Title
Optimal Control for Learning with Applications in Dynamic MRI

Author
Maidens, John Norman

Publication Date
2017

Peer reviewed|Thesis/dissertation

Optimal Control for Learning with Applications in Dynamic MRI

by

John Norman Maidens

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Murat Arcak, Chair
Professor Peder E. Z. Larson
Professor Ali Mesbah

Summer 2017

Optimal Control for Learning with Applications in Dynamic MRI

Abstract

Optimal Control for Learning with Applications in Dynamic MRI

by

John Norman Maidens

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Murat Arcak, Chair

Unknown parameters in models of dynamical systems can be learned reliably only when the system is excited such that the measured output data is informative. Given a statistical model that specifies the dependence of the measured data on the state of the dynamical system, the design of maximally informative inputs to the system can be formulated as a mathematical optimization problem using the Fisher information as an objective function.

Optimal design in the time domain is hard in general, but efficient approximation algorithms have been developed in some special cases. In this dissertation, we present new approaches to solving this problem using optimal control algorithms based on convex relaxations, and exploiting geometric structure in the underlying optimization problem.

Magnetic resonance imaging (MRI) serves as a motivating application problem throughout. We highlight two successes of these methods in the design of dynamic MRI experiments: magnetic resonance fingerprinting (MRF) for accelerated anatomic imaging, and hyperpolarized carbon-13 MRI for noninvasively monitoring cancer metabolism. In particular, we use optimal experiment design algorithms to compute optimized flip angle sequences for MRF and hyperpolarized carbon-13 acquisitions as well as optimized tracer injection inputs for estimating metabolic rate parameters in hyperpolarized carbon-13 acquisitions. In the final chapter, we present results on constrained reconstruction of metabolism maps from experimental data, closing the path from experiment design to data collection to synthesis of interpretable information.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to acknowledge my PhD advisor, Murat Arcak, for giving me unrivalled freedom to follow my interests, and for the trust he has put in me throughout my PhD. I thank Murat for the encouragement to pursue new interests and for allowing me to shape my own research direction as I dabbled in statistics, machine learning, and medical imaging. In Murat's mind his students are capable of the very best, and this pushed me to do justice to the faith he put in me.

Special thanks to Peder Larson from UCSF, for his close research collaboration, for investing his time and energy to teach me an entirely new field, and for his patience with me as I learned. And thank you to Jeremy Gordon for his endless generosity with his time and expertise, which helped me immeasurably as I learned my way. Also to Prof. Miki Lustig for the spark that originally got me interested in MRI, and to Christine Leon Swisher, who first introduced me to hyperpolarized carbon 13 and kick started my collaboration with USCF.

Thank you to Ali Mesbah for serving on both my qualifying exam and dissertation committees. His expertise and feedback during my qualifying exam helped to shape my research direction, and will no doubt also help shape this dissertation. Thanks also to Axel Barrau and Silvère Bonnabel, for their collaboration on the symmetry reduction project, and to Andy Packard, Chris Meissen, and Octavio Narváez-Aroche whose discussions and suggestions helped shape my research toward the end of my PhD.

Thank you to everyone in the Arcak group: Marcella Gomez, Andras Gyorgy, Eric Kim, Mindy Perkins, Stan Smith, Justin Hsia, Ana Rufino Ferreira, Sam Coogan, Jon Tu, and Yusef Shafi for your camaraderie in the many group meetings, reading groups, conferences, and research discussions that shaped my education. Further thanks to Michael Li, Meeko Oishi, Ian Michell and Guy Dumont, who inspired me to love research, and whose mentorship and guidance are the reason I ever made it to the PhD program in the first place.

Finally, thank you to my parents at home in Edmonton whose love, care, and patience have been the foundation of my education. And to Jesse and Fiona, my loving and supportive family here in Berkeley.

Chapter 1

Introduction

In this dissertation, we consider the problem of estimating the value of unknown parameters in a mathematical model of a dynamical system from noisy output data. For such a system, the reliability of the parameter estimates depends on the choice of input used to excite the system, as some inputs provide much greater information about the parameters than others. This problem is known as optimal experiment design and much work has been done on this problem in the last 50 years [35, 39, 58, 105].

Historically, a great deal of work on optimal experiment design has taken a frequency domain approach, where the input to the system is designed based on its power spectrum. Here, we will approach this problem in the time domain, hoping to be able to perform experiment design for systems with nonlinear dynamics, and to put time domain constraints on the admissible inputs.

My interest in optimal design stems from applied problems in medical imaging, in particular to solving quantitative imaging problems in magnetic resonance imaging (MRI). Data collected in MRI experiments is typically noisy due to thermal movement of electrons in the receiver coil and the object being imaged. This makes it challenging to estimate model parameters from dynamic data sets when the signal to noise ratio is small. This challenge can be addressed by designing experimental parameters with the goal of maximizing the information about unknown model parameters contained in the data collected.

After introducing the problem of optimal experiment design in Section 1.1, we will introduce two such MRI problems. In Section 1.2 we intoduce hyperpolarized carbon-13 MRI, an emerging imaging technology that enables *in vivo* metabolism monitoring with chemical specificity. Then in Section 1.3 we introduce magnetic resonance fingerprinting, a recently-developed method for quantitative anatomic imaging. In Section 1.4, we summarize the contributions of this dissertation to give the reader a high-level overview of what is to come in the subsequent chapters.

1.1 Optimal Design of Dynamic Experiments

Problem Description

We consider a discrete-time dynamical system with noisy observations

$$\begin{aligned} x_{t+1} &= f(t, x_t, u_t, \theta) \\ Y_t &\sim P_{x_t} \end{aligned} \tag{1.1}$$

where $x_t \in \mathbb{R}^n$ denotes the system's state, $u_t \in \mathbb{R}^m$ is a sequence of inputs to be designed and $\theta \in \mathbb{R}^p$ is a vector of unknown parameters that we wish to estimate. Observations are drawn independently from a known distribution that is parametrized by the system state x_t. We assume that for all $x_t \in \mathbb{R}^n$ the probability distribution P_{x_t} is absolutely continuous with respect to some measure μ and we denote its density with respect to μ by $p_{x_t}(y_t)$. We consider this system over a finite horizon $0 \leq t \leq N$. Our goal is to design a sequence u that provides a maximal amount of information about the unknown parameter vector θ. This problem can be addressed by maximizing the Fisher information about θ.

Fisher Information

An important notion in frequentist statistics is the Fisher information matrix for the vector of model parameters θ. The Fisher information is fundamental in the analysis of numerous statistical estimators from unbiased estimation to maximum-likelihood estimation. We begin with a definition.

Definition 1. *Let $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ be a family of probability distributions parametrized by θ in an open set $\Omega \subseteq \mathbb{R}^p$ and dominated by some measure μ. Denote the probability densities with respect to μ by p_θ and assume that the densities are differentiable with respect to θ. We define the Fisher information matrix as the $p \times p$ matrix $\mathcal{I}(\theta)$ with (i, j)-th entry defined as*

$$\mathcal{I}(\theta)_{i,j} = \mathbb{E}\left[\frac{\partial \log p_\theta(Y)}{\partial \theta_i} \frac{\partial \log p_\theta(Y)}{\partial \theta_j}\right]$$

where $Y \sim P_\theta$.

The Cramér-Rao Inequality for Unbiased Estimators

The Cramér-Rao inequality provides a lower bound on the achievable covariance of any unbiased estimator of θ in terms of the Fisher information matrix.

Theorem 1 (Cramér-Rao Bound)**.** *Let $\delta(Y)$ be any unbiased estimator of the parameter θ. Under the conditions that*

- *$\mathcal{I}(\theta)$ exists and is nonsingular*

- $\frac{\partial}{\partial\theta}\left[\int \delta(y)p_\theta(y)dy\right] = \int \delta(y)\left[\frac{\partial}{\partial\theta}p_\theta(y)\right]dy$

the estimator δ satisfies

$$cov(\delta(Y)) \geq \mathcal{I}(\theta)^{-1}$$

where $A \geq B$ indicates that $A - B$ is positive semidefinite.

Thus, as the Fisher information contained in a data set increases, it becomes possible to find more reliable estimates of the parameter vector θ. The following result shows that asymptotically, this bound is tight.

Asymptotic distribution of the Maximum-Likelihood Estimator

Let $Y^1, \ldots Y^n$ be n random variables drawn independently from the probability distribution P_θ. Given that some mild regularity conditions are satisfied, as $n \to \infty$ the maximum-likelihood estimator

$$\hat{\theta}_{MLE} = \arg\max_\theta \sum_{i=1}^{n} \log p_\theta(Y^i).$$

converges in distribution to a normal distribution with covariance $\mathcal{I}(\theta)^{-1}$.

Theorem 2 (Asymptotics of the MLE)**.** *Suppose that*

- *the MLE is consistent*
- *$p_\theta(Y)$ is bounded and C^2 in θ*
- *$\mathcal{I}(\theta)$ exists and is nonsingular*
- *$\mathbb{E}[\|\log H_\theta(Y)\|] < \infty$ where H denotes the Hessian of $p_\theta(Y)$ with respect to θ.*

Then $\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, \mathcal{I}(\theta)^{-1})$.

Scalar Measures of Positive Semidefinite Matrices

In order to use the Fisher information as an objective function for choosing an input sequence u, we must choose a method of summarizing $\mathcal{I}(\theta)$ by a scalar quantity. We survey some of the most common choices here.

- **Nonnegative linear functions**

  The most convenient class of objective functions are those that are linear functions of the information matrix. For any positive semidefinite matrix $K \in \mathbb{R}^{p\times p}$ we can define the objective function

  $$\phi(\mathcal{I}) = \mathrm{tr}(K\mathcal{I}).$$

Particular cases are the "T-optimal" design criterion (when K is the identity matrix) and the c-optimal design criterion (when $K = cc^T$ has rank 1), which is optimal for the scalar parameter $c^T\theta$.

- **Other convex functions**

  A number of other functions of the Fisher information matrix are also commonly used as scalar measures of its size. Choices include:

  - $\phi_D(\mathcal{I}) = \log\det(\mathcal{I})$ know as "D-optimal design,"
  - $\phi_E(\mathcal{I}) = \lambda_{\min}(\mathcal{I})$ known as "E-optimal design", and
  - $\phi_A(\mathcal{I}) = 1/\operatorname{tr}(\mathcal{I}^{-1})$ known as "A-optimal design."

To understand the geometric meaning of each of these optimal design criteria, consider an asymptotic confidence region associated with the MLE of θ. Such a region is described by the ellipsoid $\mathcal{E} = \{\tilde{\theta} : n(\tilde{\theta} - \hat{\theta}_{MLE})^T\mathcal{I}(\tilde{\theta} - \hat{\theta}_{MLE}) \leq \epsilon\}$ for some $\epsilon > 0$. The volume of this ellipsoid is proportional to $\sqrt{\det(\mathcal{I}^{-1})}$ and therefore the D-optimal design criterion corresponds to minimizing the volume of the confidence ellipsoid $\mathcal{E}$. If the spectrum of $\mathcal{I}$ is denoted $\lambda_1, \ldots, \lambda_p$ then the length of the axes of $\mathcal{E}$ are $\frac{1}{\sqrt{\lambda_1}}, \ldots, \frac{1}{\sqrt{\lambda_p}}$. So the E-optimal design criterion corresponds to minimizing the length of the longest axis of the confidence ellipsoid and the A-optimal design criterion corresponds to minimizing the average squared length of the confidence ellipsoid axes.

Optimizing Under Parametric Uncertainty

Given a particular scalar-valued function ϕ with which we will measure the size of the Fisher information, the optimal experiment design problem is given by

$$\begin{array}{ll} \textbf{maximize}_u & \phi(\mathcal{I}(\theta)) \\ \textbf{subject to} & u \in \mathcal{U} \end{array} \tag{1.2}$$

for some set $\mathcal{U}$ of admissible signals. In general the objective $\phi(\mathcal{I}(\theta))$ depends on the true value of the parameter $\theta \in \Omega$, which is unknown to the experimenter. Several approaches are commonly taken to address this problem.

The simplest approach, and the one that we will use, is to assume a nominal value $\theta_0 \in \Omega$ for the model parameters and to solve the problem

$$\begin{array}{ll} \textbf{maximize}_u & \phi(\mathcal{I}(\theta_0)) \\ \textbf{subject to} & u \in \mathcal{U} \end{array} \tag{1.3}$$

about this nominal value. This is referred to by some authors as "local design" [105], though we will refer to it as "nominal design" and reserve the terms local and global to refer to local and global solutions of the optimization problem. If multiple experiments can be performed,

this approach can also be extended to an iterative procedure where we start with some nominal value of the parameter vector and design an optimal experiment using this value. Data are then collected, a new value of the parameter vector is estimated from the data, and then the experiment design is performed again using the updated value of the parameter vector.

Another approach is to consider a minimax criterion, solving the worst-case optimization problem

$$\begin{array}{ll} \textbf{maximize}_u & \min\limits_{\theta\in\Omega} \phi(\mathcal{I}(\theta)) \\ \textbf{subject to} & u \in \mathcal{U}. \end{array} \tag{1.4}$$

This approach ensures a minimum level of estimator performance regardless of the true value of θ.

Another alternative is to consider a prior distribution $\pi(\theta)$ on the values that the parameter vector may take. The experiment may then be designed with respect to the average criterion

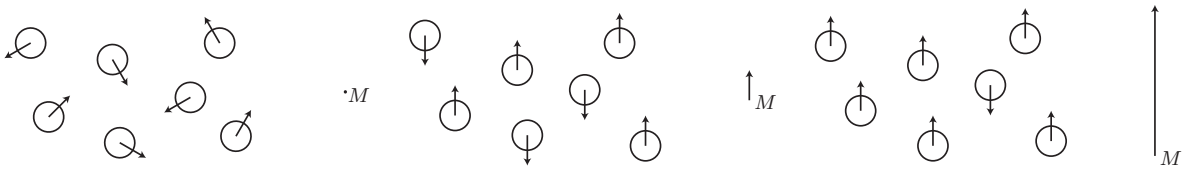
$$\begin{array}{ll} \textbf{maximize}_u & \mathbb{E}_\pi[\phi(\mathcal{I}(\theta))] \\ \textbf{subject to} & u \in \mathcal{U}. \end{array} \tag{1.5}$$

1.2 Hyperpolarized Carbon-13 MRI for Imaging Metabolism

Carbon is arguably the most important element in biochemistry. It forms the basis of all organic molecules that make up the human body, yet only recently have we begun to be able to quickly image carbon *in vivo* using magnetic resonance imaging (MRI). The emerging technology that makes this possible is known as hyperpolarized carbon-13 MRI, and it has enabled *in vivo* imaging with spatial, temporal and chemical specificity for the first time. This development is leading to new insights into the spatial distribution of metabolic activity through the analysis of dynamic image sequences.

The processes that are imaged in hyperpolarized carbon-13 MRI are inherently dynamic, resulting from blood flow, tissue perfusion, metabolic conversion, and polarization decay. Thus there is an opportunity for control researchers to improve the dynamic models, excitation inputs and estimation algorithms used in hyperpolarized carbon-13 MRI.

The measurable signal in MRI arises from radio-frequency electromagnetic waves generated by oscillating atomic nuclei. Nuclei containing an odd number of protons and/or neutrons possess a nuclear spin angular momentum, each giving rise to a small magnetic moment. Thus nuclei such as carbon (^{12}C) and oxygen (^{16}O) are invisible to MRI, while hydrogen (^{1}H) and the carbon-13 isotope (^{13}C) exhibit magnetic resonance (MR). Hydrogen MR, sometimes known as proton MR, is currently the most commonly-used in clinical settings due to the high abundance of hydrogen atoms in the human body (largely in the form of



(a) Spins in the absence of an external magnetic field are oriented randomly, leading to a net magnetization $M = 0$.

(b) In a magnetic field, spins orient themselves with the field, leading to a small net magnetization M.

(c) Hyperpolarizing the sample causes the spins to coordinate direction, leading to a greater net magnetization M.

Figure 1.1: Polarization of a collection of spins leads to a net magnetization M. The magnitude of the vector M can be increased by hyperpolarization.

H_2O) and its high sensitivity [80]. Conventional hydrogen MRI is pervasive for noninvasive imaging of anatomic structure, but provides little functional information. In this work, we focus on carbon-13 MR, which can be used to provide information about metabolic function.

Polarization

Each carbon-13 nucleus in a sample gives rise to a small magnetic moment, which we think of as a vector in three-dimensional space. In the absence of a magnetic field these magnetic moments, or spins, are oriented randomly in space. In this state, where the spins are oriented randomly, the net magnetization $M = (M_x, M_y, M_z)$ of the sample (computed by summing the magnetic moments from all nuclei in the sample) is zero (Figure 1.1a).

In the presence of a magnetic field, the spins will orient themselves to the magnetic field with some aligned parallel (n^+) with the magnetic field lines and some aligned anti-parallel (n^-) with the magnetic field lines. Due to a small energy difference between the two states, there is a slight bias toward the lower energy state aligned parallel to the magnetic field. This bias leads to a polarization of the sample, defined as the excess of spins in the lower energy state. At thermal equilibrium, the polarization P is given in terms of the applied magnetic field strength B_0, ambient temperature T, gyromagnetic ratio γ of the nucleus (10.705 MHz/T for ^{13}C), and the Boltzmann (k_B) and reduced Planck ($\hbar$) constants as

$$P = \frac{n^+ - n^-}{n^+ + n^-} = \tanh\left(\frac{\gamma\hbar B_0}{2k_B T}\right).$$

This slight bias leads to a small net magnetization of the sample at thermal equilibrium (Figure 1.1b) given by

$$M = \begin{bmatrix} 0 \\ 0 \\ M_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{N\gamma^2\hbar^2 I_z(I_z+1)B_0}{3k_B T} \end{bmatrix}$$

where N is the number of spins in the sample and I_z is the spin operator. Note that we use the standard convention that the z axis is chosen such that it points in the direction of the applied B_0 field.

Due to the low natural abundance of carbon-13 in the body and its low gyromagnetic ratio, the thermal equilibrium magnetization is insufficient to achieve sufficient signal for imaging. Thus carbon-13-based imaging relies on hyperpolarization technology to increase the polarization beyond the equilibrium level (Figure 1.1c).

Chemical shift

The unique aspect of hyperpolarized carbon-13 MRI, when compared to competing metabolic imaging technologies such as positron emission tomography (PET), is that it is the only technique that provides chemical specificity. It is possible to infer chemical information from MRI data due to a phenomenon known as chemical shift.

Chemical shift results in a small change in the resonant frequency of spins. This change is caused by shielding of the nuclei from the main magnetic field B_0 due to nearby electron orbitals [80]. The resulting frequency shift can be exploited to selectively excite specific metabolites [56], or distinguish between metabolites produced. This gives hyperpolarized carbon-13 MRI the unique ability to quantify metabolic flux in specific pathways.

Hyperpolarization using DNP

Hyperpolarized carbon-13 MRI has been enabled by new technologies for hyperpolarizing carbon-13-containing substrates in liquid state, leading to a greater than 10000× increase in signal-to-noise ratio (SNR) when imaging carbon-13. This technology relies on dissolution dynamic nuclear polarization (D-DNP) to achieve significant polarization gains [4].

Dynamic nuclear polarization relies on transferring polarization to carbon-13 nuclei from electrons using microwave radiation. In this procedure, a sample is doped with a small quantity of stable electron radical. The sample is then cooled to cryogenic temperature and placed in a strong magnet. At this temperature and magnetic field strength, electrons become nearly 100% polarized. Then by irradiating the sample with microwaves, polarization is transferred from the electrons to the carbon-13 nuclei in a biochemical substrate of interest. To prepare the sample for injection and *in vivo* imaging, it is then rapidly dissolved in warm water, neutralized to a safe pH and the electron radical is removed before injection [77].

Polarization decay in hyperpolarized substrates

Upon warming and removal from the magnet, the magnetization induced by hyperpolarization begins to decay over time toward the thermal equilibrium magnetization due to a phenomenon known as T_1 relaxation. The dynamics of the magnetization vector are governed

by a system of state equations know as the rotating frame Bloch equations:

$$\frac{d}{dt}\begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} = \begin{bmatrix} -\frac{1}{T_2} & u_2 & 0 \\ -u_2 & -\frac{1}{T_2} & u_1 \\ 0 & -u_1 & \frac{1}{T_1} \end{bmatrix} \begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{M_0}{T_1} \end{bmatrix} \tag{1.6}$$

with initial condition $M(0) = (0, 0, M_z(0))$. Here, the evolution of the state M is dependent on a sequence of control inputs u_1 and u_2 corresponding to the amplitude and frequency of the applied radio-frequency (RF) electromagnetic excitation pulse (known as the B_1 field) that rotates the vector M about the origin, and T_1 and T_2 parameters that govern the relaxation time in the longitudinal (z) and transverse (x, y) directions respectively.

When the sample is hyperpolarized we have $M_z(0) \gg M_0$, therefore the contribution of the affine term in (1.6) is negligible. Thus in the absence of RF excitation, the longitudinal magnetization exhibits exponential the decay

$$M_z(t) = M_z(0)e^{-t/T_1}.$$

In addition to T_1 relaxation, magnetization also decays due to repeated RF excitation. Throughout this dissertation, in hyperpolarized carbon-13 models we will assume that the RF pulse occurs on a time scale much faster than T_1 and T_2, therefore it can be modelled as an instantaneous state reset that rotates M to some angle α away from the z axis, known as the flip angle. We also assume that a spoiled gradient echo pulse sequence [10] is used, thus between RF pulses a strong magnetic field gradient is applied to dephase the transverse magnetization ensuring that $M_x = M_y = 0$. Thus at a time t^+ immediately after an RF pulse, the magnetization is given in terms of the magnetization at time t^- immediately before the RF pulse as

$$\begin{aligned} M_z(t^+) &= \cos(\alpha)M_z(t^-) \\ M_{xy}(t^+) &:= \sqrt{M_x(t^+)^2 + M_y(t^+)^2} = \sin(\alpha)M_z(t^-). \end{aligned}$$

It now follows that at a time t following a sequence of RF pulses with flip angles $\alpha_0, \ldots, \alpha_{N-1}$ the longitudinal magnetization remaining has decayed to

$$M_z(t) = M_z(0)e^{-t/T_1}\prod_{k=0}^{N-1}\cos(\alpha_k).$$

Metabolism and Disease

Hyperpolarized carbon-13 MRI makes it possible to track the metabolism of injected substrates with spatial, temporal and chemical specificity. In practice, a commonly-used substrate is [1-^{13}C]pyruvate, which consists of pyruvate molecules labeled with a carbon-13 atom at the first carbon position (Figure 1.2).

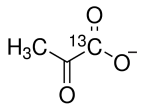


Figure 1.2: [1-^{13}C]pyruvate molecule.

Pyruvate plays an important role in cellular metabolism, as the end product of glycolysis. Following glycolysis, under anaerobic conditions pyruvate is typically converted to lactate via lactic acid fermentation, which serves as a short-term means of producing energy. Activity in this pathway can be quantified using carbon-13 MRI via observed [1-^{13}C]lactate signal. Under aerobic conditions, pyruvate can also serve as an input to the citric acid cycle, which produces cellular energy in the form of adenosine triphosphate (ATP) through cellular respiration in the mitochondria. Activity in this pathway can be quantified via pyruvate flux through the pyruvate dehydrogenase complex (PDC), which is proportional to observed ^{13}C-bicarbonate signal in carbon-13 MRI studies [74].

Pyruvate metabolism changes in cancer

Cancer results in changes to metabolism that can be used to diagnose and monitor treatment response. In particular, a phenomenon known as the Warburg Effect results in increased lactic acid fermentation in place of cellular respiration in cancerous tissues even under aerobic conditions [34]. Thus the rate of production of [1-^{13}C]lactate from injected hyperpolarized [1-^{13}C]pyruvate can be used to monitor cancer metabolism *in vivo*. The ability of hyperpolarized carbon-13 MRI to distinguish cancerous tissue from healthy tissue has been demonstrated in animal model [29] and clinical human prostate cancer [77] studies.

Pyruvate metabolism changes in heart failure

Metabolic changes in the heart have also been proposed as a significant contributing factor to congestive heart failure [99]. Hyperpolarized carbon-13 MRI studies are being used to study this hypothesis by quantifying metabolic fluxes in the heart. Hyperpolarized carbon-13 MRI has been used to demonstrate that the development of heart failure leads to decreased [1-^{13}C]pyruvate flux through the PDC in a porcine model of dilated cardiomyopathy [94]. Recent clinical studies have also demonstrated feasibility of measuring PDC flux in heart tissue using hyperpolarized [1-^{13}C]pyruvate in healthy human subjects [27].

Quantifying Metabolic Flux

Hyperpolarized carbon-13 MRI enables dynamic experiments that show metabolic activity with spatial, temporal and chemical specificity. This enables quantifying the spatial distribution of the activity of specific metabolic pathways. In this section, we discuss model-based methods of fusing this information into spatial maps of metabolic activity. This is done by

estimating kinetic parameters in a model describing the evolution of the MR signal observed in each spatial volume element (voxel).

Kinetic models of hyperpolarized MRI signal in a single voxel

Hyperpolarized carbon-13 MRI researchers commonly rely on linear compartmental models for describing the evolution of signal in a voxel [18, 47, 52]. These models describe the magnetization exchange from the pool of injected hyperpolarized substrate to pools corresponding to various metabolic products. In its simplest form, this amounts to the irreversible metabolic conversion of the substrate S to a single product P performed at a characteristic kinetic rate k_{SP}:

$$S \xrightarrow{k_{SP}} P.$$

Throughout this dissertation, we will focus on extremely simple pathways of this form, though extension to multiple products or bidirectional conversion is straightforward. Note that the clinically-relevant pathways discussed in Section 1.2 can both be modelled in this simple form.

In the absence of external RF excitation, magnetization in a particular voxel i evolves via T_1 decay and label exchange according to the differential equations

$$\frac{d}{dt}\begin{bmatrix} M_{z,i,S}(t) \\ M_{z,i,P}(t) \end{bmatrix} = \begin{bmatrix} -R_{1,i,S} - k_{SP,i} & 0 \\ k_{SP,i} & -R_{1,i,P} \end{bmatrix} \begin{bmatrix} M_{z,i,S}(t) \\ M_{z,i,P}(t) \end{bmatrix} + \begin{bmatrix} k_{TRANS,i} \\ 0 \end{bmatrix} u(t) \tag{1.7}$$

where the states $M_{z,i,S}$ and $M_{z,i,P}$ represent the longitudinal magnetization in voxel i in the substrate and product compartments respectively, the input u models an arterial input function (AIF) describing the arrival of substrate from the circulatory system, and the parameters $k_{SP,i}$, $R_{1,i,S}$, $R_{1,i,P}$, and k_{TRANS} describe the metabolic rate, T_1 decay rate in the substrate pool, and T_1 decay rate in the product pool, and perfusion rate respectively.

When a constant flip angle excitation sequence and repetition time is used for imaging, decay due to RF excitation can be modelled by replacing $R_{1,i,X}$ by an effective decay rate

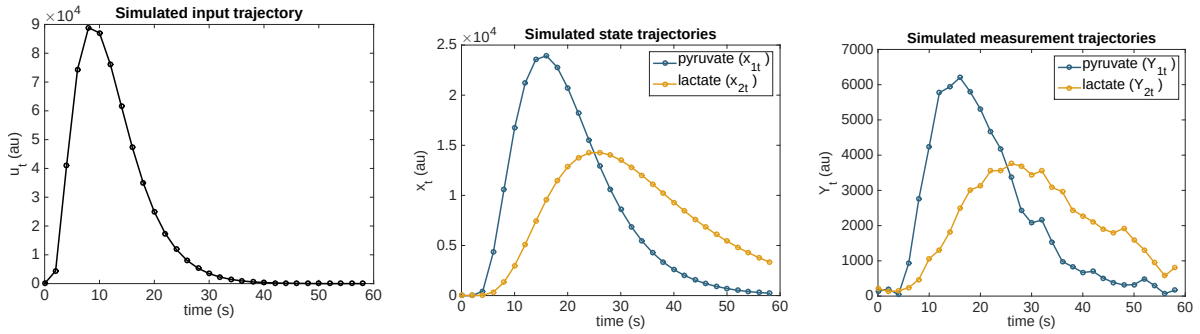
$$R_{1,i,X,\text{effective}} = R_{1,i,X} - \frac{\log(\cos\alpha)}{T_R}$$

where α is the flip angle and T_R is the repetition time, and X denotes an arbitrary compound (either S or P) [97]. However, when a variable flip angle sequence is used, signal decay due to RF excitation must be accounted for as in Section 1.2. This leads to a discrete time model for the transverse and longitudinal magnetization immediately preceding excitation k given by

$$\begin{bmatrix} M_{z,i,S}[k+1] \\ M_{z,i,P}[k+1] \end{bmatrix} = A_d \begin{bmatrix} \cos\alpha_S[k] & 0 \\ 0 & \cos\alpha_P[k] \end{bmatrix} \begin{bmatrix} M_{z,i,S}[k] \\ M_{z,i,P}[k] \end{bmatrix} + B_d u[k] \tag{1.8}$$

where A_d and B_d are computed by discretizing (1.7) assuming a zero order hold with sampling time T_R. A model for the transverse magnetization immediately following excitation k given by

$$M_{xy,i,X}[k] = \sin\alpha_X[k] M_{z,i,X}[k]. \tag{1.9}$$



(a) input u (b) states M_z (c) Noisy measurements Y

Figure 1.3: Simulated trajectories for a pyruvate to lactate conversion model using a constant flip angle sequence with $\alpha_S[k] = \alpha_P[k] = 15°$.

This transverse magnetization leads to the observable signal which we measure as an output from voxel i at time k. In the case of normally-distributed measurements, we model the generated data as

$$Y_{i,X}[k] \sim M_{xy,i,X}[k] + \epsilon_{i,X}[k]$$

where ϵ is independent identically distributed gaussian noise with a known variance σ^2. Simulated trajectories of this model are shown in Figure 1.3.

Estimation of unknown model parameters

Estimating metabolic rate parameters θ_i from experimental data collected from voxel i involves minimizing a statistical loss function $L(\theta_i|Y_i)$ that describes how well a signal model fits the observed data Y_i. Using the model equations (1.8)–(1.9) as the basis of a signal model describing the predicted measurement

$$y_i(\theta_i) = \begin{bmatrix}[M_{xy,i,S}[1] & M_{xy,i,P}[1] & \dots & M_{xy,i,S}[N] & M_{xy,i,P}[N]\end{bmatrix}$$

in terms of the vector model parameters θ_i. Loss functions include:

- the least squares loss

  $$L(\theta_i|Y_i) = \|Y_i - y_i(\theta)\|^2$$

  which corresponds to a nonlinear least squares estimation problem and

- the negative log likelihood loss

  $$L(\theta_i|Y_i) = -\log p_{\theta_i}(Y_i)$$

  which corresponds to a maximum likelihood estimation problem. Unlike the least squares loss function, this loss requires a that a probability density function describing the joint distribution of Y_i be specified. Common choices are $Y_i \sim y_i + \epsilon$ where ϵ is independent, identically-distributed (iid) Gaussian noise or independent Rician noise with location parameters given by y_i [44].

1.3 Magnetic Resonance Fingerprinting

In Chapter 3 of this dissertation we also consider an anatomic imaging technique know as magnetic resonance fingerprinting, based on more traditional proton (^{1}H) MRI.

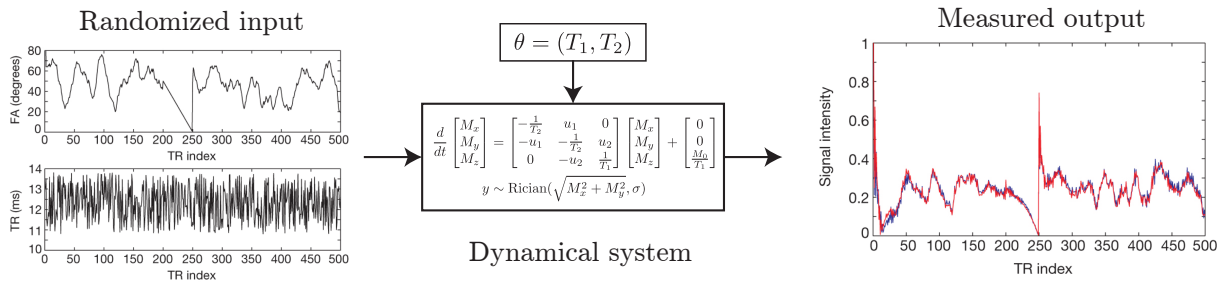
Magnetic resonance imaging (MRI) has traditionally focused on acquisition sequences that are static, in the sense that sequences typically wait for magnetization to return to equilibrium between acquisitions. Recently, researchers have demonstrated promising results based on dynamic acquisition sequences, in which spins are continuously excited by a sequence of random input pulses, without allowing the system to return to equilibrium between pulses. Model parameters corresponding to T_1 and T_2 relaxation, off-resonance and spin density are then estimated from the sequence of acquired data. This technique, termed magnetic resonance fingerprinting (MRF), has been shown to increase the sensitivity, specificity and speed of magnetic resonance studies [61, 28].

Magnetic resonance fingerprinting experiments are conducted in a fashion similar to dynamic system identification. Spins are excited via a sequence of randomized excitation pulses, and dynamic time series of the tissue response are measured (Figure 1.4a). The time series acquired differ based on the MR properties of the tissue, providing a "fingerprint" characterizing the rate parameters of the tissue. Following the experiment, fingerprints are then matched to a dictionary of simulated outputs, allowing the spatial distribution of tissue parameters to be estimated (Figure 1.4b). Further, this technique can be successfully applied to heavily undersampled data in the spatial frequency domain, allowing accurate, quantitative parameter maps to be estimated from relatively short acquisition sequences.

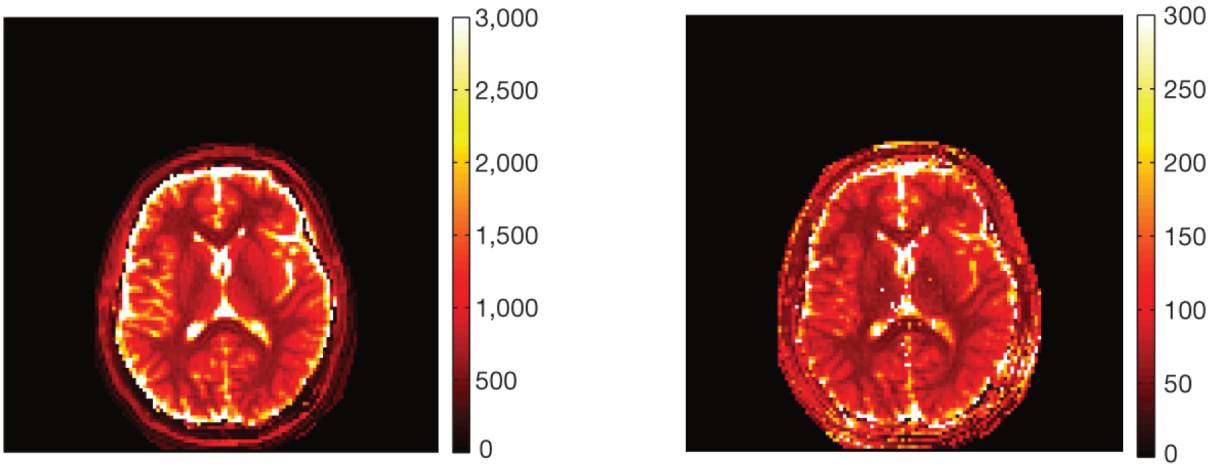
In this dissertation, we study magnetic resonance fingerprinting from an experiment design perspective. In particular, we formulate an optimization problem to replace the randomized flip angle excitation sequences with a sequence optimized for informativeness about model parameters. We demonstrate in Section 3.4 that this results in a significant improvement in estimate accuracy for acquisitions of comparable length.

1.4 Summary of Contributions

In the next four chapters, we present four optimization problems that arise in the design of dynamic MRI experiments and the subsequent data analysis. These problems and our contribution to their algorithmic solution are summarized in Table 1.5. The first involves the design of substrate injection inputs to generate maximally informative data, a problem in which the control input enters linearly. The second and third involve the design of optimized flip angle sequences, again for generating maximally informative data. In contrast with the first, these problems involve a nonlinear control system model, which is significantly more difficult to analyze globally. The fourth problem involves estimating the spatial distribution of metabolic flux parameters from the acquired data. This problem, studied in the final chapter, completes the experimental sequence from experimental design to data acquisition to data analysis.



(a) Spins are excited via a randomized sequence then the measured signal is matched to a simulated dictionary.



(b) This dictionary matching procedure allows quantitative maps of tissue parameters to be generated. Here we show T_1 (left) and T_2 (right) relaxation parameter maps (in milliseconds) generated from a numerical fingerprinting simulation.

Figure 1.4: Outline of a magnetic resonance fingerprinting experiment, adapted from the original paper [61].

Chapter 2: Substrate Injection Design for Hyperpolarized Carbon-13 MRI

In Chapter 2 we consider the optimal design of the injection input subject to constraints on the maximum injection rate and volume. This results in an dynamic optimal experiment design problem of the form discussed in Section 1.1. More formally, we consider the dynamic model defined in Equation (1.8) with an output defined in Equation (1.9) which is corrupted by iid additive Gaussian noise. The problem is to design an injection input $u[k]$ to maximize the Fisher information about the parameter of interest k_{SP} contained in the data generated from a finite number of samples under this model. The input is constrained such that both the maximum injection rate $\|u\|_\infty$ and the maximum injection volume $\|u\|_1$ are upper

Chapter	Domain	Algorithm	Application
2	Linear systems	SDP reformulation w/ exact recovery guarantees	Injection design for 13C MRI (fast globally optimal solutions)
3	Nonlinear systems w/ symmetries	Dynamic programming with symmetry reduction	Flip angle design for MR fingerprinting (75× computational speedup)
4	Nonlinear systems	Derivative-free optimization & nonlinear programming	Flip angle design for 13C MDI (20% improvement in metabolism estimates)
5	Spatially-distributed system identification	Constrained parameter mapping using nonconvex ADMM	13C MRI of prostate cancer (enables reconstruction in low SNR)

Figure 1.5: Summary of Contributions

bounded by some positive constant.

This chapter is based on the paper *Semidefinite relaxations in optimal experiment design with application to substrate injection for hyperpolarized MRI* written in collaboration with my advisor Murat Arcak, that I presented at the 2016 American Control Conference [64]. In this chapter, we show that this problem can be reformulated as a nonconvex quadratic program (QP). We then develop a procedure for approximating the global solution of the QP using a semidefinite programming relaxation. This method allows us to compute approximate solutions to particular instances of this problem as well as bounds on the global solution. In particular, for an instance with realistic values for model parameters, we find that the optimal input consists of a bolus applied at the beginning of the experiment injected at the maximum rate until the volume budget is reached (Figure 1.6). Based on the semidefinite relaxation, we then show that this input achieves an objective function value at least 98.7% of the global optimum, for these particular values of the model parameters.

Chapter 3: Flip Angle Sequence Design for MR Fingerprinting

In Chapter 3, we consider the problem of computing optimal control policies for systems possessing symmetries. This problem is motivated by the design of optimized flip angle sequences for magnetic resonance fingerprinting, a system whose dynamics (governed by the Bloch equations) possess a rotational symmetry about the axis of the main magnetic field B_0. We formulate this problem in terms of maximizing the Fisher information about the T_1 relaxation parameter, and solve it using a symmetry reduction method developed in the chapter.

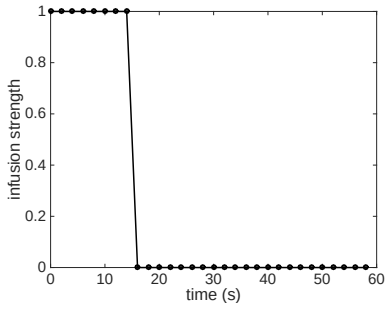


Figure 1.6: Solution to a particular instance of the injection design problem. The optimal input sequence $u[k]$ applies a bolus injection at the maximum allowable rate until the total input budget is reached.

Chapter 3 is based on a manuscript entitled *Symmetry reduction for dynamic programming* that I wrote in collaboration with Axel Barrau, Silvère Bonnabel, and Murat Arcak. This manuscript is currently under review and may be published at a later date. A preliminary version of this manuscript was also presented at the 2017 American Control Conference [68]. In this chapter we present a general method for expoiting symmetries of a dynamical system in order to reduce the computational burden of computing optimal control polices via the dynamic programming algorithm. We then apply this technique to design optimized flip angle sequences for MR fingerprinting. The resulting optimized sequence significantly improves the accuracy of parameter estimates, when compared against a randomized sequence.

Chapter 4: Flip Angle Sequence Design for Hyperpolarized Carbon-13 MRI

In Chapter 4 we again design experimental parameters to maximize the Fisher information about unknown rates in a carbon-13 MRI model. Here, we consider the problem of designing optimal flip angle excitation sequences.

Again we use the model defined in Equation (1.8) with an output defined in Equation (1.9) corrupted by iid noise. We wish to select a sequence of flip angles $\alpha_S[k]$ and $\alpha_P[k]$ used to excite each of the chemical species. Here the choice of $\alpha_S[k]$ and $\alpha_P[k]$ at each time is unconstrained. Since the flip angles enter the model in a nonlinear fashion, the resulting optimization problem is no longer a QP like in Chapter 2, so other optimization techniques must be used.

Chapter 4 is based on the paper *Optimizing flip angles for metabolic rate estimation in hyperpolarized carbon-13 MRI* that I wrote in collaboration with Jeremy W. Gordon, Murat Arcak, and Peder E. Z. Larson, and published in IEEE Transactions on Medical Imaging [67]. In this chapter, we solve the flip angle design problem to local optimality under additional smoothness constraints in using a nonlinear programming approach. The resulting optimized flip angle sequence is shown in Figure 1.7. This flip angle sequence results in a 20% decrease

in the uncertainty of metabolic rate estimates, when compared against the best existing sequences.

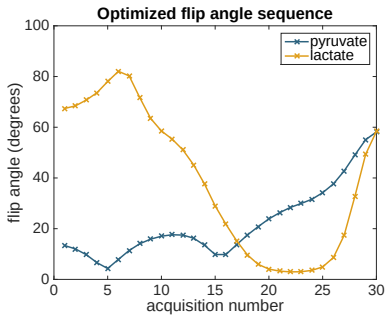


Figure 1.7: Optimized input sequence for the flip angle sequence design problem.

Chapter 5: Constrained Parameter Mapping

The final chapter investigates the computation of maps of metabolic activity from experimental data. Here we assume that we are given a statistical model for the data as well as a loss function, as described in Section 1.2. The challenge is to summarize the spatial, temporal and chemical information contained in the dynamic experimental data into a single spatial map of metabolic activity. We do so by estimating a value for the metabolic rate parameter $\theta_i = k_{SP,i}$ for each voxel i in space.

Since the objects imaged often contain spatial structure, this structure can be exploited to improve the quality of the estimated parameter maps. This can be achieved by adding regularization to the objective function that is optimized. Formally, we solve an optimization problem of the form

$$\textbf{minimize} \sum_i L(\theta_i|Y_i) + \lambda r(\theta)$$

where L is a loss function that depends on the data Y_i collected in each voxel i, and r is a regularization term that couples nearby voxels thereby enforcing spatial structure in the estimated maps. Possible choices of regularization used to enforce smoothness, sparsity and edge preservation include ℓ_2, ℓ_1 and total variation penalties. By including such penalties to exploit spatial correlations in the data, we show that better image quality can be achieved compared with independently fitting each voxel.

This chapter is based on recently-completed work, and will form the basis of a manuscript to be submitted for publication at a later date. It is loosely based on an accepted conference abstract that was presented at the 2017 Annual Meeting of the International Society for Magnetic Resonance in Medicine [69].

Chapter 2

Semidefinite Relaxations for Optimal Experiment Design in Linear Systems

2.1 Introduction

Optimal experiment design

In this chapter we consider the problem of estimating uncertain parameters in a state space model from noisy output data. For such a system, the reliability of the parameter estimates depends on the choice of input used to excite the system, as some inputs provide greater information about the parameters than others. The problem of designing an input that is maximally informative is known as optimal experiment design and much work has been done on this problem in the last 50 years [35, 39, 58, 105].

Historically, most work on optimal experiment design for dynamic systems has focused on frequency domain techniques, where an optimal input is designed based on its power spectrum. Here, we approach this problem in the time domain, allowing us to impose amplitude (ℓ_∞) and ℓ_1 constraints on the admissible inputs. Amplitude-constrained optimal experiment design is NP-hard in general [70], but semidefinite relaxation techniques can be used to generate approximate solutions and to bound the suboptimality of such solutions [70, 71]. In contrast with [70, 71], in this chapter we restrict our attention to linear measures of the information allowing us to 1) write the objective function as a quadratic function of the input sequence and 2) apply an exactness result for quadratically constrained quadratic programming to give sufficient conditions under which the semidefinite relaxation recovers the global solution. We also present the results in terms of a state space model enabling us to model uncertainty in the initial state.

Metabolic MRI using hyperpolarized substrates

We are motivated by a problem in metabolic magnetic resonance imaging (MRI) using hyperpolarized substrates. Hyperpolarized carbon-13 MRI has enabled the real-time observation

of perfusion and metabolism in preclinical and clinical studies [37, 29, 52, 77, 5, 101]. This technology is made possible by techniques for dynamic nuclear polarization (DNP) that have led to signal-to-noise ratio (SNR) increases of four to five orders of magnitude compared with endogenous signal in dissolved ^{13}C-labelled molecules [4, 38]. Injected [1-^{13}C] pyruvate is frequently used as a substrate in metabolism experiments and its rate of conversion to [1-^{13}C] lactate has been shown to distinguish between healthy and cancerous tissues in animal [29], and recently human [77], studies.

The goal of a metabolic MRI experiment is to learn the spatial distribution of metabolic rates, as this indicates the regions of the body where a particular metabolic pathway is active. Noise in the observed image data leads to uncertainty in estimates of metabolic rate parameters, but the amount of uncertainty can be mitigated with experiment design.

The problem of optimally designing an image acquisition sequence is considered in [65, 67]. In this chapter we investigate optimal substrate injection profiles. After hyperpolarization the substrate must be injected into the test subject, where metabolism occurs and a sequence of images are acquired. Greater injection volumes lead to better signal-to-noise ratio, but for safety reasons the injection profile is limited by the rate at which substrate can be injected and by the total amount of fluid injected. Thus determining the optimal injection profile is of clinical interest.

Outline

We begin in Section 2.2 by introducing the time domain optimal experiment design problem for linear systems, and discuss relevant relaxations from the literature. Then in Section 2.3 we present a mathematical model of hyperpolarized MRI and use semidefinite relaxation techniques to design optimal injection inputs for hyperpolarized substrates. We constrain the total amount of fluid injected using the ℓ_2 or ℓ_1 norms of the input signal. In the ℓ_2-constrained case, the relaxation is tight and we are able to recover a globally optimal solution to the original problem. In the ℓ_1-constrained case we do not recover a solution from the relaxation, but the objective value of the relaxation can be used to bound the optimality gap and show that a simple boxcar injection achieves at least 98.7% of the global optimum. Matlab code to reproduce the results in this chapter is available at `https://github.com/maidens/ACC-2016`.

2.2 Semidefinite Relaxations in Optimal Experiment Design

Consider a discrete-time linear system

$$\begin{aligned} x_{t+1} &= A(\theta)x_t + B(\theta)u_t \\ y_t &= Cx_t \\ Y_t &\sim N(y_t, \Sigma) \end{aligned} \tag{2.1}$$

over a horizon $0 \leq t \leq N$ where the initial state x_0 and time-invariant state dynamics matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n_u}$ are dependent on some unknown parameter vector $\theta \in \mathbb{R}^p$. We wish to choose an input sequence u to maximize some measure of the Fisher information matrix

$$\mathcal{I}(\theta) = \sum_{t=0}^{N} (\nabla_\theta x_t)^T C^T \Sigma^{-1} C (\nabla_\theta x_t) \tag{2.2}$$

about the parameter vector θ contained in the data Y. The Fisher information is a positive semidefinite matrix that is used in experiment design for parameter estimation as a measure of the informativeness of an experiment [87, 105]. The inverse of the information gives an upper bound on the covariance of an arbitrary unbiased estimator θ via the Cramer-Rao inequality. See [63] for a derivation of (2.2).

Quadratic objective function

We consider nonnegative, linear information metrics of the form

$$\varphi\big(\mathcal{I}(\theta)\big) = \operatorname{tr}\big(K\mathcal{I}(\theta)\big) \tag{2.3}$$

where $K \succeq 0$ is positive semidefinite. Particular cases include the T-optimal design criterion (where K is the $p\times p$ identity matrix) [87] and the c-optimal design criterion (where $K = cc^T$) which is optimal for the scalar parameter $c^T\theta$ [88].

To formulate the problem of choosing the sequence u_k for $k = 0, \ldots, N-1$ we define the stacked vector $\mathbf{u} = [u_0^T \ldots u_{N-1}^T]^T$ and denote its components $\mathbf{u}^{(j,k)}$ where $(j,k) = kn_u + j$ defines a reverse lexicographic ordering, j ranges over the input index from 1 to n_u and k ranges over the time index from 0 to $N-1$. With this notation established, we can state the following result.

Proposition 1. *The objective (2.3) is a quadratic function*

$$\varphi\big(\mathcal{I}(\theta)\big) = \mathbf{u}^T Q(\theta)\mathbf{u} + 2q(\theta)^T\mathbf{u} + q_0(\theta)$$

of the design variable $\mathbf{u}$*. The entries of the matrices appearing in the objective function are computed as*

$$Q(\theta)_{(j',k')(j,k)} = \sum_{t=\max\{k,k'\}+1}^{N} \sum_{h,h'=1}^{n} \sum_{i,i'=1}^{p} K^{ii'} M^{h'}_{k'i'j'}(t) S_{hh'} M^{h}_{kij}(t)$$

$$q(\theta)_{(j,k)} = \sum_{t=k+1}^{N} \sum_{h,h'=1}^{n} \sum_{i,i'=1}^{p} K^{ii'} m^{h'}_{i'}(t) S_{hh'} M^{h}_{kij}(t)$$

$$q_0(\theta) = \sum_{t=0}^{N} \sum_{h,h'=1}^{n} \sum_{i,i'=1}^{p} K^{ii'} m^{h'}_{i'}(t) S_{hh'} m^{h}_{i}(t)$$

where $S_{hh'}$ is the hh'-th entry of $C^T\Sigma^{-1}C$, $m_i^h(t)$ is the h-th entry of the vector

$$A^t\frac{\partial x_0}{\partial\theta_i}+\sum_{\ell=0}^{t-1}A^{t-\ell-1}\frac{\partial A}{\partial\theta_i}A^\ell x_0$$

and $M_{kij}^h(t)$ is the h,j-th entry of the matrix

$$A^{t-k-1}\frac{\partial B}{\partial\theta_i}+\sum_{\ell=k+1}^{t-1}A^{t-\ell-1}\frac{\partial A}{\partial\theta_i}A^{\ell-k-1}B.$$

We see that for a linear dynamical system with Gaussian-distributed measurements, optimal experiment design is a nonconvex quadratic programming problem. If there are no constraints on the admissible inputs the optimal value is infinite, as by choosing $\mathbf{u}^*$ as an eigenvector of $Q(\theta)$ that corresponds to a nonzero eigenvalue, letting $\mathbf{u}=\alpha\mathbf{u}^*$ for $\alpha>0$, we can make $\mathbf{u}^TQ(\theta)\mathbf{u}+2q(\theta)^T\mathbf{u}+q_0(\theta)$ arbitrarily large.

Constrained quadratic programming is NP-hard in general [33]. But certain quadratic programming problems lend themselves to polynomial-time approximation algorithms using a semidefinite programming relaxation [36, 59]. We first give a proof of Proposition 1, before moving on to semidefinite relaxations of the optimal experiment design problem.

Proof. Unrolling the recursion relation (2.1), we can write

$$x_t=A^tx_0+\sum_{k=0}^{t-1}A^{t-k-1}Bu_k.$$

Applying the chain rule to (2.1) we get a recursion relation for the sensitivities

$$\frac{\partial}{\partial\theta_i}x_{t+1}=A\frac{\partial}{\partial\theta_i}x_t+\frac{\partial A}{\partial\theta_i}x_t+\frac{\partial B}{\partial\theta_i}u_t$$

which can be unrolled as

$$\begin{aligned}\frac{\partial}{\partial\theta_i}x_t&=A^t\frac{\partial x_0}{\partial\theta_i}+\sum_{\ell=0}^{t-1}A^{t-\ell-1}\left[\frac{\partial A}{\partial\theta_i}x_\ell+\frac{\partial B}{\partial\theta_i}u_\ell\right].\\&=A^t\frac{\partial x_0}{\partial\theta_i}+\sum_{\ell=0}^{t-1}A^{t-\ell-1}\left[\frac{\partial A}{\partial\theta_i}\left(A^\ell x_0+\sum_{k=0}^{\ell-1}A^{\ell-k-1}Bu_k\right)+\frac{\partial B}{\partial\theta_i}u_\ell\right]\\&=\left(A^t\frac{\partial x_0}{\partial\theta_i}+\sum_{\ell=0}^{t-1}A^{t-\ell-1}\frac{\partial A}{\partial\theta_i}A^\ell x_0\right)\\&\quad+\sum_{k=0}^{t-1}\left(A^{t-k-1}\frac{\partial B}{\partial\theta_i}+\sum_{\ell=k+1}^{t-1}A^{t-\ell-1}\frac{\partial A}{\partial\theta_i}A^{\ell-k-1}B\right)u_k\end{aligned}$$

To simplify notation, from this point forward we will use the Einstein summation convention, where upper and lower repeated indices denotes an implicit summation over those indices. In this notation, we see that the sensitivity matrix $\nabla_\theta x_t$ has entries

$$(\nabla_\theta x_t)_i^h = M_{kij}^h(t)\mathbf{u}^{jk} + m_i^h(t)$$

where M and m are defined as in the statement of the proposition and the index k runs from 0 to $t-1$. Thus the objective function is expressed as

$$\begin{aligned}
&\operatorname{tr}\Big(K\mathcal{I}(\theta)\Big) = K^{ii'}\mathcal{I}(\theta)_{ii'} \\
&= K^{ii'}\sum_{t=0}^{N}(\nabla_\theta x_t)_{i'}^{h'} S_{hh'}(\nabla_\theta x_t)_i^h \\
&= K^{ii'}\sum_{t=0}^{N}\Big(M_{k'i'j'}^{h'}(t)\mathbf{u}^{j'k'} + m_{i'}^{h'}(t)\Big)S_{hh'}\Big(M_{kij}^{h}(t)\mathbf{u}^{jk} + m_i^h(t)\Big) \\
&= \sum_{k,k'=0}^{N-1}\mathbf{u}^{j'k'}\left(\sum_{t=\max\{k,k'\}+1}^{N} K^{ii'}M_{k'i'j'}^{h'}(t)S_{hh'}M_{kij}^{h}(t)\right)\mathbf{u}^{jk} \\
&\quad + 2\sum_{k=0}^{N-1}\left(\sum_{t=k+1}^{N} K^{ii'}m_{i'}^{h'}(t)S_{hh'}M_{kij}^{h}(t)\right)\mathbf{u}^{jk} \\
&\quad + \sum_{t=0}^{N} K^{ii'}m_{i'}^{h'}(t)S_{hh'}m_i^h(t) \\
&= \mathbf{u}^T Q(\theta)\mathbf{u} + 2q(\theta)^T\mathbf{u} + q_0(\theta)
\end{aligned}$$

□

Semidefinite relaxation of quadratic programs

Semidefinite relaxations of indefinite quadratic programming problems have been the subject of considerable study over the past two decades [36, 59, 79, 78, 107, 108]. The essential idea is to take a quadratic program

$$\begin{aligned}
&\textbf{maximize} && u^TQu + 2q^Tu + q_0 \\
&\textbf{subject to} && u \in \mathcal{P} \subseteq \mathbb{R}^d
\end{aligned} \tag{2.4}$$

over a linear vector space and transform it to a linear problem over a quadratic space

$$\begin{aligned}
&\textbf{maximize} && \operatorname{tr}\left(\begin{bmatrix} Q & q \\ q^T & q_0 \end{bmatrix}\begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix}\right) \\
&\textbf{subject to} && \begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \in \tilde{\mathcal{P}} \subseteq \mathbb{R}^{(d+1)\times(d+1)} \\
& && U = uu^T
\end{aligned} \tag{2.5}$$

by introducing a rank 1 semidefinite variable $U = uu^T$ and translating the constraint set $\mathcal{P}$ from the the linear vector space to a convex constraint set $\tilde{\mathcal{P}}$ in the semidefinite matrix space. The problem (2.5) in non convex, but the equality constraint $U = uu^T$ can be relaxed to the inequality $U \succeq uu^T$ then transformed to semidefinite constraint $\begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \succeq 0$ via Schur complement. Thus (2.5) can be relaxed to the convex problem

$$\begin{aligned} \textbf{maximize} \quad & \operatorname{tr}\left(\begin{bmatrix} Q & q \\ q^T & q_0 \end{bmatrix}\begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix}\right) \\ \textbf{subject to} \quad & \begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \in \tilde{\mathcal{P}} \\ & \begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \succeq 0. \end{aligned} \tag{2.6}$$

The value of the convex program (2.6) can then be used as an upper bound on the solution to (2.4). Further if a solution $\begin{bmatrix} U^* & u^* \\ u^{*T} & 1 \end{bmatrix}$ to (2.6) happens to have rank 1 then u^* is a global solution to (2.4).

To translate constraints $u \in \mathcal{P} \subseteq \mathbb{R}^d$ on u to constraints $U \in \tilde{\mathcal{P}} \subseteq \mathbb{R}^{d\times d}$ on U there are a number of methods.

- Quadratic constraints on u of the form
$$\mathcal{P} = \{u : u^T R u + 2r^T u + r_0 \leq 0\}$$
are translated to constraints of the form
$$\tilde{\mathcal{P}} = \{U : \operatorname{tr}\left(\begin{bmatrix} R & r \\ r^T & r_0 \end{bmatrix}\begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix}\right) \leq 0\}.$$
For example, the ℓ_2 constraint $\|u\|_2 \leq c$ becomes $\operatorname{tr}(U) \leq c^2$.

- Amplitude constraints on u of the form $\mathcal{P} = \{u : |u_t| \leq c_t \; t = 1, \ldots, d\}$ for particular constants c_t are modelled as d homogeneous quadratic constraints $\mathcal{P} = \{u : u_t^2 \leq c_t^2 \; t = 1, \ldots, d\}$ which correspond to $\tilde{\mathcal{P}} = \{U : U_{tt} \leq c_t^2 \; t = 1, \ldots, d\}$.

- Box constraints of the form $\mathcal{P} = \{u : 0 \leq u_t \leq c_t\}$ can be translated as $\tilde{\mathcal{P}} = \{U : U_{tt} \leq c_t^2 \; t = 1, \ldots, d \quad \wedge \quad U_{st} \geq 0 \; s, t = 1, \ldots, d\}$.

- For u normalized such that $0 \leq u_t \leq 1$, the relaxation can be tightened by noting that $u_i^2 \leq u_i$. Thus the constraints $U_{ii} \leq u_i \;\; i = 1, \ldots, d$ can be added to tighten the relaxation.

- If $u_t \geq 0$ and a is a vector with $a_t \geq 0$ then the linear constraint $\mathcal{P} = \{u : a^T u \leq b\}$ can be translated by noting that $0 \leq a^T u \leq b$ implies that $\operatorname{tr}(aa^T U) = \operatorname{tr}(a^T U a) = a^T u u^T a = (a^T u)^2 \leq b^2$. This results in a tighter approximation than adding the constraint $a^T u \leq b$ (see Lemma 1 of[49]). In particular, if we denote the $d \times d$ matrix of ones by E then the ℓ_1-norm constraint $\|u\|_1 \leq b$ can be translated as $\operatorname{tr}(EU) \leq b^2$.

A result on exact recovery from semidefinite relaxation

For a vector $u \in \mathbb{R}^d$ let u^2 denote the vector obtained by squaring the entries of u componentwise and consider a problem of the form

$$\begin{aligned} \textbf{maximize} \quad & u^T Q u + 2q^T u + q_0 \\ \textbf{subject to} \quad & u^2 \in \mathcal{F}. \end{aligned} \tag{2.7}$$

The following result gives sufficient conditions for the semidefinite relaxation

$$\begin{aligned} \textbf{maximize} \quad & \operatorname{tr}\left(\begin{bmatrix} Q & q \\ q^T & q_0 \end{bmatrix} \begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \right) \\ \textbf{subject to} \quad & \operatorname{diag}(U) \in \mathcal{F} \\ & \begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \succeq 0 \end{aligned} \tag{2.8}$$

to recover the global solution of (2.7).

Proposition 2 (Adapted from Theorem 2 of [108])**.** *If $Q_{ij} \geq 0$ for all $i \neq j$, $q_i \geq 0$ for all i and $\mathcal{F} \subseteq \mathbb{R}^d$ is a closed convex set then the values of (2.7) and (2.8) coincide. Moreover, if $\tilde{U}^*$ is a solution of (2.8) then $\sqrt{\operatorname{diag}(\tilde{U}^*)}$ is a solution of (2.7).*

Thus if $u \in \mathcal{P}$ can be expressed in the form $u^2 \in \mathcal{F}$ for some convex $\mathcal{F}$, and the entries of Q and q are nonnegative, we can globally solve (2.4) via the convex relaxation.

2.3 Infusion Input Design for Substrate Injection in Hyperpolarized Carbon-13 MRI

We consider a linear model of magnetization exchange resulting from the injection of a hyperpolarized substrate, observed using a flip angle sequence $\alpha_{k,t}$ [53]:

$$\frac{dx}{dt}(t) = \begin{bmatrix} -k_{PL} - R_{1P} - \frac{1-\cos(\alpha_{1,t})}{\Delta t} & 0 \\ k_{PL} & -R_{1L} - \frac{1-\cos(\alpha_{2,t})}{\Delta t} \end{bmatrix} x(t) + \begin{bmatrix} k_{TRANS} \\ 0 \end{bmatrix} AIF(t) \tag{2.9}$$

where $AIF(t)$ is an arterial input function. The result of a bolus (impulse) injection of substrate is often modelled as the arterial input function is of the form [75]

$$AIF(t) = A_0 t^\gamma e^{-t/\beta}.$$

In the case $\gamma = 2$ the samples of this AIF can be modelled as the impulse response of the $\gamma + 1 = 3$rd order system

$$\begin{aligned} z_{t+1} &= \begin{bmatrix} 3e^{-(\Delta t)/\beta} & -3e^{-2(\Delta t)/\beta} & e^{-3(\Delta t)/\beta} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} z_t + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_t \\ AIF_t &= A_0 \begin{bmatrix} e^{-(\Delta t)/\beta} & e^{-2(\Delta t)/\beta} & 0 \end{bmatrix} z_t. \end{aligned} \tag{2.10}$$

$$x_{t+1} = \begin{bmatrix} 3e^{-(\Delta t)/\beta} & -3e^{-2(\Delta t)/\beta} & e^{-3(\Delta t)/\beta} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ A_0 e^{-(\Delta t)/\beta}\bar{B}_1 & A_0 e^{-2(\Delta t)/\beta}\bar{B}_1 & 0 & \bar{A}_{11} & \bar{A}_{12} \\ A_0 e^{-(\Delta t)/\beta}\bar{B}_2 & A_0 e^{-2(\Delta t)/\beta}\bar{B}_2 & 0 & \bar{A}_{21} & \bar{A}_{22} \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t \tag{2.12}$$

$$y_t = \begin{bmatrix} 0 & 0 & 0 & \sin(\alpha_{1,t}) & 0 \\ 0 & 0 & 0 & 0 & \sin(\alpha_{2,t}) \end{bmatrix} x_t$$

We discretize (3.3) with step $\Delta t = 2$ s assuming a zero-order hold on the AIF, yielding a model

$$\begin{aligned} \bar{x}_{t+1} &= \bar{A}\ \bar{x}_t + \bar{B}\ AIF_t \\ y_t &= \begin{bmatrix} 0 & 0 & 0 & \sin(\alpha_{1,t}) & 0 \\ 0 & 0 & 0 & 0 & \sin(\alpha_{2,t}) \end{bmatrix} \bar{x}_t. \end{aligned} \tag{2.11}$$

Combining (2.10) and (2.11), we get a model of the full system (2.12) mapping the infusion input u to the observed signals.

We now solve an example instance of this system with model parameters taken from [67] which are shown in Table 4.1 , along with noise covariance matrix $\Sigma = I$, horizon of $N = 30$ samples and a constant flip angle sequence $\alpha_{k,t} = 15°$. Computation times for solving this problem are given in Table 2.2.

R_{1P}	R_{1L}	k_{PL}	k_{TRANS}	t_0	γ	β	A_0
1/10	1/10	0.07	0.055	3.2596	2.1430	3.4658	1.0411 $\times 10^4$

Table 2.1: Nominal parameter values used

problem	time to generate SDP	SDP decision variable size	number of SDP constraints	time to solve SDP
ℓ_2 constrained	12.0 s	30 × 30 symmetric	31	0.58 s
ℓ_1 constrained		31 × 31 symmetric	497	5.51 s

Table 2.2: Computation time to solve semidefinite programming relaxations. The SDP was solved using CVX [42] with the SeDuMi backend in MATLAB v8.4.0 running on a Macbook laptop (2.3 GHz quad-core Intel Core i7 Ivy Bridge processor, 8GB memory).

ℓ_2 constrained input

The substrate injection is constrained to limit the rate of injection to $|u_t| \leq 1$ and to limit the ℓ_2 norm of the injection to $\|u\|_2 \leq 4$. Thus we wish to solve the quadratically-constrained

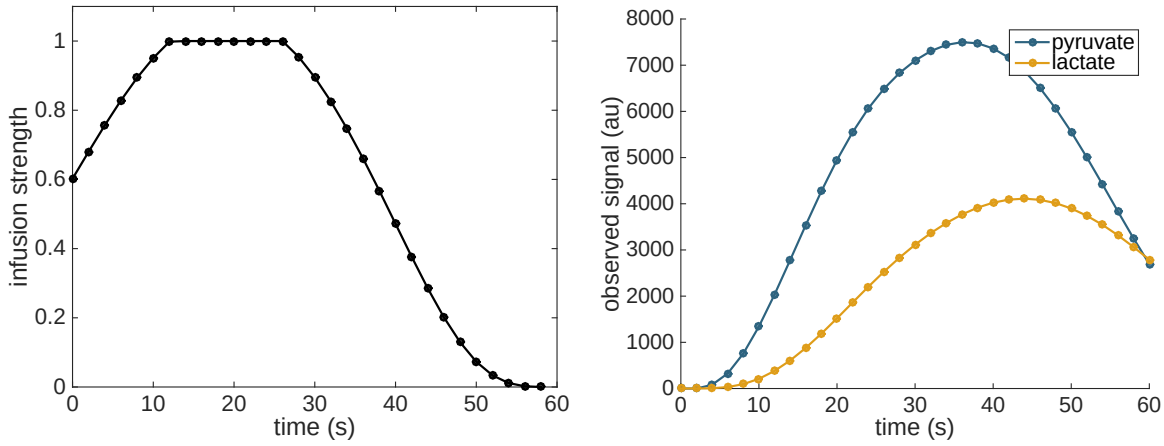
quadratic program

$$\begin{aligned} \textbf{maximize} \quad & u^T Q(\theta) u \\ \textbf{subject to} \quad & u_t^2 \leq 1 \\ & \sum_{t=0}^{N-1} u_t^2 \leq 16. \end{aligned} \tag{2.13}$$

The matrix Q is nonnegative and the constraints are of the form $u^2 \in \mathcal{F}$ where $\mathcal{F}$ is a closed convex set. Therefore this problem yields a semidefinite relaxation

$$\begin{aligned} \textbf{maximize} \quad & \operatorname{tr}\Big(Q(\theta)U\Big) \\ \textbf{subject to} \quad & U_{tt} \leq 1 \\ & \operatorname{tr}(U) \leq 16. \end{aligned} \tag{2.14}$$

whose solution has rank 1 (by Proposition 2). So we can extract a globally optimal solution to (2.13) from its semidefinite relaxation. The resulting optimal input trajectory and the corresponding output trajectories are shown in Fig. 2.1.



(a) Optimal input (b) Corresponding outputs

Figure 2.1: Optimal solution to (2.13) computed using semidefinite relaxation.

ℓ_1 constrained input

We now replace the ℓ_2-norm constraint with an ℓ_1-norm constraint $\|u\|_1 \leq 8$ that limits the total amount of substrate injected. This constraint is more clinically relevant than the ℓ_2-norm constraint, as the total substrate than can be injected is often limited due to safety-

related concerns. This leads to a linearly-constrained QP of the form

$$\begin{aligned} \textbf{maximize} \quad & u^T Q(\theta) u \\ \textbf{subject to} \quad & 0 \leq u_t \leq 1 \\ & \sum_{t=0}^{N-1} u_t \leq 8. \end{aligned} \tag{2.15}$$

This QP can be relaxed to the semidefinite program

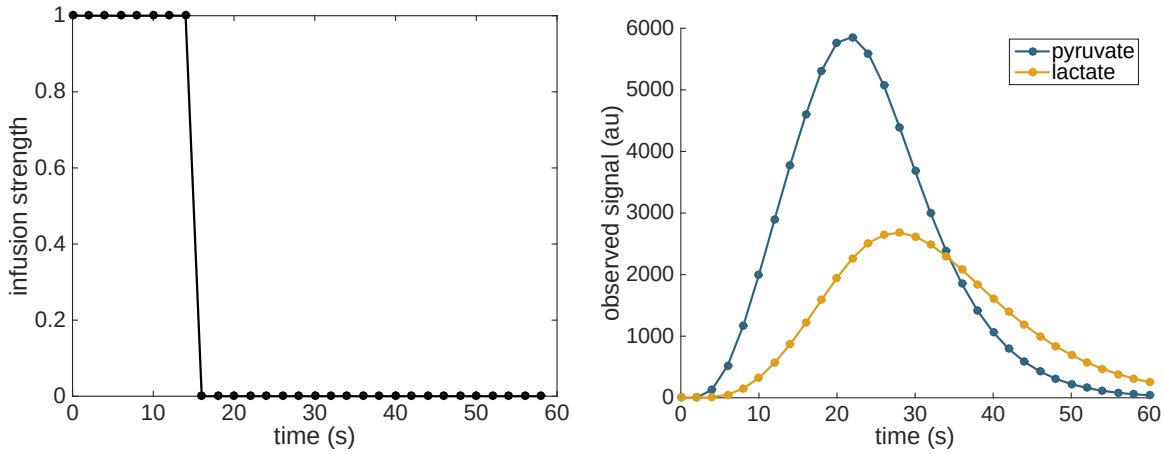
$$\begin{aligned} \textbf{maximize} \quad & \mathrm{tr}\left(Q(\theta)U\right) \\ \textbf{subject to} \quad & \begin{bmatrix} U & u \\ u^T & 1 \end{bmatrix} \succeq 0 \\ & U_{tt} \leq u_t \\ & \mathrm{tr}\left(EU\right) \leq 64 \\ & U_{st} \geq 0 \end{aligned} \tag{2.16}$$

where both U and u are decision variables and E is the $N \times N$ matrix of ones. In this case, the constraint set is more complex than in the ℓ_2-norm constrained case. Therefore the solution to the semidefinite program does not have rank 1, so we cannot extract the solution to (2.15). But the optimal value of (2.16) is an upper bound on the optimal value of (2.15), so given a proposed solution of (2.15) we can bound the optimality gap using the value of (2.16).

We conjecture that the global solution of (2.15) is the boxcar shown in Fig. 2.2. This input is what is currently used in practice: the substrate is injected at the maximum rate until the total allowable volume has been injected. Comparing the optimal value 1.2888×10^{10} of the relaxation with the objective value 1.2724×10^{10} we see that the boxcar input achieves a value of a factor of at least 0.9873 the optimal value. Thus even if our conjecture is incorrect and the boxcar is not optimal, the improvement that may be achieved by the optimal input is negligible. This observation helps to validate the current practice in hyperpolarized MRI.

2.4 Conclusion

We have found that semidefinite relaxation can be used to compute an optimal hyperpolarized substrate infusion input profile for estimating uncertain metabolic rate parameters in metabolic MRI. Future work will focus on investigating the relationship between the rank 1 recovery that we see in the ℓ_2-norm-constrained case and properties of the dynamic system such as positivity and passivity, and attempting to extend these results to nonlinear measures of the information such as the D- E- and A-optimality criteria used in [71].



(a) Boxcar input

(b) Corresponding outputs

Figure 2.2: The boxcar input function, which we conjecture to be optimal for the ℓ_1-constrained problem.

Chapter 3

Symmetry Reduction for Optimal Control of Nonlinear Systems

3.1 Introduction

The dynamic programming algorithm for computing optimal control policies has, since its development, been known to suffer from the "curse of dimensionality" [8]. Its applicability in practice is typically limited to systems with four or five continuous state variables because the number of points required to grid a space of n continuous state variables increases exponentially with the state dimension n. This complexity has led to a collection of algorithms for approximate dynamic programming, which scale to systems with larger state dimension but lack the guarantees of global optimality of the solution associated with the original dynamic programming algorithm [9, 12, 85, 86].

In practice, many real-world systems exhibit symmetries that can be exploited to reduce the complexity of system models. Symmetry reduction has found applications in fields ranging from differential equations [24, 15] to model checking [32, 54]. In control engineering, symmetries have been exploited to improve control of mechanical systems [13, 19], develop more reliable state estimators [7], study the controllability of multiagent systems [89] and to reduce the complexity of stability and performance certification for interconnected systems [3, 92]. Symmetry reduction has also been applied to the computation of optimal control policies for continuous-time systems in [43, 83] and Markov decision processes (MDPs) in [110, 76].

In this chapter, we present a theory of symmetry reduction for the optimal control of discrete-time, stochastic nonlinear systems with continuous state variables. This reduction allows dynamic programming to be performed in a lower-dimensional state space. Since the computational complexity of a dynamic programming iteration increases exponentially with state dimension, this reduction significantly decreases computational burden. Further, our proposed method does not rely on an explicit transformation of the state update equations, making the method applicable in situations where a such a transformation is difficult or

impossible to find analytically.

We present two theorems that summarize our method of symmetry reduction. Theorem 3 describes how symmetries of the system dynamics imply symmetries of the optimal cost and optimal policy functions. Theorem 4 then describes a method of computing the cost function based on reduced coordinate system that depends on fewer state variables.

This chapter is organized as follows: in Section 3.2 we introduce notation and provide background information both on dynamic programming for optimal control, and on the mathematical theory of symmetries. In Section 3.3, we derive our main theoretical results, that is, we prove that control system symmetries induce symmetries of the optimal cost function and optimal control policy, and then leverage the result to present a general method of performing dynamic programming in reduced coordinates. In Section 3.4 we apply symmetry reduction to compute the solution of an optimal control problem arising in dynamic MRI acquisition. In Section 3.5 we apply the algorithm to the problem of coordination on Lie groups. Simulations are performed on a cooperative control problem for two Dubins vehicles. Code to reproduce the computational results in this chapter is available at `https://github.com/maidens/Automatica-2017`.

3.2 Dynamic Programming and Symmetries

In this section, we first recall the main features of dynamic programming for optimal control of stochastic discrete time systems. Then we introduce our problem and provide the reader with a primer on the classical theory of symmetries. We also introduce the notion of invariant control systems with invariant costs.

Dynamic programming for optimal control of stochastic systems

We consider a discrete-time dynamical system

$$x_{k+1} = f_k(x_k, u_k, w_k), \quad k = 0, 1, \ldots, N-1 \tag{3.1}$$

where $x_k \in \mathcal{X} \subseteq \mathbb{R}^n$ is the system state, $u_k \in \mathcal{U} \subseteq \mathbb{R}^m$ is the control variable to be chosen at time k, $w_k \in \mathcal{W} \subseteq \mathbb{R}^\ell$ are independent continuous random variables each with density p_k, and $N \in \mathbb{Z}_+$ is a finite control horizon. Associated with this system is an additive cost function

$$g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$$

that we wish to minimize through our choice of u_k. We define a *control system* to be a tuple $\mathcal{S} = (\mathcal{X}, \mathcal{U}, \mathcal{W}, p, f, g, N)$ where $p = \prod_{k=0}^{N-1} p_k$ is the joint density of the random variables w_k.

We consider a class of control policies $\pi = \{\mu_0, \ldots, \mu_{N-1}\}$ where $\mu_k : \mathcal{X} \to \mathcal{U}$ maps observed states to admissible control inputs. Given an initial state x_0 and a control policy

π, we define the expected cost under this policy as

$$J_\pi(x_0) = \mathbb{E}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right].$$

An optimal policy π^* is defined as one that minimizes the expected cost:

$$J_{\pi^*}(x_0) = \min_{\pi\in\Pi} J_\pi(x_0)$$

where Π denotes the set of all admissible control policies. The optimal cost function, denoted $J^*(x_0)$, is defined to be the expected cost corresponding to an optimal policy.

An optimal policy π^* and the optimal cost function J^* can be computed using the dynamic programming algorithm. We quote the following result due to Bellman from [11]:

Proposition 3 (Dynamic Programming)**.** *For every initial state x_0, the optimal cost $J^*(x_0)$ of the basic problem is equal to $J_0(x_0)$, given by the last step of the following algorithm, which proceeds backward in time from period $N-1$ to period 0:*

$$\begin{aligned} J_N(x_N) &= g_N(x_N) \\ J_k(x_k) &= \min_{u_k\in\mathcal{U}} \mathbb{E}\Big[g_k(x_k, u_k, w_k) + J_{k+1}\Big(f_k(x_k, u_k, w_k)\Big)\Big] \\ &\qquad k = 0, 1, \ldots, N-1, \end{aligned} \tag{3.2}$$

where the expectation is taken with respect to the probability distribution of w_k. Furthermore, if $u_k^* = \mu_k^*(x_k)$ minimizes the right hand side of (3.2) for each x_k and k, then the policy $\pi^* = \{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is optimal.

Invariant system with invariant costs

We first recall the definition of a transformation group for a control system, as in [72, 51, 90]. See [84] for the more general theory.

Definition 2 (Transformation group)**.** *A transformation group on $\mathcal{X}\times\mathcal{U}\times\mathcal{W}$ is set of tuples $h_\alpha = (\phi_\alpha, \chi_\alpha, \psi_\alpha)$ parametrized by elements α of a Lie group $\mathcal{G}$ having dimension r, such that the functions $\phi_\alpha : \mathcal{X} \to \mathcal{X}$, $\chi_\alpha : \mathcal{U} \to \mathcal{U}$ and $\psi_\alpha : \mathcal{W} \to \mathcal{W}$ are all C^1 diffeomorhpisms and satisfy:*

- *$\phi_e(x) = x$, $\chi_e(u) = u$, $\psi_e(w) = w$ when e is the identity of the group $\mathcal{G}$ and*
- *$\phi_{a*b}(x) = \phi_a \circ \phi_b(x)$, $\chi_{a*b}(u) = \chi_a \circ \chi_b(u)$, $\psi_{a*b}(x) = \psi_a \circ \psi_b(x)$ for all $a, b \in \mathcal{G}$ where $*$ denotes the group operation and $\circ$ denotes function composition.*

To simplify notation we will sometimes suppress the subscripts α. In the present chapter, we will consider the following class of systems and cost functions.

Definition 3. *(Invariant control system with invariant costs) A control system $\mathcal{S}$ is $\mathcal{G}$-invariant with $\mathcal{G}$-invariant costs if for all $\alpha \in \mathcal{G}$, $x_k \in \mathcal{X}$, $u_k \in \mathcal{U}$ and $w_k \in \mathcal{W}$ we have:*

$$\begin{aligned}
\phi^{-1} \circ f_k(\phi(x_k), \chi(u_k), \psi(w_k)) &= f_k(x_k, u_k, w_k), \\
&\qquad k = 0, 1, \ldots, N-1 \\
g_k(\phi(x_k), \chi(u_k), \psi(w_k)) &= g_k(x_k, u_k, w_k), \\
&\qquad k = 0, 1, \ldots, N-1, \\
g_N(\phi(x_N)) &= g_N(x_N), \ \textit{and} \\
p_k(\psi(w_k))|\det D\psi(w_k)| &= p_k(w_k) \\
&\qquad k = 0, 1, \ldots, N-1.
\end{aligned}$$

The rationale is simple: For any fixed $\alpha \in \mathcal{G}$, consider the change of variables $X_k = \phi_\alpha(x_k)$, $U_k = \chi_\alpha(u_k)$, $W_k = \psi_\alpha(w_k)$. Then, we have

$$X_{k+1} = f_k(X_k, U_k, W_k), \quad k = 0, 1, \ldots, N-1,$$

and for $k = 0, 1, \ldots, N-1$ we have also $g_k(X_k, U_k, W_k) = g_k(x_k, u_k, w_k)$. As a result, if $u_1, \ldots, u_{N-1}$ is a series of controls that minimize $J(x_0)$, then one can expect $U_1, \ldots, U_{N-1}$ to minimize $J(X_0)$, under some assumptions on the noise. As a result, the optimal control problem needs only be solved once for all initial conditions belonging to the set $\{\phi_\alpha(x_0) | \alpha \in \mathcal{G}\}$, reducing the initial n dimensional problem to a $n-r$ dimensional problem. The present chapter derives a proper theory for such symmetry reduction in dynamic programming, and provides various examples of engineering interest.

Cartan's moving frame method

To find a reduced coordinate system in which to perform dynamic programming, we will use the moving frame method of Cartan [20]. In general, this method only results in a local coordinate transformation as it relies on the implicit function theorem. However, for many practical problems the transformation computed using this method extends globally. To simplify the exposure we will present a "global" version of the method.

We briefly introduce the moving frame method following the presentation in [16]. Consider an r-dimensional transformation group (with $r \leq n$) acting on $\mathcal{X}$ via the diffeomorphisms $(\phi_\alpha)_{\alpha \in \mathcal{G}}$. Assume we can split ϕ_α as $(\phi^a_\alpha, \phi^b_\alpha)$ with r and $n-r$ components respectively so that ϕ^a_α is an invertible map. Then, for some c in the range of ϕ^a, we define a coordinate cross section to the orbits $\mathcal{C} = \{x : \phi^a_e(x) = c\}$. This cross section is an $n-r$-dimensional submanifold of $\mathcal{X}$. Assume moreover that for any point $x \in \mathcal{X}$, there is a unique group element $\alpha \in \mathcal{G}$ such that $\phi_\alpha(x) \in \mathcal{C}$. Such α will be denoted $\gamma(x)$, and the map $\gamma : \mathcal{X} \to \mathcal{G}$ will be called moving frame.

A moving frame can be computed by solving the normalization equations:

$$\phi^a_{\gamma(x)}(x) = c.$$

Define the following map $\rho : \mathcal{X} \to \mathbb{R}_{n-r}$ as

$$\rho(x) = \phi^b_{\gamma(x)}(x).$$

Note that, for all $\alpha \in \mathcal{G}$ we have $\rho(\phi_\alpha(x)) = \rho(x)$, that is, the components of ρ are *invariant* to the group action on the state space. Further, due to our assumptions, the restriction of ρ to $\mathcal{C}$ is injective. We denote this restricted function $\bar{\rho}$, and it will serve as a reduced coordinate system to solve the invariant optimal control problem.

3.3 Main Results

In order to combat the "curse of dimensionality" associated with performing dynamic programming in high-dimensional systems, we describe a method to reduce the system's dimension by exploiting symmetries in the dynamics and stage costs.

Symmetries imply equivalence classes of optimal policies

Theorem 3. *(Symmetries of the optimal cost and policy) Let $\mathcal{G}$ be a group and let $\mathcal{S}$ be a $\mathcal{G}$-invariant control system with $\mathcal{G}$-invariant costs. Then the optimal cost functions $J_k(x_0)$ satisfy the symmetry relations*

$$J_k = J_k \circ \phi_\alpha$$

for any $k = 0, \dots, N$ and any $\alpha \in \mathcal{G}$. Furthermore, if $\pi^* = \{\mu_0^*, \dots, \mu_{N-1}^*\}$ is an optimal policy then so is $\tilde{\pi}^* := \{\chi_\alpha \circ \mu_0^* \circ \phi_\alpha^{-1}, \dots, \chi_\alpha \circ \mu_{N-1}^* \circ \phi_\alpha^{-1}\}$ for any $\alpha \in \mathcal{G}$.

Proof. First, note that

$$J_N(x_N) = g_N(x_N) = g_N(\phi(x_N)) = J_N(\phi(x_N)).$$

Now, suppose that for some $k \in \{0, \dots, N-1\}$ we have $J_{k+1}(x_{k+1}) = J_{k+1}(\phi(x_{k+1}))$ for all $x_{k+1} \in \mathcal{X}$. Then for any $x_k \in X$, and $u_k \in \mathcal{U}$ we have

$$\begin{aligned}
&\mathbb{E}\Big[g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))\Big] \\
&= \int_{\mathcal{W}} \Big[g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))\Big] p_k(w_k) dw_k \\
&= \int_{\mathcal{W}} \Big[g_k(\phi(x_k), \chi(u_k), \psi(w_k)) + J_{k+1}(\phi^{-1} \circ f_k(\phi(x_k), \chi(u_k), \psi(w_k)))\Big] p_k(w_k) dw_k \\
&= \int_{\mathcal{W}} \Big[g_k(\phi(x_k), \chi(u_k), \psi(w_k)) + J_{k+1}(f_k(\phi(x_k), \chi(u_k), \psi(w_k)))\Big] p_k(w_k) dw_k \\
&= \int_{\mathcal{W}} \Big[g_k(\phi(x_k), \chi(u_k), \psi(w_k)) + J_{k+1}(f_k(\phi(x_k), \chi(u_k), \psi(w_k)))\Big] p_k(\psi(w_k)) |\det D\psi(x_k)| dw_k
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\psi(\mathcal{W})} \Big[g_k(\phi(x_k), \chi(u_k), \tilde{w}_k) + J_{k+1}(f_k(\phi(x_k), \chi(u_k), \tilde{w}_k)) \Big] p_k(\tilde{w}_k) d\tilde{w}_k \\
&= \int_{\mathcal{W}} \Big[g_k(\phi(x_k), \chi(u_k), w_k) + J_{k+1}(f_k(\phi(x_k), \chi(u_k), w_k)) \Big] p_k(w_k) dw_k \\
&= \mathbb{E}\Big[g_k(\phi(x_k), \chi(u_k), w_k) + J_{k+1}(f_k(\phi(x_k), \chi(u_k), w_k)) \Big]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
J_k(x_k) &= \min_{u_k \in \mathcal{U}} \mathbb{E}\Big[g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k)) \Big] \\
&= \min_{u_k \in \mathcal{U}} \mathbb{E}\Big[g_k(\phi(x_k), \chi(u_k), w_k) + J_{k+1}(f_k(\phi(x_k), \chi(u_k), w_k)) \Big] \\
&= \min_{\tilde{u}_k \in \chi(\mathcal{U})} \mathbb{E}\Big[g_k(\phi(x_k), \tilde{u}_k, w_k) + J_{k+1}(f_k(\phi(x_k), \tilde{u}_k, w_k)) \Big] \\
&= J_k(\phi(x_k)).
\end{aligned}
$$

Thus $J^* = J^* \circ \phi$. Now, if $\pi^* = \{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is an optimal policy and we denote $\tilde{x}_k = \phi(x_k)$ then for any $k \in \{0, \ldots, N-1\}$ we have

$$
\begin{aligned}
J_k(\tilde{x}_k) &= J_k(x_k) \\
&= \mathbb{E}\Big[g(x_k, \mu_k^*(x_k), w_k) + J_{k+1}(f_k(x_k, \mu_k^*(x_k), w_k)) \Big] \\
&= \mathbb{E}\Big[g_k(\phi(x_k), \chi(\mu_k^*(x_k)), w_k) + J_{k+1}(f_k(\phi(x_k), \chi(\mu_k^*(x_k)), w_k)) \Big] \\
&= \mathbb{E}\Big[g_k(\phi(x_k), \chi \circ \mu_k^* \circ \phi^{-1}(\phi(x_k)), w_k) + J_{k+1}(f_k(\phi(x_k), \chi \circ \mu_k^* \circ \phi^{-1}(\phi(x_k)), w_k)) \Big] \\
&= \mathbb{E}\Big[g_k(\tilde{x}_k, \chi \circ \mu_k^* \circ \phi^{-1}(\tilde{x}_k), w_k) + J_{k+1}(f_k(\tilde{x}_k, \chi \circ \mu_k^* \circ \phi^{-1}(\tilde{x}_k), w_k)) \Big]
\end{aligned}
$$

Thus $\tilde{\pi}^* := \{\chi \circ \mu_0^* \circ \phi^{-1}, \ldots, \chi \circ \mu_{N-1}^* \circ \phi^{-1}\}$ is an optimal policy. □

Dynamic programming can be preformed using reduced coordinates

Theorem 3 readily implies the problem can be reduced, as all states along an orbit of $\mathcal{G}$ are equivalent in terms of cost, and that there are equivalence classes of optimal policies. So it

suffices to only consider the cost corresponding to a single representative of each equivalence class, and to find a single representative of the optimal policy within each class. This can now easily be done using the injective map $\bar{\rho} : \mathcal{C} \to \mathbb{R}^{n-r}$.

For $\bar{x} \in \bar{\rho}(\mathcal{C}) \subset \mathbb{R}^{n-r}$, let $z \in \mathcal{C}$ be such that $\bar{x} = \bar{\rho}(z)$, and define

$$\bar{J}_k(\bar{x}) = J_k(z).$$

The following result shows that the functions J_k on the n-dimensional space $\mathcal{X} \subseteq \mathbb{R}^n$ are completely determined by the values of $\bar{J}_k$ on the subset $\bar{\rho}(\mathcal{C})$ of $\mathbb{R}^{n-r}$.

Corollary 1. *For any $x \in \mathcal{X}$ and $k = 0, \dots, N$, the cost function J_k for the full problem can be computed in terms of the lower-dimensional cost function $\bar{J}_k$ as*

$$J_k(x) = J_k(\phi_{\gamma(x)}(x)) = \bar{J}_k(\bar{x}),$$

where $\bar{x} := \bar{\rho}(\phi_{\gamma(x)}(x))$ is well defined as $\phi_{\gamma(x)}(x) \in \mathcal{C}$.

It is thus sufficient to have evaluated $\bar{J}$ at all points of $\bar{\rho}(\mathcal{C}) \subset \mathbb{R}^{n-r}$ to be able to instantly evaluate J at any point of $\mathcal{X}$.

Theorem 4. *(Dynamic programming in reduced coordinates) The reduced coordinates are in one to one correspondance with the cross-section $\mathcal{C}$. For any $\bar{x}$, let $z \in \mathcal{C}$ satistisfy $\bar{\rho}(z) = \bar{x}$. Then in the reduced coordinates, the sequence $\bar{J}_k$ can be computed recursively via*

$$\bar{J}_k(\bar{x}) = \min_{u_k \in \mathcal{U}} \mathbb{E}\left[g_k(z, u_k, w_k) + \bar{J}_{k+1}(\rho(f_k(z, u_k, w_k))) \right].$$

Proof. We have

$$\begin{aligned} \bar{J}_k(\bar{x}) = J_k(z) &= \min_{u_k \in \mathcal{U}} \mathbb{E}\left[g_k(z, u_k, w_k) + J_{k+1}\Big(f_k(z, u_k, w_k) \Big) \right] \\ &= \min_{u_k \in \mathcal{U}} \mathbb{E}\left[g_k(z, u_k, w_k) + \bar{J}_{k+1}\Big(\rho \circ f_k(z, u_k, w_k) \Big) \right]. \end{aligned}$$

□

Case of equivariant costs

So far, we have considered the costs to be invariant. However, equivariance can obviously also be considered. Indeed, the cost g_k is said to be equivariant if there exists a family of diffeomorphisms φ_α such that $g_k(\phi_\alpha(x_k), \chi_\alpha(u_k), \psi_\alpha(w_k)) = \varphi_\alpha \circ g_k(x_k, u_k, w_k)$. As we want the cost function J to be equivariant too, we will need $\varphi_\alpha(\cdot)$ to be linear. Thus we will simply assume that φ_α is of the form $\varphi_\alpha(J) = l(\alpha)J$, that is, it is a scaling of the cost, where

$l : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$. For simplicity's sake, we consider here the problem to be noise free. Along the lines of the preceding sections it is easily proved that

$$J_k(\phi(x)) = \varphi \circ J_k(x)$$

as already noticed in [2] for the case of homogeneous costs. Symmetry reduction can then be applied. We now give two tutorial examples.

Example 1. *Consider the linear system*

$$x_{x+1} = Ax_k + Bu_k$$

with quadratic costs $g_k = x_k^T Q x_k + u_k^T R u_k$. *The system is invariant to scalings,* $\phi_\alpha(x) = \alpha x$, $\chi_\alpha(u) = \alpha u$, *and the cost is equivariant letting* $\varphi_\alpha(J) = \alpha J$, *where* $\alpha \in \mathcal{G} = \mathbb{R}_{>0}$. *The unit sphere is a cross section to the orbits, and the normalization equations yield* $\gamma(x) = 1/||x||$. *Applying the results above, we see that the controls that minimize* $J(x_0)$ *are* $||x_0||u_1^*, \cdots, ||x_0||u_{N-1}^*$, *where* $u_1^*, \cdots, u_{N-1}^*$ *are those minimizing* $J(\frac{x_0}{||x_0||})$. *This agrees with the well known fact that the optimal controller for the problem above is the linear quadratic controller, and is indeed of the linear form* $u_k = -F_k x_k$.

Example 2. *Consider the following system and costs*

$$x_{x+1} = Ax_k + Bu_k, \quad g_k = h(x_k) + ||u_k||_1$$

where $||u_k||_1$ *denotes the* L^1 *norm of* u_k *and* h *is a map satisfying* $h(ax) = ah(x)$ *for* $a > 0$. *Such costs may arise when one tries to force some controls to zero to create sparsity, a method known as* L^1 *regularization. This problem is challenging, particularly for nonconvex* h. *But according to the theory above, it is sufficient to solve it numerically for initial conditions lying on the unit sphere of the state space.*

3.4 Application to Optimal Experiment Design for MR Fingerprinting

Magnetic resonance imaging (MRI) has traditionally focused on acquisition sequences that are static, in the sense that sequences typically wait for magnetization to return to equilibrium between acquisitions. Recently, researchers have demonstrated promising results based on dynamic acquisition sequences, in which spins are continuously excited by a sequence of random input pulses, without allowing the system to return to equilibrium between pulses. Model parameters corresponding to T_1 and T_2 relaxation, off-resonance and spin density are then estimated from the sequence of acquired data. This technique, termed magnetic resonance fingerprinting (MRF), has been shown to increase the sensitivity, specificity and speed of magnetic resonance studies [61, 28].

This technique could be further improved by replacing randomized input pulse sequences with sequences that have been optimized for informativeness about model parameters. To this end, we present a model of MR spin dynamics that describes the measured data as a function of T_1 and T_2 relaxation rates and the sequence of radio-frequency (RF) input pulses, used to excite the spins.

The following model was introduced in the conference paper [68]. In this chapter, an optimal control was computed via dynamic programming on a very sparse six-dimensional grid. Now using our symmetry reduction technique, we we exploit symmetry reduction to provide a much more accurate optimal input sequence computed on a finer five-dimensional grid.

We model the spin dynamics via the equations

$$\mathbf{x}_{k+1} = U_k \begin{bmatrix} \theta_2 & 0 & 0 \\ 0 & \theta_2 & 0 \\ 0 & 0 & \theta_1 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} 0 \\ 0 \\ 1-\theta_1 \end{bmatrix} \tag{3.3}$$

where the states $x_{1,k}$ and $x_{2,k}$ describe the transverse magnetization (orthogonal to the applied magnetic field) and $x_{3,k}$ describes the longitudinal magnetization (parallel to the applied magnetic field). To simplify the presentation, off-resonance is neglected in this model. Control inputs $U_k \in \mathrm{SO}(3)$ describe flip angles corresponding to RF excitation pulses that rotate the state about the origin. Between acquisitions, transverse magnetization decays according to the parameter $\theta_2 = e^{-\Delta t/T_2}$ and the longitudinal magnetization recovers to equilibrium (normalized such that the equilibrium is $x_0 = [0 \quad 0 \quad 1]^T$) according to the parameter $\theta_1 = e^{-\Delta t/T_1}$ where Δt is the sampling interval.

$$\begin{aligned} f_k(x_k, U_k, w_k) &= \begin{bmatrix} & & & 0 & 0 & 0 \\ & U_k & & 0 & 0 & 0 \\ & & & 0 & 0 & 0 \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & U_k & \\ 0 & 0 & 0 & & & \end{bmatrix} \begin{bmatrix} \theta_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_2 & 0 \\ 0 & 0 & 1 & 0 & 0 & \theta_1 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 0 \\ 1-\theta_1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \\ g_k(x_k, U_k, w_k) &= -x_k^T \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\gamma} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\gamma} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} x_k \end{aligned} \tag{3.4}$$

We assume that data are acquired immediately following the RF pulse, allowing us to make a noisy measurement of the transverse magnetization. We also assume that the mea-

sured data are described by a multivariate Gaussian random variable

$$\mathbf{y}_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k + v_k$$

where v_k is a zero-mean Gaussian noise with covariance $\begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}$. This model results from a time discretization of the Bloch equations [14, 80] under a time scale separation assumption that specifies that the RF excitation pulses act on a much faster time scale than the relaxation time constants T_1 and T_2. A simplified two-state version of this model was considered in [66], where the transverse magnetization was modelled using a single state describing the magnitude of $[x_{1,k}, x_{2,k}]^T$.

We see from the model (3.3) that magnetization in the transverse direction decays while magnetization in the longitudinal direction grows. However only the transverse component of the magnetization can be measured. Thus there is a trade-off between making measurements (which leads to loss of magnetization) and magnetization recovery. This is the trade-off that we hope to manage through the optimal design of an input sequence U_k.

We wish to quantify the informativeness of an acquisition sequence based on the information about the T_1 relaxation parameter θ_1 that is contained in the resulting data set. More formally, we wish to choose $U_k \in \mathrm{SO}(3)$ to maximize the Fisher information about θ_1 contained in the joint distribution of $Y = (\mathbf{y}_0, \ldots, \mathbf{y}_N)$. The Fisher information $\mathcal{I}$ can be expressed as a quadratic function of the sensitivities of $\mathbf{x}_k$ with respect to θ_1:

$$\mathcal{I} = \sum_{k=0}^{N} \frac{\partial}{\partial \theta_1} \mathbf{x}_k^T \begin{bmatrix} 1/\gamma & 0 & 0 \\ 0 & 1/\gamma & 0 \\ 0 & 0 & 0 \end{bmatrix} \frac{\partial}{\partial \theta_1} \mathbf{x}_k$$

where the sensitivities $\frac{\partial}{\partial \theta_1} \mathbf{x}_k$ satisfy the following sensitivity equations:

$$\frac{\partial}{\partial \theta_1} \mathbf{x}_{k+1} = U_k \begin{bmatrix} \theta_2 & 0 & 0 \\ 0 & \theta_2 & 0 \\ 0 & 0 & \theta_1 \end{bmatrix} \frac{\partial}{\partial \theta_1} \mathbf{x}_k + U_k \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

It should be noted that for system (3.3), the objective function $\mathcal{I}$ has many local optima as a function of the input sequence U_k. Thus, in contrast with [67] which consider optimal experiment design for hyperpolarized MRI problems, for this model, local search methods provide little insight into what acquisition sequences are good. In contrast with the MRI model presented in [64], where global optimal experiment design heuristics are developed for linear dynamical systems, in this model the decision variables U_k multiply the state vector $\mathbf{x}_k$, making the output $\mathbf{y}_k$ a nonlinear function of the sequence $U = (U_0, \ldots U_{k-1})$. Thus we must use dynamic programming to find a solution.

Model

To present this problem in the formalism we have introduced, we define an augmented state vector

$$x_k = \begin{bmatrix} \mathbf{x}_k \\ \frac{\partial}{\partial \theta_1} \mathbf{x}_k \end{bmatrix} \in \mathbb{R}^6.$$

We can write the dynamics of the augmented state as a control system with f and g defined in Equation (3.4). This system has a one-dimensional group of symmetries defined by

$$\begin{aligned}
\phi_\alpha(x_k) &= \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 & 0 & 0 & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos(\alpha) & -\sin(\alpha) & 0 \\ 0 & 0 & 0 & \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} x_k \\
\chi_\alpha(U_k) &= \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} U_k \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
\psi_\alpha(w_k) &= w_k
\end{aligned}$$

for any $\alpha \in \mathbb{R}/2\pi\mathbb{Z}$.

Dynamic programming in reduced coordinates

To perform dynamic programming in a reduced coordinate system, we begin by defining the cross-section $\mathcal{C} = \{x : x_1 = 0, x_2 > 0\}$, and computing the moving frame $\gamma(x)$. To do so, we solve

$$0 = \phi^a_{\gamma(x)}(x) = x_1 \cos\gamma(x) - x_2 \sin\gamma(x).$$

Isolating γ yields

$$\gamma(x) = \operatorname{atan2}(x_1, x_2).$$

Next, we compute the invariants $\rho(x)$ using

$$\begin{aligned}
\rho(x) &= \phi^b_{\gamma(x)} \\
&= \begin{bmatrix} \sin(\text{atan2}(x_1,x_2)) & \cos(\text{atan2}(x_1,x_2)) & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos(\text{atan2}(x_1,x_2)) & -\sin(\text{atan2}(x_1,x_2)) & 0 \\ 0 & 0 & 0 & \sin(\text{atan2}(x_1,x_2)) & \cos(\text{atan2}(x_1,x_2)) & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} x \\
&= \begin{bmatrix} \sqrt{x_1^2+x_2^2} \\ x_3 \\ \frac{1}{\sqrt{x_1^2+x_2^2}}(x_2x_4 - x_1x_5) \\ \frac{1}{\sqrt{x_1^2+x_2^2}}(x_1x_4 + x_2x_5) \\ x_6 \end{bmatrix}
\end{aligned}$$

Further, ρ restricted to the cross-section $\mathcal{C}$ is injective with inverse $\bar{\rho}^{-1} : \mathbb{R}_+ \times \mathbb{R}^4 \to \mathcal{C}$ given by $\bar{\rho}^{-1}(\bar{x}) = \begin{bmatrix} 0 & \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \bar{x}_4 & \bar{x}_5 \end{bmatrix}^T$. The theory above tells us we can thus solve the optimal stochastic control problem in a 5 dimensional state space, reducing the original 6 dimensional problem of 1 dimension.

Results

To implement this algorithm, we discretize the reduced five-dimensional state space and two-dimensional input space via grids of size $6 \times 10 \times 15 \times 15 \times 15$ and 16×8 respectively. The code was written in the Julia language and parallelized to allow evaluation of J_k in parallel across grid points [66]. The implementation is publicly available at `https://github.com/maidens/Automatica-2017`.

Optimal input and state trajectories for the model corresponding to the initial condition at the equilibrium $x_0 = [0 \;\; 0 \;\; 1 \;\; 0 \;\; 0 \;\; 0]^T$ are plotted in Figures 3.1 and 3.2.

In contrast with the results from [66] where we considered a simplified version of the model, for this full model we no longer find that the optimal flip angle sequence converges to a cyclic pattern, rather it appears irregular. However, state sequence of longitudinal magnetizations and transverse magnetization magnitudes appears to converge to a constant sequence. This is likely because in this work we assumed Gaussian noise in the inputs in contrast with the Rician noise assumed in the previous work, therefore it is no longer necessary to conserve magnetization across multiple time steps before generating a reliable measurement.

3.5 Optimal Formation Control on Lie Groups

We now apply the theory presented in Section 3.2 where the state space $\mathcal{X} \subseteq \mathbb{R}^n$ is the Cartesian product of matrix Lie groups. Note that, straightforward modifications arise

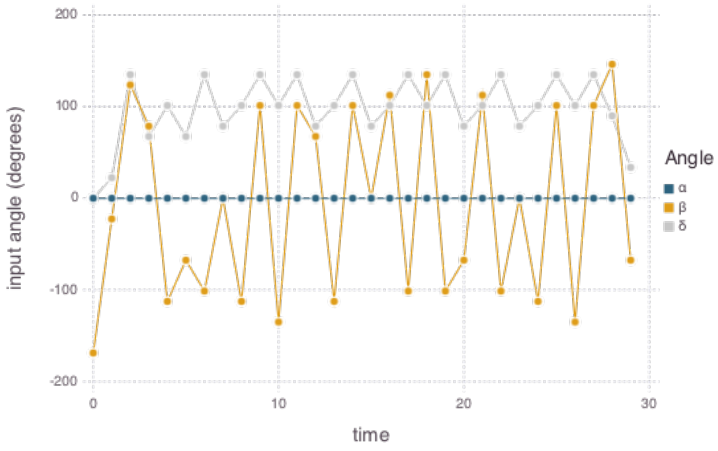
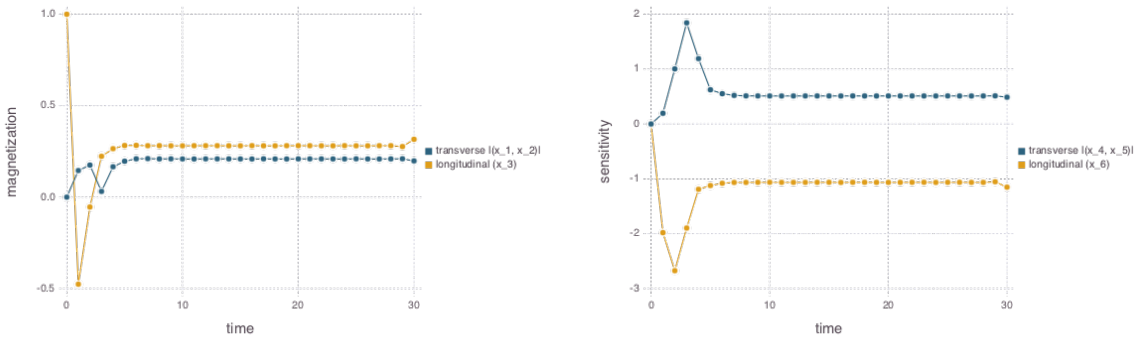


Figure 3.1: Optimal input sequence for the MR fingerprinting model. The angles α, β and δ represent rotations about the z, y and x axes respectively, resulting in an control input $U_k = R_z(\alpha_k)R_y(\beta_k)R_x(\gamma_k)$.



(a) Magnetizations (b) Sensitivities

Figure 3.2: Optimal state sequence for the MR fingerprinting model. Here we have plotted the longitudinal and transverse components of both the magnetization (states x_1, x_2, and x_3) and the sensitivities (states x_4, x_5, and x_6) where the transverse component is computed as the Euclidean norm of the vectors (x_1, x_2) and (x_4, x_5) respectively.

along the way as the state space and noise space are not vector spaces as in the theory above. The methodology is then applied to the synchronization of two non-holonomic cars in the presence of uncertainties.

General problem

We model the system as a collection of K agents, where the state of each agent evolves on a r-dimensional matrix Lie group $\mathcal{G}$. We assume that the evolution of the state of agent j proceeds according to the equation

$$X_{k+1}^j = X_k^j M(u_k^j) W_k^j \tag{3.5}$$

where X_k, $M(u_k)$, W_k are all square matrices belonging to $\mathcal{G}$, u_k is a control that lives in some finite dimensional vector space, and W_k is the noise. The control objective is to reach a desired configuration, that is, a desired value for the relative configurations of the agents $(X^1)^{-1}X^2, \ldots, (X^{K-1})^{-1}X^K$, see e.g., [93] for more information.

Systems of this form are naturally invariant to left multiplication of all X^j by some matrix $A \in \mathcal{G}$:

$$\phi_A(X) = \begin{bmatrix} AX^1 \\ \vdots \\ AX^K \end{bmatrix}$$

where $X = (X^1, \ldots, X^k) \in \mathcal{G}^K$. Letting $\chi(u^1, \ldots, u^K) \equiv (u^1, \ldots, u^K)$, $\psi(W^1, \ldots, W^K) \equiv (W^1, \ldots, W^K)$, and the costs be of the form $\tilde{g}((X^1)^{-1}X^2, \ldots, (X^{K-1})^{-1}X^K) + h(u^1, \ldots, u^K)$, we get an invariant system with invariant costs.

One can define a cross section to the orbits by letting the first agent coordinates be equal to the identity matrix, that is, $\mathcal{C} = \{X \in \mathcal{G}^K : X^1 = I\}$. The normalization equation are given by $I = \phi^a_{\gamma(X)}(X) = \gamma(X)X^1$, hence the moving frame is given by $\gamma(X) = (X^1)^{-1}$. The invariants are computed as

$$\rho(X) = \phi^b_{\gamma(X)} = \begin{bmatrix} (X^1)^{-1}X^2 \\ \vdots \\ (X^1)^{-1}X^K \end{bmatrix}$$

The optimal stochastic control problem can then be solved in the reduced coordinate system defined by ρ, reducing the state space from dimension Kr to $(K-1)r$.

Application: cooperative formation control for two stochastic Dubins vehicles

We consider two identical Dubins vehicles each with dynamics

$$\begin{aligned} z_{k+1} &= z_k + v_k \cos\theta_k \\ y_{k+1} &= y_k + v_k \sin\theta_k \\ \theta_{k+1} &= \theta_k + \frac{1}{L} v_k \tan s_k + w_k \end{aligned}$$

where y_k and z_k denote the two-dimensional position of the vehicle, θ_k denotes the heading of the vehicle, v_k is a velocity input, s_k is a steering angle input, and w_k is independent, identically-distributed zero-mean Gaussian noise with variance σ^2, and L is a parameter that determines the vehicle's steering radius.

These dynamics can be embedded in the three-dimensional special Euclidean matrix Lie group $\mathcal{G} = SE(2)$, by defining the state

$$X_k = \begin{bmatrix} \cos\theta_k & -\sin\theta_k & z_k \\ \sin\theta_k & \cos\theta_k & y_k \\ 0 & 0 & 1 \end{bmatrix},$$

input matrix

$$M(v_k, s_k) = \begin{bmatrix} \cos(\frac{1}{L} v_k \tan s_k) & -\sin(\frac{1}{L} v_k \tan s_k) & v_k \\ \sin(\frac{1}{L} v_k \tan s_k) & \cos(\frac{1}{L} v_k \tan s_k) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and noise matrix

$$W_k = \begin{bmatrix} \cos w_k & -\sin w_k & 0 \\ \sin w_k & \cos w_k & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

with state update equation of the form (3.5).

We wish to compute a control policy for a two-vehicle system, with states X^1 and X^2, where the controls can only take a finite number of values, and with terminal cost

$$J(X_0^1, X_0^2) = \mathbb{E}\left[g_N\big((X_N^1)^{-1} X_N^2\big) \right]$$

where $g_N(X) = \arccos(X_{11})^2 + |\sqrt{X_{13}^2 + X_{23}^2} - 1|$, that is, we want the vehicles to have the same heading, and follow each other at unit distance. Thanks to the theory developed above, the stochastic control problem is reduced from problem with a six dimensional state space to a problem with a three dimensional state space only.

For numerical simulations, the cost functions $\bar{J}_k$ were computed on a fixed grid of dimension $51 \times 51 \times 65$ using turning radius parameter $L = 1$, input sets $v_k \in \{-0.1, 0, -0.1\}$ and $s_k \in \{-1, 0, -1\}$ Globally optimal input and state trajectory sequences corresponding to the initial condition $x_0 = \begin{bmatrix} 0.1 & 0 & \frac{1}{2}\pi & -0.1 & 0 & \frac{3}{2}\pi \end{bmatrix}^T$ are shown in Figures 3.3 and 3.4. These are compared against a deterministic version of the model with $w_k = 0$ in Figures 3.5 and 3.6.

3.6 Conclusion

We have presented a method of reducing the complexity of dynamic programming for systems in which the state dynamics, stage costs and transition probabilities are invariant under a

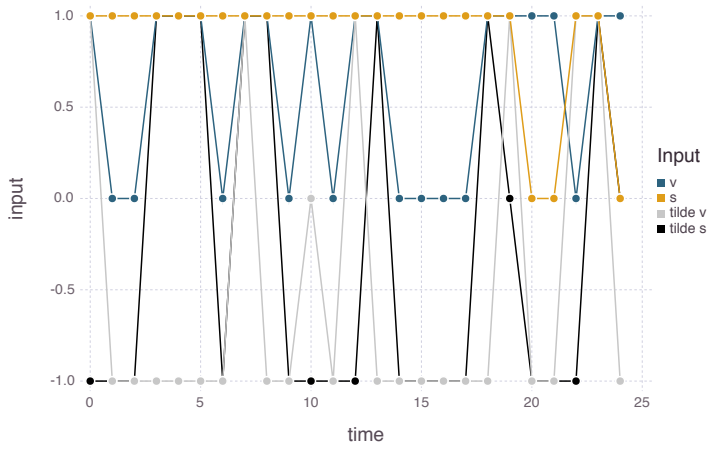


Figure 3.3: Optimal input sequence for cooperative stochastic Dubins vehicle model with $\sigma = 0.3$.

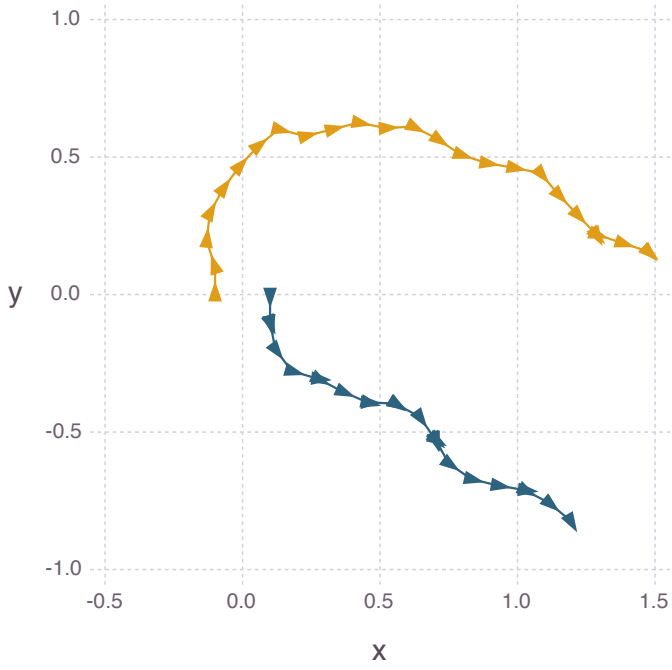


Figure 3.4: Optimal state sequence for the cooperative stochastic Dubins vehicle model with $\sigma = 0.3$.

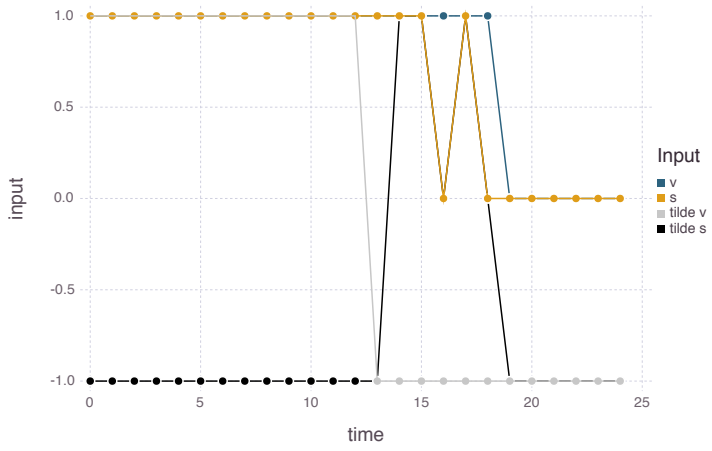


Figure 3.5: Optimal input sequence for cooperative deterministic Dubins vehicle model.

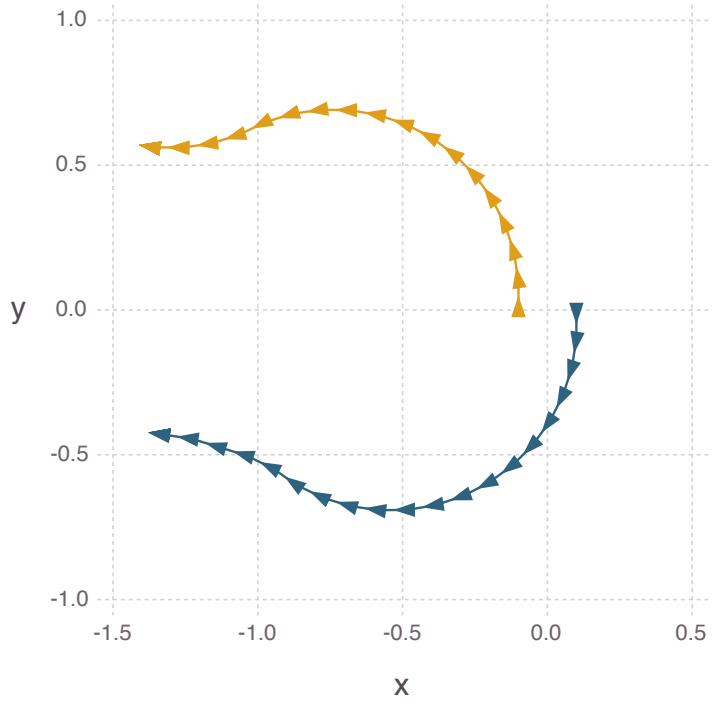


Figure 3.6: Optimal state sequence for the cooperative deterministic Dubins vehicle model.

group of symmetries. This allows us to compute globally optimal control policies for systems of moderate state dimension. We have applied this technique to compute globally optimal

trajectories to a six-dimensional original MRI model with a one-dimensional group of symmetries and for a six-dimensional stochastic Dubins vehicle model with a three-dimensional group of symmetries by reducing the dimension of the state space to five and three dimensions respectively. Since computation time for dynamic programming depends exponentially on the state space dimension, this technique enables the computation of optimal control policies for systems in which it was previously infeasible.

Chapter 4

Optimizing Flip Angles for Metabolic Rate Estimation in Hyperpolarized Carbon-13 MRI

Hyperpolarized carbon-13 magnetic resonance imaging (MRI) has enabled the real-time observation of perfusion and metabolism in preclinical and clinical studies [37, 29, 52, 77, 5, 101]. This technology is made possible by techniques for dynamic nuclear polarization (DNP) that have led to signal-to-noise ratio (SNR) increases of four to five orders of magnitude compared with endogenous signal in dissolved ^{13}C-labelled molecules [4, 38]. Injected [1-^{13}C] pyruvate is frequently used as a substrate in metabolism experiments and its rate of conversion to [1-^{13}C] lactate has been shown to distinguish between healthy and diseased tissues in animal [29], and recently human [77], studies.

In contrast with conventional MRI, magnetization is a non-renewable resource in hyperpolarized MRI. Conventional imaging relies only on thermal equilibrium polarization, therefore an arbitrary number of acquisitions can be performed if we allow time for the magnetization to return to equilibrium between acquisitions. In contrast, hyperpolarization can only be performed before a ^{13}C-labeled substrate is injected into the body, and once injected the magnetization decays due to T_1 relaxation and rapid T_2 relaxation following radio frequency (RF) excitation. Thus, the choice of excitation sequence is important for managing the trade-off between present and future measurement quality.

In typical practice a constant flip angle sequence is used for excitation, with typical values ranging from 5–30 degrees. Alternative time-varying acquisition sequences include sequences that attempt to maintain constant observed signal over time [106], maximize the cumulative observed lactate signal over time [62], or saturate the lactate signal in each acquisition [96].

In this chapter our goal is to design a time-varying flip angle sequence to achieve maximally reliable quantitative estimates of the metabolic rate that can be compared between tissue regions, across subjects, or over time. To achieve this we develop a statistical model of the observed data as a function of the flip angle sequence and design flip angles to maximize the Fisher information about the metabolic rate parameter.

We begin by presenting a mathematical model of the magnetization dynamics in the observed tissue in Section 4.1. In Section 4.2, we introduce a flip angle optimization procedure and present an optimal sequence. Next, we validate this result with computer simulation studies to demonstrate that our optimized sequence yields more reliable metabolic rate estimates than commonly-used flip angle sequences in Section 4.3. Finally, in Section 4.4, we demonstrate the feasibility of this procedure *in vivo* and demonstrate the well-foundedness of our mathematical model with experiments in a prostate cancer mouse model.

The software and experimental data required to generate the figures in this chapter are available at: `https://github.com/maidens/TMI-2015`.

4.1 Mathematical Model

Linear differential equations are well-established as models of metabolic flux measured using MR spectroscopy [29, 18]. It is shown in [47] that a first-order, two-site model with unidirectional flux of pyruvate to lactate is sufficient to accurately model the appearance of lactate, when observed by hyperpolarized MR. This work also showed that increasing the fidelity of the model to incorporate bidirectional flux, or transport of lactate outside the cell, did not significantly improve the fit to hyperpolarized MR data. Therefore, we consider a two-dimensional system of ordinary differential equations

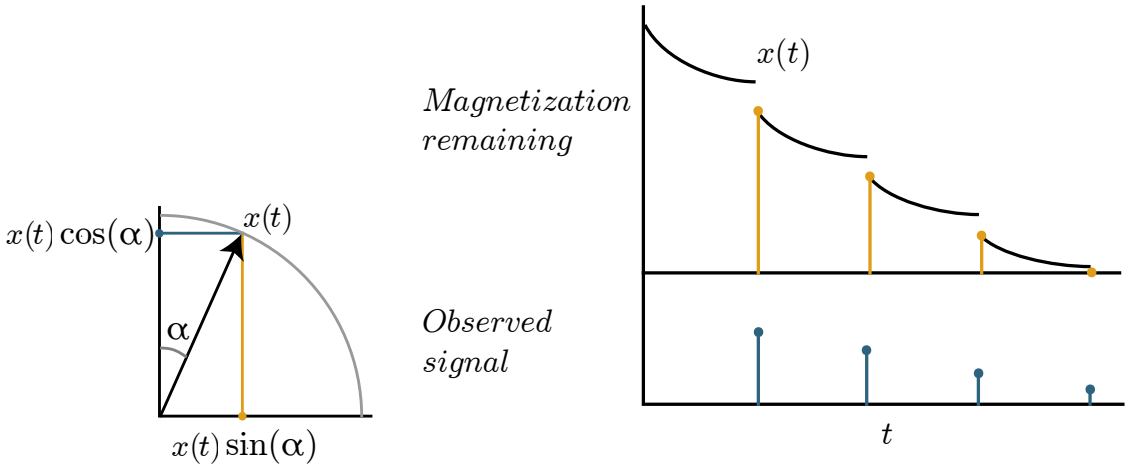
$$\frac{dx}{dt}(t) = \begin{bmatrix} -k_{PL} - R_{1P} & 0 \\ k_{PL} & -R_{1L} \end{bmatrix} x(t) + \begin{bmatrix} k_{TRANS} \\ 0 \end{bmatrix} u(t) \tag{4.1}$$

that models the magnetization dynamics in a tissue with an arterial input function $u(t)$ and uni-directional conversion from the substrate (pyruvate) to a metabolic product (lactate), which has been commonly applied for hyperpolarized ^{13}C pyruvate experiments. The state $x_1(t)$ denotes the longitudinal magnetization of pyruvate contained in a particular voxel in the tissue and $x_2(t)$ the longitudinal magnetization of lactate in the same voxel. The rate of metabolism of pyruvate to lactate is denoted k_{PL}, the perfusion rate from the arterial input to the tissue is denoted k_{TRANS}, and R_{1P} and R_{1L} are lumped parameters that account for T_1 decay in the magnetization along with other effects, such as metabolism of pyruvate into products other than lactate as well as flow of magnetization out of the slice. The input to the system $u(t)$ is an unmeasured arterial input function (AIF) resulting from the injection of hyperpolarized [1-^{13}C] pyruvate. In an experimental setting an AIF will be estimated based on the data collected, but for the purposes of designing a flip angle sequence, it will be assumed to be of gamma-variate shape

$$u(t) = A_0(t - t_0)^{\gamma} e^{-(t-t_0)/\beta}$$

with parameters t_0, γ, β, A_0 given in Table 4.1.

We acquire data at N time points separated by intervals of length T_R. Each time t an acquisition is made, we must choose a flip angle $\alpha_{k,t}$ for each compound k to be measured.



(a) Flip angle selection (b) Hybrid (continuous-discrete) dynamics

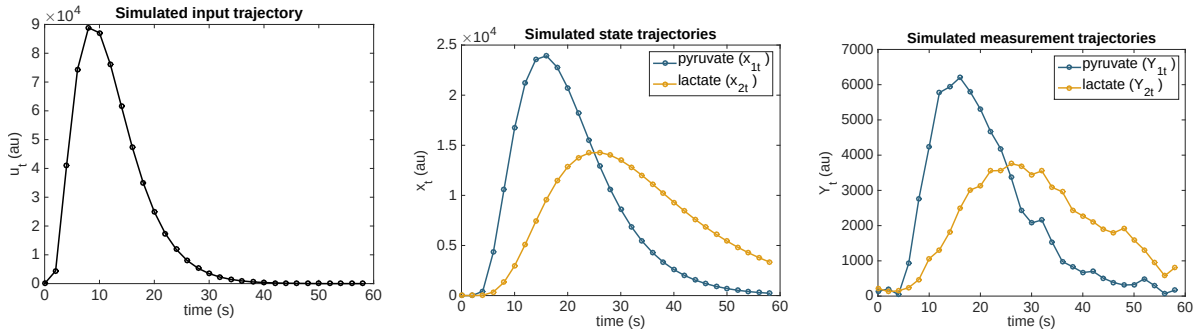
Figure 4.1: Illustration of the trade-off between present and future image intensity in a single compound. (a) Each acquisition relies on choosing an angle α to perturb the longitudinal magnetization into the transverse plane, allowing a measurement of magnitude $x(t)\sin(\alpha)$, after which $x(t)\cos(\alpha)$ longitudinal magnetization remains for future acquisitions. (b) Repeated excitation leads to repeated discrete jumps in the system state, depleting the remaining magnetization.

If the magnetization of the k-th compound before the acquisition is x_k, then this choice of flip angle allows us to measure a signal of magnitude $\sin(\alpha_{k,t})x_k$, after which $\cos(\alpha_{k,t})x_k$ magnetization remains for future acquisitions (Fig. 4.1a). This causes discrete jumps, or resets, in the system state, leading to a hybrid dynamical system [60] (Fig. 4.1b). Since we are only interested in the system's state at acquisition times, we can avoid technicalities associated with hybrid system modelling by discretizing the system in time and considering a discrete-time dynamical system that simultaneously captures the evolution of (4.1) between acquisitions and the discrete jumps induced by the acquisitions. We define the transition matrices A_d and B_d

$$A_d = \exp\left(T_R \begin{bmatrix} -k_{PL} - R_{1P} & 0 \\ k_{PL} & -R_{1L} \end{bmatrix}\right)$$
$$B_d = \begin{bmatrix} -k_{PL} - R_{1P} & 0 \\ k_{PL} & -R_{1L} \end{bmatrix}^{-1} (A_d - I) \begin{bmatrix} k_{TRANS} \\ 0 \end{bmatrix}$$

that correspond to the discretization of (4.1) assuming a zero-order hold on the input between each acquisition [22].

We will construct metabolite maps using magnitude image data, which necessitates a Rician noise model. Using magnitude images allows us to avoid modelling sources of phase in the image which would require additional states and parameters to estimate. It would also



(a) input u (b) states M_z (c) Noisy measurements Y

Figure 4.2: Simulated trajectories for a pyruvate to lactate conversion model using a constant flip angle sequence with $\alpha_S[k] = \alpha_P[k] = 15°$.

be possible to perform the parameter mapping using complex images with Gaussian noise, but this would require revising the model to account for phase, or modifying the image reconstruction to estimate and remove phase from the acquired images.

Accordingly, we model the measurements as independent Rician-distributed random variables [44], which have probability density

$$p_{x,\sigma}(y) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2+x^2}{2\sigma^2}\right) I_0\left(\frac{yx}{\sigma^2}\right)$$

where I_ν denotes the modified Bessel function of the first kind of order ν. All together, we have the discrete-time model

$$\begin{aligned} x_0 &= 0 \\ x_{t+1} &= A_d(\theta) \begin{bmatrix} \cos\alpha_{1,t} & 0 \\ 0 & \cos\alpha_{2,t} \end{bmatrix} x_t + B_d(\theta) u_t(\theta) \\ \tilde{x}_{k,t} &= \sin(\alpha_{k,t}) x_{k,t} \quad k = 1,2 \\ Y_{k,t} &\sim Rice(\tilde{x}_{k,t}, \sigma_k) \quad k = 1,2. \end{aligned} \tag{4.2}$$

Simulated trajectories of this model are shown in Fig. 4.2.

The model parameters are

$$\theta = [R_{1P}, R_{1L}, k_{PL}, k_{TRANS}, u_0, \ldots, u_{N-1}]$$

and we have the freedom to choose

$$\alpha = \begin{bmatrix} \alpha_{1,1} & \ldots & \alpha_{1,N} \\ \alpha_{2,1} & \ldots & \alpha_{2,N} \end{bmatrix}$$

to generate the best possible estimate of the unknown parameters. The noise parameters σ_k for $k = 1, 2$ can be estimated separately from a measurement of the background and are therefore assumed to be known. We fix a sampling interval of $T_R = 2$ seconds, though this could in principle be included as a decision variable.

4.2 Optimal Experiment Design

In this section, we present a methodology for designing flip angle sequences to provide maximal information about the metabolic rate parameter k_{PL}. This methodology is based on an optimization problem that can be solved to local optimality using nonlinear programming. To facilitate the implementation of this methodology, we have released an open source MATLAB toolbox for the design of optimal flip angle sequences. This toolbox is available at `https://github.com/maidens/Flip-Angle-Design-Toolbox`.

Theory

The goal of optimal experiment design is to estimate parameters in a statistical model from observed data with minimum variance in the estimates [87]. The most commonly used optimization criteria are scalar-valued functions of the Fisher information matrix. For a statistical model described by a family of probability density functions $p_\theta(y)$ parametrized by a vector of parameters $\theta \in \mathbb{R}^p$, the Fisher information is a $p \times p$ symmetric, positive semidefinite matrix defined as

$$\mathcal{I}_{ij}(\theta) = \mathbb{E}\left[\frac{\partial \log p_\theta(Y)}{\partial \theta_i}\frac{\partial \log p_\theta(Y)}{\partial \theta_j}\right]$$

where the expectation is computed under the distribution p_θ.

In linear regression models with Gaussian-distributed measurements, the maximum-likelihood estimator of θ is unbiased and has covariance equal to the inverse of the Fisher information matrix. Thus the variance of estimates of θ can be minimized by maximizing the Fisher information. For general models, the Cramér-Rao inequality

$$\mathrm{cov}(\hat{\theta}) \geq \mathcal{I}^{-1} \tag{4.3}$$

gives a lower bound on the covariance of any unbiased estimator $\hat{\theta}$ of the parameter θ in terms of the Fisher information matrix. In general, finite-sample efficient estimators do not exist, that is, there is no estimator that can achieve the Cramér-Rao bound based on a single experiment, or even a finite number of independent experiments. Nonetheless, the Fisher information is commonly used in optimal experiment design in nonlinear models as it is general, easy to compute, provides good results in practice and can be justified mathematically via asymptotic analysis[105].

In this chapter, we will estimate unknown model parameters using the maximum likelihood estimator (MLE) defined as

$$\hat{\theta}_{MLE} \in \arg\max_{\theta \in \mathbb{R}^p} p_\theta(Y)$$

where Y is the observed data. Under mild assumptions, the maximum likelihood estimator is asymptotically efficient [25], that is,

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}) \tag{4.4}$$

as the number n of independent data sets used to compute the MLE tends to infinity. Here $\mathcal{N}(0, \mathcal{I}^{-1})$ denotes the multivariate normal distribution with mean zero and covariance $\mathcal{I}^{-1}$ and $\xrightarrow{d}$ denotes convergence in distribution. Thus, we see that asymptotically the MLE achieves the Cramér-Rao bound. While it is not necessarily the case in general that the MLE based on a single data set, or any finite number of data sets, has covariance equal to $\mathcal{I}^{-1}$, it is often a good approximation. Thus, we attempt to design an experiment to reliably estimate θ by maximizing the Fisher information.

The Fisher information is a function of the true value of the parameter vector θ, which is unknown *a priori*. There are commonly three approaches to overcome this difficulty:

1. Minimax/maximin optimal design: Given a range Θ of values for the parameter vector choose the design α to maximize the worst-case information among all potential values of the parameter $\theta \in \Theta$ (i.e. $\alpha^* = \arg\max_{\alpha \in \mathcal{A}} \min_{\theta \in \Theta} \mu(\mathcal{I}(\theta, \alpha))$ for some measure μ of the size of the information matrix). This approach is advantageous because it provides a guaranteed lower bound on the information gained from an experiment despite parametric uncertainty. But it typically leads to a design α^* that is optimized for a "corner case" in the parameter space, which may be overly conservative in practice. In addition, minimax/maximin objective functions are non-differentiable and can be difficult to optimize numerically [82, 31].

2. Bayesian optimal design: Given a prior distribution $p_0(\theta)$ on the space of possible parameter values, maximize the expected information (i.e. $\alpha^* = \arg\max_{\alpha \in \mathcal{A}} \int \mu(\mathcal{I}(\theta, \alpha)) p_0(\theta) d\theta$). This approach handles parametric uncertainty nicely, but can lead to difficulties when it in unclear how to choose an appropriate prior, or when the parameter space is large and hence the computation of the high-dimensional integral is numerically intractable [21].

3. Choose a nominal value θ_0 of the parameter vector at which to optimize the information (i.e. $\alpha^* = \arg\max_{\alpha \in \mathcal{A}} \mu(\mathcal{I}(\alpha, \theta_0))$). This approach is conceptually simple and numerically tractable, but may suffer from a lack of robustness to the choice of θ_0.

In this dissertation, we take the third approach. We address the potential lack of robustness to the choice of the nominal parameter value θ_0 by a systematic numerical study of the robustness to parametric uncertainty in Section 4.3. These experiments demonstrate that for this particular model, an experiment designed using nominal values of the model parameters performs well across a wide range of values of the true parameter. It is possible that the results presented here could be improved further based on a minimax or Bayesian formulation, but this investigation is beyond the scope of our current study.

Computing Fisher information for our model

To find a maximum-likelihood estimate with minimum variance, we choose the sequence α to maximize the Fisher information matrix at a nominal value of the parameter vector θ.

The nominal parameter values used are given in Table 4.1. The T_1 relaxation, perfusion and metabolic rate parameters were chosen based on our typical data in a prostate cancer mouse model, and nominal values for the input shape and noise parameters were chosen based on a maximum likelihood fit to an arterial input function measured in a preliminary experiment.

Table 4.1: Nominal parameter values used to compute optimal flip angle sequences

R_{1P}	R_{1L}	k_{PL}	k_{TRANS}	t_0	γ	β	A_0	σ_k
1/20	1/20	0.07	0.055	3.2596	2.1430	3.4658	1.0411 $\times 10^4$	2.3608 $\times 10^4$

To compute the Fisher information, we use the expression derived in [63] for the (i,j)-th entry of $\mathcal{I}$:

$$\mathcal{I}_{ij} = \sum_{t=0}^{N} \sum_{k=1}^{2} \frac{1}{\sigma_k^2} \frac{\partial \tilde{x}_{k,t}}{\partial \theta_i} \frac{\partial \tilde{x}_{k,t}}{\partial \theta_j} \psi\left(\frac{\tilde{x}_{k,t}}{\sigma_k}\right) \tag{4.5}$$

where the sensitivities are computed recursively as

$$\begin{cases} \frac{\partial x_0}{\partial \theta_i} = 0 \\ \frac{\partial x_{t+1}}{\partial \theta_i} = \frac{\partial A_d}{\partial \theta_i} \begin{bmatrix} \cos\alpha_{1,t} & 0 \\ 0 & \cos\alpha_{2,t} \end{bmatrix} x_t \\ \qquad\qquad + A_d \begin{bmatrix} \cos\alpha_{1,t} & 0 \\ 0 & \cos\alpha_{2,t} \end{bmatrix} \frac{\partial x_t}{\partial \theta_i} + \frac{\partial B_d}{\partial \theta_i} u_t + B_d \frac{\partial u_t}{\partial \theta_i} \\ \frac{\partial \tilde{x}_t}{\partial \theta_i} = \begin{bmatrix} \sin\alpha_{1,t} & 0 \\ 0 & \sin\alpha_{2,t} \end{bmatrix} \frac{\partial x_t}{\partial \theta_i} \end{cases}$$

and ψ is defined in terms of the integral

$$\psi(z) = -z^2 + \int_0^\infty y^3 \frac{I_1^2(yz)}{I_0(yz)} \exp\left(-\frac{1}{2}(y^2+z^2)\right) dy.$$

Eliminating nuisance parameters

In practice, we do not necessarily need good estimates of all the unknown parameters in the model. For example, in this chapter our primary goal is to estimate the metabolic rate parameter k_{PL} which is useful for discriminating between cancerous and non-cancerous tissues [5], determining the severity of disease [1, 95] and monitoring response to therapy [29]. Thus we wish to modify our optimality criterion to maximize the sensitivity of the experiments to k_{PL} while considering the nuisance parameters only insofar as they allow us to estimate the parameters of interest. We do so by partitioning the information matrix as

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix}$$

where the first block corresponds to the parameters of interest and the second block corresponds to the nuisance parameters. The inverse of the Fisher information is given by

$$\mathcal{I}^{-1} = \begin{bmatrix} (\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21})^{-1} & -\mathcal{I}_{11}^{-1}\mathcal{I}_{12}(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \\ -\mathcal{I}_{22}^{-1}\mathcal{I}_{21}(\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21})^{-1} & (\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \end{bmatrix}.$$

Thus, optimal design for the parameters of interest can be performed by maximizing the Schur complement of $\mathcal{I}_{22}$:

$$\mathcal{S} = \mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21}$$

which corresponds to minimizing the asymptotic covariance of the marginal distribution of the MLE corresponding to the parameters of interest via (4.4), or equivalently, minimizing the Cramér-Rao bound on the parameters of interest via (4.3).

In general, if multiple parameters are of interest then $\mathcal{S}$ will be a matrix and we would be required to choose a suitable scalar criterion for measuring the size of $\mathcal{S}$. The problem of simultaneously estimating k_{PL} and k_{TRANS} is considered in [65], where the $D-$, $E-$ and $A-$ optimality criteria are compared. However, in this instance we are considering a single parameter of interest k_{PL}, therefore the Schur complement $\mathcal{S}$ is scalar-valued.

Regularization

We desire a smoothly-varying sequence of flip angles for a number of reasons including increasing robustness against model mismatch and interpretability of the resulting sequence of flip angles. We achieve smoothness in the flip angle sequence by adding a regularization term $\lambda\|\Delta\alpha\|_F$ to the objective function to penalize nonsmooth sequences where the differencing operator Δ is defined as

$$\begin{aligned} &\Delta\left(\begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,N} \\ \alpha_{2,1} & \dots & \alpha_{2,N} \end{bmatrix}\right) \\ &= \begin{bmatrix} (\alpha_{1,2} - \alpha_{1,1}) & \dots & (\alpha_{1,N} - \alpha_{1,N-1}) \\ (\alpha_{2,2} - \alpha_{2,1}) & \dots & (\alpha_{2,N} - \alpha_{2,N-1}) \end{bmatrix} \end{aligned}$$

and $\|\cdot\|_F$ denotes the Frobenius norm. The nonnegative parameter λ can be adjusted to achieve the desired degree of smoothness.

Regularization also improves the convexity of the objective function. When the regularization parameter takes the value $\lambda = 0$, the objective function possesses multiple local minima, but as $\lambda \to \infty$ the convex term $\|\Delta\alpha\|_F$ dominates, which guarantees that any stationary point is a global optimum.

We chose a particular value of λ by comparing the flip angle sequences resulting from the optimization using numerous regularization parameter values. The value $\lambda = 0.1$ was found to nicely balance between smoothness and range (between 0° and 90°) of the flip angle sequence.

Numerical optimization

To design an optimal flip angle scheme, we must solve the flip angle optimization problem

$$\begin{aligned} &\text{maximize} && \mathcal{S}(\theta, \alpha) - \lambda \|\Delta\alpha\|_F \\ &\text{subject to} && 0^\circ \leq \alpha_{k,t} \leq 90^\circ \quad k = 1, 2 \;\; t = 1, \ldots, N \end{aligned} \tag{4.6}$$

for the flip angle sequence α where θ is fixed to some nominal value for the unknown parameters. The MATLAB Optimization Toolbox [102] provides a derivative-free implementation of the quasi-Newton optimization algorithm of Broyden-Fletcher-Goldfarb-Shanno (BFGS) [81], which is well-suited to finding local optima of this objective function.

Results

A solution to the optimization problem (4.6), initialized at $\alpha_{k,t} = 5^\circ$, is given in Fig. 4.3. Simulated state and observation trajectories corresponding to this flip angle sequence are shown in Fig. 4.4.

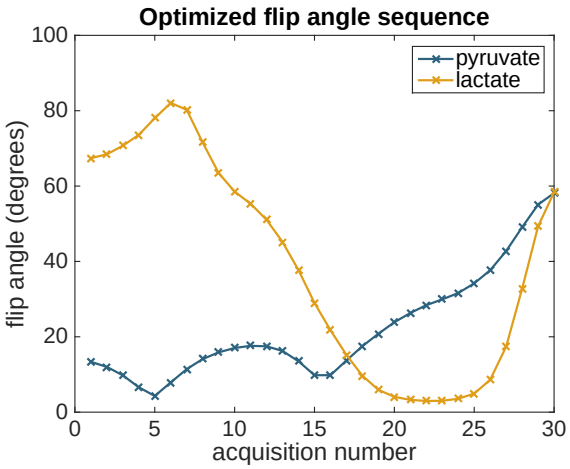
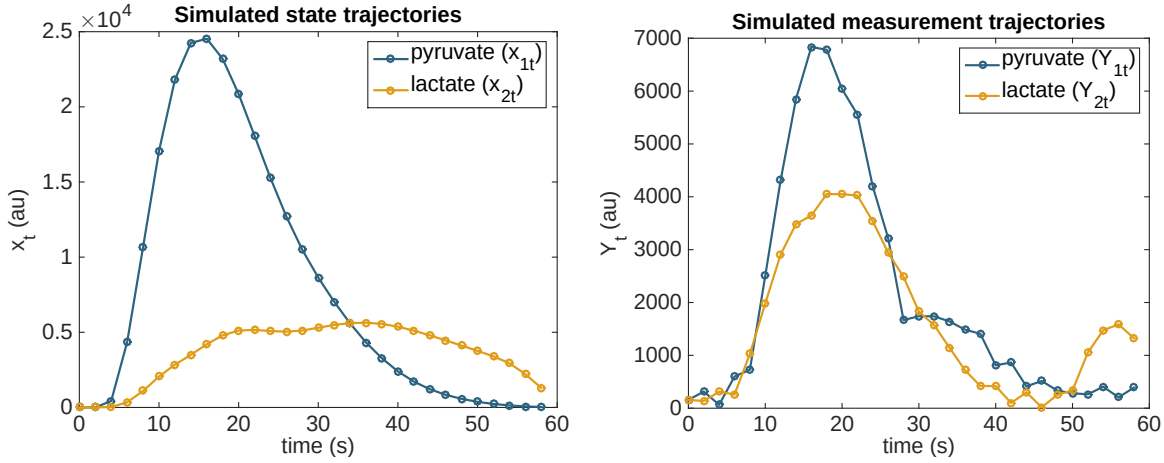


Figure 4.3: Optimized flip angle sequence for estimating the metabolic rate parameter k_{PL} using the nominal parameter values in Table 4.1 and a sampling interval $T_R = 2$ s between acquisitions.

We see that the pyruvate flip angles follow a pattern similar to flip angle sequences designed for other objectives, beginning with small flip angles to preserve magnetization for future acquisitions but increasing toward the end of the sequence [106, 62]. In contrast, the optimized flip angle sequence is much more aggressive with the lactate flip angles at the beginning of the experiment than in other variable flip angle sequences. This provides more reliable information about the leading end of the lactate time series, which contains the most information about the metabolic rate.

For the particular model and regularization parameter values used, the BFGS optimization algorithm converges to the same optimal sequence for a wide range of initializations. To



(a) states x_t (b) Rician-distributed measurements Y_t

Figure 4.4: Simulated trajectories of the model (4.2) using the optimized flip angle sequence shown in Fig. 4.3 and the arterial input function shown in Fig. 4.2a.

confirm this, we have initialized the search algorithm using three flip angle sequences with angles generated randomly between 0 and 90°. For all three initializations, the algorithm converges to the flip angle sequence shown in Fig. 4.3. This demonstrates that the flip angle sequence presented is likely a global optimum.

4.3 Validation Using Simulated Data

In this section, we demonstrate the advantage of the optimally designed flip angle sequence using computer-simulated data. Working with simulated data allows us to collect a large number of statistically independent data sets and provides us access to a "ground truth" value for the parameter vector. This makes it possible to reliably determine the parameter estimation error that results from noise in the simulated measurements. It is not feasible to acquire such a large number of data sets *in vivo*, and these would also not include ground truth values. Thus we use simulated data to demonstrate that our optimized flip angle sequence leads to smaller uncertainty in estimates of the metabolic rate parameter k_{PL}.

Two-step parameter estimation procedure

When fitting the data from *in vivo* experiments, data from different voxels will correspond to different values of the parameters k_{TRANS}, k_{PL}, R_{1P} and R_{1L} as these values change with spatial location, but all correspond to the same arterial input $u(t)$. Thus we present a fitting procedure that proceeds in two steps: first we fit a single input function $u(t)$ to the entire data set, then we fix this input function and estimate values of the remaining parameters individually for each of the voxels in the slice.

Simulation results and discussion

We wish to compare the reliability of estimates of k_{PL} between data generated using five competing flip angle sequences:

1. a T_1-effective sequence [106] that aims to keep the measured signal constant despite repeated RF excitation and magnetization exchange between chemical compounds (Fig. 4.5a)
2. an RF compensated flip angle sequence [109] that aims to keep the measured signal constant despite repeated RF excitation (Fig. 4.5b),
3. a constant flip angle sequence of 15°,
4. a sequence that maximizes the total signal-to-noise ratio in the observed signal

$$SNR_{total} = \sum_{t=0}^{N} \sum_{k=1}^{2} \frac{\tilde{x}_{k,t}}{\sigma_k}.$$

   [62] (Fig. 4.5c), and
5. our flip angle sequence that maximizes the Fisher information about k_{PL} (Fig. 4.3).

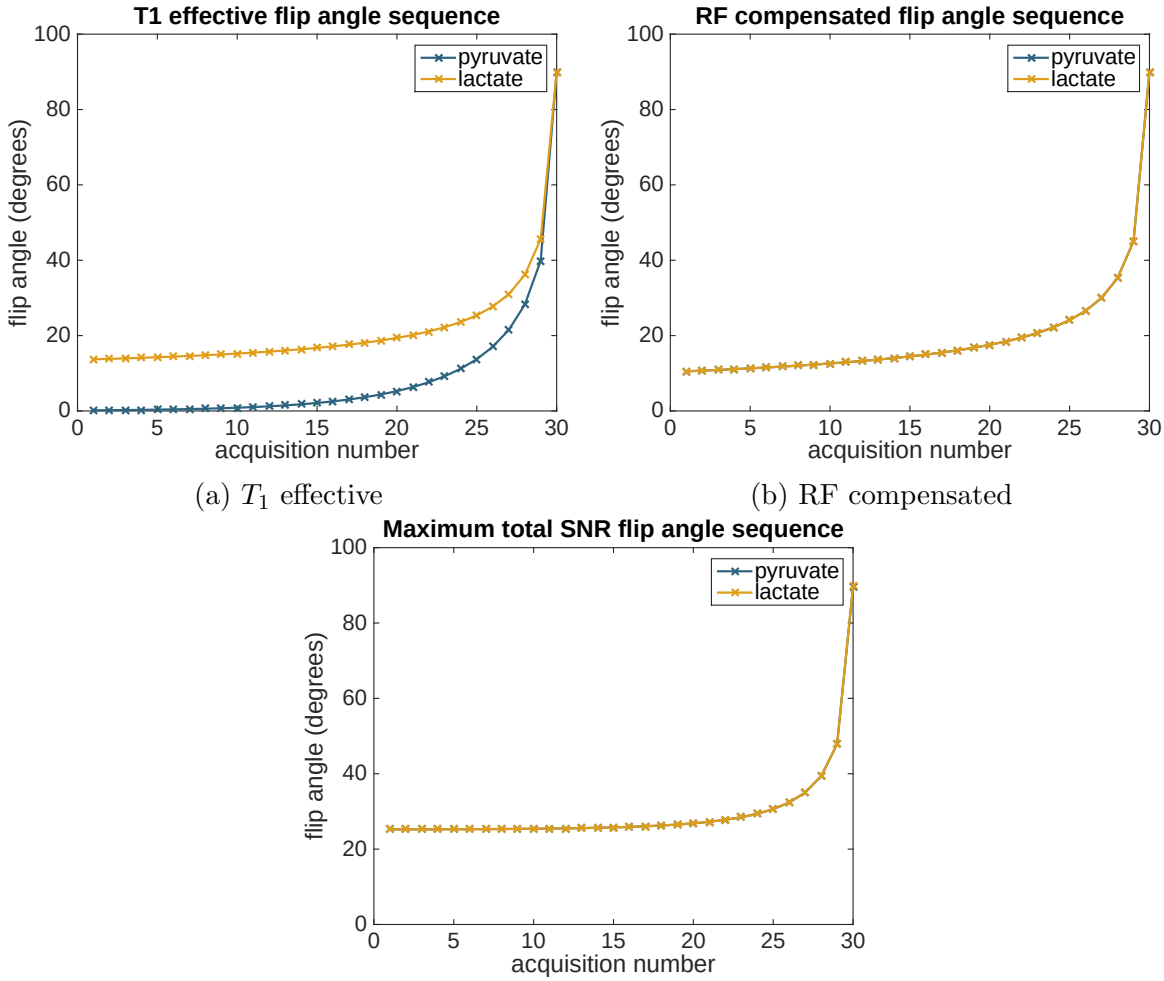
For each of the five flip angle sequences, we simulate $n = 25$ independent data sets from the model (4.2) using the parameter values given in Table 4.1. We then perform the two-step parameter estimation procedure described in Section 4.3. The resulting parameter estimates are shown in Fig. 4.6. We see that for all five flip angle sequences, the parameter estimates congregate near the ground truth value of the model parameters.

To demonstrate that our optimized flip angle sequence provides more accurate estimates of k_{PL} than the competing flip angle sequences, we compare the root mean squared (RMS) estimation error between the sequences. We repeat this experiment for various values of the noise parameter σ^2 ranging from 10^3 to 10^6 to demonstrate that the improvement in the estimates is robust to variation in the noise strength. A value of approximately 2×10^4, in the center of this range, is typical for prostate tumor mouse model experiments. For each value of σ^2 we compute the RMS error of the k_{PL} and nuisance parameter estimates across the $n = 25$ trajectories and plot these relationships in Figs. 4.7 and 4.8 respectively. The average improvement compared with competing sequences, across a range of noise parameter values $\{\sigma_p, p = 1, \ldots, 5\}$, is computed as a percentage

$$100 \times \left(\frac{1}{5} \sum_{p=1}^{5} \left[1 - \frac{RMS_{Fisher}(\sigma_p)}{RMS_{competing}(\sigma_p)} \right] \right).$$

These improvement percentages are summarized in Table 4.2.

Overall, we see that the optimized flip angle sequence provides a more reliable estimate of the parameter of interest k_{PL}, with a substantial improvement over all four competing

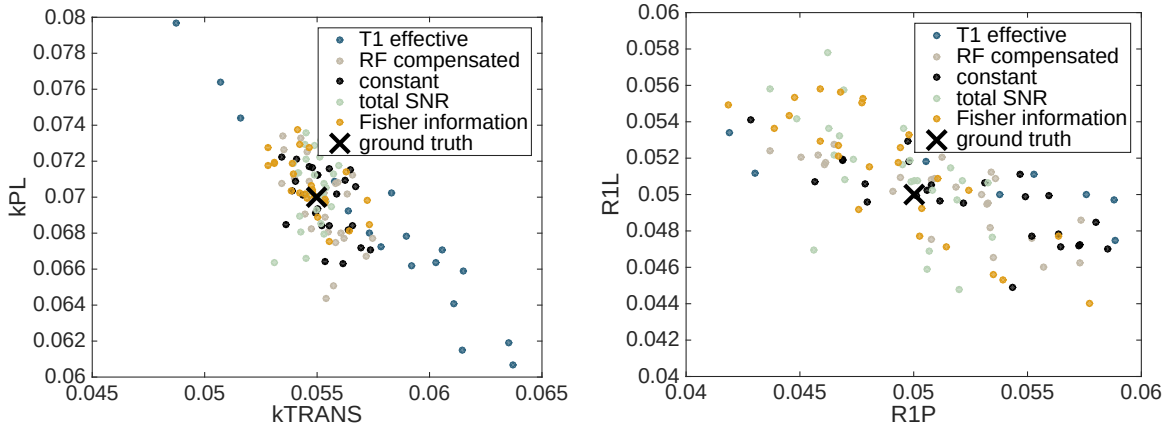


(a) T_1 effective

(b) RF compensated

(c) Maximum total SNR

Figure 4.5: Flip angle schedules compared experimentally with our optimized flip angle sequence. Note that for the RF compensated and maximum total SNR schedules, the sequences corresponding to pyruvate and lactate are identical.



(a) k_{TRANS} and k_{PL} (b) R_{1P} and R_{1L}

Figure 4.6: Maximum likelihood estimates of the parameters k_{TRANS}, k_{PL}, R_{1P} and R_{1L} for numerous independent simulated data sets compared between five flip angle sequences for $\sigma^2 = 2.3608 \times 10^4$. The ground truth value is depicted as $\times$.

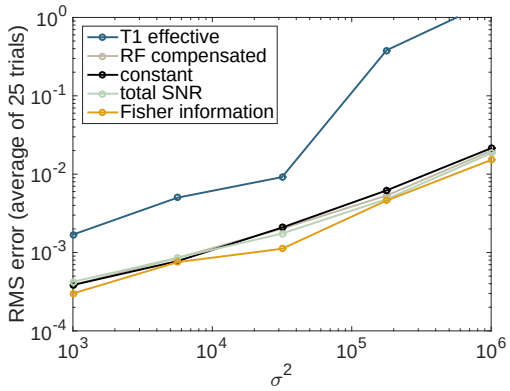
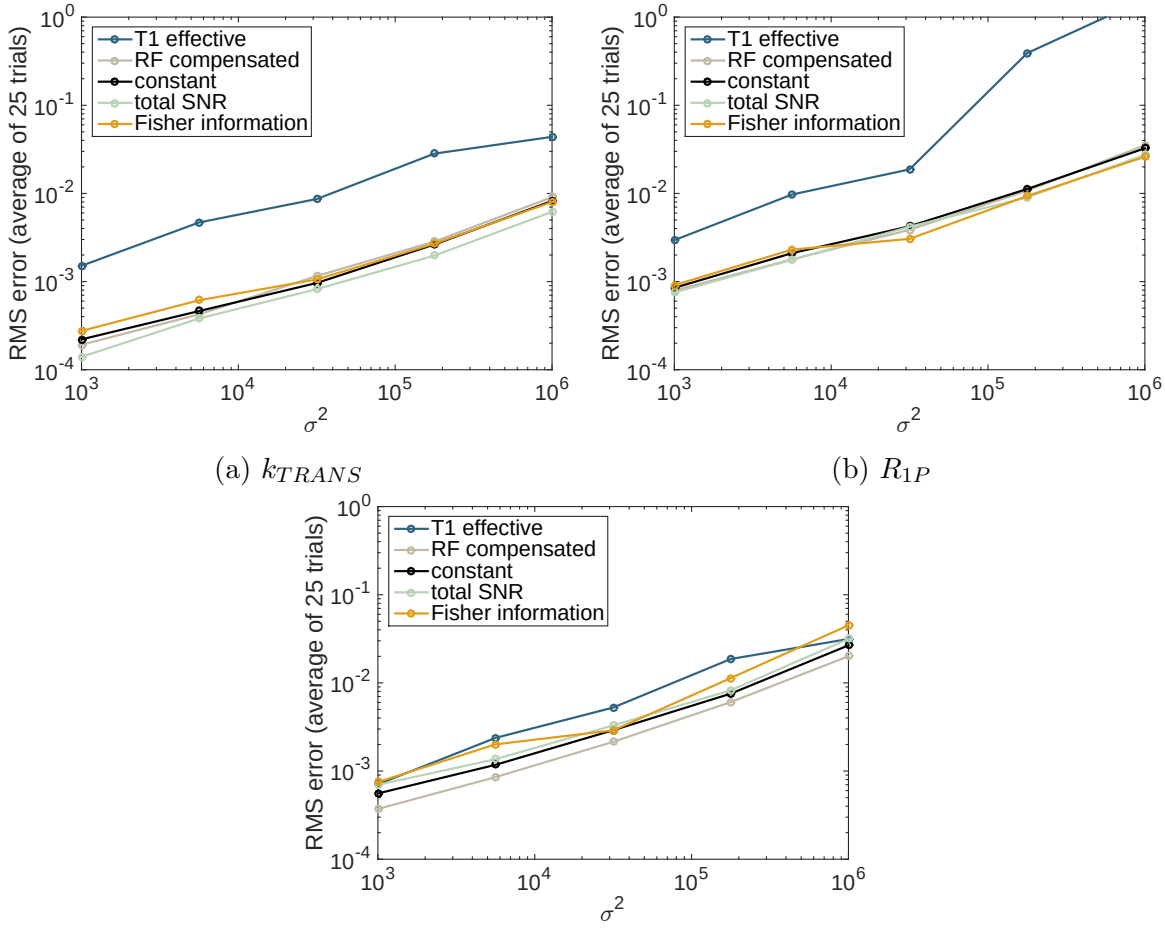


Figure 4.7: Comparison of the root mean square k_{PL} estimation error between various flip angle sequences across different values of the noise strength parameter σ^2.

Table 4.2: Improvement in metabolic rate estimate achieved by Fisher information sequence against competing sequences

Competing sequence	T_1 effective	RF compensated	Constant 15°	total SNR
Improvement achieved	90.6%	23.1%	25.3%	19.8%

flip angle schedules. The magnitude of the improvement varies from a 91% decrease in the estimation error against the T_1 effective sequence to a 20% decrease compared against the closest competitor: the maximum total SNR sequence. This improvement comes at the expense of less reliable estimates of some of the nuisance parameters. This highlights the



(a) k_{TRANS} (b) R_{1P}

(c) R_{1L}

Figure 4.8: Comparison of the root mean square nuisance parameter estimation error between various flip angle sequences across different values of the noise strength parameter σ^2.

advantage of using optimization-based methods to manage trade-offs in experiment design.

Robustness to parametric uncertainty

Based on the simulation experiments described in Section 4.3 we have argued that flip angles optimized based on the Fisher information lead to smaller error in estimates of the parameter of interest k_{PL} *when the same parameter values are used for the simulation and flip angle optimization.* In this section we dispense with the latter assumption to demonstrate that this improvement is robust to uncertainty in the model parameters. We use the flip angle sequence shown in Fig. 4.3, which was designed using the specific model parameters given in Table 4.1, along with two competing flip angle sequences to simulate data from models with different values for the parameters $k_{TRANS}, k_{PL}, R_{1P}, R_{1L}$ and t_0 as well as a factor ΔB_1 multiplying

the flip angle sequence, used to demonstrate robustness to known B_1 inhomogeneities.

The ranges of the varying parameters were chosen to represent realistic physiological ranges and to center near the values we typically observe in a prostate tumor (TRAMP) mouse model. In Fig.4.9 we plot the ensemble RMS error over n=25 simulated data sets as a function of the model parameter modified in the simulated data.

We see that the flip angles optimized based on Fisher information lead to better estimates of k_{PL} across nearly the entire range of model parameters used, despite no longer being optimal for the parameter values used to generate the data. This provides strong evidence that not exactly knowing the parameter values *a priori* does not limit the usefulness of our proposed flip angle design methodology.

4.4 In Vivo Experiments

We now move on to *in vivo* experiments. In contrast with the *in silico* experiments, here there is no ground truth value of the model parameters against which to compare our estimates, as the true rates are unknown and may vary between different regions of the tissue. However, the *in vivo* experiments can be used to show well-foundedness of the model that we have chosen and to demonstrate the feasibility of model-based parameter mapping using our optimized time-varying flip angle sequence. We show that our model can reliably reproduce observed data and achieve consistent parameter estimates across a variety of time-varying flip angle sequences.

Experimental setup

To implement this technique *in vivo*, metabolic data were acquired in a prostate tumor mouse (TRAMP) model using a 3T MRI scanner (MR750, GE Healthcare). Briefly, 24μL aliquots of [1-^{13}C] pyruvic acid doped with 15mM Trityl radical (Ox063, GE Healthcare) and 1.5mM Dotarem (Guerbet, France) were inserted into a Hypersense polarizer (Oxford Instruments, Abingdon, England) and polarized for 60 minutes. The sample was then rapidly dissoluted with 4.5g of 80mM NaOH/40mM Tris buffer to rapidly thaw and neutralize the sample. Following dissolution, 450μL of 80mM pyruvate was injected via the tail vein over 15 seconds, and data acquisition coincided with the start of injection. Metabolites from a single slice were individually excited with a singleband spectral-spatial RF pulse and encoded with a single-shot symmetric EPI readout [41], with a repetition time of 100ms, a field-of-view of 53 $\times$ 53mm, a matrix size of 16 $\times$ 16, an 8mm slice thickness, and a 2 second sampling interval. A ^{1}H image showing the anatomy contained in the slice in question is given in Fig. 4.11.

Datasets were acquired using three time-varying flip angle sequences. The T_1-effective and RF-compensated sequences, shown in Figs. 4.5a and 4.5b respectively, aim to distribute observed magnetization evenly across acquisitions, leading to roughly constant observed signals over time. The RF-compensated sequence does so by accounting for magnetization

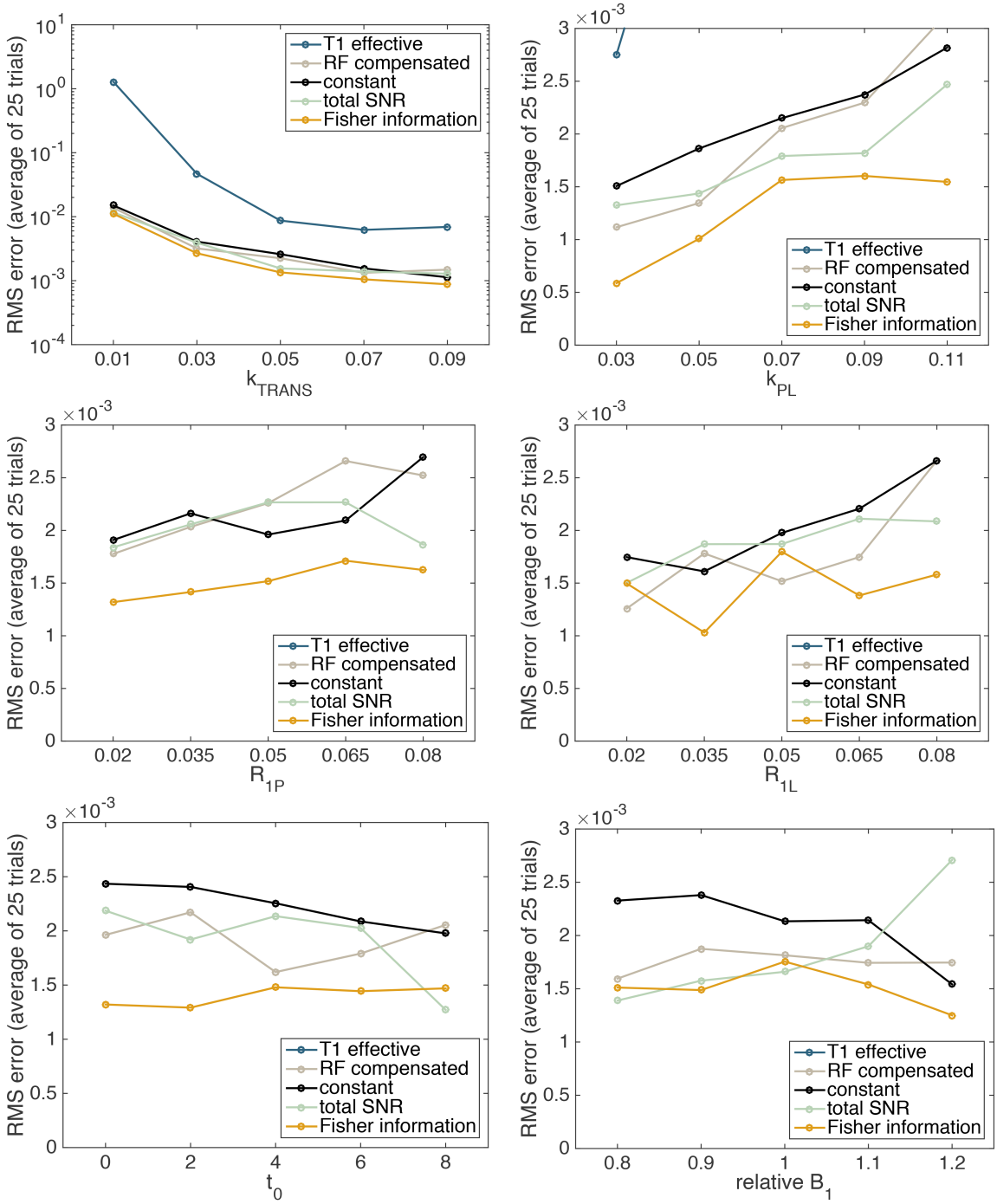


Figure 4.9: Simulated comparison of the k_{PL} estimation error across values of the model parameters. The error is compared between five flip angle sequences shown in Figs. 4.3 and 4.5. These flip angle sequences are computed based on the nominal values of the model parameters given in Table 4.1 and held fixed across all comparisons. Note that the first graph is logarithmically scaled, due to the fact that low perfusion leads to significant uncertainty in the metabolic rate estimates. Some estimation errors corresponding to the T_1 effective sequence are greater than the maximum value plotted on these axes.

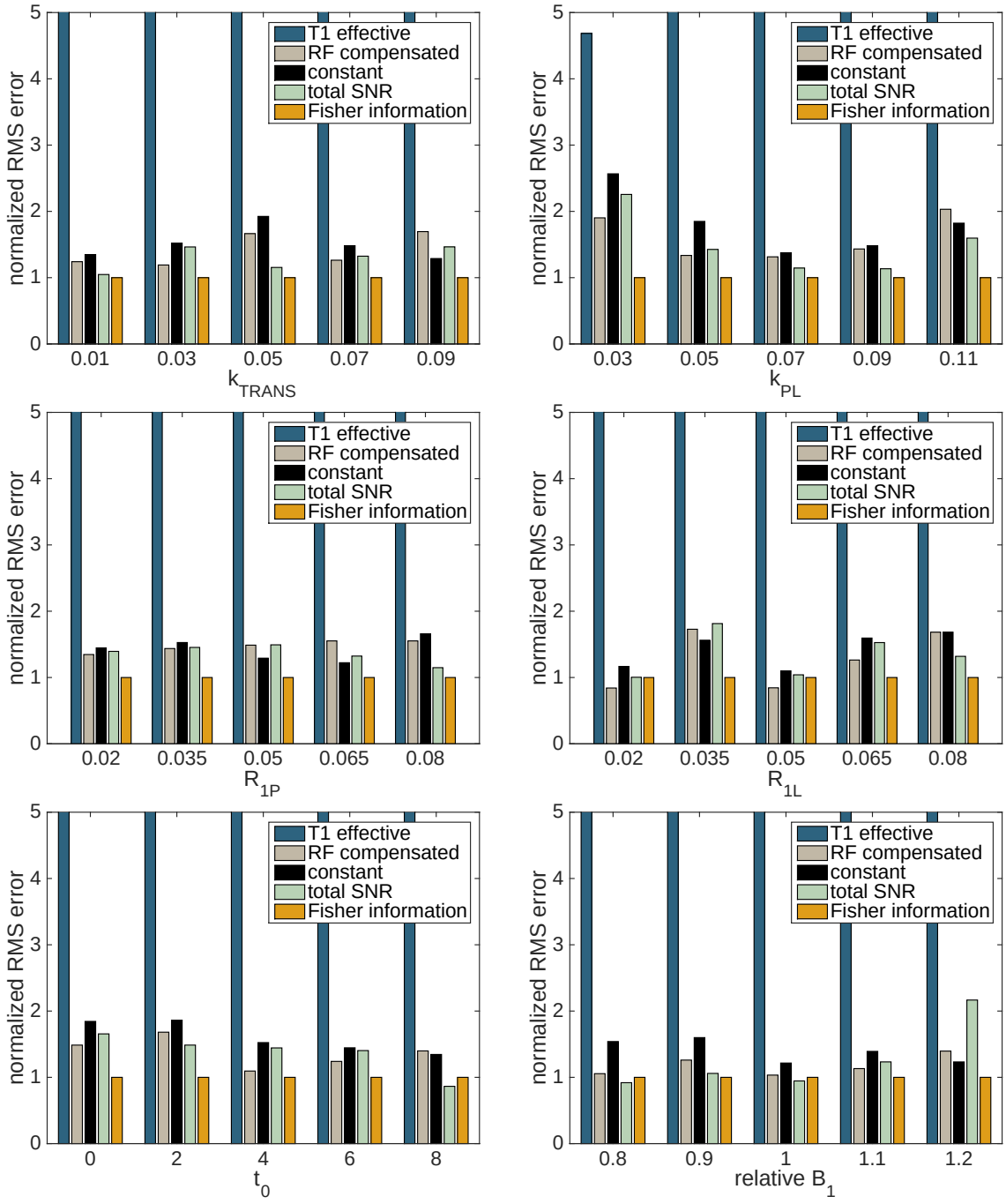


Figure 4.10: Simulated comparison of the *normalized* k_{PL} estimation error across values of the model parameters. The absolute estimation error for the Fisher information sequence ranges from approximately 10^{-3} to 10^{-2} across the different values of the model parameters.

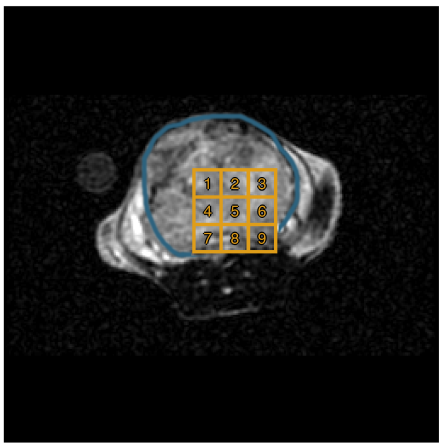


Figure 4.11: ^{1}H image of the slice of interest. A large tumor, outlined in blue, fills a significant portion of the slice. Numbered volumes used to extract trajectories for parameter estimation are outlined in gold.

lost due to repeated RF excitation, but ignoring exchange between chemical compounds and T_1 relaxation [109]. In contrast, the T_1-effective sequence accounts for exchange and T_1 relaxation as well as RF excitation in attempting to achieve a flat time profile [106]. We compare these two sequences against our sequence, shown in Fig. 4.3, that has been optimized with respect to the Fisher information about k_{PL}.

Resulting data

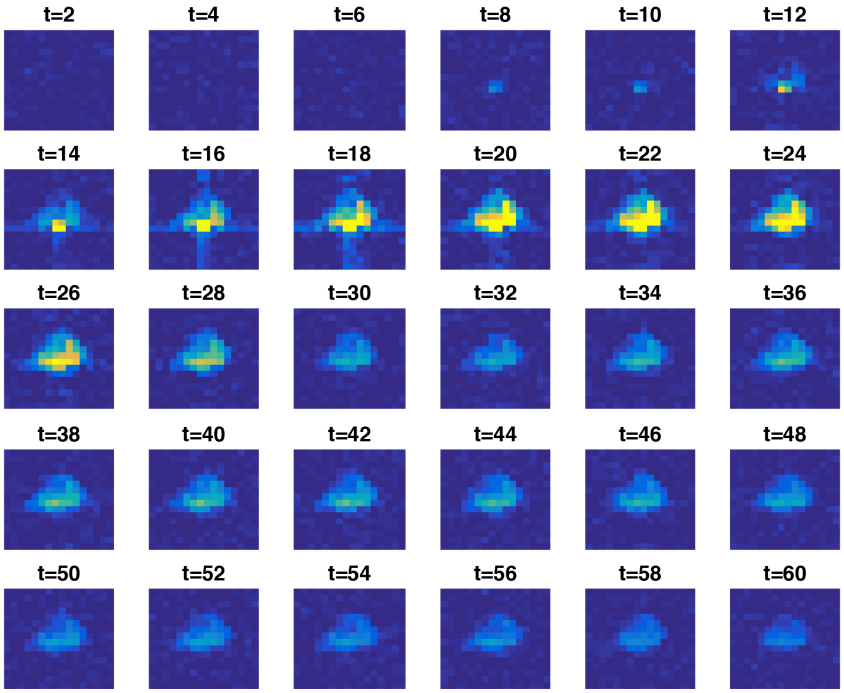
An example of the collected data, from the experiment with the Fisher information-optimized flip angles, is shown in Fig. 4.12. Experimentally-estimated values for the noise parameters are given in Table 4.3.

Table 4.3: Maximum-likelihood estimates of the noise parameter σ_k^2 for each of the three data sets collected

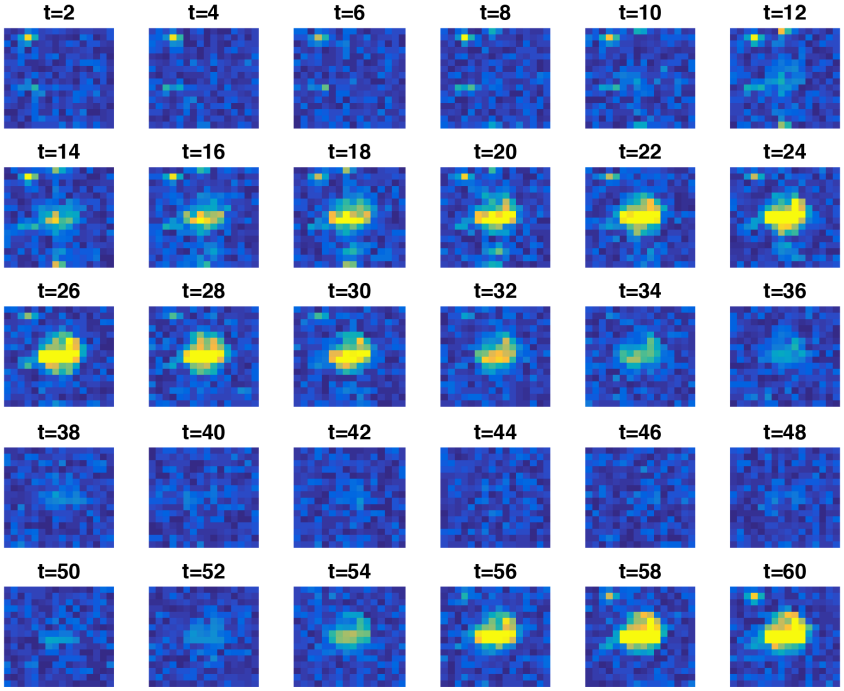
	T_1 effective	RF compensated	Fisher information
σ_1^2 (pyruvate)	1.84×10^4	2.05×10^4	2.21×10^4
σ_2^2 (lactate)	2.14×10^4	1.97×10^4	3.26×10^4

Flip angle profile modelling

Due to an imperfect (non-rectangular) slice profile, flip angles applied in practice vary spatially across the slice. This can lead to excess signal coming from regions near the boundary of the slice at later time points in the acquisitions, a phenomenon known as the slice profile



(a) pyruvate data



(b) lactate data

Figure 4.12: Data collected using the optimized flip angles shown in Fig. 4.3. The field-of-view of these images is identical to the ^{1}H image in Fig. 4.11.

effect [73, 30]. We have found that it is necessary to account for this effect in order to accurately fit the experimental data.

We consider the actual slice profile $\pi(z)$ shown in Fig. 4.13 which corresponds to the spatial response of the RF pulse used experimentally. We assume that at each time step t and for each compound k (*i.e.* $k = 1$ corresponding to pyruvate, $k = 2$ corresponding to lactate) we can choose a real parameter $\alpha_{k,t}$ such that the flip angle applied at location z is $\theta_{k,t}(z) = \alpha_{k,t}\pi(z)$. To generate a finite-dimensional model of the dynamics, we consider the magnetization dynamics at a discrete set of z coordinates $\{z_1, \ldots, z_N\}$. The magnetization at location z_i in the slice is then governed by the equations

$$x_{i,t+1} = A_d \begin{bmatrix} \cos(\alpha_{1,t}\pi(z_i)) & 0 \\ 0 & \cos(\alpha_{2,t}\pi(z_i)) \end{bmatrix} x_{i,t} + B_d u_t(p)$$

and the total magnetization measured is then assumed to be distributed

$$\tilde{x}_t = \sum_{i=1}^{N} \frac{1}{z_{i+1} - z_i} \begin{bmatrix} \sin(\alpha_{1,t}\pi(z_i)) & 0 \\ 0 & \sin(\alpha_{2,t}\pi(z_i)) \end{bmatrix} x_{i,t}$$
$$Y_{k,t} \sim Rice(\tilde{x}_{k,t}, \sigma^2).$$

This approach accounts for the slice profile effects by modelling the dynamics across the actual slice profile.

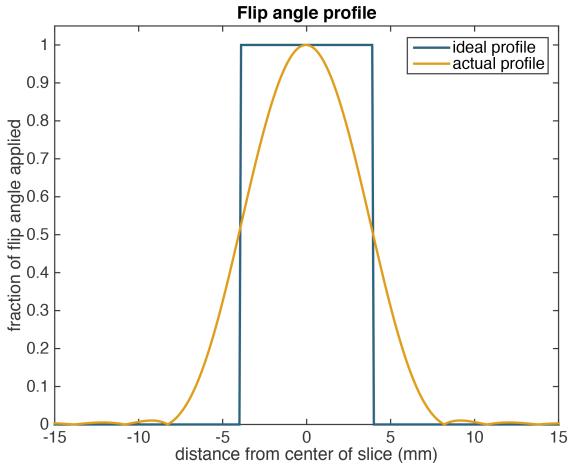


Figure 4.13: Comparison between an ideal flip angle profile across the slice and the actual profile for the RF excitation pulse used.

Parameter estimation

We begin by extracting time evolutions of the measured pyruvate and lactate signal from $n = 9$ voxels in the slice. The chosen volumes from which these signals are extracted are

illustrated in Fig. 4.11. As in Section 4.3, parameter estimation is performed in two steps. First, a single arterial input function and value for the parameter R_{1L} are estimated based on the spatial average of the time series extracted from the tumor region. Second, the estimated input function and R_{1L} value are held fixed while model parameters k_{TRANS} and k_{PL} are fit individually to the time series extracted from each of the voxels. To ensure practical identifiability of the model, the parameter R_{1P} is fixed to a value of 0.05 during both steps.

In vivo results and discussion

Estimates of the arterial input corresponding to each of the three flip angle sequences are shown in Fig. 4.14. We see that the estimated inputs are reasonably consistent between the three data sets, but have some variation due to measurement noise in the pyruvate signal. Our optimized sequence yields the most smoothly-varying input function, which suggests that it is likely the most reliable of the three estimated AIFs.

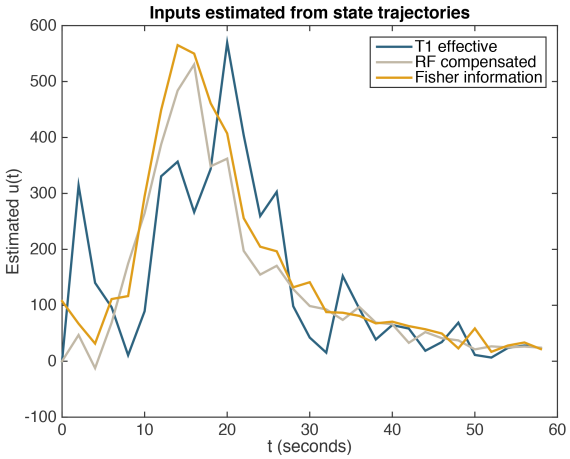


Figure 4.14: Estimated arterial input functions corresponding to each of the three flip angle sequences given in Figs. 4.3 and 4.5.

Parameter estimates corresponding to each of the nine voxels are compared between the three flip angle sequences in Fig. 4.15. Examples of the quality of the fit corresponding to a particular voxel are shown in Fig. 4.16. We see that the estimated parameter values are consistent between the three flip angle sequences and that our model is able to reliably reproduce the observed data in all three cases. This provides evidence that the model we have used in this chapter accurately describes the dynamics of magnetization exchange *in vivo*, and hence that the decision to use this model for the numerical reliability experiments of Section 4.3 is well-founded.

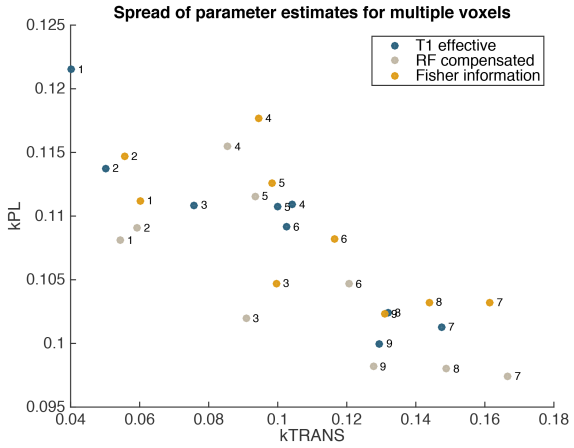


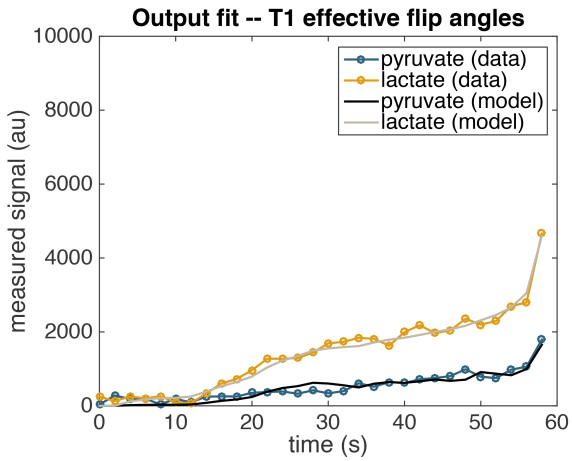
Figure 4.15: Maximum likelihood estimates of the parameters k_{TRANS} and k_{PL} for time series trajectories extracted from various voxels, labelled 1 through 9. The resulting estimates are compared between data sets collected using the three flip angle sequences shown in Figs. 4.3 and 4.5.

Parameter mapping

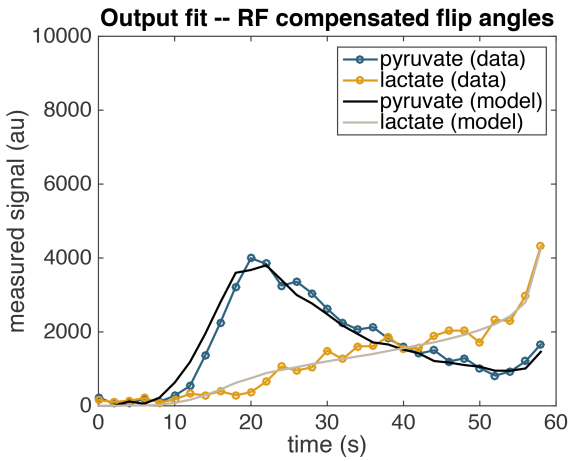
We also present maps that show the spatial distribution of estimated metabolic and perfusion rates in Fig. 4.17. We see that the range and spatial distribution of parameter estimates are consistent between acquisition sequences.

4.5 Conclusion

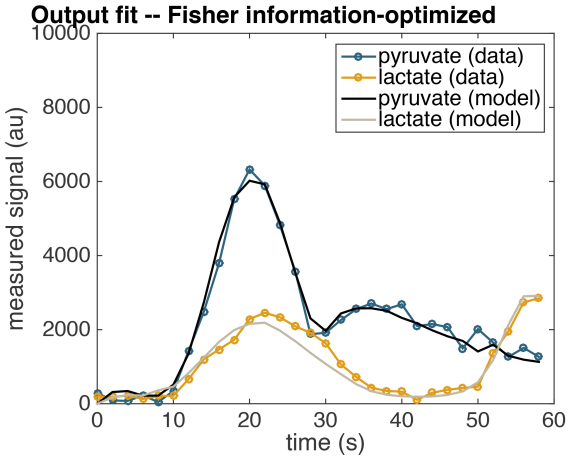
We have presented a method of generating optimal flip angle sequences for estimating the metabolic rate in a model of pyruvate metabolism. This method uses the Fisher information about the parameter of interest as the objective function that we wish to maximize. We have shown that the resulting flip angle sequence leads to smaller variance in the parameter estimates due to noise in the measured signal. We have demonstrated this *in silico* where we can explicitly compare the estimated model parameter values against the ground truth value. In this simulation experiment we demonstrated that our flip angle sequence leads to a 20% to 90% decrease in the uncertainty of the estimated metabolic rate, when compared with existing sequences. We also performed *in vivo* experiments to provide evidence that the model used in the *in silico* experiments is well-founded and demonstrate the feasibility of metabolic rate estimation and parameter mapping using this novel sequence. Based on the reliability results demonstrated *in silico* and the *in vivo* experiments demonstrating the appropriateness of the model used for the *in silico* experiments we argue that, for experiments that aim to quantitatively compare metabolic rates, optimizing flip angle sequences based on the Fisher information will probably lead to more reliable estimates of the model parameters of interest.



(a) T_1 effective flip angle sequence (Fig. 4.5a)

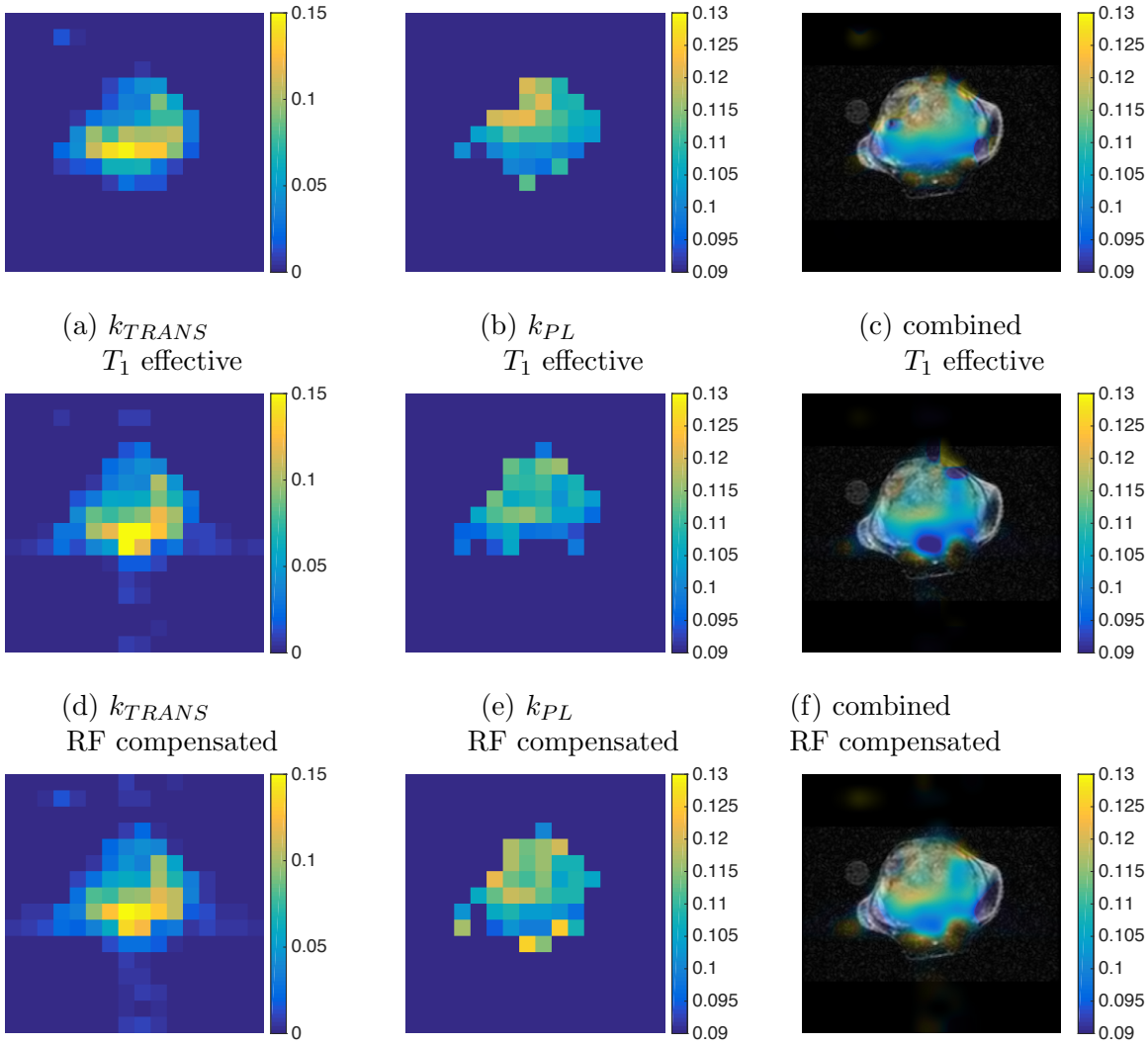


(b) RF compensated flip angle sequence (Fig. 4.5b)



(c) Fisher information optimized flip angle sequence (Fig. 4.3)

Figure 4.16: Model fit to a collection of experimentally measured time series data corresponding to voxel number 5. Each of the three data sets was collected using a different flip angle sequence.



(a) k_{TRANS} T_1 effective

(b) k_{PL} T_1 effective

(c) combined T_1 effective

(d) k_{TRANS} RF compensated

(e) k_{PL} RF compensated

(f) combined RF compensated

(g) k_{TRANS} Fisher information

(h) k_{PL} Fisher information

(i) combined Fisher information

Figure 4.17: Maps of the perfusion rate parameter k_{TRANS} and metabolic rate parameter k_{PL} corresponding to each of the three flip angle sequences. The k_{PL} maps are masked outside the perfused region using a threshold of $k_{TRANS} = 0.02$. A single map combining anatomic, perfusion and metabolism information is shown on the right. In this map, the color is determined by the estimated k_{PL} value while the transparency of the map is set using the perfusion rate parameter k_{TRANS} such that in highly-perfused tissues where the estimates of the metabolic rate parameter are more reliable the map is less transparent. The combined image data are zero-filled from 16 × 16 to 256 × 256 to match the resolution of the ^{1}H images.

Chapter 5

Spatio-Temporally Constrained Reconstruction for Hyperpolarized Carbon-13 MRI Using Kinetic Models

5.1 Introduction

Magnetic resonance imaging (MRI) using hyperpolarized carbon-13 labeled substrates has made it possible to probe metabolism *in vivo* with chemical specificity [38, 29]. This technique is increasingly being applied in the clinic, allowing researchers to investigate metabolic conditions ranging from prostate cancer [77] to heart disease [27]. In particular, experiments studying the conversion of hyperpolarized [1-^{13}C]pyruvate to [1-^{13}C]lactate are common, as the rate of conversion is upregulated in many cancers, a phenomenon known as the Warburg effect.

MRI using hyperpolarized carbon-13 is challenging due to the dynamic nature of the data collected, the low signal-to-noise ratio (SNR), and the difficulty of presenting large data sets consisting of dynamic spectroscopic images in an interpretable manner. Metabolism mapping by estimating parameters in a kinetic model from hyperpolarized MRI data has been shown to be useful for overcoming a number of these challenges [6]. Constraining the time evolution of signal in a given voxel to follow a kinetic model has been shown to allow map reconstruction from noisy, undersampled dynamic images, and to reduce the number of signal-depleting excitations required to generate images. Parameter mapping also facilitates interpretation of dynamic image data by summarizing spatial, temporal and chemical (i.e. chemical shift spectrum) information in a single spatial map.

Parameter maps are naturally a form of constrained reconstruction, as they constrain the data to lie on a manifold of trajectories of the dynamical system parametrized by the system's parameters. This constrained reconstruction reduces the sequence of dynamic images to a single map by exploiting temporal correlations within the dynamic imaging data. In this chapter, we demonstrate that we can exploit spatial correlations in addition to temporal cor-

relations by integrating prior information about the parameter map through regularization. Similar approaches have proven useful recently in the context of pharmacokinetic parameter mapping in dynamic contrast enhanced MRI [57, 45, 46].

This chapter is organized as follows. In Section 5.2 we introduce background on modelling hyperpolarized ^{13}C MRI data and existing approaches to parameter mapping. In Section 5.3 we introduce a framework for spatially-constrained parameter mapping to exploit spatial correlations in the data. In Section 5.4 we present an algorithm for efficient inference in this framework. In Section 5.5 we present the results of simulation experiments where we demonstrate the effectiveness of the method. In Section 5.6 we then apply the method to a collection of clinically-relevant data sets. Finally, Section 5.7 concludes the chapter and briefly discusses potential extensions of this work.

5.2 Background

Data Model

We model the dynamic evolution of the data Y_i collected from a single voxel i using the dynamic model

$$\frac{dx}{dt}(t) = \begin{bmatrix} -k_{PL} - R_{1P} & 0 \\ k_{PL} & -R_{1L} \end{bmatrix} x(t) + \begin{bmatrix} k_{TRANS} \\ 0 \end{bmatrix} u(t). \tag{5.1}$$

This system of ordinary differential equations (ODEs) has been widely used to model the uni-directional conversion of an injected substrate (pyruvate, in this case) to a metabolic product (lactate, in this case) [47]. The state $x_1(t)$ models the longitudinal magnetization in the substrate pool, and the state $x_2(t)$ models the longitudinal magnetization in the product pool. The parameter k_{PL} describes the rate at which the substrate is metabolized, the parameter k_{TRANS} describes the rate at which the substrate is taken up by the tissue, and the parameters R_{1P} and R_{1L} are lumped parameters that account for T_1 magnetization decay, metabolism of the substrate into unmeasured products and flow of substrate out of the voxel.

Measurements are collected at a sequence of times $\{t_1, \ldots, t_N\}$. Neglecting the effect of the input between t_k and t_{k+1}, integrating this continuous-time dynamic model and incorporating the effect of repeated radio-frequency (RF) excitation leads to a discrete-time model for the magnetization at acquisition times t_k of the form

$$\hat{L}(k+1) = e^{-R_{1L}\Delta t}\cos(\alpha_L(k))\hat{L}(k) - k_{PL}\frac{e^{-(R_{1P}+k_{PL})\Delta t} - e^{-R_{1L}\Delta t}}{R_{1P} - R_{1L} + k_{PL}}\cos(\alpha_P(k))P(k). \tag{5.2}$$

This gives a statistical model that describes the evolution of the predicted lactate signal $\hat{L}(k) = x_2(t_k)$ as a function of the measured pyruvate signal $P(k) = x_1(t_k)$ and the flip angles α_P and α_L applied to the pyruvate and lactate compartments. The predicted lactate is assumed to be $\hat{L}(0) = 0$ at the beginning of the experiment.

For the purpose of generating simulated data, the data measured at each time t_k are assumed to be independent and follow a bivariate normal distribution with mean $\delta_x\delta_y\delta_z x(t_k)$ and covariance $\sigma^2 I$ where I denotes the 2×2 identity matrix and δ_x, δ_y and δ_z describe the image resolution and slice thickness. We collect the time series data collected from voxel i into a vector $Y_i = \begin{bmatrix} P(1) & \cdots & P(N) \\ L(1) & \cdots & L(N) \end{bmatrix}$ and denote the unknown parameters to be estimated from the data $\theta_i = k_{PL}$.

Voxel-Wise Parameter Estimation

Given a collection of data Y_i from a voxel i we wish to generate an estimate of the parameter θ_i that describes the tissue in that voxel. We assume that θ_i lies in a parameter space Θ. We consider the class of "M-estimators" [50] that minimize a loss function

$$\hat{\theta}_i \in \operatorname*{argmin}_{\theta \in \Theta} \ell(\theta_i | Y_i).$$

In the present chapter, we consider the nonlinear least squares loss function

$$\ell(\theta_i | Y_i) = \|Y_i - \hat{Y}_i(\theta_i)\|_F \tag{5.3}$$

where $\hat{Y} = \begin{bmatrix} P(1) & \cdots & P(N) \\ \hat{L}(1) & \cdots & \hat{L}(N) \end{bmatrix}$ denotes the predicted signal given the pyruvate time series and $\|\cdot\|_F$ denotes the Frobenius norm (*i.e.* the ℓ_2 norm of the vectorized matrix). Under the assumption that the data collected are normally-distributed with mean proportional to $x(t_k)$, independent with identical variance, the minimum of this nonlinear least squares loss is also the maximum likelihood estimate of the parameter vector. While we consider only this loss in the present chapter, the results are applicable generally to any computationally tractable loss function.

5.3 Constrained Parameter Mapping

In order to incorporate prior information about the spatial distribution of metabolic rates and exploit spatial correlations within the data, we constrain the maps to have a desired structure through regularization. This results in an optimization problem in Lagrangian form

$$\text{minimize} \sum_{i \in \mathcal{V}} \ell(\theta_i | Y_i) + \lambda r(\theta) \tag{5.4}$$

where $\theta = (\theta_i)_{i \in \mathcal{V}}$ denotes the map of parameters across all voxels, r is a regularization term, and λ denotes a Lagrange multiplier that can be tuned in order to achieve the desired regularization strength. The choice of an appropriate regularizer depends on the desired features of the parameter map. Common choices include Tikhonov (ℓ_2) regularization, ℓ_1

regularization, and total variation regularization. We briefly summarize these three methods below.

Tikhonov regularization, or ℓ_2 regularization penalizes the size of the parameters θ_i. It involves adding a quadratic penalty term

$$r(\theta) = \|\theta\|_2^2$$

where $\|\cdot\|_2$ denotes the ordinary Euclidean norm. For linear regression problems with orthogonal covariates, this regularization leads to uniform shrinkage of the estimates [48]. For the nonlinear parameter mapping problems we consider here, using Tikhonov regularization helps to suppress large parameter values in the unperfused "background" region.

ℓ_1 regularization is another shrinkage method that penalizes parameters based on their ℓ_1 norm

$$r(\theta) = \|\theta\|_1.$$

This method induces sparsity in the resulting parameter maps, and hence also helps to suppress parameter values in the background region. It is closely-related to basis pursuit denoising [23] and lasso regression [103].

Total variation (TV) regularization is another method commonly used for image denoising [91]. The regularization term in this case is given by

$$r(\theta) = \|\nabla\theta\|_1 := \sum_{(i,j)\in\mathcal{N}} \|\theta_i - \theta_j\|_1$$

where ∇ denotes a discrete differencing operator and $\mathcal{N}$ denotes the set of all neighbouring voxels. TV regularization is known to preserve edges and large-scale structure in images while rejecting noise [100], resulting in natural-looking reconstructed images.

5.4 Iterative Algorithms for Constrained Parameter Mapping

A naive algorithm for solving this optimization problem by directly optimizing the objective function (5.4) would be inefficient because it involves solving a joint optimization over all $\{\theta_i : i \in \mathcal{V}\}$. Thus the computation time required to directly solve the optimization problem increases dramatically with matrix size, making naive approaches inefficient even for the images of moderate resolution considered here. To solve the optimization problem more efficiently, we can take advantage of the particular structure of the problem using the ADMM algorithm.

The alternating direction method of multipliers (ADMM) is an iterative optimization algorithm that is well-suited to efficiently solving such problems that can be decomposed into a sum of two terms [17]. In contrast with other distributed optimization algorithms, the ADMM algorithm is particularly well-suited to the problem formulated in this chapter as it

splits the required optimization into the sum of a set of loss functions ℓ that are complex to optimize, but can be optimized independently for each voxel, and a regularization r that is relatively simple but high-dimensional as it couples a large number of neighboring voxels. By exploiting this decomposition, ADMM allows the optimization problem to be solved efficiently. The general problem that ADMM attempts to solve is an optimization problem of the form

$$\begin{aligned} \textbf{minimize} \quad & f(x) + g(z) \\ \textbf{subject to} \quad & Ax + Bz = c. \end{aligned} \tag{5.5}$$

The algorithm does so by iteratively applying the updates

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}} \left(f(x) + \frac{\rho}{2} \|Ax - Bz^k - c + u^k\|_2^2 \right) \\ z^{k+1} &= \underset{z}{\operatorname{argmin}} \left(g(z) + \frac{\rho}{2} \|Ax^{k+1} - Bz - c + u^k\|_2^2 \right) \\ u^{k+1} &= u^k + Ax^{k+1} + Bz^{k+1} - c. \end{aligned}$$

Under the assumption that f and g are closed, proper, convex functions and that the Lagrangian

$$L(x, z, \lambda) = f(x) + g(z) + \lambda^T (Ax + Bz - c)$$

has a saddle point, it can be shown [17] that the residuals $r^k = Ax^k + Bz^k - c$ converge to zero and the values $f(x^k) + g(z^k)$ converge to the optimal value of the problem (5.5).

ADMM for iterative parameter mapping

To solve (5.4) we transform the problem to a form amenable to the ADMM algorithm by introducing a new variable $z = \theta$ and solving

$$\begin{aligned} \textbf{minimize} \quad & \sum_{i \in \mathcal{V}} -\ell(\theta_i | y_i) + \lambda r(z) \\ \textbf{subject to} \quad & \theta - z = 0. \end{aligned} \tag{5.6}$$

The ADMM iteration is then given as

$$\begin{aligned} \theta^{k+1} &= \underset{\theta}{\operatorname{argmin}} \sum_{i \in \mathcal{V}} -\ell(\theta_i | y_i) + \frac{\rho}{2} \|\theta - z^k + u^k\|_2^2 \\ z^{k+1} &= \underset{z}{\operatorname{argmin}} \, \lambda r(z) + \frac{\rho}{2} \|\theta^{k+1} - z + u^k\|_2^2 \\ u^{k+1} &= u^k + \theta^{k+1} - z^{k+1}. \end{aligned}$$

This method is sometimes known as Douglas-Rachford splitting [104]. Note that the θ update is additively separable. Introducing the proximity operator

$$\underset{f}{\text{prox}}(x) = \underset{u}{\text{argmin}}\, f(u) + \frac{1}{2}\|u - x\|_2^2$$

we can re-write this iteration as

$$\begin{aligned}
\theta_i^{k+1} &= \underset{-\frac{1}{\rho}\ell(\cdot|y_i)}{\text{prox}}\ (z_i^k - u_i^k) \quad i \in \mathcal{V} \\
z^{k+1} &= \underset{\frac{\lambda}{\rho}r}{\text{prox}}(\theta^{k+1} + u^k) \\
u^{k+1} &= u^k + \theta^{k+1} - z^{k+1}.
\end{aligned}$$

Here, the θ_i updates can be performed independently for each $i \in \mathcal{V}$, significantly decreasing time and memory required for computation and allowing the parallelization of this step.

Note that for the particular choice of loss function given in Section 5.3, $\ell(\cdot|Y_i)$ are non-convex functions and thus the formal convergence guarantees do not apply. Despite this fact, we have seen in all the experimental instances of the problem we have considered that the algorithm converges to a sensible optimum robustly for a variety of initializations.

Hybrid ℓ_2 regularization

As a particular technical detail that helps to simplify the implementation of the algorithm, we consider the case when the regularization term can be expressed as a positive combination of a Tikhonov term and another term:

$$\lambda r(\theta) = \lambda_1 r_1(\theta) + \lambda_2 \|\theta\|_2^2.$$

In this special case, the z update step can be simplified considerably. Indeed, we can express the z update as

$$\begin{aligned}
z^{k+1} &= \underset{z}{\text{argmin}}\, \lambda_1 r_1(z) + \lambda_2 \|z\|_2^2 + \frac{\rho}{2}\|\theta^{k+1} - z + u^k\|_2^2 \\
&= \underset{z}{\text{argmin}}\, \lambda_1 r_1(z) + \lambda_2 z^T z + \frac{\rho}{2} z^T z - \rho(\theta^{k+1} + u^k)^T z + \frac{\rho}{2}(\theta^{k+1} + u^k)^T(\theta^{k+1} + u^k) \\
&= \underset{z}{\text{argmin}}\, \lambda_1 r_1(z) + \left(\lambda_2 + \frac{\rho}{2}\right) z^T z - \rho(\theta^{k+1} + u^k)^T z \\
&= \underset{z}{\text{argmin}}\, \lambda_1 r_1(z) + \left(\lambda_2 + \frac{\rho}{2}\right) \|z - b\|_2^2
\end{aligned}$$

where

$$b = \frac{1}{2}\frac{\rho}{\left(\lambda_2 + \frac{\rho}{2}\right)}(\theta^{k+1} + u^k) = \frac{1}{1 + \frac{2\lambda_2}{\rho}}(\theta^{k+1} + u^k).$$

Thus the z update can be written in terms of the proximity operator of r_1 as

$$z^{k+1} = \underset{\frac{\lambda_1}{\lambda_2+\frac{\rho}{2}} r_1}{\text{prox}}\,(b) = \underset{\frac{\lambda_1}{\lambda_2+\frac{\rho}{2}} r_1}{\text{prox}} \left(\frac{1}{1+\frac{2\lambda_2}{\rho}} (\theta^{k+1} + u^k) \right). \tag{5.7}$$

5.5 Simulated Results and Discussion

To demonstrate the effectiveness of this method, we perform a sequence of experiments on simulated data. We begin with an experiment using a simple numerical phantom designed to test the robustness of metabolic parameter mapping methods to differences in perfusion, as well as their ability to reliably resolve large and small features.

A numerical phantom

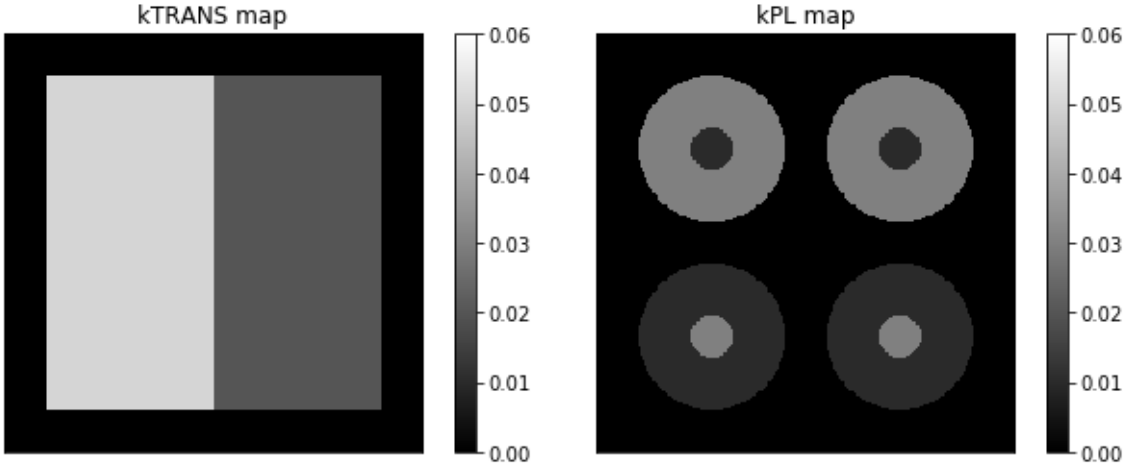
In order to compare parameter mapping algorithms, we introduce a 3D parameter phantom that describes the simple spatial distribution of perfusion rates k_{TRANS} and metabolism rates k_{PL}. This phantom is defined on the three-dimensional spatial interval $[-1,1] \times [-1,1] \times [-1,1]$. The k_{TRANS} map consists of two adjacent rectangular cuboids. The cuboid that defines the low perfusion region with $k_{TRANS} = 0.02$ is centered at $(x_0, y_0, z_0) = (0.4, 0.0, 0.0)$ and has side lengths $(L_x, L_y, L_z) = (0.8, 1.6, 1.6)$. The cuboid that defines the high perfusion region with $k_{TRANS} = 0.05$ is centered at $(x_0, y_0, z_0) = (-0.4, 0.0, 0.0)$ and also has side lengths $(L_x, L_y, L_z) = (0.8, 1.6, 1.6)$. The k_{PL} map is made up of four large spherical features nested with four smaller spherical features. Centers (x_0, y_0, z_0), radii R and k_{PL} values for these spheres are given in Table 5.1.

Table 5.1: Geometric parameters defining the numerical metabolic phantom. x_0, y_0, and z_0 define the center of the sphere, R defines the sphere radius and k_{LP} defines the metabolic rate in the sphere's interior.

R	x_0	y_0	z_0	k_{PL}
0.35	0.45	0.45	0.00	0.01
0.35	-0.45	0.45	0.00	0.01
0.35	0.45	-0.45	0.00	0.03
0.35	-0.45	-0.45	0.00	0.03
0.10	0.45	0.45	0.00	0.03
0.10	-0.45	0.45	0.00	0.03
0.10	0.45	-0.45	0.00	0.01
0.10	-0.45	-0.45	0.00	0.01

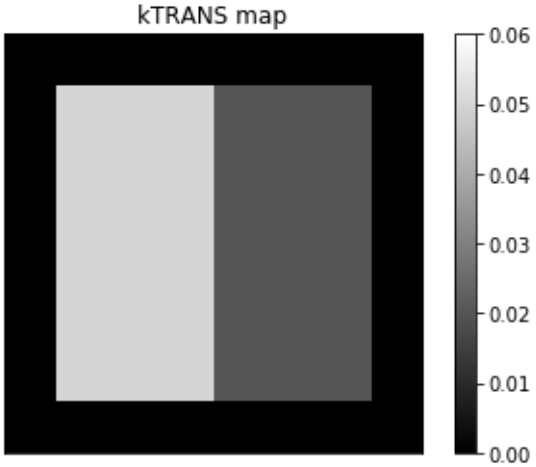
When discretizing this phantom at a particular resolution, parameters θ_i are chosen based on the value of each phantom at the center of each voxel. Thus partial voluming is

not considered in the current experiments. A central slice through $z = 0$ is shown in Figure 5.1 for $256 \times 256 \times 256$ and $16 \times 16 \times 16$ matrix sizes.

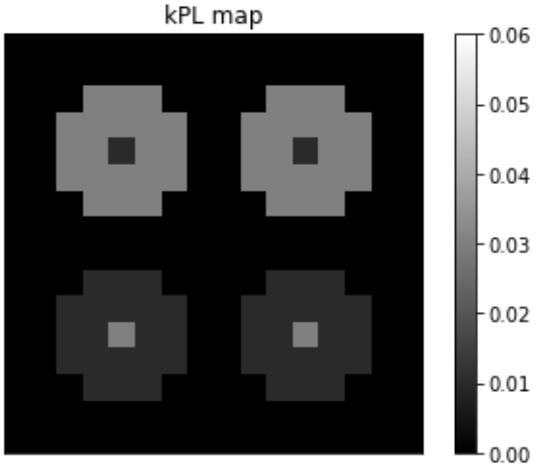


(a) k_{TRANS} map at $256 \times 256 \times 256$

(b) k_{PL} map at $256 \times 256 \times 256$



(c) k_{TRANS} map at $16 \times 16 \times 16$



(d) k_{PL} map at $16 \times 16 \times 16$

Figure 5.1: Slice through $z = 0$ of the 3D dynamic phantom at varying resolutions.

Reconstruction at a variety of noise levels

To generate simulated data for validating our algorithm, we simulate trajectories for each voxel of the $16 \times 16 \times 16$ dynamic phantom described in the previous section according to the model (5.1) with arterial input $u(t) = k_{TRANS} A_0 (t - t_0)^\gamma e^{(-(t-t_0)/\beta}$ added to the pyruvate compartment, and states scaled by $\cos(\alpha_{P/L}(k))$ and measured outputs scaled by $\sin(\alpha_{P/L}(k))$ each time that simulated data are collected, where $\alpha_{P/L}(k)$ is a spectrally-selective flip angle

applied to spins in the P or L compartment during acquisition k. An optimized dynamic flip angle sequence based on the method of [106] is used for the simulation, and shown in Figure 5.2. This same flip angle sequence is also used for a majority of the *in vivo* experiments.

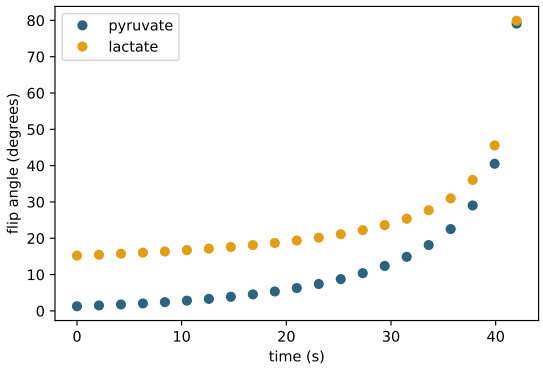


Figure 5.2: Dynamic flip angle sequence used for experimental validation

We then add independent, identically-distributed (iid) Gaussian noise at a variety of SNR levels, measured based on the SNR in the lactate channel corresponding to the peak lactate level. Simulated time series and image data are shown in Figure 5.3.

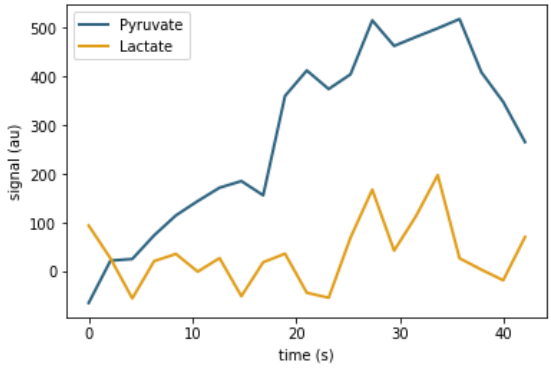
For SNR levels of 8, 4, 2, and 1, we fit the model (5.2) to the data using the loss function (5.3) and the regularization $r(\theta) = \lambda_1\|\nabla\theta\|_1 + \lambda_2\|\theta\|_2^2$ with λ_1 =1e06 and λ_2 =1e08 using the hybrid ADMM z update described in Equation (5.7). The values of λ_1 and λ_2 are selected such that the total absolute error is minimized (see Section 5.5). Before fitting, the simulated data are scaled by $1/\sin(\alpha_{P/L}(k))$ to counteract the effect of the time-varying flip angle sequence. In Figure 5.4 we compare the results of this constrained fit against two competing methods: independent voxel-wise fit (equivalent to our method with $\lambda_1 = \lambda_2 = 0$) and independent voxel-wise fit followed by total variation denoising of the resulting parameter map. We see that the constrained reconstruction allows accurate parameter maps to be generated in high noise regimes where the competing methods have difficulty.

Quantitative Improvements

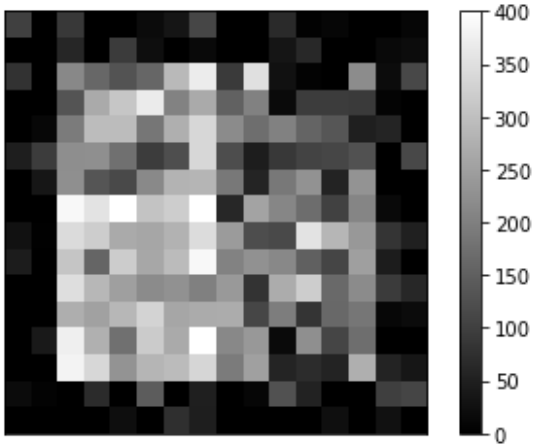
In addition to the qualitative benefits of spatial regularization demonstrated in the previous section, regularization can also lead to quantitative improvements in the estimates of dynamic parameters. In simulation experiments where we have access to the ground truth values of the model parameters, we can quantify the improvement in estimates $\hat{\theta}$ of θ via the total absolute error

$$\|\hat{\theta} - \theta\|_1 = \sum_{i \in V} |\hat{k}_{PL_i} - k_{PL_i}|.$$

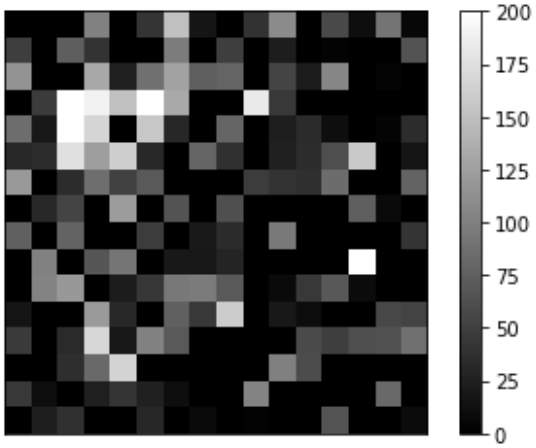
In Figure 5.5 we plot the total absolute error for various values of the regularization parameters λ_1 and λ_2. This experiment was performed using the $16 \times 16 \times 16$ phantom from



(a) Sample time series data from a high k_{TRANS}, high k_{PL} voxel



(b) Pyruvate image slice through $z = 0$



(c) Lactate image slice through $z = 0$

Figure 5.3: Simulated data generated at a maximum lactate SNR level of 2.

Figure 5.1 with a maximum lactate SNR value of 2.0. We see that small values of λ_1 and λ_2 lead to larger quantitative errors in the parameter maps than the optimized values λ_1 =1e06 and λ_2 =1e08 used in the previous section. Note that the optimal values will depend on a number of factors potentially including the geometry and sparsity of the phantom, and the noise distribution, SNR and signal amplitude in the dynamic images. Thus by appropriately choosing λ_1 and λ_2, we can achieve quantitative improvements in the parameter map in addition to the qualitative improvements we have already demonstrated.

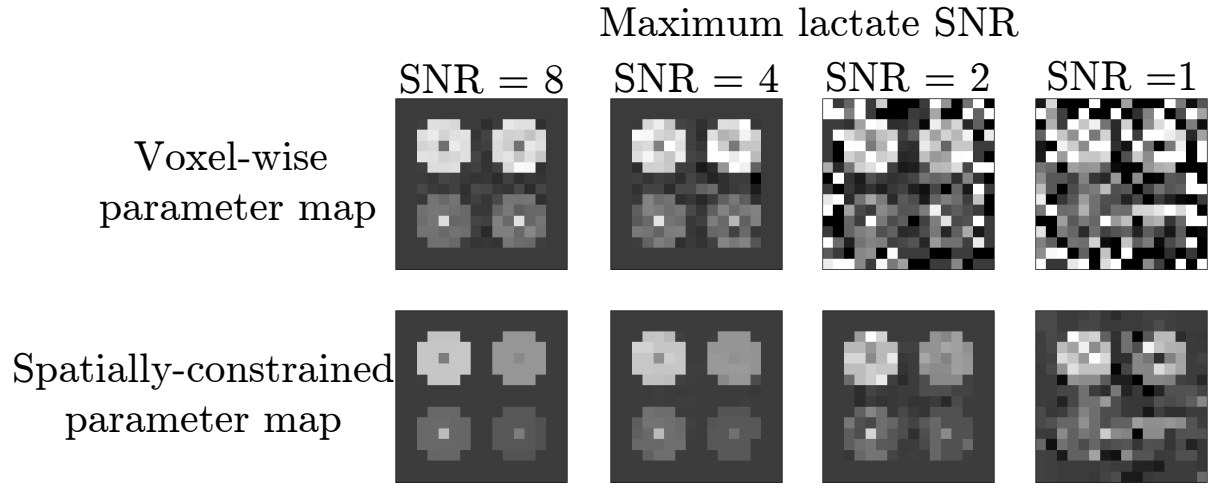


Figure 5.4: Results of simulated k_{PL} mapping experiment

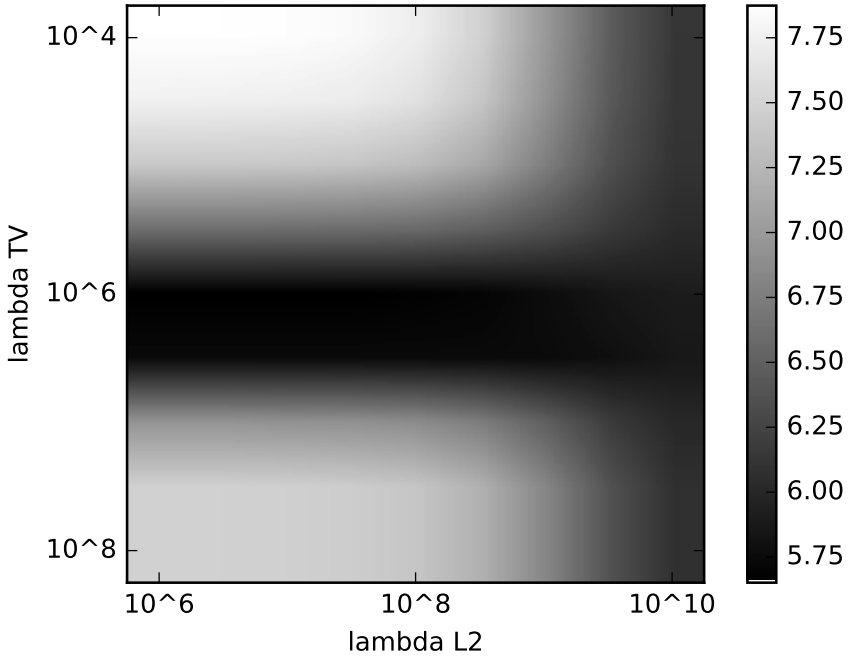


Figure 5.5: Total absolute estimation error for k_{PL} for various values of the regularization parameters λ_1 and λ_2.

5.6 In Vivo Results and Discussion

We now move on to experiments on a number of datasets collected *in vivo*. In contrast to the simulation experiments, we no longer have access to ground truth values of the model parameters to make quantitative comparisons. However, we will use the *in vivo* experiments to demonstrate that the spatially-constrained parameter mapping technique leads to qualitative improvements in the parameter maps.

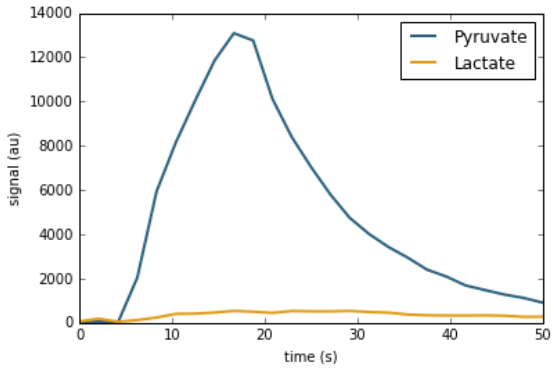
We begin with an experiment in healthy rats where we can collect high SNR data. For these data, we add artificial noise to demonstrate how the spatially-constrained parameter mapping technique can be used to allow reconstruction in low SNR regimes, for realistic anatomies. We then apply this technique to the analysis of a number of low SNR clinical datasets collected in prostate cancer patients. These experiments demonstrate that spatio-temporally constrained kinetic modelling can be used to generate improved metabolic parameter maps from low SNR experimental data.

High SNR rat kidney data analysis

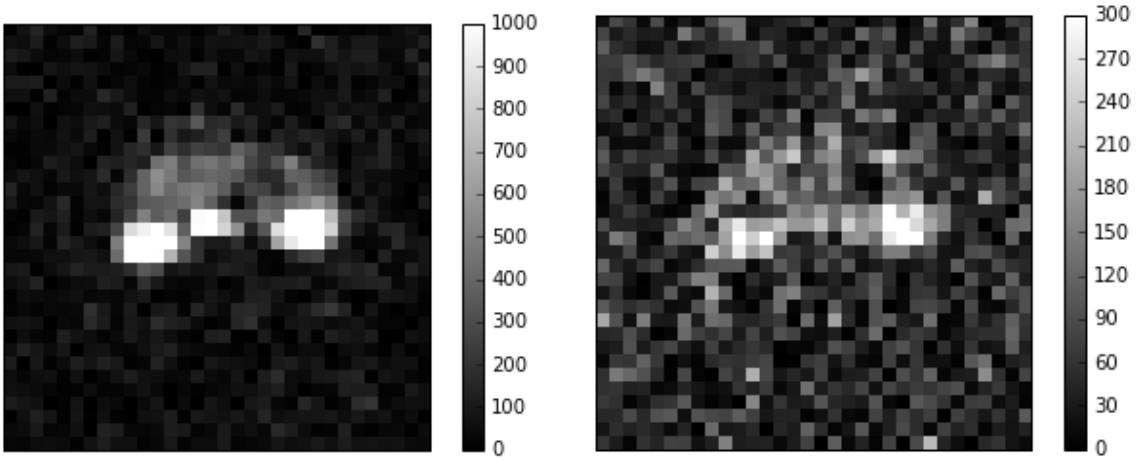
We begin by analyzing a metabolic dataset acquired in healthy Sprague-Dawley rats on a 3T MRI scanner (MR750, GE Healthcare). 2.5mL of 80mM hyperpolarized [1-13C]pyruvate was injected over 15s, and data acquisition coincided with the start of injection. Metabolites from a single slice were individually excited with a singleband spectral-spatial RF pulse and encoded with a single-shot EPI readout, an in-plane resolution of 3 x 3mm, a 15mm slice thickness centered on the kidneys, and a 2s sampling interval. The resulting dynamic image sequences, are relatively high SNR with Rician noise resulting from magnitude images, are shown in Figure 5.6.

In Figure 5.7 we compare a spatially constrained fit of the data against an independent voxel-wise fit. The voxel-wise fit is masked to only show k_{PL} fit in the highly perfused regions where the total area under the pyruvate curve (AUC) is greater than 2e04. We see that the constrained fit leads to more smoothly-varying maps. Additionally, the Tikhonov regularization helps alleviate problems with artificially high k_{PL} estimates in the background region and tissues with low perfusion, a common problem with k_{PL} mapping from Rician-distributed data. This leads to more realistic k_{PL} values in the intestinal tissue proximal to the kidneys without significantly affecting the k_{PL} estimates in the kidney voxels, and also removes the need to mask the images to the high perfusion region.

To investigate the robustness of this technique to noise, we perform a sequence of experiments in which artificial iid Gaussian noise of varying strengths is added to the *in vivo* data before fitting k_{PL}. This allows us to replicate the results of Figure 5.4 with more realistic anatomy. We see that qualitatively, the spatially-constrained fit is more robust to strong noise than the independent fit.



(a) Sample time series data from high lactate SNR voxel



(b) Pyruvate image at time $t = 50$s

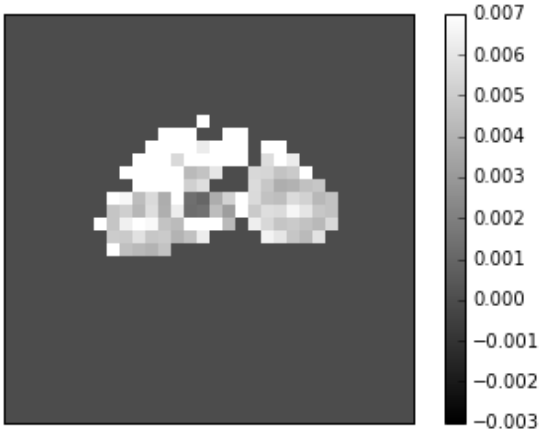
(c) Lactate image at time $t = 50$s

Figure 5.6: Dynamic metabolite images collected in the healthy rat experiment.

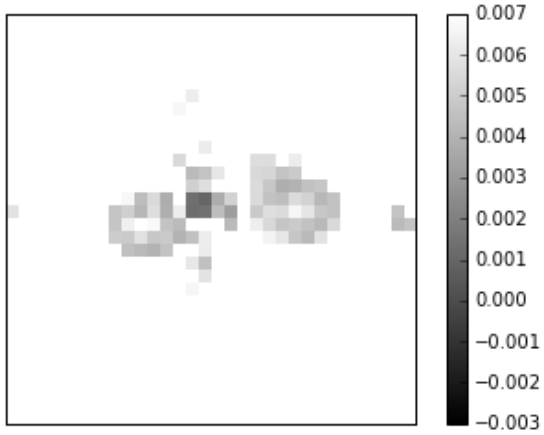
Human prostate cancer data analysis

To validate this technique on clinically-relevant data, we have analyzed three prostate cancer datasets collected during clinical experiments at UCSF. Imaging was performed using a 3T GE scanner using a abdominal clamshell ^{13}C transmission coil and an endo-rectal receive coil. The injected solution consisted of 220-260 mM [1-^{13}C]-pyruvate at a dose of 0.43 mL/kg. Dissolution DNP was performed using a 5T SpinLab polarizer (GE Healthcare). Before injection the electron paramagnetic agent OX063 (GE Healthcare) is filtered out, and automated pH, temperature, polarization, volume and EPA concentration tests were performed.

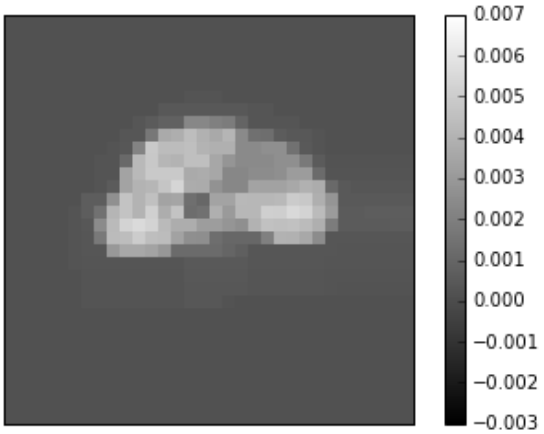
Images were encoded using two techniques. One set of images labeled "EPI" were collected using a spectrally-selective excitation with an echo-planar (EPI) readout [40]. Two sets of images labelled "EPSI" were collected using a blipped EPSI acquisition with a com-



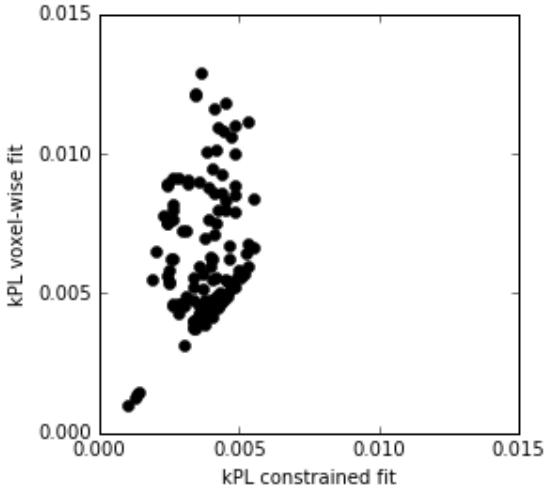
(a) Independent voxel-wise fit masked to region with pyruvate AUC > 2e04.



(b) Independent voxel-wise fit without masking exhibits high k_{PL} values in the background region.



(c) Spatially-constrained fit with λ_1 = 1e07 and λ_2 = 1e10.



(d) Scatterplot of constrained and unconstrained k_{PL} fits.

Figure 5.7: Comparison of unconstrained and constrained k_{PL} maps fit to the healthy rat dataset.

pressed sensing reconstruction [55].

Raw space/time/chemical data reconstructed from the EPI acquisition are shown in Figure 5.9. The raw data are rather noisy and also difficult to interpret for metabolic activity due to 3D spatial, temporal and chemical dimensions.

We fit 3D k_{PL} parameter maps to the data using the constrained reconstruction method. Regularization strengths λ_1 and λ_2 are selected manually based on the qualitative appearance of the parameter maps. Due to the quick parameter map estimation enabled by the

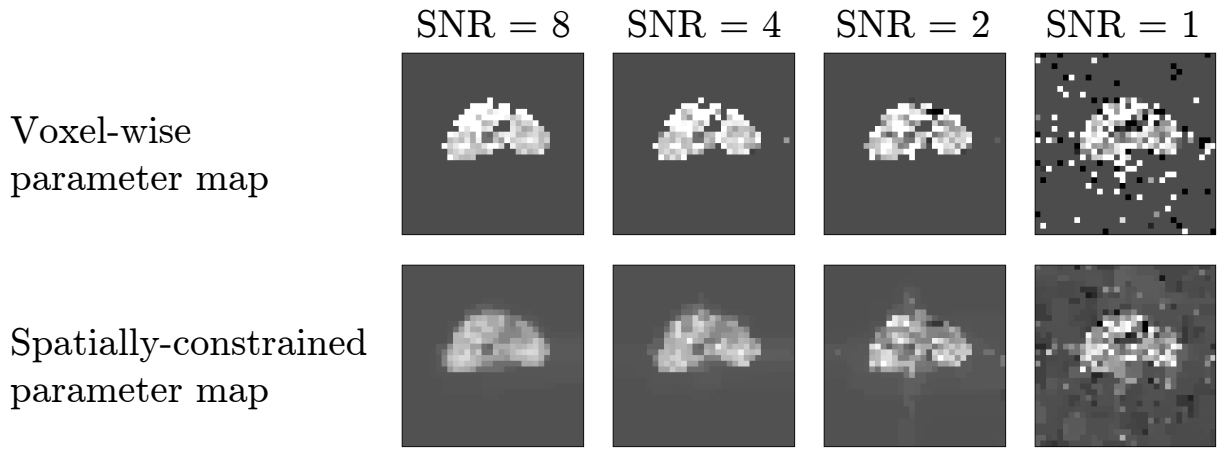
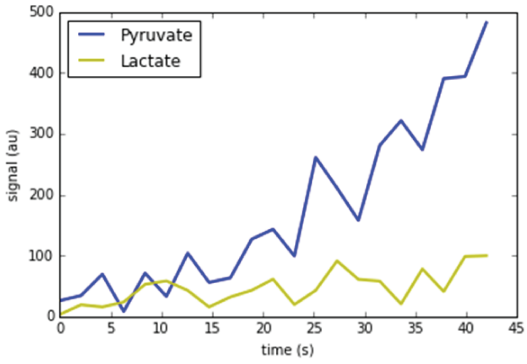
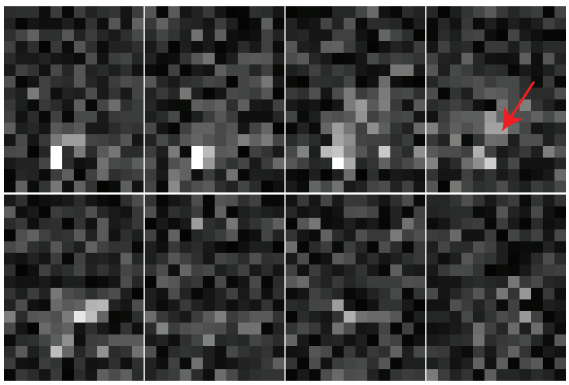


Figure 5.8: Comparison of k_{PL} maps at various artificial noise levels.



(a) Time series data at pyruvate and lactate frequencies corresponding to the voxel indicated in red.



(b) Lactate data from 8 of the 16 slices at the time of the final acquisition t=42 seconds from the start of injection.

Figure 5.9: Sample of raw EPI data collected in a prostate cancer patient.

parallelized ADMM iteration, it is possible to perform this exploration relatively efficiently. The results are shown in Figures 5.10, 5.11 and 5.12. Additionally, in Figure 5.13 we compare the resulting parameter maps for a variety of values for the regularization parameters λ_1 and λ_2. We see that with an appropriate choice of regularization, we can recover qualitatively satisfying parameter maps for a variety of datasets. Note that the regularization parameters differ significantly between the EPI and EPSI acquisitions due mainly to the different amplitudes of the raw dynamic image data.

Additionally, in Figure 5.14 we compare unconstrained and constrained fits on the single dataset from the EPI acquisition. The fits are overlaid on ^{1}H images of the anatomy using SIVIC [26]. The unconstrained fit is masked to prostate-adjacent voxel due to non-identifiability in the background region, whereas this in not necessary for the constrained

fit. Note that the strong regularization leads to significant quantitative shrinkage of the k_{PL} estimates. However, it improves the qualitative indication of the highly metabolically-active region. Elevated k_{PL} in the prostate base is consistent with biopsy proven Gleason 3+4 cancer.

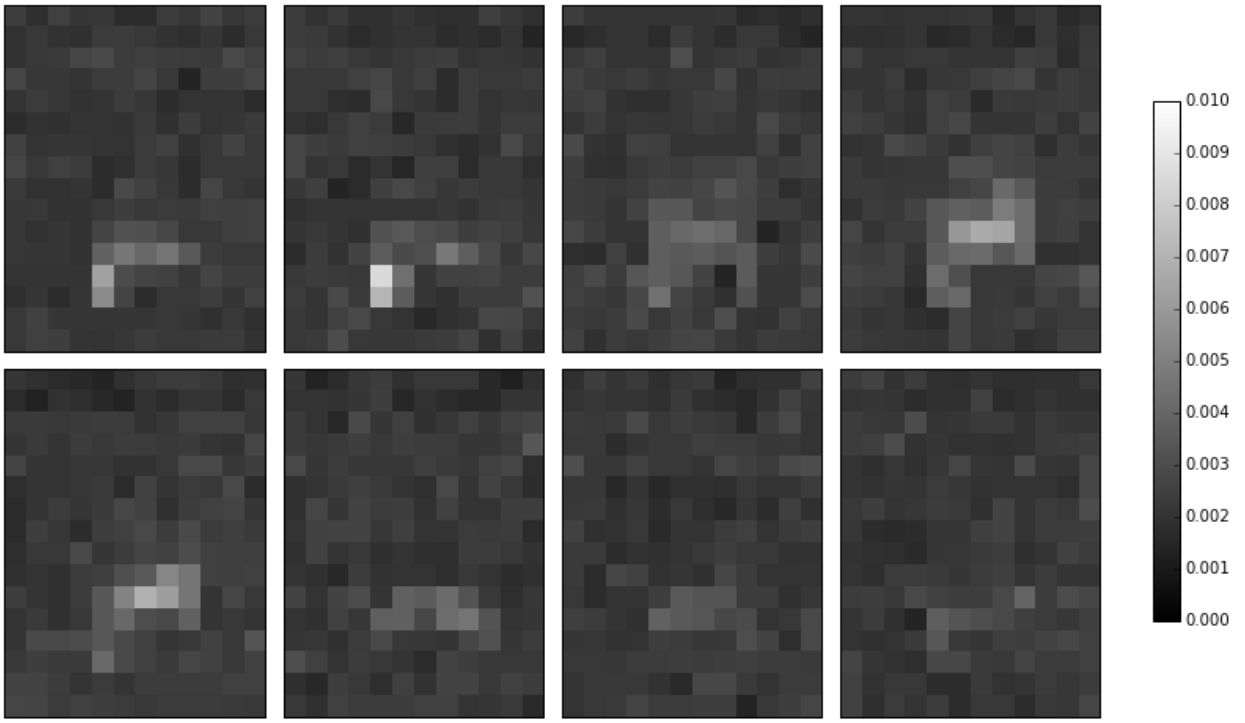


Figure 5.10: Spatially-constrained k_{PL} maps computed with $\lambda_1 = 5e04$ and $\lambda_2 = 1e09$ from the 3D EPI human prostate cancer dataset. Each image corresponds to one slice through the prostate.

5.7 Conclusion

We have demonstrated that constrained reconstruction of parameter maps via spatial regularization improves the qualitative performance of model-based parameter mapping. We have shown this first in simulated experiments where we can also demonstrate quantitative improvements in the parameter estimates. The results of the *in vivo* studies echo the qualitative benefits of constraining parameter maps through regularization, and validate that the ADMM-based algorithm we have developed enables efficient reconstruction of parameter maps for problems of practical interest by exploiting the objective function's structure.

Looking forward, the ability to exploit spatial and temporal correlations in the data for denoising could potentially help to overcome problems with low SNR in hyperpolarized 13C MRI, enabling the reconstruction of higher resolution images. Also, developing methods to

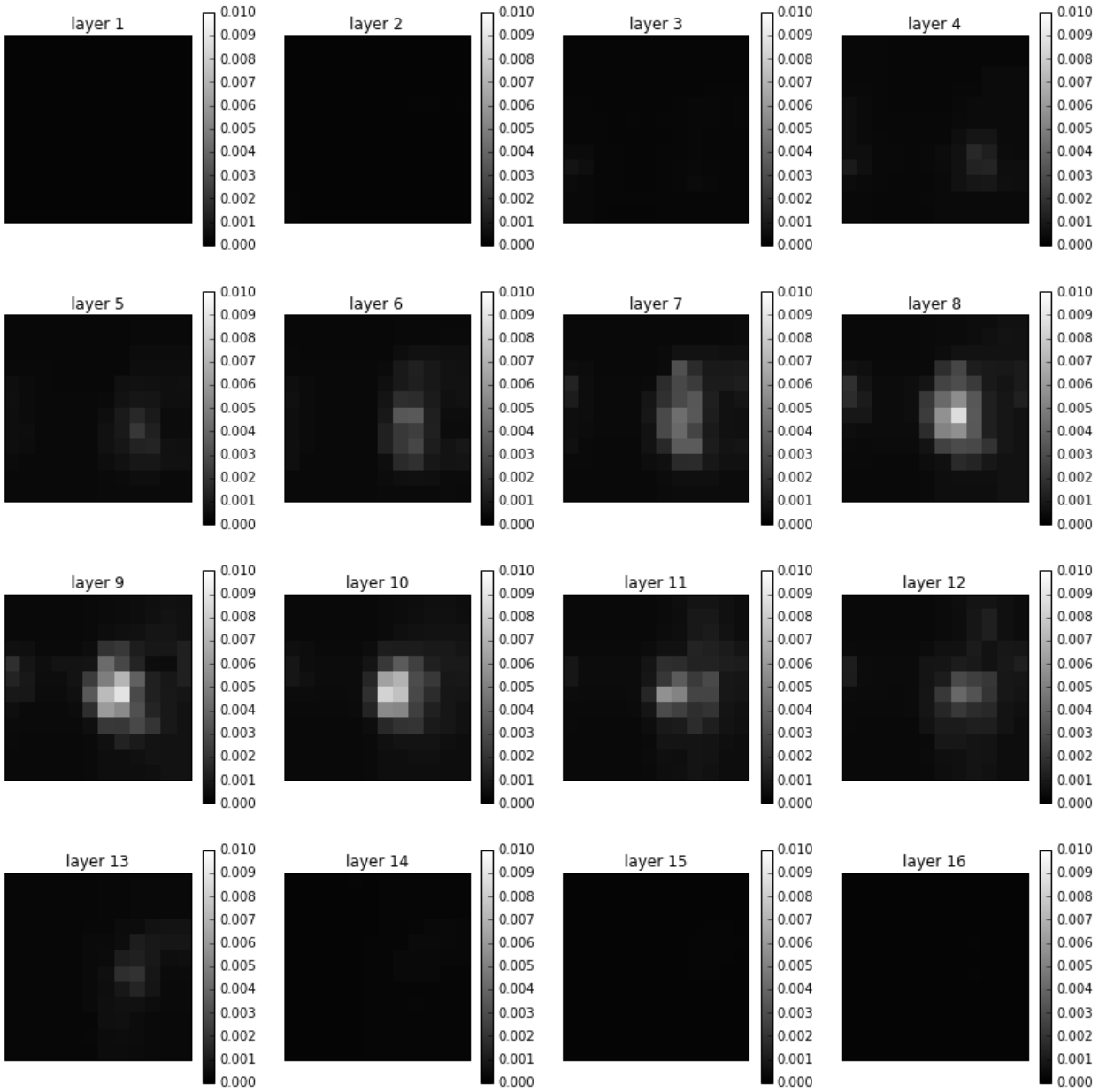


Figure 5.11: Spatially-constrained k_{PL} maps computed with $\lambda_1 = 1e18$ and $\lambda_2 = 1e20$ from the 3D EPSI human prostate cancer dataset pc9154.

choose the regularization strength hyperparameters systematically may help to improve the quantitative bias seen in some of the *in vivo* experiments. In particular, methods based on Shure's unbiased risk estimate used for selecting hyperparameters in total variation denoising applications [98] can likely be adapted to this context.

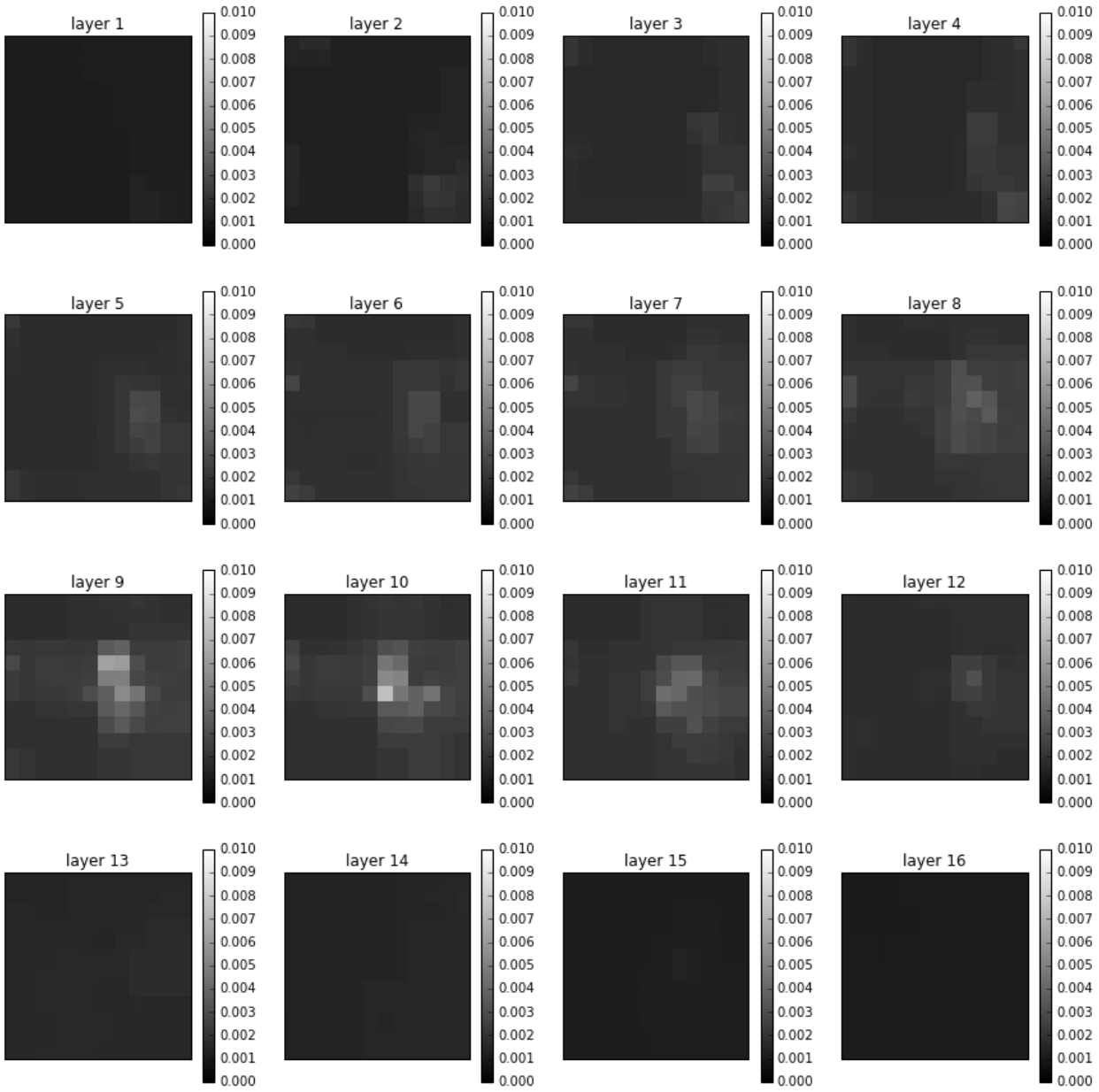


Figure 5.12: Spatially-constrained k_{PL} maps computed with $\lambda_1 = 2e17$ and $\lambda_2 = 1e14$ from the 3D EPSI human prostate cancer dataset pc9375.

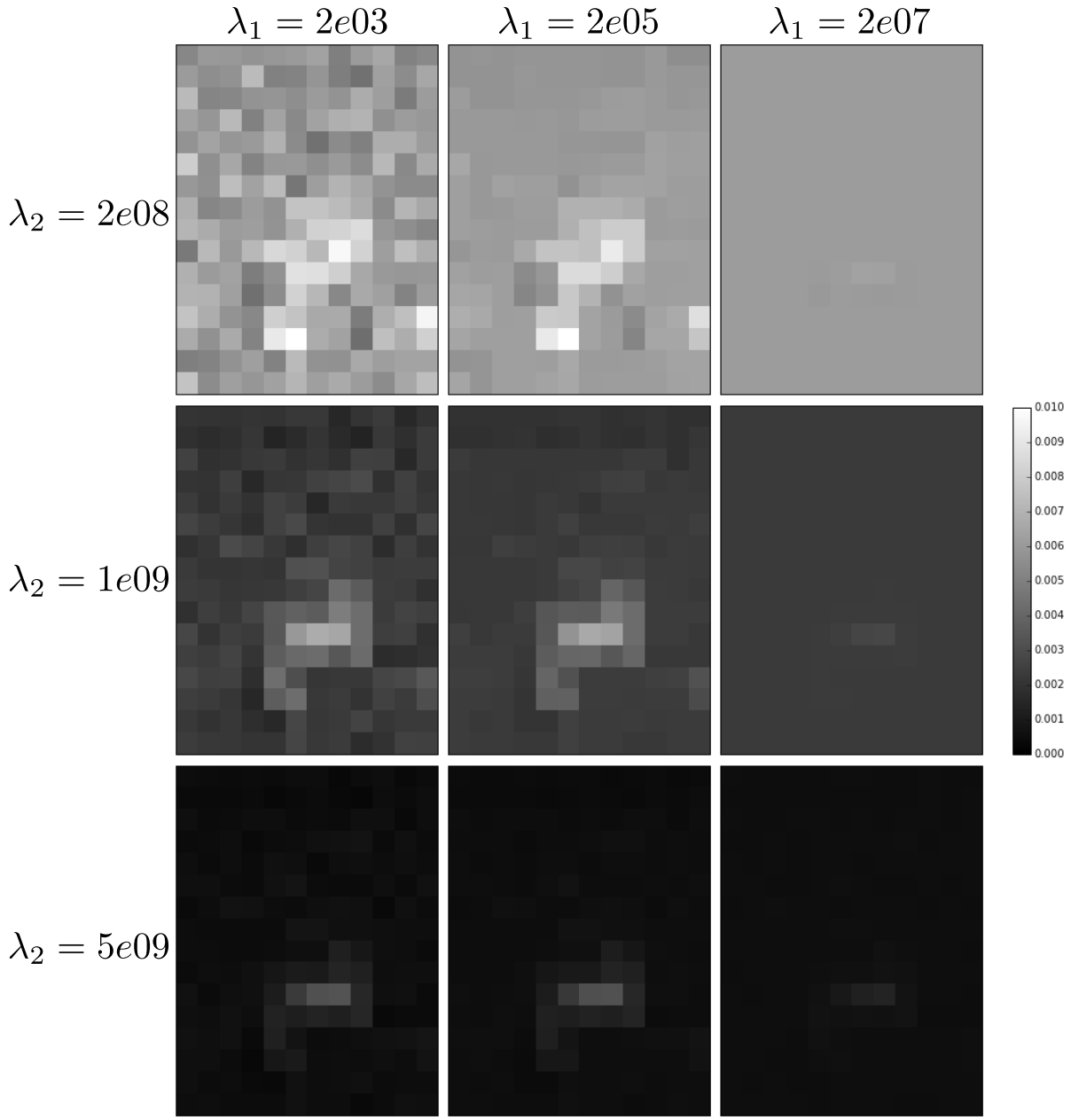
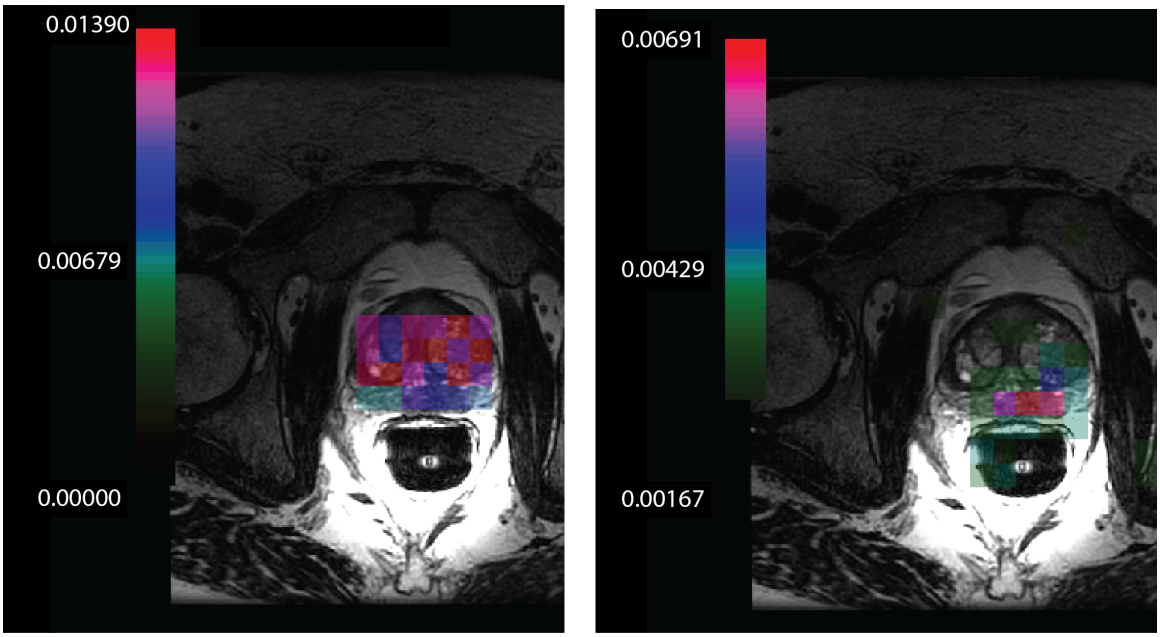


Figure 5.13: Constrained fits with different regularization strengths compared on the 3D EPI human prostate cancer dataset.



(a) Unconstrained fit ($\lambda_1 = \lambda_2 = 0$) masked to prostate region

(b) Spatially-constrained fit with $\lambda_1 = 3e06$ and $\lambda_2 = 3e09$.

Figure 5.14: Comparison of unconstrained and constrained k_{PL} maps fit to the EPI data overlaid on prostate anatomy.

Bibliography

[1] Mark J. Albers et al. "Hyperpolarized ^{13}C Lactate, Pyruvate, and Alanine: Noninvasive Biomarkers for Prostate Cancer Detection and Grading". In: *Cancer Research* 68.20 (2008), pp. 8607–8615.

[2] Fernando Alvarez and Nancy L Stokey. "Dynamic programming with homogeneous functions". In: *Journal of economic theory* 82.1 (1998), pp. 167–189.

[3] Murat Arcak, Chris Meissen, and Andrew Packard. "Symmetry Reduction". In: *Networks of Dissipative Systems: Compositional Certification of Stability, Performance, and Safety*. Springer, 2016, pp. 55–62.

[4] Jan H. Ardenkjær-Larsen et al. "Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR". In: *Proceedings of the National Academy of Sciences* 100.18 (Sept. 2003), pp. 10158–10163. ISSN: 1091-6490. DOI: 10.1073/pnas.1733835100.

[5] Naeim Bahrami et al. "Kinetic and perfusion modeling of hyperpolarized ^{13}C pyruvate and urea in cancer with arbitrary RF flip angles". In: *Quantitative Imaging in Medicine and Surgery* 4.1 (2014).

[6] James A. Bankson et al. "Kinetic modeling and constrained reconstruction of hyperpolarized 1-13C-pyruvate offers improved metabolic imaging of tumors". In: *Cancer Research* (2015).

[7] A. Barrau and S. Bonnabel. "The invariant extended Kalman filter as a stable observer". In: *Arxiv preprint* (2014), arXiv:1410.1465.

[8] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[9] Richard Bellman and Stuart Dreyfus. "Functional Approximations and Dynamic Programming". In: *Mathematical Tables and Other Aids to Computation* 13.68 (1959), pp. 247–251.

[10] Matt A. Bernstein, Kevin F. King, and Xiaohong Joe Zhou. "BASIC PULSE SEQUENCES". In: *Handbook of MRI Pulse Sequences*. Academic Press, 2004, pp. 579–647.

[11] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control, Volume I*. 3rd. Athena Scientific, 2005.

[12] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control, Volume II: Approximate Dynamic Programming*. 4th. Athena Scientific, 2012.

[13] Anthony M. Bloch et al. "Nonholonomic mechanical systems with symmetry". In: *Arch. Rational Mech. Anal.* 136 (1996), pp. 21–99.

[14] F. Bloch. "Nuclear Induction". In: *Phys. Rev.* 70 (7-8 1946), pp. 460–474.

[15] George Bluman and Sukeyuki Kumei. *Symmetries and Differential Equations*. Applied Mathematical Sciences. Springer, 2013.

[16] S. Bonnabel, P. Martin, and P. Rouchon. "Symmetry-Preserving Observers". In: *IEEE Transactions on Automatic Control* 53.11 (2008), pp. 2514–2526.

[17] Stephen Boyd et al. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". In: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122.

[18] Kevin M. Brindle. "NMR methods for measuring enzyme kinetics in vivo". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 20.3 (1988), pp. 257–293.

[19] Francesco Bullo and Richard M. Murray. "Tracking for fully actuated mechanical systems: a geometric framework". In: *Automatica* 35.1 (1999), pp. 17–34.

[20] É. Cartan. *La théorie des groupes finis et continus et la géométrie différentielle: traitées par la méthode du repère mobile*. Cahiers scientifiques. Gauthier-Villars, 1937.

[21] Kathryn Chaloner and Isabella Verdinelli. "Bayesian Experimental Design: A Review". In: *Statistical Science* 10 (1995), pp. 273–304.

[22] Chi-Tsong Chen. *Linear System Theory and Design*. 3rd. New York, NY, USA: Oxford University Press, Inc., 1998. ISBN: 0195117778.

[23] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. "Atomic decomposition by basis pursuit". In: *SIAM review* 43.1 (2001), pp. 129–159.

[24] Peter A. Clarksonz and Elizabeth L. Mansfield. "Symmetry reductions and exact solutions of a class of nonlinear heat equations". In: *Physica D: Nonlinear Phenomena* 70.3 (1994), pp. 250–288.

[25] Harald Cramér. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[26] Jason C. Crane, Marram P. Olson, and Sarah J. Nelson. "SIVIC: Open-Source, Standards-Based Software for DICOM MR Spectroscopy Workflows". In: *International Journal of Biomedical Imaging* 2013 (2013), p. 169526.

[27] Charles H. Cunningham et al. "Hyperpolarized ^{13}C Metabolic MRI of the Human Heart: Novelty and Significance". In: *Circulation Research* 119.11 (2016), pp. 1177–1182.

[28] M. Davies et al. "A Compressed Sensing Framework for Magnetic Resonance Fingerprinting". In: *SIAM Journal on Imaging Sciences* 7.4 (2014), pp. 2623–2656.

[29] Sam E Day et al. "Detecting tumor response to treatment using hyperpolarized ^{13}C magnetic resonance imaging and spectroscopy". In: *Nature Medicine* 11 (2007), pp. 1382–1387.

[30] Martin H. Deppe et al. "Slice profile effects in 2D slice-selective MRI of hyperpolarized nuclei". In: *Journal of Magnetic Resonance* 202.2 (2010), pp. 180–189.

[31] Belmiro P. M. Duarte and Weng Kee Wong. "A semi-infinite programming based algorithm for finding minimax optimal designs for nonlinear models". In: *Statistics and Computing* 24.6 (2013), pp. 1063–1080.

[32] E. Allen Emerson and A. Prasad Sistla. "Symmetry and model checking". In: *Formal Methods in System Design* 9.1 (1996), pp. 105–131.

[33] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. A Series of books in the mathematical sciences. W. H. Freeman, 1979.

[34] Robert A. Gatenby and Robert J. Gillies. "Why do cancers have high aerobic glycolysis?" In: *Nat Rev Cancer* 4.11 (2004), pp. 891–899.

[35] Michel Gevers et al. "Optimal experiment design for open and closed-loop system identification". In: *Communications in Information and Systems* 11.3 (2011), pp. 197–224.

[36] Michel X. Goemans and David P. Williamson. "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming". In: *J. Assoc. Comput. Mach.* 42 (1995), pp. 1115–1145.

[37] Klaes Golman and Stefan J. Petersson. "Metabolic Imaging and Other Applications of Hyperpolarized $^{13}C^{1}$". In: *Academic Radiology* 13.8 (Aug. 2006), pp. 932–942. DOI: 10.1016/j.acra.2006.06.001.

[38] Klaes Golman et al. "Molecular imaging with endogenous substances". In: *Proceedings of the National Academy of Sciences* 100.18 (2003), pp. 10435–10439.

[39] G.C. Goodwin and R.L. Payne. *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, 1977.

[40] Jeremy W. Gordon, Daniel B. Vigneron, and Peder E.Z. Larson. "Development of a symmetric echo planar imaging framework for clinical translation of rapid dynamic hyperpolarized 13C imaging". In: *Magnetic Resonance in Medicine* 77.2 (2017), pp. 826–832. ISSN: 1522-2594.

[41] Jeremy W Gordon et al. "Ramp-Sampled, Symmetric EPI for Rapid Dynamic Metabolic Imaging of Hyperpolarized ^{13}C Substrates on a Clinical MRI Scanner". In: *Proceedings of the 23rd Annual Meeting of ISMRM, Toronto, Ontario, Abstract 4717*. 2015.

[42] Michael Grant and Stephen Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. http://cvxr.com/cvx. Mar. 2014.

[43] J. Grizzle and S. Marcus. "Optimal control of systems possessing symmetries". In: *IEEE Transactions on Automatic Control* 29.11 (1984), pp. 1037–1040.

[44] H. Gudbjartsson and S. Patz. "The Rician distribution of noisy MRI data". In: *Magnetic Resonance in Medicine* 34.6 (1995), pp. 910–914. ISSN: 0740-3194.

[45] Yi Guo et al. "Highly Accelerated Brain DCE MRI with Direct Estimation of Pharmacokinetic Parameter Maps". In: *Proceedings of the 23rd Annual Meeting of ISMRM, Toronto, Ontario, Abstract 1093*. 2015.

[46] Yi Guo et al. "High-resolution whole-brain DCE-MRI using constrained reconstruction: Prospective clinical evaluation in brain tumor patients". In: *Medical Physics* 43.5 (2016), pp. 2013–2023.

[47] Crystal Harrison et al. "Comparison of kinetic models for analysis of pyruvate-to-lactate exchange by hyperpolarized ^{13}C NMR". In: *NMR in Biomedicine* 25.11 (2012), pp. 1286–1294.

[48] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer, 2009.

[49] C. Helmberg, F. Rendl, and R. Weismantel. "Quadratic knapsack relaxations using cutting planes and semidefinite programming". In: *Integer Programming and Combinatorial Optimization*. Ed. by William H. Cunningham, S.Thomas McCormick, and Maurice Queyranne. Vol. 1084. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1996, pp. 175–189.

[50] Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. John Wiley & Sons, Inc., 2009, pp. 1–21.

[51] Bronislaw Jakubczyk. "Symmetries of nonlinear control systems and their symbols". In: *Canadian Math. Conf. Proceed*. Vol. 25. 1998, pp. 183–198.

[52] S. M. Kazan et al. "Kinetic modeling of hyperpolarized ^{13}C pyruvate metabolism in tumors using a measured arterial input function". In: *Magnetic Resonance in Medicine* 70.4 (2013), pp. 943–953. ISSN: 1522-2594. DOI: 10.1002/mrm.24546.

[53] O. Khegai et al. "Apparent rate constant mapping using hyperpolarized [1-^{13}C]pyruvate". In: *NMR in Biomedicine* 27.10 (2014), pp. 1256–1265.

[54] Marta Kwiatkowska, Gethin Norman, and David Parker. "Symmetry Reduction for Probabilistic Model Checking". In: *Computer Aided Verification: 18th International Conference, CAV 2006, Seattle, WA, USA*. Ed. by Thomas Ball and Robert B. Jones. Springer, 2006, pp. 234–248.

[55] Peder E. Z. Larson et al. "Fast dynamic 3D MR spectroscopic imaging with compressed sensing and multiband excitation pulses for hyperpolarized 13C studies". In: *Magnetic Resonance in Medicine* 65.3 (2011), pp. 610–619.

[56] Peder E.Z. Larson et al. "Multiband excitation pulses for hyperpolarized ^{13}C dynamic chemical-shift imaging". In: *Journal of Magnetic Resonance* 194.1 (2008), pp. 121–127.

[57] Sajan Goud Lingala et al. "Accelerated DCE MRI using constrained reconstruction based on pharmaco-kinetic model dictionaries". In: *Proceedings of the 23rd Annual Meeting of ISMRM, Toronto, Ontario, Abstract 1539*. 2015.

[58] L. Ljung. *System Identification: Theory for the User*. Pearson Education, 1999.

[59] Zhi-Quan Luo et al. "Semidefinite Relaxation of Quadratic Optimization Problems". In: *IEEE Signal Processing Magazine* 27.3 (2010), pp. 20–34.

[60] John Lygeros, Claire Tomlin, and Shankar Sastry. *Hybrid Systems: Modeling, Analysis and Control*. UC Berkeley / ETH Zurich lecture notes. 2008. URL: `https://web.archive.org/web/20100723205138/http://www-inst.cs.berkeley.edu/~ee291e/sp09/handouts/book.pdf`.

[61] Dan Ma et al. "Magnetic resonance fingerprinting". In: *Nature* 495.7440 (Mar. 2013), pp. 187–192.

[62] Sonam I. Machingal. "Sampling Strategies for Hyperpolarized Carbon-13 Imaging". MA thesis. USA: University of California, San Francisco, 2014.

[63] J. Maidens and M. Arcak. "A note on optimal experiment design for nonlinear systems using dynamic programming". In: *ArXiv e-prints* (2015). URL: `http://arxiv.org/abs/1503.07232`.

[64] John Maidens and Murat Arcak. "Semidefinite relaxations in optimal experiment design with application to substrate injection for hyperpolarized MRI". In: *Proceedings of the American Control Conference (ACC)*. 2016, pp. 2023–2028.

[65] John Maidens, Peder E. Z. Larson, and Murat Arcak. "Optimal experiment design for physiological parameter estimation using hyperpolarized carbon-13 magnetic resonance imaging". In: *Proceedings of the American Control Conference (ACC)*. 2015.

[66] John Maidens, Andrew Packard, and Murat Arcak. "Parallel dynamic programming for optimal experiment design in nonlinear systems". In: *Proceedings of the Conference on Decision and Control (CDC)*. 2016, pp. 2894–2899.

[67] John Maidens et al. "Optimizing flip angles for metabolic rate estimation in hyperpolarized carbon-13 MRI". In: *IEEE Transactions on Medical Imaging* 35.11 (2016), pp. 2403–2412.

[68] John Maidens et al. "Symmetry reduction for dynamic programming and application to MRI". In: *Proceedings of the American Control Conference (ACC)*. To appear. 2017.

[69] J. Maidens et al. "Spatio-temporally constrained reconstruction for hyperpolarized carbon-13 MRI using kinetic models". In: *Proceedings of the ISMRM Annual Meeting*. 2017. URL: http://submissions.mirasmart.com/ISMRM2017/ViewSubmissionPublic.aspx?sei=GH0eaFQTF.

[70] Ian R. Manchester. "Amplitude-Constrained Input Design: Convex Relaxation and Application to Clinical Neurology". In: *SYSID2012: IFAC Symposium on System Identification*. 2012.

[71] Ian R. Manchester. "Input design for system identification via convex relaxation". In: *49th IEEE Conference on Decision and Control (CDC)*. 2010, pp. 2041–2046.

[72] Philippe Martin, Pierre Rouchon, and Joachim Rudolph. "Invariant tracking". In: *ESAIM: Control, Optimisation and Calculus of Variations* 10.1 (2004), pp. 1–13.

[73] D W McRobbie, R A Lerski, and K Straughan. "Slice profile effects and their calibration and correction in quantitative NMR imaging". In: *Physics in Medicine and Biology* 32.8 (1987), pp. 971–983.

[74] Matthew E. Merritt et al. "Hyperpolarized ^{13}C allows a direct measure of flux through a single enzyme-catalyzed step by NMR". In: *Proceedings of the National Academy of Sciences* 104.50 (2007), pp. 19773–19777.

[75] Cornelius von Morze et al. "Imaging of blood flow using hyperpolarized [^{13}C]Urea in preclinical cancer models". In: *Journal of Magnetic Resonance Imaging* 33.3 (2011), pp. 692–697. ISSN: 1522-2586. DOI: 10.1002/jmri.22484.

[76] Shravan Matthur Narayanamurthy and Balaraman Ravindran. "Efficiently Exploiting Symmetries in Real Time Dynamic Programming". In: *Proceedings of the 20th International Joint Conference on Aritifical Intelligence*. IJCAI'07. Hyderabad, India: Morgan Kaufmann, 2007, pp. 2556–2561.

[77] Sarah J. Nelson et al. "Metabolic Imaging of Patients with Prostate Cancer Using Hyperpolarized [1-^{13}C]Pyruvate". In: *Science Translational Medicine* 5.198 (2013), 198ra108.

[78] A. Nemirovski, C. Roos, and T. Terlaky. "On maximization of quadratic form over intersection of ellipsoids with common center". In: *Mathematical Programming* 86.3 (1999), pp. 463–473.

[79] Yurii Nesterov. "Semidefinite relaxation and nonconvex quadratic optimization". In: *Optimization Methods and Software* 9.1-3 (1998), pp. 141–160.

[80] Dwight G. Nishimura. *Principles of Magnetic Resonance Imaging*. Lulu, 2010.

[81] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2006.

[82] Roger Fandom Noubiap and Wilfried Seidel. "A minimax algorithm for constructing optimal symmetrical balanced designs for a logistic regression model". In: *Journal of Statistical Planning and Inference* 91.1 (2000), pp. 151–168.

[83] Tomoki Ohsawa. "Symmetry Reduction of Optimal Control Systems and Principal Connections". In: *SIAM Journal on Control and Optimization* 51.1 (2013), pp. 96–120.

[84] Peter J Olver. *Classical invariant theory*. Vol. 44. Cambridge University Press, 1999.

[85] Warren B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley-Interscience, 2007.

[86] Warren B. Powell. "Perspectives of approximate dynamic programming". In: *Annals of Operations Research* 241.1 (2016), pp. 319–356.

[87] F. Pukelsheim. *Optimal design of experiments*. Probability and mathematical statistics. Wiley, 1993.

[88] Friedrich Pukelsheim. "On c-optimal design measures". In: *Series Statistics* 12.1 (1981), pp. 13–20.

[89] Amirreza Rahmani et al. "Controllability of Multi-Agent Systems from a Graph-Theoretic Perspective". In: *SIAM Journal on Control and Optimization* 48.1 (2009), pp. 162–186.

[90] Witold Respondek and Issa Amadou Tall. "Nonlinearizable single-input control systems do not admit stationary symmetries". In: *Systems & Control Letters* 46.1 (2002), pp. 1–16.

[91] Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268.

[92] Ana Rufino Ferreira et al. "Symmetry Reduction for Performance Certification of Interconnected Systems". In: *IEEE Transactions on Control of Networked Systems* (2017).

[93] Alain Sarlette, Silvere Bonnabel, and Rodolphe Sepulchre. "Coordinated motion design on Lie groups". In: *IEEE Transactions on Automatic Control* 55.5 (2010), pp. 1047–1058.

[94] Marie A. Schroeder et al. "Hyperpolarized ^{13}C magnetic resonance reveals early- and late-onset changes to in vivo pyruvate metabolism in the failing heart". In: *European Journal of Heart Failure* 15.2 (2013), pp. 130–140.

[95] Marie A. Schroeder et al. "In vivo assessment of pyruvate dehydrogenase flux in the heart using hyperpolarized carbon-13 magnetic resonance". In: *Proceedings of the National Academy of Sciences* 105.33 (2008), pp. 12051–12056.

[96] Rolf F. Schulte et al. "Saturation-recovery metabolic-exchange rate imaging with hyperpolarized [1-^{13}C] pyruvate using spectral-spatial excitation". In: *Magnetic Resonance in Medicine* 69.5 (2013), pp. 1209–1216.

[97] Lise Vejby Søgaard et al. "In vivo measurement of apparent diffusion coefficients of hyperpolarized ^{13}C-labeled metabolites". In: *NMR in Biomedicine* 27.5 (2014), pp. 561–569.

[98] V. Solo. "Selection of regularisation parameters for total variation denoising". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 3. 1999, pp. 1653–1655.

[99] William C. Stanley, Fabio A. Recchia, and Gary D. Lopaschuk. "Myocardial Substrate Metabolism in the Normal and Failing Heart". In: *Physiological Reviews* 85.3 (2005), pp. 1093–1129.

[100] David Strong and Tony Chan. "Edge-preserving and scale-dependent properties of total variation regularization". In: *Inverse problems* 19.6 (2003), S165.

[101] Christine Leon Swisher et al. "Quantitative measurement of cancer metabolism using stimulated echo hyperpolarized carbon-13 MRS". In: *Magnetic Resonance in Medicine* 71.1 (2014), pp. 1–11.

[102] The MathWorks, Inc. *MATLAB and Optimization Toolbox Release 2013b*. Natick, Massachusetts, United States.

[103] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288.

[104] L. Vandenberghe. *Lecture 13: Douglas-Rachford method and ADMM*. 2016. URL: `https://web.archive.org/web/20170405001209/http://www.seas.ucla.edu/~vandenbe/236C/lectures/dr.pdf`.

[105] É. Walter and L. Pronzato. *Identification of parametric models from experimental data*. Communications and control engineering. Springer, 1997. ISBN: 9783540761198.

[106] Yan Xing et al. "Optimal variable flip angle schemes for dynamic acquisition of exchanging hyperpolarized substrates". In: *Journal of Magnetic Resonance* 234 (2013), pp. 75–81.

[107] Yinyu Ye. "Approximating quadratic programming with bound and quadratic constraints". In: *Mathematical Programming* 84.2 (1999), pp. 219–226.

[108] Shuzhong Zhang. "Quadratic maximization and semidefinite relaxation". In: *Mathematical Programming* 87.3 (2000), pp. 453–465.

[109] Lei Zhao et al. "Gradient-Echo Imaging Considerations for Hyperpolarized ^{129}Xe MR". In: *Journal of Magnetic Resonance, Series B* 113.2 (1996), pp. 179–183.

[110] Martin Zinkevich and Tucker Balch. "Symmetry in Markov decision processes and its implications for single agent and multi agent learning". In: *In Proceedings of the 18th International Conference on Machine Learning*. Morgan Kaufmann, 2001, pp. 632–640.