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 Engineering and Computer Science)
in the University of Michigan
2017

Doctoral Committee:

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

2017

TABLE OF CONTENTS

LIST OF FIGURES

iv

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Advances in Quantitative MRI:
Acquisition, Estimation, and Application

by

Gopal Nataraj

Co-Chairs: Jeffrey A. Fessler and Jon-Fredrik Nielsen

todo

CHAPTER 1

Introduction

Magnetic resonance imaging (MRI) is a non-invasive tool that has earned widespread clinical adoption due (among other factors) 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 comparably long scan times. For these reasons, MRI is expensive relative to other medical imaging modalities. Accordingly, one broad initiative recently advocated by the MR community is to increase the *value* of MRI examinations.

Two reasonable measures of an MRI acquisition's value are its *sensitivity* to a given disorder and its *specificity* in distinguishing it from others. The field of *quantitative MRI* (QMRI) seeks to use MRI data to build MR *biomarkers*, or measurable tissue properties that can increase the sensitivity and specificity of MRI for 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 tra-

ditionally 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 Thesis Overview

This thesis seeks to address the above challenges by building an automated workflow for QMRI. We borrow 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, **??**) 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, **??**) 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 Thesis Proposal Organization

The main body of this thesis proposal is organized as follows:

- Chapter 2 reviews relevant background material on MRI and optimization.
- Chapter 3 discusses methods for MRI parameter estimation from likelihood models and applies these methods to model-based MR relaxometry, (*i.e.*, estimation of relaxation parameters T_1, T_2), of interest for many neurological applications. It derives some content (especially regarding applications) from conference papers [3, 4].
- Chapter 4 introduces a minimax optimization approach 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, phantom studies, and *in vivo* experiments. It derives content mainly from journal paper [5], which substantially extends conference paper [6].

- Chapter **??** describes scalable MRI parameter estimation using kernel ridge regression. It derives some content from conference paper [7].
- Chapter **??** studies multi-compartmental models for relevant MR pulse sequences and proposes a new acquisition useful for myelin water fraction estimation, of interest in white matter disorders. It applies kernel-based MR parameter estimation to estimate myelin water fraction, in simulations and preliminary *in vivo* experiments. It derives some content from conference paper [8].
- Chapter **??** summarizes several items of possible future work (on both short- and long-term timescales) and presents a timeline for completion of this thesis.

In the thesis, the appendices will be organized as follows:

- Appendix A will propose an algorithm for combining multiple MRI datasets (as is necessary for many parameter estimation problems), when each dataset is acquired using multiple receiver coils.
- Appendix B will present an analysis of model mismatch due to the presence of diffusion, will show that neglecting diffusive effects during T_2 estimation can cause significant bias, and will suggest acquisition modifications for mitigating this bias.

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 [9, 10, 11].

2.1.1 Bloch Equations

The Bloch equations [12] 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

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

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

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 second-order effects such as multiple (possibly interacting) isochromat compartments (Chapter **??**) and diffusion (Appendix B).

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 $b'_{xy}(\mathbf{r}, t) = b'_x(\mathbf{r}, t) + i b'_y(\mathbf{r}, t) := b_{xy}(\mathbf{r}, t) e^{i\omega_0 t}$ 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 $m'_{xy}(\mathbf{r}, t) = m'_x(\mathbf{r}, t) + i m'_y(\mathbf{r}, t) := m_{xy}(\mathbf{r}, t) e^{i\omega_0 t}$ 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:

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

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

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 $b'_{xy}(\mathbf{r}, t) \approx \kappa^{\mathrm{t}}(\mathbf{r}) b'_{1,x}(t) \hat{i} + 0\hat{j}$, 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 $T_{\mathrm{P}} \sim 1\mathrm{ms}$) is much shorter than relaxation time constants (typically $T_1 \sim 1000\mathrm{ms}$ and $T_2 \sim 50\mathrm{ms}$ in brain tissue)

and thus neglect relaxation effects during excitation.

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

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

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

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

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

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

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

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

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:

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

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

6

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

7

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 [13] 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 [14]) 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 [16] 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 [14], 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 [15].

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:

$$\mathbf{m}'\left(\mathbf{r}, t_0 + \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right) = \mathbf{R}_{z'}\left(\phi'\left(\mathbf{r}, \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; T_\mathrm{P}\right)\right)\mathbf{E}\left(\mathbf{r}, \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; T_\mathrm{P}\right)\mathbf{m}'(\mathbf{r}, t_0 + T_\mathrm{P})$$

$$+ \mathbf{m}_0\left(\mathbf{r}, \frac{T_\mathrm{P}}{2} + T_\mathrm{E}; T_\mathrm{P}\right). \tag{2.19}$$

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

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

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

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

$$+ \mathbf{m}_0\left(\mathbf{r}, T_\mathrm{R}; \frac{T_\mathrm{P}}{2} + T_\mathrm{E}\right). \tag{2.21}$$

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 [16]. 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 [17].

[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

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

{eq:spgr-bmmp-t0}   (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

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

{eq:spgr-mxyp-t0}   (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:

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

{eq:spgr-mxyp-te}   (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.* $\phi'\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}; t_0 + \frac{T_{\mathrm{P}}}{2}\right) \approx \omega'(\mathbf{r}) T_{\mathrm{E}}$. 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 **??** 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):

{eq:spgr-int}
$$s_S\left(\mathbf{r}, t_0 + \frac{T_P}{2} + T_E\right) \propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_P}{2} + T_E\right) \mathrm{d}^3\mathbf{r} \tag{2.26}$$

$$\approx m_{xy}(\mathbf{r}, t_0 + T_P)e^{-T_E/T_2(\mathbf{r})} \int_{\mathbb{R}} e^{-i\omega' T_E} \mathsf{p}_{\omega'}(\omega')\ \mathrm{d}\omega'$$

$$= m_{xy}(\mathbf{r}, t_0 + T_P)e^{-T_E/T_2(\mathbf{r})} e^{-R'_2(\mathbf{r})T_E - i\bar\omega'(\mathbf{r})T_E}$$

{eq:spgr-model}
$$= \frac{im_0(\mathbf{r})\sin\left(\alpha(\mathbf{r})\right)\left(1 - e^{-T_R/T_1(\mathbf{r})}\right)}{1 - e^{-T_R/T_1(\mathbf{r})}\cos\left(\alpha(\mathbf{r})\right)} e^{-T_E/T_2^*(\mathbf{r})} e^{-i\bar\omega'(\mathbf{r})T_E}, \tag{2.27}$$

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

{sss,bkgrd,mri,ss,dess}

DESS [18, 19] 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 $\alpha(\mathbf{r}) := \alpha(\mathbf{r}, t_0 + rT_R + T_P; t_0 + rT_R), \forall r \in \{0, 1, 2, \dots\}$:

{eq:dess-ex}
$$\mathbf{m}'(\mathbf{r}, t_0 + T_P) = \mathbf{R}_{x'}(\alpha(\mathbf{r}))\mathbf{m}'(\mathbf{r}, t_0). \tag{2.28}$$

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:

{eq:dess-pr}
$$\mathbf{m}'(\mathbf{r}, t_0 + T_R) = \mathbf{R}_{z'}(\phi'(\mathbf{r}))\mathbf{E}(\mathbf{r}, T_R; T_P)\mathbf{m}'(\mathbf{r}, t_0 + T_P) + \mathbf{m}_0(\mathbf{r}, T_R; T_P), \tag{2.29}$$

where the abbreviation $\phi'(\mathbf{r}) := \phi'(\mathbf{r}, t_0 + (r+1)T_R; t_0 + rT_R + T_P), \forall r \in \{0, 1, 2, \dots\}$ 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)

12

gives the magnetization at the data acquisition time after RF excitation:

$$m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_P}{2} + T_E\right) = m'_{xy}(\mathbf{r}, t_0 + T_P)e^{-(T_E - T_P/2)/T_2(\mathbf{r})}e^{-i\phi'\left(\mathbf{r}, t_0 + \frac{T_P}{2} + T_E; t_0 + T_P\right)}$$

$$\approx m'_{xy}(\mathbf{r}, t_0 + T_P)e^{-T_E/T_2(\mathbf{r})}e^{-i\phi'\left(\mathbf{r}, t_0 + \frac{T_P}{2} + T_E; t_0 + \frac{T_P}{2}\right)} \quad (2.35)$$

$$\approx m'_{xy}(\mathbf{r}, t_0 + T_P)e^{-T_E/T_2(\mathbf{r})}e^{-i\omega'(\mathbf{r})T_E}, \quad (2.36)$$

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:

$$m'_{xy}(\mathbf{r}, t_0) = m'_{xy}\left(\mathbf{r}, t_0 - \left(T_E - \frac{T_P}{2}\right)\right)e^{-(T_E - T_P/2)/T_2(\mathbf{r})}e^{-i\phi'\left(\mathbf{r}, t_0; t_0 - \left(T_E - \frac{T_P}{2}\right)\right)}. \quad (2.37)$$

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

$$m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_P}{2} - T_E\right) = m'_{xy}(\mathbf{r}, t_0)e^{+(T_E - T_P/2)/T_2(\mathbf{r})}e^{+i\phi'\left(\mathbf{r}, t_0; t_0 - \left(T_E - \frac{T_P}{2}\right)\right)}$$

$$\approx m'_{xy}(\mathbf{r}, t_0)e^{+T_E/T_2(\mathbf{r})}e^{+i\phi'\left(\mathbf{r}, t_0 + \frac{T_P}{2}; t_0 + \frac{T_P}{2} - T_E\right)} \quad (2.38)$$

$$\approx m'_{xy}(\mathbf{r}, t_0)e^{+T_E/T_2(\mathbf{r})}e^{+i\omega'(\mathbf{r})T_E}. \quad (2.39)$$

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 **??** 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):

$$\{\text{eq:dess-def-int}\} \quad s_{\mathrm{D}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \mathrm{d}^3\mathbf{r} \tag{2.40}$$

$$\approx \int_{\mathbb{R}} \int_{\mathbb{R}} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} + T_{\mathrm{E}}\right) \mathsf{p}_{\phi'}(\phi') \mathsf{p}_{\omega'}(\omega') \, \mathrm{d}\phi' \, \mathrm{d}\omega'$$

$$\approx e^{-T_{\mathrm{E}}/T_2(\mathbf{r})} \int_{\mathbb{R}} m'_{xy}(\mathbf{r}, t_0 + T_{\mathrm{P}}) \mathsf{p}_{\phi'}(\phi') \, \mathrm{d}\phi' \int_{\mathbb{R}} e^{-i\omega' T_{\mathrm{E}}} \mathsf{p}_{\omega'}(\omega') \, \mathrm{d}\omega'$$

$$= +i m_0(\mathbf{r}) E_2(\mathbf{r}, T_{\mathrm{E}}) e^{-\left(R'_2(\mathbf{r}) + i\bar{\omega}'(\mathbf{r})\right) T_{\mathrm{E}}} \tan\frac{\alpha(\mathbf{r})}{2} \left(1 - \frac{\eta(\mathbf{r}, T_{\mathrm{R}})}{\xi(\mathbf{r}, T_{\mathrm{R}})}\right);$$

$$\{\text{eq:dess-def-model}\} \tag{2.41}$$

$$\{\text{eq:dess-ref-int}\} \quad s_{\mathrm{D}}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) \propto \int_{\mathbb{V}(\mathbf{r})} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) \mathrm{d}^3\mathbf{r} \tag{2.42}$$

$$\approx \int_{\mathbb{R}} \int_{\mathbb{R}} m'_{xy}\left(\mathbf{r}, t_0 + \frac{T_{\mathrm{P}}}{2} - T_{\mathrm{E}}\right) \mathsf{p}_{\phi'}(\phi') \mathsf{p}_{\omega'}(\omega') \, \mathrm{d}\phi' \, \mathrm{d}\omega'$$

$$\approx e^{+T_{\mathrm{E}}/T_2(\mathbf{r})} \int_{\mathbb{R}} m'_{xy}(\mathbf{r}, t_0) \mathsf{p}_{\phi'}(\phi') \, \mathrm{d}\phi' \int_{\mathbb{R}} e^{+i\omega' T_{\mathrm{E}}} \mathsf{p}_{\omega'}(\omega') \, \mathrm{d}\omega'$$

$$= -i m_0(\mathbf{r}) E_2^{-1}(\mathbf{r}, T_{\mathrm{E}}) e^{-\left(R'_2(\mathbf{r}) - i\bar{\omega}'(\mathbf{r})\right) T_{\mathrm{E}}} \tan\frac{\alpha(\mathbf{r})}{2} (1 - \eta(\mathbf{r}, T_{\mathrm{R}})),$$

$$\{\text{eq:dess-ref-model}\} \tag{2.43}$$

where (2.41) and (2.43) introduce intermediate variables

$$\eta(\mathbf{r}, t) := \sqrt{\frac{1 - E_2^2(\mathbf{r}, t)}{1 - E_2^2(\mathbf{r}, t)/\xi^2(\mathbf{r}, t)}};$$

$$\xi(\mathbf{r}, t) := \frac{1 - E_1(\mathbf{r}, t)\cos\alpha(\mathbf{r})}{E_1(\mathbf{r}, t) - \cos\alpha(\mathbf{r})}.$$

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.

---

[8]In theory, it suffices to design dephasing gradients to impart as few as one complete 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 such "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 model mismatch due to diffusive signal loss. Appendix B studies diffusion in DESS and discusses regimes of dephasing gradient moments which balance partial-spoiling versus diffusive sources of model mismatch.

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 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 [20] is one such iterative algorithm and repeats the following simple update:

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

where $\mathsf{P}_{\mathbb{X}}$ denotes projection onto $\mathbb{X}$; $\nabla_{\mathbf{x}} := \left[\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_L}\right]$ 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 [21] 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 $(\cdot)^{\mathsf{H}}$

---

[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 $\mathsf{P}_{\mathbb{X}}(\tilde{\mathbf{x}})$.

denotes conjugate transpose. The variable projection method [22] reduces the complexity of (2.47) when the forward model takes the partially linear structure $\mathbf{f}(\mathbf{x}) \equiv \mathbf{A}(\mathbf{x}_N)\mathbf{x}_L$ and the feasible set takes the partially unconstrained form $\mathbb{X} \equiv \mathbb{C}^{L_L} \times \mathbb{X}_N$, where $\mathbf{x}_L \in \mathbb{C}^{L_L}$; $\mathbf{x}_N \in \mathbb{X}_N$; and $\mathbf{A} : \mathbb{X}_N \mapsto \mathbb{C}^{D \times L_L}$ is a matrix function. These restrictions on (2.47) define a so-called separable least-squares problem:

{eq:sep-ls}
$$(\mathbf{x}_L^*, \mathbf{x}_N^*) \in \left\{ \arg \min_{\substack{\mathbf{x}_L \in \mathbb{C}^{L_L} \\ \mathbf{x}_N \in \mathbb{X}_N}} \|\mathbf{y} - \mathbf{A}(\mathbf{x}_N)\mathbf{x}_L\|_{\mathbf{W}}^2 \right\}. \tag{2.48}$$

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

$$\mathbf{x}_L^*(\mathbf{x}_N) = \arg \min_{\mathbf{x}_L \in \mathbb{C}^{L_L}} \|\mathbf{y} - \mathbf{A}(\mathbf{x}_N)\mathbf{x}_L\|_{\mathbf{W}}^2$$

{eq:sep-ls-lin}
$$= \left(\mathbf{W}^{1/2}\mathbf{A}(\mathbf{x}_N)\right)^{\dagger}\mathbf{W}^{1/2}\mathbf{y} \tag{2.49}$$

{eq:sep-ls-fullrnk}
$$= \left(\mathbf{A}^{\mathsf{H}}(\mathbf{x}_N)\mathbf{W}\mathbf{A}(\mathbf{x}_N)\right)^{-1}\mathbf{A}^{\mathsf{H}}(\mathbf{x}_N)\mathbf{W}\mathbf{y}, \tag{2.50}$$

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:

$$\mathbf{x}_N^* \in \left\{ \arg \min_{\mathbf{x}_N \in \mathbb{X}_N} \left\|\mathbf{y} - \mathbf{A}(\mathbf{x}_N)\left(\mathbf{A}^{\mathsf{H}}(\mathbf{x}_N)\mathbf{W}\mathbf{A}(\mathbf{x}_N)\right)^{-1}\mathbf{A}^{\mathsf{H}}(\mathbf{x}_N)\mathbf{W}\mathbf{y}\right\|_{\mathbf{W}}^2 \right\}$$

{eq:sep-ls-nonlin}
$$\equiv \left\{ \arg \max_{\mathbf{x}_N \in \mathbb{X}_N} \mathbf{y}^{\mathsf{H}}\mathbf{W}\mathbf{A}(\mathbf{x}_N)\left(\mathbf{A}^{\mathsf{H}}(\mathbf{x}_N)\mathbf{W}\mathbf{A}(\mathbf{x}_N)\right)^{-1}\mathbf{A}^{\mathsf{H}}(\mathbf{x}_N)\mathbf{W}\mathbf{y} \right\}, \tag{2.51}$$

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

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 $(\mathbf{x}_L^*(\mathbf{x}_N^*), \mathbf{x}_N^*)$ 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 $(\widehat{\mathbf{x}}_L(\widehat{\mathbf{x}}_N), \widehat{\mathbf{x}}_N)$.

QMRI Parameter Estimation
using Likelihood Models

3.1 Introduction

This brief chapter 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 $\{y_1(\mathbf{r}), \ldots, y_D(\mathbf{r})\}$, 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 $d \in \{1, \ldots, D\}$. 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 $\epsilon_d \sim \mathbb{CN}(0, \sigma_d^2)$ is assumed for simplicity[1] to be (circularly-symmetric) complex Gaussian noise [25, 26] 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 $\mathbf{y}(\mathbf{r}) := [y_1(\mathbf{r}), \ldots, y_D(\mathbf{r})]^\mathsf{T} \in \mathbb{C}^D$ 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 $\mathbf{s} := [s_1, \ldots, s_D]^\mathsf{T} : \mathbb{C}^L \times \mathbb{C}^K \times \mathbb{R}^{A \times D} \mapsto \mathbb{C}^D$ and acquisition parameter $\mathbf{P} := [\mathbf{p}_1, \ldots, \mathbf{p}_D] \in \mathbb{R}^{A \times D}$ associated with that scan profile. Here, noise $\boldsymbol{\epsilon}(\mathbf{r}) := [\epsilon_1(\mathbf{r}), \ldots, \epsilon_D(\mathbf{r})]^\mathsf{T} \in \mathbb{C}^D$ typically has diagonal covariance structure $\boldsymbol{\Sigma} := \mathrm{diag}\left([\sigma_1, \ldots, \sigma_D]^\mathsf{T}\right)$ due to independence across datasets, where $\mathrm{diag}(\cdot)$ 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 reduced to just two spatially varying latent parameters: desired parameter $T_1(\mathbf{r}) \in \mathbb{R}$ and nuisance parameter $c_1(\mathbf{r}) := im_0(\mathbf{r})e^{-T_\mathrm{E}/T_2^*(\mathbf{r})}e^{-i\bar{\omega}'(\mathbf{r})T_\mathrm{E}} \in \mathbb{C}$. We assign $\mathbf{x} \leftarrow [T_1, c_1]^\mathsf{T}$ and $\boldsymbol{\nu} \leftarrow \kappa^\mathrm{t}$ 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

---

[1]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 Cartesian, it may be preferable that y_d be estimated by more sophisticated techniques, *e.g.* [23, 24]. In such cases, reconstructed image noise is unlikely to be Gaussian-distributed.

$A \leftarrow 2$ 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: $\kappa^\mathrm{t}(\mathbf{r}) \in \mathbb{R}$, $T_1(\mathbf{r}) \in \mathbb{R}$, $\bar{\omega}'(\mathbf{r}) \in \mathbb{R}$, $c_2(\mathbf{r}) := m_0(\mathbf{r}) e^{-T_E/T_2^*(\mathbf{r})} \in \mathbb{C}$, and $T_2(\mathbf{r}) \in \mathbb{R}$. In this chapter, we take $\kappa^\mathrm{t}(\mathbf{r}) \in \mathbb{R}$ 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[2] due to Rician noise. These choices assign $\boldsymbol{\nu} \leftarrow [\kappa^\mathrm{t}, T_1]^\mathsf{T}$ 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 $\mathbf{r}_1, \dots, \mathbf{r}_V$:

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

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 $\mathbf{X} := [\mathbf{x}(\mathbf{r}_1), \dots, \mathbf{x}(\mathbf{r}_V)] \in \mathbb{C}^{L \times V}$ and known $\mathbf{N} := [\boldsymbol{\nu}(\mathbf{r}_1), \dots, \boldsymbol{\nu}(\mathbf{r}_V)] \in \mathbb{C}^{K \times V}$ parameter images (with fixed acquisition parameter $\mathbf{P}$) to reconstructed image data $\mathbf{Y} := [\mathbf{y}(\mathbf{r}_1), \dots, \mathbf{y}(\mathbf{r}_V)] \in \mathbb{C}^{D \times V}$, save for noise image $\mathbf{E} := [\boldsymbol{\epsilon}(\mathbf{r}_1), \dots, \boldsymbol{\epsilon}(\mathbf{r}_V)] \in \mathbb{C}^{D \times V}$. The goal in QMRI parameter estimation is to estimate latent parameter images $\mathbf{X}$ from MR

---

[2]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 [27], and the acquisitions we examine here are capable of producing SNR in tissue of at minimum 100 and usually more.

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:

$$\widehat{\mathbf{X}}_{\mathrm{ML}}(\mathbf{Y}; \mathbf{N}, \mathbf{P}) \in \left\{\arg\max_{\mathbf{X} \in \mathbb{X}^V} \mathsf{L}(\mathbf{X})\right\}$$

$$\equiv \left\{\arg\min_{\mathbf{X} \in \mathbb{X}^V} -\log \mathsf{L}(\mathbf{X})\right\} \tag{3.6}$$

$$= \left\{\arg\min_{\mathbf{X} \in \mathbb{X}^V} \sum_{v=1}^{V} \|\mathbf{y}(\mathbf{r}_v) - \mathbf{s}(\mathbf{x}(\mathbf{r}_v); \boldsymbol{\nu}(\mathbf{r}_v), \mathbf{P})\|_{\boldsymbol{\Sigma}^{-1}}^2\right\}$$

$$= \left\{\arg\min_{\mathbf{X} \in \mathbb{X}^V} \left\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \mathbf{S}(\mathbf{X}; \mathbf{N}, \mathbf{P}))\right\|_{\mathrm{F}}^2\right\}, \tag{3.7}$$

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

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

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

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 [28, 29] and extrapolated other nuisance latent object parameters m_0 and T_2^* from measurements at 1.5T [30]. For simplicity, we assumed no flip angle variation $\kappa^t \leftarrow 1$ 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 [31, 30] 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 $T_R \leftarrow 12.2\mathrm{ms}$ and echo times $T_E \leftarrow 4.67\mathrm{ms}$ 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 [32], the ML estimator (3.7), and the RL estimator (3.8). The MOM, ML, and RL estimators respectively took $0.11\mathrm{s}$, $0.75\mathrm{s}$, and $31\mathrm{s}$. 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.* [32, 33] 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 $(10^{1.5}, 10^{3.5})\mathrm{ms}$) 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
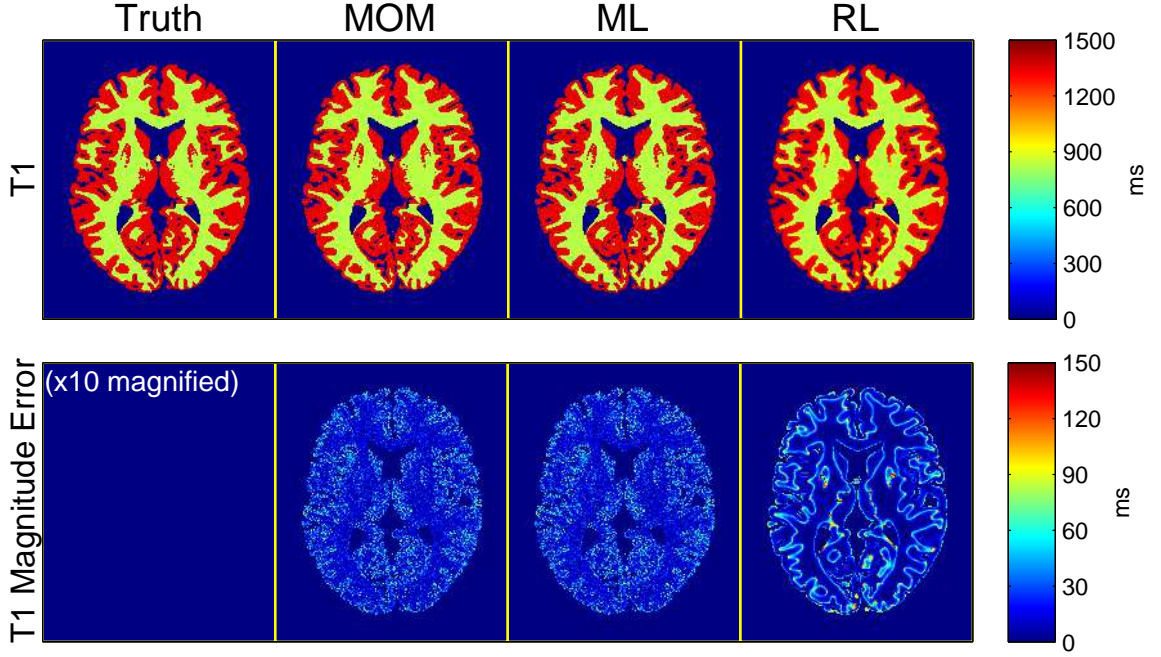
Figure 3.1: $\mathbf{T}_1$ MOM, ML, and RL estimates and corresponding error images, from two simulated SPGR scans. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Table 3.1 presents corresponding sample statistics.

{fig:relax,sim,t1}

the first five iterations and fixed thereafter. We employed a diagonal preconditioner to retain the linear convergence guarantees of GPM [34] 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 [35] 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

{eq:relax,conv-crit}
$$\left\|\boldsymbol{\Omega}^{-1}\big(\mathbf{X}^{(i)} - \mathbf{X}^{(i-1)}\big)\right\|_{\mathrm{F}} < 10^{-7}\left\|\boldsymbol{\Omega}^{-1}\big(\mathbf{X}^{(i)}\big)\right\|_{\mathrm{F}} \tag{3.12}$$

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

	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 [2]. T_1 values are in milliseconds. c_1 values are unitless. Fig. 3.1 presents corresponding images.

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.

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 $T_\mathrm{R} \leftarrow 17.5\mathrm{ms}$, and symmetric echo times $T_\mathrm{E} \leftarrow 4.67\mathrm{ms}$. 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 [19], 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)

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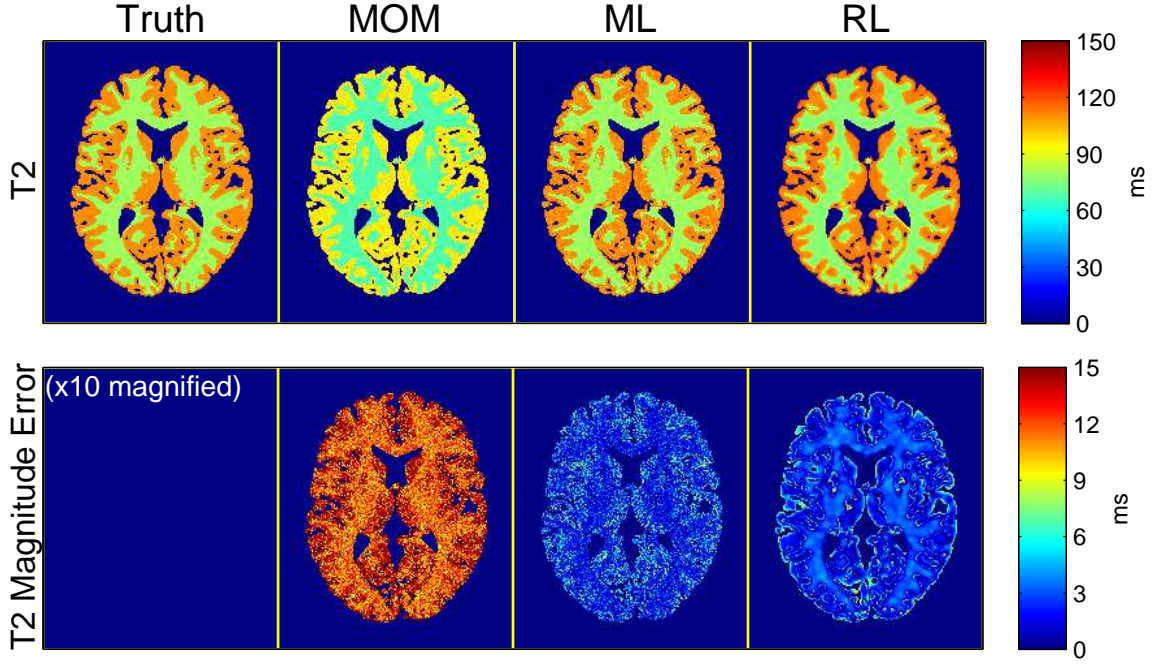
Figure 3.2: $\mathbf{T}_2$ MOM, ML, and RL estimates and corresponding error images, from one simulated DESS scan. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Table 3.2 presents corresponding sample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

{fig:relax,sim,t2}

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

26

	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 [2]. T_2 values are in milliseconds. c_2 values are unitless. Fig. 3.2 presents corresponding images.

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.

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 will 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 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 [33, 36]. 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

{s,relax,conc}

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.

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 [37]. Neurological applications include: lesion classification in multiple sclerosis [38]; tumor characterization [39, 40]; and symptom onset prediction in stroke [41, 42]. In addition, T_1, T_2 have shown promise for detecting hip and knee cartilage degeneration [43, 44] and for assessing cardiac dysfunction due to iron overload [45] or edema [46]. Motivated by this broad interest in T_1, T_2 mapping, this chapter 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 [47], multi-SE trains [48], or fast $\mathbf{k}$-space trajectories [49, 50, 51] have been proposed to accelerate T_1 [52, 53, 54, 55] and T_2 [56, 57, 58, 59] relaxometry with these classical sequences. These techniques are more sensitive to model non-idealities [60, 61, 62], and are still speed-limited by the long T_R required for (near)-complete T_1 recovery.

Steady-state (SS) pulse sequences [13, 14] 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 (*nuisance*) parameters, complicating quantification. Furthermore, some such methods [33, 63] 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 [64, 33, 63, 65, 66, 67, 68, 36, 69, 70].

The dual-echo steady-state (DESS) sequence [19] was recently proposed as a promising SS imaging technique for T_2 estimation [68]. Because it produces two distinct signals per excitation, the DESS sequence can reduce scan repetition requirements by recording twice as much data per scan. As with other SS methods, the resulting signals [71, 72] 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 [68, 36] or separate scans to first estimate nuisance parameters [4]. 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 [73, 74], flip angles for spoiled gradient-recalled echo (SPGR) sequences [75], and echo times for SE experiments [76]. More recent studies have considered additional scan design challenges, including scan time constraints [77], multiple latent parameters [78], multiple scan parameter types [79], and latent parameter spatial variation [80, 81, 82].

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 optimization. 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 discusses practical challenges and suggests future directions. Section 4.6 summarizes key contributions.

4.2 A CRB-Inspired Scan Selection Method

{s,scn-dsgn,crb}

4.2.1 The CRB and its Relevance to QMRI

{ss,scn-dsgn,crb,sig}

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

{n-dsgn,mod-vec-abbrev}
$$\mathbf{y} = \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) + \boldsymbol{\epsilon}, \tag{4.1}$$

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 $\boldsymbol{\epsilon} \sim \mathbb{CN}(\mathbf{0}, \boldsymbol{\Sigma})$, the likelihood function (3.4) is (to within constants independent of $\mathbf{x}$)

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

Under suitable[1] regularity conditions, the Fisher information matrix $\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) \in \mathbb{C}^{L \times L}$ [83] characterizes the imprecision of unbiased estimates of $\mathbf{x}$ from $\mathbf{y}$, given $\boldsymbol{\nu}$ and $\mathbf{P}$:

$$\mathbf{F}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}) := \mathsf{E}_{\mathbf{y}}\left((\nabla_{\mathbf{x}} \log \mathsf{L}(\mathbf{x}|\mathbf{y}))^{\mathsf{H}} \nabla_{\mathbf{x}} \log \mathsf{L}(\mathbf{x}|\mathbf{y})\right)$$

{eq:scn-dsgn,fisher}
$$= (\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}))^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\nu}, \mathbf{P}), \tag{4.3}$$

---

[1]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 [84] ensures that any unbiased[2] 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 [85], 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 $\widetilde{\Psi}^{\mathsf{t}}$, as viewed over "tight" object parameter ranges $\mathbb{X}^{\mathsf{t}}$ and $\mathbb{N}^{\mathsf{t}}$:

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

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

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

---

[2]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 [27]), 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

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

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 $\mathbb{N}^{\mathrm{b}} \supseteq \mathbb{N}^{\mathrm{t}}$:

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

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

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 [64, 63, 65] and subsequent T_2 mapping from other sequences [33, 4]. 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 $\mathbf{p} \leftarrow [\alpha_0, T_R]^\mathsf{T}$ 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. 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

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

to satisfy a total time constraint, $\|\mathbf{T}_R\|_1 \leq T_{\max}$. 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 [75, 33] or 1 DESS [68] 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 $S_{\mathrm{DESS}} \geq 1$ 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 $\mathbf{W} \leftarrow \mathrm{diag}(0.1, 1, 0)$ 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 $\mathbb{X}^{\mathrm{t}} \leftarrow \mathbb{T}_1^{\mathrm{t}} \times \mathbb{T}_2^{\mathrm{t}} \times 1$. Here, $\mathbb{T}_1^{\mathrm{t}} \leftarrow [800, 1400]$ms and $\mathbb{T}_2^{\mathrm{t}} \leftarrow [50, 120]$ms correspond with WM and GM regions of interest (ROIs) at 3T [28, 29]. We take $\mathbb{N}^{\mathrm{t}} \leftarrow [0.9, 1.1]$ to account for 10% transmit field spatial variation. Broadened ranges $\mathbb{X}^{\mathrm{b}} \leftarrow [400, 2000]$ms $\times [40, 200]$ms $\times 1$ and $\mathbb{N}^{\mathrm{b}} \leftarrow [0.5, 2]$ 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 $\widetilde{\Psi}^{\mathrm{t}}\!\left(\widehat{\mathbf{P}}\right)$ within 1% of global optimum $\widetilde{\Psi}^{\mathrm{t}}\!\left(\breve{\mathbf{P}}\right)$.

4.3.2 Scan Profile Comparisons

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

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

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

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 $\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}}(\widehat{\mathbf{P}})$ and $\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}}(\widehat{\mathbf{P}})$, Table 4.1 presents analogous worst-case standard deviations $\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{b}}(\widehat{\mathbf{P}})$ and $\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{b}}(\widehat{\mathbf{P}})$ over $\mathbb{X}^{\mathrm{b}} \times \mathbb{N}^{\mathrm{b}}$ to show how each estimator degrades over the broadened object parameter range. When viewed over tight range $\mathbb{X}^{\mathrm{t}} \times \mathbb{N}^{\mathrm{t}}$, the $(0,2)$ profile provides a 11.5% reduction in worst-case cost over the other choices. Extending to broadened range $\mathbb{X}^{\mathrm{b}} \times \mathbb{N}^{\mathrm{b}}$, this reduction

Scan	$(2,1)$	$(1,1)$	$(0,2)$
$\widehat{\alpha}_0^{\mathrm{spgr}}$	$(15,5)^\circ$	15°	$-$
$\widehat{\alpha}_0^{\mathrm{dess}}$	30°	10°	$(35,10)^\circ$
$\widehat{T}_R^{\mathrm{spgr}}$	$(12.2, 12.2)$	13.9	$-$
$\widehat{T}_R^{\mathrm{dess}}$	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 [68], 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.
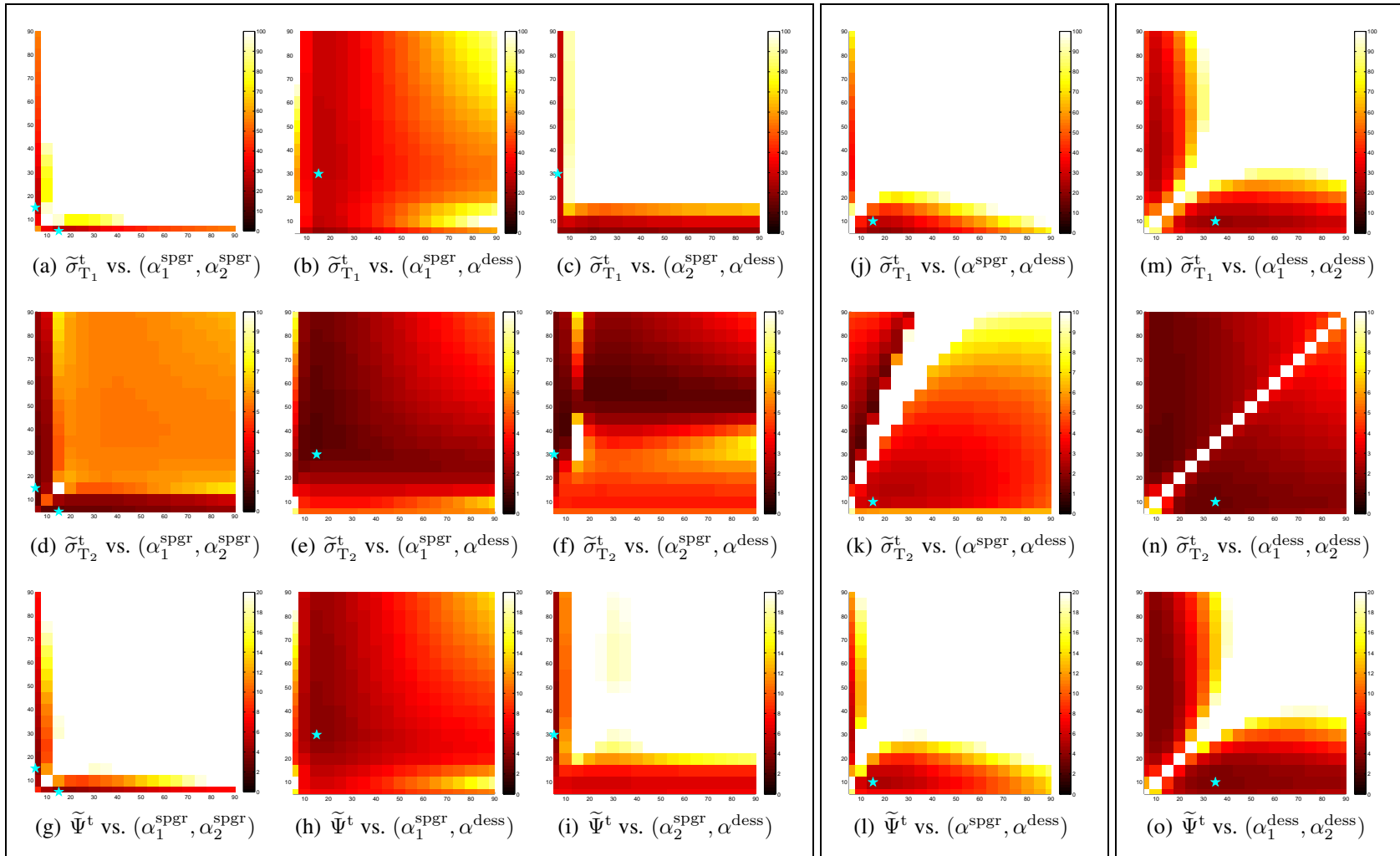
Figure 4.1: Worst-case standard deviations $\widetilde{\sigma}_{T_1}^t$ (top), $\widetilde{\sigma}_{T_2}^t$ (middle), and cost $\widetilde{\Psi}^t$ (bottom), versus pairs of nominal flip angles, holding other scan parameters fixed at selected profile $\widehat{P}$. Subfigures (a)-(i), (j)-(l), and (m)-(o) 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 $\widetilde{\sigma}_{T_1}^t$, $\widetilde{\sigma}_{T_2}^t$, and $\widetilde{\Psi}^t$ subfigures, respectively. The optimized $(0, 2)$ profile appears most robust to transmit field spatial variation.

lsgn,heat-map}

4.4 Experimental Validation and Results

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 [28, 29] and extrapolate other nuisance latent object parameters m_0 and T_2^* from measurements at 1.5T [30]. We assign these parameter values to the discrete anatomy of the BrainWeb digital phantom [31, 30] 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 $T_\mathrm{E} = 4.67$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 $\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}}$ and $\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}}$ reported in Table 4.1. All values are reported in milliseconds.

{tab:scn-dsgn,sim}

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[3] that, despite model nonlinearity and Rician noise, estimation bias in WM- and GM-like voxels is negligible. Sample standard deviations are consistent with $\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}}$ and $\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}}$ (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 $\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}}$ and $\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}}$. The separations between these distribution pairs visually depict how estimator

---

[3]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 $T_2 \leftarrow 110$
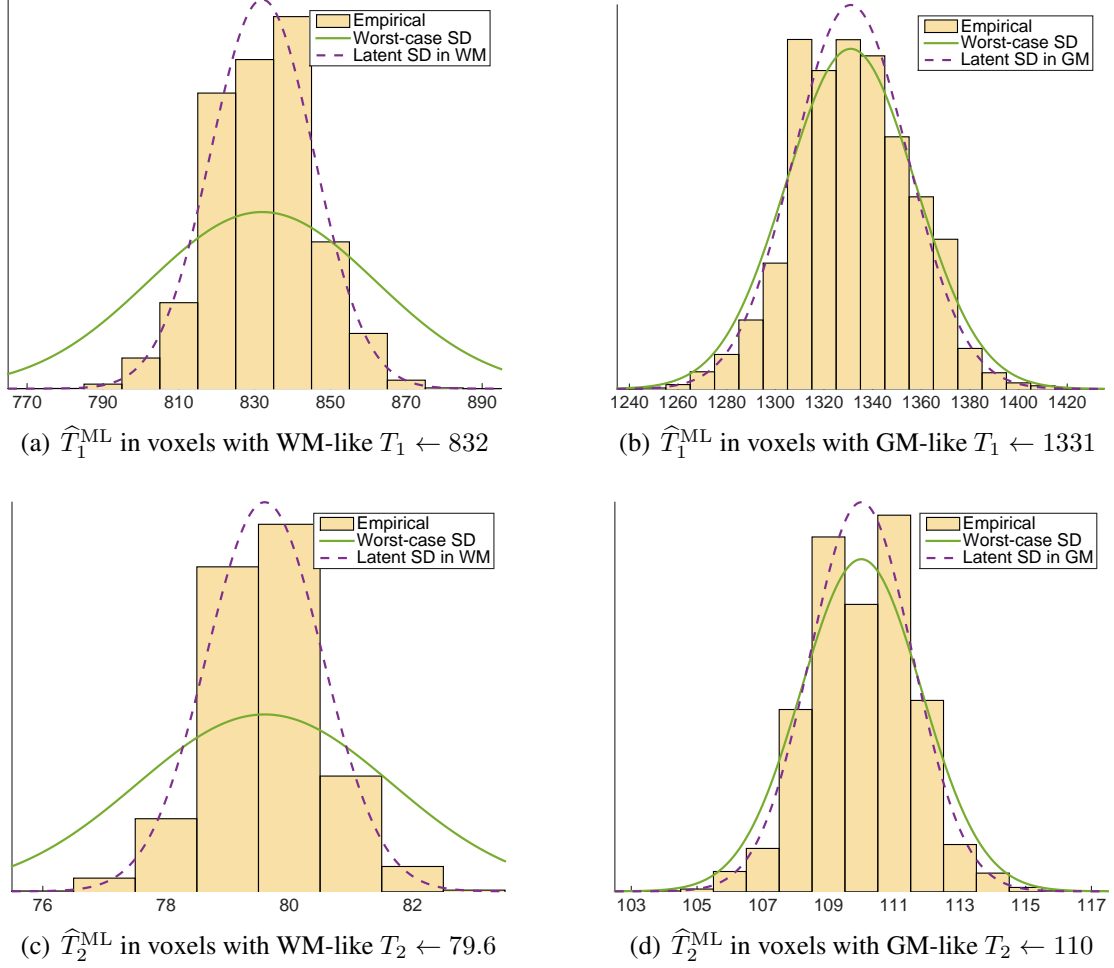
Figure 4.2: Histograms of T_1 and T_2 estimates from noisy independent measurements of a *single* nominal WM or GM value. In each plot, two normal distributions are overlaid, each with latent means T_1 and T_2. In (a)-(b) and (c)-(d), the solid green curve is $\mathcal{N}\left(T_1, (\widetilde{\sigma}_{\mathrm{T}_1}^{\mathrm{t}})^2)\right)$ and $\mathcal{N}\left(T_2, (\widetilde{\sigma}_{\mathrm{T}_2}^{\mathrm{t}})^2\right)$, 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.

{fig:scn-dsgn,normal}

40

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) [1]. 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 $\widehat{\alpha}_0$ and repetition times $\widehat{\mathbf{T}}_R$ 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 [86], 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 $T_E \leftarrow 4.67$ms before or after RF excitations. To avoid slice-profile effects, we sample 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 $(T_R, T_E) \leftarrow (1400, 14)$ms fixed and vary (adiabatic) inversion time $T_I \in \{50, 150, 450, 1350\}$ms across scans. For SE, we similarly hold $T_R \leftarrow 1000$ms fixed and vary echo time $T_E \in \{10, 30, 60, 150\}$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 [87]. 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 [88] from the resulting pair of datasets. We normalize this transmit field map estimate by the 0.075G peak Fermi pulse amplitute to estimate transmit coil spatial variation map $\mathbf{s}_t$. After calibration via separate measurements, we take $\mathbf{s}_t$ 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 [89] 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++ [90] 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 $\widehat{\mathbf{X}}_{\mathrm{ML}}(\mathbf{N}, \widehat{\mathbf{P}})$. We subsequently solve RL problem (3.8) with initialization $\widehat{\mathbf{X}}_{\mathrm{ML}}(\mathbf{N}, \widehat{\mathbf{P}})$ to obtain regularized estimates $\widehat{\mathbf{X}}_{\mathrm{RL}}(\mathbf{N}, \widehat{\mathbf{P}})$ 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[4] $\mathbf{T}_1$, then $\mathbf{T}_2$ estimation from IR and SE reference scans.

---

[4] 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 [91].

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 $\mathbf{s}_t$ as is used for SPGR/DESS profiles, we estimate $\mathbf{T}_1$ using a signal model similar to one proposed in [92], which accounts for imperfect excitation/refocusing and imperfect inversion. We then take both $\mathbf{T}_1$ and $\mathbf{s}_t$ 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 $\mathbb{X}^t$ and broad $\mathbb{X}^b$ 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 $\mathbb{X}^t$ and $\mathbb{X}^b$. $\mathbf{T}_1$ 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 [1] across the vial range. $\mathbf{T}_2$ estimates from all profiles are also in good agreement with NIST for vials within $\mathbb{X}^b$. SPGR/DESS profiles likely underestimate large T_2 values ($\geq$200ms) due to greater influence of diffusion in DESS [93, 94, 95], (studied further in Appendix B). SPGR/DESS profiles possibly overestimate and the IR/SE profile likely underestimates short ($\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
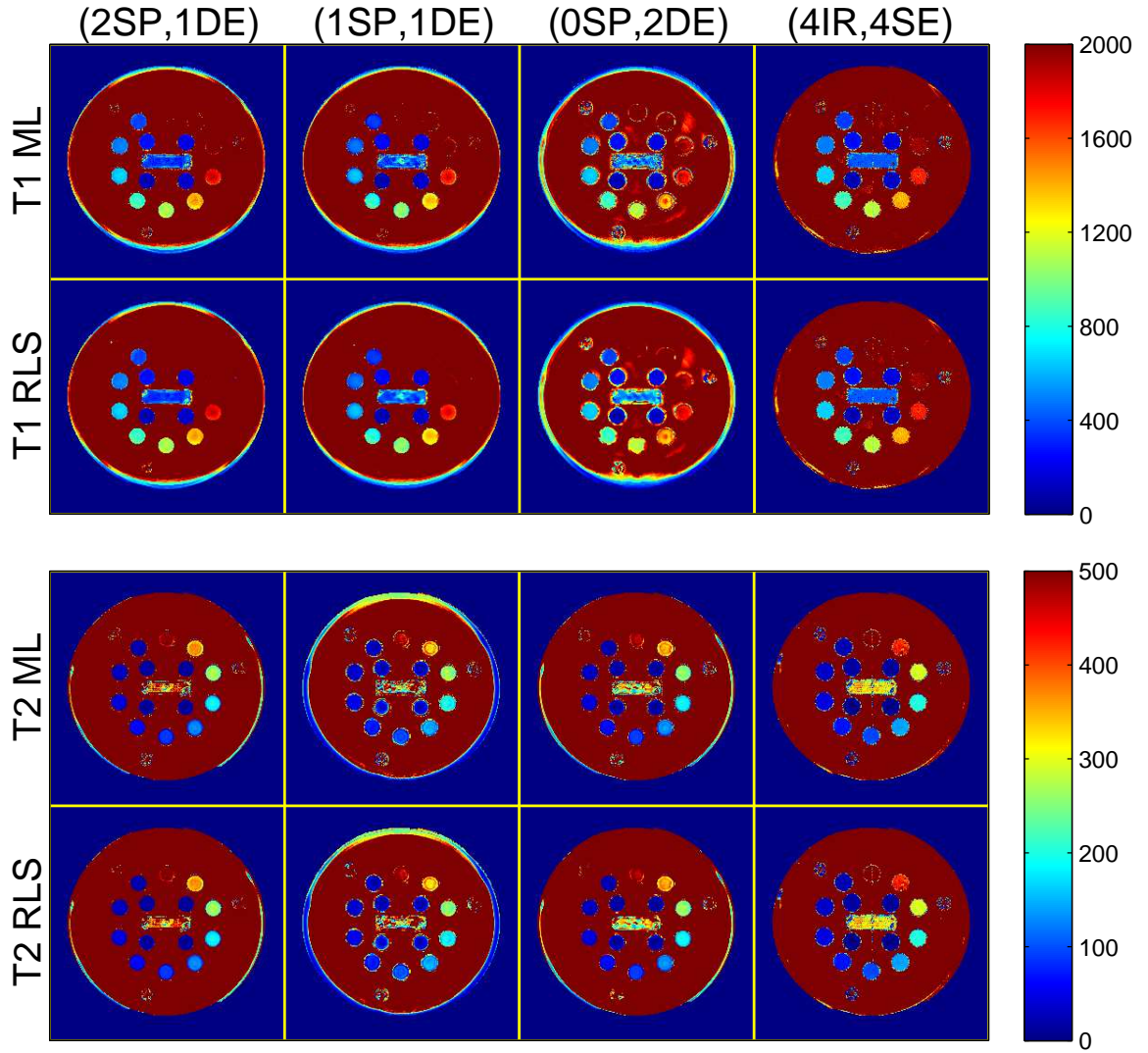
Figure 4.3: Colorized $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from an HPD® quantitative phantom. Columns correspond to scan profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Fig. 4.4 provides identical grayscale images that enumerate vials. Colorbar ranges are in milliseconds.

{fig:scn-dsgn,hpd,jet}

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 [2]) for phantom vials within (or nearly within) tight design range $\mathbb{X}^{\mathrm{t}}$ (marked orange in Fig. 4.4). Due to error propagation
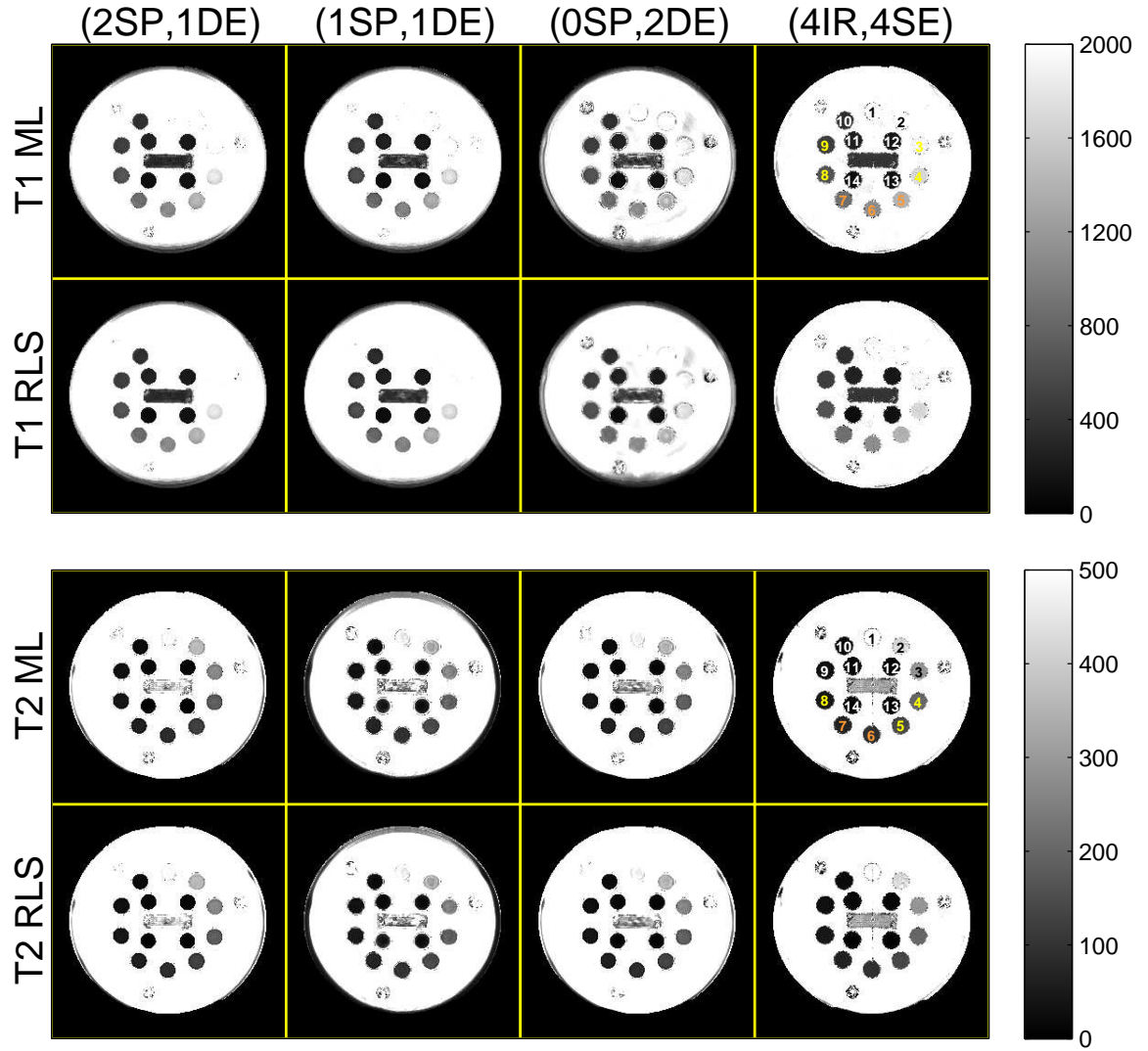
Figure 4.4: Grayscale $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from an HPD® quantitative phantom. Columns correspond to scan profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and 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.

{fig:scn-dsgn,hpd,gray}

from coil combination and $\mathbf{s}_t$ 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
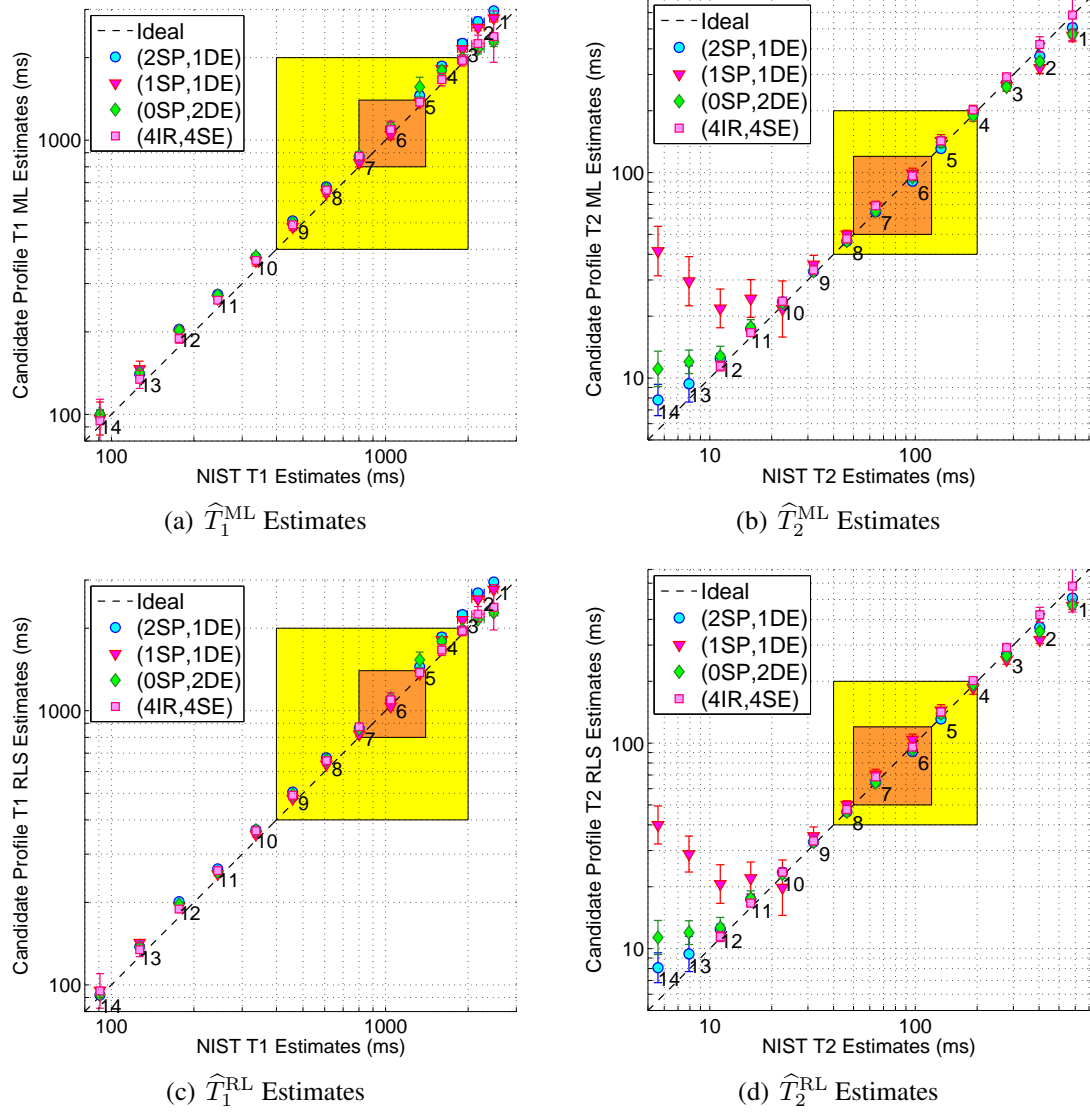
(d) $\widehat{T}_2^{\mathrm{RL}}$ Estimates

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 [1]. 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.

fig:scn-dsgn,hpd,ml-rls}

46

	(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 [1]). 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[5] 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.

---

[5]For each coil-combined dataset, we compute a separate 2D rigid transformation (with respect to the $T_{\mathrm{I}} = 50\mathrm{ms}$ 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.
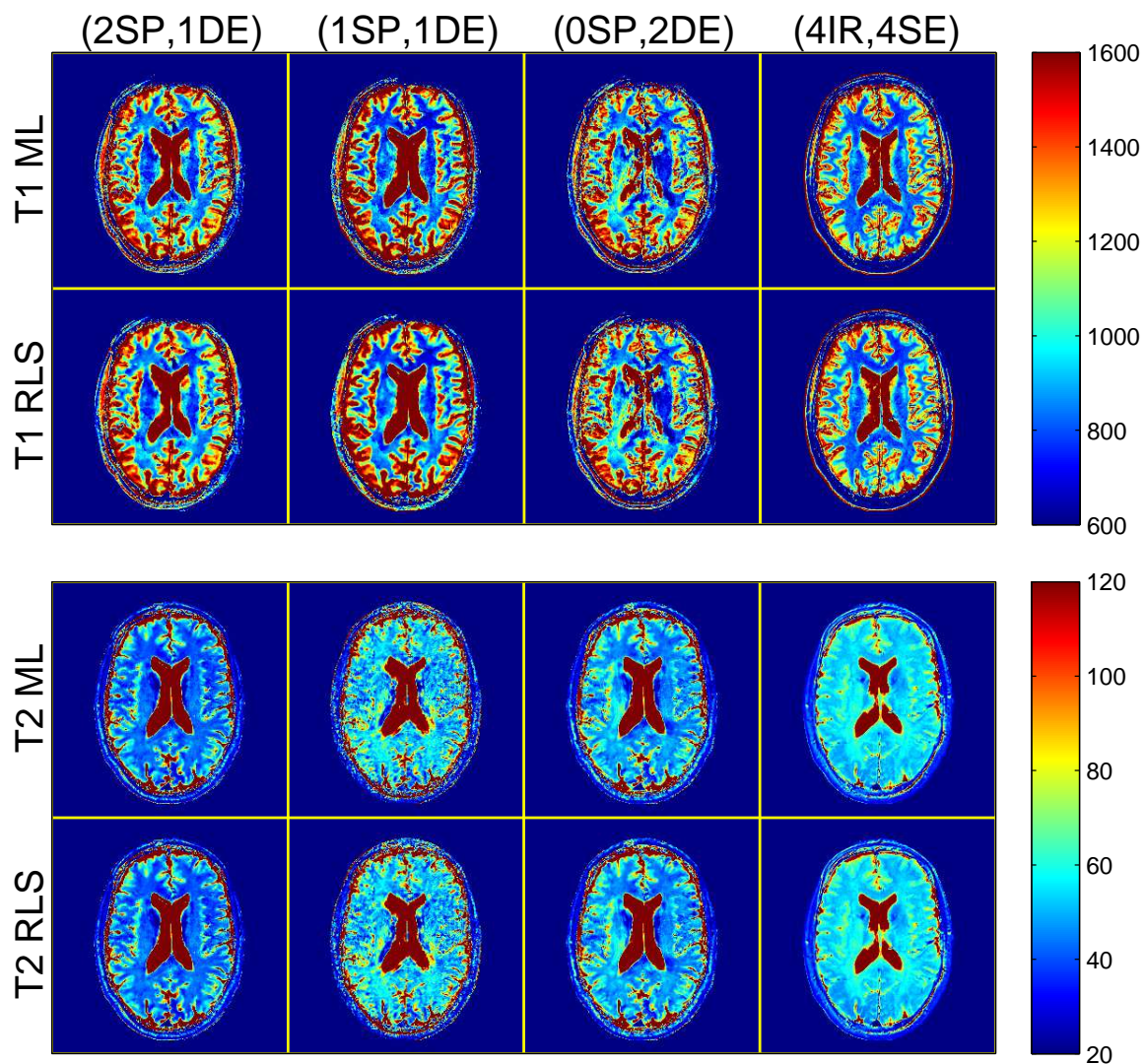
Figure 4.6: Colorized $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimates from the brain of a healthy volunteer. Columns correspond to profiles consisting of (2 SPGR, 1 DESS), (1 SPGR, 1 DESS), (0 SPGR, 2 DESS), and (4 IR, 4 SE) acquisitions. Rows distinguish $\mathbf{T}_1$ and $\mathbf{T}_2$ ML and RL estimators. Table 4.5 presents corresponding WM/GM within-ROI sample statistics. Colorbar ranges are in milliseconds.

{fig:scn-dsgn,brain,jet}

	(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.

b:scn-dsgn,hpd,std-dev}

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 computed[6] 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.

---

[6]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.
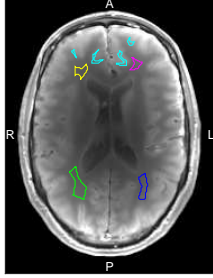
	ROI	(2SP,1DE)	(1SP,1DE)	(0SP,2DE)	(4IR,4SE)
	AR WM	840 ± 32	770 ± 31	840 ± 43	780 ± 22
	AL WM	740 ± 61	660 ± 45	740 ± 55	760 ± 24
$\widehat{T}_1^{\mathrm{ML}}$	PR WM	890 ± 88	860 ± 72	960 ± 84	810 ± 26
	PL WM	$860 \pm 70.$	850 ± 61	880 ± 79	820 ± 37
	A GM	1200 ± 210	1200 ± 230	1300 ± 230	1300 ± 180
	AR WM	840 ± 24	$770 \pm 20.$	840 ± 43	$780 \pm 20.$
	AL WM	740 ± 51	670 ± 37	740 ± 54	760 ± 23
$\widehat{T}_1^{\mathrm{RL}}$	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
	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
$\widehat{T}_2^{\mathrm{ML}}$	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
	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
$\widehat{T}_2^{\mathrm{RL}}$	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

{scn-dsgn,brain,roi,gray}

{tab:scn-dsgn,brain}

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 $\widehat{T}_1^{\mathrm{ML}}$ 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 $\widehat{T}_1^{\mathrm{ML}}$ estimates that are more consistently similar to IR WM $\widehat{T}_1^{\mathrm{ML}}$ 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 $\widehat{T}_2^{\mathrm{ML}}$ 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 $\widehat{T}_2^{\mathrm{ML}}$ than the reference IR/SE profile, though the $(0, 2)$ profile is in reasonable agreement with

other steady-state estimates [96]. These discrepancies may due to differences in sensitivity to multi-compartmental relaxation [97]. 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. Chapter **??** studies multi-compartmental relaxation in much greater detail.

4.5 Discussion and Future Work

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 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 [76, 74, 77, 78], 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 [80, 81] have addressed object parameter spatial variation by instead constructing cost functions related to the Bayesian CRB [98], 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 [80]. However, prior distributions are generally unknown, and may need to be estimated from data, as in [81].

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 er-

rors 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 [4].) 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. However, it is apparent that the nonlinear estimation procedures required for many mapping problems can amplify the influence of these secondary effects, often inducing substantial 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 [58, 99]. In practice, difficulties arise in efficient computation of signal gradients required in (4.3), which may demand more specialized techniques, as in [100]. Designing scan profiles involving such complex signal models would likely necessitate optimization techniques more involved than the simple grid searches used in this work.

4.6 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 $(2, 1)$ [4] 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 neglected higher-order effects. Nevertheless, the example application studied in this chapter illustrates that scan optimization can enable new parameter mapping techniques from established pulse sequences.

<div align="center">

CHAPTER 5

QMRI Parameter Estimation
via Regression with Kernels (PERK)

</div>

5.1 Introduction

In quantitative magnetic resonance imaging (QMRI), one seeks to estimate latent parameter images from suitably informative data. Since MR acquisitions are tunably sensitive to many physical processes (*e.g.*, relaxation [12], diffusion [101], and chemical exchange [102]), MRI parameter estimation is important for many QMRI applications (*e.g.*, relaxometry [103], diffusion tensor imaging [104], and multi-compartmental imaging [97]). Motivated by widespread applications, this chapter introduces a general method for fast MRI parameter estimation.

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, many works (*e.g.*, [105, 106, 92, 107, 99, 108, 109, 110, 58, 111, 112, 113, 5, 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 [104] or conductivity [117]) field mapping

techniques require estimation of at minimum 4 and 7 latent parameters per voxel, respectively. Phase-based longitudinal [118] or transverse [119, 87] field mapping could avoid noise-amplifying algebraic manipulations on reconstructed image data that are conventionally used to reduce signal dependencies on nuisance latent parameters. Compartmental fraction mapping [97, 8] from steady-state pulse sequences requires estimation of at least 7 [120] and as many as 10 [121] 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 (Section **??** 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 5.9 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

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 $\boldsymbol{\epsilon} \sim \mathbb{CN}(\mathbf{0}_D, \boldsymbol{\Sigma})$ 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 single 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 $\{(\mathbf{x}_1, \boldsymbol{\nu}_1, \boldsymbol{\epsilon}_1), \ldots, (\mathbf{x}_N, \boldsymbol{\nu}_N, \boldsymbol{\epsilon}_N)\}$ and corresponding measurements $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. 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[1] 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

---

[1]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

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

$$\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{n'=1}^{N} a_{l,n'} k(\mathbf{q}_n, \mathbf{q}_{n'}) + b_l - x_{l,n}\right)^2 \tag{5.8}$$

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

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

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

where $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the Gram matrix consisting of entries $k(\mathbf{q}_n, \mathbf{q}_{n'})$ for $n, n' \in \{1, \ldots, N\}$; $\mathbf{M} := \mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\mathsf{T} \in \mathbb{R}^{N \times N}$ is a de-meaning operator; $\boldsymbol{x}_l := [x_{l,1}, \ldots, x_{l,N}]^\mathsf{T}$; $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ 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 $\widehat{\mathbf{x}}$:

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

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

When $\rho_l > 0 \; \forall l \in \{1, \ldots, L\}$, 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

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

where positive definite matrix bandwidth $\boldsymbol{\Lambda} \in \mathbb{R}^{Q \times Q}$ controls the length scales in $\mathbf{q}$ over which the estimator $\widehat{\mathbf{x}}$ smooths and $\|\cdot\|_{\boldsymbol{\Gamma}} \equiv \left\|\boldsymbol{\Gamma}^{1/2}(\cdot)\right\|_2$ is a weighted ℓ^2-norm with PSD matrix weights $\boldsymbol{\Gamma}$. 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

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

(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 $d \in \{1, \ldots, D\}$ [27]. Second, we assume that the Gaussian kernel bandwidth matrix $\boldsymbol{\Lambda}$ has the block diagonal structure

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

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

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

{eq,exp-bmk}

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 $\boldsymbol{\Sigma} \to \mathbf{0}_{D \times D}$ for fixed $\boldsymbol{\Lambda_\alpha}$, $\mathsf{E}_{\boldsymbol{\alpha}|\mathbf{x},\boldsymbol{\nu}}(\mathbf{k}(\boldsymbol{\alpha},\boldsymbol{\nu})) \to \mathbf{k}(\boldsymbol{\mu},\boldsymbol{\nu})$ 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

{ss,perk,perf,cov}

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

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

{eq,cov}

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

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

where $\widetilde{\boldsymbol{\alpha}}_n := \boldsymbol{\mu} - \boldsymbol{\alpha}_n$ and $\boldsymbol{\Delta}(t) := \frac{1}{2}\big((t\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^{-2}\boldsymbol{\Sigma} + \mathbf{I}_D)\big)^{-1}\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^{-2}$ 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 $\tilde{z}(\mathbf{v}, s; \mathbf{q}) \leftarrow \cos\big(2\pi(\mathbf{v}^\mathsf{T} \mathbf{q} + s)\big)$. 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

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

For symmetric $\mathsf{p}_{\mathbf{v}}$, (5.22) exists [134] and is a Fourier transform. Thus choosing $\mathsf{p}_{\mathbf{v}} \leftarrow \mathcal{N}\big(\mathbf{0}_Q, (2\pi\mathbf{\Lambda})^{-2}\big)$ 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

sss,perk,pract,mod,dist}

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

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

where $\nabla_{\mathbf{x}}(\cdot)$ denotes row gradient with respect to $\mathbf{x}$ and the approximation holds well for moderately high-SNR measurements [27]. 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 $\mathbf{x}$ [84]. For good performance, it is thus reasonable to ensure $\mathbf{F}(\mathbf{x}, \boldsymbol{\nu})$ is well-conditioned over $\mathsf{supp}(p_{\mathbf{x},\boldsymbol{\nu}})$.

There are many strategies one could employ to control the condition number of $\mathbf{F}(\mathbf{x}, \boldsymbol{\nu})$ over $\mathsf{supp}(p_{\mathbf{x},\boldsymbol{\nu}})$. In our experiments, we used data [5] 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 $\mathsf{supp}(p_{\mathbf{x},\boldsymbol{\nu}})$ 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 $(\mathbf{x}, \boldsymbol{\nu}) \in \mathsf{supp}(p_{\mathbf{x},\boldsymbol{\nu}})$ can induce PERK estimation error over larger subsets of $\mathsf{supp}(p_{\mathbf{x},\boldsymbol{\nu}})$.

If many separate prior parameter estimates are available, one can estimate the particular shape of $p_{\mathbf{x},\boldsymbol{\nu}}$ empirically and then clip and renormalize $p_{\mathbf{x},\boldsymbol{\nu}}$ 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 $p_{\mathbf{x},\boldsymbol{\nu}} \equiv p_{\mathbf{x}}p_{\boldsymbol{\nu}}$ in which case one can still estimate $p_{\boldsymbol{\nu}}$ empirically but must set $p_{\mathbf{x}}$ 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 $\mathbf{C}_{\tilde{z}\tilde{z}}$ 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 **??**.

5.5.2.3 Choosing Kernel Bandwidth

It is desirable to choose the Gaussian kernel's bandwidth matrix $\boldsymbol{\Lambda}$ 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)

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

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, $T_{\mathrm{E}} \leftarrow 30\mathrm{ms}$ is the echo time and $\epsilon \sim \mathbb{CN}(0, 0.01^2)$ 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
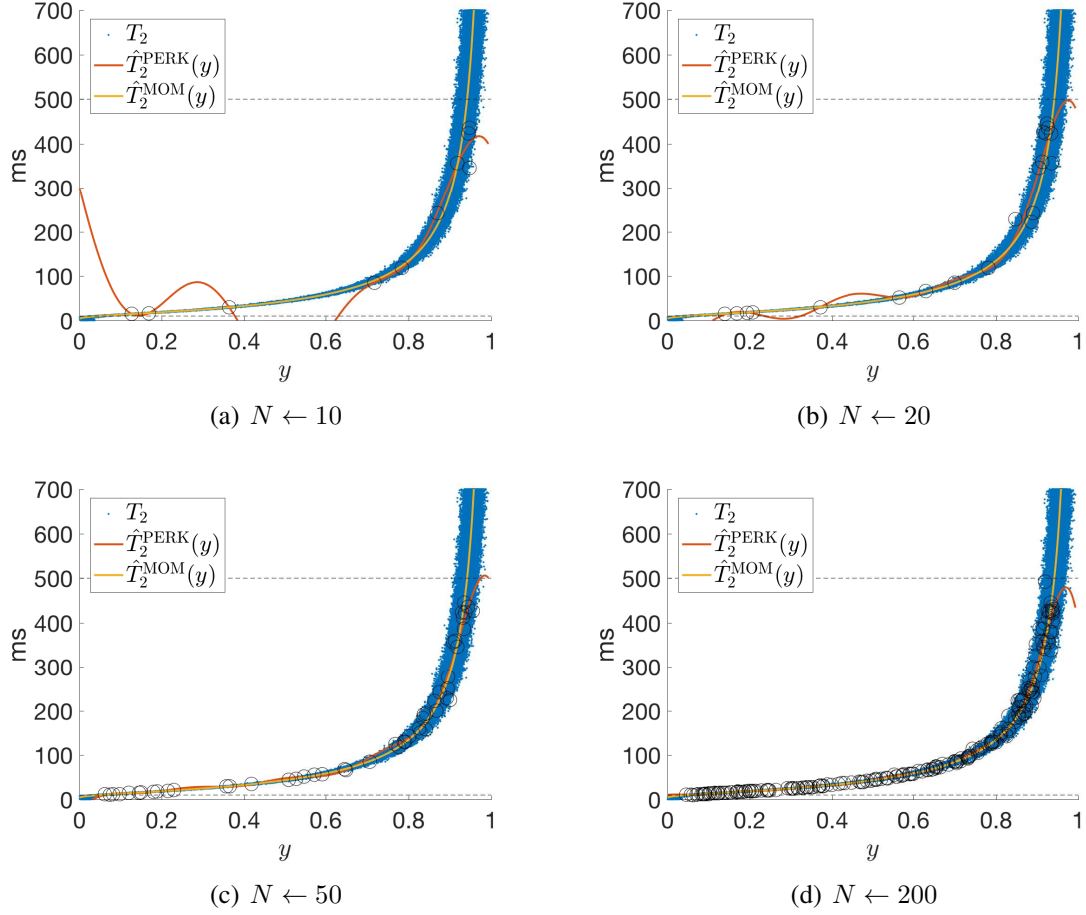
Figure 5.1: Illustrations of PERK for T_2 estimation from a single spin echo measurement. Subfigures vary the number N of PERK training points, marked with black circles. The orange and yellow curves plot PERK $\widehat{T}_2^{\mathrm{PERK}}$ and MOM $\widehat{T}_2^{\mathrm{MOM}}$ estimators evaluated at test points, marked with blue dots. Dashed black lines denote the sampling distribution support $\mathrm{supp}(\mathsf{p}_{T_2})$ 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 $\mathrm{supp}(\mathsf{p}_{T_2})$.

{fig,toy}

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 $\widehat{T}_2^{\mathrm{MOM}}(\cdot) := -T_{\mathrm{E}}/\log|\cdot|$ (yellow curves). As N increases, $\widehat{T}_2^{\mathrm{PERK}}$ appears more similar to $\widehat{T}_2^{\mathrm{MOM}}$ within well-sampled regions of $\mathrm{supp}(\mathsf{p}_{T_2})$ (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 [1] for external validation and because it is easier to validate PERK estimates against gold-standard grid search estimates in problems involving few unknowns (Ch. **??** 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 [5] consisting of two spoiled gradient-recalled echo (SPGR) [16] and one dual-echo steady-state (DESS) [19] 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 [5], wherein repetition times and flip angles were optimized for precise T_1 and T_2 estimation in cerebral tissue at 3T field strength [5] and echo times were fixed across scans. We used standard magnitude[2] 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

---

[2]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 [5], 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 [5] 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 $\mathsf{p}_{T_1} \leftarrow \mathrm{logunif}(400, 2000)\mathrm{ms}$, $\mathsf{p}_{T_2} \leftarrow \mathrm{logunif}(40, 200)\mathrm{ms}$ 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 [5] optimized the acquisition. We assumed noise covariance $\mathbf{\Sigma}$ of form $\sigma^2 \mathbf{I}_4$ (as in [5]) 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 $\boldsymbol{\nu}, s\ Z \leftarrow 10^3$ times to construct approximate feature map $\tilde{\mathbf{z}}$. For each lth latent parameter where $l \in \{1, \ldots, L\}$, 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

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

where $t \in \{1, \ldots, T\}$ 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 $\mathbf{W}$.

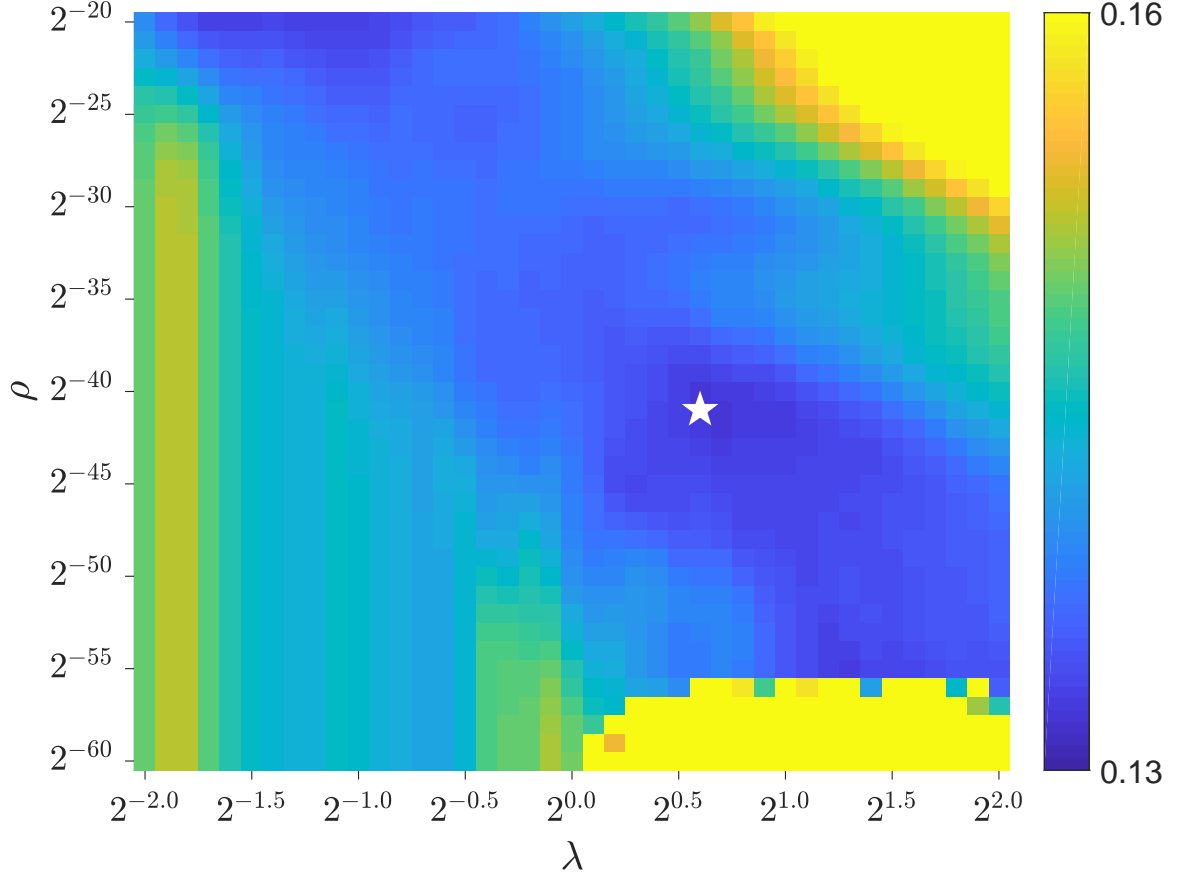Fig. 5.2 plots $\Psi(\lambda, \rho)$ for $T \leftarrow 10^5$ test points and $\mathbf{W} \leftarrow \mathrm{diag}\left([0, 0.5, 0.5]^{\mathsf{T}}\right)$ selected

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 $p_{\mathbf{x},\boldsymbol{\nu}}$ and the weighting places emphasis on good T_1, T_2 estimation performance. A white star marks the minimizer $\left(\widehat{\lambda}, \widehat{\rho}\right) \leftarrow (2^{0.6}, 2^{-41})$.

{fig,holdout}

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) [22], a popular technique that has been used in many QMRI algorithms and applications (see *e.g.* [138, 105, 106, 92, 99, 109, 108, 111, 113, 139, 5]). Following [5], we clustered flip angle scaling map voxels into 20 clusters via k-means++ [90] 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) [20]. 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 [34] yet accelerate practical performance. We initialized PGPM via conventional method-of-moments estimators of m_0, T_1 from 2 SPGR scans [32] and T_2 from 1 DESS scan [19] (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

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

was satisfied, where $\mathbf{X}$ collects latent parameter voxels in its columns, $(\cdot)^{(i)}$ denotes the ith iterate, $\mathbf{\Omega} := \mathrm{diag}\big(\mathrm{med}\big(\mathbf{X}^{(0)}\big)\big)$ 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 [28] to the 81st slice of the BrainWeb digital phantom [31] 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).
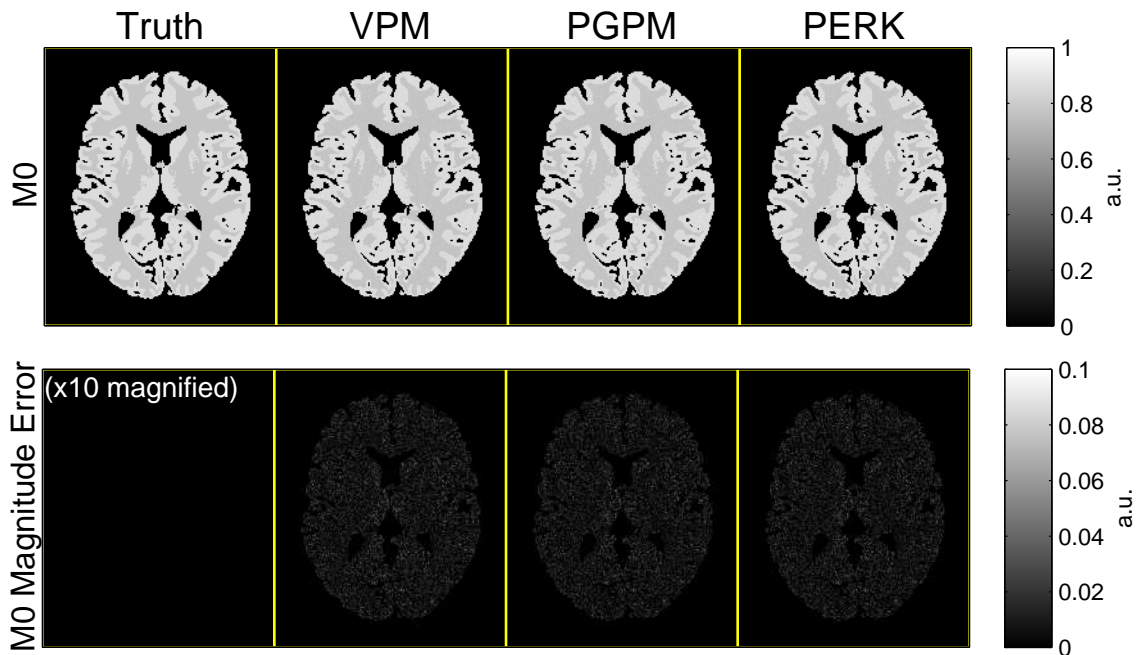
Figure 5.3: m_0 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three m_0 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics.

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

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 [2]. 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
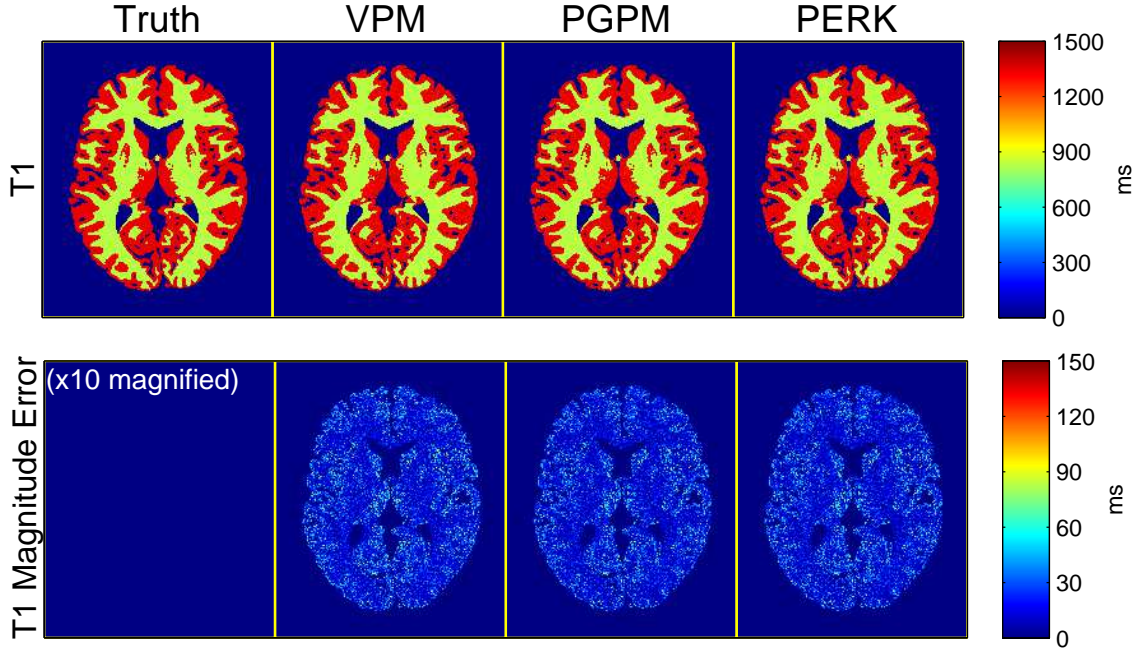
Figure 5.4: T_1 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three T_1 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics.

{fig,sim,t1}

time; and two Bloch-Siegert (BS) scans [87] (for separate flip angle scaling estimation). Nominal flip angles were achieved by scaling a 2cm slab-selective Shinnar-Le Roux RF pulse [86] 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 [5].

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 [89] 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 [88] 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
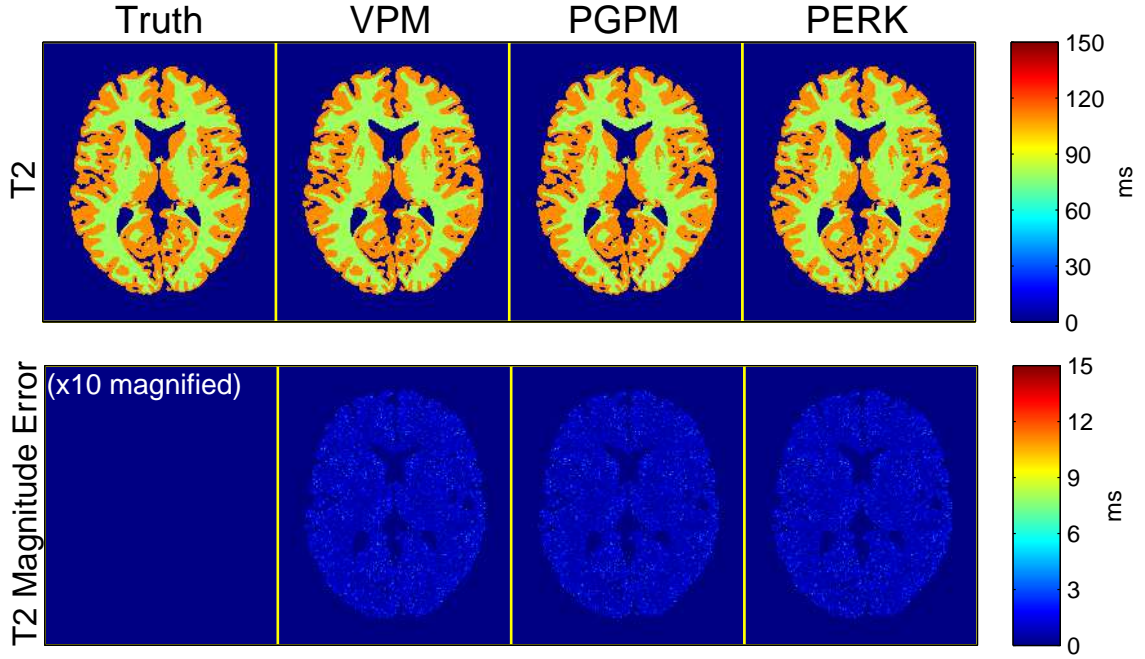
Figure 5.5: T_2 VPM, PGPM, and PERK estimates and corresponding error images, in simulation. Magnitude error images are $10\times$ magnified. Voxels not assigned WM- or GM-like relaxation times are masked out in post-processing for display. Difference images demonstrate that all three T_1 estimates exhibit low estimation error. Table 5.1 presents corresponding sample statistics.

{fig,sim,t2}

support $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ (as measured by NIST NMR reference measurements [1]) 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 [5]. 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) [1]. 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 ??, expanding $\mathsf{supp}(\mathsf{p}_{\mathbf{x},\boldsymbol{\nu}})$ well beyond the acquisition design parameter range of
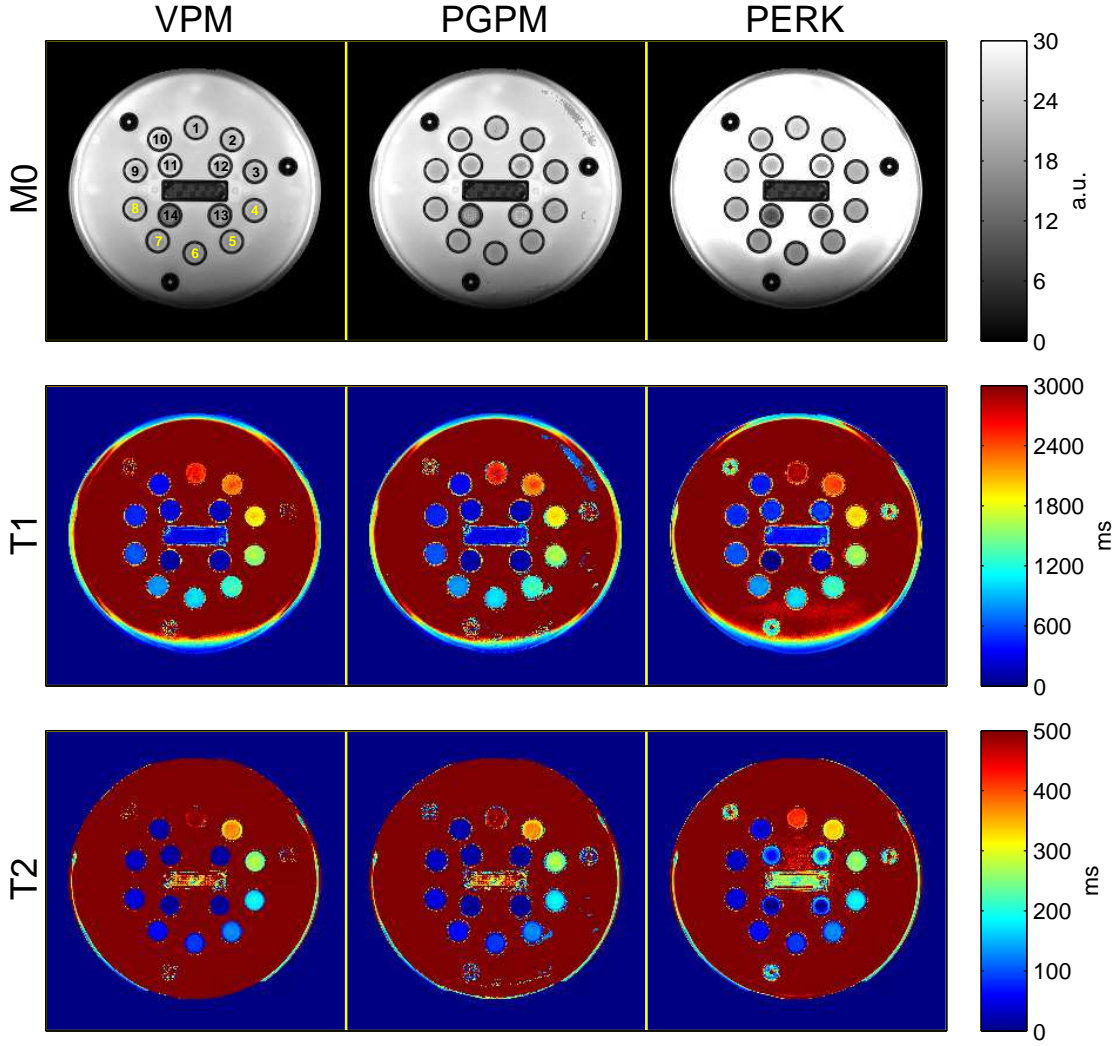
Figure 5.6: VPM, PGPM, and PERK m_0, T_1, T_2 estimates in a quantitative phantom. Vials are enumerated and highlighted to correspond with markers and colored boxes in Fig. 5.7. PERK has only been trained to accurately estimate within vials 4-8; within these vials, VPM, PGPM, and PERK estimates appear 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 [5]).
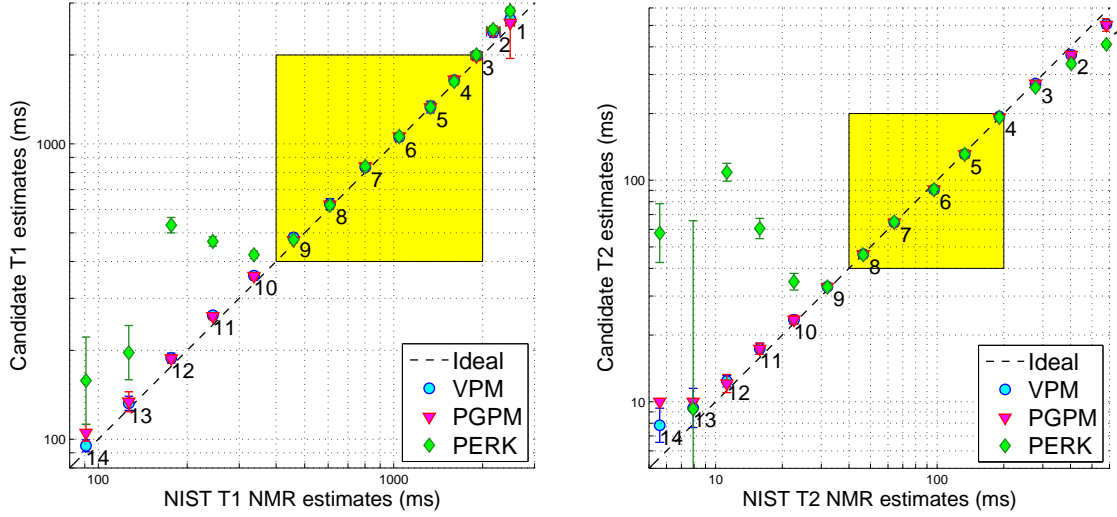
	NMR	VPM	PGPM	PERK
V4 T_1	1604 ± 7.2	1645 ± 48	1649 ± 48	1626 ± 46
V5 T_1	1332 ± 0.8	1335 ± 61	1331 ± 41	$1332 \pm 40.$
V6 T_1	1044 ± 3.2	1055 ± 28	$1060. \pm 29$	1061 ± 29
V7 T_1	801.7 ± 1.70	834 ± 21	$840. \pm 23$	839 ± 23
V8 T_1	608.6 ± 1.03	627 ± 25	623 ± 12	$620. \pm 13$
V4 T_2	190.94 ± 0.011	194 ± 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 [1]. 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 $\text{supp}(\mathsf{p}_{\mathbf{x},\nu})$. Missing markers lie outside axis limits. Corresponding tables replicate sample means $\pm$ sample standard deviations for vials within $\text{supp}(\mathsf{p}_{\mathbf{x},\nu})$. Each value is rounded off to the highest place value of its (unreported) standard error, computed via formulas in [2]. 'V#' indicates vial numbers. All values are reported in milliseconds. Within $\text{supp}(\mathsf{p}_{\mathbf{x},\nu})$, VPM, PGPM, and PERK estimates agree excellently with each other and reasonably with NMR measurements.

{fig,hpd-tight,plot}

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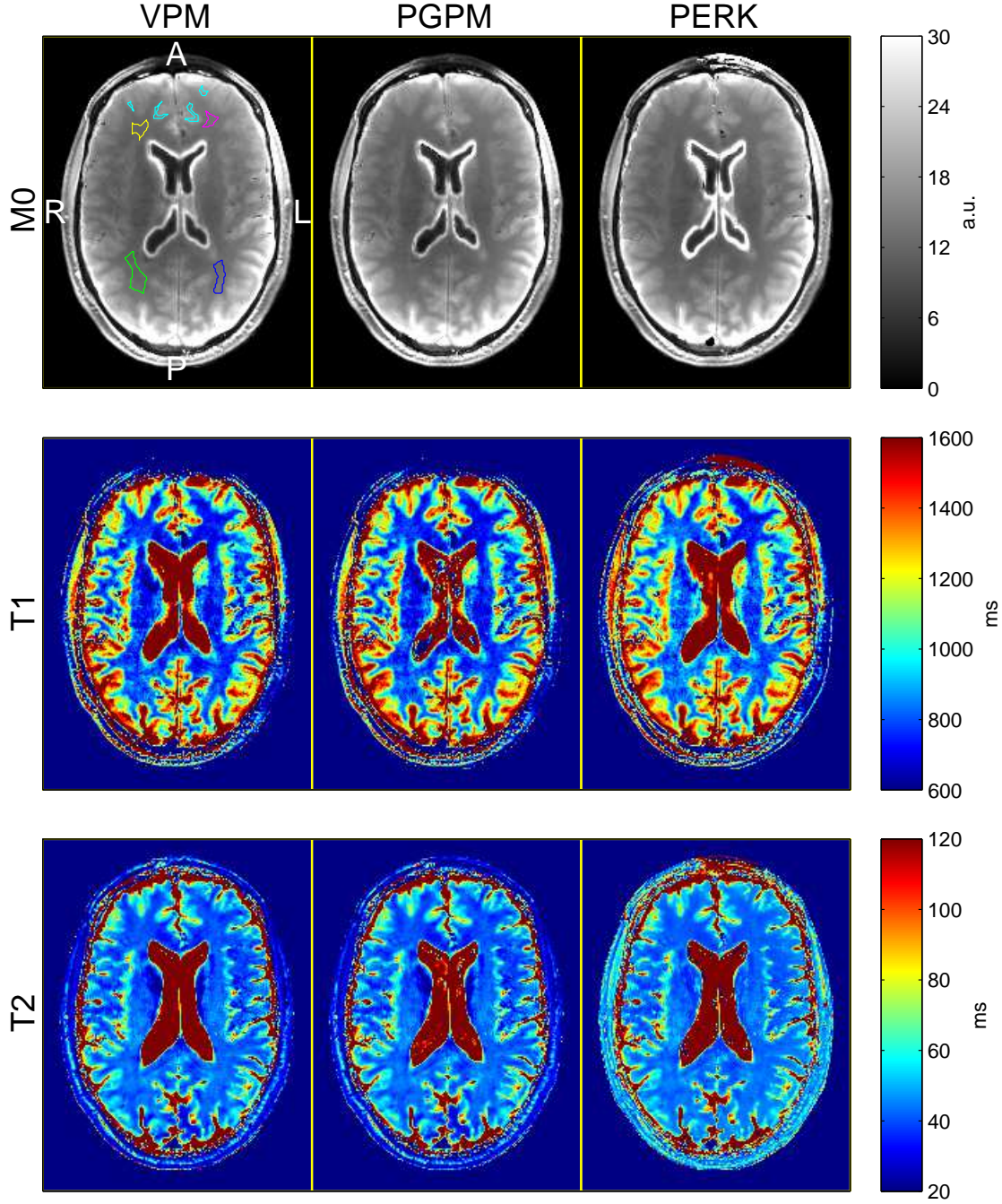
Figure 5.8: VPM, PGPM and PERK estimates of m_0, T_1, T_2 in the brain of a healthy volunteer. Separate WM ROIs are distinguished by anterior/posterior (A/P) and right/left (R/L) directions. Four small anterior cortical GM polygons are pooled into a single GM ROI. Images are cropped in post-processing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

{fig,brain}

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
	AR WM	778 ± 28	779 ± 27	832 ± 31
	AL WM	731 ± 37	713 ± 33	725 ± 41
T_1	PR WM	805 ± 52	796 ± 51	831 ± 51
	PL WM	789 ± 40	788 ± 38	815 ± 42
	A GM	1120 ± 180	1120 ± 180	$1150 \pm 170.$
	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
T_2	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 [2]. All values are in milliseconds.

{tab,brain}

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.* [28, 29]).

5.8   Robustness Studies

{s,perk,robust}

5.9   Discussion

{s,perk,disc}

5.10   Conclusion

{s,perk,conc}

APPENDIX A

Coil Data Combination from Multiple Datasets

{a,cc-multi}

In the completed thesis, this appendix will describe an unpublished algorithm for combining multiple MRI datasets when each dataset is acquired using multiple receiver coils with fixed coil geometry. Such sequences of coil datasets arise naturally in many quantitative MRI applications, including the ones studied in Chapters 4-**??**. The algorithm extends [89] (which considers joint image reconstruction and coil sensitivity estimation in the case of a single dataset) to exploit the fixed coil sensitivity across several datasets, thereby improving estimation conditionality as compared to separate coil combination across datasets.

APPENDIX B

DESS in the Presence of Diffusion

{a,dess-diff}

In the completed thesis, this appendix will present an unpublished analysis of model mismatch in single-compartment DESS signal models (2.41) and (2.43) in the presence of diffusion. We first develop models that describe the single-compartment DESS signal in the presence of diffusion. We then use these models to show through phantom experiments that neglecting diffusive effects during T_2 estimation (as in Chapters 3-4) induces significant estimation bias. We conclude with recommendations for MR acquisition settings (specifically, for dephasing gradient area) that reduce diffusion-related estimation bias without excessively imparting other bias. These recommendations were used to guide DESS acquisition design for all relevant experiments considered in the main body of this thesis.

BIBLIOGRAPHY

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] A. Macovski, *Medical imaging systems*. New Jersey: Prentice-Hall, 1983.

[10] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, *Magnetic resonance imaging: Physical principles and sequence design*. New York: Wiley, 1999.

[11] D. G. Nishimura, "Principles of magnetic resonance imaging," 1996. Unpublished textbook.

[12] F. Bloch, "Nuclear induction," *Phys. Rev.*, vol. 70, pp. 460–74, Oct. 1946.

[13] W. S. Hinshaw, "Image formation by nuclear magnetic resonance: The sensitive point method," *J. Appl. Phys.*, vol. 47, p. 3709, Aug. 1976.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] T. W. Redpath and R. A. Jones, "FADE-A new fast imaging sequence," *Mag. Res. Med.*, vol. 6, pp. 224–34, Feb. 1988.

[19] 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.

[20] 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.

[21] C. Byrne, "A unified treatment of some iterative algorithms in signal processing and image reconstruction," *Inverse Prob.*, vol. 20, pp. 103–20, Feb. 2004.

[22] 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.

[23] 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.

[24] 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.

[25] A. Macovski, "Noise in MRI," *Mag. Res. Med.*, vol. 36, pp. 494–7, Sept. 1996.

[26] T. Lei, "Statistics of MR signals: revisited," in *Proc. SPIE 6510 Medical Imaging: Phys. Med. Im.*, p. 651052, 2007.

[27] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Mag. Res. Med.*, vol. 34, pp. 910–4, Dec. 1995.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Cont. Opt.*, vol. 20, no. 2, pp. 221–46, 1982.

[35] J. A. Fessler, "Michigan image reconstruction toolbox (MIRT) for Matlab," 2016. Available from web.eecs.umich.edu/ fessler.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] H. Wang and Y. Cao, "Spatially regularized T1 estimation from variable flip angles MRI," *Med. Phys.*, vol. 39, pp. 4139–48, July 2012.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] R. Heule, C. Ganter, and O. Bieri, "Triple echo steady-state (TESS) relaxometry," *Mag. Res. Med.*, vol. 71, pp. 230–7, Jan. 2014.

[71] M. L. Gyngell, "The steady-state signals in short-repetition-time sequences," *J. Mag. Res.*, vol. 81, pp. 474–83, Feb. 1989.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Philosophical Society*, vol. 22, pp. 700–25, July 1925.

[84] H. Cramér, *Mathematical methods of statistics*. Princeton: Princeton Univ. Press, 1946.

[85] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, vol. 24, pp. 586–602, Dec. 1953.

[86] 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.

[87] 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.

[88] 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.

[89] L. Ying and J. Sheng, "Joint image reconstruction and sensitivity estimation in SENSE (JSENSE)," *Mag. Res. Med.*, vol. 57, pp. 1196–1202, June 2007.

[90] 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.

[91] 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.

[92] 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.

[93] 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.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

[98] 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.

[99] 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.

[100] B. Zhao, J. Haldar, K. Setsompop, and L. L. Wald, "Optimal experiment design for magnetic resonance fingerprinting," in *embc*, pp. 453–6, 2016.

[101] H. C. Torrey, "Bloch equations with diffusion terms," *Phys. Rev.*, vol. 104, pp. 563–5, 1956.

[102] H. M. McConnell, "Reaction rates by nuclear magnetic resonance," *J. of Chemical Phys.*, vol. 28, pp. 430–31, Mar. 1958.

[103] N. Bloembergen, E. M. Purcell, and R. V. Pound, "Relaxation effects in nuclear magnetic resonance absorption," *Phys. Rev.*, vol. 73, pp. 679–712, Apr. 1948.

[104] 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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.*, 2017.

[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.*, 2017.

[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.

[117] 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.

[118] 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.

[119] G. R. Morrell, "A phase-sensitive method of flip angle mapping," *Mag. Res. Med.*, vol. 60, pp. 889–94, Oct. 2008.

[120] 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.

[121] 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 multicomponent driven equilibrium single pulse observation of T1 and T2," *Mag. Res. Med.*, vol. 70, pp. 147–54, July 2013.

[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.

[123] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, May 1950.

[124] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–97, Sept. 1995.

[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.

[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.

[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.

[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.

[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.

[130] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, Feb. 1970.

[131] I. Steinwart and A. Christmann, *Support vector machines*. Springer, 2008.

[132] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *NIPS*, 2007.

[133] D. J. Sutherland and J. Schneider, "On the error of random Fourier features," in *Proc. Intl. Conf. on Uncertainty in AI*, 2015.

[134] Z. Wu, "Generalized Bochner's theorem for radial function," *Approximation Theory and its Applications*, vol. 13, no. 3, pp. 47–57, 1997.

[135] M. A. Woodbury, "Inverting modified matrices," 1950. Tech. Report 42, Stat. Res. Group, Princeton Univ.

[136] V. N. Vapnik, *Statistical learning theory*. Wiley-Interscience, 1998.

[137] M. M. Siddiqui, "Statistical inference for Rayleigh distributions," *RADIO SCIENCE Journal of Research NBS/USNC-URSI*, vol. 68D, pp. 1005–10, Sept. 1964.

[138] J. Gong and J. P. Hornak, "A fast T1 algorithm," *Mag. Res. Im.*, vol. 10, no. 4, pp. 623–6, 1992.

[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.