R}^L}} \Psi\left(\mathbf{h}, \mathbf{b}; \{(\mathbf{x}_n, \mathbf{q}_n)\}_1^N\right), \text{ where} \tag{5.4}$$

$$\Psi\left(\mathbf{h}, \mathbf{b}; \{(\mathbf{x}_n, \mathbf{q}_n)\}_1^N\right) = \sum_{l=1}^{L} \Psi_l\left(h_l, b_l; \{(x_{l,n}, \mathbf{q}_n)\}_1^N\right); \tag{5.5}$$

$$\Psi_l(h_l, b_l; \{(x_{l,n}, \mathbf{q}_n)\}_1^N) = \rho_l \|h_l\|_{\bar{\mathbb{H}}}^2 + \frac{1}{N} \sum_{n=1}^{N} (h_l(\mathbf{q}_n) + b_l - x_{l,n})^2. \tag{5.6}$$

Here, each $h_l : \mathbb{R}^Q \mapsto \mathbb{R}$ is a scalar function that maps to the lth component of the output of $\mathbf{h}$; each $b_l, x_{l,n} \in \mathbb{R}$ are scalar components of $\mathbf{b}, \mathbf{x}_n$; $\bar{\mathbb{H}}$ is an RKHS whose norm $\|\cdot\|_{\bar{\mathbb{H}}}$ is induced by inner product $\langle \cdot, \cdot \rangle_{\bar{\mathbb{H}}} : \bar{\mathbb{H}}^2 \mapsto \mathbb{R}$; and each ρ_l controls for regularity in h_l.

Since (5.5) is separable in the components of $\mathbf{h}$ and $\mathbf{b}$, it suffices to consider optimizing each (h_l, b_l) by separately minimizing (5.6) for each $l \in \{1, \ldots, L\}$. Remarkably, a generalization of the Representer Theorem [129], restated as is relevant here for completeness, reduces minimizing (5.6) to a finite-dimensional optimization problem.

Theorem 1 (Generalized Representer, [129]) *Define $k : \mathbb{R}^Q \times \mathbb{R}^Q \mapsto \mathbb{R}$ to be the SPD kernel associated with RKHS $\bar{\mathbb{H}}$, such that reproducing property $h_l(\mathbf{q}) = \langle h_l, k(\cdot, \mathbf{q}) \rangle_{\bar{\mathbb{H}}}$ holds for all $h_l \in \bar{\mathbb{H}}$ and $\mathbf{q} \in \mathbb{R}^Q$. Then any minimizer $(\widehat{h}_l, \widehat{b}_l)$ of (5.6) over $\bar{\mathbb{H}} \times \mathbb{R}$ admits a representation for $\widehat{h}_l$ of the form*

$$\widehat{h}_l(\cdot) \equiv \sum_{n=1}^{N} a_{l,n} k(\cdot, \mathbf{q}_n), \tag{5.7}$$

where each $a_{l,n} \in \mathbb{R}$ for $n \in \{1, \ldots, N\}$.

Thm. 1 ensures that any solution to the component-wise $(N+1)$-dimensional problem

$$\begin{aligned}(\widehat{\mathbf{a}}_l, \widehat{b}_l) \in \arg \min_{\substack{\mathbf{a}_l \in \mathbb{R}^N \\ b_l \in \mathbb{R}}} \rho_l \left\| \sum_{n'=1}^{N} a_{l,n'} k(\cdot, \mathbf{q}_{n'}) \right\|_{\bar{\mathbb{H}}}^2 + \\ \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{n'=1}^{N} a_{l,n'} k(\mathbf{q}_n, \mathbf{q}_{n'}) + b_l - x_{l,n} \right)^2\end{aligned} \tag{5.8}$$

corresponds via (5.7) to a minimizer of (5.6) over $\bar{\mathbb{H}} \times \mathbb{R}$, where $\mathbf{a}_l := [a_{l,1}, \ldots, a_{l,N}]^\mathsf{T}$.

Fortunately, a solution of (5.8) exists uniquely for $\rho_l > 0$ and can be expressed as

$$\widehat{\mathbf{a}}_l = ((\mathbf{MKM} + N\rho_l \mathbf{I}_N))^{-1} \mathbf{M}\boldsymbol{x}_l; \tag{5.9}$$

$$\widehat{b}_l = \frac{1}{N} \mathbf{1}_N^\mathsf{T} (\boldsymbol{x}_l - \mathbf{K}\widehat{\mathbf{a}}_l), \tag{5.10}$$

where $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the Gram matrix consisting of entries $k(\mathbf{q}_n, \mathbf{q}_{n'})$ for $n, n' \in \{1, \dots, N\}$; $\mathbf{M} := \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N \mathbf{1}_N^\mathsf{T} \in \mathbb{R}^{N \times N}$ is a de-meaning operator; $\boldsymbol{x}_l := [x_{l,1}, \dots, x_{l,N}]^\mathsf{T}$; $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix; and $\mathbf{1}_N \in \mathbb{R}^N$ is a vector of ones. Substituting (5.9) into (5.7) yields an expression for the lth entry $\widehat{x}_l$ of MRI parameter estimator $\widehat{\mathbf{x}}$:

$$\widehat{x}_l(\cdot) \leftarrow \boldsymbol{x}_l^\mathsf{T} \left(\frac{1}{N} \mathbf{1}_N + \mathbf{M}((\mathbf{MKM} + N\rho_l \mathbf{I}_N))^{-1} \mathbf{k}(\cdot) \right), \tag{5.11}$$

where $\mathbf{k}(\cdot) := [k(\cdot, \mathbf{q}_1), \dots, k(\cdot, \mathbf{q}_N)]^\mathsf{T} - \frac{1}{N}\mathbf{K}\mathbf{1}_N : \mathbb{R}^Q \mapsto \mathbb{R}^N$ is an embedding operator.

When $\rho_l > 0 \; \forall l \in \{1, \dots, L\}$, estimator $\widehat{\mathbf{x}}(\cdot)$ with entries (5.11) minimizes (5.5) over $\bar{\mathbb{H}}^L \times \mathbb{R}^L$. However, the utility of $\widehat{\mathbf{x}}(\cdot)$ depends on the choice of kernel k, which induces a choice on the RKHS $\bar{\mathbb{H}}$ and thus the function space $\bar{\mathbb{H}}^L \times \mathbb{R}^L$ over which (5.4) optimizes. For example, if k was selected as the canonical dot product $k(\mathbf{q}, \mathbf{q}') \leftarrow \langle \mathbf{q}, \mathbf{q}' \rangle_{\mathbb{R}^Q} := \mathbf{q}^\mathsf{T}\mathbf{q}'$ (for which RKHS $\bar{\mathbb{H}} \leftarrow \mathbb{R}^Q$), then (5.11) would reduce to affine ridge regression [130] which is optimal over $\mathbb{R}^Q \times \mathbb{R}$ but is unlikely to be useful when signal model $\mathbf{s}$ is nonlinear in $\mathbf{x}$. Since we expect a useful estimate $\widehat{\mathbf{x}}(\mathbf{q})$ to depend nonlinearly (but smoothly) on $\mathbf{q}$ in general, we instead use an SPD kernel k that is likewise nonlinear in its arguments and thus corresponds to an RKHS much richer than $\mathbb{R}^Q$. Specifically, we use a Gaussian kernel

$$k(\mathbf{q}, \mathbf{q}') \leftarrow \exp\left(-\frac{1}{2} \|\mathbf{q} - \mathbf{q}'\|^2_{\mathbf{\Lambda}^{-2}} \right), \tag{5.12}$$

where positive definite matrix bandwidth $\mathbf{\Lambda} \in \mathbb{R}^{Q \times Q}$ controls the length scales in $\mathbf{q}$ over which the estimator $\widehat{\mathbf{x}}$ smooths and $\|\cdot\|_{\mathbf{\Gamma}} \equiv \left\| \mathbf{\Gamma}^{1/2}(\cdot) \right\|_2$ is a weighted ℓ^2-norm with PSD matrix weights $\mathbf{\Gamma}$. We use a Gaussian kernel over other candidates because it is a *universal kernel*, meaning weighted sums of the form $\sum_{n=1}^N a_n k(\cdot, \mathbf{q}_n)$ can approximate $\mathcal{L}^2$ functions to arbitrary accuracy for N sufficiently large [131].

Interestingly, the RKHS associated with Gaussian kernel (5.12) is infinite-dimensional. Thus, Gaussian kernel regression can be interpreted as first "lifting" via a nonlinear *feature map* $\mathbf{z} : \mathbb{R}^Q \mapsto \bar{\mathbb{H}}$ each $\mathbf{q}$ into an infinite-dimensional *feature* $\mathbf{z}(\mathbf{q}) = k(\cdot, \mathbf{q}) \in \bar{\mathbb{H}}$, and then performing regularized affine regression on the features via dot products of the form $\langle k(\cdot, \mathbf{q}), k(\cdot, \mathbf{q}') \rangle_{\bar{\mathbb{H}}} = k(\mathbf{q}', \mathbf{q})$. From this perspective, the challenges of nonlinear estimation via likelihood models are avoided because we *select* (through the choice of kernel) characteristics of the nonlinear dependence that we wish to model and need only *estimate* via (5.8) the linear dependence of each entry in $\widehat{\mathbf{x}}$ on the corresponding features.

5.4 Bias and Covariance Analysis

This section presents expressions for the bias and covariance of Gaussian PERK estimator $\widehat{\mathbf{x}}(\cdot)$, conditioned on object parameters $\mathbf{x}, \boldsymbol{\nu}$. We focus on these conditional statistics to enable study of estimator performance as $\mathbf{x}, \boldsymbol{\nu}$ are varied. Though not mentioned explicitly hereafter, both expressions treat the training sample $\{(\mathbf{x}_1, \mathbf{q}_1), \ldots, (\mathbf{x}_N, \mathbf{q}_N)\}$ and regularization parameters $\rho_1, \ldots, \rho_L$ as fixed.

5.4.1 Conditional Bias

The conditional bias of $\widehat{\mathbf{x}} \equiv \widehat{\mathbf{x}}(\boldsymbol{\alpha}, \boldsymbol{\nu})$ is written as

$$\begin{aligned}\mathrm{bias}(\widehat{\mathbf{x}}|\mathbf{x}, \boldsymbol{\nu}) &:= \mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\widehat{\mathbf{x}}(\boldsymbol{\alpha}, \boldsymbol{\nu})) - \mathbf{x} \\ &= \mathbf{R}\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\mathbf{k}(\boldsymbol{\alpha}, \boldsymbol{\nu})) + (\mathbf{m}_{\mathbf{x}} - \mathbf{x}),\end{aligned} \tag{5.13}$$

where $\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\cdot)$ denotes expectation with respect to $\boldsymbol{\alpha} := |\mathbf{y}|$ and conditioned on $\mathbf{x}, \boldsymbol{\nu}$. Here, the lth row of $\mathbf{R} \in \mathbb{R}^{L \times N}$ and lth entry of regressand sample mean $\mathbf{m}_{\mathbf{x}} \in \mathbb{R}^L$ respectively are $\tilde{\mathbf{x}}l^{\mathsf{T}}\mathbf{M}((\mathbf{MKM} + N\rho_l\mathbf{I}_N))^{-1}$ and $\frac{1}{N}\tilde{\mathbf{x}}l^{\mathsf{T}}\mathbf{1}_N$ for $l \in \{1, \ldots, L\}$. To proceed analytically, we make two mild assumptions. First, we assume that $\mathbf{y} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_D, \boldsymbol{\Sigma})$ has sufficiently high signal-to-noise ratio (SNR) such that its complex modulus $\boldsymbol{\alpha}$ is approximately Gaussian-distributed. We specifically consider the typical case where covariance matrix $\boldsymbol{\Sigma}$ is diagonal with diagonal entries $\sigma_1^2, \ldots, \sigma_D^2$, in which case measurement amplitude conditional distribution $\mathsf{p}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}$ is simply approximated as $\mathsf{p}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}} \leftarrow \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^D$ has dth coordinate $\sqrt{|s_d(\mathbf{x}, \boldsymbol{\nu})|^2 + \sigma_d^2}$ for $d \in \{1, \ldots, D\}$ [30]. Second, we assume that the Gaussian kernel bandwidth matrix $\boldsymbol{\Lambda}$ has the block diagonal structure

$$\boldsymbol{\Lambda} \leftarrow \begin{bmatrix} \boldsymbol{\Lambda}_{\boldsymbol{\alpha}} & \mathbf{0}_{D \times K} \\ \mathbf{0}_{K \times D} & \boldsymbol{\Lambda}_{\boldsymbol{\nu}}. \end{bmatrix} \tag{5.14}$$

where $\mathbf{\Lambda}_{\boldsymbol{\alpha}} \in \mathbb{R}^{D\times D}$ and $\mathbf{\Lambda}_{\boldsymbol{\nu}} \in \mathbb{R}^{K\times K}$ are positive definite. With these simplifying assumptions, the nth entry of the expectation in (5.13) is well approximated as

$$\begin{aligned}
\left[\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\mathbf{k}(\boldsymbol{\alpha},\boldsymbol{\nu}))\right]_n &= \int_{\mathbb{R}^D} e^{-\frac{1}{2}\|\mathbf{q}-\mathbf{q}_n\|^2_{\mathbf{\Lambda}^{-2}}}\,\mathsf{p}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu})\,\mathrm{d}\,\boldsymbol{\alpha} \\
&\approx \frac{e^{-\frac{1}{2}\|\boldsymbol{\nu}-\boldsymbol{\nu}_n\|^2_{\mathbf{\Lambda}_{\boldsymbol{\nu}}^{-2}}}}{\sqrt{(2\pi)^D\det(\mathbf{\Sigma})}}\int_{\mathbb{R}^D} e^{-\frac{1}{2}\left(\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_n\|^2_{\mathbf{\Lambda}_{\boldsymbol{\alpha}}^{-2}}+\|\boldsymbol{\alpha}-\boldsymbol{\mu}\|^2_{\mathbf{\Sigma}^{-1}}\right)}\,\mathrm{d}\,\boldsymbol{\alpha} \\
&= \frac{e^{-\frac{1}{2}\left(\|\boldsymbol{\nu}-\boldsymbol{\nu}_n\|^2_{\mathbf{\Lambda}_{\boldsymbol{\nu}}^{-2}}+\|\boldsymbol{\mu}-\boldsymbol{\alpha}_n\|^2_{\left(\mathbf{\Lambda}_{\boldsymbol{\alpha}}^{-2}\mathbf{\Sigma}+\mathbf{I}_D\right)^{-1}\mathbf{\Lambda}_{\boldsymbol{\alpha}}^{-2}}\right)}}{\sqrt{\det\left(\mathbf{\Lambda}_{\boldsymbol{\alpha}}^{-2}\mathbf{\Sigma}+\mathbf{I}_D\right)}},
\end{aligned} \tag{5.15}$$

where $\det(\cdot)$ denotes determinant and the Gaussian integral follows after completing the square of the integrand's exponent. It is clear from (5.15) that as $\mathbf{\Sigma} \to \mathbf{0}_{D\times D}$ for fixed $\mathbf{\Lambda}_{\boldsymbol{\alpha}}$, $\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\mathbf{k}(\boldsymbol{\alpha},\boldsymbol{\nu})) \to \mathbf{k}(\boldsymbol{\mu},\boldsymbol{\nu})$ and therefore

$$\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\widehat{\mathbf{x}}(\boldsymbol{\alpha},\boldsymbol{\nu})) \to \widehat{\mathbf{x}}\big(\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\boldsymbol{\alpha}),\boldsymbol{\nu}\big) \equiv \widehat{\mathbf{x}}(\boldsymbol{\mu},\boldsymbol{\nu}) \tag{5.16}$$

which perhaps surprisingly means that the conditional bias asymptotically approaches the noiseless conditional estimation error $\widehat{\mathbf{x}}(\boldsymbol{\mu},\boldsymbol{\nu}) - \mathbf{x}$ despite $\widehat{\mathbf{x}}$ being nonlinear in $\boldsymbol{\alpha}$.

5.4.2 Conditional Covariance

The conditional covariance of $\widehat{\mathbf{x}} \equiv \widehat{\mathbf{x}}(\boldsymbol{\alpha},\boldsymbol{\nu})$ is written as

$$\begin{aligned}
\mathrm{cov}(\widehat{\mathbf{x}}|\mathbf{x},\boldsymbol{\nu}) &:= \mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}\Big(\big(\widehat{\mathbf{x}}-\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\widehat{\mathbf{x}})\big)\big(\widehat{\mathbf{x}}-\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\widehat{\mathbf{x}})\big)^{\mathsf{T}}\Big) \\
&= \mathbf{R}\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}\Big(\widetilde{\mathbf{k}}(\boldsymbol{\alpha},\boldsymbol{\nu})\widetilde{\mathbf{k}}(\boldsymbol{\alpha},\boldsymbol{\nu})^{\mathsf{T}}\Big)\mathbf{R}^{\mathsf{T}},
\end{aligned} \tag{5.17}$$

where $\widetilde{\mathbf{k}}(\boldsymbol{\alpha},\boldsymbol{\nu}) := \mathbf{k}(\boldsymbol{\alpha},\boldsymbol{\nu}) - \mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\mathbf{k}(\boldsymbol{\alpha},\boldsymbol{\nu}))$. To proceed analytically, we take the same high-SNR and block-diagonal bandwidth assumptions as in Subsection 5.4.1. Then after straightforward manipulations similar to those yielding (5.15), the (n,n')th entry of the expectation in (5.17) is well approximated as

$$\begin{aligned}
&\left[\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}\Big(\widetilde{\mathbf{k}}(\boldsymbol{\alpha},\boldsymbol{\nu})\widetilde{\mathbf{k}}(\boldsymbol{\alpha},\boldsymbol{\nu})^{\mathsf{T}}\Big)\right]_{n,n'} = e^{-\frac{1}{2}\left(\|\boldsymbol{\nu}-\boldsymbol{\nu}_n\|^2_{\mathbf{\Lambda}_{\boldsymbol{\nu}}^{-2}}+\|\boldsymbol{\nu}-\boldsymbol{\nu}_{n'}\|^2_{\mathbf{\Lambda}_{\boldsymbol{\nu}}^{-2}}\right)} \\
&\times\left(\frac{e^{-\frac{1}{2}\left(\|\widetilde{\boldsymbol{\alpha}}_n-\widetilde{\boldsymbol{\alpha}}_{n'}\|^2_{\mathbf{\Delta}(0)}+\|\widetilde{\boldsymbol{\alpha}}_n+\widetilde{\boldsymbol{\alpha}}_{n'}\|^2_{\mathbf{\Delta}(2)}\right)}}{\sqrt{\det\left(2\mathbf{\Lambda}_{\boldsymbol{\alpha}}^{-2}\mathbf{\Sigma}+\mathbf{I}_D\right)}} - \frac{e^{-\frac{1}{2}\left(\|\widetilde{\boldsymbol{\alpha}}_n-\widetilde{\boldsymbol{\alpha}}_{n'}\|^2_{\mathbf{\Delta}(1)}+\|\widetilde{\boldsymbol{\alpha}}_n+\widetilde{\boldsymbol{\alpha}}_{n'}\|^2_{\mathbf{\Delta}(1)}\right)}}{\det\left(\mathbf{\Lambda}_{\boldsymbol{\alpha}}^{-2}\mathbf{\Sigma}+\mathbf{I}_D\right)}\right),
\end{aligned} \tag{5.18}$$

where $\widetilde{\boldsymbol{\alpha}}_n := \boldsymbol{\mu} - \boldsymbol{\alpha}_n$ and $\boldsymbol{\Delta}(t) := \frac{1}{2}\big(\big(t\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^{-2}\boldsymbol{\Sigma} + \mathbf{I}_D\big)\big)^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^{-2}$ for $t \in \mathbb{N}$. The emergence of $\widetilde{\boldsymbol{\alpha}}_n \pm \widetilde{\boldsymbol{\alpha}}_{n'}$ terms in (5.18) show that the conditional covariance (unlike the conditional bias) is directly influenced not only by the individual expected test point distances to each of the training points $\widetilde{\boldsymbol{\alpha}}_1, \ldots, \widetilde{\boldsymbol{\alpha}}_N$ but also by the local training point sampling density.

5.5 Implementation Considerations

This section focuses on important practical implementation issues. Subsection 5.5.1 discusses a conceptually intuitive approximation of PERK estimator (5.11) that in many problems can significantly improve computational performance. Subsection 5.5.2 describes strategies for data-driven model selection.

5.5.1 A Kernel Approximation

In practical problems with even moderately large ambient dimension Q, the necessarily large number of training samples N complicates storage of (dense) $N \times N$ Gram matrix $\mathbf{K}$. Using a kernel approximation can mitigate storage and processing issues. Here we employ *random Fourier features* [132], a recent method for approximating translation-invariant kernels having form $k(\mathbf{q}, \mathbf{q}') \equiv k(\mathbf{q} - \mathbf{q}')$. This subsection reviews the main result of [132] for the purpose of constructing an intuitive and computationally efficient approximation of (5.11).

The strategy of [132] is to construct independent probability distributions $\mathsf{p}_{\mathbf{v}}$ and p_s associated with random $\mathbf{v} \in \mathbb{R}^Q$ and random $s \in \mathbb{R}$ as well as a function (that is parameterized by $\mathbf{q}$) $\tilde{z}(\cdot, \cdot; \mathbf{q}) : \mathbb{R}^Q \times \mathbb{R} \times \mathbb{R}^Q \mapsto \mathbb{R}$, such that

$$\mathsf{E}_{\mathbf{v},s}(\tilde{z}(\mathbf{v}, s; \mathbf{q})\tilde{z}(\mathbf{v}, s; \mathbf{q}')) = k(\mathbf{q} - \mathbf{q}'), \tag{5.19}$$

where $\mathsf{E}_{\mathbf{v},s}(\cdot)$ denotes expectation with respect to $\mathsf{p}_{\mathbf{v}}\mathsf{p}_s$. When such a construction exists, one can build approximate feature maps $\tilde{\mathbf{z}}$ by concatenating and normalizing evaluations of $\tilde{z}$ on Z samples $\{(\mathbf{v}_1, s_1), \ldots, (\mathbf{v}_Z, s_Z)\}$ of $(\mathbf{v}, s)$ (drawn jointly albeit independently), to produce approximate features

$$\tilde{\mathbf{z}}(\mathbf{q}) := \sqrt{\frac{2}{Z}}[\tilde{z}(\mathbf{v}_1, s_1; \mathbf{q}), \ldots, \tilde{z}(\mathbf{v}_Z, s_Z; \mathbf{q})]^\mathsf{T} \tag{5.20}$$

for any $\mathbf{q}$. Then by the strong law of large numbers,

$$\lim_{Z\to\infty} \langle \tilde{\mathbf{z}}(\mathbf{q}), \tilde{\mathbf{z}}(\mathbf{q}') \rangle_{\mathbb{R}^Z} \overset{a.s.}{\to} k(\mathbf{q}, \mathbf{q}') \qquad \forall \mathbf{q}, \mathbf{q}' \tag{5.21}$$

which, in conjunction with strong performance guarantees for finite Z [132, 133], justifies interpreting $\tilde{\mathbf{z}}$ as an approximate (and now finite-dimensional) feature map.

We use the Fourier construction of [132] that assigns $\tilde{z}(\mathbf{v}, s; \mathbf{q}) \leftarrow \cos\left(2\pi\left(\mathbf{v}^\mathsf{T}\mathbf{q} + s\right)\right)$. If also $\mathsf{p}_s \leftarrow \mathrm{unif}(0, 1)$, then $\mathsf{E}_{\mathbf{v},s}(\tilde{z}(\mathbf{v}, s; \mathbf{q})\tilde{z}(\mathbf{v}, s; \mathbf{q}'))$ simplifies to

$$\int_{\mathbb{R}^Q} \cos\left(2\pi\mathbf{v}^\mathsf{T}(\mathbf{q} - \mathbf{q}')\right)\mathsf{p}_\mathbf{v}(\mathbf{v}) \, \mathrm{d}\, \mathbf{v}. \tag{5.22}$$

For symmetric $\mathsf{p}_\mathbf{v}$, the integral in (5.22) exists [134] and is a Fourier transform. Thus choosing $\mathsf{p}_\mathbf{v} \leftarrow \mathcal{N}\left(\mathbf{0}_Q, (2\pi\mathbf{\Lambda})^{-2}\right)$ satisfies (5.19) for Gaussian kernel (5.12), where $\mathbf{0}_Q \in \mathbb{R}^Q$ is a vector of zeros.

Sampling $\mathsf{p}_\mathbf{v}, \mathsf{p}_s$ Z times and subsequently constructing $\tilde{\mathbf{Z}} := [\tilde{\mathbf{z}}(\mathbf{q}_1), \ldots, \tilde{\mathbf{z}}(\mathbf{q}_N)] \in \mathbb{R}^{Z\times N}$ via repeated evaluations of (5.20) gives for $Z \ll N$ a low-rank approximation $\tilde{\mathbf{Z}}^\mathsf{T}\tilde{\mathbf{Z}}$ of Gram matrix $\mathbf{K}$. Substituting this approximation into (5.11) and applying the matrix inversion lemma [135] yields

$$\widehat{x}_l(\cdot) \leftarrow m_{x_l} + \mathbf{c}_{\mathbf{z}x_l}^\mathsf{T}((\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} + \rho_l\mathbf{I}_Z))^{-1}(\tilde{\mathbf{z}}(\cdot) - \mathbf{m}_{\tilde{\mathbf{z}}}), \tag{5.23}$$

where $m_{x_l} := \frac{1}{N}\tilde{\mathbf{x}}l^\mathsf{T}\mathbf{1}_N$ and $\mathbf{m}_{\tilde{\mathbf{z}}} := \frac{1}{N}\tilde{\mathbf{Z}}\mathbf{1}_N$ are sample means; and $\mathbf{c}_{\mathbf{z}x_l} := \frac{1}{N}\tilde{\mathbf{Z}}\mathbf{M}\tilde{\mathbf{x}}l$ and $\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}} := \frac{1}{N}\tilde{\mathbf{Z}}\mathbf{M}\tilde{\mathbf{Z}}^\mathsf{T}$ are sample covariances. Estimator (5.23) is an affine minimum mean-squared error estimator on the approximate features, and illustrates that Gaussian PERK via estimator (5.11) is asymptotically (in Z) equivalent to regularized affine regression after nonlinear, high-dimensional feature mapping.

5.5.2 Tuning Parameter Selection

This subsection proposes guidelines for data-driven selection of user-selectable parameters. Our goal here is to use problem intuition to automatically choose as many tuning parameters as possible, thereby leaving as few parameters as possible to manual selection. In this spirit, we focus on "online" model selection, where one chooses tuning parameters for training the estimator $\widehat{\mathbf{x}}(\cdot)$ *after* acquiring (unlabeled) real test data. This online approach can be considered a form of *transductive learning* [136, Ch. 8] since we train our estimator with knowledge of unlabeled test data in addition to labeled training data. Observe that since many voxel-wise separable MRI parameter estimation problems are comparatively low-dimensional, PERK estimators can often be quickly trained using only a moderate number of simulated training examples; in fact, training can in some problems take comparable or even less time than evaluating the PERK estimator on full-volume high-resolution measurement images. For these reasons, online PERK model selection is often practical.

5.5.2.1 Choosing Sampling Distribution

For reasonable PERK performance, it is important to choose the joint distribution of latent and known parameters $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ such that latent parameters can be estimated precisely over the joint distribution's support $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. For continuously differentiable magnitude signal model $\boldsymbol{\mu}$, we quantify precision at a single point $(\mathbf{x}, \boldsymbol{\nu})$ using the Fisher information matrix

$$\begin{aligned}\mathbf{F}(\mathbf{x}, \boldsymbol{\nu}) &:= \mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}\left(\left(\nabla_{\mathbf{x}} \log \mathsf{p}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}\right)^{\mathsf{T}} \nabla_{\mathbf{x}} \log \mathsf{p}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}\right) \\ &\approx \left(\nabla_{\mathbf{x}} \boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\nu})\right)^{\mathsf{T}} (\boldsymbol{\Sigma})^{-1} \nabla_{\mathbf{x}} \boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\nu})\end{aligned} \tag{5.24}$$

where $\nabla_{\mathbf{x}}(\cdot)$ denotes row gradient with respect to $\mathbf{x}$ and the approximation holds well for moderately high-SNR measurements [30]. When it exists, the inverse of $\mathbf{F}(\mathbf{x}, \boldsymbol{\nu})$ provides a lower-bound on the conditional covariance of any unbiased estimator of $\mathbf{x}$ [87]. For good performance, it is thus reasonable to ensure $\mathbf{F}(\mathbf{x}, \boldsymbol{\nu})$ is well-conditioned over $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$.

There are many strategies one could employ to control the condition number of $\mathbf{F}(\mathbf{x}, \boldsymbol{\nu})$ over $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. In our experiments, we used data [6] from acquisitions designed to *minimize* a cost function related to the *maximum* of $\mathbf{F}^{-1}(\mathbf{x}, \boldsymbol{\nu})$ over bounded latent and known parameter ranges of interest (Subsection 5.7.1 provides application-specific details). We then assigned $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ to coincide with the support of these acquisition design parameter ranges of interest. Assessing worst-case imprecision via the conservative minimax criterion is appropriate here because point-wise poor conditioning at any $(\mathbf{x}, \boldsymbol{\nu}) \in \mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ can induce PERK estimation error over larger subsets of $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$.

If many separate prior parameter estimates are available, one can estimate the particular shape of $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ empirically and then clip and renormalize $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ so as to assign nonzero probability only within an appropriate support. When prior estimates are unavailable, it may in certain problems be reasonable to instead assume a separable distributional structure $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}} \equiv \mathsf{p}_{\mathbf{x}}\mathsf{p}_{\boldsymbol{\nu}}$ in which case one can still estimate $\mathsf{p}_{\boldsymbol{\nu}}$ empirically but must set $\mathsf{p}_{\mathbf{x}}$ manually based on typical ranges of latent parameters.

5.5.2.2 Choosing Regularization Parameters

As presented, PERK estimator (5.11) and its approximation (5.23) leave freedom to select different regularization parameters $\rho_1, \ldots, \rho_L$ for estimating each of the L latent parameters. However, the respective unitless matrices $\mathbf{MKM}$ and $\mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$ whose condition numbers are influenced by $\rho_1, \ldots, \rho_L$ do not vary with l. Thus it is reasonable to assign each $\rho_l \leftarrow \rho \; \forall l \in \{1, \ldots, L\}$ some fixed $\rho > 0$. This simplification significantly reduces training computation to just one rather than L large matrix inversions. We select the scalar regularization parameter ρ using the holdout process described in Subsection 5.7.1.1.

5.5.2.3 Choosing Kernel Bandwidth

It is desirable to choose the Gaussian kernel's bandwidth matrix $\mathbf{\Lambda}$ such that PERK estimates are invariant to the overall scale of test data. We use (after observing test data, and for both training and testing)

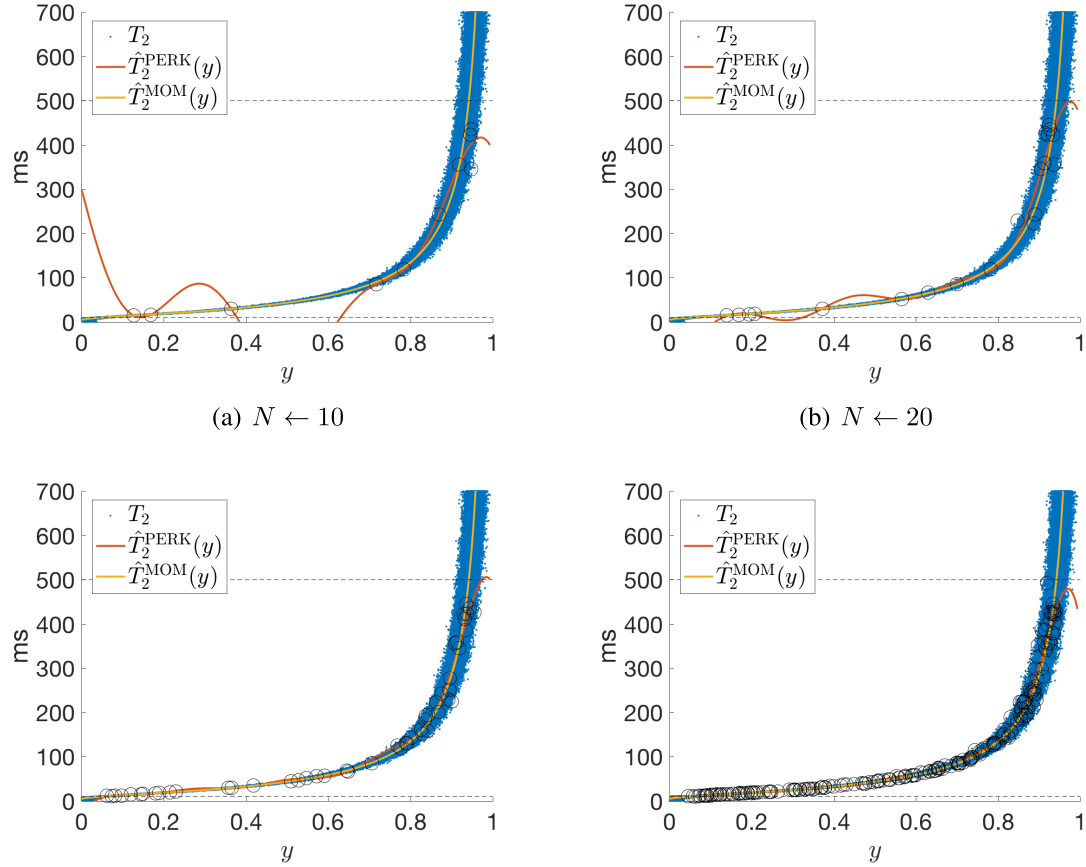
$$\mathbf{\Lambda} \leftarrow \lambda \operatorname{diag}\left(\left[\mathbf{m}_{\boldsymbol{\alpha}}^{\mathsf{T}}, \mathbf{m}_{\boldsymbol{\nu}}^{\mathsf{T}}\right]^{\mathsf{T}}\right), \tag{5.25}$$

where $\mathbf{m}_{\boldsymbol{\alpha}} \in \mathbb{R}^D$ and $\mathbf{m}_{\boldsymbol{\nu}} \in \mathbb{R}^K$ are sample means across voxels of magnitude test image data and known parameters, respectively; and $\operatorname{diag}(\cdot)$ assigns its argument to the diagonal entries of an otherwise zero matrix. We select the only scalar bandwidth parameter $\lambda > 0$ using holdout as well.

5.6 PERK Demonstration in a 1-D Toy Problem

To build intuition and for ease of visualization, we first apply PERK in a one-dimensional toy problem, namely T_2 estimation from a single spin-echo measurement. We generated training data using a mono-exponential (unity-m_0) signal model $y = e^{-T_\mathrm{E}/T_2} + \epsilon$, where y is a complex spin-echo measurement, $T_\mathrm{E} \leftarrow 30$ms is the echo time and $\epsilon \sim \mathbb{C}\mathcal{N}(0, 0.01^2)$ is complex Gaussian noise. We sampled $N \leftarrow 10, 20, 50, 200$ regressands from T_2 sampling distribution $\mathsf{p}_{T_2} \leftarrow \operatorname{logunif}(10, 500)$ and took the magnitude of noisy complex signal model evaluations to generate corresponding magnitude regressors. We trained PERK separately using each of the four labeled training datasets, holding fixed hyperparameters $(\lambda, \rho) \leftarrow (2^{-1.5}, 2^{-20})$ that were manually chosen to aid in illustrating PERK's typical behavior.

Fig. 5.1 illustrates the 1-D PERK estimator $\widehat{T}_2^{\mathrm{PERK}}$ and shows how its performance improves as N is increased. To produce each subfigure, we uniformly sampled 100,000 true (latent) T_2 values, evaluated the noisy signal model as in training to generate magnitude



(a) $N \leftarrow 10$ (b) $N \leftarrow 20$

(c) $N \leftarrow 50$ (d) $N \leftarrow 200$

Figure 5.1: Illustrations of PERK for T_2 estimation from a single spin echo measurement. Subfigures vary the number N of PERK training points, marked with black circles. The orange and yellow curves plot PERK $\widehat{T}_2^{\text{PERK}}$ and MOM $\widehat{T}_2^{\text{MOM}}$ estimators evaluated at test points, marked with blue dots. Dashed black lines denote the sampling distribution support $\mathsf{supp}(\mathsf{p}_{T_2})$ over which each PERK estimator was trained. As N increases, $\widehat{T}_2^{\text{PERK}}$ appears more similar to $\widehat{T}_2^{\text{MOM}}$ within well-sampled regions of $\mathsf{supp}(\mathsf{p}_{T_2})$.

test points (blue dots), and evaluated each PERK estimator at the unlabeled test points (orange curves). For comparison, subfigures within Fig. 5.1 also plot the intuitive method-of-moments (MOM) estimator $\widehat{T}_2^{\text{MOM}}(\cdot) := -T_{\text{E}}/\log|\cdot|$ (yellow curves). As N increases, $\widehat{T}_2^{\text{PERK}}$ appears more similar to $\widehat{T}_2^{\text{MOM}}$ within well-sampled regions of $\mathsf{supp}(\mathsf{p}_{T_2})$ (marked by dashed black lines). PERK will be more useful in nonlinear estimation problems where such a minimally biased and low-dimensional MOM estimator is unavailable.

5.7 Experimentation

This section demonstrates PERK for quantifying MR relaxation parameters T_1 and T_2, a well-studied application. We studied this relatively simple problem instead of the more complicated problems that motivated our method because we had access to reference T_1, T_2 phantom NMR measurements [2] for external validation and because it is easier to validate PERK estimates against gold-standard grid search estimates in problems involving few unknowns (Ch. 6 applies PERK to a more challenging problem). Subsection 5.7.1 describes implementation details that were fixed in all simulations and experiments. Subsection 5.7.2 studies estimator statistics in numerical simulations. Subsection 5.7.3 and Subsection 5.7.4 respectively compare PERK performance in phantom and *in vivo* experiments.

5.7.1 Methods

In all simulations and experiments, we used data arising from a fast acquisition [6] consisting of two spoiled gradient-recalled echo (SPGR) [19] and one dual-echo steady-state (DESS) [22] scans. Since each SPGR (DESS) scan generates one (two) signal(s) per excitation, this acquisition yielded $D \leftarrow 4$ datasets. We fixed scan parameters to be identical to those in [6], wherein repetition times and flip angles were optimized for precise T_1 and T_2 estimation in cerebral tissue at 3T field strength [6] and echo times were fixed across scans. We used standard magnitude[3] SPGR and DESS signal models expressed as a function of four free parameters per voxel: flip angle spatial variation (due to transmit field inhomogeneity) κ; longitudinal and transverse relaxation time constants T_1 and T_2; and a pure-real proportionality constant m_0. We assumed prior knowledge of $K \leftarrow 1$ known parameter $\boldsymbol{\nu} \leftarrow \kappa$ (in experiments, through separate acquisition and estimation of flip angle scaling maps) and collected the remaining $L \leftarrow 3$ latent parameters as $\mathbf{x} \leftarrow [m_0, T_1, T_2]^\mathsf{T}$.

We used the same PERK training and testing process across all simulations and experiments. We assumed a separable prior distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}} \leftarrow \mathsf{p}_{m_0,T_1,T_2,\kappa} \equiv \mathsf{p}_{m_0}\mathsf{p}_{T_1}\mathsf{p}_{T_2}\mathsf{p}_{\kappa}$ and estimated flip angle scaling marginal distribution p_κ from known κ map voxels via kernel density estimation (implemented using the built-in MATLAB® function `fitdist` with default options). To match the scaling of training and test data, we set m_0 marginal distribution $\mathsf{p}_{m_0} \leftarrow \mathrm{unif}(2.2 \times 10^{-16}, u)$, with u set as $6.67\times$ the maximum value of magnitude test

[3] Standard complex DESS signal models depend on a fifth free parameter associated with phase accrual due to off-resonance effects. Because the first and second DESS signals depend differently on off-resonance phase accrual [6], off-resonance related phase (unlike signal loss) cannot be collected into the (now complex) proportionality constant. To avoid (separate or joint) estimation of an off-resonance field map, we followed [6] and used magnitude SPGR and DESS signal models. We accounted for consequently Rician-distributed noise in magnitude image data during training.

data. We chose the supports of T_1, T_2 marginal distributions $\mathsf{p}_{T_1} \leftarrow \text{logunif}(400, 2000)$ms, $\mathsf{p}_{T_2} \leftarrow \text{logunif}(40, 200)$ms and clipped the support of p_κ to assign nonzero probability only within $[0.5, 2]$ such that these supports coincided with the supports over which [6] optimized the acquisition. We assumed noise covariance $\mathbf{\Sigma}$ of form $\sigma^2 \mathbf{I}_4$ (as in [6]) and estimated the (spatially invariant) noise variance σ^2 from Rayleigh-distributed regions of magnitude test data, using estimators described in [137]. We sampled $N \leftarrow 10^5$ latent and known parameter realizations from these distributions and evaluated SPGR and DESS signal models to generate corresponding noiseless measurements. After adding complex Gaussian noise realizations, we concatenated the (Rician) magnitude of these noisy measurements with known parameter realizations to construct pure-real regressors. We separately selected and then held fixed free parameters $\lambda \leftarrow 2^{0.6}$ and $\rho \leftarrow 2^{-41}$ via a simple holdout process, described in Subsection 5.7.1.1. We set Gaussian kernel bandwidth matrix $\mathbf{\Lambda}$ from test data via (5.25). We sampled $\mathbf{v}, s$ $Z \leftarrow 10^3$ times to construct approximate feature map $\tilde{\mathbf{z}}$. For each lth latent parameter where $l \in \{1, \dots, L\}$, we applied $\tilde{\mathbf{z}}$ to training data; computed sample means $m_{x_l}, \mathbf{m}_{\tilde{\mathbf{z}}}$ and sample covariances $\mathbf{c}_{\mathbf{z}x_l}, \mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$; and evaluated (5.23) on test image data and the known flip angle scaling map on a per-voxel basis.

5.7.1.1 Model Selection via Holdout

We selected Gaussian kernel bandwidth scaling parameter λ and regularization parameter ρ using the following offline holdout procedure in simulation. We discretized (λ, ρ) over a finely spaced grid spanning many orders of magnitude. As described in Subsection 5.7.1, we trained a PERK estimator $\widehat{\mathbf{x}}_{\lambda,\rho}$ for each candidate model parameter setting. We evaluated each PERK estimator on a separate simulated dataset consisting of many samples from the training prior distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$. We selected model parameters by exhaustively seeking a minimizer $\left(\widehat{\lambda}, \widehat{\rho}\right)$ of the "holdout" cost function

$$\Psi(\lambda, \rho) := \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left\| ([\operatorname{diag}(\mathbf{x}_t)])^{-1} (\widehat{\mathbf{x}}_{\lambda,\rho}(\mathbf{q}_t) - \mathbf{x}_t) \right\|_{\mathbf{W}}^2} \tag{5.26}$$

where $t \in \{1, \dots, T\}$ indexes T test points; each $\mathbf{x}_t$ is the true latent parameter corresponding to holdout test data point $\mathbf{q}_t$; and $\mathbf{W}$ is a diagonal unit-trace weighting matrix. Intuitively, $\Psi(\lambda, \rho)$ is the weighted normalized root mean squared error of PERK estimator $\widehat{\mathbf{x}}_{\lambda,\rho}$, where the mean approximates an expectation with respect to $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ and the latent parameter weighting is specified by $\mathbf{W}$.

Fig. 5.2 plots $\Psi(\lambda, \rho)$ for $T \leftarrow 10^5$ test points and $\mathbf{W} \leftarrow \operatorname{diag}\left([0, 0.5, 0.5]^\mathsf{T}\right)$ selected

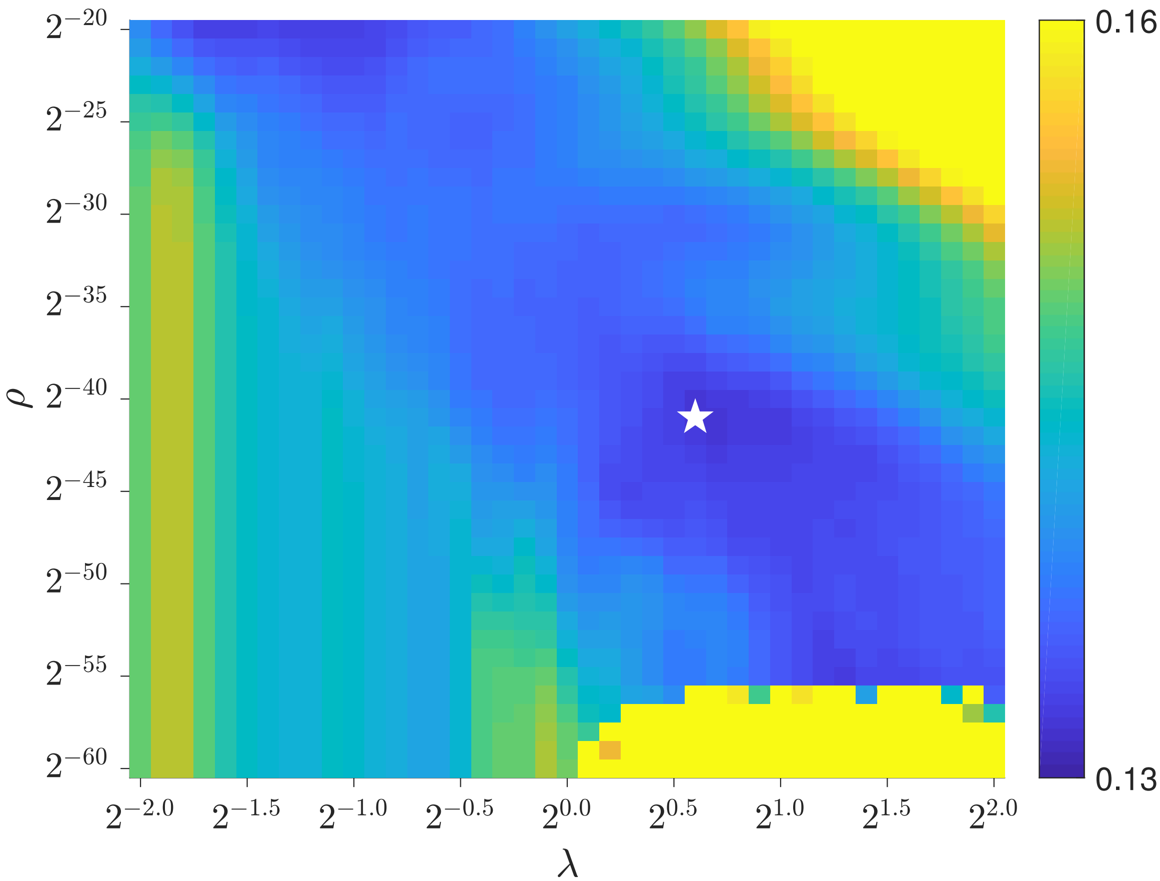


Figure 5.2: Holdout criterion $\Psi(\lambda, \rho)$ versus Gaussian kernel bandwidth scaling parameter λ and regularization parameter ρ. Each pixel is the weighted normalized root mean squared error of a candidate PERK estimator, where the empirical mean over 10^5 test points approximates an expectation with respect to training prior distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ and the weighting places emphasis on good T_1, T_2 estimation performance. A white star marks the minimizer $\left(\widehat{\lambda}, \widehat{\rho}\right) \leftarrow (2^{0.6}, 2^{-41})$.

to place equal emphasis on T_1, T_2 estimation. We chose our fine grid search range using a preliminary coarse grid search spanning a much wider range of (λ, ρ) values. Overall, we observe a broad range of (λ, ρ) values that yield similar cost function values. Holdout cost $\Psi(\lambda, \rho)$ gracefully increases with larger (λ, ρ) values due to under-fitting. For very small ρ values, $\Psi(\lambda, \rho)$ can be large because poorly conditioned matrix inversions cause machine imprecision to dominate estimation error. In all simulations and experiments, we fixed free model parameters to the minimizer $\left(\widehat{\lambda}, \widehat{\rho}\right) \leftarrow (2^{0.6}, 2^{-41})$, indicated by a white star.

5.7.1.2 Evaluation

We evaluated PERK latent parameter estimates against maximum-likelihood (ML) estimates computed via two well-suited algorithms that we describe here in turn. We first implemented a grid search estimator accelerated by the variable projection method (VPM) [25], a popular technique that has been used in many QMRI algorithms and applications (see *e.g.* [138, 105, 106, 95, 103, 109, 108, 111, 113, 139, 6]). Following [6], we clustered flip angle scaling map voxels into 20 clusters via k-means++ [93] and used each of the 20 cluster means along with 500 T_1 and T_2 values logarithmically spaced between $\left(10^{1.5}, 10^{3.5}\right)$ and $\left(10^{0.5}, 10^{3}\right)$ to compute 20 dictionaries, each consisting of $250,000$ signal vectors (fewer clusters introduced noticeable errors in experiments). Iterating over clusters, we generated each cluster's dictionary and applied VPM and grid search over magnitude image data voxels assigned to that cluster.

We also compared PERK to iterative ML optimization via a preconditioned variant of the classical gradient projection method (PGPM) [23]. We designed the preconditioner as the inverse of a positive definite diagonal majorizer of the negative log-likelihood cost function's Hessian matrix, updated for the first five iterations and fixed thereafter. We employed a diagonal preconditioner to retain the linear convergence rate guarantees of GPM [37] yet accelerate practical performance. We initialized PGPM via conventional method-of-moments estimators of m_0, T_1 from 2 SPGR scans [35] and T_2 from 1 DESS scan [22] (the method-of-moments T_2 estimator is strongly biased). We used the MATLAB® Symbolic Toolbox to generate cumbersome but analytical expressions for the gradient and Hessian of the magnitude SPGR and DESS signal models. At each PGPM iteration, we used these expressions to compute a preconditioned descent direction, update the iterate, and project each voxel's T_1 and T_2 iterate to lie within $[100, 3000]$ms and $[10, 700]$ms, respectively. We continued iterations until the convergence criterion

$$\left\|(\boldsymbol{\Omega})^{-1}\left(\mathbf{X}^{(i)} - \mathbf{X}^{(i-1)}\right)\right\|_{\mathrm{F}} < 10^{-7}\left\|(\boldsymbol{\Omega})^{-1}\left(\mathbf{X}^{(i-1)}\right)\right\|_{\mathrm{F}} \tag{5.27}$$

was satisfied, where $\mathbf{X}$ collects latent parameter voxels in its columns, $(\cdot)^{(i)}$ denotes the ith iterate, $\boldsymbol{\Omega} := \mathrm{diag}\left(\mathrm{med}\left(\mathbf{X}^{(0)}\right)\right)$ is a fixed latent parameter weighting matrix, and $\mathrm{med}(\cdot)$ takes the median across the columns of its argument.

To ensure monotone local convergence in cost, we implemented PGPM to include a simple step-halving line search at each iteration. In early experiments however, we observed even in simulation and even with preconditioning that attempting to update all voxels simultaneously using a single line search resulted in large errors due to excessive step-halving and subsequent early termination of iterations. To circumvent separate line searches for every voxel, we first clustered latent parameter initializations and flip angle scaling map voxels into 50 clusters and then ran PGPM separately on each cluster (fewer clusters reintroduced early stopping).

We performed all simulations and experiments running MATLAB® R2013a on a 3.5GHz desktop computer equipped with 32GB RAM. Because our experiments use a single slice of image data, we report PERK training and testing times separately and note that only the latter time would scale linearly with the number of voxels (the former would scale negligibly due only to online model selection). In the interest of reproducible research, code and data will be freely available at https://gitlab.eecs.umich.edu/fessler/qmri.

5.7.2 Numerical Simulations

We assigned typical T_1, T_2 values in white matter (WM) and grey matter (GM) at 3T [31] to the 81st slice of the BrainWeb digital phantom [34] to produce ground truth m_0, T_1, T_2 maps. We simulated 217×181 noiseless single-coil SPGR and DESS image data, modeling (and then assuming as known) 20% flip angle spatial variation. We corrupted noiseless datasets with additive complex Gaussian noise to yield noisy complex datasets with SNR ranging from 94-154 in WM and 82-154 in GM, where SNR is defined

$$\mathsf{SNR}(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\epsilon}}) := \|\tilde{\mathbf{y}}\|_2 / \|\tilde{\boldsymbol{\epsilon}}\|_2 \tag{5.28}$$

for image data voxels $\tilde{\mathbf{y}}$ and noise voxels $\tilde{\boldsymbol{\epsilon}}$ corresponding to a region of interest (ROI) within a single SPGR/DESS dataset. We estimated m_0, T_1, T_2 from noisy magnitude images and known κ maps using VPM, PGPM, and PERK. VPM took 791s; PGPM took 1821s; and PERK training and testing respectively took 3.6s and 1.5s.

Figs. 5.3, 5.4, and 5.5 compare VPM, PGPM, and PERK estimates of m_0, T_1, T_2 respectively, alongside $10\times$ magnified absolute difference images with respect to the ground truth. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that within WM- and GM-like voxels, all three methods exhibit low estimation error.

Table 5.1 compares sample statistics of VPM, PGPM, and PERK m_0, T_1, T_2 estimates, computed over 7810 WM-like and 9162 GM-like voxels. Overall, all three methods achieve excellent performance. PERK estimates are slightly more precise but slightly less accurate than gold-standard VPM estimates. Results suggest that at least in WM- and GM-like voxels, PGPM is capable of descending the ML cost towards a desirable solution; in fact, PGPM achieves slightly better precision than either VPM or PERK. All three methods exhibit comparable root mean squared errors (RMSEs).

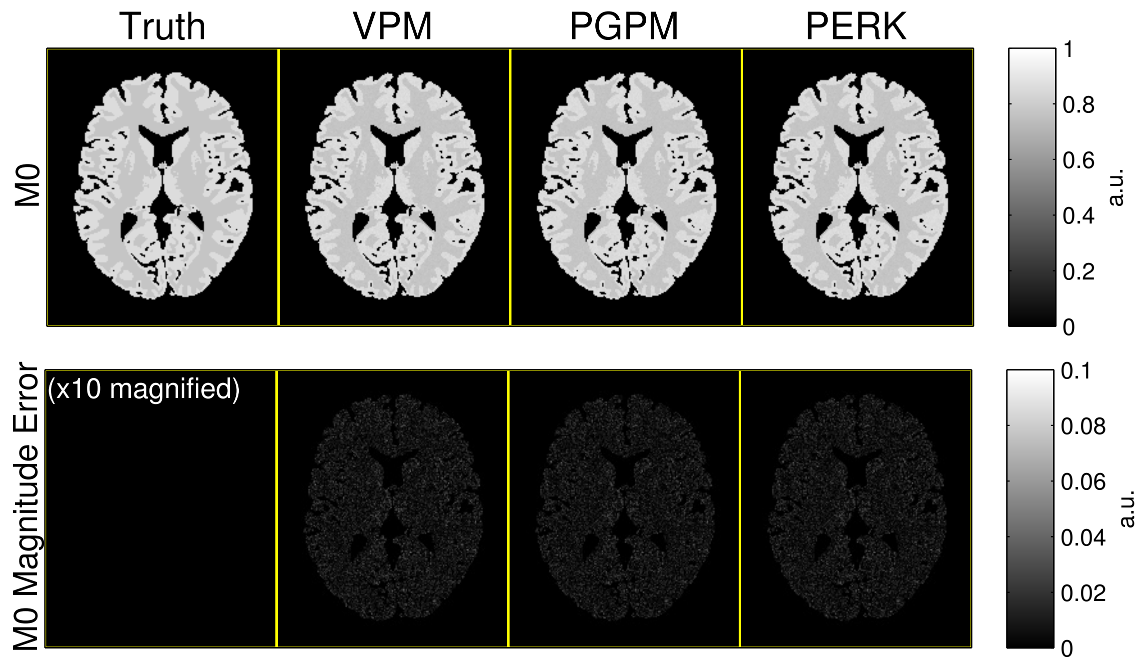


Figure 5.3: m_0 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three m_0 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics.

	Truth	VPM	PGPM	PERK
WM m_0	0.77	0.7700 ± 0.00919 (0.0092)	0.76999 ± 0.00871 (0.00871)	0.77002 ± 0.00873 (0.00873)
GM m_0	0.86	0.8601 ± 0.01192 (0.0119)	0.8600 ± 0.01142 (0.0114)	0.8613 ± 0.01147 (0.0133)
WM T_1	832	832.1 ± 17.2 (17.2)	832.1 ± 16.2 (16.2)	833.0 ± 16.5 (16.5)
GM T_1	1331	1331.5 ± 31.1 (31.1)	1331.2 ± 29.7 (29.7)	1332.1 ± 30.4 (30.4)
WM T_2	79.6	79.61 ± 0.988 (0.988)	79.60 ± 0.952 (0.952)	79.46 ± 0.978 (0.989)
GM T_2	110.	110.02 ± 1.40 (1.40)	110.02 ± 1.35 (1.35)	109.91 ± 1.35 (1.35)

Table 5.1: Sample means $\pm$ sample standard deviations (RMSEs) of VPM, PGPM, and PERK m_0, T_1, T_2 estimates, computed in simulation over 7810 WM-like and 9162 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. m_0 values are unitless. T_1, T_2 values are in milliseconds. Figs. 5.3, 5.4, and 5.5 present corresponding images.

5.7.3 Phantom Experiments

Phantom experiments used datasets from fast coronal scans of a High Precision Devices® MR system phantom T_2 array acquired on a 3T GE Discovery™ scanner with an 8-channel receive head array. This acquisition consisted of: two SPGR scans with $5, 15°$ flip angles and $12.2, 12.2$ms repetition times; one DESS scan with $30°$ flip angle and 17.5ms repetition

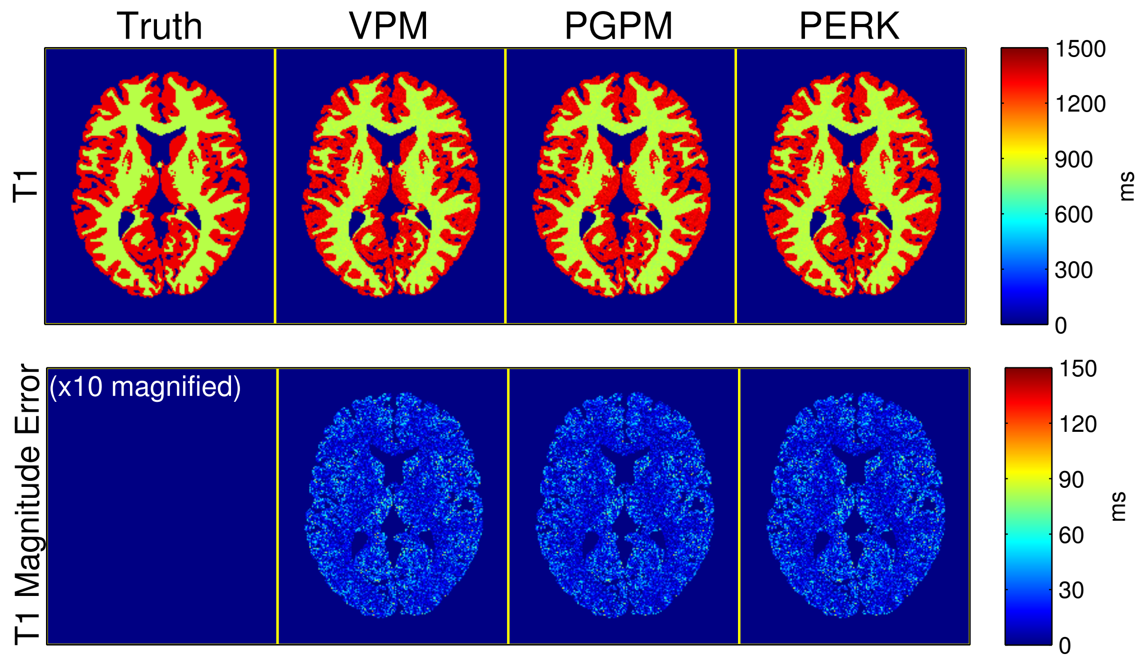


Figure 5.4: T_1 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three T_1 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics.

time; and two Bloch-Siegert (BS) scans [90] (for separate flip angle scaling estimation). Nominal flip angles were achieved by scaling a 2cm slab-selective Shinnar-Le Roux RF pulse [89] of duration 1.28ms and time-bandwidth product 4. All scans collected fully-sampled 3D Cartesian data using 4.67ms echo times with a $256 \times 256 \times 8$ matrix over a $24 \times 24 \times 4\text{cm}^3$ field of view. Scan time totaled 3m17s. The scan room temperature was recorded as 293K at the beginning of the exam. Further acquisition details are in [6].

For each SPGR, DESS, and BS dataset, we reconstructed raw coil images via 3D Fourier transform and subsequently processed only one image slice centered within the excitation slab. We combined SPGR and DESS coil images using a natural extension of [92] to the case of multiple datasets. We similarly (but separately) combined BS coil images and estimated κ maps by normalizing and calibrating regularized transmit field estimates [91] from complex coil-combined BS images. We estimated m_0, T_1, T_2 from magnitude SPGR/DESS images and κ maps using VPM, PGPM, and PERK. VPM took 928s; PGPM took 1257s; and PERK training and testing respectively took 4.2s and 1.9s.

Fig. 5.6 compares VPM, PGPM, and PERK m_0, T_1, T_2 estimates. Vials are enumerated in descending T_1, T_2 order. Vials whose T_1, T_2 values are within sampling distribution

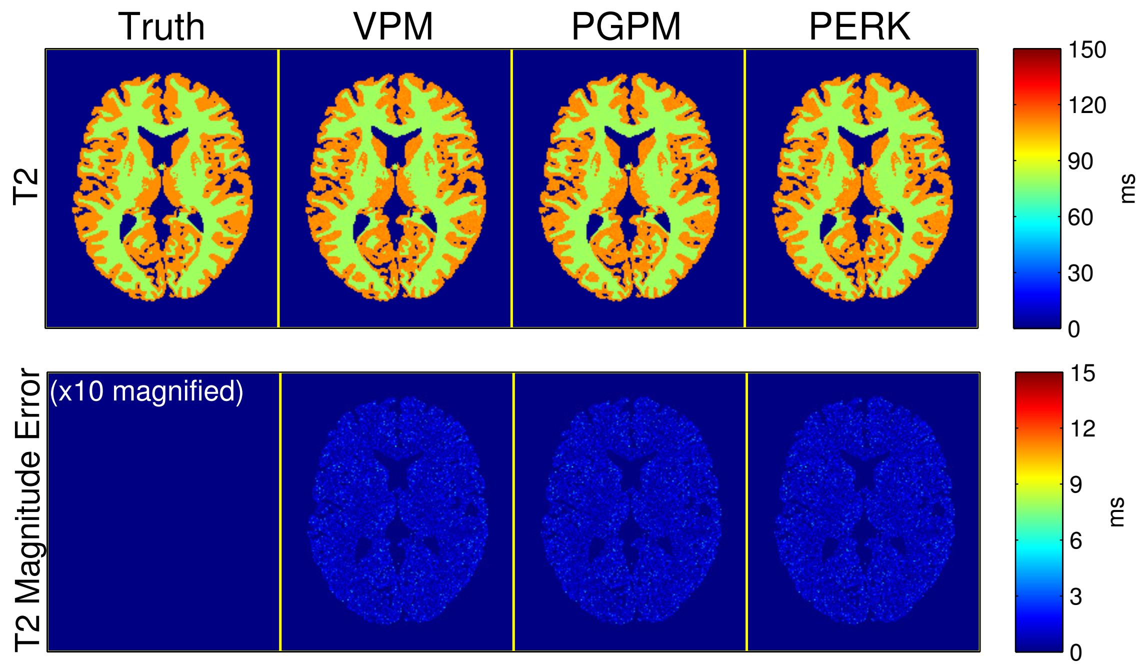


Figure 5.5: T_2 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three T_1 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics.

support $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ (as measured by NIST NMR reference measurements [2]) have labels highlighted with yellow numbers. Here, $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ was chosen to reflect the ranges of latent parameter values for which the SPGR/DESS scan parameters were optimized in [6]. Circular ROIs are selected well away from vial encasings and correspond with sample statistics presented in Fig. 5.7. Distilled water surrounds the encased vials. Within the highlighted vials of interest, VPM, PGPM, and PERK estimates appear visually similar.

Fig. 5.7 compares sample means and sample standard deviations computed within ROIs of VPM, PGPM, and PERK T_1, T_2 estimates against nuclear magnetic resonance (NMR) reference measurements reported at 293.00K from the National Institute for Standards of Technology (NIST) [2]. Yellow box boundaries indicate projections of the PERK sampling distribution's support $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. ROI labels correspond with vial markers depicted in Fig. 5.6. Within $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$, corresponding tables demonstrate that VPM, PGPM, and PERK estimates agree excellently with each other and reasonably with NMR measurements. We do not expect good PERK performance outside $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ and indeed observe poor ability to extrapolate. As discussed in Subsection 5.5.2.1 and demonstrated in Subsection 5.8.2, expanding $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ well beyond the acquisition design parameter range of

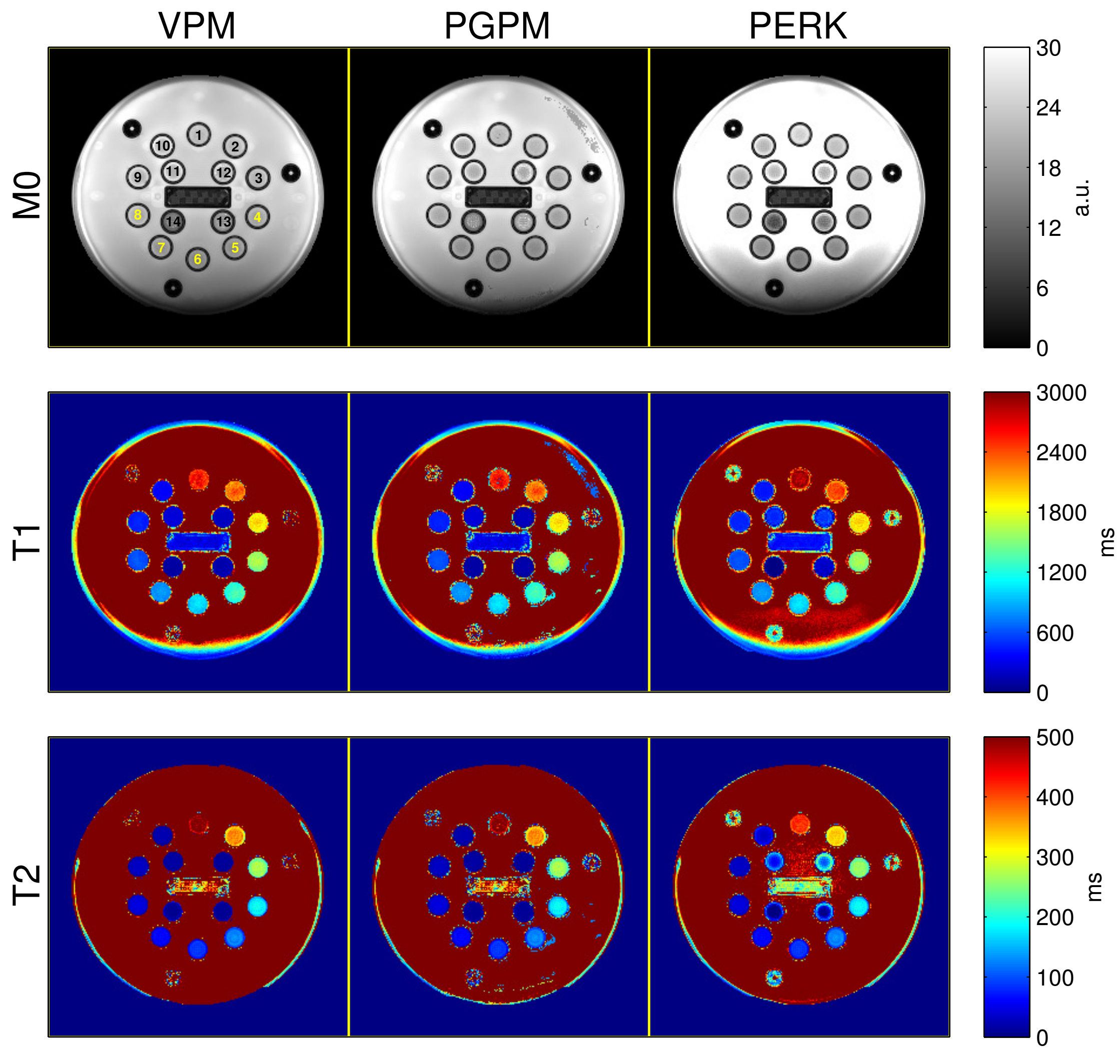
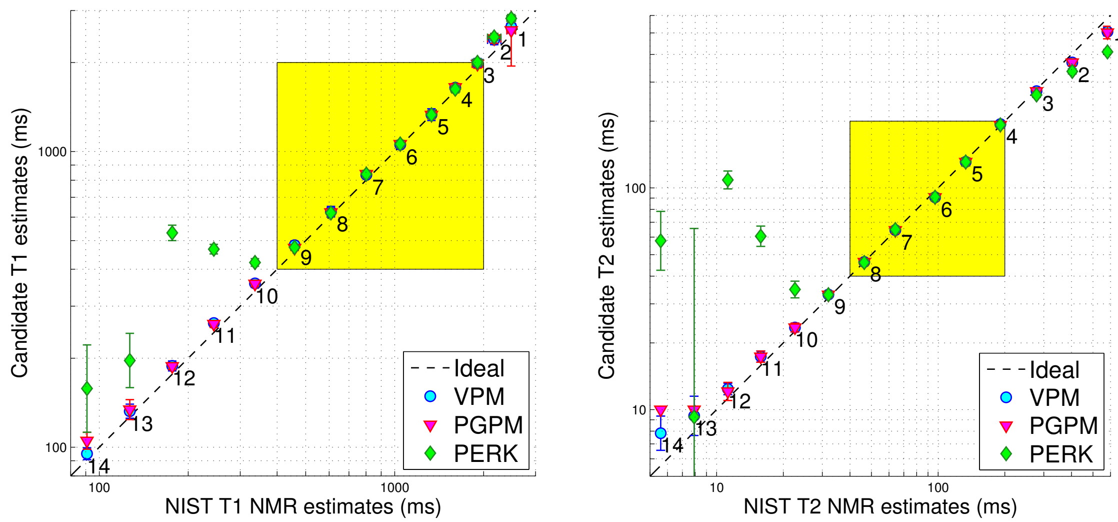


Figure 5.6: VPM, PGPM, and PERK m_0, T_1, T_2 estimates in a quantitative phantom. Vials are enumerated and highlighted to correspond with markers and colored boxes in Fig. 5.7. PERK has only been trained to accurately estimate within vials 4-8; within these vials, VPM, PGPM, and PERK estimates appear visually similar.

interest can reduce PERK performance for typical T_1, T_2 WM and GM values.

5.7.4 *In vivo* Experiments

In vivo experiments used datasets from axial scans of a healthy volunteer acquired with a 32-channel Nova Medical® receive head array. To address bulk motion between scans, we rigidly registered coil-combined images to a reference before parameter estimation. All other data acquisition, image reconstruction, and parameter estimation details are the same as in phantom experiments (acquisition and reconstruction details are reported in [6]).



	NMR	VPM	PGPM	PERK
V4 T_1	1604 ± 7.2	1645 ± 48	1649 ± 48	1626 ± 46
V5 T_1	1332 ± 0.8	1335 ± 61	1331 ± 41	$1332 \pm 40.$
V6 T_1	1044 ± 3.2	1055 ± 28	$1060. \pm 29$	1061 ± 29
V7 T_1	801.7 ± 1.70	834 ± 21	$840. \pm 23$	839 ± 23
V8 T_1	608.6 ± 1.03	627 ± 25	623 ± 12	$620. \pm 13$
V4 T_2	190.94 ± 0.011	194 ± 5.5	192.4 ± 5.2	192.5 ± 4.9
V5 T_2	133.27 ± 0.073	131.2 ± 5.3	131 ± 5.5	131 ± 5.5
V6 T_2	96.89 ± 0.049	90.8 ± 3.5	90.8 ± 3.5	90.9 ± 3.5
V7 T_2	64.07 ± 0.034	64.6 ± 2.2	64.5 ± 2.1	65.0 ± 2.1
V8 T_2	46.42 ± 0.014	46.4 ± 1.5	46.4 ± 1.5	46.1 ± 1.5

Figure 5.7: Phantom sample statistics of VPM, PGPM, and PERK T_1, T_2 estimates and NIST NMR reference measurements [2]. Plot markers and error bars indicate sample means and sample standard deviations computed over ROIs within the 14 vials labeled and color-coded in Fig. 5.6. Yellow box boundaries indicate projections of the PERK sampling distribution's support $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. Missing markers lie outside axis limits. Corresponding tables replicate sample means $\pm$ sample standard deviations for vials within $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. 'V#' indicates vial numbers. All values are reported in milliseconds. Within $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$, VPM, PGPM, and PERK estimates agree excellently with each other and reasonably with NMR measurements.

VPM took 838s; PGPM took 2178s; and PERK training and testing took 4.2s and 1.6s.

Fig. 5.8 compares VPM, PGPM, and PERK m_0, T_1, T_2 estimates. The PERK m_0 estimate appears smoothed (although no spatial regularization was used) but is otherwise very similar to the VPM and PGPM m_0 estimates. Narrow display ranges emphasize that VPM, PGPM, and PERK T_1, T_2 estimates discern cortical WM/GM boundaries similarly, though

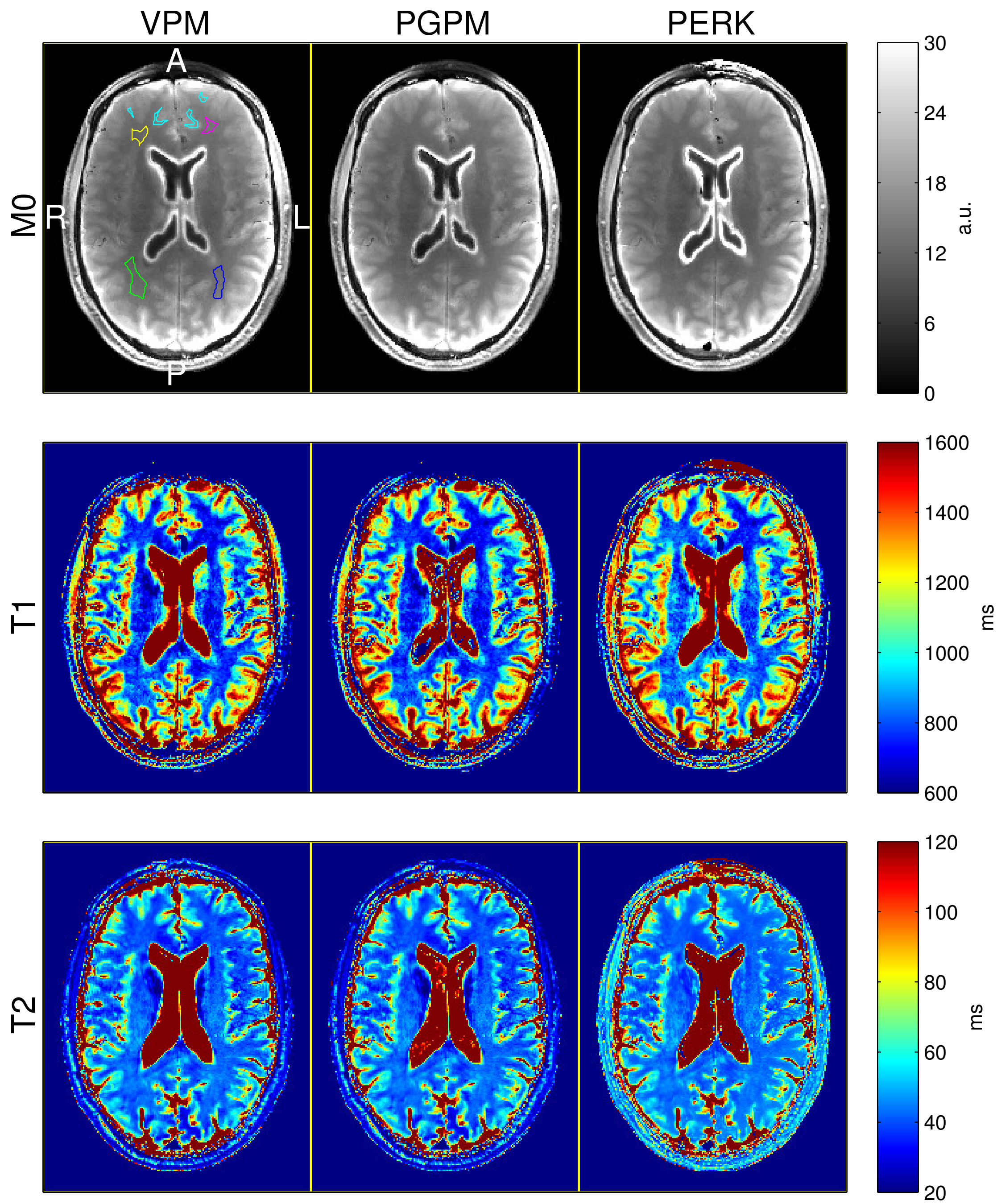


Figure 5.8: VPM, PGPM and PERK estimates of m_0, T_1, T_2 in the brain of a healthy volunteer. Separate WM ROIs are distinguished by anterior/posterior (A/P) and right/left (R/L) directions. Four small anterior cortical GM polygons are pooled into a single GM ROI. Images are cropped in post-processing for display.

PERK T_1 estimates are noticeably highest in some WM regions. VPM, PGPM, and PERK T_2 estimates are nearly indistinguishable in lateral regions but disagree somewhat in medial regions close to cerebrospinal fluid (CSF). We neither expect nor observe reasonable PERK

	ROI	VPM	PGPM	PERK
T_1	AR WM	778 ± 28	779 ± 27	832 ± 31
	AL WM	731 ± 37	713 ± 33	725 ± 41
	PR WM	805 ± 52	796 ± 51	831 ± 51
	PL WM	789 ± 40	788 ± 38	815 ± 42
	A GM	1120 ± 180	1120 ± 180	$1150 \pm 170.$
T_2	AR WM	40.0 ± 1.29	40.0 ± 1.27	41.18 ± 0.94
	AL WM	39.7 ± 1.7	39.7 ± 1.7	41.3 ± 1.02
	PR WM	43.0 ± 2.7	43.0 ± 2.7	43.7 ± 2.6
	PL WM	43.0 ± 1.8	43.0 ± 1.8	43.5 ± 1.36
	A GM	53.5 ± 11.8	53.4 ± 11.7	53.3 ± 11.6

Table 5.2: *In vivo* sample means $\pm$ sample standard deviations of VPM, PGPM, and PERK T_1, T_2 estimates, computed over color-coded ROIs indicated in Fig. 5.8. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. All values are in milliseconds.

performance in voxels containing CSF.

Table 5.2 summarizes sample statistics of VPM, PGPM, and PERK T_1, T_2 estimates, computed over four separate WM ROIs containing 96, 69, 224, and 148 voxels and one pooled cortical anterior GM ROI containing 156 voxels. Overall, VPM, PGPM, and PERK T_1, T_2 estimates are comparable. T_1 estimates in GM and T_2 estimates in WM/GM do not differ significantly. PERK T_1 estimates are significantly higher than VPM and PGPM T_1 estimates in one WM ROI; however, all T_1 estimates are well within the range of typical literature measurements at 3T (see *e.g.* [31, 32]).

5.8 Robustness Studies

This section investigates PERK robustness to two types of non-idealities that may be encountered in other applications. Subsection 5.8.1 studies PERK performance sensitivity to mismatch between training and testing noise variance. Subsection 5.8.2 studies PERK performance degradation when trained with latent parameter distributions that have wider support than the parameter ranges used for optimizing the scan design in [6].

5.8.1 Mismatch in Training vs. Testing Noise Statistics

To assess the importance of training PERK with appropriately noisy training data, we investigated PERK's performance sensitivity to the standard deviation σ of the noise distribution from which noise realizations are drawn to generate training data. Instead of

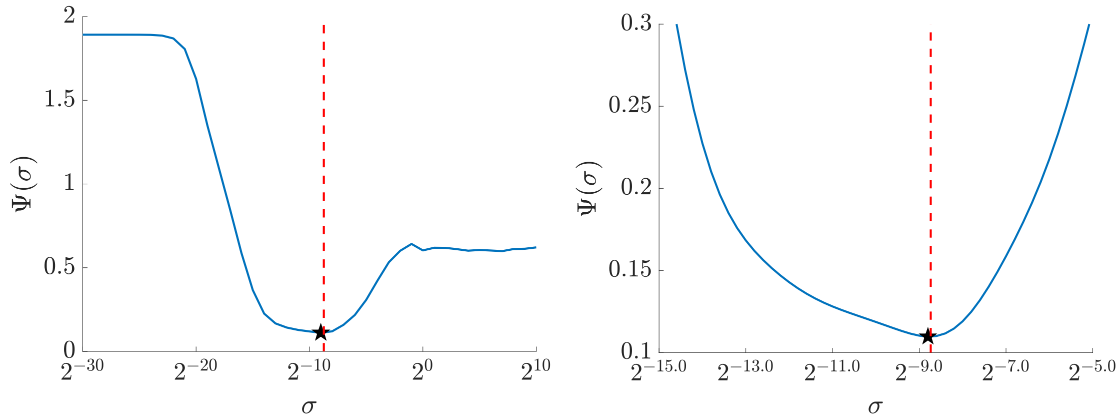


Figure 5.9: Performance criterion $\Psi(\sigma)$ versus PERK training noise standard deviation σ, over two different scales. Similar to Fig. 5.2, each point on the blue curve is the weighted normalized root mean squared error of a separately trained PERK estimator. In each subplot, a black star marks the performance criterion minimizer $\widehat{\sigma} \leftarrow 2^{-8.8}$ while a dashed red line marks the (latent) test data noise standard deviation $\sigma^\star \leftarrow 2^{-8.736}$. To within quantization error, PERK performs best when trained with training data whose noise statistics match test data noise statistics. Performance degradation as measured by Ψ does not exceed 10% for σ values within a factor of two of $\sigma^\star$. This result suggests that PERK is somewhat robust to moderate misspecification of the noise level.

setting σ online as in other experiments, here we fixed σ offline to one of many discretized values spanning many orders of magnitude. Otherwise as described in Subsection 5.7.1, we trained a PERK estimator for each σ setting. Similar to Subsection 5.7.1.1, we tested each PERK estimator on a separate simulated dataset consisting of 10^5 samples from training prior distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$. We assessed performance sensitivity by comparing evaluations $\Psi(\sigma)$ of holdout cost (5.26) at each σ setting.

Fig. 5.9 plots $\Psi(\sigma)$ as σ is varied over two different scales. In each subplot, a black star marks the minimizer $\hat{\sigma} \leftarrow 2^{-8.8}$ while a dashed red line marks the (latent) test data noise standard deviation $\sigma^\star \leftarrow 2^{-8.736}$. To within quantization error, PERK performs best when trained with training data whose noise statistics match test data noise statistics. As measured by Ψ, PERK performance degrades by at most 10% for choices of $\sigma \in [2^{-10.2}, 2^{-8}]$. Results suggest that for good PERK performance, it is desirable to set σ to within about a factor of two of the test data noise standard deviation $\sigma^\star$. Nevertheless, there is a zone near $\sigma^\star$ where PERK performance is reasonably similar, indicating that PERK is somewhat robust to some misspecification of the noise level.

5.8.2 Mismatch in Scan Design vs. Sampling Distribution Support

Although the SPGR/DESS acquisition was optimized in [6] for a certain range of T_1, T_2 values, it is interesting to investigate how well PERK can perform outside that parameter range if presented (simulated) training data over a wider range of latent parameters. It is also interesting to explore whether using such a wider range of latent parameters for training degrades performance for the parameter range of primary interest. Thus, we repeated the phantom experiment described in Subsection 5.7.3 except now using a PERK estimator trained using a sampling prior distribution with broader support. We still assume a separable prior distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}} \leftarrow \mathsf{p}_{m_0}\mathsf{p}_{T_1}\mathsf{p}_{T_2}\mathsf{p}_{\kappa}$ (with p_{m_0} and p_{κ} set as before) but now set $\mathsf{p}_{T_1} \leftarrow \text{logunif}(10^{1.5}, 10^{3.5})$ and $\mathsf{p}_{T_2} \leftarrow \text{logunif}(10^{0.5}, 10^{3.5})$ to have wider supports. These support endpoints now match the grid search support used by the VPM. All other training and testing details are unchanged from before.

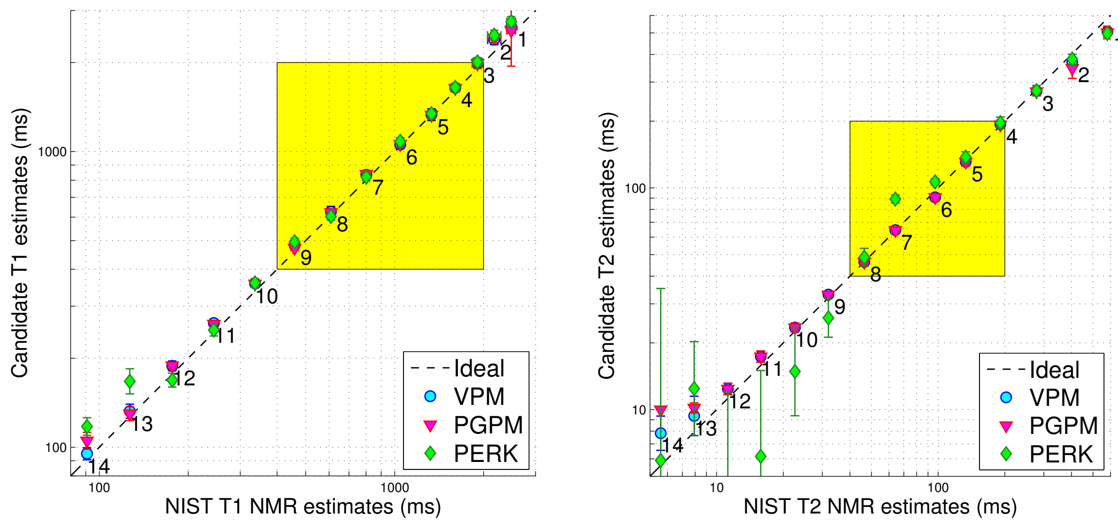
Fig. 5.10 is analogous to Fig. 5.7 in that it plots sample means and sample standard deviations computed within ROIs of VPM, PGPM, and PERK T_1, T_2 estimates, except now using a PERK estimator trained over the broader sampling distribution. Fig. 5.11 presents corresponding images. The yellow boxes are unchanged from Fig. 5.7 and so their boundaries no longer correspond to projections of the PERK sampling distribution's support. Rather, they serve to clearly highlight that PERK estimator performance can significantly deteriorate even over the parameter range of interest, when trained using a range of parameters that exceeds the design criteria of the acquisition.

Fig. 5.10 also tabulates sample means and sample standard deviations computed within ROIs of vials 4-8. Comparing again with Fig. 5.7, PERK T_2 estimation accuracy is more severely affected than T_1 estimation accuracy (interestingly, T_1 estimation accuracy is in fact improved for many vials). PERK T_1, T_2 estimation precision is consistently worse in vials 4-8 when trained over the broader sampling range.

These observations highlight the importance of considering acquisition design and parameter estimation in tandem, and with consideration of the latent parameter ranges of interest in a given application.

5.9 Discussion

The single-slice experiments show that PERK can achieve similar WM/GM T_1, T_2 estimation performance as dictionary-based grid search via VPM or iterative optimization via PGPM, but in more than 2 orders of magnitude less time. This acceleration factor will grow to at least 3 orders of magnitude for T_1, T_2 estimation over a typical full imaging volume



	NMR	VPM	PGPM	PERK
V4 T_1	1604 ± 7.2	1645 ± 48	1639 ± 48	1649 ± 51
V5 T_1	1332 ± 0.8	1335 ± 61	1331 ± 41	$1343 \pm 40.$
V6 T_1	1044 ± 3.2	1055 ± 28	$1060. \pm 29$	1083 ± 32
V7 T_1	801.7 ± 1.70	834 ± 21	$840. \pm 23$	821 ± 25
V8 T_1	608.6 ± 1.03	627 ± 25	623 ± 12	604 ± 18
V4 T_2	190.94 ± 0.011	194 ± 5.5	193.1 ± 5.2	197 ± 11
V5 T_2	133.27 ± 0.073	131.2 ± 5.3	131 ± 5.5	138 ± 8
V6 T_2	96.89 ± 0.049	90.8 ± 3.5	90.8 ± 3.5	106.6 ± 3.6
V7 T_2	64.07 ± 0.034	64.6 ± 2.2	64.5 ± 2.1	89.2 ± 3.7
V8 T_2	46.42 ± 0.014	46.4 ± 1.5	46.4 ± 1.5	48.9 ± 4.6

Figure 5.10: Phantom sample statistics of VPM, PGPM, and PERK T_1, T_2 estimates and NIST NMR reference measurements [2] in the case of strong mismatch between acquisition design versus PERK training distribution support. Unlike analogous results in Fig. 5.7, here the PERK estimator was trained with a sampling distribution whose support extended well beyond the range of T_1, T_2 values for which the acquisition was optimized. Comparing to Fig. 5.7, we find that PERK estimator performance degrades within the highlighted T_1, T_2 range of interest. Plot markers and error bars indicate sample means and sample standard deviations computed over ROIs within the 14 vials labeled and color-coded in Fig. 5.11. Corresponding tables replicate sample means $\pm$ sample standard deviations for vials within the highlighted range. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. All values are in milliseconds.

(because PERK training time scales negligibly with the number of voxels) and may grow even higher for full-volume parameter estimation in problems involving more unknowns per voxel (see [9] for a demonstration in simulation). Even with recent low-rank dictionary approximations [109, 112, 114, 115] dictionary-based methods are unlikely to achieve the

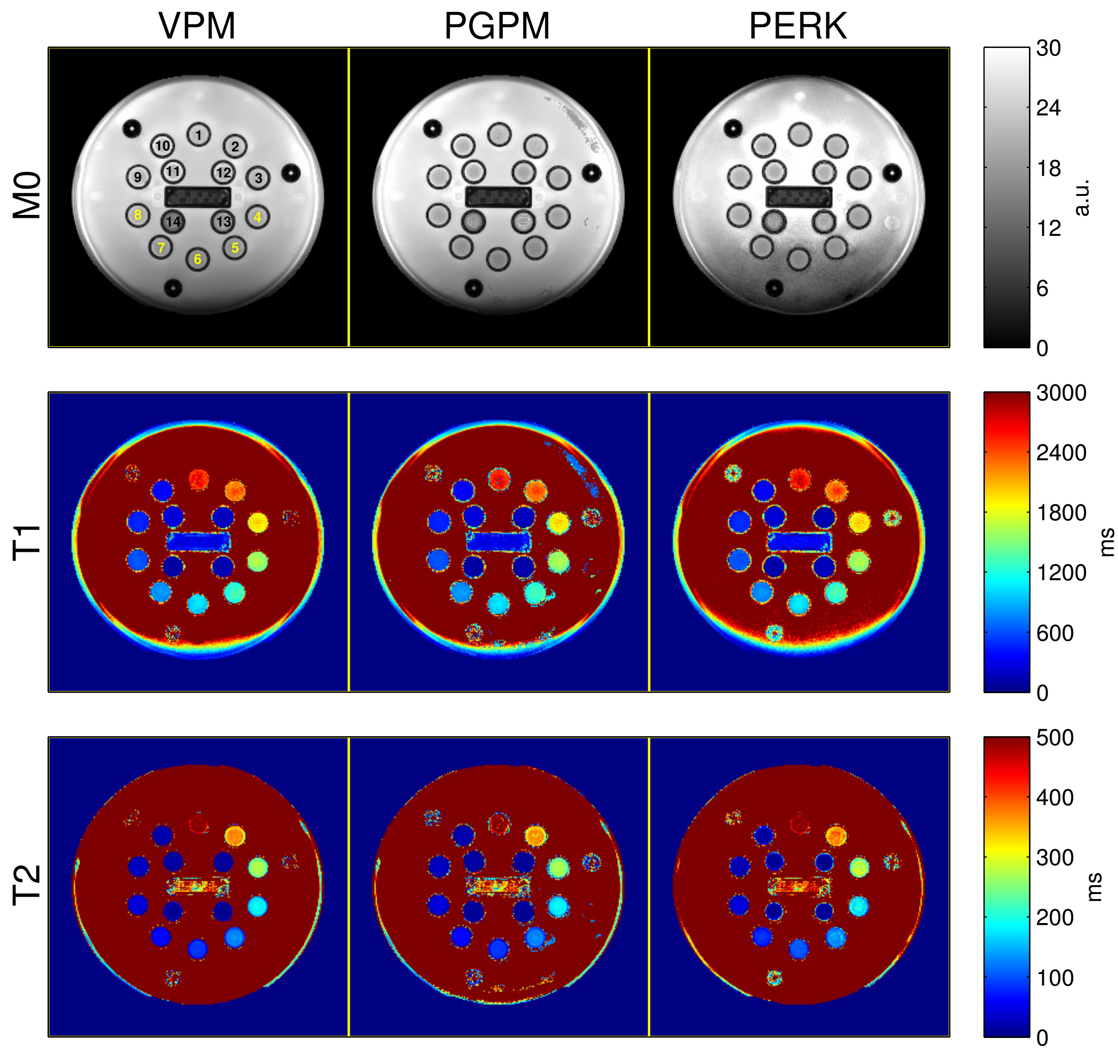


Figure 5.11: More aggressively trained VPM, PGPM, and PERK m_0, T_1, T_2 estimates in a quantitative phantom. Here the PERK estimator was trained with a sampling distribution whose support extended over less well identified T_1, T_2 values. Comparing with analogous images in Fig. 5.6, PERK performance within vials 4-8 degrades, though in other vials performance clearly improves. Vials are enumerated and highlighted to correspond with markers and colored boxes in Fig. 5.10.

large-scale speed of PERK.

PERK also handles known parameters $\boldsymbol{\nu}$ more naturally than does dictionary-based grid search. Grid search necessitates pre-clustering $\boldsymbol{\nu}$ voxel values and generating one dictionary per cluster; however, it is in general unclear *a priori* how many clusters are needed to balance accuracy and computation. In contrast, PERK simply considers the coordinates of each $\boldsymbol{\nu}$ sample as additional regressor dimensions. As the Gaussian PERK estimator is continuous in $\boldsymbol{\nu}$ (and $\boldsymbol{\alpha}$), Gaussian PERK does not suffer from either cluster (or grid) quantization bias.

Interestingly, PERK storage requirements grow more directly with regressor dimension Q than with regressand dimension L. Using formulas for rank-one covariance matrix updates, constructing $\widehat{\mathbf{x}}(\cdot)$ element-wise via L evaluations of (5.23) can be implemented to use $O(Z^2)$ memory units when $\rho_l \leftarrow \rho \; \forall l \in \{1, \dots, L\}$ (as recommended in Subsection 5.5.2.2). Direct application of [133, Proposition 4] to the case of Gaussian kernel (5.12) reveals that Z should be scaled subquadratically but superlinearly with Q to conservatively maintain a given threshold of maximal kernel approximation error. Thus, PERK memory requirements need grow no faster than $O(Q^4)$ to maintain a given level of kernel approximation error.

The $O(Q^4)$ PERK memory requirement ensures improvement over large-scale grid search in modestly overdetermined estimation problems, *i.e.* when $Q \approx L$. In applications where the number of measurements far exceeds L (*e.g.*, MR fingerprinting [103]), PERK may still provide performance gains if images are projected [109] or directly reconstructed [114] into a low-dimensional measurement subspace prior to per-voxel processing. Using this idea, we recently applied PERK to MR fingerprinting in [1].

Phantom experiments most clearly demonstrate that while PERK T_1, T_2 estimates are accurate within a properly selected training range, PERK may extrapolate poorly outside the sampling distribution's support (an improperly selected support can significantly degrade performance; see Subsection 5.8.2 for a demonstration). If more graceful degradation is desired, it may be helpful to additionally fit coefficients of a low-order polynomial and thereby form estimates of form, *e.g.*, $\widehat{x}_l(\mathbf{q}) := \widehat{h}_l(\mathbf{q}) + \widehat{b}_l + \widehat{\mathbf{c}}_l^{\mathsf{T}}\mathbf{q}$. However, greater model complexity may require more training samples to prevent overfitting.

In vivo experiments demonstrated that VPM, PGPM, and PERK T_1, T_2 estimates are overall comparable in WM and GM regions of interest. Nevertheless, small but consistently unidirectional discrepancies persist between the ML and PERK T_1 estimates in WM, one of which is statistically significant. These subtle discrepancies may indicate that ML and PERK estimators behave differently in regions with increased model mismatch. One possible source of *in vivo* model mismatch could be diffusive signal loss, to which DESS is especially sensitive [97, 96]. In particular, unaccounted diffusive signal loss could reduce the DESS second echo's already low SNR in WM to a point where non-Gaussian noise statistics become important to consider. Whereas PERK was trained with simulated data corrupted by Rician-distributed noise, the ML estimators used in this work take a (standard) Gaussian noise assumption and may thus be more prone than PERK to noise-related bias at low SNR. Taking these statements together, unaccounted diffusive effects might bias Gaussian ML estimators more than a properly trained PERK estimator and might explain minor discrepancies between ML and PERK T_1 estimates in WM.

The present formulation constructs separate scalar estimators for each coordinate of $\widehat{\mathbf{x}}$. A natural extension might instead seek to construct vector estimators that consist of linear combinations of vector features that reside in an RKHS of vector-valued functions (see [140] for a review). Here, the associated reproducing kernel would now be matrix-valued and might encode expected dependencies among the outputs of $\widehat{\mathbf{x}}$. With enough training points, the resulting vector estimator could achieve improved estimator performance in terms of accuracy and precision, at the expense of tuning more model parameters and increased computational burden.

In this work, we trained PERK using simulated training data corrupted by noise realizations drawn from a single noise distribution, whose statistics were estimated once from background regions of unlabeled test image data. This training strategy produced reasonable results perhaps in part because our experiments used fully-sampled Cartesian data, for which coil-combined images exhibit little spatial variation in the noise distribution due to receive coil sensitivity spatial variation [141]. To apply PERK in applications where input measurement images exhibit large spatial variation in the noise variance (*e.g.*, multiple-coil acquisitions with parallel imaging acceleration), it may be advantageous to train PERK using simulated training data corrupted by noise realizations drawn from an appropriate distribution over noise distributions. If noise variance maps are available, one could alternately train several PERK estimators with training datasets corrupted by different amounts of noise and apply each estimator to correspondingly noisy measurement image voxels.

Because there is ambiguity in MR data scale due to receive gains and other amplitude scaling factors, it is desirable to construct an estimator that is unaffected by changes in data scale between training and testing. In experiments, we address scaling ambiguity by setting the marginal m_0 sampling distribution p_{m_0} based on test measurements, thereby matching simulated training measurement scale to test measurement scale. This strategy would require retraining between acquisitions that are different in scale but are otherwise identical, which may be undesirable in practice. Alternatively, one could preprocess each noisy training regressor and each noisy test measurement by rescaling each such that (without loss of generality) its first entry is unity, is subsequently uninformative, and can thus be safely pruned to reduce problem dimensionality. Training and testing estimators (for latent parameters other than m_0) using these preprocessed regressors and test points is then largely invariant to the support of p_{m_0} [1]. One drawback to this approach is that normalization by noisy training regressors and test measurements could increase estimation variance.

As explained further in Subsection 5.5.2, we chose to train PERK after observation of unlabeled test data, a strategy that permits automatic selection of some tuning parameters but requires training at test time. Other applications may require many more training points than was required in our experiments for reasonable PERK performance, in which case such online training might be less practical. Using our PERK implementation, offline training would require additional selection of test measurement scale, known object parameter distribution p_ν, and noise variance σ^2. Test measurement scale selection could be avoided using the scale-invariant training strategy discussed in the previous paragraph. As emphasized in Subsection 5.5.2.1 and demonstrated in Subsection 5.8.2, PERK performance is quite sensitive to the object parameter distribution's support, and so at least the support of p_ν would need to be carefully selected based on separate prior parameter estimates or problem-specific intuition. As demonstrated in Subsection 5.8.1, PERK performs best when training and testing data noise statistics coincide but degrades gracefully with mild levels of mismatch, so σ^2 could be selected based on separate SNR approximations.

As an alternative to PERK, researchers have recently proposed MRI parameter estimation via deep neural network learning [142, 143]. Deep learning requires enormous numbers of training points to train many model parameters without overfitting, and its limited theoretical basis renders its practical use largely an art. Here, we have introduced and investigated PERK with an emphasis on its simplicity and its relatively intuitive model selection (see Subsection 5.5.2); a thorough comparison with deep learning is a possible topic for future work.

5.10 Conclusion

This paper has introduced PERK, a fast and general method for dictionary-free MRI parameter estimation. PERK first uses prior parameter/noise distributions and a general nonlinear MR signal model to simulate many parameter-measurement training points and then constructs a nonlinear regression function from these training points using linear combinations of nonlinear kernels. We have demonstrated PERK for T_1, T_2 estimation from optimized SPGR/DESS acquisitions [6], a simple application where it is straightforward to validate PERK estimates against gold-standard VPM estimates, iterative PGPM estimates, and NIST reference measurements. Numerical simulations showed that PERK achieves T_1, T_2 RMSE comparable to VPM and PGPM in WM- and GM-like voxels. Phantom experiments showed that within a properly chosen sampling distribution support, VPM, PGPM, and PERK estimates agree excellently with each other and reasonably with NIST NMR measurements. *In vivo* experiments showed that VPM, PGPM, and PERK produce comparable T_1 estimates and nearly indistinguishable T_2 estimates in WM and GM ROIs. PERK used identical model selection parameters across all simulations and experiments and consistently provided at least a 140$\times$ acceleration over VPM and PGPM. This acceleration factor may increase by several orders of magnitude for estimation problems involving more latent parameters per voxel [9, 11].

CHAPTER 6

Fast Myelin Water Imaging via Acquisition Design and PERK

6.1 Introduction

This chapter[1] adapts acquisition design (developed in Ch. 4) and applies parameter estimation (developed in Ch. 5) to design a new method for imaging a specific MR biomarker of clinical interest. In particular, we study a biomarker for *myelin content* in the human brain.

Myelin is a lipid-rich material that forms an insulating sheath encasing neuronal axons predominantly in white matter (WM) regions of the human brain [144]. Demyelination (*i.e.*, myelin loss) is central to the development of several neurodegenerative disorders such as multiple sclerosis (MS) [145]. Non-invasive myelin quantification in WM is thus desirable for monitoring the onset and progression of neurodegenerative disease.

MR relaxation time constants (especially spin-spin time constant T_2) depend on the macromolecular environment surrounding excited water molecules. In nerve tissue, these environments vary spatially on scales much smaller than the millimeter-scale resolutions used in typical MR imaging experiments. Thus, there is significant variation of relaxation times within a typical imaging voxel containing nerve tissue.

Many researchers have attempted to characterize tissue microstructure by estimating the distribution of MR relaxation time constants and associating certain ranges of time constants with particular "compartments" or "pools" of water molecules that exist in similar macromolecular environments. *In vitro* NMR studies of nerve animal tissue ascribed a fast-relaxing[2] water compartment with T_2 ~10-40ms initially to general protein and phospholipid structures [146] and later more specifically to water trapped between the phos-

[1] This chapter partially derives content from in-preparation journal paper [10] that extends conference paper [11].

[2] The fast-relaxing compartment bears its name with reference to the portion of the T_2 distribution in water (from about 10ms to at least 1000ms) that is typically observable in MRI.

pholipid bilayers of myelin [147, 148]. Shortly thereafter, the first MR images of so-called *myelin water fraction* (MWF), defined as the proportion of MR signal arising from the fast-relaxing water compartment relative to total MR signal, were demonstrated *in vivo* in the human brain [100]. More recently, MWF has been shown to correlate well with histological measurements of myelin content in animal models of nerve injury [149] and demyelination [150]. In humans, MWF has been measured to be significantly lower in "normally appearing" WM of MS patients versus controls [151], and to correlate strongly with post-mortem histological measurements of myelin content in MS patients [152]. Thus, there is strong evidence that MWF as defined in [100] is a specific biomarker for WM myelin content.

All of the aforementioned studies estimate MWF images from a multi-echo spin echo (MESE) MRI pulse sequence [51] with long repetition time $T_\mathrm{R} \geq 2$s to ensure sufficient recovery of the longitudinal magnetization in nerve tissue. Whole-brain MWF imaging using such long-T_R MESE acquisitions at a typical imaging resolution would require hours of scan time. To enable more clinically practical scan times, researchers have more recently shown that MESE-based MWF imaging can be accelerated without significantly changing the resulting MWF images [153, 154] by acquiring multiple gradient echoes per refocusing pulse [155]. However, these and other acquisition modifications used in [154] do not address the fundamental long-T_R requirement of MESE acquisitions and thus would still require long scan times for whole-brain MWF imaging at millimeter-scale resolution. Furthermore, estimating a T_2 distribution from MESE data constitutes a poorly-conditioned estimation problem that continues to demand high SNR [156, 157], so the need remains for a more SNR-efficient acquisition for myelin water imaging. As an alternative to MESE acquisitions, scan profiles consisting of short-T_R steady-state (SS) sequences were proposed for whole-brain myelin water imaging in about 30m scan time [121]. Despite recent further refinements [158, 101], myelin water images from SS pulse sequences have thus far been shown to be incomparable with MWF images from MESE pulse sequences [159], likely due at least in part to insufficient precision [160].

Inspired by [6], we reconsidered myelin water imaging from SS pulse sequences from the perspective of statistical experiment design. In [6], we optimized several combinations of spoiled gradient-recalled echo (SPGR) [19] and dual-echo steady-state (DESS) scans [21, 22] for single-compartment T_1, T_2 estimation and found that different optimized scan combinations gave rise to significantly different *in vivo* T_2 estimates (but comparable phantom T_2 estimates), indicating *in vivo* sensitivity to model nonidealities (see Ch. 4). Further simulation studies suggested that these inconsistencies may be explained by multi-compartmental relaxation. This chapter shows that this apparent SPGR/DESS sensitivity to multi-compartmental relaxation can be exploited for fast myelin water imaging.

This chapter introduces a fast SS MRI acquisition for precise myelin water imaging. The acquisition consists of a combination of SPGR and DESS scans that enables maximally precise estimation of the faster-relaxing signal fraction f_F from two-compartment SPGR/DESS signal models, subject to a scan time constraint. The precision-optimized acquisition is as fast as the SS acquisition proposed in [158] but enables $\sim$40% expected coefficient of variation in unbiased f_F estimates. (Similar calculations for [121, 158] found that f_F coefficients of variation frequently exceeded 100% [160].) To our knowledge, *in vivo* experiments are the first to demonstrate (lateral) WM myelin water content estimates from a fast SS acquisition that are similar to conventional MWF estimates from a slower MESE acquisition.

The remainder of this chapter is organized as follows. Section 6.2 reviews and develops simple two-compartment signal models for SPGR and DESS pulse sequences, respectively. Section 6.3 adapts and then applies QMRI acquisition design (developed in Chapter 4) to design a fast DESS acquisition for precise f_F estimation. Section 6.4 demonstrates myelin water content imaging from a fast precision-optimized DESS acquisition in numerical simulations and *in vivo* experiments. Section 6.5 discusses advantages, challenges, and future work. Section 6.6 summarizes key contributions.

6.2 Multi-Compartmental Models for SS Sequences

This section develops multi-compartmental signal models for the SPGR and DESS pulse sequences. Subsection 6.2.1 reviews and extends a concise Bloch-matrix derivation [161] of an SPGR signal model that accounts for exchange between multiple compartments. Subsection 6.2.2 applies the Bloch-matrix representation to derive analogous (but previously unpublished) multi-compartmental DESS signal models. Though the derivations below focus for simplicity on only two exchanging compartments, the Bloch-matrix formulation generalizes readily to greater numbers of interacting compartments.

6.2.1 A Two-Compartment SPGR Signal Model

The McConnell equations [162] extend the Bloch equations [15] to account for physical exchange[3] between two or more intra-voxel compartments. Here, we consider the interaction

[3]The word "exchange" is often a source of confusion because it is used in different contexts to refer to a variety of transport phenomena. As originally described in [162], *chemical exchange* specifically refers to the rapid reversible transfer of a nucleus between two or more molecular environments (*e.g.*, hydrogen bonding in water or proton exchange between water and a macromolecule). More recently, exchange has also been loosely used to describe several other processes that can be characterized with similar physical equations but where explicit nuclear transfer does not occur. One such process that we denote for clarity as *physical*

of a fast-relaxing water compartment (characterized by comparatively short spin-lattice $T_{1,\mathrm{F}}$ and spin-spin $T_{2,\mathrm{F}}$ relaxation times) with a slow-relaxing water compartment (characterized by longer relaxation times $T_{1,\mathrm{S}}, T_{2,\mathrm{S}}$). In primed coordinates rotating clockwise about the longitudinal z-axis at the Larmor frequency, the dynamics of corresponding fast-relaxing and slow-relaxing compartmental magnetization vectors $\mathbf{m}'_{\mathrm{F}} := \left[m'_{x,\mathrm{F}}, m'_{y,\mathrm{F}}, m'_{z,\mathrm{F}}\right]^{\mathsf{T}}$ and $\mathbf{m}'_{\mathrm{S}} := \left[m'_{x,\mathrm{S}}, m'_{y,\mathrm{S}}, m'_{z,\mathrm{S}}\right]^{\mathsf{T}}$ are coupled via first-order exchange rates $r_{\mathrm{F}\to\mathrm{S}}$ (from fast to slow compartment) and $r_{\mathrm{S}\to\mathrm{F}}$ (vice-versa). During periods when no RF excitation is present, these magnetization dynamics separate in transverse and longitudinal components. The two-compartment transverse equations extend (2.9) to read

$$\frac{\partial}{\partial t} m'_{xy,\mathrm{F}}(\mathbf{r},t) = -i\gamma m'_{xy,\mathrm{F}}(\mathbf{r},t)\omega'_{\mathrm{F}}(\mathbf{r}) - \frac{m'_{xy,\mathrm{F}}(\mathbf{r},t)}{T_{2,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r})m'_{xy,\mathrm{F}}(\mathbf{r},t) + r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})m'_{xy,\mathrm{S}}(\mathbf{r},t); \tag{6.1}$$

$$\frac{\partial}{\partial t} m'_{xy,\mathrm{S}}(\mathbf{r},t) = -i\gamma m'_{xy,\mathrm{S}}(\mathbf{r},t)\omega'_{\mathrm{S}}(\mathbf{r}) - \frac{m'_{xy,\mathrm{S}}(\mathbf{r},t)}{T_{2,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})m'_{xy,\mathrm{S}}(\mathbf{r},t) + r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r})m'_{xy,\mathrm{F}}(\mathbf{r},t), \tag{6.2}$$

where $m'_{xy,\mathrm{F}}(\mathbf{r},t) := m'_{x,\mathrm{F}}(\mathbf{r},t) + im'_{y,\mathrm{F}}(\mathbf{r},t)$ and $m'_{xy,\mathrm{S}}(\mathbf{r},t) := m'_{x,\mathrm{S}}(\mathbf{r},t) + im'_{y,\mathrm{S}}(\mathbf{r},t)$ are complex compartmental representations of the transverse magnetization at position $\mathbf{r}$ and time t; $\omega'_{\mathrm{F}}(\mathbf{r}) \in \mathbb{R}$ and $\omega'_{\mathrm{S}}(\mathbf{r}) \in \mathbb{R}$ allow for compartment-specific but time-invariant off-resonance effects; and $\gamma \in \mathbb{R}$ is the gyromagnetic ratio. During periods when no excitation is present, analogous two-compartment longitudinal equations extend (2.10) to read

$$\frac{\partial}{\partial t} m'_{z,\mathrm{F}}(\mathbf{r},t) = -\frac{m'_{z,\mathrm{F}}(\mathbf{r},t) - f_{\mathrm{F}}(\mathbf{r})m_0(\mathbf{r})}{T_{1,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r})m'_{z,\mathrm{F}}(\mathbf{r},t) + r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})m'_{z,\mathrm{S}}(\mathbf{r},t); \tag{6.3}$$

$$\frac{\partial}{\partial t} m'_{z,\mathrm{S}}(\mathbf{r},t) = -\frac{m'_{z,\mathrm{S}}(\mathbf{r},t) - f_{\mathrm{S}}(\mathbf{r})m_0(\mathbf{r})}{T_{1,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})m'_{z,\mathrm{S}}(\mathbf{r},t) + r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})m'_{z,\mathrm{F}}(\mathbf{r},t), \tag{6.4}$$

where $f_{\mathrm{F}}(\mathbf{r}) \in [0,1]$ and $f_{\mathrm{S}}(\mathbf{r}) \in [0,1]$ denote fast- and slow-relaxing compartmental fractions and $m_0(\mathbf{r})$ denotes the total[4] equilibrium magnetization.

Equations (6.1)-(6.4) comprise a first-order affine dynamical system and can be equiv-

exchange involves the transfer of intact molecules between different compartments (*e.g.*, water molecules across a membrane). In the ensuing derivations, we refer specifically to physical exchange but note that the McConnell equations describe chemical exchange as well.

[4]The numerous parameters introduced here are often interdependent. In regions with only two water compartments present, $f_{\mathrm{F}}(\mathbf{r}) + f_{\mathrm{S}}(\mathbf{r}) = 1$. In regions where only two water pools exhibit exchange and that exchange is in chemical equilibrium, $f_{\mathrm{F}}(\mathbf{r})r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) = f_{\mathrm{S}}(\mathbf{r})r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})$.

alently and concisely written in matrix form as

$$\frac{\partial}{\partial t}\mathbf{m}'(\mathbf{r},t) = \mathbf{A}(\mathbf{r})\mathbf{m}'(\mathbf{r},t) + \mathbf{c}(\mathbf{r}), \tag{6.5}$$

where $\mathbf{m}'(\mathbf{r},t) := \operatorname{vec}\Big([\mathbf{m}'_{\mathrm{F}}(\mathbf{r},t), \mathbf{m}'_{\mathrm{S}}(\mathbf{r},t)]^{\mathsf{T}}\Big) \in \mathbb{R}^6$ collects the compartmental magnetization vectors and $\operatorname{vec}(\cdot)$ denotes vectorization. Here, system matrix $\mathbf{A}(\mathbf{r}) \in \mathbb{R}^{6\times 6}$ admits a block diagonal form[5] $\mathbf{A}(\mathbf{r}) \equiv [\mathbf{A}_{xy}(\mathbf{r}), \mathbf{0}_{4\times 2}; \mathbf{0}_{2\times 4}, \mathbf{A}_z(\mathbf{r})]$, with $\mathbf{A}_{xy}(\mathbf{r}) :=$

$$\begin{bmatrix} -\frac{1}{T_{2,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) & \omega'_{\mathrm{F}}(\mathbf{r}) & 0 \\ r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & -\frac{1}{T_{2,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) & 0 & \omega'_{\mathrm{S}}(\mathbf{r}) \\ -\omega'_{\mathrm{F}}(\mathbf{r}) & 0 & -\frac{1}{T_{2,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \\ 0 & -\omega'_{\mathrm{S}}(\mathbf{r}) & r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & -\frac{1}{T_{2,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \end{bmatrix} \tag{6.6}$$

collecting transverse dynamics;

$$\mathbf{A}_z(\mathbf{r}) := \begin{bmatrix} -\frac{1}{T_{1,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \\ r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & -\frac{1}{T_{1,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \end{bmatrix} \tag{6.7}$$

collecting longitudinal dynamics; and $\mathbf{c}(\mathbf{r}) := \Big[0,0,0,0,\frac{f_{\mathrm{F}}(\mathbf{r})m_0(\mathbf{r})}{T_{1,\mathrm{F}}(\mathbf{r})},\frac{f_{\mathrm{S}}(\mathbf{r})m_0(\mathbf{r})}{T_{1,\mathrm{S}}(\mathbf{r})}\Big]^{\mathsf{T}}$. If $\mathbf{A}(\mathbf{r})$ is invertible, a matrix exponential solution to (6.5) exists and reads

$$\mathbf{m}'(\mathbf{r},t) = e^{(t-t_0)\mathbf{A}(\mathbf{r})}\mathbf{m}'(\mathbf{r},t_0) + \big(e^{(t-t_0)\mathbf{A}(\mathbf{r})} - \mathbf{I}_6\big)\big(\mathbf{A}(\mathbf{r})\big)^{-1}\mathbf{c}(\mathbf{r}) \qquad \forall t \geq t_0, \tag{6.8}$$

where $\mathbf{m}'(\mathbf{r},t_0)$ is the magnetization at an initial time t_0 and $\mathbf{I}_6 \in \mathbb{R}^{6\times 6}$ is an identity matrix. In chemical equilibrium (described in Footnote 4), a direct calculation involving compartmental equilibrium magnetization $\mathbf{m}_0(\mathbf{r}) := [0,0,0,0,f_{\mathrm{F}}(\mathbf{r})m_0(\mathbf{r}),f_{\mathrm{S}}(\mathbf{r})m_0(\mathbf{r})]^{\mathsf{T}}$ reveals that $\mathbf{A}(\mathbf{r})\mathbf{m}_0(\mathbf{r}) = -\mathbf{c}(\mathbf{r})$, in which case (6.8) simplifies to

$$\mathbf{m}'(\mathbf{r},t) = e^{(t-t_0)\mathbf{A}(\mathbf{r})}\mathbf{m}'(\mathbf{r},t_0) + \big(\mathbf{I}_6 - e^{(t-t_0)\mathbf{A}(\mathbf{r})}\big)\mathbf{m}_0(\mathbf{r}) \qquad \forall t \geq t_0, \tag{6.9}$$

a form that resembles the matrix operator solution (2.14) to the single-compartment Bloch equations in the presence of only free precession and relaxation.

[5] For conciseness we use here the MATLAB®-like shorthand

$$[\mathbf{M}_{1,1},\mathbf{M}_{1,2};\mathbf{M}_{2,1},\mathbf{M}_{2,2}] \equiv \begin{bmatrix} \mathbf{M}_{1,1}\mathbf{M}_{1,2} \\ \mathbf{M}_{2,1}\mathbf{M}_{2,2} \end{bmatrix}$$

for all appropriately sized submatrices $\mathbf{M}_{1,1},\mathbf{M}_{1,2},\mathbf{M}_{2,1},\mathbf{M}_{2,2}$.

To develop a two-compartment SPGR signal model, we modify the derivation of the single-compartment SPGR model (presented in Subsection 2.1.2.1) to account for inter-compartmental exchange (via (6.9)) during intervals of free precession and relaxation. As before, let $\mathbf{m}'(\mathbf{r}, t_0)$ denote the magnetization at an initial time t_0 selected well into the steady-state and just prior to RF excitation. The SPGR sequence first applies an RF excitation with pulse duration T_P. If we neglect relaxation, off-resonance, and exchange during excitation (which is reasonable for sufficiently short T_P) and further assume that each compartment experiences this excitation with equal transmit field sensitivity, we may model RF excitation as a rotation by a compartment-wise constant (but spatially varying) nutation angle $\alpha(\mathbf{r}, t_0 + T_\mathrm{P}; t_0)$ (defined in (2.7)), abbreviated $\alpha(\mathbf{r})$ hereafter. For clockwise rotations about the x'-axis, such near-instantaneous rotation may be represented as

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r}))\mathbf{m}'(\mathbf{r}, t_0), \tag{6.10}$$

where $\mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r})) := \mathbf{R}_{x'}(\alpha(\mathbf{r})) \otimes \mathbf{I}_2 \in \mathbb{R}^{6\times 6}$ is a compartment-wise rotation operation; $\mathbf{R}_{x'}(\alpha(\mathbf{r}))$ is defined in (2.6); and $\otimes$ denotes the Kronecker product.

The compartments exchange while their magnetization vectors precess and relax as per (6.8) until data acquisition at echo time $T_\mathrm{E} \in \left[\frac{T_\mathrm{P}}{2}, T_\mathrm{R}\right]$ after the midpoint of RF excitation:

$$\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) = e^{\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})}\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) + \left(\mathbf{I}_6 - e^{\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})}\right)\mathbf{m}_0(\mathbf{r}). \tag{6.11}$$

Following signal reception (which we assume as in Footnote 3 of Chapter 2 to negligibly influence magnetization), the SPGR sequence spoils the remaining transverse magnetization in both compartments while leaving unaffected the longitudinal magnetization components in each compartment. We model ideal spoiling in both compartments as

$$\mathbf{S}^{\otimes 2}\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right), \tag{6.12}$$

where $\mathbf{S}^{\otimes 2} := \mathbf{S} \otimes \mathbf{I}_2 \in \mathbb{R}^{6\times 6}$ is a compartment-wise spoiling operator and $\mathbf{S} \in \mathbb{R}^{3\times 3}$ is the diagonal binary matrix defined in (2.20). After spoiling, compartments continue to exchange while their longitudinal magnetization components partially recover until $t \leftarrow$

$t_0 + T_\mathrm{R}$, completing one repetition cycle:

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{R}) = e^{\left(T_\mathrm{R} - \left(\frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right)\right)\mathbf{A}(\mathbf{r})} \mathbf{S}^{\otimes 2} \mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) + \left(\mathbf{I}_6 - e^{\left(T_\mathrm{R} - \left(\frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right)\right)\mathbf{A}(\mathbf{r})}\right)\mathbf{m}_0(\mathbf{r}). \tag{6.13}$$

In steady-state, one cycle of excitation, acquisition, spoiling, and recovery returns the magnetization back to its initial state. We enforce this through the steady-state condition

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r}))\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{R}), \tag{6.14}$$

which yields an algebraic system of equations. When it exists, the solution is

$$\begin{aligned}
\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) &= \left(\mathbf{I}_6 - \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r}))\mathbf{S}^{\otimes 2} e^{(T_\mathrm{R} - T_\mathrm{P})\mathbf{A}(\mathbf{r})}\right)^{-1} \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r})) && (6.15)\\
&\times \left(\mathbf{I}_6 - \mathbf{S}^{\otimes 2} e^{(T_\mathrm{R} - T_\mathrm{P})\mathbf{A}(\mathbf{r})} - e^{\left(T_\mathrm{R} - \left(\frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right)\right)\mathbf{A}(\mathbf{r})}\left(\mathbf{I}_6 - \mathbf{S}^{\otimes 2}\right)\right)\mathbf{m}_0(\mathbf{r})\\
&= \left(\mathbf{I}_6 - \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r}))\mathbf{S}^{\otimes 2} e^{(T_\mathrm{R} - T_\mathrm{P})\mathbf{A}(\mathbf{r})}\right)^{-1} \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r})) && (6.16)\\
&\times \left(\mathbf{I}_6 - \mathbf{S}^{\otimes 2} e^{(T_\mathrm{R} - T_\mathrm{P})\mathbf{A}(\mathbf{r})}\right)\mathbf{m}_0(\mathbf{r}),
\end{aligned}$$

where (6.15) is due to straightforward matrix operations and (6.16) makes use of the ideal spoiling property $(\mathbf{I}_6 - \mathbf{S}^{\otimes 2})\mathbf{m}_0(\mathbf{r}) = \mathbf{0}_6 \forall \mathbf{r}$. Substituting (6.16) into (6.11) yields an expression for the compartmental magnetization at the echo time.

The matrix exponentials in (6.16) would be cumbersome to expand in general but fortunately always arise prepended by spoiling matrix $\mathbf{S}^{\otimes 2}$. The combined form reduces to

$$\mathbf{S}^{\otimes 2} e^{(T_\mathrm{R} - T_\mathrm{P})\mathbf{A}(\mathbf{r})} \equiv \left[\mathbf{0}_{4\times 4}, \mathbf{0}_{4\times 2}; \mathbf{0}_{2\times 4}, e^{(T_\mathrm{R} - T_\mathrm{P})\mathbf{A}_z(\mathbf{r})}\right] \tag{6.17}$$

and admits an explicit representation for (6.16) that we find via standard solvers[6]. Note that if RF pulses are assumed instantaneous (*i.e.*, $T_\mathrm{P} \leftarrow 0$), then the longitudinal components of (6.16) are equivalent to [161, Eq. 34]. Another instructive (and much simpler) special case

[6] Since we require many signal evaluations for acquisition design and parameter estimation, it is well worthwhile from a computational standpoint to symbolically solve algebraic systems of equations whenever possible. However, for complicated equation systems such as (6.16), the manipulations required are tedious and error-prone. In MATLAB®, we used the Symbolic Toolbox and the function `matlabFunction` to symbolically solve (6.16) and automatically generate memory-friendly function handles.

neglects exchange (*i.e.*, $r_{\text{F}\to\text{S}} \leftarrow 0$ and $r_{\text{S}\to\text{F}} \leftarrow 0$), in which case (6.16) reduces to

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\text{P}) = \begin{bmatrix} 0 \\ 0 \\ \frac{f_\text{F}(\mathbf{r}) \sin(\alpha(\mathbf{r}))\left(1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{F}}(\mathbf{r})}\right)}{1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{F}}(\mathbf{r})}\cos(\alpha(\mathbf{r}))} \\ \frac{f_\text{S}(\mathbf{r}) \sin(\alpha(\mathbf{r}))\left(1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{S}}(\mathbf{r})}\right)}{1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{S}}(\mathbf{r})}\cos(\alpha(\mathbf{r}))} \\ \frac{f_\text{F}(\mathbf{r}) \cos(\alpha(\mathbf{r}))\left(1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{F}}(\mathbf{r})}\right)}{1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{F}}(\mathbf{r})}\cos(\alpha(\mathbf{r}))} \\ \frac{f_\text{S}(\mathbf{r}) \cos(\alpha(\mathbf{r}))\left(1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{S}}(\mathbf{r})}\right)}{1-e^{-(T_\text{R}-T_\text{P})/T_{1,\text{S}}(\mathbf{r})}\cos(\alpha(\mathbf{r}))} \end{bmatrix} m_0(\mathbf{r}), \tag{6.18}$$

a two (non-exchanging) compartment extension to the SPGR steady-state solution (2.23).

The received signal is approximately proportional to the integrated transverse magnetization arising from both compartments. We take further assumptions to derive expressions, the first two of which are two-compartment extensions of assumptions used in §2.1.2.1:

1. We assume that the signal is localized to a scale over which there is within-voxel variation of each compartment's off-resonance frequency, but minimal intra-voxel variation of other space-varying parameters $m_0, f_\text{F}, f_\text{S}, T_{1,\text{F}}, T_{1,\text{S}}, T_{2,\text{F}}, T_{2,\text{S}}, \alpha, r_{\text{F}\to\text{S}}, r_{\text{S}\to\text{F}}$. This assumption effectively fixes all compartmental properties other than ω'_F and ω'_S over the volume $\mathbb{V}$ of a sufficiently small voxel.
2. We assume that the fast-relaxing and slow-relaxing compartments' off-resonance frequencies are independently distributed within a localized voxel with marginal distributions $\mathsf{p}_{\omega'_\text{F}} := \text{Cauchy}\left(\bar{\omega}'_\text{F}, R'_{2,\text{F}}\right)$ and $\mathsf{p}_{\omega'_\text{S}} := \text{Cauchy}\left(\bar{\omega}'_\text{S}, R'_{2,\text{S}}\right)$, where $\bar{\omega}'_\text{F}, \bar{\omega}'_\text{S}$ are median off-resonance frequencies and $R'_{2,\text{F}}, R'_{2,\text{S}}$ are broadening bandwidths.
3. We assume for short echo time T_E that minimal exchange occurs between excitation and signal reception. We use this strong assumption only to facilitate first-order expansion of the matrix exponential in (6.20), thereby separating broadening integrals by compartment. Without this assumption, (6.20) remains valid.

With these assumptions, the noiseless two-compartment steady-state SPGR signal model for a voxel centered at position $\mathbf{r}$ and with volume defined by $\mathbb{V}(\mathbf{r})$ is (to within constants):

$$s_\mathrm{S}\left(\mathbf{r}, t_0+\frac{T_\mathrm{P}}{2}+T_\mathrm{E}\right) \propto \int_{\mathbb{V}(\mathbf{r})} [1,1,i,i,0,0]\mathbf{m}'\left(\mathbf{r}, t_0+\frac{T_\mathrm{P}}{2}+T_\mathrm{E}\right) \mathrm{d}^3\,\mathbf{r} \tag{6.19}$$

$$\approx \left(\int_{\mathbb{R}^2} [1,1,i,i,0,0] e^{\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})} \mathsf{p}_{\omega'_\mathrm{F}}(\omega'_\mathrm{F})\mathsf{p}_{\omega'_\mathrm{S}}(\omega'_\mathrm{S})\,\mathrm{d}\,\omega'_\mathrm{F}\,\mathrm{d}\,\omega'_\mathrm{S}\right)\mathbf{m}'(\mathbf{r}, t_0+T_\mathrm{P}) \tag{6.20}$$

$$\approx \begin{bmatrix} e^{-\left(1/T_{2,\mathrm{F}}(\mathbf{r})+R'_{2,\mathrm{F}}(\mathbf{r})+i\bar{\omega}'_\mathrm{F}(\mathbf{r})\right)\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)} \\ e^{-\left(1/T_{2,\mathrm{S}}(\mathbf{r})+R'_{2,\mathrm{S}}(\mathbf{r})+i\bar{\omega}'_\mathrm{S}(\mathbf{r})\right)\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)} \\ ie^{-\left(1/T_{2,\mathrm{F}}(\mathbf{r})+R'_{2,\mathrm{F}}(\mathbf{r})+i\bar{\omega}'_\mathrm{F}(\mathbf{r})\right)\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)} \\ ie^{-\left(1/T_{2,\mathrm{S}}(\mathbf{r})+R'_{2,\mathrm{S}}(\mathbf{r})+i\bar{\omega}'_\mathrm{S}(\mathbf{r})\right)\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)} \\ 0 \\ 0 \end{bmatrix}^\mathsf{T} \mathbf{m}'(\mathbf{r}, t_0+T_\mathrm{P}) \tag{6.21}$$

$$= m'_{xy,\mathrm{F}}(\mathbf{r}, t_0+T_\mathrm{P})e^{-\left(1/T_{2,\mathrm{F}}(\mathbf{r})+R'_{2,\mathrm{F}}(\mathbf{r})+i\bar{\omega}'_\mathrm{F}(\mathbf{r})\right)\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)} + m'_{xy,\mathrm{S}}(\mathbf{r}, t_0+T_\mathrm{P})e^{-\left(1/T_{2,\mathrm{S}}(\mathbf{r})+R'_{2,\mathrm{S}}(\mathbf{r})+i\bar{\omega}'_\mathrm{S}(\mathbf{r})\right)\left(T_\mathrm{E}-\frac{T_\mathrm{P}}{2}\right)}. \tag{6.22}$$

Observe that (6.20) uses Assumption 1 to leave $\mathbf{m}'(\mathbf{r}, t_0+T_\mathrm{P})$ outside of broadening integrals defined by Assumption 2 and (6.21) uses Assumption 3 to expand the matrix exponential and evaluate broadening integrals component-wise. Eq. (6.22) naturally extends the single-compartment SPGR model (2.27) and shows clearly that compartmental signals simply add if exchange between excitation and signal reception is neglected.

Though the manipulations leading to (6.22) require strong assumptions, they serve to demonstrate the nontrivial two-compartment SPGR signal dependence on off-resonance distributions. Unlike in the single-compartment case, the signal decay and dephasing terms due to off-resonance effects $e^{-\left(R'_{2,\mathrm{F}}+i\bar{\omega}'_\mathrm{F}\right)T_\mathrm{E}}$ and $e^{-\left(R'_{2,\mathrm{S}}+i\bar{\omega}'_\mathrm{S}\right)T_\mathrm{E}}$ may *not* be absorbed into an apparent spin density (in fact, to first order these terms would affect the apparent compartmental fractions of typical interest). Since off-resonance distributions do often differ significantly in cerebral tissue [163, 164], accurate *in vivo* parameter estimation from this model requires either joint estimation of compartmental distributions or acquisition with echo times much shorter than compartmental broadening timescale differences.

6.2.2 A Two-Compartment DESS Signal Model

To develop two-compartment DESS signal models, we first adapt the McConnell equation solutions of Subsection 6.2.1 to describe compartmental magnetization evolution in the presence of time-dependent field inhomogeneities (such as the unbalanced dephasing gradient used in DESS). We then modify corresponding derivations of single-compartment DESS models (presented in Subsection 2.1.2.2), to account for inter-compartmental exchange during intervals of free precession and relaxation.

As discussed in Subsection 2.1.2.2, the DESS pulse sequence interlaces fixed RF excitations with fixed dephasing gradients to produce two distinct signals per excitation. Because these dephasing gradients create time-dependence in compartmental off-resonance effects, DESS signal dynamics cannot in general be described by matrix exponential solution (6.9). Instead, the dynamics (during periods without excitation) arise as a solution to

$$\frac{\partial}{\partial t}\mathbf{m}'(\mathbf{r},t) = \mathbf{A}(\mathbf{r},t)\mathbf{m}'(\mathbf{r},t) + \mathbf{c}(\mathbf{r}), \tag{6.23}$$

where $\mathbf{A}(\mathbf{r},t) \equiv [\mathbf{A}_{xy}(\mathbf{r},t), \mathbf{0}_{4\times 2}; \mathbf{0}_{2\times 4}, \mathbf{A}_z(\mathbf{r})]$ contains now time-dependent compartmental off-resonance effects $\omega'_{\mathrm{F}}(\mathbf{r},t), \omega'_{\mathrm{S}}(\mathbf{r},t)$ that appear only in submatrix $\mathbf{A}_{xy}(\mathbf{r},t) :=$

$$\begin{bmatrix} -\frac{1}{T_{2,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) & \omega'_{\mathrm{F}}(\mathbf{r},t) & 0 \\ r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & -\frac{1}{T_{2,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) & 0 & \omega'_{\mathrm{S}}(\mathbf{r},t) \\ -\omega'_{\mathrm{F}}(\mathbf{r},t) & 0 & -\frac{1}{T_{2,\mathrm{F}}(\mathbf{r})} - r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \\ 0 & -\omega'_{\mathrm{S}}(\mathbf{r},t) & r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) & -\frac{1}{T_{2,\mathrm{S}}(\mathbf{r})} - r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \end{bmatrix}. \tag{6.24}$$

If $\mathbf{A}(\mathbf{r},t)$ is invertible, an exponential series solution to (6.23) exists and could be expressed exactly using the Magnus expansion [165]. Here, we utilize a first-order Magnus expansion that in chemical equilibrium yields a simple approximate[7] solution for all $t \geq t_0$:

$$\mathbf{m}'(\mathbf{r},t) \approx e^{\int_{t_0}^{t} \mathbf{A}(\mathbf{r},t')\,\mathrm{d}t'}\mathbf{m}'(\mathbf{r},t_0) + \left(e^{\int_{t_0}^{t} \mathbf{A}(\mathbf{r},t')\,\mathrm{d}t'} - \mathbf{I}_6\right)\mathbf{m}_0(\mathbf{r}). \tag{6.27}$$

Higher-order expansions would better capture compartmental magnetization interactions due to off-resonance effects and are not considered hereafter for simplicity.

[7] Approximation (6.27) is exact if commutation relation

$$\mathbf{A}(\mathbf{r},t_1)\mathbf{A}(\mathbf{r},t_2) = \mathbf{A}(\mathbf{r},t_2)\mathbf{A}(\mathbf{r},t_1)$$

holds for all $t_1, t_2 \geq t_0$, which in turn holds pointwise if and only if

$$r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r})(\omega'_{\mathrm{F}}(\mathbf{r},t_1) - \omega'_{\mathrm{S}}(\mathbf{r},t_1) - (\omega'_{\mathrm{F}}(\mathbf{r},t_2) - \omega'_{\mathrm{S}}(\mathbf{r},t_2))) = 0; \tag{6.25}$$
$$r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r})(\omega'_{\mathrm{F}}(\mathbf{r},t_1) - \omega'_{\mathrm{S}}(\mathbf{r},t_1) - (\omega'_{\mathrm{F}}(\mathbf{r},t_2) - \omega'_{\mathrm{S}}(\mathbf{r},t_2))) = 0. \tag{6.26}$$

Conditions (6.25)-(6.26) hold exactly in the special cases of a time-independent difference in compartmental off-resonance frequencies (*i.e.*, $\omega'_{\mathrm{F}}(\mathbf{r},t_1) - \omega'_{\mathrm{S}}(\mathbf{r},t_1) = \omega'_{\mathrm{F}}(\mathbf{r},t_2) - \omega'_{\mathrm{S}}(\mathbf{r},t_2)\ \forall \mathbf{r}, \forall t_1, t_2 \geq t_0$) or no exchange (*i.e.*, $r_{\mathrm{F}\to\mathrm{S}}(\mathbf{r}) \leftarrow 0$ and $r_{\mathrm{S}\to\mathrm{F}}(\mathbf{r}) \leftarrow 0\,\forall \mathbf{r}$).

We next use (6.27) to develop a two-compartment DESS signal model. As before, let $\mathbf{m}'(\mathbf{r}, t_0)$ denote the two-compartment magnetization at an initial time t_0 selected well into the steady-state and just prior to RF excitation. The DESS sequence first applies a fixed RF excitation, which we again assume is of sufficiently short duration T_P as to permit neglect of within-pulse relaxation, off-resonance, and exchange effects. Further assuming that each compartment experiences excitation with equal transmit field sensitivity, we model RF excitation as a simple rotation operation by angle $\alpha(\mathbf{r})$:

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r}))\mathbf{m}'(\mathbf{r}, t_0). \tag{6.28}$$

The transverse components of $\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P})$ contribute to a first acquired signal; dephase (but do not spoil completely) due to gradient dephasing, and contribute again to a second (smaller, but nonzero) acquired signal. We assume that the dephasing gradient is of sufficiently small gradient area so as to contribute mainly to compartmental off-resonance phase accrual and negligibly to compartmental magnetization attenuation due to self-diffusion. Then, compartmental magnetization evolution during data acquisition and gradient dephasing and until repetition time T_R is reasonably described by (6.27):

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{R}) \approx e^{\int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \mathbf{A}(\mathbf{r},t') \,\mathrm{d}\,t'}\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) + \left(e^{\int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \mathbf{A}(\mathbf{r},t') \,\mathrm{d}\,t'} - \mathbf{I}_6\right)\mathbf{m}_0(\mathbf{r}). \tag{6.29}$$

Conveniently, approximation (6.29) depends on compartmental off-resonance effects entirely through compartmental phase functions $\phi'_\mathrm{F}(\mathbf{r}) := \int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \omega'_\mathrm{F}(\mathbf{r}, t') \,\mathrm{d}\,t'$ and $\phi'_\mathrm{S}(\mathbf{r}) := \int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \omega'_\mathrm{S}(\mathbf{r}, t') \,\mathrm{d}\,t'$, which will later aid intra-voxel integration.

In steady state, one cycle of excitation, first acquisition, gradient spoiling, second acquisition, and partial recovery returns the compartmental magnetization back to its initial state. We enforce this through the usual steady-state condition

$$\mathbf{m}'(\mathbf{r}, t_0) = \mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{R}) \tag{6.30}$$

which yields an algebraic system of equations. The solution (if it exists) gives the steady-state compartmental magnetization just prior to RF excitation:

$$\mathbf{m}'(\mathbf{r}, t_0) = \left(\mathbf{I}_6 - e^{\int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \mathbf{A}(\mathbf{r},t') \,\mathrm{d}\,t'}\mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r}))\right)^{-1}\left(e^{\int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \mathbf{A}(\mathbf{r},t') \,\mathrm{d}\,t'} - \mathbf{I}_6\right)\mathbf{m}_0(\mathbf{r}). \tag{6.31}$$

Substituting (6.31) into (6.28) would yield a similar expression for the steady-state compartmental magnetization immediately following RF excitation. Equivalently, one may substitute (6.30) into (6.28) before solving for $\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P})$ directly, which gives

$$\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) = \left(\mathbf{I}_6 - \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r})) e^{\int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \mathbf{A}(\mathbf{r},t')\,\mathrm{d}\,t'}\right)^{-1} \times \mathbf{R}_{x'}^{\otimes 2}(\alpha(\mathbf{r})) \left(e^{\int_{t_0+T_\mathrm{P}}^{T_\mathrm{R}} \mathbf{A}(\mathbf{r},t')\,\mathrm{d}\,t'} - \mathbf{I}_6\right) \mathbf{m}_0(\mathbf{r}). \tag{6.32}$$

Unlike the SPGR two-compartment magnetization (6.16), analogous DESS expressions (6.31)-(6.32) depend strongly on transverse submatrix $\int_{t_0}^{T_\mathrm{P}} \mathbf{A}_{xy}(\mathbf{r}, t')\,\mathrm{d}\,t'$. Consequently, it remains challenging (even with computer solvers) to expand corresponding matrix exponentials and thereby find explicit representations of (6.31)-(6.32) in general.

Frequently, the DESS signals are acquired at symmetric echo times T_E before and after the center of each RF pulse. Since no gradient dephasing occurs during intervals between refocusing and defocusing echoes that contain excitations, we can reasonably assume linear off-resonance phase accrual over these intervals and therefore use (6.9) (instead of (6.27)) to evolve the magnetization accordingly. Substituting (6.32) into (6.9) gives the magnetization at the data acquisition time after RF excitation:

$$\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) = e^{\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})} \mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P}) + \left(\mathbf{I}_6 - e^{\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})}\right) \mathbf{m}_0(\mathbf{r}). \tag{6.33}$$

To compute the magnetization at the acquisition time before excitation, we consider the free precession, relaxation, and exchange that occurs between signal reception and excitation:

$$\mathbf{m}'(\mathbf{r}, t_0) = e^{\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})} \mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} - T_\mathrm{E}\right) + \left(\mathbf{I}_6 - e^{\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)\mathbf{A}(\mathbf{r})}\right) \mathbf{m}_0(\mathbf{r}). \tag{6.34}$$

Inserting (6.31) into (6.34) and rearranging gives an expression for $\mathbf{m}'\big(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} - T_\mathrm{E}\big)$.

The received signal is approximately proportional to the integrated transverse magnetization arising from both compartments. To derive expressions, we take assumptions very similar to those used in Subsection 6.2.1 and append additional distributional assumptions on the compartmental phase accrual functions $\phi'_\mathrm{F}(\mathbf{r})$ and $\phi'_\mathrm{S}(\mathbf{r})$:

1. We assume that the signal is localized to a scale over which there is within-voxel variation of compartmental off-resonance effects, but minimal intra-voxel variation of other space-varying parameters $m_0, f_\mathrm{F}, f_\mathrm{S}, T_{1,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{F}}, T_{2,\mathrm{S}}, \alpha, r_{\mathrm{F}\to\mathrm{S}}, r_{\mathrm{S}\to\mathrm{F}}$. This assumption effectively fixes all compartmental properties other than ϕ'_F, ϕ'_S, ω'_F, and ω'_S over the volume $\mathbb{V}$ of a sufficiently small voxel.
2. We assume that the dephasing gradient imparts a sufficiently large integral number

n_{cyc} of intra-voxel phase cycles such that full-repetition compartmental phase accruals ϕ'_{F} and ϕ'_{S} are distributed essentially uniformly as $\mathsf{p}_{\phi'_{\mathrm{F}}} \leftarrow \mathrm{unif}(0, 2\pi n_{\mathrm{cyc}})$ and $\mathsf{p}_{\phi'_{\mathrm{S}}} \leftarrow \mathrm{unif}(0, 2\pi n_{\mathrm{cyc}})$, where $n_{\mathrm{cyc}} \in \{1, 2, 3, \dots\}$.

3. We again assume that compartmental off-resonance phase accrues linearly between each excitation and its adjacent data acquisition periods, and that off-resonance frequencies are independently distributed within a localized voxel with marginal distributions $\mathsf{p}_{\omega'_{\mathrm{F}}} := \mathrm{Cauchy}\big(\bar{\omega}'_{\mathrm{F}}, R'_{2,\mathrm{F}}\big)$ and $\mathsf{p}_{\omega'_{\mathrm{S}}} := \mathrm{Cauchy}\big(\bar{\omega}'_{\mathrm{S}}, R'_{2,\mathrm{S}}\big)$, where $\bar{\omega}'_{\mathrm{F}}, \bar{\omega}'_{\mathrm{S}}$ are median off-resonance frequencies and $R'_{2,\mathrm{F}}, R'_{2,\mathrm{S}}$ are broadening bandwidths.
4. We assume for short echo times T_{E} that negligible exchange occurs between each excitation and its adjacent data acquisition periods. This assumption facilitates expansion of the matrix exponentials explicitly visible in (6.33)-(6.34) and separates off-resonance frequency broadening integrals by compartment.

With these assumptions, the noiseless two-compartment steady-state DESS signal models for a voxel centered at position $\mathbf{r}$ and with volume defined by $\mathbb{V}(\mathbf{r})$ are (to within constants):

$$
\begin{aligned}
s_{\mathrm{D}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) &\propto \int_{\mathbb{V}(\mathbf{r})} [1,1,i,i,0,0]\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \mathrm{d}^3\,\mathbf{r} && (6.35)\\
&\approx \left(\int_{\mathbb{R}^2} [1,1,i,i,0,0] e^{\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)\mathbf{A}(\mathbf{r})} \mathsf{p}_{\omega'_{\mathrm{F}}}(\omega'_{\mathrm{F}})\mathsf{p}_{\omega'_{\mathrm{S}}}(\omega'_{\mathrm{S}})\ \mathrm{d}\,\omega'_{\mathrm{F}}\ \mathrm{d}\,\omega'_{\mathrm{S}}\right)\\
&\quad \times \int_{\mathbb{R}^2} \mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{P}})\mathsf{p}_{\phi'_{\mathrm{F}}}(\phi'_{\mathrm{F}})\mathsf{p}_{\phi'_{\mathrm{S}}}(\phi'_{\mathrm{S}})\ \mathrm{d}\,\phi'_{\mathrm{S}}\ \mathrm{d}\,\phi'_{\mathrm{F}} && (6.36)\\
&= \begin{bmatrix} e^{-\left(1/T_{2,\mathrm{F}}(\mathbf{r}) + R'_{2,\mathrm{F}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{F}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ e^{-\left(1/T_{2,\mathrm{S}}(\mathbf{r}) + R'_{2,\mathrm{S}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{S}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ ie^{-\left(1/T_{2,\mathrm{F}}(\mathbf{r}) + R'_{2,\mathrm{F}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{F}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ ie^{-\left(1/T_{2,\mathrm{S}}(\mathbf{r}) + R'_{2,\mathrm{S}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{S}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ 0 \\ 0 \end{bmatrix}^{\mathsf{T}}\\
&\quad \times \int_{\mathbb{R}^2} \mathbf{m}'(\mathbf{r}, t_0 + T_{\mathrm{P}})\mathsf{p}_{\phi'_{\mathrm{F}}}(\phi'_{\mathrm{F}})\mathsf{p}_{\phi'_{\mathrm{S}}}(\phi'_{\mathrm{S}})\ \mathrm{d}\,\phi'_{\mathrm{S}}\ \mathrm{d}\,\phi'_{\mathrm{F}}; && (6.37)\\
s_{\mathrm{D}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) &\propto \int_{\mathbb{V}(\mathbf{r})} [1,1,i,i,0,0]\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) \mathrm{d}^3\,\mathbf{r} && (6.38)\\
&\approx \left(\int_{\mathbb{R}^2} [1,1,i,i,0,0] e^{-\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)\mathbf{A}(\mathbf{r})} \mathsf{p}_{\omega'_{\mathrm{F}}}(\omega'_{\mathrm{F}})\mathsf{p}_{\omega'_{\mathrm{S}}}(\omega'_{\mathrm{S}})\ \mathrm{d}\,\omega'_{\mathrm{F}}\ \mathrm{d}\,\omega'_{\mathrm{S}}\right)\\
&\quad \times \int_{\mathbb{R}^2} \mathbf{m}'(\mathbf{r}, t_0)\mathsf{p}_{\phi'_{\mathrm{F}}}(\phi'_{\mathrm{F}})\mathsf{p}_{\phi'_{\mathrm{S}}}(\phi'_{\mathrm{S}})\ \mathrm{d}\,\phi'_{\mathrm{S}}\ \mathrm{d}\,\phi'_{\mathrm{F}} && (6.39)\\
&= \begin{bmatrix} e^{+\left(1/T_{2,\mathrm{F}}(\mathbf{r}) - R'_{2,\mathrm{F}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{F}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ e^{+\left(1/T_{2,\mathrm{S}}(\mathbf{r}) - R'_{2,\mathrm{S}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{S}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ ie^{+\left(1/T_{2,\mathrm{F}}(\mathbf{r}) - R'_{2,\mathrm{F}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{F}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ ie^{+\left(1/T_{2,\mathrm{S}}(\mathbf{r}) - R'_{2,\mathrm{S}}(\mathbf{r}) + i\bar{\omega}'_{\mathrm{S}}(\mathbf{r})\right)\left(T_{\mathrm{E}} - \frac{T_{\mathrm{P}}}{2}\right)} \\ 0 \\ 0 \end{bmatrix}^{\mathsf{T}}\\
&\quad \times \int_{\mathbb{R}^2} \mathbf{m}'(\mathbf{r}, t_0)\mathsf{p}_{\phi'_{\mathrm{F}}}(\phi'_{\mathrm{F}})\mathsf{p}_{\phi'_{\mathrm{S}}}(\phi'_{\mathrm{S}})\ \mathrm{d}\,\phi'_{\mathrm{S}}\ \mathrm{d}\,\phi'_{\mathrm{F}}. && (6.40)
\end{aligned}
$$

Observe that (6.36) and (6.39) use Assumption 1 to separate the phase and frequency broadening integrals respectively defined in Assumptions 2 and 3. Similar to SPGR calculations in Subsection 6.2.1, (6.37) and (6.40) use Assumption 4 to evaluate the frequency broadening integral compartment-wise. However, the phase broadening integrals in (6.37) and (6.40) do not separate. Thus, even with the simple first-order Magnus expansion taken in

(6.27), we find in DESS that exchange causes magnetization compartments to interact not only via exchange rates but also through differences across compartments in off-resonance phase accrual. In other words, additional assumptions on the relationship between compartmental off-resonance distributions (*e.g.*, that they are the same) would affect the DESS models *even if* the signal were received immediately following excitation (at $T_\mathrm{E} \leftarrow \frac{T_\mathrm{P}}{2}$).

In the special case where exchange is altogether neglected (which is a stronger assumption than Assumption 4, especially for longer T_R), the phase broadening integrals in (6.37) and (6.40) separate across compartments and admit the closed-form expressions

$$\begin{aligned} s_\mathrm{D}\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) &\propto +im_0(\mathbf{r}) \tan\frac{\alpha(\mathbf{r})}{2} \\ &\quad\times \left(f_\mathrm{F}(\mathbf{r}) \left(1 - \frac{\eta_\mathrm{F}(\mathbf{r}, T_\mathrm{R} - T_\mathrm{P})}{\xi_\mathrm{F}(\mathbf{r}, T_\mathrm{R} - T_\mathrm{P})}\right) e^{-\left(1/T_{2,\mathrm{F}}(\mathbf{r}) + R'_{2,\mathrm{F}}(\mathbf{r}) + i\bar{\omega}'_\mathrm{F}(\mathbf{r})\right)\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)} \right. \\ &\quad \left. + f_\mathrm{S}(\mathbf{r}) \left(1 - \frac{\eta_\mathrm{S}(\mathbf{r}, T_\mathrm{R} - T_\mathrm{P})}{\xi_\mathrm{S}(\mathbf{r}, T_\mathrm{R} - T_\mathrm{P})}\right) e^{-\left(1/T_{2,\mathrm{S}}(\mathbf{r}) + R'_{2,\mathrm{S}}(\mathbf{r}) + i\bar{\omega}'_\mathrm{S}(\mathbf{r})\right)\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)} \right); \end{aligned} \tag{6.41}$$

$$\begin{aligned} s_\mathrm{D}\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} - T_\mathrm{E}\right) &\propto -im_0(\mathbf{r}) \tan\frac{\alpha(\mathbf{r})}{2} \\ &\quad\times \left(f_\mathrm{F}(\mathbf{r}) (1 - \eta_\mathrm{F}(\mathbf{r}, T_\mathrm{R} - T_\mathrm{P})) e^{+\left(1/T_{2,\mathrm{F}}(\mathbf{r}) - R'_{2,\mathrm{F}}(\mathbf{r}) + i\bar{\omega}'_\mathrm{F}(\mathbf{r})\right)\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)} \right. \\ &\quad \left. + f_\mathrm{S}(\mathbf{r}) (1 - \eta_\mathrm{S}(\mathbf{r}, T_\mathrm{R} - T_\mathrm{P})) e^{+\left(1/T_{2,\mathrm{S}}(\mathbf{r}) - R'_{2,\mathrm{S}}(\mathbf{r}) + i\bar{\omega}'_\mathrm{S}(\mathbf{r})\right)\left(T_\mathrm{E} - \frac{T_\mathrm{P}}{2}\right)} \right), \end{aligned} \tag{6.42}$$

where η_F, η_S, ξ_F, and ξ_S are intermediate variables defined as

$$\begin{aligned} \eta_\mathrm{F}(\mathbf{r}, t) &:= \sqrt{\frac{1 - (\exp(-t/T_{2,\mathrm{F}}(\mathbf{r})))^2}{1 - (\exp(-t/T_{2,\mathrm{F}}(\mathbf{r}))/\xi_\mathrm{F}(\mathbf{r}, t))^2}}; \\ \xi_\mathrm{F}(\mathbf{r}, t) &:= \frac{1 - \exp(-t/T_{1,\mathrm{F}}(\mathbf{r}))\cos\alpha(\mathbf{r})}{\exp(-t/T_{1,\mathrm{F}}(\mathbf{r})) - \cos\alpha(\mathbf{r})}; \\ \eta_\mathrm{S}(\mathbf{r}, t) &:= \sqrt{\frac{1 - (\exp(-t/T_{2,\mathrm{S}}(\mathbf{r})))^2}{1 - (\exp(-t/T_{2,\mathrm{S}}(\mathbf{r}))/\xi_\mathrm{S}(\mathbf{r}, t))^2}}; \\ \xi_\mathrm{S}(\mathbf{r}, t) &:= \frac{1 - \exp(-t/T_{1,\mathrm{S}}(\mathbf{r}))\cos\alpha(\mathbf{r})}{\exp(-t/T_{1,\mathrm{S}}(\mathbf{r})) - \cos\alpha(\mathbf{r})}. \end{aligned}$$

Eqs. (6.41)-(6.42) naturally extend single-compartment models (2.41) and (2.43), and elucidate the intuitive result that compartmental signals simply add if exchange is neglected.

6.3 A Fast SS Acquisition for Precise Myelin Imaging

This section develops a new acquisition consisting of fast SPGR and DESS scans for precise myelin water imaging. Subsection 6.3.1 adapts acquisition design (introduced in Ch. 4) to enable larger-scale scan optimization. Subsection 6.3.2 applies scalable acquisition design to design a fast SPGR/DESS acquisition that enables precise estimation of the faster-relaxing signal fraction f_{F} in two-compartment SPGR/DESS models. Subsequent sections take f_{F} as a quantitative measure of myelin water content in SS acquisitions.

6.3.1 Scalable Acquisition Design

Recall from Subsection 4.2.1 that the inverse of the Fisher information $\mathbf{F}(\mathbf{x};\boldsymbol{\nu},\mathbf{P}) \in \mathbb{C}^{L\times L}$ lower-bounds the covariance of unbiased estimates of L latent object parameters $\mathbf{x} \in \mathbb{C}^L$, given K known object parameters $\boldsymbol{\nu} \in \mathbb{C}^K$ and A tunable acquisition parameters for each of D datasets $\mathbf{P} \in \mathbb{R}^{A\times D}$. As before, we continue to focus on minimizing a weighted average of the latent parameter variances and thus study the objective function

$$\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P}) := \mathrm{tr}\big(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x};\boldsymbol{\nu},\mathbf{P})\mathbf{W}^{\mathsf{T}}\big), \tag{6.43}$$

where $\mathbf{W}$ is a diagonal weighting matrix and $\mathrm{tr}(\cdot)$ denotes the matrix trace operation. For scan design, we seek to minimize Ψ with respect to acquisition parameters $\mathbf{P}$.

In Subsection 4.2.2, we addressed the dependance of Ψ on space-varying object parameters $\mathbf{x}$ and $\boldsymbol{\nu}$ through a min-max optimization problem. The associated "worst-case" design criterion requires only weak assumptions on object parameter distributions but is non-differentiable in $\mathbf{P}$. For the relatively simple application described in Section 4.3, the min-max criterion was studied through exhaustive search and so non-differentiability did not matter. However, myelin water imaging requires estimation of several more latent parameters and thus necessitates scan parameter selection for a greater number of datasets and thus over a larger search space. Since exhaustive search over this higher-dimensional search space is prohibitively expensive computationally, we study here an alternate design criterion that is differentiable in $\mathbf{P}$ and is thus amenable to gradient-based local optimization. Specifically, we seek an acquisition parameter $\breve{\mathbf{P}}$ that minimizes the *expected* weighted average of latent parameter variances over a search space $\mathbb{P}$:

$$\breve{\mathbf{P}} \in \left\{\arg\min_{\mathbf{P}\in\mathbb{P}} \bar{\Psi}(\mathbf{P})\right\}, \text{ where} \tag{6.44}$$

$$\bar{\Psi}(\mathbf{P}) := \mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P})) \tag{6.45}$$

and $\mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(\cdot)$ denotes joint expectation with respect to prior joint distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ on $\mathbf{x}, \boldsymbol{\nu}$.

Unlike min-max cost (4.7), expected cost (6.45) is often differentiable in $\mathbf{P}$. We next construct the gradient matrix $\nabla_{\mathbf{P}}\bar{\Psi}(\mathbf{P}) \in \mathbb{R}^{A\times D}$ and provide sufficient conditions for when this gradient matrix exists. Our simple strategy involves first constructing $\nabla_{\mathbf{P}}\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P})$ element-wise for fixed $\mathbf{x},\boldsymbol{\nu}$ and then relating $\nabla_{\mathbf{P}}\bar{\Psi}(\mathbf{P})$ to $\nabla_{\mathbf{P}}\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P})$. Let $\frac{\partial}{\partial p_{a,d}}$ be the (a,d)th element of matrix operator $\nabla_{\mathbf{P}}$. By standard matrix derivative identities, we have

$$\begin{aligned}\frac{\partial}{\partial p_{a,d}}\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P}) &= \frac{\partial}{\partial p_{a,d}}\operatorname{tr}\bigl(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x};\boldsymbol{\nu},\mathbf{P})\mathbf{W}^{\mathsf{T}}\bigr)\\ &= -\operatorname{tr}\left(\mathbf{W}\mathbf{F}^{-1}(\mathbf{x};\boldsymbol{\nu},\mathbf{P})\frac{\partial}{\partial p_{a,d}}(\mathbf{F}(\mathbf{x};\boldsymbol{\nu},\mathbf{P}))\mathbf{F}^{-1}(\mathbf{x};\boldsymbol{\nu},\mathbf{P})\mathbf{W}^{\mathsf{T}}\right). \end{aligned} \tag{6.46}$$

For image data corrupted by additive complex Gaussian noise with zero mean and covariance $\boldsymbol{\Sigma}$, the Fisher information (repeated from (4.3) for clarity) is given by

$$\mathbf{F}(\mathbf{x};\boldsymbol{\nu},\mathbf{P}) = (\nabla_{\mathbf{x}}\mathbf{s}(\mathbf{x};\boldsymbol{\nu},\mathbf{P}))^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\nabla_{\mathbf{x}}\mathbf{s}(\mathbf{x};\boldsymbol{\nu},\mathbf{P}), \tag{6.47}$$

where $\mathbf{s} := [s_1,\ldots,s_D]^{\mathsf{T}}$ is the noiseless signal model. Furthermore, if elements of each measurement vector are assumed independent (as is typical), $\boldsymbol{\Sigma}$ takes the diagonal structure $\boldsymbol{\Sigma} \leftarrow \operatorname{diag}(\sigma_1^2,\ldots,\sigma_D^2)$ and

$$\begin{aligned}\frac{\partial}{\partial p_{a,d}}(\mathbf{F}(\mathbf{x};\boldsymbol{\nu},\mathbf{P})) &= \frac{\partial}{\partial p_{a,d}}\sum_{d'=1}^{D}\frac{1}{\sigma_{d'}^2}(\nabla_{\mathbf{x}}s_{d'}(\mathbf{x};\boldsymbol{\nu},\mathbf{p}_{d'}))^{\mathsf{H}}\nabla_{\mathbf{x}}s_{d'}(\mathbf{x};\boldsymbol{\nu},\mathbf{p}_{d'})\\ &= \frac{1}{\sigma_d^2}\frac{\partial}{\partial p_{a,d}}\Bigl((\nabla_{\mathbf{x}}s_d(\mathbf{x};\boldsymbol{\nu},\mathbf{p}_d))^{\mathsf{H}}\nabla_{\mathbf{x}}s_d(\mathbf{x};\boldsymbol{\nu},\mathbf{p}_d)\Bigr). \end{aligned} \tag{6.48}$$

Substituting (6.47)-(6.48) into (6.46) gives expressions in terms of signal model derivatives for each element of $\nabla_{\mathbf{P}}\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P})$. These expressions are well-defined if $\mathbf{F}$ is invertible and if mixed partial derivatives $\nabla_{\mathbf{p}_1}(\nabla_{\mathbf{x}}s_1)^{\mathsf{T}},\ldots,\nabla_{\mathbf{p}_D}(\nabla_{\mathbf{x}}s_D)^{\mathsf{T}}$ exist and are continuous in $\mathbf{x},\mathbf{P}$. Further assuming that $\nabla_{\mathbf{P}}\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P})$ remains bounded for all $\mathbf{x},\boldsymbol{\nu}$,

$$\begin{aligned}\nabla_{\mathbf{P}}\bar{\Psi}(\mathbf{x};\boldsymbol{\nu},\mathbf{P}) &= \nabla_{\mathbf{P}}\mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P}))\\ &= \mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(\nabla_{\mathbf{P}}\Psi(\mathbf{x};\boldsymbol{\nu},\mathbf{P})), \end{aligned} \tag{6.49}$$

which provides an expression for the gradient of the expected cost, as desired.

If (6.49) exists, one could solve (6.44) iteratively for convex search space $\mathbb{P}$ via updates

$$\mathbf{P}^{(i)} \leftarrow \mathsf{P}_{\mathbb{P}}\bigl(\mathbf{P}^{(i-1)} - \nabla_{\mathbf{P}}\bar{\Psi}\bigl(\mathbf{P}^{(i-1)}\bigr)\bigr), \tag{6.50}$$

where $\mathsf{P}_{\mathbb{P}}(\cdot)$ denotes projection onto $\mathbb{P}$ and i indexes iteration. Since $\bar{\Psi}$ is non-convex in $\mathbf{P}$ in general, such iterations achieve only locally optimal convergence in cost (further discussed in Subsection 2.2.1) and the locally convergent minimizer depends on initialization.

6.3.2 SPGR/DESS Scan Design Implementation Details

This subsection applies scalable scan design problem (6.44) to optimize fast scan profiles consisting of SPGR and DESS scans for precise myelin water imaging. Intuitively, we study SPGR/DESS scan profiles over previously studied SS scan combinations [121, 158, 101] because SPGR/DESS signals are relatively insensitive to off-resonance related non-idealities, which as demonstrated in Section 6.2 can be difficult to model accurately in multi-compartmental systems and in fact here (as in [158, 101]) are largely neglected.

For the feasibility studies discussed in the remainder of this chapter, we assumed that signal arises from two non-exchanging water compartments (*i.e.*, $r_{\mathrm{F}\to\mathrm{S}} \leftarrow 0$, $r_{\mathrm{S}\to\mathrm{F}} \leftarrow 0$, and $f_{\mathrm{F}} + f_{\mathrm{S}} = 1$) with identical broadening distributions (*i.e.*, $R'_{2,\mathrm{F}} \equiv R'_{2,\mathrm{S}}$ and $\bar{\omega}'_{\mathrm{F}} \equiv \bar{\omega}'_{\mathrm{S}}$). These simplifications provided closed-form expressions for the SPGR (6.22) and DESS (6.41)-(6.42) signal models as well as their gradients. We used magnitude signal models for scan design to reduce SPGR/DESS signal dependence on off-resonance effects, noting that Rician distributed noise in corresponding magnitude image data is well-approximated as Gaussian for sufficiently large SNR [30]. We fixed $T_{\mathrm{P}}, T_{\mathrm{E}}$ across scans and thereby reduced model dependencies to seven free object parameters per voxel: f_{F}, $T_{1,\mathrm{F}}$, $T_{2,\mathrm{F}}$, $T_{1,\mathrm{S}}$, $T_{2,\mathrm{S}}$, κ^{t}, and $c_3 := m_0 e^{-R'_{2,\mathrm{F}} T_{\mathrm{E}}} \equiv m_0 e^{-R'_{2,\mathrm{S}} T_{\mathrm{E}}}$; and two acquisition parameters per dataset: $\mathbf{p}_d \leftarrow [\alpha_0, T_{\mathrm{R}}]^{\mathsf{T}}, \forall d \in \{1, \ldots, D\}$. We assumed prior knowledge of transmit field sensitivity $\boldsymbol{\nu} \leftarrow \kappa^{\mathrm{t}}$ (which in practice can be estimated from separate fast acquisitions, *e.g.* [90]) and collected the remaining $L \leftarrow 6$ latent parameters as $\mathbf{x} \leftarrow [f_{\mathrm{F}}, T_{1,\mathrm{F}}, T_{2,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{S}}, c_3]$.

We took fast-relaxing compartmental fraction f_{F} to be a quantitative measure of myelin water content and tailored scan design cost (6.45) to encourage scan combinations that enable precise f_{F} estimation. Specifically, we set weight matrix $\mathbf{W} \leftarrow \mathrm{diag}\big((\mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(f_{\mathrm{F}}))^{-1}, \mathbf{0}_5\big)$ to penalize only f_{F} imprecision, thereby ignoring nuisance parameter imprecision. Here, fast-fraction variance weight $(\mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(f_{\mathrm{F}}))^{-1}$ assigns interpretable meaning to $\sqrt{\bar{\Psi}(\mathbf{P})}$ as a unitless measure of the expected relative standard deviation (alternately, expected coefficient of variation) afforded by $\mathbf{P}$ in asymptotically unbiased estimates of f_{F}.

We approximated expectations of form $\mathsf{E}_{\mathbf{x},\boldsymbol{\nu}}(\cdot)$ by taking empirical averages using samples of $\mathbf{x}, \boldsymbol{\nu}$ drawn from a prior distribution. We used a coordinate-wise separable prior distribution, modeling $f_{\mathrm{F}} \sim \mathrm{unif}(0.03, 0.21)$ to conservatively contain state-of-the-art MESE MWF measurements in WM [159] and modeling $T_{1,\mathrm{F}}$, $T_{2,\mathrm{F}}$, $T_{1,\mathrm{S}}$, and $T_{2,\mathrm{S}}$ to be Gaussian

	Optimized flip angles (deg)	Optimized repetition times (ms)
SPGR	–	–
DESS	$[33.0, 18.3, 15.1]^{\mathsf{T}}$	$[17.5, 30.2, 60.3]^{\mathsf{T}}$

Table 6.1: SPGR/DESS flip angles and repetition times that comprise $\widehat{\mathbf{P}}$, a scan parameter matrix designed under total time budget $\sum_{d=1}^{D} T_{\mathrm{R},d} \leq 108$ms for precise f_{F} estimation in WM. For our noise variance measurements, this acquisition is expected to yield 42.5% relative standard deviation in asymptotically unbiased f_{F} estimates from two-compartment signal models. Interestingly, the optimized scan profile omits SPGR scans entirely.

distributed with means 400ms, 20ms, 1000ms, and 80ms selected from literature measurements [100, 158] and standard deviations that are 20% of corresponding means. Since $\mathbf{W}$ placed zero weight on estimating c_3, it sufficed to fix $c_3 \leftarrow 1$ and to assign noise variance $\mathbf{\Sigma} \leftarrow (1.49 \times 10^{-7})\mathbf{I}_{10}$ based on separate measurements in unit-normalized image data (*cf.* Section 4.4.2.1 for acquisition details). Lastly, we modeled $\kappa^{\mathrm{t}} \sim \mathrm{unif}(0.9, 1.1)$ to account for 10% flip angle spatial variation.

We constrained our search space $\mathbb{P}$ to reflect hardware, safety, and model-accuracy limitations and to avoid undesirably long acquisitions. To control RF energy deposition, we restricted DESS flip angles to range between $1°$ and $60°$. We further restricted SPGR flip angles to be between $1°$ and $40°$ to avoid excessive model mismatch due to partial spoiling effects [19]. To comply with other fixed pulse sequence timing requirements, we required DESS and SPGR repetition times to be no less than 17.5ms and 11.8ms respectively. We constrained each pair of DESS defocusing- and refocusing-echo datasets be assigned the same flip angle and repetition time. Lastly, we imposed a somewhat aggressive total scan time constraint $\sum_{d=1}^{D} T_{\mathrm{R},d} \leq 108$ms that ensured all feasible points described acquisitions at least as fast as the state-of-the-art fast steady-state acquisition proposed in [158]. These constraints together defined a convex search space over which we optimized $\bar{\Psi}$.

We separately optimized (6.45) for each of the 36 nontrivial candidate SPGR/DESS scan profiles that are feasible under the above time constraint. For a candidate profile containing D datasets, we separately solved (6.44) with $200D$ initializations selected uniformly randomly within the feasible set. For each profile and each initialization, we solved (6.44) using the built-in MATLAB® function `fmincon` with the `active-set` algorithm, a cost function convergence tolerance of 10^{-7}, a maximum of 500 iterations, and otherwise default options. We performed scan optimization running MATLAB® R2017a with a pool of 12 workers on two Xeon-X5650 2.67GHz hexa-core CPUs.

Table 6.1 summarizes the optimized scan parameter $\widehat{\mathbf{P}}$ that minimizes (6.45) over all profiles and all initializations. Interestingly, the optimized scan profile contains DESS scans only. We find that $\sqrt{\bar{\Psi}(\widehat{\mathbf{P}})} = 0.425$, meaning that at a realistic noise level, the acquisition defined by $\widehat{\mathbf{P}}$ is expected to yield 42.5% relative standard deviation in asymptotically unbiased f_{F} estimates from non-exchanging two-compartment signal models. Similar calculations for previous SS myelin water imaging acquisitions [121, 158] found that f_{F} relative standard deviations frequently exceeded 100% [160].

6.4 Experimentation

This section demonstrates myelin water imaging using the precision-optimized SS acquisition developed in Section 6.3 and fast PERK estimation (developed in Ch. 5). Subsection 6.4.1 describes implementation details common to both simulations and experiments. Subsection 6.4.2 studies PERK and three other estimators' statistics in two numerical simulations. Subsection 6.4.3 compares PERK-based f_{F} maps to MWF maps from conventional MESE acquisitions through *in vivo* experiments.

6.4.1 Methods

In all simulations and experiments, we used data arising from the fast SS scan profile described in Table 6.1. Since this scan profile consisted of three DESS scans and each DESS scan generates two signals per excitation, this acquisition yielded $D \leftarrow 6$ datasets. We assumed non-exchanging two-compartment DESS signal models (6.41)-(6.42) and took the same assumptions as in Subsection 6.3.2 to reduce model dependencies to $L \leftarrow 6$ latent parameters $\mathbf{x} \leftarrow [f_{\mathrm{F}}, T_{1,\mathrm{F}}, T_{2,\mathrm{F}}, T_{1,\mathrm{S}}, T_{2,\mathrm{S}}, c_3]^{\mathsf{T}}$ and $K \leftarrow 1$ known parameter $\boldsymbol{\nu} \leftarrow \kappa^{\mathrm{t}}$ per voxel. We focused on precisely estimating f_{F} in WM from this fast DESS acquisition. We considered the other five latent parameters to be nuisance parameters and thus did not evaluate the performance of their (possibly imprecise) estimators.

We estimated f_{F} using the PERK estimator developed in Ch. 5. We trained, tuned, and tested PERK in much the same way as in Subsection 5.7.1. We assumed a separable prior distribution $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}} \leftarrow \mathsf{p}_{f_{\mathrm{F}}} \mathsf{p}_{T_{1,\mathrm{F}}} \mathsf{p}_{T_{2,\mathrm{F}}} \mathsf{p}_{T_{1,\mathrm{S}}} \mathsf{p}_{T_{2,\mathrm{S}}} \mathsf{p}_{c_3} \mathsf{p}_{\kappa^{\mathrm{t}}}$. We set fast-relaxing fraction marginal distribution $\mathsf{p}_{f_{\mathrm{F}}} \leftarrow \mathrm{unif}(-0.1, 0.4)$ and note that we sample negative values[8] with nonzero probability to ensure sampling bias does not contribute significantly to f_{F} estimation bias, especially in low-f_{F} regions. We chose relaxation parameter marginal distributions $\mathsf{p}_{T_{1,\mathrm{F}}} \leftarrow \mathrm{logunif}(50, 700)$ms, $\mathsf{p}_{T_{2,\mathrm{F}}} \leftarrow \mathrm{logunif}(5, 50)$ms, $\mathsf{p}_{T_{1,\mathrm{S}}} \leftarrow \mathrm{logunif}(700, 2000)$ms, $\mathsf{p}_{T_{2,\mathrm{S}}} \leftarrow \mathrm{logunif}(50, 300)$ms similar to those used for scan design but with finite support. To match

[8]Our two-compartment signal models are linear in f_{F} and are therefore well-defined for zero or even negative f_{F} values.

the scaling of training and testing data, we set $\mathsf{p}_{c_3} \leftarrow \mathrm{unif}(2.2 \times 10^{-16}, u)$, with u set as $10\times$ the maximum value of magnitude test data. We estimated flip angle scaling marginal distribution $\mathsf{p}_{\kappa^{\mathrm{t}}}$ from known κ^{t} map voxels via kernel density estimation (implemented using the built-in MATLAB® function `fitdist` with default options) and then clipped the support of $\mathsf{p}_{\kappa^{\mathrm{t}}}$ to assign nonzero probability only within $[0.5, 2]$. We assumed noise covariance $\mathbf{\Sigma}$ of form $\sigma^2 \mathbf{I}_6$ and estimated the (spatially invariant) noise variance σ^2 from Rayleigh-distributed regions of magnitude test data, using estimators described in [137]. We sampled $N \leftarrow 10^6$ latent and known parameter realizations from these distributions and evaluated two-compartment DESS signal models (6.41)-(6.42) to generate corresponding complex noiseless measurements. After adding complex Gaussian noise realizations, we concatenated the (Rician) magnitude of these noisy measurements with known parameter realizations to construct pure-real regressors. We used Gaussian kernel (5.12) with bandwidth matrix $\mathbf{\Lambda}$ set based on test data scale via (5.25). We separately selected and then held fixed bandwidth scaling parameter $\lambda \leftarrow 2^{0.3}$ and regularization parameter $\rho \leftarrow 2^{-19}$ via holdout. In particular, we optimized holdout cost function (5.26) with holdout weighting matrix $\mathbf{W} \leftarrow \mathrm{diag}\left(\left[1, \mathbf{0}_5^{\mathsf{T}}\right]^{\mathsf{T}}\right)$ set to focus only on f_{F} estimation and with $T \leftarrow 10^5$ holdout test points drawn from a distribution very similar to $\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}}$ except with $\mathsf{p}_{f_{\mathrm{F}}} \leftarrow \mathrm{unif}(0.03, 0.21)$ to avoid division-by-zero issues. All other holdout details were identical to those described in Subsection 5.7.1.1. We sampled random feature realizations $\mathbf{v}, s$ $Z \leftarrow 10^3$ times to construct approximate feature map $\tilde{\mathbf{z}}$ via (5.20). For each lth latent parameter where $l \in \{1, \dots, L\}$, we applied $\tilde{\mathbf{z}}$ to training data; computed sample means $m_{x_l}, \mathbf{m}_{\tilde{\mathbf{z}}}$ and sample covariances $\mathbf{c}_{\tilde{\mathbf{z}}x_l}, \mathbf{C}_{\tilde{\mathbf{z}}\tilde{\mathbf{z}}}$; and evaluated approximate PERK estimator (5.23) on test image data and the known flip angle scaling map on a per-voxel basis.

We compared PERK f_{F} estimates from the aforementioned DESS acquisition to two conventional MWF estimates from a MESE acquisition. The first conventional MWF estimate [100] is related to the solution of a nonnegative least-squares (NNLS) problem [166]

$$\widehat{\mathbf{x}}(\mathbf{y}) \in \left\{ \arg\min_{\mathbf{x} \in \mathbb{X}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\}, \tag{6.51}$$

where $\mathbf{y} \in \mathbb{R}^D$ denotes MESE echo amplitudes at D echo times; $\mathbf{A} \in \mathbb{R}^{D \times L}$ models the D MESE signals as weighted sums of L discrete T_2 component signals; $\mathbb{X} \subset \mathbb{R}^L$ is the nonnegative orthant; and $\mathbf{x} \in \mathbb{X}$ denotes L component weights. Whereas solutions to (6.51) tend to be sparse for $L > D$ as is typical, researchers have suggested that spectral distributions are more likely smooth in biological tissue [167]. For smoother *in vivo* spectrum estimates and for improved problem conditioning, we also compared to a second MWF estimate [168] that is related to the solution of a regularized NNLS problem (RNNLS)

$$\widehat{\mathbf{x}}(\mathbf{y}) \in \left\{ \arg \min_{\mathbf{x} \in \mathbb{X}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \|\mathbf{x}\|_2^2 \right\}, \tag{6.52}$$

where β is positive regularization parameter. Similar to [100] or [168], each of the two conventional MWF estimators are then respectively defined as $\widehat{f}_{\mathrm{M}} := \left(\mathbf{i}^{\mathsf{T}}\widehat{\mathbf{x}}(\cdot)\right)/\|\widehat{\mathbf{x}}(\cdot)\|_1$, where $\widehat{\mathbf{x}}(\cdot)$ is given by (6.51) or (6.52) and $\mathbf{i} \in \{0,1\}^L$ is in both cases nonzero only in entries corresponding to $T_2 \in [15, 40]$ms. As recommended in recent MESE MWF imaging literature [169], we computed MESE signal amplitudes using the extended phase graph (EPG) formalism [170] to account for stimulated echo signal contributions that persist in MESE due to non-ideal refocusing pulses. We accounted for non-ideal refocusing by assuming κ^{t} is known. We also partially accounted for incomplete recovery by assuming bulk T_1 is known. To circumvent separate EPG simulations for every voxel, we clustered known κ^{t}, T_1 map voxels into 100 clusters via k-means++ [93] and ran only 100 EPG simulations using each of the cluster means. Iterating over clusters, we generated each cluster's system matrix and solved (6.51) and (6.52) for MESE image voxels assigned to that cluster. We constructed NNLS and RNNLS MWF estimates by estimating $L \leftarrow 100$ spectral component images from $D \leftarrow 32$ MESE measurement images. We manually fixed RNNLS regularization parameter $\beta \leftarrow 2^{-13}$ as a small value that provided reasonable *in vivo* results. We solved (6.51) and (6.52) using the MATLAB® function `lsqnonneg` with default options.

We performed simulations and experiments running MATLAB® R2013a on a 3.5GHz desktop computer with 32GB RAM. In the interest of reproducible research, code and data will be freely available at https://gitlab.eecs.umich.edu/fessler/qmri.

6.4.2 Numerical Simulations

This subsection studies PERK and maximum likelihood (ML) f_{F} estimator statistics as well as NNLS and RNNLS f_{M} estimator statistics in two increasingly realistic simulations. Simulation 6.4.2.1 investigates estimator performance when voxel data is simulated to arise from two water compartments that recover with the same bulk T_1 time constant, in which case none of the above estimators incur bias due to model mismatch. Simulation 6.4.2.2 investigates estimator performance when voxel data is more realistically simulated as arising from three water compartments that recover with compartment-specific T_1 time constants, in which case any of the above estimators could incur bias due to model mismatch.

6.4.2.1 Two-Compartment Simulation without Model Mismatch

We first simulated data to arise from two non-exchanging water compartments with different fast $T_{2,\mathrm{F}} \leftarrow 20$ms and slow $T_{2,\mathrm{S}} \leftarrow 80$ms transverse relaxation time constants (selected based on [100, 158]) but the same bulk longitudinal relaxation time constant $T_1 \leftarrow 832$ms in WM and $T_1 \leftarrow 1331$ms in GM (selected based on [31]). With this two-compartment ground truth, MWF f_M and fast-relaxing fraction f_F are equivalent and directly comparable. We assigned $f_\mathrm{M} \equiv f_\mathrm{F} \leftarrow 0.15$ in WM and $f_\mathrm{M} \equiv f_\mathrm{F} \leftarrow 0.03$ in GM and constrained corresponding slow-compartment fractions as $1 - f_\mathrm{F}$. We prescribed these parameter values to the 81st slice of the BrainWeb digital phantom [34] to produce 217×181 ground truth parameter maps. We generated κ^t to model 20% flip angle spatial variation. Using extended phase graphs to model stimulated echo signal contributions due to non-ideal refocusing, we simulated noiseless single-coil 32-echo MESE image data with nominal 90 degrees excitation and nominal 180 degrees refocusing flip angles; $T_\mathrm{E} \leftarrow 10$ms echo interval spacing; $T_\mathrm{R} \leftarrow 600$ms repetition time; and two cycles of gradient dephasing before and after each refocusing pulse. We corrupted noiseless MESE images with additive complex Gaussian noise to yield noisy complex datasets with SNR ranging from 17-868 in WM and 15-697 in GM, where SNR is defined

$$\mathsf{SNR}(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\epsilon}}) := \|\tilde{\mathbf{y}}\|_2 / \|\tilde{\boldsymbol{\epsilon}}\|_2 \tag{6.53}$$

for image data voxels $\tilde{\mathbf{y}}$ and noise voxels $\tilde{\boldsymbol{\epsilon}}$ corresponding to a region of interest (ROI) within one image. We estimated f_M from noisy magnitude MESE images and known T_1, κ^t maps by solving NNLS (6.51) and RNNLS (6.52) problems as explained in Subsection 6.4.1. NNLS and RNNLS respectively took 40.3s and 49.6s.

Using non-exchanging two-compartment models (6.41)-(6.42), we also simulated noiseless single-coil DESS image data using the precision-optimized nominal flip angles and repetition times presented in Table 6.1 and fixed symmetric defocusing and refocusing echo times $T_\mathrm{E} \leftarrow 5.29$ms. We corrupted noiseless DESS images with additive complex Gaussian noise to yield noisy complex datasets with SNR ranging from 22-222 in WM and 25-242 in GM, where SNR is computed via (6.53). We estimated f_F from noisy magnitude DESS images and known κ^t maps as detailed in Subsection 6.4.1. In simulations only, we compared PERK f_F estimates to ML f_F estimates achieved via the variable projection method [25] and grid search. The ML estimate utilizes 20 dictionaries each consisting of nearly 8×10^6 signal vectors computed using finely spaced samples on an unrealistically narrow feasible region consisting of a hypercube with boundaries set as $[-0.1, 0.4]$ in f_F and $\pm 20\%$ away from the truth in other latent parameter dimensions. PERK training and

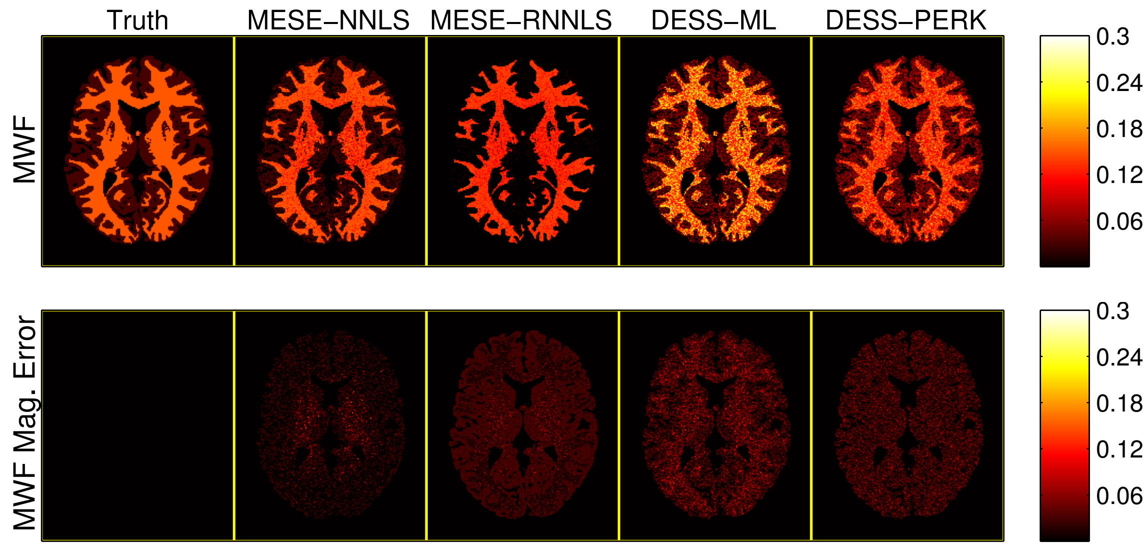


Figure 6.1: NNLS/RNNLS MESE f_{M} and ML/PERK DESS f_{F} estimates alongside corresponding magnitude error images, in a two-compartment simulation where none of the associated estimators incur bias due to model mismatch. Voxels not assigned WM- or GM-like compartmental fractions are masked out in post-processing for display. Table 6.2 presents corresponding sample statistics.

testing respectively took 33.8s and 1.0s while ML estimation took 17726s (nearly 5h).

Fig. 6.1 compares NNLS and RNNLS f_{M} estimates as well as ML and PERK f_{F} estimates alongside magnitude difference images with respect to the ground truth $f_{\mathrm{M}} \equiv f_{\mathrm{F}}$ map. Unlike both f_{F} estimates, both f_{M} estimates visibly exhibit systematic error due to flip angle spatial variation despite perfect knowledge of κ^{t}. The RNNLS f_{M} estimate exhibits greater error than the NNLS f_{M} estimate in both WM- and GM-like voxels due to regularization. The PERK f_{F} estimate visibly exhibits less error in WM-like voxels than the ML f_{F} estimate, perhaps in part because PERK tuning parameters (λ, ρ) were optimized via holdout for estimating WM-like f_{F} values.

Table 6.2 compares samples statistics of NNLS and RNNLS f_{M} estimates as well as ML and PERK f_{F} estimates, computed over 7810 WM-like and 9162 GM-like voxels. With the exception of the MESE-RNNLS GM f_{M} estimate, all other estimates agree with true values to within one standard deviation. The MESE-NNLS WM and GM f_{F} estimates achieve the least root mean-squared errors (RMSEs) overall. The RNNLS f_{M} estimate is more precise but less accurate than the NNLS f_{M} estimate due to regularization. The PERK f_{F} estimate is more precise but less accurate than the ML f_{F} estimate, perhaps also due to regularization. PERK f_{F} estimates exhibit better WM RMSE and slightly worse GM RMSE than ML f_{F} estimates.

	WM	GM
True $f_{\mathrm{M}} \equiv f_{\mathrm{F}}$	0.15	0.03
MESE-NNLS $\widehat{f_{\mathrm{M}}}$	$0.1375 \pm 0.0187\,(0.0225)$	$0.0203 \pm 0.01296\,(0.0162)$
MESE-RNNLS $\widehat{f_{\mathrm{M}}}$	$0.1285 \pm 0.0146\,(0.0260)$	$0.00207 \pm 0.00524\,(0.02841)$
DESS-ML $\widehat{f_{\mathrm{F}}}$	$0.1590 \pm 0.0433\,(0.0442)$	$0.0334 \pm 0.0272\,(0.0274)$
DESS-PERK $\widehat{f_{\mathrm{F}}}$	$0.1352 \pm 0.0267\,(0.0305)$	$0.0436 \pm 0.0267\,(0.0299)$

Table 6.2: Sample means $\pm$ sample standard deviations (RMSEs) of NNLS/RNNLS MESE f_{M} estimates and ML/PERK DESS f_{F} estimates in a two-compartment simulation where none of the associated estimators incur bias due to model mismatch. Sample statistics are computed over 7810 WM-like and 9162 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. Fig. 6.1 presents corresponding images.

6.4.2.2 Three-Compartment Simulation with Model Mismatch

We next simulated data to arise from three non-exchanging water compartments with myelin water-like $(500, 20)$ms, cellular water-like $(1000, 80)$ms, and free water-like $(3000, 3000)$ms (longitudinal, transverse) relaxation time constants selected based on [100, 158]. With this three-compartment ground truth, the aforementioned MESE MWF estimators could incur bias due to their bulk-T_1 assumption and the aforementioned DESS fast-fraction estimators could incur bias due to their two-compartment assumption. Thus f_{M} and f_{F} are not equivalent here and their estimates need not necessarily be comparable. We assigned (myelin, cellular, free) water-like fractions of $(0.15, 0.82, 0.03)$ in WM, and $(0.03, 0.94, 0.03)$ in GM. We simulated data otherwise exactly as detailed in Simulation 6.4.2.1 to yield MESE image datasets with SNR ranging from $24 - 795$ in WM and $29 - 862$ in GM and to yield DESS image datasets with SNR ranging from $24 - 221$ in WM and $30 - 241$ in GM, where SNR is computed via (6.53). We estimated f_{M} from noisy magnitude MESE images and known bulk T_1 and κ^{t} maps by solving NNLS (6.51) and RNNLS (6.52) problems as explained in Subsection 6.4.1. We estimated f_{F} from noisy magnitude DESS images and known κ^{t} maps using PERK and ML estimators, as explained in Subsections 6.4.1-6.4.2.1 respectively. NNLS and RNNLS respectively took 42.7s and 69.2s. PERK training and testing respectively took 34.2s and 1.1s while ML estimation took 17681s (nearly 5h).

Fig. 6.2 compares NNLS and RNNLS f_{M} estimates as well as ML and PERK f_{F} estimates alongside magnitude difference images with respect to the ground truth MWF. The PERK f_{F} estimator achieves the lowest errors in WM but overestimates in GM (as does the ML f_{F} estimator), causing reduced WM/GM contrast relative to other estimators. Unlike both f_{F} estimates, both f_{M} estimates visibly exhibit systematic error due to flip angle spatial variation, despite perfect knowledge of κ^{t}. All estimates are higher (though to vary-

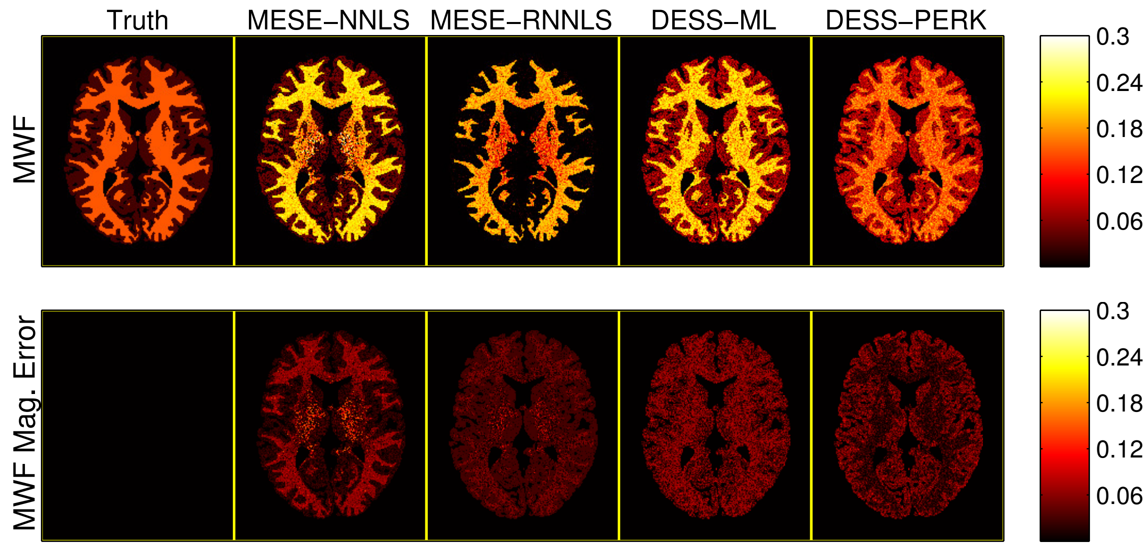


Figure 6.2: NNLS/RNNLS MESE f_{M} and ML/PERK DESS f_{F} estimates alongside corresponding magnitude error images, in a three-compartment simulation where any of the associated estimators could incur bias due to model mismatch. Voxels not assigned WM- or GM-like compartmental fractions are masked out in post-processing for display. Table 6.3 presents corresponding sample statistics.

	WM	GM
True $f_{\mathrm{M}} \equiv f_{\mathrm{F}}$	0.15	0.03
MESE-NNLS $\widehat{f}_{\mathrm{M}}$	0.1910 ± 0.0463 (0.0618)	0.0349 ± 0.0192 (0.0198)
MESE-RNNLS $\widehat{f}_{\mathrm{M}}$	0.1699 ± 0.0354 (0.0406)	0.00272 ± 0.00673 (0.02809)
DESS-ML $\widehat{f}_{\mathrm{F}}$	0.1987 ± 0.0275 (0.0559)	0.0632 ± 0.0280 (0.0434)
DESS-PERK $\widehat{f}_{\mathrm{F}}$	0.1576 ± 0.0243 (0.0254)	0.0754 ± 0.0231 (0.0510)

Table 6.3: Sample means $\pm$ sample standard deviations (RMSEs) of NNLS/RNNLS MESE f_{M} estimates and ML/PERK DESS f_{F} estimates in a three-compartment simulation where any of the associated estimators could incur bias due to model mismatch. Sample statistics are computed over 7810 WM-like and 9162 GM-like voxels. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. Fig. 6.2 presents corresponding images.

ing degrees) than corresponding estimates presented in Fig. 6.1, indicating some sensitivity to model mismatch. Except for PERK f_{F} estimates in WM and RNNLS f_{M} estimates in GM, all estimates exhibit greater error than corresponding estimates presented in Fig. 6.1, indicating that in most cases model mismatch is detrimental to estimation performance.

Table 6.3 compares sample statistics of NNLS and RNNLS f_{M} estimates as well as ML and PERK f_{F} estimates, computed over the same WM-like and GM-like ROIs as in

Table 6.2. Several estimates now differ from true values by more than one standard deviation, indicating significant bias due to model mismatch in these cases. The PERK f_F estimator is most accurate and achieves the lowest RMSE in WM, but also suffers from the highest RMSE in GM. The NNLS f_M estimator is most accurate and achieves the lowest RMSE in GM, but also suffers from the highest RMSE in WM. RNNLS f_M (PERK f_F) estimates are now both more accurate and more precise than NNLS f_M (ML f_F) estimates in WM, suggesting that regularization may be beneficial in cases of model mismatch. Perhaps surprisingly, RNNLS f_M and PERK f_F estimates do not differ significantly in WM (but do differ in GM) suggesting that these WM estimates may be comparable even when characterizing 3-compartment systems, at least for the nominal values selected here.

6.4.3 *In vivo* Experiments

We acquired *in vivo* data using a GE Discovery™ MR750 3.0T scanner with a 32-channel Nova Medical® receive head array. In a single scan session involving a healthy volunteer, we collected the DESS acquisition described in Subsection 6.3.2; a MESE acquisition for validation; an SPGR acquisition for separate bulk T_1 estimation; and a Bloch-Siegert (BS) acquisition for separate κ^t estimation. Each of these acquisitions is described next in turn.

We acquired DESS data by prescribing the optimized nominal flip angles and repetition times presented in Table 6.1 and holding all other scan parameters fixed across DESS scans. We achieved desired nominal flip angles by scaling a 9.0mm slab-selective Shinnar-Le Roux (SLR) pulse [89] of duration 3.0ms and time-bandwidth product 6. We interleaved RF excitations with 2 gradient dephasing phase cycles over a 3mm slice thickness to distinguish the DESS echoes. We acquired DESS data with a $200 \times 200 \times 8$ matrix over a $240 \times 240 \times 24$mm^3 field of view (FOV). Using a 31.25kHz readout bandwidth, we acquired 3D axial DESS data at minimum $T_\mathrm{E} \leftarrow 5.29$ms before and after RF excitations. To avoid slice-profile effects, we sampled $\mathbf{k}$-space over a 3D Cartesian grid. Including time to reach steady-state, the DESS acquisition took 3m15s scan time.

We acquired MESE data with nominally $90°$ excitation flip angles, achieved by scaling the same SLR pulse shape as above. A sequence of 32 identical nominally $180°$ refocusing pulses succeeded each excitation, where the time between excitation and first refocusing pulse peaks was fixed to the minimum possible $\frac{T_\mathrm{E}}{2} \leftarrow 4.6$ms and subsequent refocusing pulse peaks were separated by echo spacing $T_\mathrm{E} \leftarrow 9.2$ms. We designed each refocusing pulse as a 21.0mm slab-selective SLR pulse of duration 2.0ms and time-bandwidth product 2. We elected to use shaped refocusing pulses instead of shorter hard pulses to suppress unwanted signal outside the excitation slab due to imperfect refocusing. To suppress stimulated echo signal contributions, we flanked each refocusing pulse with a symmetric gradient crusher pair, where each crusher imparted 14 phase cycles across the 21.0mm refocusing slab. Immediately following the refocusing pulse train, we imparted 8 gradient dephasing phase cycles over a 3mm slice thickness to suppress residual transverse magnetization. To reduce scan time, we used a repetition time $T_{\mathrm{R}} \leftarrow 600$ms that is shorter than those used in recent works (*e.g.*, [154, 159]) and used separate bulk T_1 estimates to account for incomplete recovery. We acquired 3D MESE data over the same imaging volume and with the same resolution, readout bandwidth, and $\mathbf{k}$-space trajectory as was used for the DESS acquisition. We repeated the MESE scan twice to permit averaging in postprocessing for increased SNR. Including three prepended repetitions to approach steady-state, each MESE scan took 16m2s for a total MESE acquisition time of 32m4s.

We acquired SPGR data for separate bulk T_1 estimation. We varied across nine scans nominal flip angles from $5°$ to $45°$ with even increments and fixed all other scan parameters across scans. We achieved desired nominal flip angles by scaling the same SLR pulse shape used in the DESS acquisition. We acquired 3D data at minimal echo time $T_{\mathrm{E}} \leftarrow 5.1$ms over the same imaging volume and with the same resolution, readout bandwidth, and $\mathbf{k}$-space trajectory as was used in the DESS acquisition. We implemented RF spoiling by imparting 8 gradient dephasing phase cycles over a 3mm slice thickness immediately following each readout and by RF phase cycling with an RF phase increment that increases by $117°$ each $T_{\mathrm{R}} \leftarrow 13.1$ms-long repetition [19]. Including time to reach steady-state, the SPGR acquisition took 3m32s scan time.

We acquired a pair of BS-shifted SPGR scans [90] for separate flip angle scaling κ^{t} estimation. We modified the 3D SPGR sequence just described by inserting a ± 4kHz off-resonant Fermi pulse (of 9.0ms duration and with 0.05G peak amplitude) immediately following on-resonant excitation and immediately prior to readout. This extended the echo time to $T_{\mathrm{E}} \leftarrow 15.0$ms. We also conservatively extended the repetition time to $T_{\mathrm{R}} \leftarrow 300$ms to prevent excess RF heating. We used a small $5°$ nominal excitation flip angle for reduced contrast in BS images and thereby smoother κ^{t} estimates. We acquired BS data with a reduced $200 \times 50 \times 8$ matrix. All other scan parameters were the same as for the SPGR acquisition. Including time to reach steady-state, the BS acquisition took 4m30s scan time.

We reconstructed all raw coil images via 3D Fourier transform and subsequently processed only one image slice centered within the excitation slab. We upsampled BS coil images along the phase-encoding direction to the same image size as other coil images, using intermediate zero-padding to suppress ringing. We jointly coil-combined all coil images *within* each of the four acquisitions using a natural extension of [92] for the case of multiple datasets, detailed in Appendix A; however, we did not coil-combine sharing coil

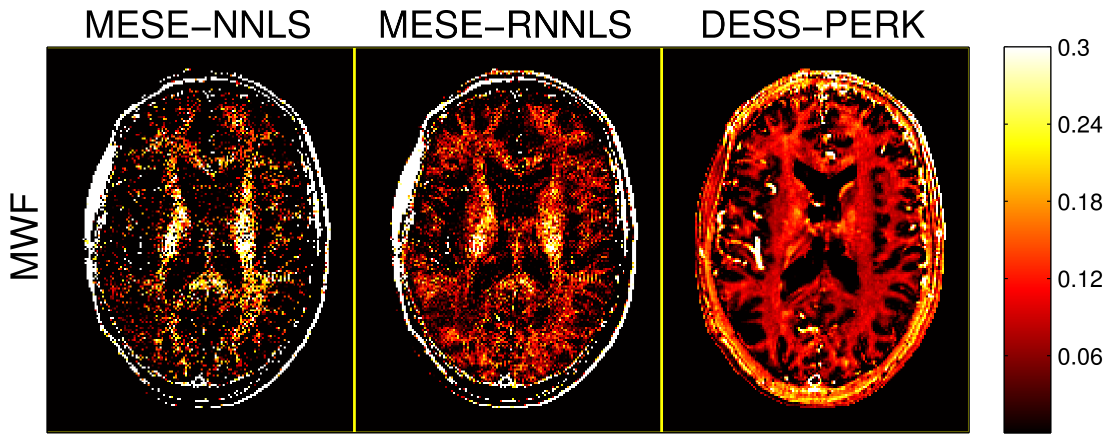


Figure 6.3: Representative NNLS and RNNLS f_{M} estimates from a MESE acquisition alongside a PERK f_{F} estimate from a precision-optimized DESS acquisition, in the brain of a healthy volunteer. Using similar signal reception imaging parameters, MESE f_{M} estimates required 40m6s total scan time while DESS f_{F} estimates required 7m45s total scan time. PERK f_{F} estimates exhibit less WM variation and more clearly delineate cortical WM/GM boundaries than MESE f_{M} estimates. Table 6.4 presents corresponding sample statistics computed over manually selected WM and GM ROIs.

data *across* acquisitions. We estimated flip angle spatial variation κ^{t} maps by normalizing and calibrating regularized transmit field estimates [91] from complex coil-combined BS images. We estimated bulk T_1 maps from magnitude coil-combined SPGR images and κ^{t} maps using variable projection method [25] and grid search (the one-dimensional grid search used 50 dictionaries, each computed using 1000 logarithmically-spaced T_1 samples between 10ms and 3000ms). To address bulk motion between acquisitions, we rigidly registered coil-combined MESE and DESS images as well as κ^{t}, T_1 maps to one coil-combined MESE first-echo image. After registration, we averaged MESE images voxel-by-voxel across scan repetitions to increase effective SNR. We estimated f_{M} from magnitude averaged MESE images and κ^{t}, T_1 maps by solving NNLS (6.51) and RNNLS (6.52) problems as explained in Subsection 6.4.1. We estimated f_{F} from magnitude DESS images and κ^{t} maps by applying PERK as explained in Subsection 6.4.1. NNLS and RNNLS respectively took 57.4s and 122.9s. PERK training and testing respectively took 37.1s and 1.0s.

Fig. 6.3 compares NNLS and RNNLS f_{M} estimates from MESE scans as well as PERK f_{F} estimates from optimized DESS scans. PERK f_{F} estimates exhibit less WM variation and more clearly delineate cortical WM/GM boundaries than MESE f_{M} estimates. RNNLS f_{M} estimates are visibly lower than NNLS f_{M} estimates due to regularization but exhibit reduced WM variation, somewhat improving visualization of WM tracts. RNNLS f_{M} and PERK f_{F} estimates appear visually similar in lateral WM regions, but both NNLS and

ROI	MESE-NNLS $\widehat{f}_{\mathrm{M}}$	MESE-RNNLS $\widehat{f}_{\mathrm{M}}$	DESS-PERK $\widehat{f}_{\mathrm{F}}$
AR WM	0.09 ± 0.096	0.074 ± 0.055	0.117 ± 0.019
AL WM	0.066 ± 0.086	0.054 ± 0.041	0.100 ± 0.0119
PR WM	0.047 ± 0.074	0.044 ± 0.041	0.093 ± 0.019
PL WM	0.116 ± 0.098	0.075 ± 0.050	0.0870 ± 0.0114
IC WM	0.211 ± 0.133	0.178 ± 0.083	0.111 ± 0.0241
AC GM	0.007 ± 0.024	0.010 ± 0.017	0.019 ± 0.045

Table 6.4: *Left*: WM/GM ROIs, overlaid on a representative anatomical MESE first-echo image. Separate lateral WM ROIs are distinguished by anterior-right (AR), anterior-left (AL), posterior-right (PR), and posterior-left (PL) directions and are respectively comprised of 90, 79, 182, and 201 voxels. Two internal capsule (IC) polygons are pooled into a single medial WM ROI comprised of 347 voxels. Three small anterior cortical (AC) GM polygons are pooled into a single GM ROI comprised of 78 voxels. *Right*: Sample means $\pm$ sample standard deviations of NNLS/RNNLS f_{M} estimates from a MESE acquisition as well as PERK f_{F} estimates from an optimized DESS acquisition, computed over WM/GM ROIs. Each sample statistic is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [3]. Fig. 6.3 presents corresponding images.

RNNLS f_{M} estimates are elevated in medial regions. MESE $\widehat{f}_{\mathrm{M}}$ overestimation in internal capsules (IC) has been attributed to overlap in NNLS T_2 spectrum estimates of the myelin water and cellular water T_2 peaks [159]. We additionally observe that MESE f_{M} estimates exhibit similar spatial variation here versus in simulations (*cf.* Figs. 6.1-6.2) suggesting that some WM spatial variation in MESE f_{M} estimates may be attributable to flip angle variation, despite compensation for transmit field inhomogeneity.

Table 6.4 summarizes sample statistics of NNLS/RNNLS f_{M} estimates from MESE scans and PERK f_{F} estimates from optimized DESS scans, separately computed over four lateral WM ROIs containing 90, 79, 182, and 201 voxels; one pooled medial IC WM ROI containing 347 voxels; and one pooled anterior cortical (AC) GM ROI containing 78 voxels. PERK f_{F} estimates exhibit the lowest variation within WM ROIs and the most similar sample means across WM ROIs. NNLS and RNNLS f_{M} sample means are significantly higher in the IC WM ROI than in lateral WM ROIs, possibly due to overlap in NNLS T_2 spectrum peaks and/or to flip angle spatial variation (as described in the previous paragraph). PERK f_{F} sample means differ significantly from RNNLS f_{M} sample means in several WM ROIs, though PERK WM f_{F} sample means are consistently comparable with literature MESE measurements (*e.g.*, see [156] for a review). Neither the NNLS/RNNLS $\widehat{f}_{\mathrm{M}}$ nor PERK $\widehat{f}_{\mathrm{F}}$ estimators measured significant myelin water content in AC GM.

6.5 Discussion

Simulations and experiments demonstrate the feasibility of myelin water content quantification from a fast precision-optimized DESS MR acquisition. It remains challenging to assess whether DESS f_{F} estimates are comparable to MESE f_{M} estimates *in vivo* because of high MESE f_{M} estimation variation. For greater confidence in comparisons, future *ex vivo* studies in excised brain tissue would allow for longer MESE scans without increasing motion-induced errors and would thereby enable more precise MESE f_{M} estimation. Studies at higher field strengths would also enable more precise estimation due to higher SNR. Nevertheless, the experiments described herein are the first to demonstrate *in vivo* lateral WM myelin water content estimates from a fast SS MR acquisition that are at all similar to conventional MWF estimates from a slower MESE MR acquisition.

Taken together with the results presented in Ch. 5, experiments herein also provide evidence that the PERK estimator [8] can maintain reasonable accuracy while scaling gracefully with the number of latent parameters per voxel. In simulations, PERK consistently took at least $500\times$ less time and consistently achieved lower WM RMSE than an ML estimator achieved via unrealistically narrow grid search around the ground truth. In preliminary experiments on other precision-optimized *in vivo* datasets from the same healthy volunteer, PERK took comparable time and produced similar f_{F} estimates as reported here while a more realistically constrained grid search took about 68 CPU-days running on 24 nodes of a computing cluster and did not produce reasonable f_{F} estimates. To avoid the confounding possibility of *in vivo* ML errors due to multiple global minima, we included narrowly constrained ML grid search estimates from a 3-compartment simulation here and from poor ML accuracy conclude that the ML estimator is likely more sensitive to model mismatch than the PERK estimator for this application. These results suggest that PERK scales much better than grid search with the number of latent parameters per voxel and that PERK may be more robust than grid search to multi-compartmental modeling errors.

Despite freedom to design arbitrary combinations of SPGR and DESS scans, the optimized acquisition used here as well as several other unreported acquisitions designed under different total time constraints consisted either entirely or mostly of DESS scans. Since the two-compartment SPGR signal models used in acquisition design depend on $T_{1,\mathrm{F}}, T_{1,\mathrm{S}}$ but not $T_{2,\mathrm{F}}, T_{2,\mathrm{S}}$, DESS-dominated scan designs suggest that multi-compartmental T_2 effects give rise to f_{F} sensitivity in SS sequences more so than multi-compartmental T_1 effects. Somewhat surprisingly, reported and unreported precision-optimized acquisitions also exhibit substantial T_{R} variation across scans, even at the expense of fewer scans than possible under time constraints. In further unreported studies, we investigated this phenomenon by repeating scan optimization while implicitly constraining repetition times to be minimal. We consistently observed substantial ($\sim$10-20$\%$) degradation in expected f_{F} relative standard deviation, suggesting that T_{R} variation (in addition to flip angle variation) is important for designing acquisitions that enable precise f_{F} estimation.

Though Subsection 6.2 modeled the effects of first-order physical exchange in two-compartment SPGR and DESS models, experiments thereafter used non-exchanging models to work with closed-form signal models and signal gradients during acquisition design and to keep consistent with standard MESE model assumptions. There is growing evidence however that some significant physical exchange across the myelin bilayers likely persists in cerebral WM (see [157] for a very recent review). A thorough investigation of the sensitivity of f_{F} estimates from non-exchanging two-compartment SPGR/DESS models to realistic physical exchange rates is a topic for further research.

Even with high SNR, differences in model assumptions, objective functions, and estimation algorithms may limit the quantitative comparability of DESS f_{F} imaging as implemented here and MESE f_{M} imaging as implemented here and in original works [100, 168]. For more similar model assumptions, one could attempt to estimate from a suitably optimized DESS acquisition a T_2 (or joint (T_1, T_2)) distribution using two-, three-, or higher-compartment models and correspondingly estimate from MESE data a more coarsely sampled T_2 (or joint (T_1, T_2)) distribution. If κ^{t}, T_1 maps are known and non-exchanging additive models are employed, one could estimate T_2 distributions from both MESE and DESS data using NNLS or RNNLS objective functions. With weaker model assumptions that cause signal models to be nonlinear in unknowns, one could estimate distributions using PERK. This work focused on demonstrating the feasibility of myelin water quantification using a simple two-compartment model of a fast DESS acquisition; estimating more unknowns from more complicated (MESE or DESS) models will at least necessitate more scans and may still prove challenging in practice.

6.6 Conclusion

This chapter has introduced a fast SS MRI acquisition for precise myelin water imaging. The acquisition consists of three DESS scans whose flip angles and repetition times have been optimized under an aggressive time constraint to enable precise estimation of the faster-relaxing signal fraction f_{F} in a simple two-compartment DESS model. Simulations without model mismatch demonstrate that PERK and ML f_{F} estimates from the proposed DESS acquisition exhibit comparable RMSEs, but PERK is more than $500\times$ faster. Simulations with modest levels of model mismatch demonstrate that conventional MESE f_{M}

estimates are sensitive to unaccounted variable T_1-recovery rates across compartments and accounted flip angle spatial variation while DESS f_{F} estimates are sensitive to relaxation in an unaccounted third compartment, suggesting limited quantitative comparability of MESE f_{M} and DESS f_{F} WM estimates. *In vivo* experiments are nevertheless the first to demonstrate lateral WM myelin water content estimates from a fast (3m15s) SS acquisition that are similar to conventional MWF estimates from a slower (32m4s) MESE acquisition.

CHAPTER 7

Future Work

This chapter suggests avenues for further QMRI research, focusing on relatively broad future directions that this thesis has not explored. Discussion sections within main body chapters offer more focused ideas for topic-specific extensions, whereas the appendices organize partially investigated but still immature topics that are less related to QMRI.

7.1 Combine PERK with Image Reconstruction

This thesis has considered parameter estimation to be separate from image reconstruction. This separation affords fast data processing but may leave room for improved estimation performance, especially when raw data is undersampled. One could instead seek to estimate parameters directly from raw data. One approach to combining image reconstruction with PERK-based estimation might seek to solve the joint optimization problem

$$\left(\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}\right) \in \left\{ \arg \min_{\substack{\mathbf{X} \in \mathbb{C}^{L \times V} \\ \mathbf{Y} \in \mathbb{C}^{D \times V}}} \|\mathbf{D} - \mathbf{Y}\mathbf{A}\|_{\mathrm{F}}^2 + \beta \sum_{v=1}^{V} \|\mathbf{h}(\mathbf{y}_v, \boldsymbol{\nu}_v) + \mathbf{b} - \mathbf{x}_v\|_2^2 \right\}, \tag{7.1}$$

where $\mathbf{X} := [\mathbf{x}_1, \ldots, \mathbf{x}_V] \in \mathbb{C}^{L \times V}$ and $\mathbf{Y} := [\mathbf{y}_1, \ldots, \mathbf{y}_V] \in \mathbb{C}^{D \times V}$ respectively collect L latent parameters and D image datasets at V voxels; $\mathbf{h} : \mathbb{C}^{D+N} \mapsto \mathbb{C}^L$ and $\mathbf{b} \in \mathbb{C}^L$ together denote a pre-trained PERK regression function with offset; $\mathbf{D} \in \mathbb{C}^{D \times K}$ collects D raw $\mathbf{k}$-space datasets each acquired with K samples; $\mathbf{A} \in \mathbb{C}^{V \times K}$ denotes the MRI system matrix (that models receive coil sensitivity spatial variation and $\mathbf{k}$-space sampling); $\boldsymbol{\nu}_v$ denotes a known parameter at the vth voxel; and β is a free parameter that balances cost function terms. Here, the first term enforces image fidelity to raw data and the second term entangles image reconstruction and parameter estimation.

For continuously differentiable kernels, (7.1) is amenable to iterative local optimization via alternating minimization. One simple algorithm iterates the following updates:

$$\mathbf{X}^{(i+1)} \leftarrow \left[\mathbf{h}\left(\mathbf{y}_1^{(i)}, \boldsymbol{\nu}_1\right) + \mathbf{b}, \ldots, \mathbf{h}\left(\mathbf{y}_V^{(i)}, \boldsymbol{\nu}_1\right) + \mathbf{b}\right] \tag{7.2}$$

$$\mathbf{Y}^{(i+1)} \leftarrow \arg \min_{\mathbf{Y} \in \mathbb{C}^{D \times V}} \|\mathbf{D} - \mathbf{YA}\|_{\mathrm{F}}^2 + \beta \sum_{v=1}^{V} \left\|\mathbf{h}(\mathbf{y}_v, \boldsymbol{\nu}_v) + \mathbf{b} - \mathbf{x}_v^{(i+1)}\right\|_2^2 \tag{7.3}$$

where $(\cdot)^{(i)}$ denotes the ith iterate. Latent parameter update (7.2) applies PERK voxel-wise at the current image iterate. Image update (7.3) enforces consistency not only with data but also with latent parameter iterates. Locally solving the inner problem within (7.3) via gradient-based optimization requires (often simple) kernel gradients but does not require signal model gradients. Thus, (7.1) or similar variations may be useful even when analytical signal models are cumbersome or altogether unavailable.

7.2 Exploit Off-Resonance for Myelin Water Imaging

Though early sections in Chapter 6 provided a simple model of off-resonance distribution variation across intravoxel compartments, experiments therein ultimately used simpler magnitude signal models that neglected off-resonance effects. However, off-resonance distributions do often differ significantly across compartments in cerebral tissue [163, 164], so accounting for compartment-specific off-resonance effects could aid in better distinguishing compartments and could thereby enable further-improved myelin water imaging. For designing off-resonance-informed myelin water imaging acquisitions, it is reasonable to consider pulse sequences whose acquisition parameters can strongly influence signal sensitivity to off-resonance effects. In this respect, the small-tip fast recovery (STFR) sequence [171] may be well-suited for off-resonance-informed myelin water imaging because its tip-up pulse magnitude and phase provide additional degrees of freedom by which to sensitize acquisitions to compartmental off-resonance effects. Off-resonance-informed myelin water imaging using (spoiled) STFR sequences is an active area of research in our group.

7.3 Correlate with Other Myelin Biomarkers

As discussed in Section 6.5, fundamental differences between DESS f_{F} and MESE MWF imaging may limit their quantitative comparability. To build evidence that DESS f_{F} imaging is nevertheless a specific biomarker for intact myelin content, we are also interested in how f_{F} correlates with other myelin biomarkers. One contending noninvasive marker arises from MR pulse sequences sensitized to the inhomogeneous magnetization transfer (ihMT) effect [172], which has recently been shown to be specific to the large membrane lipids that comprise much of myelin [173, 174]. Multi-compartmental and ihHT MRI markers could be successively compared through *ex vivo*, healthy volunteer, and patient studies. Outside MRI, invasive measurements from histology have been used to study myelin (as in *e.g.*, [149, 150]) and could serve as a gold-standard *in situ* marker. MRI and histological markers could be compared by correlating respective *ex vivo* and *in situ* studies.

7.4 Translate into a Clinical Setting

Moving towards clinical translation of PERK-based DESS myelin water imaging would first require building greater evidence that f_{F} is a specific biomarker for intact myelin content. This would require *in situ* studies to more definitively establish whether or not f_{F} estimates correlate with gold-standard histological markers. Repetition studies in one or more healthy volunteers would provide evidence of f_{F} reproducibility. Longitudinal research studies in infants or patients would respectively provide evidence of whether DESS f_{F} images are specifically indicative of myelination or demyelination.

Enabling clinical studies would require packaging PERK into a streamlined and automatic workflow. At least with fully sampled acquisitions, PERK could be implemented to separately follow image reconstruction and transmit field mapping, taking coil-combined images and flip angle maps as inputs and giving f_{F} (and possibly other parameter) estimates as outputs. Though the experiments herein used fast (less than 1m) online training, PERK could be trained offline with minor implementation modifications detailed in Section 5.9, so long as approximate prior knowledge of noise statistics is available.

APPENDIX A

Multiple-Dataset Complex Coil Combination

A.1 Introduction

This appendix introduces an unpublished algorithm for simultaneously coil-combining a collection of MR coil image datasets without prior knowledge of coil sensitivity maps. The algorithm expects that MR coil datasets arise from a single acquisition with little inter-scan motion but allows for variable MR contrasts across datasets. Such multiple contrast, multiple receive coil acquisitions arise naturally in many QMRI applications, including the ones studied in Chapters 4-6. The algorithm extends a similar sensitivity-blind method for coil-combining a single dataset [92] to exploit coil sensitivity redundancy across datasets and thereby improve problem conditioning over coil-combining multiple datasets separately.

A.2 Problem Formulation

Suppose that we are presented image-domain coil data from C receive coils, D contrasts, and V voxel locations. A simple model for each single-coil, single-contrast image reads

$$\mathbf{y}_{c,d} = \mathrm{diag}(\boldsymbol{\kappa}_c^{\mathrm{r}})\mathbf{x}_d + \boldsymbol{\epsilon}_{c,d}, \tag{A.1}$$

where $\mathbf{y}_{c,d} \in \mathbb{C}^V$ denotes a noisy image dataset at the cth coil and dth contrast; $\boldsymbol{\kappa}_c^{\mathrm{r}} \in \mathbb{C}^V$ denotes the unknown sensitivity of the cth receive coil; $\mathbf{x}_d \in \mathbb{C}^V$ denotes the unknown magnetization of the dth contrast; $\boldsymbol{\epsilon}_{c,d} \in \mathbb{C}\mathcal{N}(\mathbf{0}_V, \boldsymbol{\Sigma})$ is complex Gaussian noise; $c \in \{1, \dots, C\}$; and $d \in \{1, \dots, D\}$. This model allows for spatial correlations but for ease of exposition ignores correlations across coils and correlations across contrasts.

We seek to estimate $\mathbf{x}_1, \dots, \mathbf{x}_D$ from $\mathbf{y}_{1,1}, \dots, \mathbf{y}_{C,D}$ without knowledge of $\boldsymbol{\kappa}_1^{\mathrm{r}}, \dots, \boldsymbol{\kappa}_C^{\mathrm{r}}$. We utilize a cost function that not only penalizes data inconsistency but also encourages spatially smooth sensitivity profiles through quadratic regularization:

$$\Psi(\boldsymbol{\kappa}_1^{\mathrm{r}}, \ldots, \boldsymbol{\kappa}_C^{\mathrm{r}}, \mathbf{x}_1, \ldots, \mathbf{x}_D) := \sum_{c=1}^{C} \left(\sum_{d=1}^{D} \|\mathbf{y}_{c,d} - \mathrm{diag}(\boldsymbol{\kappa}_c^{\mathrm{r}})\mathbf{x}_d\|_{\boldsymbol{\Sigma}^{-1}}^2 + \beta_c \|\mathbf{D}\boldsymbol{\kappa}_c^{\mathrm{r}}\|_2^2, \right) \quad \text{(A.2)}$$

where $\mathbf{D}$ is a pure-real first-order finite differencing matrix and β_c is a regularization parameter. Observe that Ψ is not coercive in $\mathbf{x}_1, \ldots, \mathbf{x}_D$: scaling each $\mathbf{x}_1, \ldots, \mathbf{x}_D$ by α while scaling each $\boldsymbol{\kappa}_1^{\mathrm{r}}, \ldots, \boldsymbol{\kappa}_C^{\mathrm{r}}$ by $\frac{1}{\alpha}$ only reduces Ψ for arbitrarily large scale factor α. To resolve this scale ambiguity, we seek to minimize (A.2) under constraints:

$$\begin{aligned} (\widehat{\boldsymbol{\kappa}}_1^{\mathrm{r}}, \ldots, \widehat{\boldsymbol{\kappa}}_C^{\mathrm{r}}, \widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_D) &\in \left\{ \arg \min_{\substack{\boldsymbol{\kappa}_1^{\mathrm{r}}, \ldots, \boldsymbol{\kappa}_C^{\mathrm{r}} \in \mathbb{C}^V \\ \mathbf{x}_1, \ldots, \mathbf{x}_D \in \mathbb{C}^V}} \Psi(\boldsymbol{\kappa}_1^{\mathrm{r}}, \ldots, \boldsymbol{\kappa}_C^{\mathrm{r}}, \mathbf{x}_1, \ldots, \mathbf{x}_D) \right\} \\ \text{such that} \qquad & \mathbf{1}_V^{\mathsf{T}} \mathbf{x}_d = c_d \qquad \forall d \in \{1, \ldots, D\}, \end{aligned} \quad \text{(A.3)}$$

where $\mathbf{1}_V$ is length-V vector of ones and $c_d \in \mathbb{C}$ is a fixed constant.

A.3 Alternating Minimization Algorithm

We solve (A.3) by alternating between updating coil sensitivity variables while holding contrast variables fixed and updating contrast variables while holding coil sensitivity variables fixed. We next describe each of these updates in turn.

Because we assumed for simplicity that noise is uncorrelated across receive coils, each of the C coil updates can be updated in parallel. We update the cth coil variable by minimizing a quadratic cost function:

$$\boldsymbol{\kappa}_c^{\mathrm{r}(i+1)} \leftarrow \arg \min_{\boldsymbol{\kappa}_c^{\mathrm{r}} \in \mathbb{C}^V} \sum_{d=1}^{D} \left\| \mathbf{y}_{c,d} - \mathrm{diag}\left(\mathbf{x}_d{}^{(i)}\right) \boldsymbol{\kappa}_c^{\mathrm{r}} \right\|_{\boldsymbol{\Sigma}^{-1}}^2 + \beta_c \|\mathbf{D}\boldsymbol{\kappa}_c^{\mathrm{r}}\|_2^2. \quad \text{(A.4)}$$

where $(\cdot)^{(i)}$ denotes the ith iterate. For moderate V, (A.4) can be solved directly:

$$\boldsymbol{\kappa}_c^{\mathrm{r}(i+1)} \leftarrow \left(\mathrm{diag}\left(\mathbf{x}_d{}^{(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathrm{diag}\left(\mathbf{x}_d{}^{(i)}\right) + \beta_c \mathbf{D}^{\mathsf{H}} \mathbf{D} \right)^{\dagger} \mathrm{diag}\left(\mathbf{x}_d{}^{(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{y}_{c,d}. \quad \text{(A.5)}$$

For larger V with non-diagonal $\boldsymbol{\Sigma}$, more sophisticated inner updates may be necessary.

Because we assumed that noise is also uncorrelated across contrasts, each of the D contrast variables can be updated in parallel. We update the dth contrast variable at the

$(i+1)$th iteration by solving a constrained problem:

$$\mathbf{x}_d^{(i+1)} \leftarrow \arg \min_{\mathbf{x}_d \in \mathbb{C}^V} \sum_{c=1}^{C} \left\| \mathbf{y}_{c,d} - \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right) \mathbf{x}_d \right\|_{\boldsymbol{\Sigma}^{-1}}^2 \text{ such that } \mathbf{1}_V^{\mathsf{T}} \mathbf{x}_d = c_d. \tag{A.6}$$

We solve (A.6) using the method of Lagrange multipliers. The Lagrangian is

$$\Lambda(\mathbf{x}_d, \lambda_d) := \sum_{c=1}^{C} \left\| \mathbf{y}_{c,d} - \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right) \mathbf{x}_d \right\|_{\boldsymbol{\Sigma}^{-1}}^2 + 4 \operatorname{Re}\left(\bar{\lambda}_d \left(\mathbf{1}_V^{\mathsf{T}} \mathbf{x}_d - c_d\right)\right), \tag{A.7}$$

where $\lambda_d \in \mathbb{C}$ is a Lagrange multiplier; $\operatorname{Re}(\cdot)$ extracts the real component of its argument; and $\bar{(\cdot)}$ denotes complex conjugate. Setting $\nabla_{\mathbf{x}_d} \Lambda(\mathbf{x}_d, \lambda_d) = \mathbf{0}_V$ yields the update

$$\mathbf{x}_d^{(i+1)} \leftarrow \left(\sum_{c=1}^{C} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right) \right)^{\dagger} \left(\sum_{c=1}^{C} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{y}_{c,d} - \lambda_d \mathbf{1}_V \right) \tag{A.8}$$

where $(\cdot)^{\dagger}$ denotes pseudoinverse and λ_d is at each iteration updated such that the constraint remains satisfied:

$$\lambda_d^{(i+1)} \leftarrow \frac{\mathbf{1}_V^{\mathsf{T}} \left(\sum_{c=1}^{C} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right) \right)^{\dagger} \left(\sum_{c=1}^{C} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{y}_{c,d} \right) - c_d}{\mathbf{1}_V^{\mathsf{T}} \left(\sum_{c=1}^{C} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right)^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \operatorname{diag}\left(\boldsymbol{\kappa}_c^{\mathrm{r}(i)}\right) \right)^{\dagger} \mathbf{1}_V}. \tag{A.9}$$

Observe that matrix divisions required to implement (A.9) need not depend on contrast index d. Thus D contrast updates require $D+1$ matrix division subproblems per iteration, of which D divisions are amenable to parallelization.

Since (A.3) is non-convex, solutions will depend on initialization. We initialize each contrast separately using the square root of the sum over that contrast's magnitude coil images squared, a popular conventional estimator:

$$\mathbf{x}_d^{(0)} \leftarrow \operatorname{diag}\left(\sqrt{\sum_{c=1}^{C} \left| \mathbf{y}_{c,d} \right|^2} \right) e^{j \angle \mathbf{y}_{1,d}}, \tag{A.10}$$

where $\sqrt{\cdot}$, $|\cdot|$, and $\angle(\cdot)$ respectively denote element-wise square root, absolute value, and phase angle operators. Here, assigning phase to ${\mathbf{x}_d}^{(0)}$ using the phase of the first coil's data causes the phase of coil sensitivity estimates to be relative to the first coil's phase.

APPENDIX B

SS-Informed RF Pulse Design

B.1 Introduction

Modern radiofrequency (RF) pulse design methods often relate a desired magnetization pattern to the underlying RF pulse and excitation gradients. Conventional techniques assume negligible residual transverse magnetization immediately prior to excitation to show that under the small-excitation approximation, this relation is linear with respect to the RF pulse [175]. For fixed gradients, this facilitates rapid pulse design algorithms using linear filter design principles [176, 89] or iterative algorithms [177, 178]. In practice, pulse designers can realize near-complete transverse relaxation prior to excitation with long repetition times, large unbalanced spoiler gradients, or tip-up pulses [171]. However, all of these methods require additional scan time beyond excitation and data acquisition.

This appendix investigates the small-excitation RF pulse design problem when the usual assumption of near-complete decay of residual transverse magnetization is not taken. Omitting this assumption reveals the influence of RF pulses on the steady-state (SS) transverse magnetization rather than on the conventionally modeled single-repetition transverse magnetization. Because the former is more directly related to the received signal, such SS-informed RF pulse design might allow for more accurate excitation patterns.

B.2 Signal Model

We begin with the Bloch equations in a non-inertial reference frame rotating clockwise about the z-axis at the Larmor frequency:

$$\frac{\partial}{\partial t}m'_{xy}(\mathbf{r},t) = i\gamma\big(m'_z(\mathbf{r},t)b'_{xy}(\mathbf{r},t) - m'_{xy}(\mathbf{r},t)b'_z(\mathbf{r},t)\big) - \frac{m'_{xy}(\mathbf{r},t)}{T_2(\mathbf{r})}; \tag{B.1}$$

$$\frac{\partial}{\partial t}m'_z(\mathbf{r},t) = \gamma\big(m'_x(\mathbf{r},t)b'_y(\mathbf{r},t) - m'_y(\mathbf{r},t)b'_x(\mathbf{r},t)\big) - \frac{m'_z(\mathbf{r},t) - m_0(\mathbf{r})}{T_1(\mathbf{r})}. \tag{B.2}$$

Here, $m'_{xy}(\mathbf{r},t) \equiv m'_x(\mathbf{r},t) + im'_y(\mathbf{r},t)$ and $m'_z(\mathbf{r},t)$ are respectively the rotating-frame transverse and longitudinal magnetization at position $\mathbf{r} \in \mathbb{R}^3$ and time $t \geq 0$; $b'_{xy}(\mathbf{r},t) \equiv b'_x(\mathbf{r},t) + ib'_y(\mathbf{r},t)$ and $b'_z(\mathbf{r},t)$ are respectively the apparent transverse and apparent longitudinal magnetic field; $T_1(\mathbf{r})$ and $T_2(\mathbf{r})$ are spin-lattice and spin-spin relaxation time constants; $m_0(\mathbf{r})$ is the equilibrium magnetization; γ is the gyromagnetic ratio; and $i := \sqrt{-1}$. These coupled differential equations are challenging to solve outright. To proceed, we take two simplifying assumptions:

- We assume that RF pulses do not strongly perturb the longitudinal magnetization from its initial state at time $t \leftarrow t_0$, *i.e.* $m'_z(\mathbf{r},t)b'_{xy}(\mathbf{r},t) \approx m'_z(\mathbf{r},t_0)b'_{xy}(\mathbf{r},t)$ for all $t \in [t_0, t_0 + T_{\mathrm{P}}]$ where $T_{\mathrm{P}} \geq 0$ denotes RF pulse duration. This small-excitation assumption differs from the conventional small-excitation assumption [175] (that instead approximates $m'_z(\mathbf{r},t) \approx m_0(\mathbf{r})$ for all $t \in [t_0, t_0 + T_{\mathrm{P}}]$) in that it allows for memory of the longitudinal magnetization prior to excitation.
- We assume that the apparent transverse magnetic field separates in position and time, *i.e.* $b'_{xy}(\mathbf{r},t) \approx \kappa^{\mathrm{t}}(\mathbf{r})b_1(t)$ where $\kappa^{\mathrm{t}}(\mathbf{r}) \in \mathbb{C}$ is the RF coil spatial variation and $b_1(t) \equiv b_{1,x}(t) + ib_{1,y}(t)$ is the RF excitation envelope.

With these assumptions, the rotating-frame Bloch equations read

$$\frac{\partial}{\partial t}m'_{xy}(\mathbf{r},t) = i\gamma\big(m'_z(\mathbf{r},t_0)\kappa^{\mathrm{t}}(\mathbf{r})b_1(t) - m'_{xy}(\mathbf{r},t)b'_z(\mathbf{r},t)\big) - \frac{m'_{xy}(\mathbf{r},t)}{T_2(\mathbf{r})}; \tag{B.3}$$

$$\frac{\partial}{\partial t}m'_z(\mathbf{r},t) = \gamma\Big(\mathrm{Re}\big(\bar{\kappa}^{\mathrm{t}}(\mathbf{r})m'_{xy}(\mathbf{r},t)\big)b_{1,y}(t) - \mathrm{Im}\big(\bar{\kappa}^{\mathrm{t}}(\mathbf{r})m'_{xy}(\mathbf{r},t)\big)b_{1,x}(t)\Big) - \frac{m'_z(\mathbf{r},t) - m_0(\mathbf{r})}{T_1(\mathbf{r})}, \tag{B.4}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ respectively extract real and imaginary components and $\bar{(\cdot)}$ denotes complex conjugation. Expressing (B.3)-(B.4) in terms of $\tilde{m}'_{xy}(\mathbf{r},t) \equiv \tilde{m}'_x(\mathbf{r},t) + i\tilde{m}'_y(\mathbf{r},t) := \bar{\kappa}^{\mathrm{t}}(\mathbf{r})m'_{xy}(\mathbf{r},t)$ allows further simplification:

$$\frac{\partial}{\partial t}\tilde{m}'_{xy}(\mathbf{r},t) = i\gamma\Big(m'_z(\mathbf{r},t_0)\big|\kappa^{\mathrm{t}}(\mathbf{r})\big|^2 b_1(t) - \tilde{m}'_{xy}(\mathbf{r},t)b'_z(\mathbf{r},t)\Big) - \frac{\tilde{m}'_{xy}(\mathbf{r},t)}{T_2(\mathbf{r})}; \tag{B.5}$$

$$\frac{\partial}{\partial t}m'_z(\mathbf{r},t) = \gamma\big(\tilde{m}'_x(\mathbf{r},t)b_{1,y}(t) - \tilde{m}'_y(\mathbf{r},t)b_{1,x}(t)\big) - \frac{m'_z(\mathbf{r},t) - m_0(\mathbf{r})}{T_1(\mathbf{r})}. \tag{B.6}$$

The small-excitation assumption above decouples (B.5) from (B.6). Thus we can solve directly for $\tilde{m}'_{xy}(\mathbf{r},t)$ and then obtain an expression for $m'_z(\mathbf{r},t)$ via substitution into (B.6).

We solve (B.5) using the method of integrating factors. The solution for times $t \geq t_0$ is

$$\begin{aligned} \tilde{m}'_{xy}(\mathbf{r}, t) &= \tilde{m}'_{xy}(\mathbf{r}, t_0) e^{i\phi(\mathbf{r}, t; t_0)} e^{-(t-t_0)/T_2(\mathbf{r})} \\ &\quad + i\gamma m'_z(\mathbf{r}, t_0) \left|\kappa^{\mathrm{t}}(\mathbf{r})\right|^2 \int_{t_0}^{t} b'_{xy}(t') e^{i\phi(\mathbf{r}, t; t')} e^{-(t-t')/T_2(\mathbf{r})} \, \mathrm{d}\, t', \end{aligned} \tag{B.7}$$

where $\phi(\mathbf{r}, t; t') := -\gamma \int_{t'}^{t} b'_z(\mathbf{r}, \tau) \, \mathrm{d}\, \tau$ denotes the phase accumulation from t' to t due to off-resonance effects. Note that if one assumes negligible transverse magnetization from prior excitations (*i.e.*, $\tilde{m}'_{xy}(\mathbf{r}, t_0) \approx 0$) and complete longitudinal recovery (*i.e.*, $m'_z(\mathbf{r}, t_0) \approx m_0(\mathbf{r})$), one recovers the conventional linear relation between magnetization and RF field for small excitations [175]. Without these assumptions, the longitudinal magnetization must also be considered. The solution to (B.6) expressed in terms of (B.7) for $t \geq t_0$ is

$$\begin{aligned} m'_z(\mathbf{r}, t) &= m'_z(\mathbf{r}, t_0) e^{-(t-t_0)/T_1(\mathbf{r})} + m_0(\mathbf{r}) \big(1 - e^{-(t-t_0)/T_1(\mathbf{r})}\big) \\ &\quad + \gamma \int_{t_0}^{t} e^{-(t-t')/T_1(\mathbf{r})} \big(\tilde{m}'_x(\mathbf{r}, t') b'_y(t') - \tilde{m}'_y(\mathbf{r}, t') b'_x(t')\big) \, \mathrm{d}\, t'. \end{aligned} \tag{B.8}$$

We next impose a steady-state condition to solve for the magnetization at an initial time t_0 selected well into the steady-state. After many periodic repetition cycles, the magnetization following one full repetition will equilibrate under certain mild assumptions [17] to the initial magnetization. We specifically assume that t_0 marks the beginning of a steady-state repetition interval of length $T_{\mathrm{R}} \geq T_{\mathrm{P}}$ during which RF excitation may be nonzero for $t \in [t_0, t_0 + T_{\mathrm{P}}]$ but $b'_{xy}(\mathbf{r}, t) = 0$ for free precession period $t \in (t_0 + T_{\mathrm{P}}, t_0 + T_{\mathrm{R}})$. Under these assumptions, the steady-state relations

$$\tilde{m}'_{xy}(\mathbf{r}, t_0 + T_{\mathrm{R}}) = \tilde{m}'_{xy}(\mathbf{r}, t_0) \tag{B.9}$$

$$m'_z(\mathbf{r}, t_0 + T_{\mathrm{R}}) = m'_z(\mathbf{r}, t_0) \tag{B.10}$$

provide an algebraic system of equations for the initial magnetization components. With some algebra and after reversion from intermediate variable $\tilde{m}'_{xy}(\mathbf{r}, t_0)$ to apparent transverse magnetization $m'_{xy}(\mathbf{r}, t_0)$, we find that steady-state initial magnetization is

$$m'_{xy}(\mathbf{r}, t_0) = \frac{i\gamma m_0(\mathbf{r}) \kappa^{\mathrm{t}}(\mathbf{r}) (1 - E_1(\mathbf{r}, T_{\mathrm{R}})) \int_{t_0}^{t_0 + T_{\mathrm{P}}} b_1(t') e^{i\phi(\mathbf{r}, t_0 + T_{\mathrm{R}}; t')} e^{-(t_0 + T_{\mathrm{R}} - t')/T_2(\mathbf{r})} \, \mathrm{d}\, t'}{(1 - E_2(\mathbf{r}, T_{\mathrm{R}}) e^{i\phi(\mathbf{r}, t_0 + T_{\mathrm{R}}; t_0)}) ((1 - E_1(\mathbf{r}, T_{\mathrm{R}})) + q(\mathbf{r}))}; \tag{B.11}$$

$$m'_z(\mathbf{r}, t_0) = \frac{m_0(\mathbf{r}) (1 - E_1(\mathbf{r}, T_{\mathrm{R}}))}{(1 - E_1(\mathbf{r}, T_{\mathrm{R}})) + q(\mathbf{r})}, \tag{B.12}$$

where $E_1(\mathbf{r}, t) := e^{-t/T_1(\mathbf{r})}$; $E_2(\mathbf{r}, t) := e^{-t/T_2(\mathbf{r})}$; and

$$\begin{aligned} q(\mathbf{r}) &= \gamma^2 \big|\kappa^{\mathrm{t}}(\mathbf{r})\big|^2 \int_{t_0}^{t_0+T_{\mathrm{P}}} \int_{t_0}^{t'} e^{-(T_{\mathrm{R}}+t_0-t')/T_1(\mathbf{r})} e^{-(t'-\tau)/T_2(\mathbf{r})} \operatorname{Re}\Big(b_1(t') b_{xy}'^{*}(\tau) e^{-i\phi(\mathbf{r},t';\tau)}\Big) \,\mathrm{d}\,\tau \,\mathrm{d}\,t' \\ &= \frac{\gamma^2 \big|\kappa^{\mathrm{t}}(\mathbf{r})\big|^2}{2} \int_{t_0}^{t_0+T_{\mathrm{P}}} \int_{t_0}^{t_0+T_{\mathrm{P}}} b_1(t') b_{xy}'^{*}(\tau) e^{-(T_{\mathrm{R}}+t_0-\max(\tau,t'))/T_1(\mathbf{r})} e^{-|t'-\tau|/T_2(\mathbf{r})} e^{-i\phi(\mathbf{r},t';\tau)} \,\mathrm{d}\,\tau \,\mathrm{d}\,t'. \end{aligned} \tag{B.13}$$

It is intuitive to design RF excitations that achieve a desired magnetization pattern at echo time $T_{\mathrm{E}} \geq T_{\mathrm{P}}$ following the start of excitation. We obtain the magnetization at echo time by substituting (B.11)-(B.12) into (B.7) and evaluating at time $t \leftarrow t_0 + T_{\mathrm{E}}$:

$$m_{xy}'(\mathbf{r}, t_0 + T_{\mathrm{E}}) = \frac{i\gamma m_0(\mathbf{r}) \kappa^{\mathrm{t}}(\mathbf{r})(1 - E_1(\mathbf{r}, T_{\mathrm{R}})) \int_{t_0}^{t_0+T_{\mathrm{P}}} b_1(t') e^{i\phi(\mathbf{r},t_0+T_{\mathrm{E}};t')} e^{-(t_0+T_{\mathrm{E}}-t')/T_2(\mathbf{r})} \,\mathrm{d}\,t'}{(1 - E_2(\mathbf{r}, T_{\mathrm{R}}) e^{i\phi(\mathbf{r},t_0+T_{\mathrm{R}};t_0)})((1 - E_1(\mathbf{r}, T_{\mathrm{R}})) + q(\mathbf{r}))}. \tag{B.14}$$

Comparing with (B.11), we reassuringly recover the SS condition as T_{E} approaches T_{R}:

$$\lim_{T_{\mathrm{E}} \to T_{\mathrm{R}}} m_{xy}'(\mathbf{r}, t_0 + T_{\mathrm{E}}) = m_{xy}'(\mathbf{r}, t_0 + T_{\mathrm{R}}) = m_{xy}'(\mathbf{r}, t_0). \tag{B.15}$$

The received signal is approximately proportional to the integrated transverse magnetization over a voxel volume $\mathbb{V}$. Further approximating the received signal as directly proportional to the steady-state magnetization (as is typical in RF pulse design) requires both standard and nonstandard additional assumptions:

- We first assume that the signal is localized to a scale over which there is minimal variation of $m_0(\mathbf{r})$, $T_1(\mathbf{r})$, and $T_2(\mathbf{r})$, a common assumption in RF pulse design.
- We next assume that $\phi(\mathbf{r}, t'; \tau)$ is a slowly-varying function with $\mathbf{r}$ over the voxel volume for all fixed time point pairs $(\tau, t') \in \{(\tau, t') | t_0 \leq \tau \leq t' \leq t_0 + T_{\mathrm{P}}\}$. In regions where through-voxel field-inhomogeneity gradients are reasonably small, this can be accomplished by restricting excitation gradient amplitudes from being too large and by balancing the excitation gradients to have zero net area.
- We last assume that $\phi(\mathbf{r}, t_0 + T_{\mathrm{E}}; t_0 + T_{\mathrm{P}})$ and $\phi(\mathbf{r}, t_0 + T_{\mathrm{R}}; t_0)$ also exhibit slow spatial variation. If excitation gradients are balanced, the former condition only further requires imaging gradients to balance at the echo time as is typical. However, the latter condition requires *all* gradients to be balanced, and thus restricts the analysis hereafter to unspoiled or weakly spoiled sequences.

With these additional assumptions, the received steady-state signal for a typically sized voxel centered at position $\mathbf{r}$ is (to within constants):

$$s(\mathbf{r}, t_0 + T_\mathrm{E}) \propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{E}) \, \mathrm{d}^3 \, \mathbf{r}' \tag{B.16}$$

$$\approx |\mathbb{V}(\mathbf{r})| m'_{xy}(\mathbf{r}, t_0 + T_\mathrm{E}), \tag{B.17}$$

where $|\cdot|$ here denotes the volume of a set. By including the contributions of previous (unspoiled) excitations, we have derived a *nonlinear* dependence of SS signal on small-excitation RF fields. Exploiting this nonlinear dependence will require specialized algorithms to efficiently design SS-informed RF pulses and/or excitation gradients.

Conventional small-excitation RF pulse design utilizes an assumption that off-resonance phase grows linearly in time to develop convenient Fourier-type relations between the excited magnetization and the excitation $\mathbf{k}$-space trajectory. Because the SS magnetization's dependence on off-resonance phase is nonlinear, a direct Fourier-type relation no longer applies. However, the integrals within SS magnetization (B.14) may *individually* be expressed as Fourier-type relations, so the possibility remains for signal model implementation using multiple Fourier transforms. With the affine off-resonance assumption $b'_z(\mathbf{r}, t) \approx \frac{\omega'(\mathbf{r})}{\gamma} + \mathbf{r} \cdot \mathbf{g}(t)$ the integral in the numerator of (B.14) describes the conventional Fourier-type relation [175], where $\omega'(\mathbf{r})$ denotes off-resonance frequency and $\mathbf{g}(t)$ denotes a linear excitation gradient trajectory. We next focus on developing a Fourier-type relation for the integral in the denominator of (B.14), contained within $q(\mathbf{r})$. Defining *k*-space trajectory $\mathbf{k}(t) := -\frac{\gamma}{2\pi} \int_{t_0}^{t} \mathbf{g}(t') \, \mathrm{d}\, t'$, phase accrual $\phi(\mathbf{r}, t'; \tau)$ can be approximated as

$$\phi(\mathbf{r}, t'; \tau) \approx -\omega'(\mathbf{r})(t' - \tau) + 2\pi \mathbf{r} \cdot (\mathbf{k}(t') - \mathbf{k}(\tau)). \tag{B.18}$$

Substituting (B.18) into (B.13) and simplifying yields

$$q(\mathbf{r}) \approx \frac{\gamma^2 |\kappa^\mathrm{t}(\mathbf{r})|^2}{2} \int_{t_0}^{t_0+T_\mathrm{P}} \int_{t_0}^{t_0+T_\mathrm{P}} b_1(t') b'^*_{xy}(\tau) \zeta(\mathbf{r}, t'; \tau) e^{-2\pi i \mathbf{r} \cdot (\mathbf{k}(t') - \mathbf{k}(\tau))} \, \mathrm{d}\, \tau \, \mathrm{d}\, t' \tag{B.19}$$

$$= \frac{\gamma^2 |\kappa^\mathrm{t}(\mathbf{r})|^2}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} p(\mathbf{r}, \mathbf{k_1}, \mathbf{k_2}) e^{2\pi i \mathbf{r} \cdot (\mathbf{k}_1 + \mathbf{k}_2)} \, \mathrm{d}\, \mathbf{k}_1 \, \mathrm{d}\, \mathbf{k}_2, \tag{B.20}$$

where $\zeta(\mathbf{r}, t', \tau)$ abbreviates relaxation and constant off-resonance effects as

$$\zeta(\mathbf{r}, t'; \tau) := e^{-(T_\mathrm{R} + t_0 - \max(\tau, t'))/T_1(\mathbf{r})} e^{-|t' - \tau|/T_2(\mathbf{r})} e^{+i\omega'(\mathbf{r})(t' - \tau)} \tag{B.21}$$

and $p(\mathbf{r}, \mathbf{k}_1, \mathbf{k}_2)$ defines a path in a six-dimensional k-space:

$$p(\mathbf{r}, \mathbf{k}_1, \mathbf{k}_2) := \int_{t_0}^{t_0+T_\mathrm{P}} \int_{t_0}^{t_0+T_\mathrm{P}} b_1(t') b'^{*}_{xy}(\tau) \zeta(\mathbf{r}, t'; \tau) \delta_3(\mathbf{k}(t') + \mathbf{k}_1) \delta_3(\mathbf{k}(\tau) - \mathbf{k}_2) \, \mathrm{d}\,\tau \, \mathrm{d}\,t', \tag{B.22}$$

where $\delta_n(\cdot)$ denotes an n-dimensional Dirac delta function. Eq. (B.20) is not a true Fourier relation because $p(\mathbf{r}, \mathbf{k}_1, \mathbf{k}_2)$ depends on $\mathbf{r}$. In the special case where relaxation and spatial variation in $\omega'(\mathbf{r})$ are neglected, $\zeta(\mathbf{r}, t'; \tau)$ separates in t', τ and (B.20) simplifies to a true Fourier transform for which fast methods (*e.g.*, [177]) to separately evaluate integrals over $\mathbf{k}_1, \mathbf{k}_2$ are available. In the general case, a similar Fourier relation could be constructed if $\zeta(\mathbf{r}, t'; \tau)$ were approximated with a "rank"-L tensor-product expansion of the form

$$\zeta(\mathbf{r}, t'; \tau) \approx \sum_{l=1}^{L} u_l(\mathbf{r}) v_l(t') \bar{v}_l(\tau), \tag{B.23}$$

where $u_l(\mathbf{r}) \in \mathbb{R}$ and $v_l(t) \in \mathbb{C}$ are the lth spatial and temporal basis functions. Inserting (B.23) into (B.19) would then yield a true Fourier relation similar to the one in [179]:

$$\begin{aligned} q(\mathbf{r}) &\approx \frac{\gamma^2 |\kappa^\mathrm{t}(\mathbf{r})|^2}{2} \sum_{l=1}^{L} u_l(\mathbf{r}) \left| \int_{t_0}^{t_0+T_\mathrm{P}} v_l(t') b_1(t') e^{-2\pi i \mathbf{r}\cdot\mathbf{k}(t')} \, \mathrm{d}\,t' \right|^2 \\ &= \frac{\gamma^2 |\kappa^\mathrm{t}(\mathbf{r})|^2}{2} \sum_{l=1}^{L} u_l(\mathbf{r}) \left| \mathcal{F}^{-1}_{3,\mathbf{k}} \left[\int_{t_0}^{t_0+T_\mathrm{P}} v_l(t') b_1(t') \delta_3(\mathbf{k}(t') + \mathbf{k}) \, \mathrm{d}\,t' \right] (\mathbf{r}) \right|^2, \end{aligned} \tag{B.24}$$

where $\mathcal{F}^{-1}_{3,\mathbf{k}}[\cdot](\mathbf{r})$ denotes an 3-dimensional inverse Fourier transform with respect to $\mathbf{k}$ and whose output is a function of $\mathbf{r}$. Unfortunately, (B.23) is a *partially-complex* tensor-product expansion, and would require a tailored decomposition algorithm. Alternately, we could expand the pure-real relaxation and complex off-resonance terms of $\zeta(\mathbf{r}, t'; \tau)$ separately:

$$\zeta(\mathbf{r}, t'; \tau) =: \xi(\mathbf{r}, t'; \tau) e^{+i\omega'(\mathbf{r})t'} e^{-i\omega'(\mathbf{r})\tau} \tag{B.25}$$

$$\approx \left(\sum_{l=1}^{L} u_l(\mathbf{r}) v_l(t') v_l(\tau) \right) \left(\sum_{k=1}^{K} u'_k(\mathbf{r}) v'_k(t') \right) \left(\sum_{k'=1}^{K} \bar{u}'_{k'}(\mathbf{r}) \bar{v}'_{k'}(\tau) \right), \tag{B.26}$$

where now $u_l(\mathbf{r}), v_l(t) \in \mathbb{R}$ and $u'_k(\mathbf{r}), v'_k(t) \in \mathbb{C}$. Inserting (B.26) into (B.19) would yield

$$\begin{aligned} q(\mathbf{r}) &\approx \frac{\gamma^2 |\kappa^{\mathrm{t}}(\mathbf{r})|^2}{2} \sum_{l=1}^{L} u_l(\mathbf{r}) \left| \sum_{k=1}^{K} u'_k(\mathbf{r}) \int_{t_0}^{t_0+T_{\mathrm{P}}} v_l(t') v'_k(t') b_1(t') e^{-2\pi i \mathbf{r}\cdot\mathbf{k}(t')} \, \mathrm{d}\, t' \right|^2 \\ &= \frac{\gamma^2 |\kappa^{\mathrm{t}}(\mathbf{r})|^2}{2} \sum_{l=1}^{L} u_l(\mathbf{r}) \left| \sum_{k=1}^{K} u'_k(\mathbf{r}) \mathcal{F}^{-1}_{3,\mathbf{k}} \left[\int_{t_0}^{t_0+T_{\mathrm{P}}} v_l(t') v'_k(t') b_1(t') \delta_3(\mathbf{k}(t') + \mathbf{k}) \, \mathrm{d}\, t' \right] (\mathbf{r}) \right|^2 . \end{aligned} \tag{B.27}$$

Though (B.25) simplifies the tensor product to work over $\mathbb{R}$ only, its tensor decomposition of $\xi(\mathbf{r}, t'; \tau)$ may still be challenging in practice. For ease of exposition in this early work, we revert to partially-complex expansion (B.23) and choose convenient (complex) temporal basis functions $v_l(t)$ for $l = 1, \ldots, L$, as in prior works [180, 181]. We then propose to estimate pure-real spatial basis functions $u_l(t)$ for $l = 1, \ldots, L$ via linear least-squares.

We next discretize SS signal model (B.14) for computer-aided SS-informed RF pulse design. Let $\mathbf{b} := [b_1(T_1), \ldots, b_1(t_J)]^{\mathsf{T}} \in \mathbb{C}^J$ discretize the excitation waveform into J timepoints from $t \leftarrow t_0$ to $t \leftarrow t_0 + T_{\mathrm{P}}$. Let

$$\mathbf{m} := \left[\frac{s(\mathbf{r}_1, t_0 + T_{\mathrm{E}})}{m_0(\mathbf{r}_1)|\mathbb{V}(\mathbf{r}_1)|}, \ldots, \frac{s(\mathbf{r}_N, t_0 + T_{\mathrm{E}})}{m_0(\mathbf{r}_N)|\mathbb{V}(\mathbf{r}_1)|} \right]^{\mathsf{T}} \in \mathbb{C}^N \tag{B.28}$$

discretize the (unitless) relative magnetization into N voxels, respectively. Let

$$a_{nj} := \frac{i(\gamma \kappa^{\mathrm{t}}(\mathbf{r}_n)\Delta t)(1 - E_1(\mathbf{r}_n)) e^{i\phi(\mathbf{r}_n, t_0 + T_{\mathrm{E}}; t_j)} e^{-(t_0 + T_{\mathrm{E}} - t_j)/T_2(\mathbf{r}_n)}}{(1 - E_2(\mathbf{r}_n) e^{i\phi(t_0 + T_{\mathrm{R}}; t_0)})}; \tag{B.29}$$

$$d_{nn'} := \delta[n - n'](1 - E_1(\mathbf{r}_n)); \tag{B.30}$$

$$\begin{aligned} s_{jj'n} &:= \frac{|\gamma \kappa^{\mathrm{t}}(\mathbf{r}_n)\Delta t|^2}{2} \zeta(\mathbf{r}_n, t_j, t_{j'}) e^{-2\pi i \mathbf{r}_n \cdot (\mathbf{k}(t_{j'}) - \mathbf{k}(t_j))} \\ &\approx \frac{|\gamma \kappa^{\mathrm{t}}(\mathbf{r}_n)\Delta t|^2}{2} \sum_{l=1}^{L} u_l(\mathbf{r}_n) v_l(t_{j'}) v_l^*(t_j) e^{-2\pi i \mathbf{r}_n \cdot (\mathbf{k}(t_{j'}) - \mathbf{k}(t_j))}, \end{aligned} \tag{B.31}$$

respectively denote scalar elements of the matrices $\mathbf{A} \in \mathbb{C}^{N \times J}$; $\mathbf{D} \in \mathbb{R}^{N \times N}$; and tensor $\mathcal{S} \in \mathbb{C}^{J \times J \times N}$. Here $\Delta t := T_{\mathrm{P}}/J$ denotes the RF pulse sampling interval and $\delta[\cdot]$ denotes the Kronecker delta function. Then the noiseless discretized signal model reads

$$\mathbf{m} = \left(\mathbf{D} + \operatorname{diag}\left(\left[\mathbf{b}^{\mathsf{H}} \mathbf{S}_1 \mathbf{b}, \ldots, \mathbf{b}^{\mathsf{H}} \mathbf{S}_N \mathbf{b} \right]^{\mathsf{T}} \right) \right)^{-1} \mathbf{A} \mathbf{b}, \tag{B.32}$$

where $\mathbf{S}_n \in \mathbb{C}^{J \times J}$ denotes the nth layer of $\mathcal{S}$ for each $n \in \{1, \ldots, N\}$.

B.3 Two Iterative Algorithms

This section develops two iterative algorithms for designing SS-informed RF pulses. Subsection B.3.1 employs a perturbative expansion approach that takes inspiration from a similar algorithm for large-tip angle RF pulse design [182]. Subsection B.3.2 employs an optimization approach that uses variable splitting and the alternating direction method of multipliers (ADMM) algorithm [183]. Both subsections focus on designing the RF pulse only and assume the excitation gradient trajectory is fixed.

B.3.1 Perturbative Expansions

Given an RF pulse design at the ith iteration $\mathbf{b}^{(i)}$, one can predict the magnetization $\mathbf{m}^{(i)}$ via either (B.32) or Bloch simulations. If the deviation $\Delta\mathbf{m}^{(i+1)} := \mathbf{m} - \mathbf{m}^{(i)}$ from the desired pattern is small, it should be possible to reduce the subsequent deviation with small perturbation $\Delta\mathbf{b}^{(i+1)}$ to $\mathbf{b}^{(i)}$. We can then iteratively improve the RF pulse by accumulating small perturbations, *i.e.* $\mathbf{b}^{(i+1)} \leftarrow \mathbf{b}^{(i)} + \Delta\mathbf{b}^{(i+1)} \equiv \mathbf{b}^{(0)} + \sum_{\iota=1}^{i+1} \Delta\mathbf{b}^{(\iota)}$.

Here we apply a first-order perturbative expansion to approximate how the $(i+1)$th incremental excitation $\Delta\mathbf{b}^{(i+1)}$ relates to the $(i+1)$th pattern deviation $\Delta\mathbf{m}^{(i+1)}$:

$$\mathbf{A}\big(\mathbf{b}^{(i)} + \Delta\mathbf{b}^{(i+1)}\big) = \Big(\mathbf{D} + \operatorname{diag}_n\Big(\big(\mathbf{b}^{(i)} + \Delta\mathbf{b}^{(i+1)}\big)^{\mathsf{H}} \mathbf{S}_n \big(\mathbf{b}^{(i)} + \Delta\mathbf{b}^{(i+1)}\big)\Big)\Big)\big(\mathbf{m}^{(i)} + \Delta\mathbf{m}^{(i+1)}\big) \tag{B.33}$$

$$\begin{aligned} &\approx \big(\mathbf{D} + \operatorname{diag}_n\big((\mathbf{b}^{(i)})^{\mathsf{H}}\mathbf{S}_n\mathbf{b}^{(i)}\big)\big)\big(\mathbf{m}^{(i)} + \Delta\mathbf{m}^{(i+1)}\big) \\ &+ 2\operatorname{diag}_n\big(\operatorname{Re}\big((\mathbf{b}^{(i)})^{\mathsf{H}}\mathbf{S}_n\Delta\mathbf{b}^{(i+1)}\big)\big)\mathbf{m}^{(i)}, \end{aligned} \tag{B.34}$$

where $\operatorname{diag}_n\big(\mathbf{b}^{\mathsf{H}}\mathbf{S}_n\mathbf{b}'\big) := \operatorname{diag}\Big(\big[\mathbf{b}^{\mathsf{H}}\mathbf{S}_1\mathbf{b}', \ldots, \mathbf{b}^{\mathsf{H}}\mathbf{S}_n\mathbf{b}'\big]^{\mathsf{T}}\Big)$ and the approximation drops all higher-order terms. Canceling out zeroth-order terms

$$\mathbf{A}\mathbf{b}^{(i)} \approx \big(\mathbf{D} + \operatorname{diag}_n\big((\mathbf{b}^{(i)})^{\mathsf{H}}\mathbf{S}_n\mathbf{b}^{(i)}\big)\big)\mathbf{m}^{(i)} \tag{B.35}$$

that either are exactly equal (if $\mathbf{m}^{(i)}$ is computed via (B.32)) or nearly equal (if $\mathbf{m}^{(i)}$ is computed via Bloch simulations) and rearranging yields the update

$$\begin{aligned} \operatorname{Re}\big(\Delta\mathbf{b}^{(i+1)}\big) &\leftarrow \Big(\mathbf{A} - 2\operatorname{diag}\big(\mathbf{m}^{(i)}\big)\operatorname{Re}\Big(\boldsymbol{\Gamma}^{(i)}\Big)\Big)^{\dagger}\big(\mathbf{D} + \operatorname{diag}_n\big((\mathbf{b}^{(i)})^{\mathsf{H}}\mathbf{S}_n\mathbf{b}^{(i)}\big)\big)\Delta\mathbf{m}^{(i+1)}; \\ \operatorname{Im}\big(\Delta\mathbf{b}^{(i+1)}\big) &\leftarrow \Big(i\mathbf{A} + 2\operatorname{diag}\big(\mathbf{m}^{(i)}\big)\operatorname{Im}\Big(\boldsymbol{\Gamma}^{(i)}\Big)\Big)^{\dagger}\big(\mathbf{D} + \operatorname{diag}_n\big((\mathbf{b}^{(i)})^{\mathsf{H}}\mathbf{S}_n\mathbf{b}^{(i)}\big)\big)\Delta\mathbf{m}^{(i+1)}, \end{aligned} \tag{B.36}$$

where $\boldsymbol{\Gamma}^{(i)} := \left[\mathbf{S}_1\mathbf{b}^{(i)}, \dots, \mathbf{S}_n\mathbf{b}^{(i)}\right]^{\mathsf{H}} \in \mathbb{C}^{N \times J}$. Each perturbative correction update requires solving two size-J linear least-squares problems.

B.3.2 Variable Splitting and ADMM

Alternately, we can approach SS-informed RF pulse design by solving a suitable optimization problem. We would like to solve the non-convex optimization problem

$$\widehat{\mathbf{b}} \in \left\{ \arg \min_{\mathbf{b} \in \mathbb{C}^J} \left\| \left(\mathbf{D} + \operatorname{diag}_n\left(\mathbf{b}^{\mathsf{H}}\mathbf{S}_n\mathbf{b}\right)\right)^{-1} \mathbf{A}\mathbf{b} - \mathbf{m} \right\|_{\mathbf{W}}^2 + \beta \|\mathbf{C}\mathbf{b}\|_2^2 \right\} \tag{B.37}$$

where $\mathbf{W} \in \mathbb{R}^{N \times N}$ denotes a spatial weighting matrix; $\mathbf{C}$ is a pure-real first-order finite differencing matrix; and $\beta \in \mathbb{R}$ is a regularization parameter. Since solving (B.37) is challenging, we instead study a related constrained problem:

$$\begin{aligned}
\left(\widehat{\mathbf{b}}, \widehat{\mathbf{z}}_1, \widehat{\mathbf{z}}_2\right) \; &\in \left\{ \arg \min_{\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{C}^J} \Psi(\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2) \right\} && \text{subject to} && \text{(B.38)} \\
\mathbf{z}_1 \; &= \mathbf{f}(\mathbf{b}, \mathbf{z}_2) + \mathbf{d} && \text{and} && \\
\mathbf{z}_2 \; &= \mathbf{b}, && \text{where} && \\
\Psi(\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2) &:= \|\mathbf{A}\mathbf{b} - \operatorname{diag}(\mathbf{m})\mathbf{z}_1\|_{\mathbf{W}}^2 + \beta\|\mathbf{C}\mathbf{b}\|_2^2; &&&& \text{(B.39)}
\end{aligned}$$

$\mathbf{f}(\mathbf{b}, \mathbf{b}') := \left[\mathbf{b}^{\mathsf{H}}\mathbf{S}_1\mathbf{b}', \dots, \mathbf{b}^{\mathsf{H}}\mathbf{S}_N\mathbf{b}'\right]$ evaluates $\mathbf{f} : \mathbb{C}^J \times \mathbb{C}^J \mapsto \mathbb{C}^N$ for arbitrary $\mathbf{b}, \mathbf{b}'$; and $\mathbf{d} := \mathbf{D}\mathbf{1}_N \in \mathbb{R}^N$. We solve (B.38) via ADMM [183]. The augmented Lagrangian is

$$\begin{aligned}
\Lambda(\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\nu}_1, \boldsymbol{\nu}_2) := \Psi(\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2) + \operatorname{Re}\left(\boldsymbol{\nu}_1^{\mathsf{H}}(\mathbf{f}(\mathbf{b}, \mathbf{z}_2) + \mathbf{d} - \mathbf{z}_1)\right) + \frac{\rho_1}{2}\|\mathbf{f}(\mathbf{b}, \mathbf{z}_2) + \mathbf{d} - \mathbf{z}_1\|_2^2 \\
+ \operatorname{Re}\left(\boldsymbol{\nu}_2^{\mathsf{H}}(\mathbf{b} - \mathbf{z}_2)\right) + \frac{\rho_2}{2}\|\mathbf{b} - \mathbf{z}_2\|_2^2,
\end{aligned} \tag{B.40}$$

where $\boldsymbol{\nu}_1 \in \mathbb{C}^N$ and $\boldsymbol{\nu}_2 \in \mathbb{C}^J$ are dual variables and $\rho_1, \rho_2 > 0$ are constraint penalty parameters. Rescaling the dual variables as $\mathbf{u}_1 := \frac{\boldsymbol{\nu}_1}{\rho_1}$ and $\mathbf{u}_2 := \frac{\boldsymbol{\nu}_2}{\rho_2}$, completing the square, and dropping constants reveals an alternate but equivalent form of the augmented Lagrangian that leads to simpler variable updates:

$$\begin{aligned}
\Lambda'(\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2, \mathbf{u}_1, \mathbf{u}_2) := \Psi(\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2) + \frac{\rho_1}{2}\|\mathbf{f}(\mathbf{b}, \mathbf{z}_2) + \mathbf{d} - \mathbf{z}_1 + \mathbf{u}_1\|_2^2 - \frac{\rho_1}{2}\|\mathbf{u}_1\|_2^2 \\
+ \frac{\rho_2}{2}\|\mathbf{b} - \mathbf{z}_2 + \mathbf{u}_2\|_2^2 - \frac{\rho_2}{2}\|\mathbf{u}_2\|_2^2.
\end{aligned} \tag{B.41}$$

ADMM cycles through updating the primal variables $\mathbf{b}, \mathbf{z}_1, \mathbf{z}_2$ followed by gradient ascent on the scaled dual variables $\mathbf{u}_1, \mathbf{u}_2$, holding other variables fixed from previous iterations.

Though ρ_1 and ρ_2 influence (local) convergence rates, they do not affect the solution. We next describe each of these updates.

1. The $\mathbf{b}$ update involves a size-J linear least-squares subproblem:

$$\begin{aligned}\mathbf{b}^{(i+1)} &\leftarrow \arg\min_{\mathbf{b}\in\mathbb{C}^J} \Lambda'(\mathbf{b},\mathbf{z}_1^{(i)},\mathbf{z}_2^{(i)},\mathbf{u}_1^{(i)},\mathbf{u}_2^{(i)}) \\ &= \arg\min_{\mathbf{b}\in\mathbb{C}^J} \Psi\Big(\mathbf{b},\mathbf{z}_1^{(i)},\mathbf{z}_2^{(i)}\Big) \qquad + \frac{\rho_1}{2}\left\|\mathbf{f}(\mathbf{b},\mathbf{z}_2^{(i)}) - \Big(\mathbf{z}_1^{(i)} - \mathbf{d} - \mathbf{u}_1^{(i)}\Big)_2^2\right\| \\ &\qquad + \frac{\rho_2}{2}\left\|\mathbf{b} - \Big(\mathbf{z}_2^{(i)} - \mathbf{u}_2^{(i)}\Big)\right\|_2^2. \end{aligned} \tag{B.42}$$

2. The $\mathbf{z}_1$ update involves a trivial linear least-squares subproblem:

$$\begin{aligned}\mathbf{z}_1^{(i+1)} &\leftarrow \arg\min_{\mathbf{z}_1\in\mathbb{C}^N} \Lambda'(\mathbf{b}^{(i+1)},\mathbf{z}_1,\mathbf{z}_2^{(i)},\mathbf{u}_1^{(i)},\mathbf{u}_2^{(i)}) \\ &= \arg\min_{\mathbf{z}_1\in\mathbb{C}^N} \left\|\mathrm{diag}(\mathbf{m})\mathbf{z}_1 - \mathbf{A}\mathbf{b}^{(i+1)}\right\|_{\mathbf{W}}^2 + \frac{\rho_1}{2}\left\|\mathbf{z}_1 - \Big(\mathbf{f}(\mathbf{b}^{(i+1)},\mathbf{z}_2^{(i)}) + \mathbf{d} + \mathbf{u}_1^{(i)}\Big)\right\|_2^2. \end{aligned} \tag{B.43}$$

3. The $\mathbf{z}_2$ update involves another size-J linear least-squares subproblem:

$$\begin{aligned}\mathbf{z}_2^{(i+1)} &\leftarrow \arg\min_{\mathbf{z}_2\in\mathbb{C}^J} \Lambda'(\mathbf{b}^{(i+1)},\mathbf{z}_1^{(i+1)},\mathbf{z}_2,\mathbf{u}_1^{(i)},\mathbf{u}_2^{(i)}) \\ &= \arg\min_{\mathbf{z}_2\in\mathbb{C}^J} \frac{\rho_1}{2}\left\|\mathbf{f}(\mathbf{b}^{(i+1)},\mathbf{z}_2) - \Big(\mathbf{z}_1^{(i+1)} - \mathbf{d} - \mathbf{u}_1^{(i)}\Big)\right\|_2^2 + \frac{\rho_2}{2}\left\|\mathbf{z}_2 - \Big(\mathbf{b}^{(i+1)} + \mathbf{u}_2^{(i)}\Big)\right\|_2^2. \end{aligned} \tag{B.44}$$

4. The scaled dual variables $\mathbf{u}_1, \mathbf{u}_2$ are updated via one gradient ascent iteration:

$$\mathbf{u}_1^{(i+1)} \leftarrow \mathbf{u}_1^{(i)} + \rho_1\Big(\mathbf{f}(\mathbf{b}^{(i+1)},\mathbf{z}_2^{(i+1)}) + \mathbf{d} - \mathbf{z}_1^{(i+1)}\Big); \tag{B.45}$$

$$\mathbf{u}_2^{(i+1)} \leftarrow \mathbf{u}_2^{(i)} + \rho_2\Big(\mathbf{b}^{(i+1)} - \mathbf{z}_2^{(i+1)}\Big). \tag{B.46}$$

If a saddle point local to the initialization exists, ADMM will converge to that saddle point and the solution will correspond to a local solution to constrained problem (B.38).

BIBLIOGRAPHY

[1] G. Nataraj, M. Gao, J. Assländer, C. Scott, and J. A. Fessler, "Shallow learning with kernels for dictionary-free magnetic resonance fingerprinting," in *ISMRM Workshop on MR Fingerprinting*, 2017. iii, 3, 56, 85, 86

[2] K. E. Keenan, K. F. Stupic, M. A. Boss, S. E. Russek, T. L. Chenevert, P. V. Prasad, W. E. Reddick, K. M. Cecil, J. Zheng, P. Hu, and E. F. Jackson, "Multi-site, multi-vendor comparison of T1 measurement using ISMRM/NIST system phantom," in *Proc. Intl. Soc. Mag. Res. Med.*, p. 3290, 2016. ix, x, xi, xii, 42, 44, 47, 48, 69, 76, 78, 83

[3] S. Ahn and J. A. Fessler, "Standard errors of mean, variance, and standard deviation estimators," Tech. Rep. 413, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, July 2003. x, xi, xii, xiii, xiv, 25, 27, 45, 74, 78, 80, 83, 113, 114, 118

[4] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "Regularized, joint estimation of T1 and M0 maps," in *Proc. Intl. Soc. Mag. Res. Med.*, p. 3128, 2014. 2, 18

[5] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "Model-based estimation of T2 maps with dual-echo steady-state MR imaging," in *Proc. IEEE Intl. Conf. on Image Processing*, pp. 1877–81, 2014. 2, 18, 31, 35, 54, 55

[6] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "Optimizing MR scan design for model-based T1, T2 estimation from steady-state sequences," *IEEE Trans. Med. Imag.*, vol. 36, pp. 467–77, Feb. 2017. 2, 30, 56, 66, 69, 70, 72, 75, 76, 77, 80, 82, 87, 90

[7] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "A min-max CRLB optimization approach to scan selection for relaxometry," in *Proc. Intl. Soc. Mag. Res. Med.*, p. 1672, 2015. 2, 30

[8] G. Nataraj, J.-F. Nielsen, C. Scott, and J. A. Fessler, "Dictionary-free MRI PERK: Parameter estimation via regression with kernels," *IEEE Trans. Med. Imag.*, 2018. To appear. 3, 56, 119

[9] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "Dictionary-free MRI parameter estimation via kernel ridge regression," in *Proc. IEEE Intl. Symp. Biomed. Imag.*, pp. 5–9, 2017. 3, 56, 83, 88

[10] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "Fast myelin water imaging via Bayesian experimental design and PERK," 2018. In preparation. 3, 89

[11] G. Nataraj, J.-F. Nielsen, and J. A. Fessler, "Myelin water fraction estimation from optimized steady-state sequences using kernel ridge regression," in *Proc. Intl. Soc. Mag. Res. Med.*, p. 5076, 2017. 3, 57, 88, 89

[12] A. Macovski, *Medical imaging systems*. New Jersey: Prentice-Hall, 1983. 4

[13] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, *Magnetic resonance imaging: Physical principles and sequence design*. New York: Wiley, 1999. 4

[14] D. G. Nishimura, "Principles of magnetic resonance imaging," 1996. Unpublished textbook. 4

[15] F. Bloch, "Nuclear induction," *Phys. Rev.*, vol. 70, pp. 460–74, Oct. 1946. 4, 91

[16] W. S. Hinshaw, "Image formation by nuclear magnetic resonance: The sensitive point method," *J. Appl. Phys.*, vol. 47, p. 3709, Aug. 1976. 8, 30

[17] K. Scheffler, "A pictorial description of steady-states in rapid magnetic resonance imaging," *Concepts in Magnetic Resonance*, vol. 11, no. 5, pp. 291–304, 1999. 8, 30, 130

[18] B. A. Hargreaves, S. S. Vasanawala, J. M. Pauly, and D. G. Nishimura, "Characterization and reduction of the transient response in steady-state MR imaging," *Mag. Res. Med.*, vol. 46, pp. 149–58, July 2001. 8

[19] Y. Zur, M. L. Wood, and L. J. Neuringer, "Spoiling of transverse magnetization in steady-state sequences," *Mag. Res. Med.*, vol. 21, pp. 251–63, Oct. 1991. 8, 9, 69, 90, 107, 116

[20] V. Denolin, C. Azizieh, and T. Metens, "New insights into the mechanisms of signal formation in RF-spoiled gradient echo sequences," *Mag. Res. Med.*, vol. 54, pp. 937–54, Oct. 2005. 9

[21] T. W. Redpath and R. A. Jones, "FADE-A new fast imaging sequence," *Mag. Res. Med.*, vol. 6, pp. 224–34, Feb. 1988. 11, 90

[22] H. Bruder, H. Fischer, R. Graumann, and M. Deimling, "A new steady-state imaging sequence for simultaneous acquisition of two MR images with clearly different contrasts," *Mag. Res. Med.*, vol. 7, pp. 35–42, May 1988. 11, 25, 31, 69, 72, 90

[23] J. B. Rosen, "The gradient projection method for nonlinear programming, Part I: Linear constraints," *SIAM J. Appl. Math.*, vol. 8, no. 1, pp. 181–217, 1960. 16, 72

[24] C. Byrne, "A unified treatment of some iterative algorithms in signal processing and image reconstruction," *Inverse Prob.*, vol. 20, pp. 103–20, Feb. 2004. 16

[25] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Prob.*, vol. 19, pp. R1–26, Apr. 2003. 16, 72, 111, 117

[26] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Sig. Proc.*, vol. 51, pp. 560–74, Feb. 2003. 19

[27] M. J. Muckley, D. C. Noll, and J. A. Fessler, "Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA)," *IEEE Trans. Med. Imag.*, vol. 34, pp. 578–88, Feb. 2015. 19

[28] A. Macovski, "Noise in MRI," *Mag. Res. Med.*, vol. 36, pp. 494–7, Sept. 1996. 19

[29] T. Lei, "Statistics of MR signals: revisited," in *Proc. SPIE 6510 Medical Imaging: Phys. Med. Im.*, p. 651052, 2007. 19

[30] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Mag. Res. Med.*, vol. 34, pp. 910–4, Dec. 1995. 20, 33, 62, 66, 106

[31] J. P. Wansapura, S. K. Holland, R. S. Dunn, and W. S. Ball, "NMR relaxation times in the human brain at 3.0 Tesla," *J. Mag. Res.*, vol. 9, pp. 531–8, Apr. 1999. 23, 36, 39, 73, 80, 111

[32] G. J. Stanisz, E. E. Odrobina, J. Pun, M. Escaravage, S. J. Graham, M. J. Bronskill, and R. M. Henkelman, "T_1, T_2 relaxation and magnetization transfer in tissue at 3T," *Mag. Res. Med.*, vol. 54, pp. 507–12, Sept. 2005. 23, 36, 39, 80

[33] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, pp. 1085–97, Nov. 1999. 23, 39, 52

[34] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. Med. Imag.*, vol. 17, pp. 463–8, June 1998. 23, 39, 52, 73, 111

[35] R. K. Gupta, "A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation times using Fourier transform NMR," *J. Mag. Res.*, vol. 25, pp. 231–5, Jan. 1977. 23, 72

[36] S. C. L. Deoni, B. K. Rutt, and T. M. Peters, "Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state," *Mag. Res. Med.*, vol. 49, pp. 515–26, Mar. 2003. 23, 28, 31, 35

[37] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Cont. Opt.*, vol. 20, no. 2, pp. 221–46, 1982. 24, 72

[38] J. A. Fessler, "Michigan image reconstruction toolbox (MIRT) for Matlab," 2016. Available from web.eecs.umich.edu/ fessler. 24

[39] R. Heule, C. Ganter, and O. Bieri, "Rapid estimation of cartilage T2 with reduced T1 sensitivity using double echo steady state imaging," *Mag. Res. Med.*, vol. 71, pp. 1137–43, Mar. 2014. 28, 31

[40] H.-L. M. Cheng, N. Stikov, N. R. Ghugre, and G. A. Wright, "Practical medical applications of quantitative MR relaxometry," *J. Mag. Res. Im.*, vol. 36, pp. 805–24, Oct. 2012. 30

[41] H. B. W. Larsson, J. Frederiksen, L. Kjaer, O. Henriksen, and J. Olesen, "In vivo determination of T1 and T2 in the brain of patients with severe but stable multiple sclerosis," *Mag. Res. Med.*, vol. 7, pp. 43–55, May 1988. 30

[42] T. Kurki, N. Lundbom, M. Komu, and M. Kormano, "Tissue characterization of inter cranial tumors by magnetization transfer and spin-lattice relaxation parameters in vivo," *J. Mag. Res. Im.*, vol. 6, pp. 573–9, Aug. 1996. 30

[43] E. Englund, A. Brun, Z. Gyorffy-Wagner, E. Larsson, and B. Persson, "Relaxation times in relation to grade of malignancy and tissue necrosis in astrocytic gliomas," *Mag. Res. Im.*, vol. 4, no. 5, pp. 425–9, 1986. 30

[44] S. Siemonsen, K. Mouridsen, B. Holst, T. Ries, J. Finsterbusch, G. Thomalia, L. Ostergaard, and J. Fiehler, "Quantitative T2 values predict time from symptom onset in acute stroke patients," *Stroke*, vol. 40, pp. 1612–6, May 2009. 30

[45] L. D. DeWitt, J. P. Kistler, D. C. Miller, E. P. Richardson, and F. S. Buonanno, "NMR-neuropathologic correlation in stroke," *Stroke*, vol. 18, no. 2, pp. 342–51, 1987. 30

[46] S. J. Matzat, J. V. Tiel, G. E. Gold, and E. H. G. Oei, "Quantitative MRI techniques of cartilage composition," *Quant. Imaging Med. Surg.*, vol. 3, pp. 162–74, June 2013. 30

[47] T. J. Mosher and B. J. Dardzinski, "Cartilage MRI T2 relaxation time mapping: overview and applications," *Semin. Musculoskelet. Radiol.*, vol. 8, no. 4, pp. 355–68, 2004. 30

[48] H. Guo, W.-Y. Au, J. S. Cheung, D. Kim, J. H. Jensen, P.-L. Khong, Q. Chan, K. C. Chan, C. Tosti, H. Tang, T. R. Brown, W. W. M. Lam, S.-Y. Ha, G. M. Brittenham, and E. X. Wu, "Myocardial T2 quantification in patients with iron overload at 3 Tesla," *J. Mag. Res. Im.*, vol. 30, pp. 394–400, Aug. 2009. 30

[49] S. Giri, Y. C. Chung, A. Merchant, G. Mihai, S. Rajagopalan, S. V. Raman, and O. P. Simonetti, "T2 quantification for improved detection of myocardial edema," *Cardiovasc. Magn. Reson.*, vol. 11, no. 1, pp. 56–68, 2009. 30

[50] D. C. Look and D. R. Locker, "Time saving in measurement of NMR and EPR relaxation times," *Rev Sci Instrum*, vol. 41, pp. 250–1, Feb. 1970. 30

[51] H. Y. Carr and E. M. Purcell, "Effects of diffusion on free precession in nuclear magnetic resonance experiments," *Phys. Rev.*, vol. 94, pp. 630–8, May 1954. 30, 90

[52] M. K. Stehling, R. Turner, and P. Mansfield, "Echo-planar imaging: magnetic resonance imaging in a fraction of a second," *Science*, vol. 254, pp. 43–50, Oct. 1991. 30

[53] C. B. Ahn, J. H. Kim, and Z. H. Cho, "High-speed spiral-scan scho planar NMR imaging - I," *IEEE Trans. Med. Imag.*, vol. 5, pp. 2–7, Mar. 1986. 30

[54] C. H. Meyer, B. S. Hu, D. G. Nishimura, and A. Macovski, "Fast spiral coronary artery imaging," *Mag. Res. Med.*, vol. 28, pp. 202–13, Dec. 1992. 30

[55] I. Kay and R. M. Henkelman, "Practical Implementation and Optimization of One-shot T1 imaging," *Mag. Res. Med.*, vol. 22, pp. 414–24, Dec. 1991. 30

[56] P. A. Gowland and M. O. Leach, "Fast and accurate measurements of T1 using a multi-readout single inversion-recovery sequence," *Mag. Res. Med.*, vol. 26, pp. 79–88, July 1992. 30

[57] D. R. Messroghli, A. Radjenovic, S. Kozerke, D. M. Higgins, M. U. Sivananthan, and J. P. Ridgway, "Modified Look-Locker inversion recovery (MOLLI) for high-resolution T_1 mapping of the heart," *Mag. Res. Med.*, vol. 52, pp. 141–6, July 2004. 30

[58] M. K. Stehling, R. J. Ordidge, R. Coxon, and P. Mansfield, "Inversion-recovery Echo-planar imaging (IR-EPI) at 0.5T," *Mag. Res. Med.*, vol. 13, pp. 514–7, Mar. 1990. 30

[59] J.-M. Bonny, M. Zanca, J.-Y. Boire, and A. Veyre, "T2 maximum likelihood estimation from multiple spin-echo magnitude images," *Mag. Res. Med.*, vol. 36, pp. 287–93, Aug. 1996. 30

[60] D. Kumar, T. D. Nguyen, S. A. Gauthier, and A. Raj, "Bayesian algorithm using spatial priors for multiexponential T2 relaxometry from multiecho spin echo MRI," *Mag. Res. Med.*, vol. 68, pp. 1536–43, Nov. 2012. 30

[61] N. Ben-Eliezer, D. K. Sodickson, and K. T. Block, "Rapid and accurate T2 mapping from multi–spin-echo data using Bloch-simulation-based reconstruction," *Mag. Res. Med.*, vol. 73, pp. 809–17, Feb. 2015. 30, 55, 56

[62] T. D. Nguyen, C. Wisnieff, M. A. Cooper, D. Kumar, A. Raj, P. Spincemaille, Y. Wang, T. Vartanian, and S. A. Gauthier, "T2prep three-dimensional spiral imaging with efficient whole brain coverage for myelin water quantification at 1.5 tesla," *Mag. Res. Med.*, vol. 67, pp. 614–21, Mar. 2012. 30

[63] S. Majumdar, S. C. Orphanoudakis, A. Gmitro, M. O'Donnell, and J. C. Gore, "Error in the measurements of T2 using multiple-echo MRI techniques: 1. Effect of radiofrequency pulse imperfections," *Mag. Res. Med.*, vol. 3, pp. 397–417, June 1986. 30

[64] S. Majumdar, S. C. Orphanoudakis, A. Gmitro, M. O'Donnell, and J. C. Gore, "Error in the measurements of T2 using multiple-echo MRI techniques: 2. Effects of static field inhomogeneity," *Mag. Res. Med.*, vol. 3, pp. 562–74, Aug. 1986. 30

[65] F. Farzaneh, S. J. Riederer, and N. J. Pelc, "Analysis of T2 limitations and off-resonance effects on spatial resolution and artifacts in echo-planar imaging," *Mag. Res. Med.*, vol. 14, pp. 123–39, Apr. 1990. 30

[66] L.-C. Chang, C. G. Koay, P. J. Basser, and C. Pierpaoli, "Linear least-squares method for unbiased estimation of T1 from SPGR signals," *Mag. Res. Med.*, vol. 60, pp. 496–501, Aug. 2008. 31, 35

[67] E. K. Fram, R. J. Herfkens, G. A. Johnson, G. H. Glover, J. P. Kaaris, A. Shimakawa, T. G. Perkins, and N. J. Pelc, "Rapid calculation of T1 using variable flip angle gradient refocused imaging," *Mag. Res. Im.*, vol. 5, no. 3, pp. 201–8, 1987. 31, 35

[68] H. Wang and Y. Cao, "Spatially regularized T1 estimation from variable flip angles MRI," *Med. Phys.*, vol. 39, pp. 4139–48, July 2012. 31, 35

[69] S. C. L. Deoni, H. A. Ward, T. M. Peters, and B. K. Rutt, "Rapid T_2 estimation with phase-cycled variable nutation steady-state free precession," *Mag. Res. Med.*, vol. 52, pp. 435–9, Aug. 2004. 31

[70] S. C. L. Deoni, "Transverse relaxation time (T2) mapping in the brain with off-resonance correction using phase-cycled steady-state free precession imaging," *J. Mag. Res. Im.*, vol. 30, pp. 411–7, Aug. 2009. 31

[71] G. H. Welsch, K. Scheffler, T. C. Mamisch, T. Hughes, S. Millington, M. Deimling, and S. Trattnig, "Rapid estimation of cartilage T2 based on double echo at steady state (DESS) with 3 Tesla," *Mag. Res. Med.*, vol. 62, pp. 544–9, Aug. 2009. 31, 35, 37

[72] T. Stöcker, F. Keil, K. Vahedipour, D. Brenner, E. Pracht, and N. J. Shah, "MR parameter quantification with magnetization-prepared double echo steady-state (MP-DESS)," *Mag. Res. Med.*, vol. 72, pp. 103–11, July 2014. 31

[73] R. Heule, C. Ganter, and O. Bieri, "Triple echo steady-state (TESS) relaxometry," *Mag. Res. Med.*, vol. 71, pp. 230–7, Jan. 2014. 31

[74] M. L. Gyngell, "The steady-state signals in short-repetition-time sequences," *J. Mag. Res.*, vol. 81, pp. 474–83, Feb. 1989. 31

[75] W. Hänicke and H. U. Vogel, "An analytical solution for the SSFP signal in MRI," *Mag. Res. Med.*, vol. 49, pp. 771–5, Apr. 2003. 31

[76] G. H. Weiss, R. K. Gupta, J. A. Ferretti, and E. D. Becker, "The choice of optimal parameters for measurement of spin-lattice relaxation times. I. Mathematical formulation," *J. Mag. Res.*, vol. 37, pp. 369–79, Feb. 1980. 31

[77] Y. Zhang, H. N. Yeung, M. O'Donnell, and P. L. Carson, "Determination of sample time for T1 measurement," *J. Mag. Res. Im.*, vol. 8, pp. 675–81, May 1998. 31, 54

[78] H. Z. Wang, S. J. Riederer, and J. N. Lee, "Optimizing the precision in T1 relaxation estimation using limited flip angles," *Mag. Res. Med.*, vol. 5, pp. 399–416, Nov. 1987. 31, 35

[79] J. A. Jones, P. Hodgkinson, A. L. Barker, and P. J. Hore, "Optimal sampling strategies for the measurement of spin-spin relaxation times," *J. Mag. Res. B*, vol. 113, pp. 25–34, Oct. 1996. 31, 54

[80] J. Imran, François. Langevin, and Hervé. Saint-Jalmes, "Two-point method for T1 estimation with optimized gradient-echo sequence," *Mag. Res. Im.*, vol. 17, pp. 1347–56, Nov. 1999. 31, 54

[81] S. C. L. Deoni, T. M. Peters, and B. K. Rutt, "Determination of optimal angles for variable nutation proton magnetic spin-lattice, T_1, and spin-spin, T_2, relaxation times measurement," *Mag. Res. Med.*, vol. 51, pp. 194–9, Jan. 2004. 31, 54

[82] L. Fleysher, R. Fleysher, S. Liu, W. Zaaraoui, and O. Gonen, "Optimizing the precision-per-unit-time of quantitative MR metrics: Examples for T_1, T_2, and DTI," *Mag. Res. Med.*, vol. 57, pp. 380–7, Feb. 2007. 31

[83] M. Akçakaya, S. Weingärtner, Sébastien. Roujol, and R. Nezafat, "On the selection of sampling points for myocardial T1 mapping," *Mag. Res. Med.*, vol. 73, pp. 1741–53, May 2015. 31, 54

[84] C. M. Lewis, S. A. Hurley, M. E. Meyerand, and C. G. Koay, "Data-driven optimized flip angle selection for T1 estimation from spoiled gradient echo acquisitions," *Mag. Res. Med.*, vol. 76, pp. 792–802, Sept. 2016. 31, 54

[85] Y. Liu, J. R. Buck, and V. N. Ikonomidou, "Generalized min-max bound-based MRI pulse sequence design framework for wide-range T1 relaxometry: A case study on the tissue specific imaging sequence," *PLoS One*, vol. 12, pp. 1–20, 02 2017. 31

[86] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Philosophical Society*, vol. 22, pp. 700–25, July 1925. 32

[87] H. Cramér, *Mathematical methods of statistics*. Princeton: Princeton Univ. Press, 1946. 33, 66

[88] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, vol. 24, pp. 586–602, Dec. 1953. 33

[89] J. Pauly, P. Le Roux, D. Nishimura, and A. Macovski, "Parameter relations for the Shinnar-Le Roux selective excitation pulse design algorithm," *IEEE Trans. Med. Imag.*, vol. 10, pp. 53–65, Mar. 1991. 42, 75, 115, 128

[90] L. I. Sacolick, F. Wiesinger, I. Hancu, and M. W. Vogel, "B1 mapping by Bloch-Siegert shift," *Mag. Res. Med.*, vol. 63, pp. 1315–22, May 2010. 43, 56, 75, 106, 116

[91] H. Sun, W. A. Grissom, and J. A. Fessler, "Regularized estimation of Bloch-Siegert B1+ Maps in MRI," in *Proc. IEEE Intl. Conf. on Image Processing*, pp. 3646–50, 2014. 43, 75, 117

[92] L. Ying and J. Sheng, "Joint image reconstruction and sensitivity estimation in SENSE (JSENSE)," *Mag. Res. Med.*, vol. 57, pp. 1196–1202, June 2007. 43, 75, 116, 125

[93] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annual ACM-SIAM Symp. Disc. Alg. (SODA)*, pp. 1027–35, 2007. 43, 72, 110

[94] L. R. Frank, E. C. Wong, and R. B. Buxton, "Slice profile effects in adiabatic inversion: Application to multislice perfusion imaging," *Mag. Res. Med.*, vol. 38, pp. 558–64, Oct. 1997. 43

[95] J. K. Barral, E. Gudmundson, N. Stikov, M. Etezadi-Amoli, P. Stoica, and D. G. Nishimura, "A robust methodology for in vivo T1 mapping," *Mag. Res. Med.*, vol. 64, pp. 1057–67, Oct. 2010. 44, 56, 72

[96] C. E. Carney, S. T. S. Wong, and S. Patz, "Analytical solution and verification of diffusion effect in SSFP," *Mag. Res. Med.*, vol. 19, pp. 240–6, June 1991. 44, 85

[97] E. X. Wu and R. B. Buxton, "Effect of diffusion on the steady-state magnetization with pulsed field gradients," *J. Mag. Res.*, vol. 90, pp. 243–53, Nov. 1990. 44, 85

[98] R. Kaiser, E. Bartholdi, and R. R. Ernst, "Diffusion and field-gradient effects in NMR Fourier spectroscopy," *J. Chem. Phys.*, vol. 60, pp. 2966–79, Apr. 1974. 44

[99] R. Heule, P. Bär, C. Mirkes, K. Scheffler, S. Trattnig, and O. Bieri, "Triple-echo steady-state T2 relaxometry of the human brain at high to ultra-high fields," *NMR in Biomedicine*, vol. 27, pp. 1037–45, Sept. 2014. 52

[100] A. Mackay, K. Whittall, J. Adler, D. Li, D. Paty, and D. Graeb, "In vivo visualization of myelin water in brain by magnetic resonance," *Mag. Res. Med.*, vol. 31, pp. 673–7, June 1994. 52, 57, 90, 107, 109, 110, 111, 113, 120

[101] S. C. L. Deoni, L. Matthews, and S. H. Kolind, "One component? Two components? Three? The effect of including a nonexchanging "free" water component in multi-component driven equilibrium single pulse observation of T1 and T2," *Mag. Res. Med.*, vol. 70, pp. 147–54, July 2013. 52, 57, 90, 106

[102] R. D. Gill and B. Y. Levit, "Applications of the van Trees inequality: A Bayesian Cramér-rao bound," *Bernoulli*, vol. 1, no. 1/2, pp. 59–79, 1995. 54

[103] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, "Magnetic resonance fingerprinting," *Nature*, vol. 495, pp. 187–93, Mar. 2013. 55, 56, 72, 85

[104] B. Zhao, J. Haldar, K. Setsompop, and L. L. Wald, "Optimal experiment design for magnetic resonance fingerprinting," in *embc*, pp. 453–6, 2016. 55

[105] J. P. Haldar, J. Anderson, and S. W. Sun, "Maximum likelihood estimation of T1 relaxation parameters using VARPRO," in *Proc. Intl. Soc. Mag. Res. Med.*, p. 41, 2007. 56, 72

[106] D. Hernando, J. P. Haldar, B. P. Sutton, J. Ma, P. Kellman, and Z.-P. Liang, "Joint estimation of water/fat images and field inhomogeneity map," *Mag. Res. Med.*, vol. 59, pp. 571–80, Mar. 2008. 56, 72

[107] E. Staroswiecki, K. L. Granlund, M. T. Alley, G. E. Gold, and B. A. Hargreaves, "Simultaneous estimation of T2 and apparent diffusion coefficient in human articular cartilage in vivo with a modified three-dimensional double echo steady state (DESS) sequence at 3 T," *Mag. Res. Med.*, vol. 67, no. 4, pp. 1086–96, 2012. 56

[108] J. D. Trzasko, P. M. Mostardi, S. J. Riederer, and A. Manduca, "Estimating T1 from multichannel variable flip angle SPGR sequences," *Mag. Res. Med.*, vol. 69, no. 6, pp. 1787–94, 2013. 56, 72

[109] D. F. McGivney, E. Pierre, D. Ma, Y. Jiang, H. Saybasili, V. Gulani, and M. A. Griswold, "SVD compression for magnetic resonance fingerprinting in the time domain," *IEEE Trans. Med. Imag.*, vol. 33, pp. 2311–22, Dec. 2014. 56, 72, 83, 85

[110] B. Zhao, F. Lam, and Z.-P. Liang, "Model-based MR parameter mapping with sparsity constraints: parameter estimation and performance bounds," *IEEE Trans. Med. Imag.*, vol. 33, pp. 1832–44, Sept. 2014. 56

[111] B. Zhao, W. Lu, T. K. Hitchens, F. Lam, C. Ho, and Z.-P. Liang, "Accelerated MR parameter mapping with low-rank and sparsity constraints," *Mag. Res. Med.*, vol. 74, pp. 489–98, Aug. 2015. 56, 72

[112] S. F. Cauley, K. Setsompop, D. Ma, Y. Jiang, H. Ye, E. Adalsteinsson, M. A. Griswold, and L. L. Wald, "Fast group matching for MR fingerprinting reconstruction," *Mag. Res. Med.*, vol. 74, pp. 523–8, Aug. 2015. 56, 83

[113] B. Zhao, K. Setsompop, H. Ye, S. Cauley, and L. L. Wald, "Maximum likelihood reconstruction for magnetic resonance fingerprinting," *IEEE Trans. Med. Imag.*, vol. 35, pp. 1812–23, Aug. 2016. 56, 72

[114] J. Assländer, M. A. Cloos, F. Knoll, D. K. Sodickson, Jüergen. Hennig, and R. Lattanzi, "Low rank alternating direction method of multipliers reconstruction for MR fingerprinting," *Mag. Res. Med.*, vol. 79, pp. 83–96, Jan. 2018. 56, 83, 85

[115] M. Yang, D. Ma, Y. Jiang, J. Hamilton, N. Seiberlich, M. A. Griswold, and D. McGivney, "Low rank approximation methods for MR fingerprinting with large scale dictionaries," *Mag. Res. Med.*, vol. 79, pp. 2392–400, Apr. 2018. 56, 83

[116] D. A. Feinberg, L. E. Crooks, P. Sheldon, J. H. Iii, J. Watts, and M. Arakawa, "Magnetic resonance imaging the velocity vector components of fluid flow," *Mag. Res. Med.*, vol. 2, pp. 555–66, Dec. 1985. 56

[117] D. L. Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, "Diffusion tensor imaging: Concepts and applications," *J. Mag. Res. Im.*, vol. 13, pp. 534–546, Apr. 2001. 56

[118] D. S. Tuch, V. J. Wedeen, A. M. Dale, J. S. George, and J. W. Belliveau, "Conductivity tensor mapping of the human brain using diffusion tensor MRI," *Proc. Natl. Acad. Sci.*, vol. 98, pp. 11697–701, Sept. 2001. 56

[119] K. Sekihara, S. Matsui, and H. Kohno, "NMR imaging for magnets with large nonuniformities," *IEEE Trans. Med. Imag.*, vol. 4, pp. 193–9, Dec. 1985. 56

[120] G. R. Morrell, "A phase-sensitive method of flip angle mapping," *Mag. Res. Med.*, vol. 60, pp. 889–94, Oct. 2008. 56

[121] S. C. L. Deoni, B. K. Rutt, T. Arun, C. Pierpaoli, and D. K. Jones, "Gleaning multicomponent T1 and T2 information from steady-state imaging data," *Mag. Res. Med.*, vol. 60, pp. 1372–87, Dec. 2008. 57, 90, 91, 106, 108

[122] G. S. Kimeldorf and G. A. Wahba, "A correspondence between Bayesian estimation on stochastic processes and smoothing by splines," *Ann. Math. Stat.*, vol. 41, pp. 495–502, Apr. 1970. 57

[123] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, May 1950. 57

[124] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–97, Sept. 1995. 57

[125] C. Saunders, A. Gammerman, and V. Vovk, "Ridge regression learning algorithm in dual variables," in *Proc. Intl. Conf. Mach. Learn*, pp. 515–21, 1998. 57

[126] C. Huang, C. G. Graff, E. W. Clarkson, A. Bilgin, and M. I. Altbach, "T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing," *Mag. Res. Med.*, vol. 67, pp. 1355–66, May 2012. 57

[127] C. Huang, A. Bilgin, T. Barr, and M. I. Altbach, "T2 relaxometry with indirect echo compensation from highly undersampled data," *Mag. Res. Med.*, vol. 70, pp. 1026–37, Oct. 2013. 57

[128] L. Zhao, X. Feng, and C. H. Meyer, "Direct and accelerated parameter mapping using the unscented Kalman filter," *Mag. Res. Med.*, vol. 75, pp. 1989–99, May 2016. 57

[129] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. Computational Learning Theory (COLT)*, pp. 416–426, 2001. LNCS 2111. 60

[130] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, Feb. 1970. 61

[131] I. Steinwart and A. Christmann, *Support vector machines*. Springer, 2008. 61

[132] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *NIPS*, 2007. 64, 65

[133] D. J. Sutherland and J. Schneider, "On the error of random Fourier features," in *Proc. Intl. Conf. on Uncertainty in AI*, 2015. 65, 85

[134] Z. Wu, "Generalized Bochner's theorem for radial function," *Approximation Theory and its Applications*, vol. 13, no. 3, pp. 47–57, 1997. 65

[135] M. A. Woodbury, "Inverting modified matrices," 1950. Tech. Report 42, Stat. Res. Group, Princeton Univ. 65

[136] V. N. Vapnik, *Statistical learning theory*. Wiley-Interscience, 1998. 65

[137] M. M. Siddiqui, "Statistical inference for Rayleigh distributions," *RADIO SCIENCE Journal of Research NBS/USNC-URSI*, vol. 68D, pp. 1005–10, Sept. 1964. 70, 109

[138] J. Gong and J. P. Hornak, "A fast T1 algorithm," *Mag. Res. Im.*, vol. 10, no. 4, pp. 623–6, 1992. 72

[139] J. He, Q. Liu, A. G. Christodoulou, C. Ma, F. Lam, and Z.-P. Liang, "Accelerated high-dimensional MR imaging with sparse sampling using low-rank tensors," *IEEE Trans. Med. Imag.*, vol. 35, pp. 2119–29, Sept. 2016. 72

[140] M. A. Alvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," Tech. Rep. MIT-CSAIL-TR-2011-033, MIT, CSAIL, June 2011. 86

[141] S. Aja-Fernándezand G. Vegas-Sánchez-Ferrero, *Statistical analysis of noise in MRI: modeling, filtering and estimation*. Springer, 2016. 86

[142] O. Cohen, B. Zhu, and M. S. Rosen, "Deep learning for rapid sparse MR fingerprinting reconstruction," 2017. arxiv 1710.05267. 87

[143] P. Virtue, S. X. Yu, and M. Lustig, "Better than real: Complex-valued neural nets for MRI fingerprinting," in *Proc. IEEE Intl. Conf. on Image Processing*, pp. 3953–7, 2017. 87

[144] P. Morell, *Myelin*. Springer, 1984. 89

[145] M. M. Goldenberg, "Multiple sclerosis review," *Pharmacy and Therapeutics*, vol. 37, pp. 175–84, Mar. 2012. 89

[146] V. Vasilescu, E. Katona, V. Simplaceanu, and D. Demco, "Water compartments in the myelinated nerve. III. Pulsed NMR results," *Experientia*, vol. 34, pp. 1443–4, Nov. 1978. 89

[147] R. S. Menon and P. S. Allen, "Application of continuous relaxation time distributions to the fitting of data from model systems and excised tissue," *Mag. Res. Med.*, vol. 20, pp. 214–27, Aug. 1991. 90

[148] W. A. Stewart, A. L. Mackay, K. P. Whittall, G. R. W. Moore, and D. W. Paty, "Spin-spin relaxation in experimental allergic encephalomyelitis. Analysis of CPMG data using a non-linear least-squares method and linear inverse theory," *Mag. Res. Med.*, vol. 29, pp. 767–75, June 1993. 90

[149] P. J. Gareau, B. K. Rutt, S. J. Karlik, and J. R. Mitchell, "Magnetization transfer and multicomponent T2 relaxation measurements with histopathologic correlation in an experimental model of MS," *J. Mag. Res. Im.*, vol. 11, pp. 586–95, June 2000. 90, 124

[150] S. Webb, C. A. Munro, R. Midha, and G. J. Stanisz, "Is multicomponent T2 a good measure of myelin content in peripheral nerve?," *Mag. Res. Med.*, vol. 49, pp. 628–45, Apr. 2003. 90, 124

[151] C. Laule, I. M. Vavasour, G. R. W. Moore, J. Oger, D. K. B. Li, D. W. Paty, and A. L. MacKay, "Water content and myelin water fraction in multiple sclerosis," *J. Neurol.*, vol. 251, pp. 284–93, Mar. 2004. 90

[152] C. Laule, E. Leung, D. K. B. Li, A. L. Traboulsee, D. W. Paty, A. L. MacKay, and G. R. W. Moore, "Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology," *Multiple Sclerosis J.*, vol. 12, pp. 747–53, Nov. 2006. 90

[153] M. D. Does and J. C. Gore, "Rapid acquisition transverse relaxometric imaging," *J. Mag. Res.*, vol. 147, pp. 116–20, Nov. 2000. 90

[154] T. Prasloski, A. Rauscher, A. L. MacKay, M. Hodgson, I. M. Vavasour, C. Laule, and B. Mädler, "Rapid whole cerebrum myelin water imaging using a 3D GRASE sequence," *NeuroImage*, vol. 63, pp. 533–9, Oct. 2012. 90, 116

[155] D. A. Feinberg and K. Oshio, "GRASE (gradient- and spin-echo) MR imaging: a new fast clinical imaging technique," *Radiology*, vol. 181, pp. 597–602, Nov. 1991. 90

[156] E. Alonso-Ortiz, I. R. Levesque, and G. B. Pike, "MRI-based myelin water imaging: A technical review," *Mag. Res. Med.*, vol. 73, pp. 70–81, Jan. 2015. 90, 118

[157] M. D. Does, "Inferring brain tissue composition and microstructure via MR relaxometry," *J. Neuroimag.*, 2018. 90, 120

[158] S. C. L. Deoni, "Correction of main and transmit magnetic field (B0 and B1) inhomogeneity effects in multicomponent-driven equilibrium single-pulse observation of T1 and T2," *Mag. Res. Med.*, vol. 65, pp. 1021–35, Apr. 2011. 90, 91, 106, 107, 108, 111, 113

[159] J. Zhang, S. H. Kolind, C. Laule, and A. L. MacKay, "Comparison of myelin water fraction from multiecho T2 decay curve and steady-state methods," *Mag. Res. Med.*, vol. 73, pp. 223–32, Jan. 2015. 90, 106, 116, 118

[160] C. L. Lankford and M. D. Does, "On the inherent precision of mcDESPOT," *Mag. Res. Med.*, vol. 69, pp. 127–36, Jan. 2013. 90, 91, 108

[161] R. G. Spencer and K. W. Fishbein, "Measurement of spin-lattice relaxation times and concentrations in systems with chemical exchange using the one-pulse sequence: breakdown of the Ernst model for partial saturation in nuclear magnetic resonance spectroscopy," *J. Mag. Res.*, vol. 142, pp. 120–35, Jan. 2000. 91, 95

[162] H. M. McConnell, "Reaction rates by nuclear magnetic resonance," *J. of Chemical Phys.*, vol. 28, pp. 430–31, Mar. 1958. 91

[163] K. L. Miller, "Asymmetries of the balanced SSFP profile. Part I: Theory and observation," *Mag. Res. Med.*, vol. 63, pp. 385–95, Feb. 2010. 97, 123

[164] K. L. Miller, S. M. Smith, and P. Jezzard, "Asymmetries of the balanced SSFP profile. Part II: White matter," *Mag. Res. Med.*, vol. 63, pp. 396–406, Feb. 2010. 97, 123

[165] W. Magnus, "On the exponential solution of differential equations for a linear operator," *Comm. Pure Appl. Math.*, vol. 7, pp. 649–73, Nov. 1954. 98

[166] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Prentice-Hall, 1974. 109

[167] R. M. Kroeker and R. M. Henkelman, "Analysis of biological NMR relaxation data with continuous distributions of relaxation data," *J. Mag. Res.*, vol. 69, pp. 218–35, Sept. 1986. 109

[168] K. P. Whittall and A. L. MacKay, "Quantitative interpretation of NMR relaxation data," *J. Mag. Res.*, vol. 84, pp. 134–52, Aug. 1989. 110, 120

[169] T. Prasloski, B. Mädler, Q.-S. Xiang, A. MacKay, and C. Jones, "Applications of stimulated echo correction to multicomponent T2 analysis," *Mag. Res. Med.*, vol. 67, no. 6, pp. 1803–14, 2012. 110

[170] J. Hennig, "Multiecho imaging sequences with low refocusing flip angles," *J. Mag. Res.*, vol. 88, pp. 397–407, July 1988. 110

[171] J.-F. Nielsen, D. Yoon, and D. C. Noll, "Small-tip fast recovery imaging using non-slice-selective tailored tip-up pulses and radiofrequency-spoiling," *Mag. Res. Med.*, vol. 69, pp. 657–66, Mar. 2013. 123, 128

[172] G. Varma, G. Duhamel, C. de Bazelaire, and D. C. Alsop, "Magnetization transfer from inhomogeneously broadened lines: A potential marker for myelin," *Mag. Res. Med.*, vol. 73, pp. 614–22, Feb. 2015. 124

[173] G. Varma, O. M. Girard, V. H. Prevost, G. Duhamel, and D. C. Alsop, "Interpretation of magnetization transfer from inhomogeneously broadened lines (ihMT) in tissues as a dipolar effect within motion restricted molecules," *J. Mag. Res.*, vol. 260, pp. 67–76, Nov. 2015. 124

[174] S. D. Swanson, D. I. Malyarenko, M. L. Fabiilli, R. C. Welsh, J.-F. Nielsen, and A. Srinivasan, "Molecular, dynamic, and structural origin of inhomogeneous magnetization transfer in lipid membranes," *Mag. Res. Med.*, vol. 77, pp. 1318–28, Mar. 2017. 124

[175] J. Pauly, D. Nishimura, and A. Macovski, "A k-space analysis of small-tip-angle excitation," *J. Mag. Res.*, vol. 81, pp. 43–56, Jan. 1989. 128, 129, 130, 132

[176] M. Shinnar, L. Bolinger, and J. S. Leigh, "The use of finite impulse response filters in pulse design," *Mag. Res. Med.*, vol. 12, pp. 81–7, Oct. 1989. 128

[177] C. Yip, J. A. Fessler, and D. C. Noll, "Iterative RF pulse design for multidimensional, small-tip-angle selective excitation," *Mag. Res. Med.*, vol. 54, pp. 908–17, Oct. 2005. 128, 133

[178] 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," *Mag. Res. Med.*, vol. 59, pp. 908–15, Apr. 2008. 128

[179] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi, and D. C. Noll, "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity," *IEEE Trans. Sig. Proc.*, vol. 53, pp. 3393–402, Sept. 2005. 133

[180] D. C. Noll, C. H. Meyer, J. M. Pauly, D. G. Nishimura, and A. Macovski, "A homogeneity correction method for magnetic resonance imaging with time-varying gradients," *IEEE Trans. Med. Imag.*, vol. 10, pp. 629–37, Dec. 1991. 134

[181] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities," *IEEE Trans. Med. Imag.*, vol. 22, pp. 178–88, Feb. 2003. 134

[182] W. A. Grissom, D. Xu, A. B. Kerr, J. A. Fessler, and D. C. Noll, "Fast large-tip-angle multidimensional and parallel RF pulse design in MRI," *IEEE Trans. Med. Imag.*, vol. 28, pp. 1548–59, Oct. 2009. 135

[183] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976. 135, 136